The Effects of Environmental Sound and Context on Listener Engagement with Speech-based Media

Jay Alexander Harrison

PhD

University of York

Physics, Engineering and Technology

November 2024

Abstract

In the last 20 years the advent and growth in popularity of portable digital audio players, smartphones, and podcasting have contributed to the rise in popularity of mobile speech-based media consumption using auditory headset devices. But little is known about what impact a listener's environmental context and the sounds contained within it have on their engagement with media.

This thesis investigates how listener engagement with speech-based media is affected by surrounding environmental context, the sound that emanates from it, and the mediating effects that auditory headset devices have on listeners' ability to perceive this sound.

The results of an initial study on podcast listening behaviours identified 5 factors of environmental context that are capable of both positively and negatively affecting engagement. The *Soundscape and at Work* factor suggested that some listeners experience increased engagement with podcasts when listening in the presence of moderate or high levels of human or mechanical sounds.

A study was then conducted to quantify the insertion loss and gain performance of three contemporary true wireless earbuds. The results found that all devices possessed a wide range of insertion loss and gain performance, with the most open hear-through modes significantly boosting external signals by as much as 15 dB, relative to open ear listening. The boosting was found to be centred around the frequencies considered most important to speech intelligibility.

Finally, a listening experiment was conducted to investigate how different levels of environmental noise and boosting from hear-through modes affected listeners' engagement with speech-based media consumed over earphones. The results showed that while increased exposure to noise resulted in lower engagement scores for most participants, the opposite was true for a small proportion of participants. The result suggest that adjustable hear-through in earbuds can provide listeners with an increased awareness of their surroundings, without compromising their engagement.

Contents

Abstract **2**

Contents **3**

List of Figures **6**

List of Tables **12**

Acknowledgements **14**

Declaration of Authorship **15**

1 Introduction **16**
1.1 Statement of Hypothesis 19
1.2 Aims and Objectives 20
1.3 Thesis Structure 22
1.4 Contributions to the Field 23

2 Fundamentals of Sound, Human Hearing, and Environmental Noise **25**
2.1 Sound in the Free Field 25
2.1.1 Basic Properties of Sound 26
2.2 Human Hearing 34
2.2.1 Anatomy of the Ear 34
2.2.2 Spatial Hearing 35
2.3 Environmental Noise and Soundscape 39
2.3.1 Environmental Noise 39
2.3.2 Soundscape 41
2.4 Summary 42

3 Related Work **44**
3.1 Auditory Headsets and Mobile Audio Consumption 44
3.1.1 Auditory Headsets used in the Mobile Consumption of Audio 44
3.1.2 Active External Sound Modification Functionality in Auditory Headsets . . . 56
3.1.2.1 Speech Intelligibility Metrics 70

3.2 Listener Engagement and Mobile Audio Consumption . . . 71
3.2.1 A Historical Overview of Mobile Audio Consumption . . . 72
3.2.2 Audio-based Media Listened to in the Mobile Consumption of Audio . . . 81
3.2.3 Aspects and Assessment of Listener Engagement with Spoken Narratives . . . 83
3.2.4 Listener Engagement in Mobile Audio Consumption Experiences . . . 89
3.3 Summary . . . 97

4 Environmental Context and Attentional Engagement in Podcast Listening Experiences 100
4.1 Materials and methods . . . 102
4.1.1 Participants . . . 102
4.1.2 Measures . . . 102
4.1.3 Procedure . . . 104
4.2 Results . . . 106
4.2.1 Factors of environmental context . . . 106
4.2.2 Relationships between aspects of podcast listening engagement . . . 109
4.2.3 Relationships between environmental context factor scores and aspects of podcast listening engagement . . . 120
4.3 Discussion . . . 123
4.3.1 Hypothesis 1: Factors of environmental context . . . 123
4.3.2 Hypothesis 2: Associations amongst aspects of podcast listening engagement 129
4.3.3 Hypothesis 3: Associations amongst environmental context factor scores and aspects of podcast listening engagement . . . 130
4.3.4 Implications for related research fields . . . 131
4.4 Conclusion . . . 136

5 Measurement of Insertion Loss and Gain in True Wireless Earbuds 138
5.1 Methods . . . 139
5.1.1 Inclusion criteria for devices tested . . . 139
5.1.2 Device modes tested . . . 140
5.1.3 Acoustic test fixture . . . 141
5.2 Spectral modification of broadband noise . . . 143
5.2.1 Spectral modification of broadband noise method . . . 143
5.2.2 Spectral modification of broadband noise results . . . 145
5.2.3 Presentation level linearity . . . 155
5.2.4 Time domain performance . . . 162
5.3 Discussion . . . 164
5.4 Conclusion . . . 168

6 The Effects of Hear-through Augmented Environmental Noise on Listener Engagement with Speech-based Media 169
6.0.1 Research Aims . . . 170
6.1 Methods and Materials . . . 171
6.1.1 Experiment Design . . . 171
6.1.2 Story Stimuli . . . 171

6.1.3 Noise Conditions . . . 172
6.1.4 Hear-through Simulation Noise Condition . . . 173
6.1.5 Noise Condition Soundscape Material . . . 176
6.1.6 Apparatus and Level Calibration . . . 177
6.1.7 Measures . . . 178
6.1.8 Participants . . . 180
6.1.9 Experiment Procedure . . . 180
6.2 Data Analysis and Results . . . 181
6.2.1 Immersion . . . 182
6.2.2 Narrative Absorption . . . 182
6.2.3 Enjoyment . . . 184
6.2.4 Listening Effort . . . 184
6.2.5 Exit Questionnaire: Awareness of Noise . . . 185
6.3 Discussion . . . 186
6.4 Conclusion . . . 188

7 Conclusion **189**
7.1 Summary . . . 189
7.2 Contributions to the Field . . . 192
7.3 Restatement of Hypothesis . . . 194
7.4 Future Work . . . 195
7.5 Final Remarks . . . 197

A Supplementary Material for Chapter 4 **199**

B Supplementary Material for Chapter 6 **236**

Bibliography **259**

List of Figures

2.1 The mass-spring model of sound propagation, after [1]. 26
2.2 The mass-spring model showing the propagation of a sound pulse through a material, after [1]. 27
2.3 The propagation of sine wave through a material represented by the mass-spring model, after [1]. 29
2.4 *Diagram illustrating how the intensity of a sound radiating from a point source in free field conditions is inversely proportional to the square of the distance from the source, after [2].* 32
2.5 *Diagram illustrating the anatomy of the human ear, after [3] and [2].* 35
2.6 *Diagram illustrating the head-related coordinate system and other key concepts relating to spatial hearing, adapted from [4] and [5].* 36
2.7 *Diagram showing a model of Interaural Time Difference (ITD) illustrating how the incidence angle of a sound source and the scale of the head results in a difference in the paths the sound will take to arrive at either ear, after [1].* 37
2.8 *Diagram showing how the occlusion from the head results in acoustic shadowing at higher frequencies, after [6].* 38

3.2 *Diagrams showing examples of open and closed supra-aural headphones on the pinna of an ATF, after [7].* 47
3.3 *Diagram showing an example of an open supra-concha headphone on the concha of an ATF, after [7].* 47
3.4 *Diagrams showing examples of open and closed intra-concha earphones in the concha of an ATF, after [7].* 48
3.5 *Diagrams showing examples of open and closed insert earphones in the ear canal of an ATF, after [7].* 49
3.6 *Diagram showing an example of a typical open extra-aural headphone suspended adjacent to the pinna of an ATF, inspired by [7].* 50
3.7 *Photograph showing how the AKG K1000 extra-aural headphones would be worn by a user [8].* 51
3.8 *Photograph showing how the Bose Frames are worn by a user, adapted from [9].* . . . 52
3.9 *Diagram showing how sound is received at the auditory nerve via bone conduction and air conduction, after [3].* 54
3.10 *Diagram showing a generic hear-through equalisation block diagram that illustrates how hear-through works in an in-ear earphone device, after [10] and [7].* 58
3.11 *An advert for Brandes radio headsets from 1925 [11].* 73

3.12 *A photograph of the Regency EP-1 earphone for use with the Regency TR-1 transistor radio, adapted from [12].* . . . 74
3.13 *Timeline of key technological developments in devices used for the mobile consumption of audio. Source images adapted from [13–19].* . . . 75
3.14 *A photograph of the Sony MDR 3L2 headphones for use with the Sony Walkman TPS-L2 cassette player [20].* . . . 76
3.15 *A photograph of the Apple earphones for use with the Apple iPod (1st generation) digital audio player [21].* . . . 78
3.16 *A photograph of the Apple AirPods (1st generation) earphones for use with the Apple iPhone smartphones and other compatible devices [22].* . . . 80
3.17 *A photograph of the Apple AirPods Pro (1st generation) earphones for use with the Apple iPhone smartphones and other compatible devices [23].* . . . 80

4.1 *Attentional engagement as a function of environmental context in podcast listening experiences* scale with the *factors of environmental context* produced by the EFA. The bottom 9 items are fainter to highlight the fact that they were removed during the EFA process. . . . 105

5.1 *Photographs of the devices inserted in KEMAR. From left to right: Apple Airpods Pro (2nd gen.), Bose QuietComfort Earbuds II, and Sony LinkBuds S.* . . . 140
5.2 *Diagram illustrating the spectral modification properties analysis terminology. The line plotted on this graph is not representative of any real data.* . . . 146
5.3 *Diagram illustrating the spectral modification attenuation and boosting region properties in cases where the modification is indicative of a shelf rather than a band.* . . . 147
5.4 *Passive spectral modification of diffuse pink noise at 80 dB SPL for Off modes for all devices.* . . . 148
5.5 *Total spectral modification of diffuse pink noise at 80 dB SPL for noise cancellation and hear-through modes for all devices.* . . . 151
5.6 *Active spectral modification of diffuse pink noise at 80 dB SPL for noise cancellation and hear-through modes for all devices.* . . . 152
5.7 *Passive spectral modification of diffuse pink noise at presentation levels of 65 to 105 dB SPL for Off modes for all devices.* . . . 156
5.8 *Total spectral modification of diffuse pink noise at presentation levels of 65 to 105 dB SPL for Noise cancellation modes for all devices.* . . . 158
5.9 *Spectrograms showing temporal discontinuities in the performance of devices A and B at a presentation level of 105 dB SPL. Discontinuities are visible approximately every 6-8 s in the left ear of Device A's HTM 1response and from 1-5 s in the equivalent right ear response. For Device B's NCM, discontinuities are visible as vertical lines below 5 kHz throughout the duration of the recording in the left ear response and instantaneous shifts in the frequency spectrum at approximately 1-2, 5-6, and 13-19 s in the right ear response.* . . . 159
5.10 *Total spectral modification of diffuse pink noise at presentation levels of 65 to 105 dB SPL for Hear-through 1 modes for all devices.* . . . 160

5.11 *Spectrogram showing temporal discontinuities in the spectral modification of 105 dB SPL diffuse pink noise for the Device A's HTM 1. Discontinuities are visible approximately every 6-8 s.* 161
5.12 *Spectrograms showing the onset of spectral modification of diffuse pink noise for Device A's HTM 5, Device B's HTM 1, and Device C's HTM 1. For Device A the onset is visible between 2-3 s and most prominent in the frequencies either side of the 2-4 kHz region, while the onset is instantaneous for devices B and C.* 163

6.1 Graphical summary of the noise conditions; *No Noise (NN)*, *Low Noise (LN)*, *High Noise (HN)*, and *Hear-through (HT)*. The *HN & HT* conditions had equal binaural loudness. 173
6.2 Graph showing the transfer functions of noise conditions and the target measurement for the *Hear-through (HT)* condition produced using diffuse broadband pink noise. . 174
6.3 Block diagram illustrating the processing chain that was used to create the *Hear-through (HT)* condition in the experiment. 175
6.4 DPA binaural microphone mounted on the Shure SE215 earphones, used for calibration and production of the hear-through simulation in the *hear-through* noise condition. . 176
6.5 A participant seated in the listening position. 177
6.6 *The 15-cm printed graphical line scale that participants used to provide their immersion ratings. Participants were provided with a separate scale printed on an A4 sheet of paper for each immersion rating. The A4 sheets were removed from the participant's view before the presentation of the next stimulus.* 179
6.7 Estimated marginal means of Immersion, Narrative Absorption, and Enjoyment scores for all stimuli at all noise conditions. Error bars represent 95% confidence intervals. . 183
6.8 Bootstrapped cell means (BCM) of Listening Effort scores for all stimuli at all noise conditions. Error bars represent 95% confidence intervals. 185
6.9 Participants who were aware of other sounds and how their awareness influenced their Immersion. *Number of participants who reported an awareness of other sound(s) **Other sounds were any sounds aside from the storytellers or audience reactions. . . 186

A.1 Page 1 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4. 200
A.2 Page 2 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4. 201
A.3 Page 3 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4. 202
A.4 Page 4 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4. 203
A.5 Page 5 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4. 204
A.6 Page 6 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4. 205
A.7 Page 7 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4. 206

A.8 Page 8 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4. 207
A.9 Page 9 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4. 208
A.10 Page 10 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4. 209
A.11 Page 11 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4. 210
A.12 Page 12 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4. 211
A.13 Page 13 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4. 212
A.14 Page 14 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4. 213
A.15 Page 15 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4. 214
A.16 Page 16 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4. 215
A.17 Page 17 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4. 216
A.18 Page 18 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4. 217
A.19 Page 19 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4. 218
A.20 Page 20 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4. 219
A.21 Page 21 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4. 220
A.22 Page 22 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4. 221
A.23 Page 23 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4. 222
A.24 Page 24 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4. 223
A.25 Page 25 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4. 224
A.26 Page 26 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4. 225
A.27 Page 27 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4. 226
A.28 Page 1 of 9 of the ethical approval form (Harrison101121) for the study in Chapter 4, approved by the University of York Electronic Engineering department ethics committee. 227
A.29 Page 2 of 9 of the ethical approval form (Harrison101121) for the study in Chapter 4, approved by the University of York Electronic Engineering department ethics committee. 228

A.30 Page 3 of 9 of the ethical approval form (Harrison101121) for the study in Chapter 4, approved by the University of York Electronic Engineering department ethics committee. 229
A.31 Page 4 of 9 of the ethical approval form (Harrison101121) for the study in Chapter 4, approved by the University of York Electronic Engineering department ethics committee. 230
A.32 Page 5 of 9 of the ethical approval form (Harrison101121) for the study in Chapter 4, approved by the University of York Electronic Engineering department ethics committee. 231
A.33 Page 6 of 9 of the ethical approval form (Harrison101121) for the study in Chapter 4, approved by the University of York Electronic Engineering department ethics committee. 232
A.34 Page 7 of 9 of the ethical approval form (Harrison101121) for the study in Chapter 4, approved by the University of York Electronic Engineering department ethics committee. 233
A.35 Page 8 of 9 of the ethical approval form (Harrison101121) for the study in Chapter 4, approved by the University of York Electronic Engineering department ethics committee. 234
A.36 Page 9 of 9 of the ethical approval form (Harrison101121) for the study in Chapter 4, approved by the University of York Electronic Engineering department ethics committee. 235

B.1 Page 1 of 12 of the ethical approval form (Harrison23112023) for the study in Chapter 6, approved by the University of York School of Physics, Engineering and Technology ethics committee. 237
B.2 Page 2 of 12 of the ethical approval form (Harrison23112023) for the study in Chapter 6, approved by the University of York School of Physics, Engineering and Technology ethics committee. 238
B.3 Page 3 of 12 of the ethical approval form (Harrison23112023) for the study in Chapter 6, approved by the University of York School of Physics, Engineering and Technology ethics committee. 239
B.4 Page 4 of 12 of the ethical approval form (Harrison23112023) for the study in Chapter 6, approved by the University of York School of Physics, Engineering and Technology ethics committee. 240
B.5 Page 5 of 12 of the ethical approval form (Harrison23112023) for the study in Chapter 6, approved by the University of York School of Physics, Engineering and Technology ethics committee. 241
B.6 Page 6 of 12 of the ethical approval form (Harrison23112023) for the study in Chapter 6, approved by the University of York School of Physics, Engineering and Technology ethics committee. 242
B.7 Page 7 of 12 of the ethical approval form (Harrison23112023) for the study in Chapter 6, approved by the University of York School of Physics, Engineering and Technology ethics committee. 243
B.8 Page 8 of 12 of the ethical approval form (Harrison23112023) for the study in Chapter 6, approved by the University of York School of Physics, Engineering and Technology ethics committee. 244
B.9 Page 9 of 12 of the ethical approval form (Harrison23112023) for the study in Chapter 6, approved by the University of York School of Physics, Engineering and Technology ethics committee. 245
B.10 Page 10 of 12 of the ethical approval form (Harrison23112023) for the study in Chapter 6, approved by the University of York School of Physics, Engineering and Technology ethics committee. 246

B.11 Page 11 of 12 of the ethical approval form (Harrison23112023) for the study in Chapter 6, approved by the University of York School of Physics, Engineering and Technology ethics committee. 247
B.12 Page 12 of 12 of the ethical approval form (Harrison23112023) for the study in Chapter 6, approved by the University of York School of Physics, Engineering and Technology ethics committee. 248
B.13 Page 1 of 4 of the study information sheet provided to participants in the experiment described in Chapter 6. 252
B.14 Page 2 of 4 of the study information sheet provided to participants in the experiment described in Chapter 6. 253
B.15 Page 3 of 4 of the study information sheet provided to participants in the experiment described in Chapter 6. 254
B.16 Page 4 of 4 of the study information sheet provided to participants in the experiment described in Chapter 6. 255
B.17 Page 1 of 2 of the experiment screening and consent google form that participants were required to complete before taking part in the experiment described in Chapter 6. 256
B.18 Page 2 of 2 of the experiment screening and consent google form that participants were required to complete before taking part in the experiment described in Chapter 6. 257
B.19 The four distractor tasks participants completed between stimuli in the experiment described in Chapter 6. Participants spent 4 minutes on each distractor task and were informed that their assessment in the tasks was not being assessed. Top left: "Matchsticks ˜ Matches Puzzle" Participants were handed an ipad with level 1 from the second pack of puzzles loaded on the screen and asked to complete as many puzzles as they were able in the time allowed. Top right: "Memorama • Classic" Participants were handed an ipad with level 15 loaded on the screen and asked to complete as many levels as they were able in the time allowed. Bottom left: "LEGO Set 41775-1 Rainbow Unikitty" Participants were handed a wooden tray with a box containing the disassembled parts of the LEGO set. They were shown a printed image of the completed set for 5 seconds and asked to try and build the set from memory in the time allowed. They were told to alert the researcher if they thought they had completed the set, who would then verify if they had built it correctly. Bottom right: "Spot the Difference" Participants were handed a blank A4 notebook and a A4 printout of the creative commons spot the difference image from: https://flic.kr/p/dVQwna. They were asked to find all of the differences between the two images and write down their answers on the notepad. They were told to alert the researcher if they thought they had found all of the differences. 258

List of Tables

2.1 Some common environmental sounds and their associated SPLs, after [24]. 34

4.1 Factor loadings for the EFA with direct quartimin rotation of the retained "environment context attentional engagement podcast listening" items. Analysis conducted using MIFA dataset (N = 264). 107
4.2 Pearson's product-moment correlations amongst the five factors produced by the EFA (N = 264). 109
4.3 Amount of time participants spend listening to podcasts using a smartphone on average weekdays and weekend days (N = 264). 110
4.4 Locations in which participants have consumed (N = 264) and most often consume (N = 264) podcasts using a smartphone. 111
4.5 Monitoring devices participants have used (N = 264) and most often use (N = 264) to consume podcasts with a smartphone. 113
4.6 Activities participants have engaged in (N = 264) and most often engage in (N = 238) while listening to podcasts with a smartphone. 115
4.7 Proportion of time participants spend multitasking while listening to podcasts using a smartphone on average weekdays and weekend days. (N = 264) 116
4.8 Methods participants have used (N = 264) and most often use (N = 262) to discover podcasts. 117
4.9 Spearman's rank-order correlations amongst the *aspects of podcast listening engagement*. (N = 264) 119
4.10 Pearson's product-moment correlations amongst the total monitoring devices, total multitasking activities, total listening locations, and total discovery methods. (N = 264) 119
4.11 Spearman's rank-order correlations amongst the *aspects of podcast listening engagement* & *environmental context factor scores* (N = 264). 121
4.12 Pearson's product-moment correlations amongst the total monitoring devices, total multitasking activities, total listening locations, total discovery methods, and *environmental context factor scores* (N = 264). 122

5.1 The modes tested in each device and study nomenclature used to describe equivalent modes. 141
5.2 Passive spectral modification properties analysis of results shown in Fig. 5.4. 149
5.3 Total spectral modification properties analysis of results shown in Fig. 5.5. 150
5.4 Active spectral modification properties analysis of results shown in Fig. 5.6. 155

6.1 Items of the Narrative Absorption Scale (NAS; Capturing Four Dimensions), Enjoyment, and Listening Effort. 179

Acknowledgements

My journey towards the completion of this thesis has been a long and yet incredibly rewarding one. Along the way I have had the pleasure and privilege of meeting so many kind and generous people, who have all played their part in making my PhD experience the rich tapestry that I now reflect upon it as having been. Although sadly it is not possible to mention all of these people here individually, I would like to highlight several people who have played an especially important part in this.

Firstly, I would like to thank Professor Damian Murphy for the opportunity to undertake this PhD and his subsequent supervision throughout the duration of the project. His steadfast guidance, support, and relentless optimism continually pushed me to give my very best, and has allowed me to achieve things and develop in ways I wouldn't have thought possible when I began. I will be forever grateful for the impact this has had on my life.

I would like to thank Dr. Alan Archer-Boyd for his guidance as my industry supervisor over the last two and a half years. His kindness, warmth, and wisdom has been a constant source of motivation and inspiration. I would also like thank Dr. Chris Baume, Dr. Jon Francombe, and Dr. Chris Pike for their valuable and insightful guidance as industry supervisors in the earlier stages of my PhD, in addition to all those at BBC R&D who made me feel so welcome. Thanks also to Dr. David Halliday and Dr. Gavin Kearney for the sage guidance they provided in their respective roles as my thesis advisor and progression chair. Thanks must also be extended to BBC R&D, the UK's Arts and Humanities Research Council, XR Stories, and the University of York for funding the project.

To all those at the AudioLab who have contributed to making it the wonderful, creative, and, quite frankly, daft crucible that I have known and loved it to be, thank you. Special mentions to Frank and Andrew, who first welcomed me to the lunch table in 2021 and have been a reliable source of MATLAB advocacy, bleak film recommendations, and "humour", ever since. Special thanks also to David and Joe for going through it all with me and making it look pretty easy by comparison.

Thank you to my friends and family, especially Jean, whose love and support has meant so much.

And finally, to Aoife, thank you. Thank you for always being there for me, for planting the life in winter, for making me dream wild, and for being so irrepressibly you. None of this would have been possible without your love, care, and unrelenting support.

Declaration of Authorship

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as references. I also declare that parts of this research have been presented in previous conference and journal publications, which are listed as follows:

- J. Harrison, A.W. Archer-Boyd, J. Francombe, C. Pike and D.T. Murphy, "The relationship between environmental context and attentional engagement in podcast listening experiences", *Frontiers in Psychology*, vol. 13, 2023, doi: 10.3389/fpsyg.2022.1074320
- J. Harrison, A.W. Archer-Boyd and D.T. Murphy, "The effects of hear-through augmented environmental noise on listener engagement with speech-based media", in *AES 5th International Conference on Audio for Virtual and Augmented Reality*, Redmond, USA, 2024.
- J. Harrison, A.W. Archer-Boyd and D.T. Murphy, "Measurement of insertion loss and gain in variable hear-through and noise cancellation modes of true wireless earbuds", *Acta Acustica*, vol. 9, no. 28, 2025, doi: 10.1051/aacus/2025013

Chapter 1

Introduction

In 1979 the release of the very first personal stereo (the Sony Walkman TPS-L2) sparked a revolution in how listeners consumed music and other audio-based media. Before this individuals were able to engage in the mobile consumption of linear music and speech programming using portable transistor radios, but the original Walkman was the first device to allow listeners to soundtrack their movements through the world with carefully curated media choices.

The personal stereo was in many ways the first device that allowed immersion and an augmentation of reality using a portable device. While this practice of sound tracking a walk to the shops or commute to work with a favourite piece of music is now a popular activity for many and an almost guaranteed sight on pavements, high streets, and public transport, initially attitudes towards the dramatic increase in mobile media consumption were mixed. Some considered Walkman use in public such a departure from accepted societal norms that users were often considered rude or antisocial [25].

In fact, the designers of the Walkman were themselves so concerned that listeners would be isolated from their surroundings that they included two headphone jacks so that music could be shared, and a "hotline" button, which lowered the volume of the media and activated a built-in microphone, making it easier for users to communicate with others [26]. However, as the popularity of the Walkman continued to grow these concerns would largely fade away, with the "hotline" button not featuring in any future Walkman model due to its lack of popularity.

Fast forward to today, and the technological evolution of mobile listening devices such as the smartphone, combined with rise in media consumption using streaming platforms, has cemented the place of mobile devices as one of the most popular means by which music and other audio-based media are consumed [27].

Furthermore, where previously music was the dominant form of media most closely associated with mobile media consumption, the advent and rise of podcasts [28], audio books, and on-demand radio has resulted in an explosion in the popularity of speech-based media. This trend can now also be seen in the proportion of different media consumed in mobile listening experiences, with data from a 2023 Radio Joint Audience Research (RAJAR) consumer listening survey suggested that speech-based media consumption using mobile devices is now at least on par with music consumption [27].

There has been a long history of scholarly work dating back to the release of the original Walkman analysing how listening to audio over headphones in public spaces shapes how individuals perceive and interact with their surroundings. Initially, the term "the Walkman effect" was coined to describe the control that the new phenomenon of mobile headphone listening gave to individuals [29], allowing them to move through the world without interruption and disengaged from others in their surrounding environment. Other scholars spoke enthusiastically of the theatrical listening affordances that mobile audio consumption made possible through augmentation of curated audio over listeners' experiences of the real world [30].

As the popularity of mobile media listening grew and the technology used therein also began to evolve, scholars continued to expand their study of what the practice meant for individuals and their experience of the world [31–35]. A consistent focus within this work is the notion of mobile audio listeners inhabiting an "auditory bubble", with various positive and negative symptoms for the listener attributed to this bubble by the literature.

Despite this there has been very little work conducted to investigate how environmental sound and other aspects of non-auditory environmental, situational, and social context that a mobile audio consumer experiences, relates to the engagement that they experience with the media itself. It is also the case that the vast majority of literature concerning all aspects of this practice of mobile media consumption has focused on music. However, research strongly suggests that findings from mobile music consumption research are not directly transferable to mobile speech-based media.

A further key factor in the question of how environmental sound affects listener engagement in mobile listening experiences is how the listener's aural perception of external sound in their environment is mediated by the headphones or alternate auditory headset device that is being using to listen to the media. For example, the original Walkman came packaged with a small pair of supra-concha headphones that rested against the ears of the listener. This would provide a markedly different perception of external environmental sound compared to the occlusive effects of a pair of contemporary intra-concha earphones worn inside the ears of a listener.

The advent of noise cancellation and adjustable hear-through features in contemporary headphones and earbuds are other compounding factors that can affect listeners aural perception of external sound. While hear-through functionality might at first glance appear to be a decidedly modern technological development, its conceptual origins could arguably be traced back to the original Walkman and the aforementioned "hotline" button feature, which aimed to address similar issues, albeit in a more rudimentary way. Contemporary research has shown that some hear-through users seek full acoustic transparency, where perception of external sounds is ideally identical to open ear listening [36], whereas others desire an enhanced awareness of external sounds where parts of the frequency spectrum are boosted relative to the open ear [37]. Of course it follows that these factors could also potentially have an effect on a mobile audio consumer's engagement with the media itself, however, this has yet to be considered or investigated at all within the literature.

The research presented in this thesis makes contributions that aim to address these gaps in the literature around the mobile consumption of speech-based media and how environmental context, sound, and consumer headphone technology all play a part in influencing listener engagement with the media itself. The work presented first concerns a large scale consumer survey on podcast listening behaviours that quantifies aspects of podcast listening and factors of environmental context that influence attentional engagement. An acoustic measurement study then examines how wireless earbud devices and features mediate a listener's ability to perceive their acoustic surroundings. Finally, a controlled listening experiment investigates how environmental noise affects listener engagement with speech-based media, and how this relationship is further mediated by the boosting of external sound provided by the adjustable hear-through feature in wireless earbuds.

1.1 Statement of Hypothesis

The body of work presented in this thesis spans a wide array of research topics and fields, focused mainly around listener engagement with speech-based media and environmental noise. Therefore the following overall hypothesis serves as a guiding tenet for the work contained in the thesis and the associated aims and objectives:

> *The quality of engagement that listeners' experience when consuming speech-based media using auditory headsets is mediated by elements of their surrounding environmental context, the environmental noise that emanates from within it, and the nature of occlusion provided by the auditory headset.*

Within the context of the present thesis the key elements of this hypothesis are defined as follows:

- **Quality of engagement:** This is intended as a collective term used to describe any and all examples of qualitative psychological responses that a listener might experience when consuming audio-based media. This could include, but is not limited to, attentional engagement, immersion, narrative absorption, and enjoyment. This term should also be understood as representing the full range of extent of engagement that a listener might experience, for example they might experience a strong sense of immersion when consuming one piece of media, but little to no immersion in the consumption of another.
- **Speech-based media:** This is intended to refer to any form of media that could be described as being predominately or completely comprised of spoken word content. This could include podcasts, audio books, live radio, and on-demand radio.
- **Auditory headsets:** This is used to describe any device that is worn on a users head that allows them to listen to recorded or broadcast media via means of air or bone conduction.
- **Environmental context:** This is used to describe an individuals' immediate physical surroundings, any objects, individuals, or stimuli contained within it, and relevant physical quantities such as time, temperature, light, and weather conditions. In this sense it should be understood

that environmental context encompasses a specific set of physical parameters that are distinct from other primary categories such as situational and social context.

- **Environmental noise:** This should be understood as being representative of any sound in a listener's surroundings that has the potential to capture their attention, regardless of the psychological response that the sound illicits from the listener. In this sense the term environmental noise should be considered interchangeable with the term *environmental sound*, within the context of this thesis.

- **Nature of occlusion:** This term is intended to describe any and all mediating effects that an auditory headset might have on a listener's perception of an external environmental noise signal as received at their eardrums. This could include the passive noise reduction provided by the occlusion of a pair of traditional over-ear headphones, the mixture of passive occlusion and active boosting or a pair of earbuds in transparency mode, or even the complete passive acoustic transparency provided by a bone conduction headset.

1.2 Aims and Objectives

The following series of aims and objectives were formulated in response to the hypothesis outlined in the previous section, these are as follows:

A_1 Investigate how different forms of environmental context relate to listeners' engagement with speech-based media.

(a) Design a survey instrument to capture quantitative and qualitative data responses relating to listener engagement and behaviours in podcast listening experiences.

(b) Advertise and distribute this survey to a large number of eligible participants using an online survey platform, newsletters, and social media.

(c) Analyse the survey data using appropriate statistical methods to identify factors of environmental context that influence listener engagement and quantify various aspects of podcast listening behaviour.

A_2 Quantify how auditory headset devices used in the consumption of speech-based media mediate the extent to which listeners are able to perceive environmental noise in their surroundings.

(a) Record a series of acoustic measurements capturing the transfer functions of various passive and active noise cancellation and hear-through modes of auditory headset devices.

(b) Analyse the measurement recordings to quantify the insertion loss and gain of external sound provided by each device mode measured.

A_3 Explore the relationship between environmental noise and listener engagement with speech-based media.

(a) Design and pilot a controlled listening experiment using appropriate experimental methodologies to assess how environmental noise mediates listeners' engagement with speech-based media.

(b) Advertise the experiment to recruit and collect data from a large corpus of participants.

(c) Analyse the results to establish the relationship between listener engagement and exposure to noise.

A_4 Examine how hear-through modes in auditory headset devices mediate the relationship between environmental noise and listener engagement with speech-based media.

(a) Ensure that the design of the above mentioned listening experiment also permits investigation of how the modification of noise provided by hear-through modes further mediate the relationship between listener engagement and exposure to noise.

(b) Conduct an additional analysis of experiment results focusing on the hear-through condition(s) results relative to the other noise conditions.

Successful completion of each of the objectives will meet the stated aims, and therefore a rigorous investigation of the hypothesis.

1.3 Thesis Structure

To contextualise the original work presented in the thesis, and to make it accessible to all readers, Chapters 2 and 3 first present an overview of fundamental theory and a review of literature and contemporary research across relevant fields.

Chapter 2 first provides a review of fundamental principles of sound, human hearing and environmental noise, laying the theoretical foundations for the subsequent related work presented in Chapter 3. This related work covers an array of interdisciplinary research topics branching out from the central themes of mobile audio consumption, auditory headsets, and listener engagement with speech-based media.

The findings from this review of literature are then used to inform the podcast listening behaviours and environmental context survey work presented in Chapter 4. This chapter presents details of the study methodology, survey design, and analysis of results used to quantify aspects of podcast listening engagement and identify factors of environmental context that influence attentional engagement.

Chapter 5 then details a series of acoustic measurements that are conducted to quantify the insertion loss and gain performance of true wireless earbuds that feature both noise cancellation and adjustable hear-through modes. The results from these measurements are used to inform the design of the listening experiment presented in the following chapter.

Chapter 6 then presents an overview of the study methodology and the results of this listening experiment. The experiment utilises a mixed design with a between-subjects factor of noise condition and a within-subjects factor of stimuli to assess how environmental noise and boosting from hear-through modes affect listener engagement with speech-based media.

Chapter 7 concludes the thesis with a summary of the work presented, a restatement and review of the hypothesis with respect to the findings of the original work. The contributions that this work has made to the field are also presented, in addition to a summary of potential future avenues of research that lead on from the original work in this thesis. The thesis is concluded with some final remarks from the author.

1.4 Contributions to the Field

The work contained in this thesis makes several novel contributions to the fields of applied psychology, mobile audio consumption, augmented reality, headphone technology, and media studies. These contributions are introduced here in the first instance and again as they are expanded upon at the conclusion of the thesis in Chapter 7.

- The study presented in Chapter 4 is the first of its kind to quantitatively investigate the relationship between listeners' environmental context and their attentional engagement in podcast listening experiences.
- The results presented in Chapter 4 are the first to support the hypothesis that a listener's environmental context exerts influence over their attentional engagement in podcast listening experiences.
- The study presented in Chapter 5 is the first of its kind to measure the insertion loss and gain performance of noise cancellation and adjustable hear-through modes in commercially available devices that feature both modes of functionality.
- The results presented in Chapter 5 provide the novel finding that none of the measured true wireless earbuds appeared to be attempting to achieve complete acoustic transparency in their hear-through mode functionality.
- The study presented in Chapter 6 is the first of its kind to perform controlled listening experiments investigating the effects of both passively occluded and hear-through boosted environmental noise on listener engagement with speech-based media, measured across Immersion, Narrative Engagement, and Enjoyment.
- The results presented in Chapter 6 are the first to show that for most listeners, in controlled listening experiments, engagement with speech-based media decreases and Listening Effort increases, as a result of exposure to passively occluded environmental noise.

- Further results presented in Chapter 6 also provide the novel finding that listeners' appear to respond differently to environmental noise that is boosted by hear-through, compared to noise that is perceived only through the passive occlusion of earphones.

The findings listed above are presented in detail throughout this thesis and within the following publications:

- J. Harrison, A.W. Archer-Boyd, J. Francombe, C. Pike and D.T. Murphy, "The relationship between environmental context and attentional engagement in podcast listening experiences", *Frontiers in Psychology*, vol. 13, 2023, doi: 10.3389/fpsyg.2022.1074320
- J. Harrison, A.W. Archer-Boyd and D.T. Murphy, "The effects of hear-through augmented environmental noise on listener engagement with speech-based media", in *AES 5th International Conference on Audio for Virtual and Augmented Reality*, Redmond, USA, 2024.
- J. Harrison, A.W. Archer-Boyd and D.T. Murphy, "Measurement of insertion loss and gain in variable hear-through and noise cancellation modes of true wireless earbuds", *Acta Acustica*, vol. 9, no. 28, 2025, doi: 10.1051/aacus/2025013

Chapter 2

Fundamentals of Sound, Human Hearing, and Environmental Noise

The following chapter will provide a brief overview of the key concepts relating to sound, human hearing, and environmental noise that are required to understand the research presented in the following chapters.

2.1 Sound in the Free Field

Sound can be thought of both as a sensation, perceived by the human auditory system, and as a stimulus, acting as a wave propagating through an elastic medium in a physical system. These two perspectives on sound represent two distinctly different ways in which sound can by studied. The first relates to psychoacoustics and all of the ways in which sound is perceived by a listener, whereas the second involves the consideration of sound as an objective and observable phenomenon through the lens of the field of physics [2].

2.1.1 Basic Properties of Sound

This section first introduces the fundamental properties of sound as a physical phenomenon. An explanation of how sound waves travel through air is provided, followed by an illustration of the relationship that exists between frequency and wavelength. The section concludes with a discussion of sound intensity, sound pressure level (SPL), and a summary of typical SPLs of common sounds heard in public spaces.

Propagation of Sound Waves in Air

Sound travels as a longitudinal wave phenomenon in air, that is to say the air particles move in parallel to the direction that the sound is propagating [2]. This type of wave propagation can be represented by the mass-spring model. Figure 2.1 shows a simplified one dimensional form of the mass-spring model, where air molecules are represented by a series of masses and intermolecular forces are represented by an interconnecting spring between two masses [1].

FIGURE 2.1: The mass-spring model of sound propagation, after [1].

A wave is introduced to the particles represented by the mass-spring model if a force is applied to one of the masses. To illustrate this, if a force caused the left most mass shown in Figure 2.1 to be pushed towards the other masses this would cause the spring between the two masses to become temporarily compressed. At some point the spring will reach a state of maximal compression before then beginning to stretch out (or rarefy), eventually returning the first mass back to its original position. This rarefaction of the first spring also causes a force to be applied to the next mass along. This pattern of compression and rarefaction continues to propagate through the subsequent masses and springs. An illustration of this pattern of compression and rarefaction moving through the same material as a result of an excitation from a single pulse is presented in Figure 2.2. This also illustrates how it is the actually the movement of a pressure differential between molecules, and not movement of the

molecules themselves, that facilitates the propagation of a sound wave, with compression representing regions of high pressure and rarefaction representing low pressure.

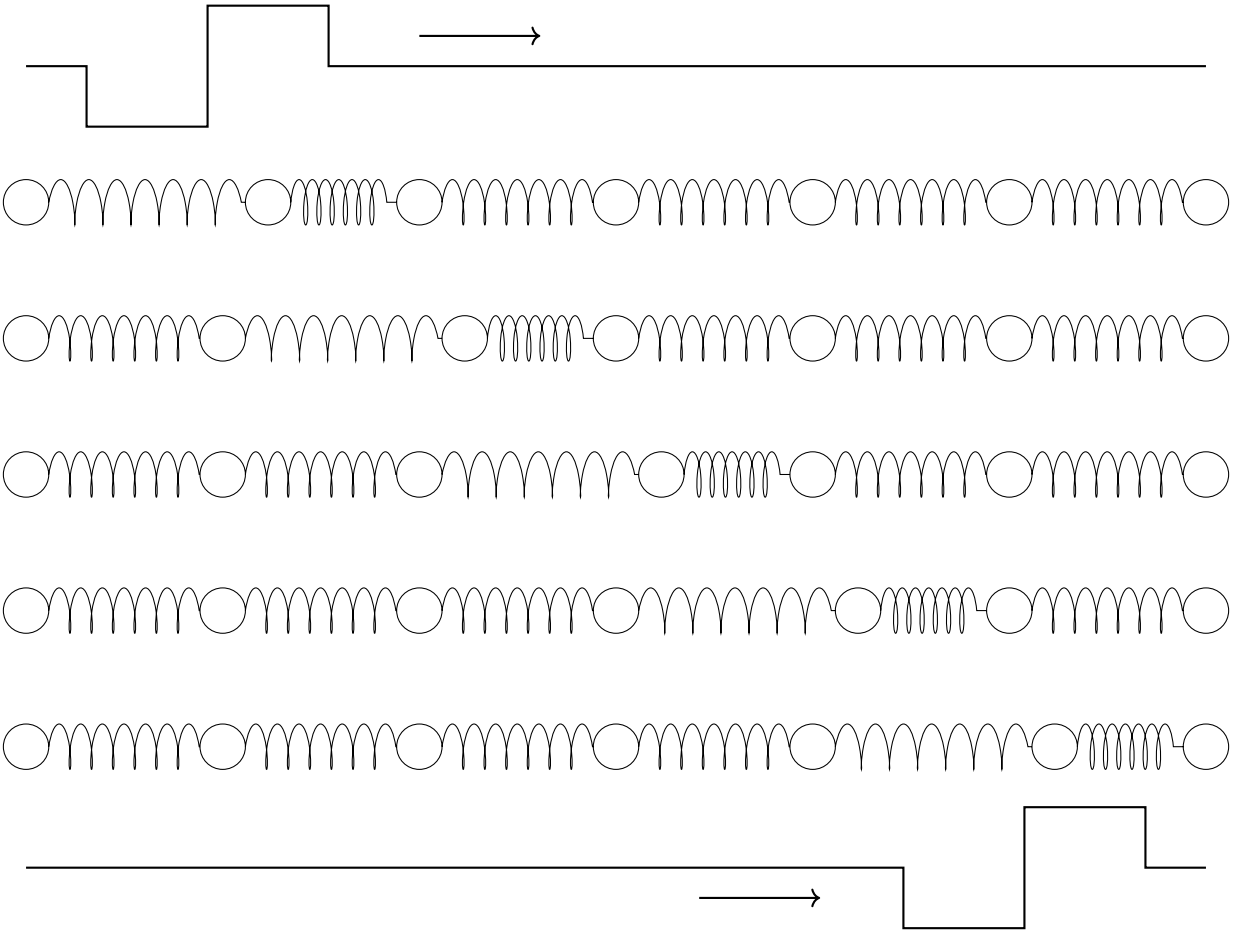


FIGURE 2.2: The mass-spring model showing the propagation of a sound pulse through a material, after [1].

Within this model of sound propagation it can be said that the stiffness and density of the media that the wave is travelling through are directly related to the speed at which the sound wave traverses the matter. The stiffer the matter, the faster sound will travel through it, and this is represented by the stiffness of the springs in the mass-spring model. Conversely, the denser the matter, the slower sound will travel through it, and this is represented by the density or mass per unit volume, of the masses in the model. The equation for the thermodynamic speed of sound c in a given material is therefore [38]:

$$c = \sqrt{\frac{B}{\rho}} \tag{2.1}$$

where B is the bulk modulus (Nm^{-2}), which describes the strength of the intermolecular forces of the matter, and ρ is the density ($\mathrm{kg\ m}^{-3}$) of the material. More specifically, the speed of sound in a gas can be calculated using the following equation [38]:

$$C_{\text{gas}} = \sqrt{\frac{\gamma RT}{M}} \quad \left(\frac{\text{m}}{\text{s}}\right) \tag{2.2}$$

where C_{gas} is the speed of sound in the gas, γ is the Abiabatic index of the specific gas (ratio of specific heats), M is the molecular mass of the specific gas ($\mathrm{kg\ mole}^{-1}$), R is the universal gas constant ($8.31\mathrm{JK}^{-1}\mathrm{mole}^{-1}$), and T is the absolute temperature of the gas in Kelvins (K). Therefore, the speed of sound in air at a temperature of 20°C can be calculated by evaluating the above equation (2.2) as follows:

$$c = \sqrt{\frac{1.4 \cdot 8.31 \cdot T}{2.89 \times 10^{-2}}} \tag{2.3}$$

$$c = 20.05\sqrt{T} \tag{2.4}$$

$$c = 20.05\sqrt{273 + 20} \tag{2.5}$$

$$c = 343.4\,\mathrm{ms}^{-1} \tag{2.6}$$

Frequency and Wavelength

The examples of sound wave propagation introduced and illustrated by the mass-spring model in the previous section are based on propagation as a result of excitation from a single pulse. However, the mass-spring model can also be used to illustrate the physics of periodic sound wave propagation, which is perceived by the ear as a continuous repetitive sound. The most basic form of periodic signal is a sine wave, which constitutes repeating vibrations at a single frequency. Figure 2.3 shows an illustration of a sound wave exhibiting sinusoidal behaviour. In the same way that the mass-spring model diagram in Figure 2.2 illustrated the movement of pressure differential between air molecules as a result of excitation from a single pulse, in Figure 2.3 points of high pressure are represented by

the compressed springs and the point of low pressure is represented by a rarefied spring, together illustrating the repeating oscillation of a periodic sine wave.

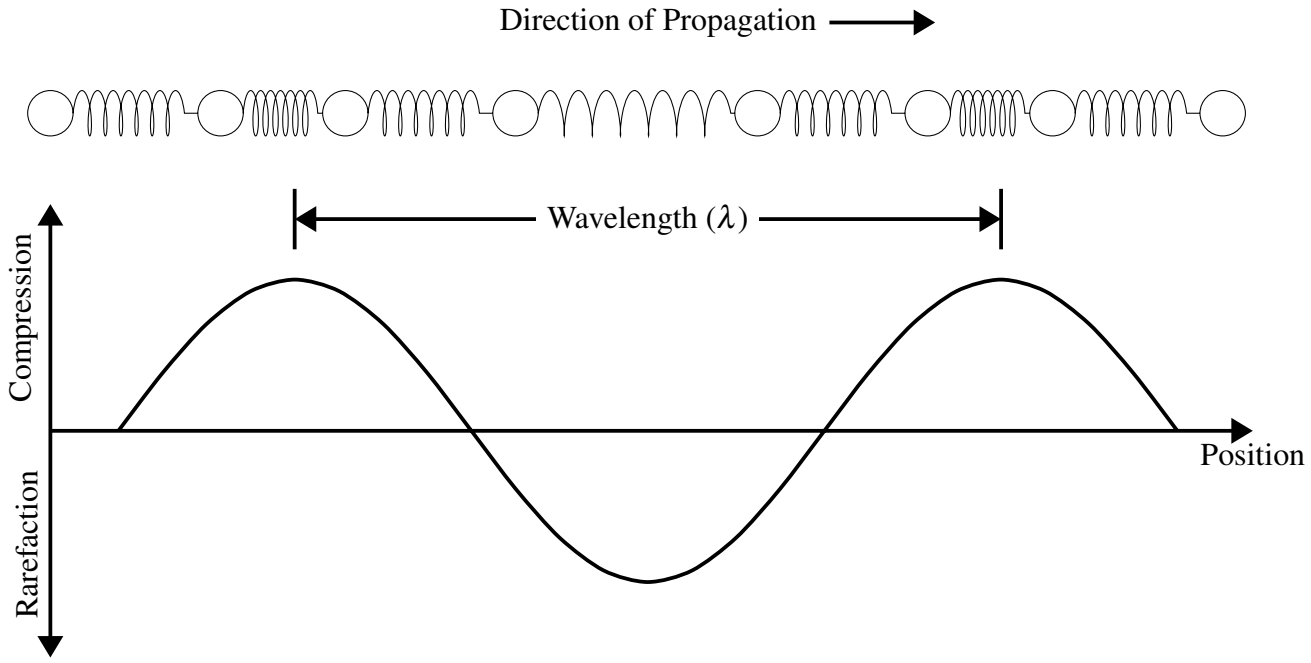


FIGURE 2.3: The propagation of sine wave through a material represented by the mass-spring model, after [1].

The distance that a single repeated section of a wave spans is referred to as the wave's wavelength, this is represented in equations by λ. The frequency f of a sine wave is directly related to its wavelength as follows:

$$c = f\lambda \tag{2.7}$$

This equation can be rearranged to calculate the wavelength of a sine wave if its frequency is known. In addition to frequency, there are two other core components that are that are required to define a sine wave, these are as follows:

- Amplitude (A), which in this context describes the difference in pressure between the maximal peak of a compression or rarefaction and the midway point of a cycle. This is perceived by a listener as the loudness of the wave.

- Phase (ϕ), which describes the initial starting point of the wave. This has an almost negligible impact on how the wave is perceived in isolation, but a much more significant bearing on how the wave is perceived when subject to interference from other sound waves.

The output of a sine wave at a given point in time $y(t)$ can therefore be calculated using the following formula:

$$y(t) = A\sin(2\pi ft + \phi) \tag{2.8}$$

It is of course the case that the vast majority of real sounds encountered by humans in the natural world are far more complex than a single sine wave. However, it is known to be the case that any periodic waveform can be modelled as a combination of a series of sine waves with nominal frequencies, amplitudes, and phase values [2]. This is Fourier's theorem, and can be described mathematically as [1]:

$$f(t) = a_0 + \sum_{n=1}^{\infty} a_n \cos(n\omega_0 t) + b_n \sin(n\omega_0 t) \tag{2.9}$$

where a_0 is the amount that the whole signal is offset from 0 (referred to as the d.c. component or bias), a_b and a_n represent the input of the nth cosine and sine components, and ω is the angular frequency in radians ($2\pi f$). By extension, the Continuous-Time Fourier Transform can be used to describe non-period waveforms [39]:

$$X(j\omega) = \int_{-\infty}^{\infty} x(t)\, e^{-j\omega t}\, dt \tag{2.10}$$

where $x(t)$ is the original time-domain signal, $X(j\omega)$ is the frequency-domain representation, ω is the angular frequency in radians/second, and $e^{-j\omega t}$ is the complex exponential. Furthermore, the

Inverse Fourier Transform can be used to reconstruct the original time domain signal from its Fourier Transform [39]:

$$x(t) = -\frac{2\pi}{T}\int_{-\infty}^{\infty} X(j\omega)e^{j\omega t}\,d\omega \tag{2.11}$$

Sound Intensity

Most sound that an individual encounters in outdoor spaces behaves approximately according to free field conditions. Free field conditions are defined when sound is free to travel unimpeded in straight lines [2]. Typically, most outdoor urban public spaces do not behave as a free field. Objects and buildings within these spaces introduce reflection, absorption, and diffraction of sound, in addition to other forms of obstacle and boundary related wave interactions. However, for the purposes of describing the propagation of direct sound in outdoor spaces, that is, any sound that reaches the receiver unimpeded without interacting with any obstacles or boundaries, it is useful to think in terms of free field conditions.

Figure 2.4 shows how in free field conditions a sound emanating from a point source[1] radiates spherically and uniformly is all directions. It also illustrates how the intensity of a sound decreases as the distance from its source increases. The diagram shows that the same amount of sound power that passes through the small area represented by the square marked $A1$, also passes through the successively larger areas marked by the further square marked by $A2$, $A3$, and $A4$. As the sound travels further from its point source origin and the radius r of propagation expands, its energy is spread over increasingly larger areas, and consequently its intensity decreases [2]. The amount that the sound intensity decreases is inversely proportional to the square of the distance from the source. This is an example of an inverse square law and is described mathematically by the following equation [2]:

[1] A sound source may be considered a point source if its largest dimension is less than one-fifth of the distance than the distance that it is being measured from [2].

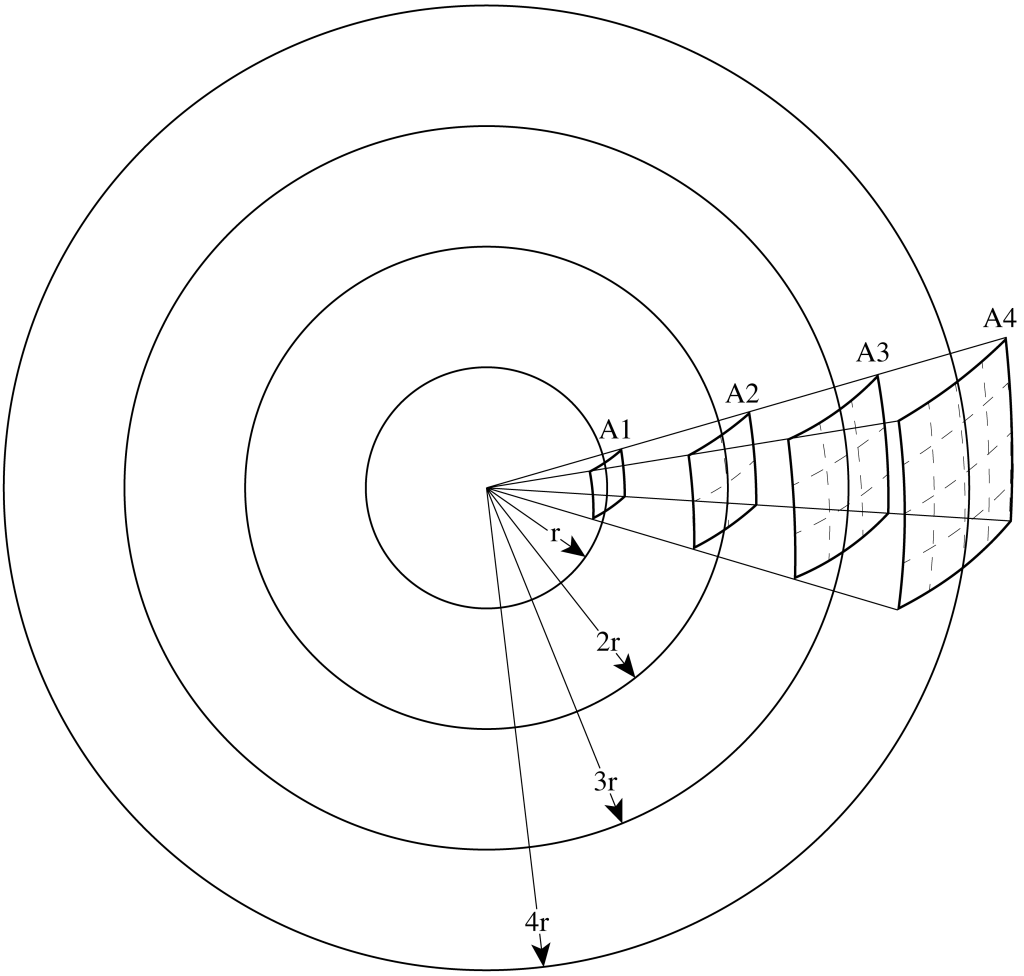


FIGURE 2.4: *Diagram illustrating how the intensity of a sound radiating from a point source in free field conditions is inversely proportional to the square of the distance from the source, after [2].*

$$I = \frac{W}{4\pi r^2} \quad \left(\frac{\mathrm{W}}{\mathrm{m}^2}\right) \tag{2.12}$$

where I is the intensity of the sound, W is the power of the sound source, and r is the distance from the source (radius).

Sound Pressure Level

Sound levels are typically measured using a decibel (dB) scale and constitute sound pressure level (SPL) values. SPL measurements quantify the loudness of a sound relative to the weakest audible sound that the human ear can perceive, approximately equal to a pressure value of 20 μPa [4]. The SPL level of a recorded RMS (root mean squared) sound pressure value is calculated using the following formula [2]:

$$\text{dB SPL} = 20\log_{10}\left[\frac{\text{RMS Sound Pressure}}{2\times10^{-5}\text{ Pa}}\right] \tag{2.13}$$

Table 2.1 is a list of some common environmental sounds and their typical measured SPL levels. It is worth noting that while it is possible for sounds to exceed 194 dB SPL, above this point the waves begin to distort and they are technically classified as shock waves. The dB scale is not linear but rather a logarithmic scale, partly to allow for the vast range of perceptible pressure values it has to be able to represent, with sound levels measured at 194 dB SPL having a sound pressure of 100,000 Pa [2]. But it is also to reflect the fact that human perception of loudness is logarithmic and not linear. Humans also perceive loudness differently across the frequency spectrum, being more sensitive to sounds in the 3-4 kHz region compared to sounds at higher and lower frequencies [2].

It is therefore sometimes appropriate to apply a frequency dependant weighting that, when applied to a measured sound pressure level, produces a final dB SPL level that accounts for how the human auditory system perceives loudness at different frequency bands and amplitudes. The most commonly used weightings are A, B, C, and D. A-weighting was originally derived from a set of equal-loudness contour curves published in a paper by Fletcher and Munson [40] and is commonly used by acousticians for the measurement of environmental and industrial noise.

Sound Pressure Levels (dB SPL)				
Weakest sound heard	0 dB		Hand drill	98 dB
Whisper quiet library at 1.8 m	30 dB		Power motor at 0.9 m	107 dB
Normal conversation at 0.9 m	60-65 dB		Sandblasting, live concert	115 dB
Telephone dial tone	80 dB		Onset of pain	125 dB
City traffic (inside car)	85 dB		Pneumatic riveter at 1.2 m	125 dB
Train whistle at 152 m	90 dB		Level at which short term exposure can cause permanent damage	140 dB
Jackhammer at 15 m	95 dB		Jet engine at 30 m	140 dB
Subway train at 61 m	95 dB		12 Gauge shotgun blast	165 dB
Level at which sustained exposure may result in hearing loss	90-95 dB		Death of hearing tissue	180 dB
			Loudest sound possible	194 dB

TABLE 2.1: Some common environmental sounds and their associated SPLs, after [24].

2.2 Human Hearing

A significant proportion of the research presented in the later chapters of this thesis are related to various aspects of how listeners perceive sound, and so it is important to have an understanding of the fundamental concepts relating to human hearing and how listeners perceive sound sources spatially.

2.2.1 Anatomy of the Ear

The human ear consists of three main parts, the outer ear, middle ear, and inner ear. Figure 2.5 is a diagram that illustrates the principle components of the anatomy of the ear. The outer ear is made up of the pinna and the auditory canal, which is terminated by the eardrum, also referred to as the tympanic membrane. The middle ear is a cavity filled with air containing three small bones called the malleus, incus, and stapes, and collectively referred to as the ossicles. The stapes attaches to the oval window of the inner ear and the malleus is attached to the eardrum. The malleus, incus, and staples join together and form a mechanical connection between the eardrum and the oval window of the cochlea in the inner ear. [2] Changes in air pressure at the ear drum are translated to vibrations in the fluid contained within the cochlea, which are in turn translated into electrical impulses via approximately 25,000 nerve endings in the auditory nerve where the inner ear is terminated [41].

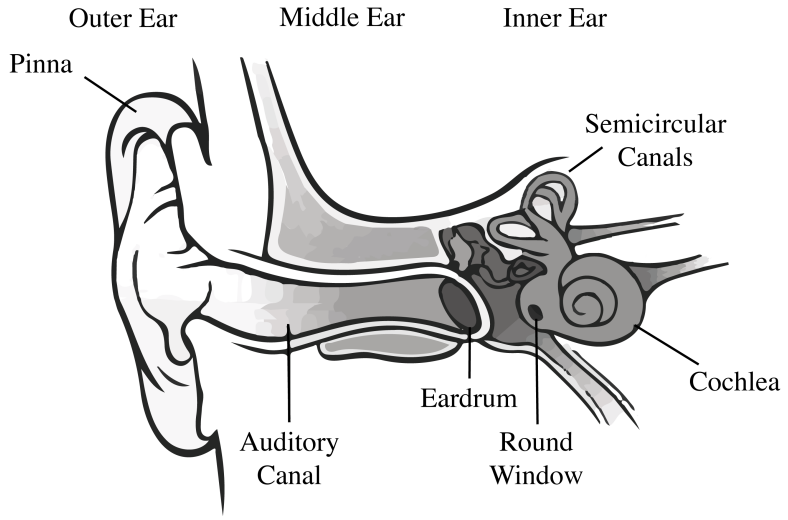


FIGURE 2.5: *Diagram illustrating the anatomy of the human ear, after [3] and [2].*

2.2.2 Spatial Hearing

Humans are able to localise the perceived origin of a sound in space as a result of their binaural hearing abilities. Figure 2.6 is a diagram illustrating key terminology relevant to describing the binaural hearing system. The concepts that these terms describe are defined as follows:

- The interaural axis is a line bisecting either ear of the head.
- Straight ahead is a line that begins at the centre of the head and extends directly forward in the same direction that the head is facing.
- The horizontal, median, and frontal planes are axes that can be used to describe the positional location of sound sources relative to the head. The horizontal or azimuthal plane is parallel to the interaural axis and the direction straight ahead of the head. The median plane is parallel to the straight ahead direction and perpendicular to the interaural axis. The frontal plane runs parallel to the interaural plane and perpendicular to the straight ahead direction.
- The sound direction describes the position of a sound source as defined by two angles. The azimuth angle (θ) is the direction of the sound source on the azimuthal plane. The elevation angle $\emptyset$ describes the elevation of the sound source.

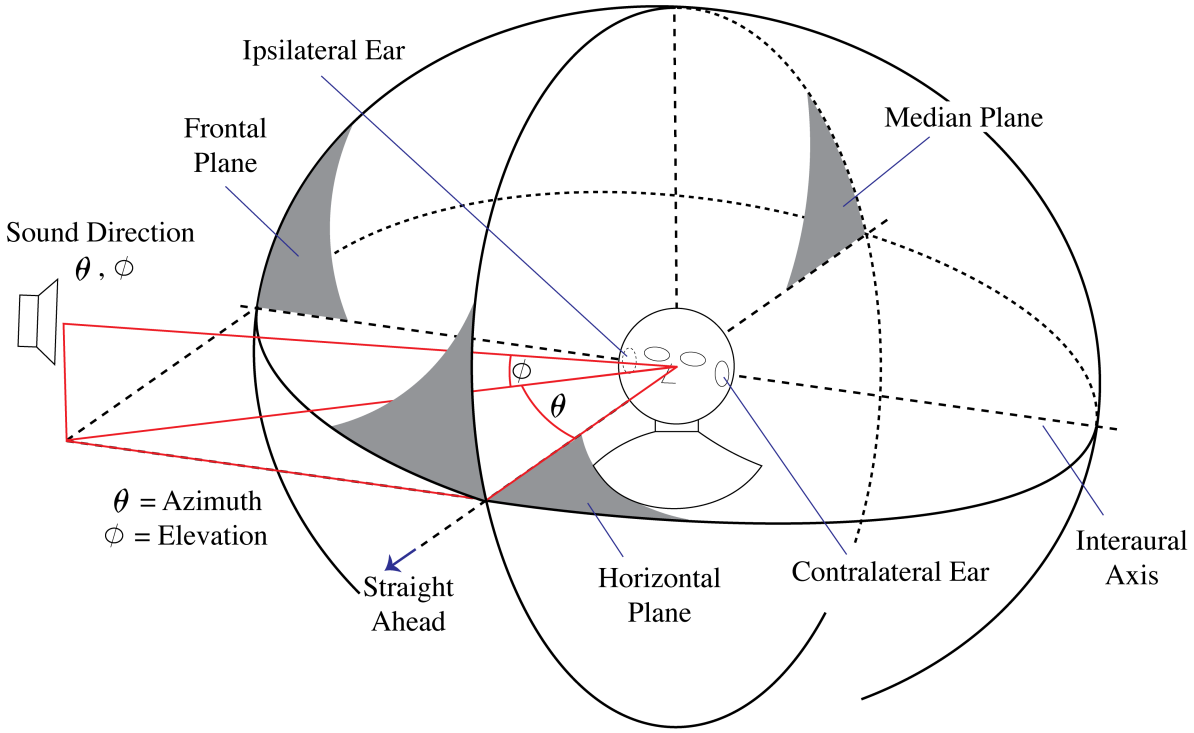


FIGURE 2.6: *Diagram illustrating the head-related coordinate system and other key concepts relating to spatial hearing, adapted from [4] and [5].*

- The terms ipsilateral and contralateral are used to describe either ear relative to the direction of a sound source on the horizontal plane. The ipsilateral ear is whichever ear is on the same side of the head as the direction of the sound at a given point in time, with the contralateral ear being the ear on the opposite side to the direction of the sound. The labelling of ipsilateral and contralateral ears will therefore swap depending on the arrival direction of the sound.

Interaural Time Difference (ITD)

Unless located exactly on the median plane, a sound in a listener's surroundings will arrive at one of their ears slightly before the other. This difference in the time that the sound takes to reach either ear is referred to as the interaural time difference (ITD). Figure 2.7 is a diagram that illustrates the extra distance that a sound from a source located on one side of the horizontal plane has to travel to reach the contralateral ear, compared to the ipsilateral ear. The human auditory system is then able to process this difference in the arrival time to determine a perceived direction of incidence for the

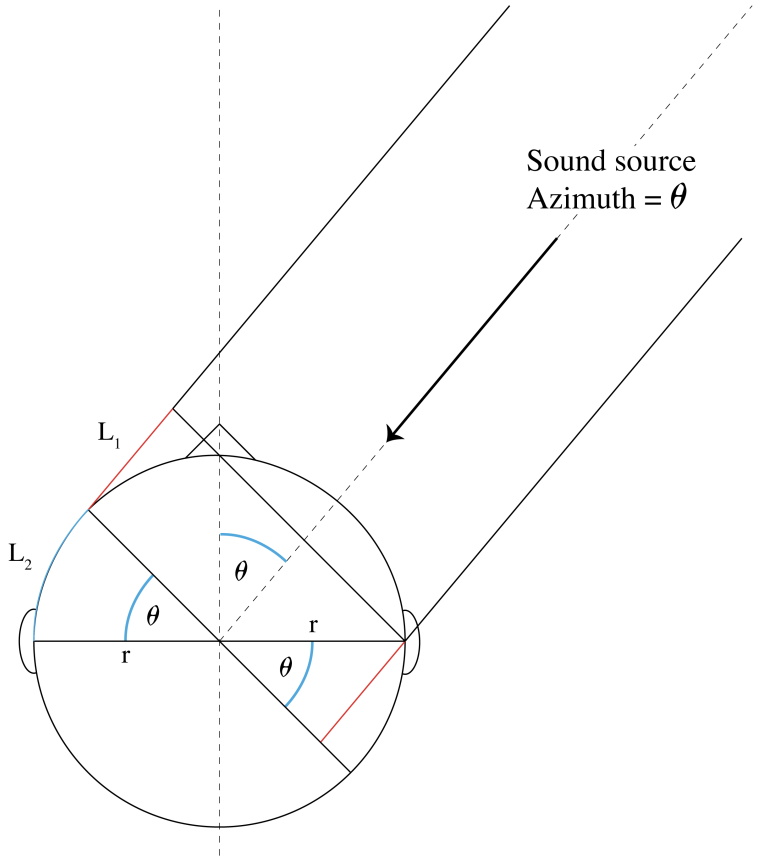


FIGURE 2.7: *Diagram showing a model of Interaural Time Difference (ITD) illustrating how the incidence angle of a sound source and the scale of the head results in a difference in the paths the sound will take to arrive at either ear, after [1].*

sound [4]. An ITD model proposed by Woodworth initially suggested that ITDs were independent of frequency [42], however subsequent work [43] has shown theoretically and experimentally that ITDs are frequency dependant, due to the diffraction of sound waves around the head. This results in low frequency ITDs being 1.5 greater than high frequency ITDs, meaning that localisation of sound sources based on ITD cues is less accurate for higher frequency sources [44].

Interaural Level Difference (ILD)

If a sound arrives at the head from such an angle that the signal received at the contralateral ear is physically occluded by the head this can result in a frequency dependent attenuation of the level of the signal. This phenomenon is known as the acoustic shadow and is illustrated by the diagram in

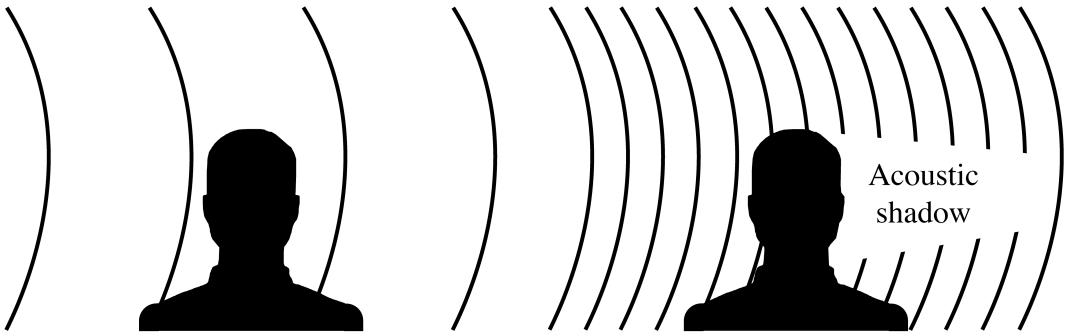


FIGURE 2.8: *Diagram showing how the occlusion from the head results in acoustic shadowing at higher frequencies, after [6].*

Figure 2.8. In the scenario presented on the left hand side of the diagram it is shown how when lower frequency sound waves (typically below 1500 Hz [41]) are occluded by a head there is little or no attenuation of the signal. Whereas, in the scenario illustrated on the right hand side of the diagram, it can be seen that higher frequency sounds, possessing wavelengths of similar dimensions to the head and features on the head, are attenuated at the contralateral ear, relative to the direct signal at the ipsilateral ear [4]. This effect is most significant for sounds occupying frequencies above 1500 Hz [41].

Spectral Cues

In instances where the direction that a sound arrives from is about the median plane the ITD and ILD are minimal and potentially even zero. Localisation of sound from interaural cues in these scenarios is therefore less precise compared to sound arriving from a direction located on the horizontal or frontal planes and not on the median plane. Despite this, localisation of sources on the median plane is made possible by evaluation of spectral differences in monaural cues [4]. Monaural cues are non-identical for each ear and both monaural and binaural cues are applied to the sound pressure received at the eardrums as a result of diffraction of the arriving sound on physical features of the head and torso. The extent of this diffraction can be quantified by the head-related transfer function (HRTF) [45, 46]. This is defined as the measured sound pressure at the eardrum or ear canal entrance divided by the measured sound pressure at the centre of the position of the head with the head and torso removed. HRTFs are also dependent on the direction of arrival (DOA) of sound and therefore a large number

of measurements are required at various DOA incidences around the head in order to accurately capture and model a real listener's spatial listening experience. This collection of measurements is referred to as a HRTF dataset and can be used to create a personalised HRTF filter that, when applied to audio experiences rendered binaurally and played back over headphones, can potentially lead to a more natural listening experience [47], improvements in localisation acuity [48], and an overall preferred listening experience [49]. However, the results of a study that compared listeners' perception of binaural audio rendered with their own individualised HRTF and generic HRTFs from acoustic test fixture manikins, found that listeners' tended to prefer the generic HRTFs over their own individualised HRTF across metrics including richness, brightness, preference, and externalisation [50]. These contrasting results reflect the ongoing debate that exists within the research community around the relative effectiveness of individualised and generic HRTFs.

2.3 Environmental Noise and Soundscape

The final key area of fundamental theory that readers of this thesis should be familiar with is the concept of environmental noise. This section will discuss what is meant by the environmental noise, how it should be understood in the context of this thesis, and how sounds that contribute to environmental noise can be categorised through the lens of soundscape studies.

2.3.1 Environmental Noise

The etymological origins of the word *"noise"* can be traced back to *noxia* and *nausea* [51]. These origins are congruent with contemporary definitions that define noise as sound that is considered *"out of place"* or *"unwanted sound"* [52]. It is important to highlight how these definitions indicate that the question of whether a sound is considered noise or not is therefore highly subjective and and may vary from listener to listener. Of course, there are other types of sound that the word noise can be used to describe, including the use of different standardised noise signals in acoustic measurement and testing, noise as a coded information carrier [2], and even the deliberate use of noise signals as a tool in musical compositions and recordings [53]. But generally speaking it is the case that the term noise

is most commonly used to describe sound that is thought of as undesirable, or at least considered at odds with other contemporaneous sounds.

Similarly, environmental noise typically tends to be thought of as undesirable sound, with the World Health Organization identifying it as an important problem since the 1970s [54], and Stansfeld and Clark having defined it as:

> *"...a pervasive environmental pollutant that leads to annoyance and can be perceived as stressful."* [55]

while in 2009 Murphy et al. defined environmental noise as:

> *"...any unwanted or harmful outdoor sound created by human activities that is detrimental to the quality of life of individuals."* [56]

Environmental noise that listeners experience in public spaces is typically made up of many different competing sound sources. A non-exhaustive list of these sounds is featured in an example taxonomy of the acoustic environment [57] and discussed later in this chapter.

Environmental noise is also known to be a cause of annoyance amongst the population, particularly in built up, densely populated urban areas [58]. Noise annoyance was defined by Ouis as:

> *"...a feeling of displeasure, nuisance, disturbance or irritation caused by a specific sound..."* [59]

Beyond being a cause of annoyance [60], the World Health Organization (WHO) has conducted studies that have examined how long term exposure to environmental noise can lead to hearing damage [61]. In addition the WHO have conducted systematic reviews which outline the detrimental effects that environmental noise has been found to have on cognition [62], sleep [63], and cardiovascular and metabolic health [64].

However, it is important to highlight that sound classified as environmental noise may not always be perceived as an annoyance or a cause of harm to those who experience it. While the term environmental

noise could be used to describe the loud and potentially harmful traffic noise experienced on an overpass above a busy road, it could equally be used to refer to the gentle array of sounds that make up the hubbub of a public park on a quiet afternoon. In the context of the present thesis the term *environmental noise* should be understood as representative of sound in a listener's surroundings that has the potential to capture their attention, regardless of the response the sound elicits.

2.3.2 Soundscape

In 1977 Schafer defined the term soundscape as representing:

> *"...any field of acoustic study."* [51, p.7]

Within this the themes of keynote sounds, signals, soundmarks and archetypal sounds were established to categorise sound within a soundscape [51]. Since its initial inception, ostensibly as a purely theoretical framework for the analysis of sounds and their perception within the environment, use of the term has significantly expanded to include further theoretical and empirical work in the fields of soundscape ecology [65], planning [66], conservation [67], preservation [68] and the complex acoustic urban soundscapes of modern cities [69]. Additionally, in contemporary literature the term tends to be most closely associated with research exploring the human perception of the surrounding sonic environment. However, within the context of the research presented in this thesis, the main purpose of the soundscape research discussed here is to serve as an illustration of the many and varied sounds that are typically encountered by mobile audio consumers while listening in both indoor and outdoor spaces, and how these sounds have tended to be broadly categorised by successive taxonomies within the field.

Soundscape Categorisation

In 1977 Schafer [51] proposed one of the first taxonomies for the categorisation of environmental sounds, which comprised of the following categories; *natural sounds*, *human sounds*, *mechanical sounds*, *quiet and silence*, *sounds and society*, and *sounds as indicators*.

In 2004 Gage et al. [70] proposed a taxonomy of acoustics which consisted of the three primary categories of *biophony*, *geophony*, and *anthrophony* under the main heading of *sound spectrum*. Biophony was defined as biological sounds in the soundscape produced by the environment's vocal organisms, geophony represented sound from the physical environmental such as wind, and anthrophony defined sounds in the soundscape created as a consequence of human related activity. Within anthrophony the further sub categories of *mechanistic* and *oral* sounds were defined, with the additional sub categories of *stationary* and *temporal* sounds ascribed to the mechanistic anthrophony.

In 2011 Brown et al. [57] devised a taxonomy of the acoustic environment that differentiated between *indoor* and *outdoor acoustic environments*. Within outdoor acoustic environments the main subcategories of *urban*, *rural*, *wilderness*, and *underwater acoustic environments* were defined. Under each of these, as well as with the indoor acoustic environment category, two further sub categories of *sounds generated by human/activity/facility* and *sounds not generated by human activity* were outlined. For *sounds not generated by human activity* two further main sub categories were defined as *nature* and *domesticated animals*. For *sounds generated by human activity/facility* there were a further 6 sub categories of *motorised transport*, *human movement*, *electro-mechanical: -stationary -mobile*, *voice & instrument*, *social/communal*, and *other human*. Finally, under the voice & instrument sub category the additional sub categories of *amplified* and *non-amplified* were defined.

In 2022 Giordano et al. [71] produced a systematic review of 36 taxonomies and ontologies mapping out everyday sounds from a wide range of research fields extending beyond soundscape research, including auditory cognition, artificial hearing, sound design, and medicine. While Giordano et al. concluded that further work was required to produce a completely comprehensive ontology of everyday sounds, nevertheless it remains the case that common to all taxonomies surveyed in their work, as well as those presented in this chapter, the main groupings are defined as natural, mechanical, and human-related sounds.

2.4 Summary

This chapter has provided a review of fundamental theory relevant to the concepts of sound, human hearing, and environmental noise. It was shown how sound propagates as a wave, how wavelength

relates to frequency, how sound decreases in intensity the further it radiates from its source, and how the level of a sound can be measured in dB SPL units. Core concepts crucial to the understanding of how humans are able to hear spatially were also discussed. Finally, a review of the concept of environmental noise and how it should be understood in the context of this thesis was presented. This review serves to provide an initial basis of core knowledge upon which an understanding of the further concepts and research presented in the subsequent chapters can be built. The next chapter will present a review of auditory headsets commonly used by consumers, provide a background on the practice of mobile audio consumption, and discuss contemporary research on listener engagement with speech based media.

Chapter 3

Related Work

3.1 Auditory Headsets and Mobile Audio Consumption

In the previous chapter, the topics of soundscape and spatial hearing were introduced and discussed. This provides a knowledge of the theoretical constructs required to understand the concept of environmental noise and how it comes to be perceived in free field listening conditions. The first section of the following *related work* chapter will introduce and discuss the array of auditory headsets and external sound modification technologies that are commonly used by listeners in mobile audio consumption experiences. Supporting theoretical concepts and a review of applied use cases are detailed. The historical and contemporary practice of mobile audio consumption by consumers is then introduced to provide an overview of research relating to listener engagement in mobile speech-based media consumption experiences.

3.1.1 Auditory Headsets used in the Mobile Consumption of Audio

There are many different types of auditory headsets that are commonly used today in the mobile consumption of audio. The most prevalent of these are headphone and earphone devices, which usually comprise of drivers mounted in enclosures, positioned either over, on, or in either ear, and connected to a media playback device using either a wired 3.5 mm mini jack connection or a wireless protocol such as Bluetooth. In 2022 a survey of user perspectives on headphone technology comprising of 406

participants found that 63.5% of those surveyed used over-ear headphones, 35.7% used wired in-ear earphones, 18.0% used wireless in-ear earbuds, and 11.1% used on-ear headphones [72].

Headphones and Earphones

Headphones or earphones can generally be described as belonging to one of six different categories, each of which differs in physical design and the positional relationship between the driver and the ear canal. Within these six categories there are also open and closed variants. ITU Recommendation P.57 [7] provides the following definitions for open and closed earphones:

> *"**Acoustically open earphones (nominally unsealed):** Earphones which intentionally provide an acoustic path between the external environment and the ear canal."* [7]

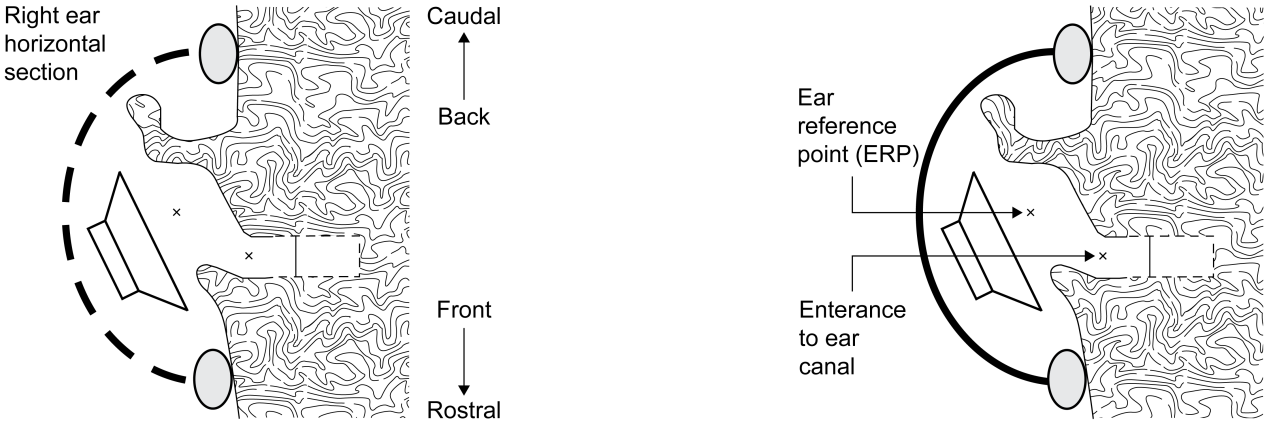
> *"**Acoustically closed earphones (nominally sealed):** Earphones which are intended to prevent any acoustic coupling between the external environment and the ear canal."* [7]

The physical design of a headset, and whether it is of open or closed design, are both significant factors in how much passive occlusion of external sound a device will provide. The five primary categories of devices that are commonly used in the mobile consumption of audio are circum-aural, supra-aural, and supra-concha headphones, and intra-concha and insert earphones. These are described in ITU-T Recommendation P.57 [7] and are defined as follows:

Circum-aural Headphones

Also commonly referred to as over-ear headphones, circum-aural earphones are defined as:

> *"Earphones which enclose the pinna and seat on the surrounding surface of the head. Contact to the head is normally maintained by compliant cushions. Circum-aural earphones may touch but not significantly compress the pinna."* [7] (see Fig. 3.1)



(A) *Circum-aural (open)* (B) *Circum-aural (closed)*

FIGURE 3.1: *Diagrams showing examples of open and closed circum-aural headphones over the pinna of an acoustic test fixture (ATF)*[1]*, after [7].*

Circum-aural headphones usually feature a wide headband that is worn over the top of the user's head, connecting either ear cup, and applying light clamping pressure to ensure secure seating on the head.

Supra-aural Headphones

Also commonly referred to as on-ear headphones, supra-aural earphones are defined as:

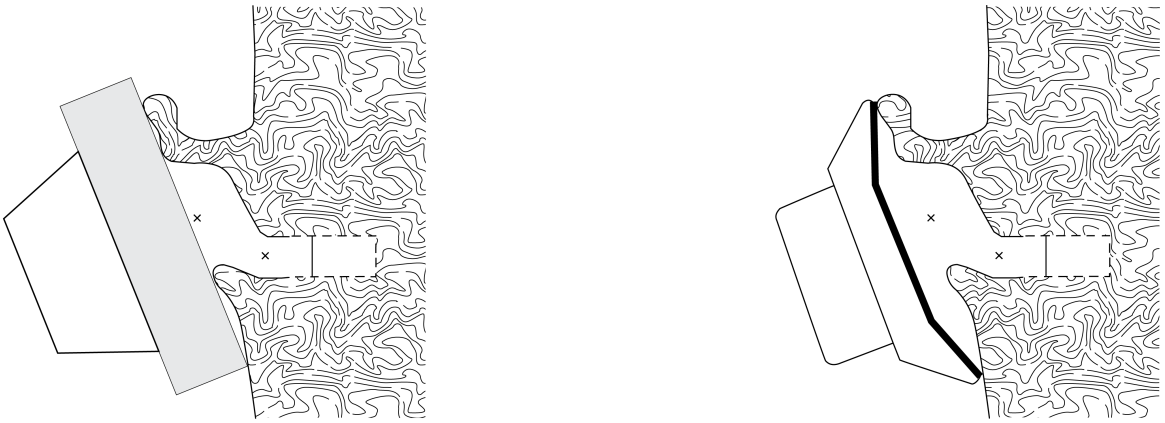
> *"Earphones which rest upon the pinna and have an external diameter (or maximum dimension) of at least 45 mm."* [7] (see Fig. 3.2)

As with circum-aural headphones, supra-aural headphones also usually feature a headband that is worn over the top of the user's head, connecting either ear cup, and applying light clamping pressure to ensure secure seating on the head.

Supra-concha Headphones

As with supra-aural earphones, supra-concha earphones are also commonly referred to as on-ear headphones, and are defined as:

[1] An ATF is a device that simulates a human head. They are commonly used when conducting acoustic measurements in instances where it is important that the acoustic effects of the head are accurately modelled.



(A) *Surpa-aural (open)* (B) *Supra-aural (closed)*

FIGURE 3.2: *Diagrams showing examples of open and closed supra-aural headphones on the pinna of an ATF, after [7].*

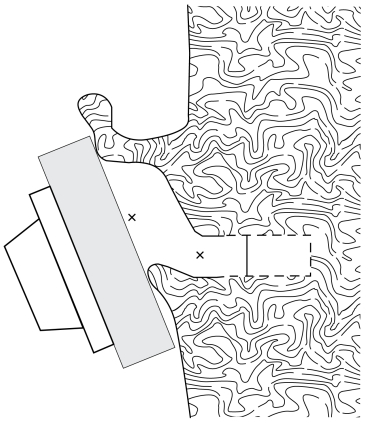


FIGURE 3.3: *Diagram showing an example of an open supra-concha headphone on the concha of an ATF, after [7].*

> *"Earphones which are intended to rest upon the ridges of the concha cavity and have an external diameter (or maximum dimension) greater than 25 mm and less than 45 mm."* [7] (see Fig. 3.3)

Supra-concha headphones usually feature a less substantial headband than circum- and supra-aural headphones. The headband is worn over the top of the user's head, connecting either ear cup, and applying light clamping pressure to ensure secure seating on the head.

Intra-concha Earphones

Also commonly referred to as Earbuds, intra-concha earphones are defined as:

> *"Earphones that are intended to rest within the concha cavity of the ear. They have an external diameter (or maximum dimension) of less than 25 mm but are not made so as to enter the ear canal."* [7] (see Fig. 3.4)

(A) *Intra-concha (open)* (B) *Intra-concha (closed)*

FIGURE 3.4: *Diagrams showing examples of open and closed intra-concha earphones in the concha of an ATF, after [7].*

Intra-concha earphones tend not to feature headbands, instead usually relying on the friction between the earphone tip and concha to maintain a secure seating in the user's ear. The strength of this seating is heavily dependant on the quality of the physical fit of the earphone in the user's concha, with this quality of fit varying drastically from user to user. It is therefore not unusual for intra-concha earphones to sometimes work themselves loose and fall out of the ear, especially in the event of sudden movement. The amount of passive occlusion of external sound that the earphones are able to provide is also highly dependent on the quality of physical fit.

Insert Earphones

Also commonly referred to as In-Ear Monitors (IEM), insert earphones are defined as:

> *"Earphones that are intended to partially or completely enter the ear canal."* [7] (see Fig. 3.5)

Insert earphones are less commonly used for the mobile consumption of audio, they are instead much more frequently used by musicians and crew for monitoring purposes in the performance and

(A) *Insert (open)* (B) *Insert (closed)*

FIGURE 3.5: *Diagrams showing examples of open and closed insert earphones in the ear canal of an ATF, after [7].*

production of live events. In such applications it is vitally important that users are able to hear the audio being played back, over the extraneous noise of the other instruments and crowd in the performance venue. Insert earphones also tend not to feature headbands, instead usually relying on the friction between the earphone tip and ear canal to maintain a secure seating in the user's ear. In-ear monitors often utilise custom ear pieces that have been molded to fit the user's ear canal perfectly, providing a very high level of passive occlusion of external sound. The user is then able to listen to a custom mix of audio that could include anything from their own instrument, their band mates' instruments, click/backing tracks, crowd microphones, or communications from production crew. This also means that insert earphones tend to be much less prone to working themselves loose and falling out of users' ears.

Extra-aural Headphones

Extra-aural headphones are the least common of the six types of headphone and earphone devices presented in this review. They are rarely used for the mobile consumption of audio, instead more often being used in audio augmented reality research applications. There is no standardized definition for extra-aural headphones, however they were described by Schneiderwind et al. [73] as follows:

"...extra-aural headphone expose the ears to the real sound field..." [73] (see Fig. 3.6)

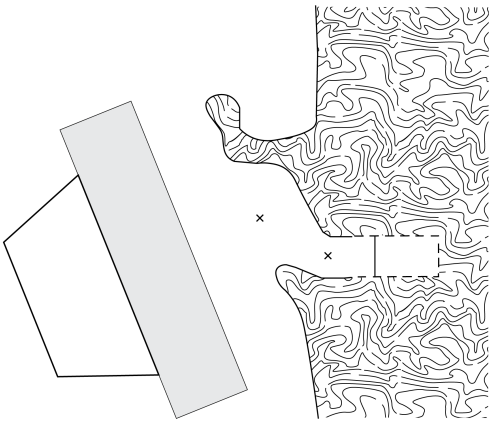


FIGURE 3.6: *Diagram showing an example of a typical open extra-aural headphone suspended adjacent to the pinna of an ATF, inspired by [7].*

Specifically referring to the AKG K1000 extra-aural headphones (pictured in Fig. 3.7) Schneiderwind et al. also stated that the fact the AKG K1000 does not enclose the ear:

> *"...implies that sound waves from the real acoustic environment can reach the ear directly for sound waves coming from the front or the back of a listener."* [73]

Many extra-aural headphones, such as the AKG K1000 [8] and BK211 [74], are similar to supra-aural headphones in acoustical design with the key difference that either ear cup is suspended away from the ear such that it does not come into direct contact with the pinna. Some extra-aural headphones such as the VIE SHAIR [73] are designed in such a way that the ear cup enclosure that houses the driver does not come into contact with the pinna, but an open trapezoidal frame is employed between the ear cups and the pinna to assist with the wearing and positioning of the headset [73].

Passive Acoustically Transparent Auditory Headsets

In 2020 McGill et al. defined *"acoustically transparent auditory headsets"* as:

> *"... auditory wearables that have the capability to, either actively or passively, intermix both real and virtual sounds."* [75, p.35]

FIGURE 3.7: *Photograph showing how the AKG K1000 extra-aural headphones would be worn by a user [8].*

There are three types of auditory headsets that are capable of providing passive acoustic transparency and are, albeit less commonly, used in the mobile consumption of audio. These are wearable non-contact directional speakers [76], bone conduction headsets [77], and acoustically transparent headphones [78], all of which leave the ear canal at least partially open to other environmental sound directly at the open ear [79]. Headsets that deliver audio to the listener in this way are described as providing passive acoustic transparency, allowing the user to consume audio while maintaining an identical or similar acoustic awareness of real sounds in their surrounding environment compared to open ear listening [79].

Wearable Non-Contact Directional Speakers

In 2000 Sawhney and Schmandt proposed the first example of an acoustically transparent headset to use wearable non-contact directional speakers, the Nomadic Radio [80]. The device was worn around the user's neck with two directional speakers on either shoulder used to provide spatialised audio localised at the open ear. In 2019 Bauer et al. presented a study [81] exploring the design of an interactive AAR experience using Bose Frames audio sunglasses (pictured in Fig. 3.8) to provide the required acoustic transparency functionality.

Produced from 2020 to 2022, Bose Frames [76] were one of the most popular consumer products

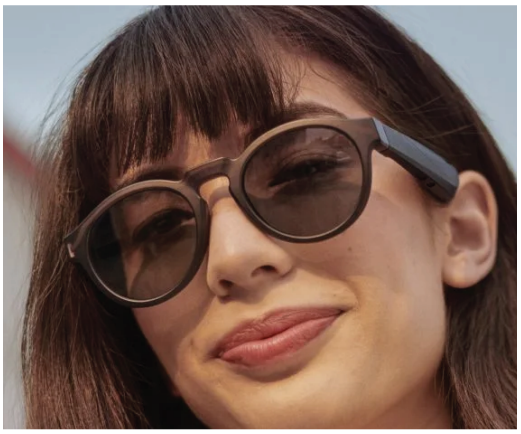


FIGURE 3.8: *Photograph showing how the Bose Frames are worn by a user, adapted from [9].*

based around the concept of wearable non-contact speakers. In 2023 Ray-Ban partnered with Meta to release a pair of smart glasses that provides users with a similar open-ear acoustically transparent listening experience as the Bose Frames [82]. Rather than being worn around the neck, both the Bose Frames and Ray-Ban Meta smart glasses instead take the form of a standard pair of glasses seated on the user's face. The directional speakers are mounted on the inside of either side of the frame at a much closer proximity to the ear, compared to other devices worn around the neck such as the Nomadic Radio [80], thereby enabling improved audio spatialisation performance. Used as an acoustically transparent auditory headset in an AAR system they exhibit largely the same perceived advantages and disadvantages as bone conduction headsets, which are discussed in the following section.

McGill et al. conducted a study in 2020 that used the passive acoustic transparency functionality provided by a pair of Bose Frames to produce an augmented television viewing system [75]. The Frames were used to provide the user with a range of personalised virtual audio stimuli in sync with the audio-visual output from the television and an auxiliary 5.1 surround speaker system.

The study identified several potential applications of the system in the delivery of private personalised audio experiences in shared television viewing scenarios. These included listening to specific channels of the surround mix to enhance the dialogue and presence respectively, the provision of individual audio description for individuals with visual impairments, listening to content in a different language to others to negate the requirement for subtitles, listening to supplementary audio material in the event that content has been viewed previously and providing the user with a mirrored stereo mix in the event they are positioned outside the sweet spot of the surround speaker system [75]. All of these

applications can be viewed as constituting personalised audio experiences that are dependent on the wider context of the circumstances in which they are being consumed.

The system performance was evaluated using questionnaires in two perceptual listening tests comprising of 12 participants. The results concluded that the use of an auditory headset to provide an augmented, mirrored stereo mix of a concurrent surround mix rendered over 5.1 leads to a sense of improved spatialisation and a more immersive and enjoyable experience. The implementation of supplementary audio content such as alternate language, director commentary and audio description was perceived by participants as being *"broadly acceptable"* [75]. In contrast to AAR experiences that are built around the natural soundscape, this research highlights the potential value of using acoustically transparent auditory headsets to deliver AAR experiences that are superimposed over existing media broadcast from separate hardware.

In 2019 Simón Gálvez et al. presented a paper detailing the creation of a novel loudspeaker system that enabled two independent personalised audio signals to be delivered to adjacent listeners in close proximity with completely localised acoustic independence between either perceived experience [83]. The system is also able to continuously maintain the spatialised audio signal provided to the listener irrespective of their relative position in the space, referred to as listener adaptive reproduction. The system achieves this by first using a camera to track the position of either listener within the environment in real time. A linear loudspeaker array then uses optimised beamforming filters to reproduce binaural 3D signals at the ears of either listener, without any cross talk between the left and right ear signals. This work forms the technological basis of the linear loudspeaker arrays produced by AudioScenic [84].

The discrete personalised regions of sound that are created using this method are referred to as personal sound zones [85]. As such the system would be capable of reproducing the same range of private personalised audio experience applications as delivered by a Bose Frames auditory headset in the study on augmented television viewing by McGill et al. [75]. It could therefore also be argued that the AudioScenic system provides the user with full passive acoustic transparency, raising the possibility of future applications for this work in the field of AAR.

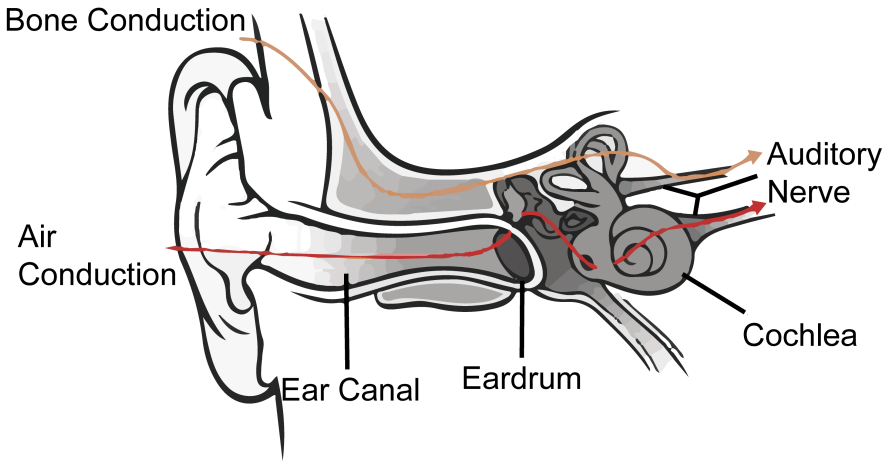


FIGURE 3.9: *Diagram showing how sound is received at the auditory nerve via bone conduction and air conduction, after [3].*

Bone Conduction Headsets

Bone conduction headsets circumvent the outer and middle ear, instead using vibration actuators to deliver sound directly to the user's skull. A diagrammatic illurstration of this is shown in Fig. 3.9. In 2007 Lindeman et al. proposed the concept of Hear-Through AR, representing the first use of a bone conduction headset to deliver acoustic transparency in an AR system [86]. The proposed system used the conduction headset to render the virtual audio component, leaving the open ear free to concurrently perceive real sounds in the surrounding acoustic environment. Whilst this method provided the user with maximum acoustic transparency and optimal ability to localise sounds in the real world due to the preservation of their physical HRTF at the open ear, it lacked the potential for any additional control over the perception of the real sounds, such as the use of active noise cancellation to selectively reject specific sound sources in the environment.

In 2019 Barde et al. conducted a study assessing the perception of elevation using binaurally spatialised audio delivered using a bone conduction headset. The study concluded that whilst bone conduction headsets were generally shown to be effective in reproducing binaural spatialisation, the localisation accuracy achieved by participants was found to be much worse compared to results achieved in equivalent binaural localisation accuracy studies that used headphones [100]. The term hear-through is still present in contemporary literature across the field, however, it is now most frequently used as a general term to describe technical affordances that are able to deliver acoustic transparency. More specifically the new term of hear-through processing [87] has been introduced to describe systems

within active hearing devices that are used to bring about a more comprehensive or fully featured acoustic transparency.

Acoustically Transparent Headphones

Acoustically transparent headphones are typically either extra-, circum- or supra-aural headphones [75] of open back acoustical design [88] that exhibit an open acoustic characteristic. Schlieper et al. defined headphones that possess an open acoustic characteristic as imposing:

> *"... the same acoustic load on the ear canal as in the case without headphones, which can be described as the radiation impedance of free air seen from the inside of the ear."* [88, p.1]

Meyer-Kahlen et al. identified the AKG K1000 circum-aural headphone (pictured in Fig. 3.7) as having frequently been used in AR research studies due to its high degree of acoustic openness [78]. In 2018 Schlieper et al. conducted a study examining three proposed methods for the determination of headphone openness [88]. The study proposed the methods of Pressure Division Ratio (PDR), which defines the openness of the headphone as a measurement of sound pressure at the entrance of the external ear and ear canal, Headphone Selection Criterion (HSC), which builds on PDR to also include sound pressure measurement at the eardrum, and the Acoustic Impedance Curve (AIC), where acoustic impedance is measured using a Kundt tube and an artificial dummy outer ear. The results demonstrated that all methods performed well with some being better suited to certain types of headphone acoustic design than others. The authors also defined Occlusion Index as a metric to represent the estimation of headphone openness in a single value [88].

In 2021 Schneiderwind et al. [73] conducted a technical evaluation of four open back and three extra-aural headphones commonly used to deliver AAR experiences. The study concluded that even the best performing extra-aural headphones, which caused the least significant perceptual and physical corruption, were perceived as being significantly different to the open ear reference.

Satonger et al. conducted a study investigating the influence of various passive headphones over a listener's ability to perceive and localise external sound sources [89]. The first section of the study

involved the collection of transfer function measurements to assess the filtering effect of several open and closed back headphones on external sound source perception. The second section constituted a listening test assessing the ability of ten participants to localise sound produced by an array of 12 loudspeakers across the horizontal plane both with and without headphones passively occluding the acoustic signal. The study concluded that wearing headphones had the effect of introducing spectral error in the measured HRTFs, with the largest spectral interaural level difference of 26.52 dB being observed in the closed back headphones and equivalent to a significant shift in perceived sound source direction. It also found that passively wearing headphones reduced the ability of listeners to accurately localise external sound sources by a standard deviation error of 0.6°, when compared to not wearing headphones [89].

This research demonstrates that in addition to occluding the listener's perception of the level of external sound sources, wearing circum-aural headphones passively can also diminish the accuracy with which users are able to localise them. As such it can be observed that passive acoustically transparent headphones generally provide lower levels of acoustic transparency than bone conduction headsets and wearable non-contact directional speakers.

3.1.2 Active External Sound Modification Functionality in Auditory Headsets

The previous section discussed the most common types of auditory headsets that are typically used in the mobile consumption of audio. This review illustrated the wide array of different physical form factors that auditory headsets can exhibit. It also demonstrated how the physical form factor of a headset determines the extent to which the listener's ear canals are passively occluded by the device, which in turn has a large effect on how much external sound is able to reach the listener's ear drums. The following section will expand on this to discuss the various active features available in auditory headsets that provide users with a means of control over how easily they are able to perceive external sound in their surrounding environment. Within the context of this thesis these active headset features are collectively referred to as external sound modification functionality.

Active Acoustically Transparent Auditory Headsets

The auditory headsets discussed in the previous section provide passive acoustic transparency by leaving the ear canal unoccluded and open to receive external sound. In contrast to this there is also a subset of devices that aim to provide the user with an acoustically transparent listening experience in instances where the ear canal is partially or fully occluded by the auditory headset. These devices usually employ microphones mounted on the enclosures at either ear of the headset to pickup external sound, process it, and feed it through to the headset drivers such that the combination of the occluded and processed sound is perceived as identical, or similar to open ear listening. This functionality is most commonly referred to as hear-through processing [87], but is also referred to as hear-through equalisation [10, 90], and also pass-through [91]. It is employed in many different types of headphone and earphone devices to provide users with an awareness of external sound. In recent years hear-through functionality has come to be increasingly closely associated with intra-concha earphone devices such as true wireless earbuds and hearables, both of which are discussed in more detail later in this section.

Hear-Through Processing

In 2020 Gupta et al. conducted a review of the challenges in the field of hear-through techniques for acoustic transparency in hearables [10]. Hear-through equalisation research was identified as being of particular relevance to acoustic transparency in hearables due to the technical challenges caused by semi-open in-ear and closed in-ear device form factors that partially or completely block the ear canal. Blocking the ear canal has the effect of stopping the quarter wavelength resonance caused by the open ear, instead causing it to be replaced by a half resonance. The first peak of the quarter resonance is approximately between 2-4 kHz whereas the equivalent first peak for the half resonance is 5-10 kHz [92]. This shift in resonant frequencies is a significant contributing factor to the spectral differences' users perceive when passively listening to external sound sources through closed in-ear hearables.

As described in the previous section, most devices that implement hear-through processing do so by picking up and processing external sound such that, when combined with the leaked real sound, the result at the ear drum is perceived as equal to the equivalent open ear response. Gupta et al.

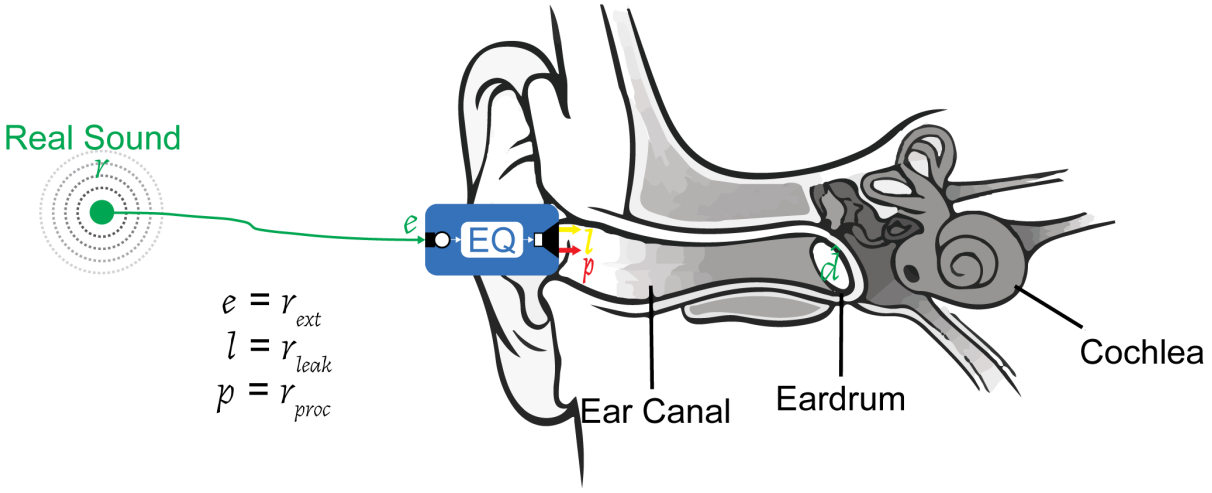


FIGURE 3.10: *Diagram showing a generic hear-through equalisation block diagram that illustrates how hear-through works in an in-ear earphone device, after [10] and [7].*

[10] provided an overview of how hear-through equalisation is typically implemented in hearable devices that employ microphones as described above. The following explanation of how hear-through equalisation typically works in this way in wireless earbud and hearable devices is based on this overview [10].

Fig. 3.10 illustrates the generic hear-through equalisation block diagram for wireless earbuds/hearables. Devices that utilise this form of hear-through equalisation must have at least one microphone mounted externally on the enclosure of either earphone to pickup the real sound that will be processed. As such, the acoustic signal that reaches the listener's ear drum $\hat{d}$ is modelled as the acoustical sum of the leaked real signal, r_{leak} and the processed real signal, r_{proc} [10]:

$$\begin{aligned} \hat{d} &= r_{leak} + r_{proc} \\ &= r * h_{leak} + r * h_{ext} * EQ * hp \end{aligned} \tag{3.1}$$

In the above equation r represents the real sound source, h_{leak} is the leaked real signal response, h_{ext} is the the response at the external microphone mounted on the earphone enclosure, EQ is the hear-through equalisation filter, and hp represents the device's driver response. The validity of this equation is subject to the assumptions that all system elements are linear and time-invariant. To determine the

necessary design of the hear-through equalisation filter, the acoustical signal received at the listener's ear drum $\hat{d}$ must closely match the equivalent open ear response, which is defined as h_{ref}.

$$r * h_{ref} \cong r * h_{leak} + r * h_{ext} * EQ * hp \tag{3.2}$$

In this equation $*$ represents convolution in the time domain. For the frequency domain the equation (3.2) can be rewritten as:

$$H_{ref}(f) \equiv H_{leak}(f) + H_{ext}(f)EQ(f)HP(f) \tag{3.3}$$

Therefore it is necessary for both equations (3.2) and (3.3) to be solved to determine the intended design of the hear-through equalisation filter such that the signal measured at the ear drum is a close match to the equivalent open ear signal. To calculate the design of the filter in this way the following information is required: the response measured at the external microphone mounting point on the earphone enclosure; the response of the device's driver; the open ear response measured in, or at the entrance to, the ear canal; and the leaked real signal response occluded by the device, measured in the ear canal. The extent to which the real sound is able to leak through to be received at the eardrum is of course heavily dependant on the quality of the device's physical fit in the ear.

Gupta et al. [10] also highlighted three studies separately proposing the application of individualised [93] and adaptive [94] cascading parametric equalisation filter systems to the hear-through signal in order to negate the spectral colouration caused by closed in-ear hearables and increase acoustic transparency. The system proposed by Riikonen et al. in 2008 included a mixer and three cascading adjustable parametric analogue filters, allowing an individualised equalisation frequency response curve [93]. The nature of the particular closed in-ear headphone that was used meant there was some leakage of the natural sound around the device. The consequent latency caused by the natural leakage summing with the delayed hear-through signal resulted in comb filtering, diminishing the level of acoustic transparency achieved. An evaluation of the system comprising of only four participants

found that the adjustable equalisation system required extensive time, resources and measurement for each individual user. In a comparison between the use of individualised and generic equalisation settings participants were found to generally prefer the generic curve. As such the authors recommend future studies investigate the use of adaptive filtering to deliver individualised calibration in real time [93].

Natural listening is a related concept to hear-through equalisation that uses an adaptive filtering system to maximise the openness and acoustic transparency of headphones [47]. This system also includes binaural reference and error microphones to filter the incoming external signal in order to render the influence of the headphone transfer function [95] such that it is virtually imperceptible. Here, natural listening headphone systems are frequently used to deliver virtual and acoustically transparent real sound sources concurrently in AAR experiences. This also represents one of several acoustically transparent headphone control strategies that is referred to by the term *hear-through processing* [87].

True Wireless Earbuds

True wireless earbuds are intra-concha earphone devices that are worn inside a user's ear canals, use wireless technology protocols such as Bluetooth to receive audio data from external hardware devices, and are not physically connected between left and right earbuds. They are now amongst the most popular devices used to listen to media with a smartphone, with a 2022 global consumer behaviours report finding that 41% of smartphone users owned a pair of true wireless earbuds with a further 20% intending to purchase a pair in the next year [96].

One of the main drivers in the rise in popularity of true wireless earbuds is the trend of consumers spending more time listening to audio content in a typical day. Shrinking form factors and improvements in battery life performance mean consumers are able to comfortably listen for longer periods of time. As an extension of a user's smartphone device true wireless earbuds are frequently used for much more than just listening to music. They are also used to play video games, watch visual media, take video/voice calls, and track user fitness/health, with listeners typically spending at least an hour a day engaging in each activity [96].

Hearables

The term *hearables* was originally conceived by Hunn in 2014 to describe wireless wearable devices worn inside the user's ears [97], then further defined in 2016 to refer to:

> *"... anything that fits in or on an ear that contains a wireless link, whether that's for audio, or remote control of audio augmentation..."* [98, p.2]

Therefore, in accordance with these definitions, all devices that are classified as true wireless earbuds would also be classified as hearables. However, not all hearable devices are also classified as true wireless earbuds, with only hearable devices that are worn inside the concha also being able to be classified as true wireless earbuds. Hearables form part of the wider field of wearables within wearable computing and are also closely related to earables and the field of earables computing, as such hearables can also be considered a manifestation of ubiquitous computing [99]. Earables are computing devices worn inside the ear that facilitate a wide variety of intelligent audio [100], biometric [101], and behavioural analytics [102] sensing capabilities. Both hearables [103], earables, and true wireless earbuds [104] can contain a wide array of signal processing, multimodal sensing, and audio related capabilities including beamforming, selective environmental noise filtering, voice activation, hearing assistance, and headtracking [105].

In 2019 Crum presented a paper exploring the potential future of multimodal applications for hearables [106]. The potential use case scenario of SuperHearing was proposed to address the challenge of individual sound source perception within a complex auditory scene, also referred to as the cocktail party problem [107]. Bearing similarity to Facebook Reality Lab's work on the concept of perceptual superpowers [108], SuperHearing proposes the use of brain wave sensing in hearables to detect when a user is having difficulty hearing those they are conversing with in a crowded and complex auditory environment. The hearable would then be able to automatically identify which sound sources in the environment constituted the user's friends using audio fingerprinting [109]. A combination of beamforming, adaptive filtering and active noise cancellation could then be used by the hearable to enhance and focus the user's perceptual hearing ability, allowing them to continue their conversation uninterrupted [106].

In 2017 Williams et al. proposed the concept of bio-responsive control for music using earable computing [110]. The authors state that existing human-computer interaction (HCI) in sound and music computing has tended to be quite tactile and explicit in nature, with common user interface tools such as mice, keyboards, control surfaces and mixing desks all utilising touch as the primary interaction modality. This is contrasted with new forms of more intuitive and fluid HCI manifested as bio-physiological interfacing. The bio sensors that facilitate this form of interfacing through monitoring of brain activity, heart rate and skin response are now beginning to become accessible to consumers via wearables and now earables.

The authors proposed that this bio-physiological human-computer interfacing, could be applied to the delivery of adaptive music and sound experiences. They also highlight a number of varied applications for this including health and well-being, adaptive video game music, indirect perceptual evaluation of audio stimuli and assisting individuals who are unable to engage with music via traditional methods due to accessibility issues in interacting with music creation.

The paper proposed a categorical approach whereby the three categories of conscious, unconscious and hybrid human-computer interaction were presented, representing forms of active and passive control. Two existing systems were referenced to demonstrate examples of conscious and unconscious interaction, the first being an emotion based generative music system controlled by galvanic skin response (unconscious) and the second being an audio mixer control system that uses brain waves to allow the user to control mixer faders (conscious). The authors state that the methods by which the bio-physiological sensor input is mapped to audio/music generation could potentially offer the user a novel affordance of HCI for music creation or curation. By way of example, a system might be capable of detecting if the user was feeling stressed and would then be able to generate calming music. It was then proposed that a machine learning system could be trained and used to analyse the consequent bio-physiological response elicited after the delivery of the generative music experience in order to iteratively improve the biosensor to music mapping performance. These examples serve to illustrate some of the potential future applications of smart technologies in hearables and earable devices.

Adjustable Hear-through and Assistive Hearing Devices

Many hearable devices now provide adjustable hear-through functionality [94], affording the listener a level of control over the extent to which their surroundings are augmented over the media. The boosting that many hearables are able to provide is such that they have been classified as *"personal sound amplification products"* under regulatory guidance, with devices such as the Apple Airpods [111] blurring the boundaries between earphone and *"assistive hearing"* devices [37]. Such devices are now often worn for longer periods throughout the day, with increased integration of smartphone features and notifications encouraging *"always-on audio"* usage [37, 79].

In 2015 Välimäki et al. [112] presented a review of applications and methods for *"assisted listening"* devices, which is a similar concept to the term *"assistive hearing"* devices used by Gilmore in 2019 [37]. In this paper Välimäki et al. outlined how consumer auditory headsets such as in-ear earbuds or over-ear headphones that feature hear-through functionality could effectively be used as a hearing aid for non hearing impaired users, helping them to perceive real sounds in environments with high levels of background noise.

Another primary application for adjustable hear-through functionality in mobile audio consumption experiences is to increase a user's sense of personal safety in public spaces. In 2022, thematic analysis of responses to a survey of user perspectives on headphone technology [72] revealed that a proportion of listeners expressed a need to be able to hear external sounds such as speech, traffic, and bus/train announcements while using headphones to consume media.

In a 2019 user attitudes and expectations survey on the use of hearables [113] 19% of participants responded that they had been in at least one potentially dangerous situation in the previous 12 months due to wearing hearables, with 7% of participants having been in at least 8 potentially dangerous situations due to wearing hearables.

In 2021 a respondent to a study on mobile media use in urban nighttime contexts revealed that their media consumption habits varied depending on the geography and context in which they were situated. The respondent stated that in one city they would only use a single earphone to consume music whilst running, while in different city they wouldn't use headphones at all when walking [114].

While in 2023 two respondents to a study on private sound environments in urban public space stated that they were less likely to use headphones in hazardous urban environments such as construction zones or areas of traffic [34].

However, if listeners' were able to possess an enhanced awareness of external sound in their surrounding environment, they might feel safe and comfortable to continue consuming audio in these circumstances. There are currently no known studies that have investigated how the perceptual effects of boosting from hear-through equipped devices might relate to an individual's ability to perceive speech, or any other potential cognitive or physiological affect.

The terminology used to describe adjustable hear-through functionality in true wireless earbuds is not standardized and therefore tends to vary between device manufacturers, with Apple having used the term *"Transparency mode"* in their Airpods Pro (2nd generation) [115] earbuds and Sony having used the term *"Ambient sound mode"* in their LinkBuds S [116] earbuds. Hear-through functionality is often provided as a separate mode, or modes, to active noise cancellation modes in devices that offer both forms of functionality. However, hear-through is sometimes presented as an extension of the noise cancellation functionality in devices. For example, rather than providing the user with a separate set of variable hear-through modes, the Bose QuietComfort II [117] earbuds instead feature an aware mode, a quiet mode, and a range of several different noise cancellation modes, which essentially serve the same purpose as variable hear-through modes in other devices. It is also often the case that devices vary in how much control a user has over the extent of environmental noise boosting in adjustable hear-through modes, depending on the intended use cases, and or technological limitations of the device.

The underlying technology used to provide adjustable hear-through in commercially available auditory headsets tends to be proprietary and not widely documented. However, in 2016 Liski et al. [94] detailed a method that used three cascading parametric digital filters and an isolation estimation algorithm to provide an adaptive hear-through equalisation system. The system also included a user adjustable control providing overall attenuation of transparency. The performance was assessed with measurement and simulations and compared against a fixed parameter equivalent. Subtle comb filtering due to latency was also detected when the volume of the hear-through signal was lowered below a certain threshold, resulting in natural sound leakage becoming more easily perceptible. However, the

application of an additional shelving filter was able to negate this issue. Whilst the adaptive algorithm was found to be able to improve equalisation in the frontal sector, external sounds from other directions of arrival were subject to some colouration [94].

Perceptual Affects of Acoustic Transparency in Auditory Headsets

In 2020 Denk et al. [118] conducted a technical evaluation investigating the acoustic transparency of hear-through modes in seven hearables [87] and two bespoke research devices. Directionally resolved impulse response measurements were carried out for each device using a KEMAR dummy head. A wide variety of performance was observed across the devices, with the best examples providing a very good account of the open ear conditions. However, some devices exhibited significant artefacts including frequency spectra colourations, comb filtering, level bias and severe degradation of binaural cues. This work demonstrates that there is still capacity for improvement in the technical competencies and consistency of acoustic transparency systems in hearables. Schepker et al. [36] published a companion study to this paper exploring the perceptual sound quality evaluation of acoustic transparency in the same nine devices, plus two additional research devices, finding that most of the commercially available hearables tested were not able to provide a perceptual sound quality that was considered comparable to the open ear. The study concluded that the main factor for determining the quality of hear-through performance was the similarity of the measured transfer functions with devices inserted to the equivalent open ear transfer function. There have been no known studies that have investigated the perceptual affects of external sound that is boosted, relative to the open ear, by adjustable hear-through modes in devices.

Active Noise Cancellation

In 2022 a global consumer behaviours report [96] highlighted how listeners are using active noise cancellation (ANC) [119–121] in true wireless earbuds to control the extent to which they are able to perceive their surrounding external soundscape, with 68% of consumers stating that ANC would influence their next purchase, and 16% of listeners specifically using earbuds to improve their hearing in noisy environments, doing so for an average of 1.4 hours per day [96]. The design of a basic single

channel feedforward ANC headphone system includes a reference microphone to detect the unwanted noise in the environment, a speaker that produces an anti-noise signal to provide cancellation, and a second microphone to assess the amount of cancellation error. Feedforward, feedback and hybrid control are the three basic systems of noise control that can be employed in all types of headphone. Typically, ANC headphones use separate control systems for either ear [122].

ANC is now an established feature that can be found in a wide array of domestic and industrial headphone, headset, and hearing protection devices, with its main purpose being to lower the perceptible affects of any disturbing or dangerous environmental noise that the listener may be exposed to. However, research has suggested that ANC may provide additional benefits to users, over and above the attenuation of unwanted extraneous noise.

Many studies have suggested that exposure to excessive environmental noise can lead to a range of impairments in an individual's cognitive performance [123–125] and general wellbeing [126–128]. While there have been some studies showing the positive effects passive hearing protection devices can have on wellbeing [129], most research in this area has focused on the positive effects of passive hearing protection on speech intelligibility and recognition [130–132]. There have been comparatively fewer studies investigating similar effects for ANC devices. Mueller et al [133] found that ANC can contribute to an improvement in several factors that are crucial to undisturbed and productive office work, including workers' perceived sense of privacy, ability to concentrate, and annoyance caused by co-workers' conversations. Niu et al. [134] found use of ANC can lead to an increase in word recognition when speech is masked by white noise, and Kim & Park [135] found that ANC provided improved speech reception threshold results.

Whilst the concept of active noise control headphones was originally conceived for the sole purpose of isolating the user from unwanted environmental noise in the external sound field, it has since expanded to include elements of selective acoustic transparency [79]. In 2020 Patel et al. detailed the design and production of a novel active noise control headphone system featuring directional hear-through capability [122]. This used a cross shaped microphone array to facilitate a highly directional beamformer, capable of picking up sound from within a small, focused zone whilst rejecting signal external to it. This functionality is able to capture the user's voice in acoustic isolation whilst simultaneously using ANC to reject environmental ambient noise. The ability to prioritise the

perception of individual sound sources in this and other ways is now a key functionality of many ANC headphones, serving to enhance the listener's inherent capability to selectively focus on specific sound sources.

Quantifying the External Attenuation Capabilities of Auditory Headsets

There has been a significant body of research aimed at quantifying the noise attenuation performance of both active [136–138] and passive noise reduction devices [139–141] under a variety of different test conditions. In instances where devices attenuate external sound relative to the equivalent open ear response, the extent of the attenuation can be measured and referred to as the insertion loss.

Insertion Loss

In 2022 the term *insertion loss* was defined by the European Telecommunications Standards Institute as:

> *"...a measurement (in dB) of the reduction in background noise caused by a device under test (DUT) measured at the listener's eardrum. Insertion loss may be due to the physical occlusion of a device (passive), algorithmic noise reduction (active), or a combination of the two (total)."* [91]

Total Insertion Loss

More specifically, the term *total insertion loss* can be used to quantify:

> *"...the attenuation of an ambient sound field when using an ANC device with noise cancellation enabled."* [91]

This is defined as follows:

"The Total Insertion Loss (IL_{total}) is the difference between the equivalent continuous sound levels of the noise source measured at the listener's eardrum:

1) *without wearing the ANC device(L_{OPEN}); and*
2) *when wearing the ANC device with ANC functionality enabled (L_{ON}).*

The IL_{total} is expressed as a function of frequency f, with 20 Hz $\leq f \leq$ 20 kHz, for each ear i and noise type n as:

$$IL_{total}(f,i,n) = (L_{OPEN}(f,i,n) - L_{ON}(f,i,n)) \text{ in dB, } \forall i \in (1{:}N)\text{" [91]}$$

The attenuation quantified by total insertion loss in devices that are equipped with active noise cancellation can be further broken down into the proportion of attenuation that is due to the passive occlusion of the ear canal by the device and the active noise control provided by the ANC processing.

Passive Insertion Loss

The term *passive insertion loss* can be used to quantify:

"...the portion of Total Insertion Loss attributed to the physical occlusion of the user's auditory system by an ANC device (e.g. through plugging of the ear canal)." [91]

This is defined as follows:

"The Passive Insertion Loss ($IL_{passive}$) is the difference between the equivalent continuous sound levels of the noise source measured at the listener's eardrum:

1) *without wearing the ANC device(L_{OPEN}); and*
2) *when wearing the ANC device with ANC functionality disabled (L_{OFF}).*

The $IL_{passive}$ is expressed as a function of frequency f, with 20 Hz $\leq f \leq$ 20 kHz, for each ear i and noise type n as:

$$IL_{passive}(f,i,n) = (L_{OPEN}(f,i,n) - L_{OFF}(f,i,n)) \text{ in dB, } \forall i \in (1{:}N)\text{" [91]}$$

Active Insertion Loss

The term *active insertion loss* quantifies:

> *"...the portion of Total Insertion Loss attributed to an ANC device's active noise cancellation."*

This is defined as follows:

> *"The Active Insertion Loss (IL_{active}) is the difference between the equivalent continuous sound levels of the noise source measured at the listener's eardrum:*
>
> *1) when wearing the ANC device with ANC functionality disabled (L_{OFF}); and*
> *2) when wearing the ANC device with ANC functionality enables (L_{ON}).*
>
> *The IL_{active} is expressed as a function of frequency f, with 20 Hz $\leq f \leq$ 20 kHz, for each ear i and noise type n as:*
>
> $$IL_{active}(f,i,n) = (L_{OFF}(f,i,n) - L_{OFF}(f,i,n)) \text{ in dB, } \forall i \in (1{:}N)\text{" [91]}$$

Quantifying the External Amplification Capabilities of Auditory Headsets

In instances where hearable or assistive hearing devices amplify external sound relative to the open ear, the extent of this amplification can be measured and referred to as the total insertion gain.

Total Insertion Gain

In 2022 the term *total insertion gain* was defined by the European Telecommunications Standards Institute as:

> *"...a frequency dependent measurement (in dB) of the pass-through amplification of a sound source or sound field presented to the user when using an ANC device in pass-through mode. It is computed as the difference between a test signal measured at the drum*

reference position (DRP) with pass-through mode enabled and the same signal measured with an open ear canal." [91]

Total Ambient Noise Insertion Gain

Furthermore, the term *total ambient noise insertion gain* can be used to quantify:

"...the amplification of an ambient noise field when using an ANC device in pass-through mode." [91]

This is defined as follows:

"The Total Ambient Noise Insertion Gain ($IG_{noise,total}$) is the difference between the equivalent continuous sound levels of a diffuse sound field measured at the listener's eardrum:

1) *when wearing the ANC device with pass-through mode enables (L_{PT}); and*
2) *without wearing the ANC device (L_{OPEN}).*

The $IG_{noise,total}$ is expressed as a function of frequency f, with 20 Hz $\leq f \leq$ 20 kHz, for each ear i and noise type n as:

$$IG_{noise,total}(f,i,n) = (L_{PT}(f,i,n) - L_{OPEN}(f,i,n)) \text{ in dB, } \forall i \in (1{:}N)\text{" [91]}$$

Use of the terms total insertion gain and total ambient noise insertion gain is congruent with use of the related terms total insertion loss, passive insertion loss, and active insertion loss [91].

3.1.2.1 Speech Intelligibility Metrics

Speech intelligibility metrics can be used as a way of assessing the impact of noise as a masking signal on the ability of listeners to understand speech signals. Various speech intelligibility metrics

have been developed for this purpose [142], with two of the most widely used objective measures being the Speech Intelligibility Index (SII) [143] and the Speech Transmission Index (STI) [144]. The SII is defined as a measure that correlates with the ability to understand speech in difficult listening situations. It is calculated as a ratio of the spoken information that is available to the listener on a scale of 1.0 to 0.0, where 1.0 indicates all spoken information is available and 0.0 indicates none [142]. The STI is defined as a measure of the extent to which speech envelope modulations are preserved in degraded listening environments [144], with the quality of speech transmission also expressed as a value between 0.0 and 1.0 [145].

3.2 Listener Engagement and Mobile Audio Consumption

In a 2020 study, Herrmann and Johnsrude [146] defined the term *"engagement"* as:

> *"...a term used in psychology and cognitive neuroscience to describe the recruitment of resources for an activity."* [147, 148]

Hermann and Johnsrude went on to clarify that within the context of their study [146], on absorption and enjoyment during listening to acoustically masked stories, the term engagement should be understood as:

> *"...a superordinate term capturing a range of immersion experiences with a narrative or story..."*

In the context of the present thesis the term *listener engagement* is intended as a subordinate term to collectively refer to any and all of the different ways in which a listener may be inclined to interact with, and or reflect upon, their listening experience when consuming audio-based media. This could include everything from quantitative measures of how frequently or under what circumstances a listener engages in media consumption, to empirical study of how individuals' perceptions of the listening experience in different ways, right across to the qualitative assessment of listeners' engagement with the narrative of the media itself.

The following section will set the scene for the remainder of this chapter by outlining a historical overview of technological advancements in devices used for the mobile consumption of audio by consumers, from the first example of individuals listening to audio using a portable device in the 1950s, through the initial period of significant growth in mobile audio consumption during the 1980s, and finally the more recent technological advancements of the 2010s that have paved the way for the contemporary research that is discussed in the rest of this thesis.

3.2.1 A Historical Overview of Mobile Audio Consumption

The concept of listeners using headphones to separate themselves from their surrounding environment has been documented for at least a century. In 1925 an advert for Brandes radio headsets (shown in Fig. 3.11) asserted that prospective customers needed a headset to: *"...listen-in without disturbing others..."* and *"...shut out the noise in the room..."*. This was of course in a time when consumers of radio or media would only ever use headphones, or headsets, to listen in the relative quiet of indoor, usually domestic, settings.

Regency TR-1 Transistor Radio

Released in 1954, the first commercially available portable device that allowed users to consume audio over an auditory headset was arguably the Regency TR-1 Transistor Radio (shown in Fig. 3.13). Being a radio it only allowed users to consume audio lineally [149] with the only audio selection control available to users being the choice of radio station, selected via tuning of radio frequency. The TR-1 featured a built in speaker, but crucially users were also able to listen using a single earphone (Model EP-1, shown in Fig. 3.12), which was available for users to purchase as an optional extra. When this earphone was plugged in it would automatically silence output from the built in speaker [150]. Suddenly, consumers were provided with the means to privately consume audio anywhere they desired, enabling a new exploration of listening affordances in an almost infinite array of different environmental contexts. The notion that use of the EP-1 earphone offered users a *"private"* listening experience was put forward by the manufacturer in owners manuals for the TR-1 line of radios. This is evidenced by the following extract taken from the manual for the Regency TR-1G radio:

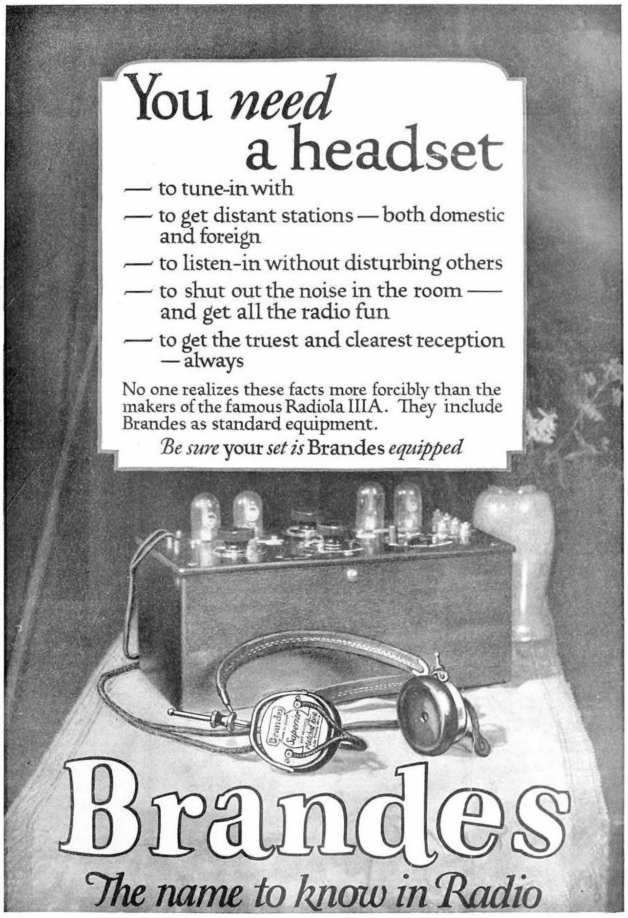


FIGURE 3.11: *An advert for Brandes radio headsets from 1925 [11].*

> *"Your Regency TR-1G has provision for the use of a special earphone (Regency Type EP-1, available from your dealer at $7.50). When the connecting plug is inserted in the small hole at left side of receiver, the loud speaker is automatically silenced and, "private" reception is obtained."* [150]

This *"private" reception* referred to the use of the earphone to allow the user to listen to their radio without others in their surroundings also being able to hear. This also allowed the user to listen without disturbing others in their surroundings. However, being only a small single earphone, users of the EP-1 would still of had a clear awareness of environmental noise in their surroundings. Nevertheless the Regency TR-1 radio, together with the EP-1 earphone, was the first step in the subsequent evolution of portable listening devices that allow users to listen to audio content augmented over the external sounds of their surrounding environment. Key developments in the evolution of portable listening devices used for the mobile consumption of audio are shown in Fig. 3.13.

FIGURE 3.12: *A photograph of the Regency EP-1 earphone for use with the Regency TR-1 transistor radio, adapted from [12].*

Sony Walkman TPS-L2 Cassette Player

Within academia, the concept of headphone listeners occupying a *"private acoustic space"* when listening to media over headphones has been discussed since at least 1977 with Schafer's use of the term to describe the intimacy of headphone listening [51]. However, it wasn't until the release of the first personal stereo device in 1979 (the Sony Walkman TPS-L2, shown in Fig. 3.13) that listeners were able to practice mobile media consumption in a way that is not dissimilar to contemporary listening practice today. In contrast to the Regency portable radio, the Walkman was the first portable listening device that allowed users the affordance of listening to recorded media with headphones. Until this point users were limited by the linear structure and choice of the broadcast radio medium. Now, with the advent of the Walkman and compact cassette tapes, users could not only exercise choice over which commercially produced cassette releases they wished to listen to, they could go further and record their own mixtapes consisting of whatever combination of self or commercially produced recordings they desired.

The TPS-L2 also included a pair of stereo on-ear headphones (Model MDR-3L2, shown in Fig. 3.14) as standard, meaning that users of the Walkman were less easily able to perceive external sounds in their surrounding environment through the passive occlusion of the headphones. In fact, the designers of the first Walkman were so concerned that listeners would be isolated from their surroundings that they included two 3.5 mm headphone jack sockets and a *"hot line"* button, which lowered the

FIGURE 3.13: *Timeline of key technological developments in devices used for the mobile consumption of audio. Source images adapted from [13–19].*

FIGURE 3.14: *A photograph of the Sony MDR 3L2 headphones for use with the Sony Walkman TPS-L2 cassette player [20].*

volume of the media and activated a microphone built-in to the body of the Walkman, allowing two users to listen in a shared listening experience, and making it easier for the users to communicate with each other while still listening. Another possible application of the hot line function was that solitary listeners could use it to increase their ability to perceive external sounds in their surrounding environment. In this sense the hot line function could potentially be thought of as the earliest example of hear-through functionality in a portable listening device, hear-through being a term that would only begin to come into popular use to describe an equivalent feature in earphones some 40 years later. The hot line button was removed from subsequent Walkman models, however the new affordances of portable listening that the Walkman enabled contributed to an explosion in the popularity of portable cassette players throughout the 1980s.

This explosion in the popularity of mobile audio consumption led scholars to begin studying how listening to media over headphones in public spaces shaped how individuals perceived and interacted with their surroundings, and other individuals located within them. The first notable example of this was *"The Walkman Effect"* [29] which described the control that a listener is granted over their environment through the curation of an individual zone of headphone listening.

Sony Discman D-50 CD Player and Sony MZ-1 MiniDisc Player

From the second half of the 1980's through to the later stages of the 1990's the original Walkman was joined by a succession of new portable cassette player devices from Sony and other manufacturers. The first examples of portable media devices that utilised digital media rather than tape were also released in this period. In 1984 Sony released the first portable compact disc (CD) player (the Sony Discman D-50, shown in Fig. 3.13) which, when paired with an optional battery pack adapter, allowed users the same affordances of mobile audio consumption available to users of the Walkman, but with the added convenience of CDs as the media format. Similarly, in 1992 Sony released the first portable MiniDisc player (the Sony MZ-1 MiniDisc player, shown in Fig. 3.13), which allowed users of the MiniDisc media format to also engage in the mobile consumption of audio. As was the case with Walkmans, both the Discman D-50 CD player and MZ-1 MiniDisc player featured 3.5 mm mini jack socket headphone output connections, meaning that they could be used with a wide array of different headphones and earphones, each of which might possess drastically different acoustical designs, which would in turn affect the extent to which the user was able to perceive external sounds in their surrounding environment while listening.

SaeHan MPMan F10 Digital Audio Player and Apple iPod (1st Generation) Digital Audio Player

In 1998 SaeHan released the first commercially available digital audio player, the SaeHan MPMan F10, shown in Fig. 3.13. While not being especially commercially successful at the time, this represented arguably the single most significant advancement in the history of technological developments in the mobile consumption of audio. The MPMan utilised solid state flash memory, meaning it could store and playback audio encoded in the new digital MP3 file format. Despite only featuring 32 mb of flash memory storage capacity, meaning there was only space for several tracks encoded at 128 kb per second, the MPMan was the first device to demonstrate the viability of MP3 encoding being used in the mobile consumption of audio. Shortly after the release of the MPMan there ensued a flood of similar solid state digital audio player devices, each offering successive improvements in the amount of data storage capacity available and hence the quantity of audio that a user could store and playback

FIGURE 3.15: *A photograph of the Apple earphones for use with the Apple iPod (1st generation) digital audio player [21].*

at any one time. However, it was the release of the Apple iPod in 2001 that truly kick-started the digital audio player revolution. The 1st generation iPod featured a mechanical hard drive with a capacity of 5 GB, meaning users could store and playback approximately 1000 songs worth of audio. This dwarfed the maximum storage capacities available in compact cassette, CD, and MiniDisc formats, managing to do so in a more compact form factor that was small enough to easily fit inside a user's pocket. The iPod also played a significant role in further popularising the use of in-ear earphones in the mobile consumption of audio. Although in-ear earphones had already been available to consumers and used with devices such as the later Walkmans, Discmans, MiniDisc players, and early mobile phone handsets, the earphones that came packaged with the 1st generation iPod (shown in Fig. 3.15), and the subsequent marketing campaigns based around them, that would go on to cement wired in-ear earphones as the dominant auditory headset used in the mobile consumption of audio throughout the 2000s and into the 2010s.

Apple iPhone (1st Generation) Smartphone

The next major technological development affecting the mobile consumption of audio was the gradual integration of telecommunication, media playback, software application, and touchscreen features in mobile phone handsets and smartphone devices that culminated in the release of the first Apple iPhone in 2007, shown in Fig. 3.13.

Despite being marketed as a *"personal digital assistant (PDA) with communication capabilities via a*

cellular connection", the IBM Simon, released in 1994, is now widely regarded as the first smartphone device. Its features included a touchscreen, email and fax provision, an array of typical PDA features, and the ability to make phone calls [151]. Further similar devices [152] would go on to be released throughout the 1990s and early 2000s, including the Ericsson R380s in 1999, which was the first device to be commercially branded as a smartphone [153].

In 2000, the Samsung SPH-M100 (Uproar) became the first mobile phone device to incorporate digital audio player functionality through the provision of MP3 file playback [154]. If a user was using the included wired in-ear earphones to listen to music and they received an incoming phone call, the music would pause, a ringtone would sound, they would be able to answer the call using a remote attached to the cable of the earphones. They would then be able to communicate with the caller using the microphone attached to the earphones cable. Despite this, however, the SPH-M100 didn't include the wider set of features required for it to now be classified as an example of a smartphone.

The early smartphones of the 1990s and 2000s were predominately marketed towards business users as devices to assist with productivity throughout the workday [152], and it wasn't until the release of the iPhone in 2007 that smartphones first broke through into popular consumer use.

The subsequent explosion in popularity of further iPhone models, and competing smartphones from other manufacturers throughout the 2010s, spearheaded the technological and cultural revolution that would lead to the current ubiquity of smartphones now seen in present day society. In some respects these smartphones very similar to the first digital audio players in their ability to store and playback digital audio files, and to later iPods in the ability to store and watch video content. However, it was the elegant integration of telecommunication features such as the ability to make phone calls, text and email contacts, and browse the internet, paired with advances in touchscreen technology, and software application support that would lead to a sea change in the affordances of mobile audio consumption.

Later iPhone models and many other smartphone devices would build on the handsfree call taking functionality introduced by the Samsung SPH-M100 by providing further integration of other smartphone system sounds and features. This marked the beginning of a trend towards smartphones and auditory headsets gradually becoming more and more closely assimilated with one another.

In 2016 Apple released the iPhone 7, which was the first iPhone not to feature a 3.5 mm headphone jack socket. While it was still possible for users to listen with wired earphones using a pair of Lightning

FIGURE 3.16: *A photograph of the Apple AirPods (1st generation) earphones for use with the Apple iPhone smartphones and other compatible devices [22].*

FIGURE 3.17: *A photograph of the Apple AirPods Pro (1st generation) earphones for use with the Apple iPhone smartphones and other compatible devices [23].*

connection compatible earphones that was included with the phone, the change was seen as a deliberate move to encourage the use of Apple's new wireless AirPod earphones, which were released in tandem with the iPhone 7. The first generation AirPods (shown in Fig. 3.16) featured enhanced integration with the iPhone 7 and other compatible Apple devices through the use of a built-in processor and Bluetooth connectivity. They also featured beamforming microphones in either earbud that could be used to filter out background noise and focus on the user's voice, improving outgoing audio quality of phone calls and the performance of voice control integration with the iPhone's digital assistant SIRI. Since the release of the iPhone 7, every subsequent version of the iPhone has also not featured a 3.5 mm headphone jack socket, as is also now the case with many smartphones from other manufacturers.

In 2019 Apple released the Airpods Pro (shown in Fig. 3.17), which were the first Apple earphones to include audio features such as Active Noise Cancellation and transparency mode. In this respect the Airpod Pro's transparency mode is very similar in design and intended application to hear-through processing functionality and by extension the aforementioned hot line button, included in the very first

personal stereo, the Sony Walkman TPS-L2 released 40 years previously.

3.2.2 Audio-based Media Listened to in the Mobile Consumption of Audio

Within the context of this thesis, the term *speech-based media* is intended to refer to audio-based media that is predominately speech-based such as podcasts, audio books, live radio, or on-demand radio, but also including examples with speech content such as music or sound design alongside the speech content. While it is recognised that radio programming can also be primarily music based, it has been included within the scope of speech-based media for this thesis as a large proportion of radio programming such as dramas, documentaries, and talk radio is predominately speech-based.

While the focus of research presented in this thesis is primarily on mobile listeners' engagement with speech-based media and podcasts more specifically, it is important to consider the role of music listening and other audio-based media in the development of contemporary mobile audio consumption practices and behaviours, and additionally how the popularity of music consumption compares with speech-based media today.

As discussed earlier in this chapter, the first consumer device to allow consumption of mobile audio over an auditory headset was the Regency TR-1 transistor radio, released in 1954. Using this device listeners were able to tune into an assortment of different radio stations, each of which would have provided the listener with an array of different audio content to choose from. Therefore, it can be concluded that right from the initial stages of mobile audio consumption, users have been able to listen to both music and speech-based media.

The Sony Walkman TPS-L2 compact cassette player, the Sony Discman D-50 CD, and the Sony MZ-1 MiniDisc player helped to bring about an explosion in the popularity of physical media being used for mobile audio consumption, with music being the driving force behind this rise in popularity of mobile audio consumption.

The advent of portable digital audio players and the growth of both legitimate digital media download stores such as iTunes, and illegitimate peer-to-peer file sharing websites, further contributed to the popularity of mobile music consumption, and this trend has continued with the advent of audio streaming services such as possible Spotify. However, the arrival of podcasts in the mid 2000s, and

their subsequent growth in popularity throughout the 2010s has led to a significant increase in the amount of speech-based media being consumed using mobile devices and auditory headsets.

This rise in the popularity of podcasting is apparent in the findings of the Infinite Dial report, an American digital media consumer behaviours survey conducted in 2024 that comprised of over 1000 participants [155]. The report found that the percentage of those surveyed who were aware of podcasts had risen from 37% in 2008 to 84% in 2024, while the percentage of people who reported ever having listened to a podcast rose from 18% in 2008 to 67%. Similarly, the percentage of people who had listened to a podcast in the last month had risen from 9% in 2008 to 47% in 2024, while the percentage of people who had listened to a podcast in the last week rose from 8% in 2008 to 34% in 2024.

In 2023 RAJAR published the latest results of their MIDAS Measurement of Internet Delivered Audio Services survey [27], which collected data from 2410 audio consumers aged 15 and over in the UK. The results showed that across the range of digital audio consumed by participants in the week prior to the survey date using all devices, 88% listened to live radio, 32% consumed music using on-demand streaming services such as Spotify or Apple music, 11% consumed music using digital downloads, while 22% had listened to podcasts, 10% had listened to Audiobooks, and 9% had listened to listen again/catch-up radio.

The study also reported the share of listening for some audio mediums split across different playback devices. Within this it showed that smartphones were used to listen to 76% of podcasts that were consumed in the the week prior to the survey date, 58% of listen again/catch-up radio, 51% of on-demand music, and 4% of the live radio that was consumed. When considered together, further analysis of these findings reveal that 16.7% of all participants surveyed used a smartphone to listen to podcasts in the week prior to the survey date, compared with 16.3% for on-demand music, 5.2% for listen again/catch-up radio, and 3.5% for live radio.

Despite these findings being based only on the results of a single UK consumer survey, and the fact that it is being known what proportion of smartphone audio consumption took place in the home as opposed to mobile listening situations, these findings suggest that currently music and speech-based media attract similar levels of listening for audio consumed using streaming services with smartphones. Based on the aforementioned findings of the 2024 Infinite Dial report [155], it is clear that this equivalence in

the popularity of music and speech-based media consumed using smartphones would not have been possible without the recent rise in popularity of podcast consumption amongst listeners.

3.2.3 Aspects and Assessment of Listener Engagement with Spoken Narratives

This section will highlight and discuss a selection of metrics that can be used to qualitatively assess listener engagement with spoken narratives and other psychological aspects of the listening experience. The central theme that links the metrics discussed in this section is the concept of an individual experiencing such a high level of engagement in the listening experience that they begin to lose awareness of their surroundings, eventually reaching a state where their attention is solely focused on the narrative of the speech-based media that they are listening to. It is important to state that this should not be considered a definitive list of all metrics present in the literature, but rather an overview of some of the key concepts and measures that readers of this thesis should be familiar with in order to understand the original research presented in the fourth and sixth chapters of this thesis.

Narrative Absorption

In 2021 Kuijpers et al. [156] provided the following definition for the term *narrative absorption*:

> *"Narrative absorption is an umbrella term for a specific mental state that, in general, can be described as altered embodied, cognitive, and emotional processes of being invested into the content of a fictional story..."*

The concept of absorption was born out of research on the concept of flow, first originating in 1975 [157] and intended for the study of engagement with *"...challenging activities that require skill"*, such as sports or literary reading [158]. The concept of flow, and how it relates to other concepts within narrative absorption, is discussed later in this chapter. However, the development of absorption as concept also has origins with research conducted in 1988 on engagement in the practice of *ludic reading* and parallels observed with various altered states of mind [159]. Subsequent research from the 1990s and continuing today has further focused on the concept of absorption in relation to different

forms of narrative media, including films, interactive stories, and books, in addition to other forms of engagement with non-narrative specific activities such as virtual reality, computer software, and games [156].

In their overview of narrative absorption and related concepts, Kuijpers et al. [156] highlighted *presence* and *cognitive absorption* as two key concepts related to absorption in non-narrative and narrative media experiences. Within this context presence was defined as:

> *"...the perceptual illusion of being in a non-mediated virtual reality environment."* [160–162]

The concept of presence is discussed in more detail later in this chapter, while cognitive absorption was defined as:

> *"...an altered sense of time, increased attention, enjoyment and curiosity while using computer software or browsing the internet."* [160, 163]

Kuijpers et al. [156] also emphasised that while concepts such as presence and flow can be freely applied to absorption-like experiences with a wide range of different media, the concept of narrative absorption is specifically related to absorption-like experiences brought about by engagement with stories. Moreover, Kuijpers et al. [156] identified the concepts of *story world absorption* [160], *transportion* [164, 165], *immersion* [166, 167], and *narrative engagement* [168] specifically as describing aspects of engagement with narrative experiences. Narrative engagement is most closely associated with engagement in audiovisual experiences, whereas story world absorption, transportation, and immersion originated from engagement with stories consumed via textual means [156]. These concepts are discussed in more detail later in this chapter and throughout the thesis. The interested reader is referred to [156] for further discussion of literature relating to the concept of narrative absorption.

There have been many different methods and scales that have been developed for the purposes of measuring different aspects of narrative absorption [156]. These include the narrative engagement scale [168], the transportation scale [165], the story world absorption scale [160], the reading flow short scale [169], and the the absorption-like states questionnaire [170]. In 2020 Herrmann and

Johnsrude [146] created the Narrative Absorption Scale (NAS) for the assessment of engagement with narratives in listening experiences. This incorporated items from the story world absorption scale [160], originally designed for assessment of engagement with written narratives, and a series of other scales related to the assessment of narrative engagement [168], emotional engagement [171], and transportion [172]. The wording used in the items of these scales was adapted by Herrmann and Johnsrude in the NAS for the assessment of listening experiences [146].

Narrative Engagement, Attentional Focus, and Flow

Busselle and Bilandzic [168] developed a scale designed to measure narrative engagement in film and television viewing experiences. Four dimensions of experiential engagement in narratives were defined including narrative understanding, attentional focus, emotional engagement, and narrative presence. Within the dimension of attentional focus, a truly engaged viewer was defined as one who is unaware of their focused attention, up until the point at which their attention drifts and they are required to refocus [168, p. 341]. When an individual reaches this level of complete attentional focus on an activity they are described as experiencing *flow* with the activity. The concept of flow was developed in the field of psychology by Mihaly Csikzentmihalyi [157, 158] and is defined as a unique state of consciousness where an individual's attention is fully focused on an activity, paired with, *"a loss of conscious awareness of oneself and one's surroundings"* [168, p. 324]. It is this type of attentional engagement, applied in the context of podcast listening experiences, that represents the primary focus of the original research presented in the fourth chapter of this thesis.

Hartmeyer et al. [173] and Sony et al. [174] conducted studies in the field of auditory neuroscience that used the term *attentional engagement* to refer to a mediating factor in individuals' performance in route planning tasks and narrative comprehension, respectively, whereby one's attentional state fluctuates between different levels of focus on an external task or stimulus. Kaya and Elhilali [175] conducted a review of studies that model auditory attention. The review found that models of auditory attention can generally be classified as being based around either bottom-up or top-down attention processing. Bottom-up attention occurs in response to external stimuli in the environment that capture the listener's attention, while top-down attention is related to goal oriented attention where an individual actively focuses their attention in order to carry out a pre-planned task or activity. Further research by White

and Shah [176], Linnell et al. [177], and Berman et al. [178] has also suggested that an individual's cognitive and attentional processes can be influenced by the nature of stimuli in their surrounding physical environment.

Gaining an understanding of how environmental context may influence attentional engagement could be highly pertinent to research in the fields of object-based media [179] and related media personalisation disciplines. Gradinar et al. [180] presented a study on the use of perceptive media in the production of adaptive storytelling experiences that highlighted the weather, temperature, and time of day as factors of a listener's environmental context that could create a deeper level of personalisation, potentially leading to higher attentional engagement. The continued growth in interest around object-based media [181] has now given rise to public trials of adaptive experiences that are perceptive to audience context. In 2022 BBC Research & Development released the Adaptive Podcasting player app and web editing tool, which enable the production of audio experiences that are personalised according to data from the listener's device and elements of their surrounding environmental context [182].

Presence and Immersion

As discussed earlier, presence and immersion are both terms which can be used to describe aspects of absorption in both non-narrative and narrative media experiences. They are perhaps most synonymous with user engagement in XR experiences, however they are also highly relevant to research on non XR-based audiovisual experiences. In reference to human engagement with media and simulation technologies, Lee defined presence as constituting:

> *"...a psychological state in which virtual (para-authentic or artificial) objects are experienced as actual objects in either sensory or nonsensory ways."* [183, p.37]

Within this central concept and based on a review of existing literature, Lee defined the three subcategories of physical, social and self-presence. Physical presence, sometimes referred to as telepresence, describes the perception of real physical objects and environments manifested as virtual physical objects, either captured by sensors or modelled artificially. Social presence relates to the perception of social personification, representative of interacting with either a real human or an artificial intelligence

that exhibits human characteristics. Similarly, social presence is considered virtual when it is perceived through the prism of technology either using sensors or artificially rendered modelling. Self-presence refers to the perception of one's own self in physical reality and its virtual reconstitution again through either sensor derived or artificially synthesised technological reproduction [183].

Extensive research has been conducted into the underlying factors that can contribute to increased feelings of social presence. In 2018 Oh et al. presented a systematic review of literature published in this area [184]. The review concluded that social presence is an essential feature in interconnected VR experiences and has been shown to result in increased levels of enjoyment and social influence. The contributing factors of social presence identified were categorised separately into immersive qualities, contextual differences and individual psychological traits. The contextual factors highlighted included aspects of social psychology as observed virtually, agency, physical proximity, task type, social and identity cues [184]. This research serves to illustrate the extent to which the qualitative user perception of presence in VR can be influenced by a wide range of contextual factors that exist within the virtual environment itself.

Within the context of VR, immersion was defined by Slater & Wilbur as the extent to which technology is:

> *"...capable of delivering an inclusive, extensive, surrounding and vivid illusion of reality to the senses of a human participant."* [185, p.3]

The key distinction therefore between presence and immersion as defined by Lee [183] and Slater & Wilbur [185] respectively, is that whilst presence is derived from the user's subjective perception of authenticity, immersion is an objective measurement of the potential for a technological system to deliver the illusion of the reality within which the presence can then be perceived. However, it still remains the case that there is no consensus as to a commonly agreed definition of immersion. This was illustrated in a paper presented by Agrawal et al. in 2020, consisting of a review and analysis of the use of the term in research conducted on audiovisual experiences [186]. The authors state that existing definitions of immersion can be separated into the two categories of an individual's psychological state as a response to an experience and the objective ability for technology to deliver the experience. The authors state that the review indicates immersion as being a cognitive construct and present the following definition:

> *"Immersion is a phenomenon experienced by an individual when they are in a state of deep mental involvement in which their cognitive processes (with or without sensory stimulation) cause a shift in their attentional state such that one may experience disassociation from the awareness of the physical world."* [186, p.407]

The further defining terms of *immersive potential* and *immersive tendency* are also proposed as a basis to for quantifying immersion in future research and experiments. Immersive potential is defined as the capability of a system or content to bring about immersion and is therefore emblematic of the earlier definition proposed by Slater & Wilbur [185]. Immersive tendency describes the propensity with which an individual is able or likely to register feelings of immersion. The authors also identified five factors capable of influencing immersion as; the physical constitution of the reproduction system content; the narrative of content; the physical environment and context derived from it; factors specific to the individual such as experience, preferences and mood, and the nature of interaction between the user and experience.

In contrast to the term immersive tendency, which describes the propensity of an individual to register feelings of immersion in an experience [186], Stecker et al. [187] proposed the novel concept of *co-immersion*, which can be understood as:

> *"... the perception of real or virtual objects as contained within or belonging to a shared multisensory scene."* [187, p.1]

Co-immersion is intended to provide a means of objective assessment across all forms of virtual and real objects, not limited to sound. It can therefore be equally applied to AR containing a combination of real and virtual objects, VR formed entirely of virtual objects, and also physical reality comprised entirely of real objects. The wider purpose of co-immersion is to provide an objective measurement of accessibility and user experience in VR and AR systems, with a particular emphasis on remote interconnected communication applications. The paper also detailed a study designed to assess the user's acuity to identify and localise overlapping talkers in a VR environment based on simulated acoustic differences. Whilst the work was presented as a proof-of-concept, the results were considered sufficiently positive to recommend a potential application as a tool for the assessment of auditory spatial awareness in XR systems [187].

3.2.4 Listener Engagement in Mobile Audio Consumption Experiences

The study of listener engagement in mobile audio consumption experiences is a multi-faceted problem that spans many different research disciplines beyond that of just acoustics or audio engineering and potential research methodologies. This section will discuss literature drawn from a diverse range of disciplines that are of relevance to this study.

Auditory Bubbles and Mobile Audio Consumption

Since the advent of the first personal stereo device in 1979 scholars have studied how listening to media over headphones in public spaces shapes how individuals perceive and interact with their surroundings. The first notable example of this was *"The Walkman Effect"* [29] which described the control that a listener is granted over their environment through the cultivation of an individual zone of headphone listening. Advancements in mobile listening technologies from the portable CD player to the iPod were reflected in developments in research on mobile headphone listening. While terms such as *"soundworlds"* and *"auditory bubbles"* were primarily devised to describe how mobile headphone listening protected the individual from the stresses of their surrounding soundscape [33, 188], other scholars posited a more complex relationship between private headphone listening and the surrounding soundscape that held an appreciation of the fact that a listener's awareness of their surrounding soundscape is often not completely drowned out by the media. This concept of *"tuning out"* [32] the soundscape with the distraction of mobile media is more akin to a fluid integration of media and soundscape, rather than a means of completely isolating oneself from it.

The majority of the literature and research relating to the concept of auditory bubbles, however, has tended to focus on how the mobile consumption of audio using headphones affects the listener's relationship with their surrounding environment and other individuals located within it. This was exemplified by Bull in 2010 [33, p. 56] who described the practice of listeners using portable listening devices and headphones to construct their *"very own auditory bubble"* within the wider public soundscape.

Most mobile headphone listening studies have used qualitative research methodologies such as phenomenology, interviews, and field studies. While Schurig [35] conducted an experiment to measure how the *"auditory bubble"* affects listeners' perception of personal space, and other studies have conducted experiments to quantify the effects of environmental sound masking on headphone listening [72], no studies have performed controlled experiments to specifically investigate listeners' perception of the auditory bubble or how listening within an auditory bubble might affect listener engagement with the media itself. The advent of recent headphone features such as head-tracked spatial audio, active noise cancellation (ANC) and hear-through has brought about further developments in the ways that mobile media listeners' manage and perceive the relationship between inner and outer soundworlds. In shutting out more of the distractions from their surroundings, ANC users could be seen as retreating further into the protection of their auditory bubble, whereas hear-through users, and their desire for a greater awareness of the world around them, practice a new kind of mobile media listening, that has yet to be addressed in the literature. While some hear-through users seek full acoustic transparency, where perception of external sounds is ideally identical to open ear listening, other users desire an enhanced awareness of external sounds where parts of the frequency spectrum are boosted relative to the open ear [10]. As referred to earlier in this chapter, the boosting that many hearables are able to provide is such that they have been classified as *"personal sound amplification products"* under regulatory guidance, with devices such as the Apple Airpods blurring the boundaries between media playback and assistive hearing devices [37]. However, with very few studies having investigated how environmental noise affects listeners' perception of the media itself in mobile media consumption [189, 190], it is not known how boosting of environmental noise from hear-through processing might influence listeners' engagement with the media, or indeed how listeners' engagement is affected by the passively occluded environmental noise experienced in a traditional auditory bubble.

Listener Engagement in Mobile Music Listening Experiences

Perhaps as a consequence of music's historical role as the dominant form of media in mobile audio consumption, the majority of the research and literature concerning listener engagement in mobile audio consumption has focused specifically on mobile music listening, with much less research having being conducted on listener engagement in mobile speech-based media consumption.

Notable examples of studies that have investigated aspects of listener engagement in mobile music listening include several papers authored by Amanda Krause. These papers cover research themes including how device usage and choice relate to music listening in everyday life [191], how location relates to listeners' everyday experiences of music [192], how activity and location relate to pleasure, arousal, dominance and judgements of music [193], how location and control over music relate to the listening experience [194], and the uses and gratifications of different music formats used to consume music in everyday life [195].

Other notable examples of research concerning mobile music listening from other authors have addressed such themes as the use of MP3 players to consume music while traveling [196], the situational and technical dimensions of mobile music listening [197], how situational characteristics of the listening experience relate to psychological affects of music such as arousal, valence, and depth [198], and a principal components analysis of the functions and experience of mobile music listening [199].

These studies have been selected to illustrate the breadth of research that has been conducted on listener engagement in mobile music listening experiences. However, the results of these studies are not presented or discussed further here as research has found that there are significant differences between listener engagement in mobile music listening experiences and mobile speech-based media consumption experiences. As such, findings of research on listener engagement in mobile music listening experiences cannot be considered transferable to the study of mobile speech-based media consumption.

Listener Engagement in Mobile Music Listening and Mobile Speech-based Media Consumption Experiences

There has been very little research that has investigated how listener engagement with music differs to listener engagement with speech-based media. However, almost all of these studies have reported significant differences in aspects of listener engagement between the mobile consumption of music and speech-based media. In 2013 Wash and Dance [190] investigated the difference in listening levels

for music and speech-based media consumed on the London underground, finding that participants tended to listen to speech-based media at significantly higher volumes compared to music.

In 2015 Nyre [189] conducted a study that compared the situational fit of music, radio and podcasting in urban headphone listening experiences. The results showed that participants preferred to engage in self curative music listening experiences and were less interested in listening to discussions of social issues over the radio or engaging in educational content presented on a podcast. The results also suggested that music was perceived as more relaxing compared to podcasts and radio, which participants felt both required more concentration to fully engage with.

In 2020 Li et al. [200] used data collected from Spotify users to understand users' audio streaming habits for music and podcasts and investigate whether music and podcasts compete with one another. The results suggested that music and podcast consumption overlaps somewhat, but one was not considered a replacement for the other. Crucially the results showed that users tended to consume podcasts in different time windows than music. Users who listened to both podcasts and music also exhibited substantial differences in the frequency and duration of listening between music and podcasts. However, in 2022 Jones and Morris [201] questioned the generalisation of findings of Nyre [189] and Li et al. [200], citing the promotion and growth of podcasts consumed on music-based streaming platforms, and the rise in popularity of popular music podcasts [202].

In 2022 Rane et al. [72] conducted a survey of user perspectives on headphone technology. Thematic analysis of qualitative responses revealed that a much higher proportion of users identified aspects of their headphone listening experience that they would like to change for spoken word content compared to music content. Two aspects that were amongst those that registered the highest disparity between responses for spoken word and music content were noise cancellation and acoustic transparency, with 48.8% of participants stating that they would like to see improvement in noise cancellation when listening to music, compared to only 12.1% when listening to spoken word content. Similarly, 8.9% reported wanting to see improvement in transparency functionality when listening to music content, compared to only 0.2% for spoken word.

In a separate study in 2022 Rane et al. [203] investigated how different parameters influence users' level of distraction and quality of listening experience when consuming different media in then presence of masking from environmental noise. The results showed that participants were less distracted when

listening to pop music and more distracted when listening to classical music and radio drama content in the presence of masking from train-based environmental noise. In 2024 Bosshard et al. [204] conducted a study that analysed how 60 participants' brains responded to music streaming, podcasts, and radio, measuring subjects' attitude, attention, memory, and engagement. The results found that music streaming brought out more positive attitudes, greater levels of memory encoding, higher attention, and increased physiological arousal compared to both radio and podcasts.

Listener Engagement in Mobile Speech-based Media Consumption Experiences

In 2011 Coens et al. [205] conducted two studies comprising of 36 and 75 participants respectively, to investigate the effect of physical exercise on the ability of listeners' to learn from an educational podcast. One experimental group listened to podcast content while walking, another while jogging, and a control group listened sitting down without engaging in exercise. In study one the test instrument used consisted of three multiple choice and two open ended questions related to the subject of physiology, which was related to a course that the participants were enrolled on. Tests were then conducted to assess how much of the educational content the participants had managed to learn. In study two the instrument consisted of 11 multiple choice and four open ended questions on dyspraxia, which was of no relation to any courses the students were enrolled on. The results of study one showed no significant differences in students' ability to learn between exercise groups, however, the results of study two found that students who were seated significantly outperformed students who were jogging while listening.

In 2014 Soto-sanfiel [206] performed an experiment comprising of 327 participants to explore the effect of mobile listening on listeners' sense of spatial presence while consuming audio recordings of stories. Spatial presence is one of the dimensions within the wider concept of presence, discussed earlier in this chapter, and is defined as:

> *"...the feeling of being physically located in a virtual setting or experiencing physical objects as if they were real."* [183, 206]

The experiment comprised of a 2x2 factorial design, with two separate fictional horror themed stories and two independent variables of listening while walking in the open around buildings, and listening while seated. Participants then completed a 35 item questionnaire [207] to assess their spatial presence while listening. The results suggested that mobile listening had a detrimental effect on participants' attention, spatial situation, and high cognitive involvement, with mobility also leading to listeners concentrating less on the story, and having less of their feelings captured by the narrative of the story. The results also showed that aspects of listeners' spatial situation, their ability to imagine the layout of a place, calculate time, and develop a mental image of the places inside a story, were further impacted by listening while walking. Finally, the results suggested that mobile listening impacts the ability of listeners' to imagine things encountered in the narrative, their relation to one another within the story world, the stimulation of thought by the story, and listeners' general perception of how useful they thought the story was.

In 2015 Soto-sanfiel [208] published additional results gathered during the same 2014 study [206], reporting on further measures that had been surveyed. Analysis of these additional results found that while mobile listening led to a reduction in attention, cognitive involvement, narrative realism, identification with characters, listeners' enjoyment of the stories was not diminished by mobile listening.

It is also important to highlight the lack of any empirical research that has specifically focused on the effect of masking from environmental noise on listener engagement in mobile speech-based media consumption experiences. While the studies discussed in the previous section did touch upon how certain aspects of listeners' perception of environmental sound can influence their engagement in speech-based media experiences, these studies also considered aspects of engagement with non-speech-based media, and no studies have specifically investigated how masking from environmental noise might influence listener engagement with speech-based media alone.

In 2020 Herrmann and Johnsrude [146] conducted a study that investigated listener absorption and enjoyment while listening to acoustically masked stories. The study comprised of four experiments; the first tested the sensitivity of a narrative absorption scale that the authors had adapted for spoken stories, the second investigated the effects of acoustic masking on story absorption, the third examined the effect of thematic knowledge on story absorption, and the fourth assessed the effect of time on story

absorption. The results for experiment one showed that the adapted narrative absorption scale was sensitive in measuring listeners' absorption. Experiment two showed that while moderate masking led to an increase in listening effort, absorption and enjoyment were only minimally impacted. Experiment three found that prior thematic knowledge of the narrative increases story absorption, comprehension, and enjoyment, while lessening listening effort. Finally, the results of experiment four showed that listeners' story absorption and their motivation to continue listening increases over time, while listening effort lowers over time.

The caveat to these results, in the context of the present thesis, is that while each of these experiments performed in the study by Herrmann and Johnsrude investigated how masking from multiple human speakers affected listeners' engagement with extracts of a storytelling podcast over headphones, the podcast content was in fact being used as a proxy to represent any generic spoken word story. In each of the experiments the storyteller podcast content was presented to participants over the same pair of headphones as the 12-talker babble that was used for the multiple human speaker masking signal. In this sense, Herrmann and Johnsrude were using the solo storytellers in each of the podcast stimuli extracts to simulate a scenario whereby participants were listening to the voice of a single human speaker recount stories in front of them, as though they were in the same acoustic space, while the 12-talker masking babble signal was intended to simultaneously simulate a crowd of human speakers conversing with one another, as though they were also surrounding the participant in the same acoustic space. As such, it could not be said that the experimental method used was entirely representative of typical acoustic scenarios in mobile speech-based media consumption experiences where the masking noise is external and affected by the occlusion of the auditory headset.

The results of the experiments conducted by Herrmann and Johnsrude represent an important contribution that could potentially be adapted and used as a basis and for the development of future studies that seek to specifically investigate listener engagement in mobile speech-based media consumption experiences. It is also notable that the adapted narrative absorption scale that was used in the study is the only known example of a questionnaire scale instrument to have been specifically developed for the assessment of listener engagement with spoken narratives.

Listener Engagement with Podcasts

Despite there having been a lack of research that has specifically focused on listener engagement in mobile speech-based media consumption experiences, there has been extensive research dedicated to the study of listener engagement with podcasts. As studies have shown that podcasts are most frequently consumed using portable listening devices such as smartphones and tablets [209] [27], it therefore follows that the body of literature concerning listener engagement with podcast listening experiences could offer valuable insights for the study of listener engagement in mobile speech-based media consumption experiences more widely.

Berry [210] and Markman [211] charted the evolution of podcasting as a medium and highlighted its similarities with radio, citing commonalities in production practices, the cultivation of parasocial relationships between host and listener, and the desire to recreate a feeling of *"liveness"* through social media and other forms of transmedia engagement [212], despite the *"time-shifted"* nature of podcast consumption. Despite the advent of on-demand radio and podcasts available through streaming platforms such as BBC Sounds [213], radio is still predominately a linear format, mainly consumed in the car, home, and workplace [214]. In contrast, podcasts are consumed across a wide variety of different environmental [215], situational [189] and social contexts [216], in some cases fluidly traversing multiple changes of context over the course of a single ubiquitous listening experience [217]. This was supported by Spinelli and Dann [218, p. 118] who characterised podcasts as having entered the *"repertoire of media used for urban personal listening"*, wherein portable listening devices are used with headphones to create isolated personal listening spaces within the wider public space in which the listener is located.

Podcasting has traditionally been considered a secondary medium that is often consumed in parallel alongside additional activities competing for the listener's attention [217]. However, a 2019 study found that 70% of podcast consumers had experience of listening to podcasts without simultaneously engaging in any additional activities [209]. Chan-Olmsted and Wang [215] argued that podcasting has now matured into its own distinct medium, separate to radio, and is consumed in different settings to fulfill different listening gratifications. Furthermore, Chan-Olmsted and Wang [215] found podcast consumption at home tended to be more active and instrumental [219], being positively associated with information seeking and negatively associated with listening as a form of escapism/pastime.

Consumption out of the home, on the other hand, was found to be more ritualized [219], positively associated with escapism/pastime and negatively associated with information seeking. However, in a study that explored the situational fit of music, radio, and podcasting in urban headphone listening experiences, Nyre [189] found that podcasts were especially popular amongst *"self-curative pedestrian headphone listeners"*.

Podcast listening ranks amongst the top media activities for holding audience attention [27]. Despite this, however, studies have suggested that as with other forms of audio-based media [220], listeners exhibit different levels of engagement as podcast consumers [221]. Existing research has predominately focused on audience engagement with the podcast medium as a function of brand connection [221], social engagement [222], parasocial relationships [223], and the amount of listening time [200], with highly engaged participants characterised as those who make regular financial contributions, develop strong parasocial relationships with podcast hosts, or pass a given threshold of regular listening. Tobin and Guadagno [222] conducted a study exploring the motivations and outcomes of why people listen to podcasts, using five aspects of podcast listening metrics that constituted different ways of engagement. These consisted of the amount of time spent listening, the settings in which episodes of listening take place, the editorial format of the podcast, the device(s) used to listen, and social aspects of listening including social and parasocial engagement. Garcia-Marin [224] conducted qualitative research based on semi-structured interviews with listeners, podcasters, and pioneers in the medium, identifying 13 factors that determine engagement in podcasting. The factors were categorised into three groups of medium-centered, user-centered and podcast-centered engagement.

3.3 Summary

This chapter has reviewed the different kinds of auditory headsets commonly used by consumers in mobile audio consumption experiences. External sound modification technologies available in contemporary auditory headset devices were discussed and broadly categorised as being capable of either active noise cancellation, acoustically transparent, or adjustable hear-through behaviour. Potential use cases of external sound modification functionality were also reviewed, highlighting a range of different applications within typical everyday mobile audio consumption scenarios. Amongst these use cases it is notable that some mobile audio consumers are employing acoustically transparent

and adjustable hear-through functionality while wearing wireless earbud and hearable devices for much longer periods throughout the day. This allows them to interact with others in their surroundings while continuing to consume media or make use of the now commonplace integration between smartphone features and earphone devices.

Historical and contemporary overviews of mobile audio consumption practices amongst audio-based media consumers were also presented. This highlighted the complex relationship that exists between the listener, the media they are consuming, and their surrounding environment, including any individuals contained within it. A review of literature related to the aspects and assessment of listener engagement with spoken narratives was also conducted. This found that self-report questionnaire scale instruments were the most widely used research tool used in the assessment of narrative-based media experiences. Despite this, there was only a single example of a scale identified in the literature that had been specifically developed for the assessment of listener engagement with spoken narratives.

A further review of literature on listener engagement in mobile audio consumption experiences uncovered extensive research on mobile music listening experiences. Research was found to have almost exclusively focused on music rather than speech-based media. Several studies have compared both media, but there are very few examples of research that have focused exclusively on listener engagement with mobile speech-based media experiences. In part this is due to the relatively recent increase in the consumption of speech based media, spearheaded by the popularity of podcasts.

Results from several studies that investigated listeners' experiences with both music and speech-based media revealed significant differences in how listeners engage with each medium. This suggests that findings cannot be simply transferred from research on engagement with music to speech-based media. Together with other findings presented in the current chapter, this illustrates a lack of knowledge that exists in many areas relating to listener engagement in mobile speech-based media consumption experiences. This is potentially further complicated by the diversity of mediums and programming formats that are encompassed by the term speech-based media. As it is not possible to investigate all forms of speech-based media, the work presented in the forthcoming chapter aims to address some of these gaps in understanding by using these findings to inform the design of a study that investigates the relationship between environmental context and attentional engagement in podcast listening experiences. Results from this initial study will also serve to identify a series of typical

mobile speech-based media consumption scenarios and listening conditions. These findings will then be used to inform the design of subsequent applied and experimental work presented in Chapters 5 and 6.

Chapter 4

Environmental Context and Attentional Engagement in Podcast Listening Experiences

This chapter presents the results of an online questionnaire on podcast listening habits and behaviors that addresses the first main aim (A_1) of the present thesis presented in Chapter 1:

A_1 Investigate how different forms of environmental context relate to listeners' engagement with speech-based media.

Previous research, discussed in Chapter 3, has established that podcasts are now amongst the most popular forms of speech-based media and are most frequently consumed using mobile listening devices across a wide variety of environmental, situational, and social contexts. Different combinations of factors within these contexts are capable of capturing a listener's attention in an almost infinite number of ways. Despite this, no studies have investigated how an individual's environmental context might influence their attentional engagement in podcast listening experiences. Improving understanding of the contexts in which episodes of listening take place, and how they might affect listener engagement, would provide a basis from which to understand the relative importance of environmental noise as an potentially influencing factor on listener engagement, and could be highly valuable to researchers and producers working in the fields of object-based and personalized media.

This study aims to identify and investigate how different factors of environmental context might relate to listeners' attentional engagement when consuming podcasts using a smartphone. Furthermore, it aims to quantitatively map out how listeners consume podcasts across several *aspects of podcast listening engagement*, and explore how these aspects relate to the proposed *factors of environmental context*. Results from this study may also be of relevance to future work in related research fields, therefore it is also a stated aim of this study to evaluate potential implications of this work in the context of podcast studies, media personalisation, and attentional processing research.

This chapter's first research question asks how different factors of environmental context relate to listeners' attentional engagement when listening to podcasts. This question was primarily concerned with the identification of different factors which would then permit further measurement, analysis, and hypothesis testing. It is hypothesised that questionnaire items will group together under simple structure criteria [225] to form factors that meaningfully define elements of environmental context that influence listeners' attentional engagement when consuming podcasts using a mobile device [H1].

The second research question asks how podcast consumers engage with podcasts across several *aspects of podcast listening engagement* and how the aspects quantitatively relate to one another. The *aspects of podcast listening engagement* investigated are the amount of listening, the locations in which episodes of listening take place, the monitoring devices used to listen, the multitasking activities engaged in while listening, and the methods used to discover podcasts. It is hypothesized there will be positive correlations observed amongst the *aspects of podcast listening engagement* [H2].

The third research question asks how the proposed *factors of environmental context* relate to the measured *aspects of podcast listening engagement*. For this question it is also hypothesised that there will be positive correlations amongst the *environmental context factor scores* and *aspects of podcast listening engagement* [H3].

It is acknowledged that other variables, related to various personal characteristics of individual listeners such as their age, the nature of their work, or whether they are paid subscribers to other forms of media, may also bare some relation to the extent of an individual's baseline level of podcast listening engagement. However, an investigation of the impact of personal characteristic variables on podcast engagement is not included in the present study, as it is consider beyond the scope of the stated research questions and main aims.

4.1 Materials and methods

4.1.1 Participants

A sample of 264 people aged 18-66+ (18-25 = 13.6%, 26-35 = 42.8%, 36-45 = 18.6%, 46-55 = 17.0%, 56-65 = 7.2%, 66+ = 0.8%) completed an online questionnaire. The majority of the sample was male (51.9%), 43.2% were female, 3.8% non-binary/third gender, and 1.1% preferred not to answer. Of the 264 participants, 134 (50.8%) resided in the United Kingdom, 51 (19.3%) in the rest of Europe, 47 (17.8%) in the United States of America, and 32 (12.1%) in the rest of the World. Participants were recruited online via various methods including newsletters, social media posts (e.g., Facebook and Twitter), research group networks, University department mail-outs, podcasts, and word of mouth. To take part in the study participants first had to confirm that they had experience of listening to podcasts using a smartphone, in this sense the sample is representative of the non-zero podcast listening population. Responses were collected between the 29th of November 2021 and the 9th of February 2022. Participant involvement in the study was exclusively on a voluntary basis.

4.1.2 Measures

Participants first provided their age, gender, and country of residence. To answer the second and third research questions, which asked how aspects of podcast listening engagement relate to one another and to the proposed *factors of environmental context*, participants were first asked a series of quantitative and qualitative questions concerning their podcast listening behaviours and habits while using a smartphone.

Participants first selected how much time they spent listening on an average weekday, and then an average weekend day, from a list of 9 options ranging from 0 minutes to more than 10 hours. They were then asked to select all of the locations in which they had listened to a podcast with a smartphone from a list of 8 options. An *Other not listed (please specify)* option was also included with a free text response box to allow participants to register additional locations that were not included in the default survey options. Participants were also asked to select the location in which they most often listened from a list that was populated by their answers to the first question.

To collect data on the monitoring devices listeners use to consume podcasts, participants selected all of the headphone- and loudspeaker-based devices they had used with a smartphone to listen, from a list of 8 options. An *Other not listed (please specify)* option was also included with a free text response box. Participants were again asked to select the monitoring device they most often used from a list that was populated by their answers to the first question. Participants were then asked to select all of the activities they had engaged in while listening to podcasts with a smartphone from a list of 12 options. An *Other not listed (please specify)* option was also included with a free text response box. Participants were also asked to select the activity they most often used from list that was populated by their answers to the first question.

Finally, participants were asked to select all of the methods they had used to discover podcasts from a list of 8 options. An *Other not listed (please specify)* option was also included with a free text response box. Participants were again asked to select the method they most often used from a list that was populated by their answers to the first question.

This data was collected as a quantitative measure of different aspects of participants' podcast listening engagement, adapted from the aspects of podcast listening engagement metrics originally presented by Tobin and Guadagno [222] and Garcia-Marin [224] in order to reflect the present study's interest in environmental context. The data collected in this part of the study was analysed to calculate the total number of responses provided by individual participants for each *aspect of podcast listening engagement*. For example, the total number of unique methods that a participant reported using to discover podcasts would provide a measure of their podcast discovery engagement level. This data was then used to answer the second and third research questions.

To answer the first and third research questions, a 30-item *Attentional engagement as a function of environmental context in podcast listening experiences* scale (Figure 4.1) was constructed based on previous literature concerning influencing factors of environmental context on emotional response to music [226], choice, and devices used in everyday music listening [191]. Participants were asked to what extent each item was representative of their observed attitudinal experience on a discrete 5-point Likert scale (1 = strongly disagree to 5 = strongly agree). An "NA" option was also included on each item, allowing participants to indicate if insufficient contextual listening experience prevented them from providing a response. The 30 items were revised from a larger collection through the removal

of unclear and repetitive items and further review by two independent music psychology experts and a podcast industry professional. All items began with the statement *"I feel actively engaged in the listening experience when using a smartphone to consume podcasts..."* with the term *"actively engaged"* defined in the survey as, *"an attentional state where the listener is fully focused on the listening experience"*.

Two optional free text questions asked participants to first describe how different factors of environmental context influenced their level of engagement when listening to podcasts using a smartphone, and then how different factors of environmental context influenced their preference for listening to specific types of podcasts using a smartphone.

The survey also included several ancillary questions intended to gather data that would inform future research. The median time participants took to complete the survey was 10 minutes and 9 seconds. There were several outliers who registered longer elapsed completion times as they completed the survey over multiple sittings. A full copy of the survey instrument and survey logic, including the participant consent form, is provided in Figures A.1-A.27 of Appendix A.

4.1.3 Procedure

The University of York Electronic Engineering department (now School of Physics, Engineering and Technology) ethics committee approved this study (approval number Harrison101121). This is shown in Figures A.28-A.36 of Appendix A. Participants accessed the study online using a link to the participant information sheet. Participants were required to provide their consent prior to viewing and completing the questionnaire using the Qualtrics online survey platform. None of the participants who took part received any remuneration for their participation in the study.

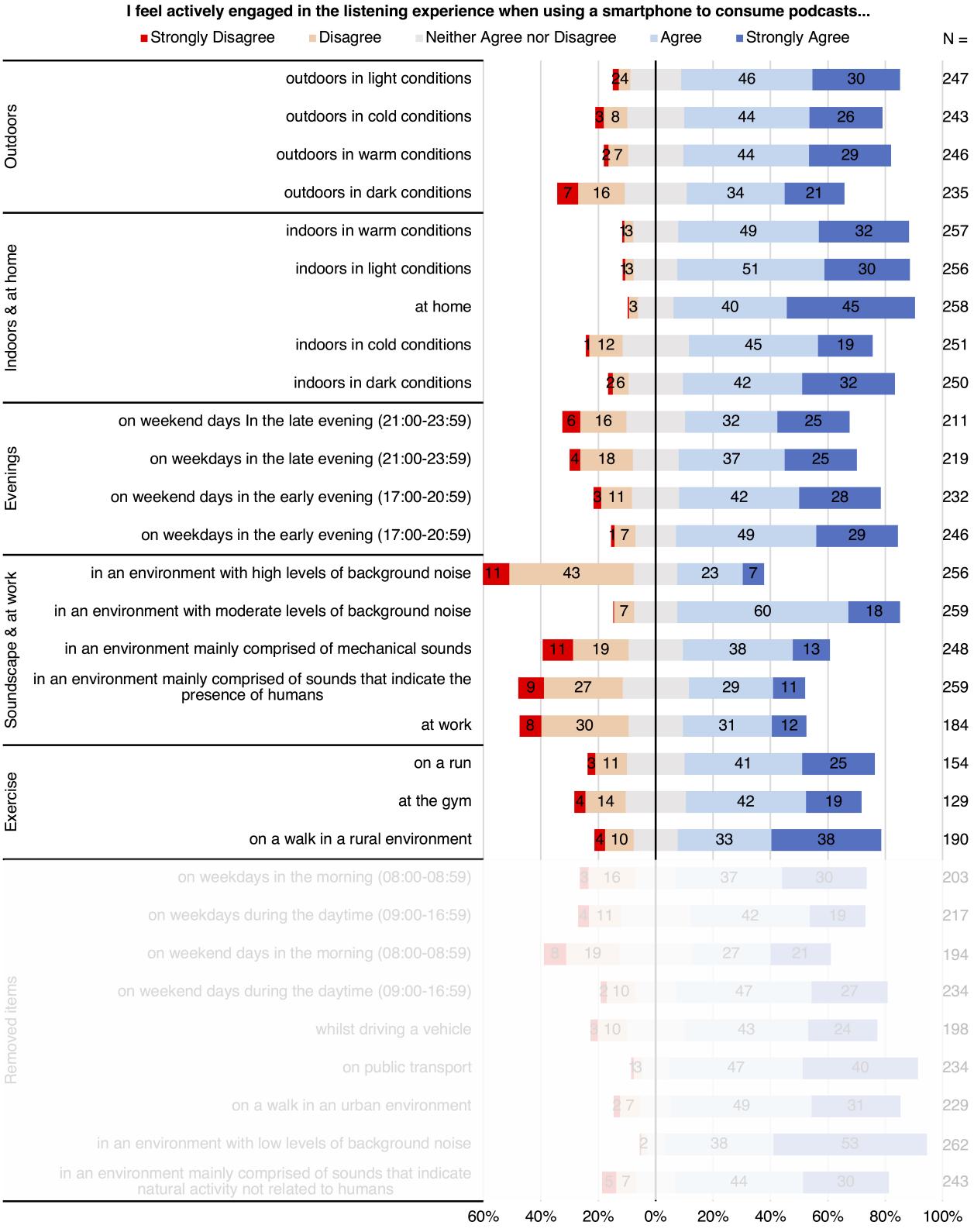


FIGURE 4.1: *Attentional engagement as a function of environmental context in podcast listening experiences* scale with the *factors of environmental context* produced by the EFA. The bottom 9 items are fainter to highlight the fact that they were removed during the EFA process.

4.2 Results

4.2.1 Factors of environmental context

To answer the first research question an exploratory factor analysis (EFA) was conducted to investigate the presence of underlying latent variables in the 30 item scale used in the online questionnaire, measuring factors of environmental context that influence listeners' attentional engagement in podcast listening experiences. The sample comprised 264 participants and included missing data due to *not applicable* item responses. 13.6% of responses across all items were classified as missing data, ranging from 51.1% to 0.8% for individual items. Missing data frequencies for each item can be attained from the sample size data reported in Figure 4.1.

Sampling adequacy tests were initially conducted using pairwise deletion, resulting in an overall Kaiser-Meyer-Olkin (KMO) sampling adequacy measure of 0.791 and individual item measures all greater than 0.633. All items were consequently classified between 'Mediocre' to 'Meritorious' as interpreted by Kaiser [227]. Bartlett's test of sphericity was significant at $p < 0.05$, suggesting it was likely the data could be factorised. Mardia's Multivariate Normality Test indicated that the data was not normally distributed (skew = 5963.36, kurtosis = 3.33) at $p < 0.001$.

The Multiple Imputation Factor Analysis (MIFA) R package [228] was used to impute missing data and indicate factor retention threshold criteria for further analysis. The incomplete dataset was imputed M = 30 times [229] using the fully conditional specification (FCS) [230] with the predictive mean matching (PMM) method. Confidence intervals for the cumulative proportion of explained variance were derived from principal component eigenvalue decomposition using Rubin's rules [231] and the average of proportions of explained variance over all imputed datasets. The EFA was then performed on an averaged estimated covariance matrix produced by the MIFA package with a Principal Axis Factoring (PAF) extraction method [232] using the *fa* function [233]. The disparity observed between the MIFA confidence intervals used to indicate factor retention threshold criteria and the consequent cumulative proportion of explained variance produced by the EFA and shown in Table 4.1 can be attributed to this distinction. However, this methodology is consistent with guidance describing its implementation

TABLE 4.1: Factor loadings for the EFA with direct quartimin rotation of the retained "environment context attentional engagement podcast listening" items. Analysis conducted using MIFA dataset (N = 264).

	Factor				
Item	**1**	**2**	**3**	**4**	**5**
outdoors in light conditions	0.806				
outdoors in cold conditions	0.753				
outdoors in warm conditions	0.735				
outdoors in dark conditions	0.610				
indoors in warm conditions		0.924			
indoors in light conditions		0.826			
at home		0.570			
indoors in cold conditions		0.477			
indoors in dark conditions		0.406			
on weekend days in the late evening (21:00-23:59)			0.859		
on weekdays in the late evening (21:00-23:59)			0.753		
on weekend days in the early evening (17:00-20:59)			0.710		
on weekdays in the early evening (17:00-20:59)			0.523		
in an environment with high levels of background noise				0.784	
in an environment with moderate levels of background noise				0.558	
in an environment mainly comprised of mechanical sounds				0.548	
in an environment mainly comprised of sounds that indicate the presence of humans				0.548	
at work				0.401	
on a run					0.970
at the gym					0.780
on a walk in a rural environment					0.423
Sum of squared (SS) loadings*	2.696	2.557	2.373	1.963	1.895
Proportion of variance** (%)	12.837	12.174	11.301	9.348	9.023
Proportion explained*** (%)	23.476	22.263	20.667	17.094	16.500

Note: Primary loadings < 0.4, alternative loadings < 0.3 and alternative loadings with a difference of < 0.2 suppressed. The five factors were labelled by the first author as *outdoors*, *indoors & at home*, *evenings*, *soundscape & at work*, and *exercise*. *SS loadings help to determine the importance of each factor in explaining the overall variance, with higher SS loadings indicating that a factor explains more of the total variance. **Proportion of variance indicates how much of the overall variance is explained by a factor. ***Proportion of variance explained describes the distribution of the variance explained within the five extracted factors.

provided by Busch and Nassiri [234] and was further confirmed in email correspondence from Nassiri (V Nassiri 2022, personal communication, 4 April).

Review of the correlation matrix showed that all items had at least one correlation with a coefficient greater than 0.30, except for *"whilst driving a vehicle"*. This item was removed, and the EFA rerun with the remaining 29 items [235]. As the dataset was not normally distributed a bootstrap, as opposed to Fieller, confidence interval was employed to determine factor retention criteria using the MIFA function [229]. As the first five factors were found to explain at least an estimated 5.0% of the total variance individually and 55.20% cumulatively, with bootstrap confidence intervals of 0.547 and 0.606, a proposed five-factor solution was considered for retention [229]. In order to support this proposed solution, separate parallel analysis and visual inspection of scree plots were conducted on the original incomplete 29 items using pairwise deletion to address missing data [236]. This initially suggested that six factors (or seven principal components) should be retained; however, when further analysis using a forced six-factor solution was carried out, the sixth factor failed to satisfy the $\geq$3 primary

loadings per factor criterion outlined by Howard [237]. Therefore, a five-factor solution was selected for retention and further analysis.

The EFA was rerun with a forced five-factor solution. As factors could potentially be correlated a direct quartimin rotation method was applied to improve interpretability, providing an equal weighting between correlated and uncorrelated factors [237]. Item factor loading criteria were drawn from a systematic review conducted by Howard [237], with items with primary factor loadings above 0.40 retained, while items with alternative loadings below 0.30 and cross-loadings with a difference of 0.20 or larger were cut. This resulted in the removal of the *weekday morning*, *weekday daytime*, *weekend morning*, *weekend daytime*, *public transport*, *urban walk environment*, *low background noise* and *human soundscape* items. These removed items are noted in the lower, faded out, section of Figure 4.1.

MIFA was rerun using the same criteria and again suggested a five-factor solution. This was supported by separate parallel analysis and visual inspection of scree plots performed on the incomplete 21 items using pairwise deletion to address missing data. This also suggested a five-factor (or five principal component) solution and consequently the EFA was rerun on the remaining 21 items using the same criteria. The overall KMO sampling adequacy measure was 0.792, with all individual measures greater than 0.638. All items were classified between 'Mediocre' to 'Meritorious' as interpreted by Kaiser [227]. The bootstrap confidence intervals for the proportion of explained variance using five factors were 0.634 and 0.690, and the equivalent Fieller's intervals were 0.613 and 0.666. The estimated proportion of explained variance for the first five factors was 0.640.

The rotated solution this produced demonstrated simple structure [225] and is shown in Table 4.1. The interpretation of the data was consistent with the proposed hypothesis and exhibited strong loadings across the five factors. Bivariate correlations between the five factors are shown in Table 4.2. Items associated with listening *outdoors* loaded on Factor 1, items describing listening *indoors & at home* on Factor 2, items relating to listening in the *evenings* on Factor 3, items pertaining to listening *soundscape & at work* on Factor 4, and listening while engaging in *exercise* on Factor 5. Cumulatively the five factors were able to explain 56.06% of the total variance across all 21 items.

An additional EFA was conducted using identical analysis criteria with pairwise deletion [238] and visual inspection of parallel analysis scree plots [239] to further support the results obtained from

TABLE 4.2: Pearson's product-moment correlations amongst the five factors produced by the EFA (*N* = 264).

		Outdoors	**Indoors & at home**	**Evenings**	**Soundscape & at work**	**Exercise**
Outdoors	PCC	–				
	Sig.	.				
Indoors & at home	PCC	**0.457****	–			
	Sig.	<0.001	.			
Evenings	PCC	**0.207****	**0.263****	–		
	Sig.	<0.001	<0.001	.		
Soundscape & at work	PCC	**0.313****	**0.258****	**0.238****	–	
	Sig.	<0.001	<0.001	<0.001	.	
Exercise	PCC	**0.244****	**0.205****	**0.146***	**0.334****	–
	Sig.	<0.001	<0.001	0.018	<0.001	.

*Correlation is significant at the 0.05 level (2-tailed). **Correlation is significant at the 0.01 level (2-tailed).

the MIFA solution. This produced the same factor loadings with the only exceptions being the *on a walk in a rural environment* item exhibiting a weak cross-loading of 0.489 and 0.330 on the *exercise* and *outdoors* listening factors respectively and Factors 4 and 5 reversing in order of proportion of variance explained due to a difference of 0.81%. Therefore, it can be determined, that as items from the Likert scale grouped together to form five *factors of environmental context*, the null hypothesis can be rejected and the alternate hypothesis [H1] accepted, that is questionnaire items will group together under simple structure criteria [225] to form factors that meaningfully define elements of environmental context that influence listeners' attentional engagement when consuming podcasts using a mobile device.

Environmental context factor scores were then computed for each participant using the method described by Busch and Nassiri [234], providing a measure of participants' attentional engagement for each of the five *factors of environmental context*. An overall mean factor score was also computed for each participant from the five *environmental context factor scores* to provide a cumulative measure of attentional engagement as a function of the *factors of environmental context* shown in Table 4.1.

4.2.2 Relationships between aspects of podcast listening engagement

To answer the second and third research questions, which collectively asked how listeners engage with podcasts across several *aspects of podcast listening engagement*, how these aspects relate to one another, and how they relate to the proposed *factors of environmental context*, participants were asked a series of questions intended as a quantitative measure of their podcast listening engagement and

TABLE 4.3: Amount of time participants spend listening to podcasts using a smartphone on average weekdays and weekend days (N = 264).

	Average weekday		Average weekend day	
Listening amount	**Frequency**	**%**	**Frequency**	**%**
More than 10 hours	2	0.76	1	0.38
5 to 10 hours	13	4.92	0	0.00
3 to 5 hours	25	9.47	11	4.17
2 to 3 hours	25	9.47	27	10.23
1 to 2 hours	69	26.14	68	25.76
30 to 60 minutes	60	22.73	74	28.03
15 to 30 minutes	45	17.05	34	12.88
Less than 15 minutes	24	9.09	29	10.98
0 minutes	1	0.38	20	7.58
Total	**264**	**100.00**	**264**	**100.00**

listening behaviours. The following section presents an outline of this data, followed by the results of correlation tests that were carried out to identify potential relationships between the aspects.

Amount of listening time

Response frequencies for the questions measuring the amount of time participants spend listening are shown in Table 4.3. The median category for the amount of listening time on an average weekday amongst all participants was 60 to 120 minutes, accounting for 26.14% of all recorded responses. The median category for an average weekend day was 30 to 60 minutes, which accounted for 28.03% of all responses. Additionally, there was only one participant who reported not listening on weekdays, compared to 20 who reported not listening on weekend days. The Likert scale *not applicable* response results presented in Figure 4.1 indicate that episodes of listening occurred fairly consistently throughout the week. Early evenings on weekdays was the most popular time to listen with 93.18% of participants reporting having listened during this time, while weekend days in the morning was the least popular with 73.48%.

Listening locations

Response frequencies for the questions concerning locations in which participants consumed podcasts are shown in Table 4.4. The data presented in the 'as a % of sample' column describes the proportion

TABLE 4.4: Locations in which participants have consumed (N = 264) and most often consume (N = 264) podcasts using a smartphone.

	Locations in which podcasts are consumed			Locations in which podcasts are most often consumed		
Location	**Frequency**	**%**	**as a % of sample**	**Frequency**	**%**	**as a % of location cases**
At home	245	20.82	92.45	126	47.73	51.43
Traveling on public transport	193	16.40	72.83	21	7.95	10.88
Walking in an urban environment	190	16.14	71.70	48	18.18	17.95
Driving a vehicle	156	13.25	58.87	28	10.61	17.95
Walking in a rural environment	108	9.18	40.75	14	5.30	12.96
On a run	88	7.48	33.21	7	2.65	7.95
At work (private)***	72	6.12	27.17	5	1.89	6.94
At the gym	61	5.18	23.02	2	0.76	3.28
At work (public)***	43	3.65	16.23	7	2.65	16.28
Other (public)*	15	1.27	5.66	5	1.89	33.33
Other (private)*	6	0.51	2.26	1	0.38	16.67
Total	**1177**	**100.00**	**445.83**	**264**	**100.00**	**-**
Listening location groups**						
Public locations	698	59.30	264.39	104	39.39	-
Private locations	479	40.70	181.44	160	60.61	-

*Computed from *Other not listed (please specify)* responses. **Computed from location responses according to analysis criteria outlined in the results section. ***Computed from at work and commuting response data.

of respondents that reported having consumed podcasts in a given location and the 'as a % of location cases' column describes locations in which podcasts are most often consumed as a proportion of the total cases for each location.

Participants' homes were reported as the location in which podcasts were most frequently consumed, with 92.45% of those surveyed having listened at home and 47.73% stating they most often listened at home. The next most popular location was listening while travelling on public transport with 72.83%, followed by listening while walking in an urban environment at 71.70%, however, listening while walking in a rural environment was much lower at 40.75%. Similarly, 18.18% of participants reported most often listening while walking in urban environments, compared to only 5.30% that most often listened while walking in rural environments.

It should also be noted, however, that these results highlight inconsistencies in the data collected by this study. While the data shown in Table 4.4 indicates that the percentage of participants who listened while walking in an urban and rural environment was 71.70% and 40.75% respectively, the *not applicable* response data provided to the Likert item questions in Figure 4.1 suggests that 86.75% of participants listened while walking in urban environments, compared to 71.67% that listened while walking in rural environments. The most likely explanation for this discrepancy is an element of

self-response bias in the Likert item response data. Despite this, it is still noteworthy that both sets of data are similarly distributed, with listening while walking in an urban environment registering a much higher proportion of cases in both instances.

Further analysis was also conducted to investigate how participants' listening episodes were distributed between private and public spaces. Criteria for listening location group categorisation was adapted from a study on the influence of location in everyday experiences of music conducted by Krause et al. [192]. The *private location* group consisted of *at home*, *at work (private)*, *driving a car* and *other (private)* responses, while the *public location* group comprised all remaining locations in Table 4.4. The *other (private)* variable was computed manually via a process of categorising the *Other not listed (please specify)* responses as either private or public according to the free text data provided by participants. The *at work (private)* and *(public)* variables were computed by filtering the *at work* responses according to whether or not the participant commuted to their workplace.

This showed that 59.30% of all locations in which participants listened were public spaces, while the remaining 41.70% were private. However, 60.61% of participants reported most often listening in a private space, compared to 39.39 % who most often listened in public. Further analysis revealed that the vast majority of participants (87.88%) had experience of listening in both private and public locations. 96.21% had experience of listening in at least one private location, while 91.67% reported having listened in at least one public location. This was contrasted by 8.33% of participants who reported only having experience of listening in private locations, and just 3.79% who had only listened in public locations.

Monitoring devices

Response frequencies for the two questions on the monitoring devices used by participants to consume podcasts are shown in Table 4.5. The data presented in the 'as a % of sample' column describes the proportion of respondents that reported having consumed podcasts using a given monitoring device and the 'as a % of device cases' column describes the monitoring devices most often used to consume podcasts as a proportion of the total cases for each device. The results showed that there was an almost equal split across monitoring devices used by the most participants, between built-in smartphone

TABLE 4.5: Monitoring devices participants have used (N = 264) and most often use (N = 264) to consume podcasts with a smartphone.

	Devices used to consume podcasts			Devices most often used to consume podcasts		
Monitoring device	**Frequency**	**%**	**as a % of sample**	**Frequency**	**%**	**as a % of device cases**
Built-in smartphone speakers	146	16.33	55.30	36	13.64	24.66
Wired in-ear headphones	145	16.22	54.92	59	22.35	40.69
Wireless in-ear headphones	141	15.77	53.41	82	31.06	58.16
Bluetooth speaker(s)	132	14.77	50.00	16	6.06	12.12
Built-in vehicle speakers	120	13.42	45.45	22	8.33	18.33
Wireless over-ear headphones	96	10.74	36.36	34	12.88	35.42
Wired over-ear headphones	92	10.29	34.85	9	3.41	9.78
Other speaker(s)*	9	1.01	3.41	2	0.76	22.22
Bone conduction headset	6	0.67	2.27	1	0.38	16.67
Other built-in device speakers*	4	0.45	1.52	1	0.38	25.00
Bluetooth hearing aids*	1	0.11	0.38	-	-	-
Other wireless headphones*	1	0.11	0.38	1	0.38	100.00
Other headphones*	1	0.11	0.38	1	0.38	100.00
Total	**894**	**100.00**	**338.64**	**264**	**100.00**	**-**
Monitoring device groups**						
Headphones devices	483	53.97	182.95	187	70.83	-
Loudspeakers devices	411	46.03	156.06	77	29.17	-

*Computed from *Other not listed (please specify)* responses. **Computed from monitoring device responses according to analysis criteria outlined in the results section.

speakers (55.30%), wired in-ear headphones (54.92%), wireless in-ear headphones (53.41%) and Bluetooth speaker(s) (50.00%). However, wireless in-ear headphones (31.06%) and wired in-ear headphones (22.35%) were the leading monitoring devices most often used by listeners by a significant margin, followed by built-in smartphone speakers (13.64%), and then wireless over-ear headphones (12.88%).

Additional analysis was conducted to explore how the monitoring devices participants used to listen were distributed between headphone- and loudspeaker-based devices. The *loudspeaker devices* group consisted of built-in smartphone speakers, Bluetooth speaker(s), built in vehicle speakers, other speaker(s) and other built-in device speakers devices. The *headphone devices* group comprised of all remaining devices listed in Table 4.5. This analysis found that 53.97% of all monitoring devices used by participants were headphones, while the remaining 46.03% were loudspeakers. Similarly, 70.83% of participants reported most often using headphones to listen, compared to 29.17% who most often used loudspeakers. 96.97% of participants reported having used at least one type of headphone monitoring device, while 18.94% of participants reported only having used headphones. In contrast, 81.06% of participants reported having used at least one loudspeaker monitoring device, while just

3.03% reported only having used loudspeakers to listen. 78.03% of participants had experience of using both headphones and loudspeakers.

Multitasking activities

Multitasking activity response frequencies are shown in Table 4.6. The data presented in the 'as a % of sample' column describes the proportion of respondents that reported having consumed podcasts while engaging in a given multitasking activity and the 'as a % of activity cases' column describes activities most often engaged in while consuming podcasts as a proportion of the total cases for each activity. The results found that 3.03% of participants didn't engage in multitasking, choosing instead to focus solely on the listening experience. There was also very little difference observed in listeners' multitasking habits between average weekdays and weekend days. *Doing housework* (76.89%), *preparing food* (73.86%) and *exercising* (61.74%) were the top three activities engaged in by the most participants. These were also the top three activities that participants reported most often engaging in, and while *doing housework* (26.89%) was still the most popular activity, the order of *exercising* (23.95%) and *preparing food* (10.92%) was reversed.

Additional analysis was also conducted to further explore how participants engaged in different multitasking activity modalities. Activity responses were separated into four categories representing *work*, *leisure*, *media* and *transit activities*. The *work activities* group consisted of the *doing housework*, *preparing food*, *working*, *shopping/running errands*, *gardening*, *DIY*, and *childcare* responses. The *media activities* group included *watching TV*, *reading*, *listening to music*, *walking on the phone*, *sending messages via phone or computer*, *using social media sites*, *watching films*, *other computer activities*, and *playing video games*. The *leisure activities* group consisted of *exercising*, *making art/crafting*, *showering/bathing*, *going to sleep*, *eating*, and *doing puzzles*. The *transit activities* group consisted of *driving*, *using public transport*, *walking*, *other travel*, *commuting*, and *cycling*.

This analysis found that most activities were *work* related, accounting for 39.82% of all activities engaged in and 42.02% of activities that participants most often engaged in. This was followed by the *media* modality which accounted for 33.64% of all activities engaged in, but only 18.49% of the activities most often engaged in. This was contrasted by the *leisure* related activities modality, which,

TABLE 4.6: Activities participants have engaged in (N = 264) and most often engage in (N = 238) while listening to podcasts with a smartphone.

	Activities engaged in while consuming podcasts			**Activities most often engaged in while consuming podcasts**		
Multitasking activity	**Frequency**	**%**	**as a % of sample**	**Frequency**	**%**	**as a % of activity cases**
Doing housework	203	18.45	76.89	64	26.89	31.53
Preparing food	195	17.73	73.86	26	10.92	13.33
Exercising	163	14.82	61.74	57	23.95	34.97
Messaging via phone/-computer	116	10.55	43.94	5	2.10	4.31
Using social media sites	112	10.18	42.42	15	6.30	13.39
Other computer activities	83	7.55	31.44	15	6.30	18.07
Playing video games	39	3.55	14.77	6	2.52	15.38
Driving*	30	2.73	11.36	11	4.62	36.67
Walking*	27	2.45	10.23	11	4.62	40.74
Making Art/Crafting*	17	1.55	6.44	5	2.10	29.41
Shopping/Running Errands*	15	1.36	5.68	2	0.84	13.33
Working*	14	1.27	5.30	6	2.52	42.86
Showering/Bathing*	12	1.09	4.55	3	1.26	25.00
Reading	11	1.00	4.17	1	0.42	9.09
Commuting*	10	0.91	3.79	3	1.26	30.00
Going to sleep*	9	0.82	3.41	2	0.84	22.22
No activity****	8	0.73	3.03	-	-	-
Watching TV	5	0.45	1.89	2	0.84	40.00
Using public transport*	5	0.45	1.89	1	0.42	20.00
Gardening*	5	0.45	1.89	1	0.42	20.00
DIY*	4	0.36	1.52	-	-	-
Other Travel*	3	0.27	1.14	-	-	-
Cycling*	3	0.27	1.14	1	0.42	33.33
Eating*	3	0.27	1.14	-	-	-
Watching films	2	0.18	0.76	-	-	-
Doing puzzles*	2	0.18	0.76	-	-	-
Childcare*	2	0.18	0.76	1	0.42	50.00
Listening to music	1	0.09	0.38	-	-	-
Talking on the phone	1	0.09	0.38	-	-	-
Total	**1100**	**100.00**	**416.67**	**238**	**100.00**	**-**
Multitasking activity groups**						
Work activities	438	39.82	165.91	100	42.02	-
Media activities	370	33.64	140.15	44	18.49	-
Leisure activities	206	18.73	78.03	67	28.15	-
Transit activities	78	7.09	29.55	27	11.34	-
No activity	8	0.73	3.03	-	-	-
Removed						
Other not listed (invalid)***		-	-	18	-	-
No activity (invalid)*****		-	-	8	-	-

*Computed from *Other not listed (please specify)* responses. **Computed from multitasking activity responses according to analysis criteria outlined in the results section. ***Removed as multiple/void free text responses were provided for activity most often engaged in. ****Participant reported not engaging in any multitasking activities. *****Removed due to participants not engaging with any multitasking activity.

TABLE 4.7: Proportion of time participants spend multitasking while listening to podcasts using a smartphone on average weekdays and weekend days. (N = 264)

	Average weekday		Average weekend day	
Multitasking proportion	**Frequency**	**%**	**Frequency**	**%**
100%	116	43.94	119	45.08
90%	41	15.53	32	12.12
80%	31	11.74	20	7.58
70%	20	7.58	19	7.20
60%	4	1.52	5	1.89
50%	22	8.33	22	8.33
40%	4	1.52	4	1.52
30%	8	3.03	9	3.41
20%	5	1.89	6	2.27
10%	3	1.14	6	2.27
0%	10	3.79	22	8.33
Total	**264**	**100.00**	**264**	**100.00**

despite representing just 18.73% of all activities engaged in, accounted for 28.15% of activities most often engaged in. Lastly, *transit* related activities received by far the lowest proportion of participant engagement, accounting for only 7.09% of all activities and 11.34% of activities most often engaged in.

The proportion of time participants spent multitasking while listening on average weekday and weekend days is also shown in Table 4.7. These results revealed that the majority of participants engage in multitasking activities for at least 90% of the time they spend listening to podcasts, while only 3.79% and 8.33% reported never engaging in multitasking activities on weekdays and weekend days respectively.

Podcast discovery methods

Podcast discovery method response frequencies are shown in Table 4.8. The data presented in the 'as a % of sample' column describes the proportion of respondents that reported having discovered podcasts via a given method and the 'as a % of method cases' column describes methods most often used to discover podcasts as a proportion of the total cases for each method. The results showed that consumers use a wide variety of methods to discover podcasts. The majority of the sample had experience of using *recommendations from friends & family* (71.29%), *listening to podcasts* (68.94%), and *streaming services* (59.09%), while a significant proportion also had experience of *searching the*

TABLE 4.8: Methods participants have used (N = 264) and most often use (N = 262) to discover podcasts.

	Methods used to discover podcasts			Methods most often used to discover podcasts		
Discovery method	**Frequency**	**%**	**as a % of sample**	**Frequency**	**%**	**as a % of method cases**
Recommendations from friends/family	189	20.09	71.59	62	23.66	32.80
Listening to podcasts	182	19.34	68.94	52	19.85	28.57
Streaming services	156	16.58	59.09	62	23.66	39.74
Searching the internet	129	13.71	48.86	28	10.69	21.71
Recommendations on social media	126	13.39	47.73	20	7.63	15.87
Listening to radio	57	6.06	21.59	-	-	-
Recommendations from YouTube creators	38	4.04	14.39	6	2.29	15.79
Recommendations from YouTube watch history	19	2.02	7.20	4	1.53	21.05
Podcast player app*	13	1.38	4.92	5	1.91	38.46
Newsletters*	9	0.96	3.41	5	1.91	55.56
Recommendations from industry media*	7	0.74	2.65	-	-	-
Recommendations from print media*	6	0.64	2.27	2	0.76	33.33
Recommendations from online media*	3	0.32	1.14	1	0.38	33.33
Recommendations from podcasters*	3	0.32	1.14	1	0.38	33.33
Recommendations from colleagues*	2	0.21	0.76	-	-	-
Recommendations from television*	1	0.11	0.38	-	-	-
Podcasting awards*	1	0.11	0.38	-	-	-
Total	**941**	**100.00**	**356.44**	**262**	**100.00**	**-**
Podcast discovery method groups**						
Other online media	332	35.28	125.76	66	25.19	-
Personal recommendations	191	20.30	70.35	62	23.66	-
Podcasts	185	19.66	70.08	53	20.23	-
Podcast apps	169	17.96	64.02	67	25.57	-
Offline media	64	6.80	24.24	14	5.34	-
Removed						
Other not listed (invalid)***	-	-	-	2	-	-

*Computed from *Other not listed (please specify)* responses. **Computed according to analysis criteria outlined in the results section. ***Removed as multiple/void free text responses were provided for discovery method most often used.

internet (48.86%), and using *recommendations on social media* (47.73%). Similarly, there was a fairly even split between *recommendations from friends & family* (23.66%), *streaming services* (23.66%), and *listening to podcasts* (19.85%) for the three methods most often used by participants.

Additional analysis was also conducted to provide a deeper insight into how listeners utilize different methods of consumption in podcast discovery. Method responses were separated into five categories based on research conducted by Insights [240], including *other online media*, *personal recommendations*, *podcasts*, *podcast apps*, and *offline media*. The *other online media* group consisted of responses to *searching the internet*, *recommendations from social media*, *YouTube creators*, *YouTube viewing history*, *online media*, *podcasting awards*, *newsletters*, and *industry media*. The *personal recommendations* group included the *recommendations from friends & family* and *colleagues' responses*. The *podcasts* group comprised responses from the *listening to podcasts* and *recommendations from podcast*

hosts items. The *podcast apps* group included responses from the *streaming services* and *podcast apps* items. The *offline media* group consisted of the *listening to radio*, *recommendations from print media*, and *recommendations from television* items.

The results from this group analysis showed that while the *other online media* group represented the clear majority share of all methods used (35.28%), the methods participants most often used were distributed relatively equally between *podcast apps* (25.57%), *other online media* (25.19%), *personal recommendations* (23.66%), and *podcasts* (20.23%). *Offline media* related methods were by far the least common amongst the sample, with only 6.8% of participants having used an *offline media* method, and 5.34% providing one as the method they had most often used.

Correlations amongst aspects of podcast listening engagement

A series of Spearman's rank-order and Pearson's product-moment correlations were run to assess all pairwise relationships amongst the *aspects of podcast listening engagement*. As the two survey questions concerning the average amount of time participants spent listening to podcasts on weekday and weekend days were measured using a non-continuous ordinal scale and the other *aspects of podcast listening engagement* metrics were continuous data, a series of Spearman's rank-order correlations were run to assess the relationships between these variables [241]. These correlations are shown in Table 4.9. Preliminary analysis, consisting of visual inspections of scatterplots, found all pairwise relationships between the amount of listening time variables and the other *aspects of podcast listening engagement* to be monotonic, with the exception of both amount of listening time variables and the total monitoring devices used pairings.

There was a statistically significant, moderate positive correlation between average amount of weekday listening time and the total number of locations in which podcasts were listened to, $r(262) = .30$, $p < .001$. There were also two statistically significant, small positive correlations between weekday listening time and the total number of activities simultaneously engaged in while listening, $r(262) = .24$, $p < .001$, and the number of methods used to discover podcasts, $r(262) = .20$, $p = .001$, respectively. A statistically significant, moderate positive correlation was also found between the average amount of weekend day listening time and total listening locations, $r(262) = .39$, $p < .001$. Similarly, there were

TABLE 4.9: Spearman's rank-order correlations amongst the *aspects of podcast listening engagement*. (N = 264)

		Weekday listening	**Weekend listening**	**Total locations**	**Total devices**	**Total activities**	**Total discovery**
Weekday	rho	–					
listening	Sig.	.					
Weekend	rho	**0.551****	–				
listening	Sig.	<0.001	.				
Total	rho	**0.303****	**0.387****	–			
locations	Sig.	<0.001	<0.001	.			
Total	rho	0.006	**0.161****	**0.415****	–		
devices	Sig.	0.926	0.009	<0.001	.		
Total	rho	**0.235****	**0.296****	**0.494****	**0.340****	–	
activities	Sig.	<0.001	<0.001	<0.001	<0.001	.	
Total	rho	**0.198****	**0.176****	**0.306****	**0.322****	**0.309****	–
discovery	Sig.	0.001	0.004	<0.001	<0.001	<0.001	.

**Correlation is significant at the 0.01 level (2-tailed).

TABLE 4.10: Pearson's product-moment correlations amongst the total monitoring devices, total multitasking activities, total listening locations, and total discovery methods. (N = 264)

		Total devices	**Total activities**	**Total locations**	**Total discovery**
Total devices	PCC	–			
	Sig.	.			
Total activities	PCC	**0.352****	–		
	Sig.	<0.001	.		
Total locations	PCC	**0.446****	**0.498****	–	
	Sig.	<0.001	<0.001	.	
Total discovery	PCC	**0.303****	**0.318****	**0.308****	–
	Sig.	<0.001	<0.001	<0.001	.

**Correlation is significant at the 0.01 level (2-tailed).

also two statistically significant, small positive correlations between weekend day listening time and the total number of activities simultaneously engaged in, r(262) = .30, $p < .001$, and the number of methods used to discover podcasts, r(262) = .18, $p = .004$, respectively.

As the computed total listening locations, total monitoring devices, total multitasking activities, and total discovery methods per participant metrics were continuous data, a series of Pearson's product-moment correlations were run to assess all pairwise relationships between these variables [241]. These correlations are shown in Table 4.10. Preliminary analysis found all pairwise relationships to be linear with all variables normally distributed, as assessed by visual inspection of Normal Q-Q Plots. Additionally, assessment of scatterplots for the bivariate combinations found there were no outliers.

There were three statistically significant, moderate positive correlations between the total number of locations in which podcasts were listened to and the number of activities simultaneously engaged

in while listening, r(262) = .50, $p < .001$, the number of monitoring devices used to listen, r(262) = .45, $p < .001$, and the number of methods used to discover podcasts, r(262) = .31, $p < .001$ respectively. There were also two statistically significant, moderate positive correlations between the total number of activities simultaneously engaged in while listening to podcasts and the total number of monitoring devices used to listen, r(262) = .35, $p < .001$, and the number of methods used to discover podcasts, r(262) = .32, $p <.001$, respectively. Finally, there was a statistically significant, moderate positive correlation between the total number of methods used to discover podcasts and the number of monitoring devices used to listen, r(262) = .30, $p < .001$.

Collectively, there were a total of twelve statistically significant relationships between the different aspects of measured podcast listening engagement, consisting of two moderate positive and four small positive Spearman's rank-order correlations, and six moderate positive Pearson's product moment correlations. Therefore, as there were several statistically significant positive correlations observed amongst the measured *aspects of podcast listening engagement* the null hypothesis can be rejected and the alternate hypothesis [H2] accepted, that is there will be positive correlations observed amongst the *aspects of podcast listening engagement*.

4.2.3 Relationships between environmental context factor scores and aspects of podcast listening engagement

To answer the third research question, which asked how the proposed *factors of environmental context* relate to the *aspects of podcast listening engagement*, a series of Spearman's rank-order and Pearson's product-moment correlations were run to assess the pairwise relationships between the *environmental context factor scores* and *aspects of podcast listening engagement*.

As the two questions that asked participants to report the average amount of time they spent listening to podcasts on weekday and weekend days were measured using a non-continuous ordinal scale and the *environmental context factor scores* were continuous data, a series of Spearman's rank-order correlations were run to assess the relationships between these variables [241]. These correlations are shown in Table 4.11. Preliminary analysis, consisting of visual inspections of scatterplots, found all pairwise relationships between the amount of listening time variables and factor scores to be monotonic, with the exception of the weekend listening and *exercise* factor score pairing.

TABLE 4.11: Spearman's rank-order correlations amongst the *aspects of podcast listening engagement & environmental context factor scores* (N = 264).

		Weekday listening	**Weekend listening**	**Total devices**	**Total activities**	**Total locations**	**Total disc.**	**Outdoors FS**	**Indoors & home FS**	**Evenings FS**	**Soundscape & work FS**	**Exercise FS**	**Overall mean FS**
Weekday	rho	–											
listening	Sig.	.											
Weekend	rho	0.551**	–										
listening	Sig.	<0.001	.										
Total	rho	0.006	**0.161****	–									
devices	Sig.	0.926	0.009	.									
Total	rho	**0.235****	**0.296****	**0.340****	–								
activities	Sig.	<0.001	<0.001	<0.001	.								
Total	rho	**0.303****	**0.387****	**0.415****	**0.494****	–							
locations	Sig.	<0.001	<0.001	<0.001	<0.001	.							
Total	rho	**0.198****	**0.176****	**0.322****	**0.309****	**0.306****	–						
discovery	Sig.	0.001	0.004	<0.001	<0.001	<0.001	.						
Outdoors	rho	0.054	0.097	-0.07	0.008	**0.128***	0.021	–					
FS	Sig.	0.386	0.116	0.258	0.901	0.038	0.73	.					
Indoors &	rho	0.074	0.126*	0.072	**0.153***	0.096	0.073	**0.509****	–				
home FS	Sig.	0.231	0.041	0.243	0.013	0.122	0.239	<0.001	.				
Evenings	rho	**0.127***	**0.250****	0.012	0.067	**0.126***	0.107	**0.252****	**0.317****	–			
FS	Sig.	0.039	<0.001	0.846	0.28	0.04	0.082	<0.001	<0.001	.			
Soundscape	rho	**0.240****	**0.305****	0.097	**0.205****	**0.295****	0.046	**0.293****	**0.236****	**0.234****	–		
& work FS	Sig.	<0.001	<0.001	0.115	<0.001	0.001	0.454	<0.001	<0.001	<0.001	.		
Exercise	rho	0.089	**0.132***	0.018	0.081	**0.219****	0.009	**0.263****	**0.233****	**0.186****	**0.335****	–	
FS	Sig.	0.148	0.032	0.774	0.188	<0.001	0.889	<0.001	<0.001	0.002	<0.001	.	
Overall	rho	**0.126***	**0.266****	0.055	**0.160****	**0.262****	0.102	**0.699****	**0.661****	**0.581****	**0.622****	**0.589****	–
mean FS	Sig.	0.041	<0.001	0.371	0.009	<0.001	0.098	<0.001	<0.001	<0.001	<0.001	<0.001	.

*Correlation is significant at the 0.05 level (2-tailed). **Correlation is significant at the 0.01 level (2-tailed). ***FS = Factor Scores

There were two statistically significant, small positive correlations between the average amount of weekday listening time, and the *soundscape & at work* factor scores, r(262) = .24, $p < .001$, and the computed *overall mean* factor scores, r(262) = .13, $p = .041$, respectively. A statistically significant, moderate positive correlation was found between the average amount of weekend day listening time and the *soundscape & at work* factor scores, r(262) = .31, $p < .001$. While there were four statistically significant, small positive correlations between weekend day listening time and *overall mean*, r(262) = .27, $p < .001$, *evenings*, r(262) = .25, $p < .001$, *exercise*, r(262) = .13, $p = .032$, and *indoors & at home* factor scores, r(262) = .13, $p = .041$, respectively.

As the *environmental context factor scores* and the computed total listening locations, total monitoring devices, total multitasking activities, and total discovery methods per participant metrics were all continuous data, a series of Pearson's product-moment correlations were run to assess all pairwise relationships between these variables [241]. These correlations are shown in Table 4.12. Preliminary analysis found all pairwise relationships were linear with all variables normally distributed, as assessed

TABLE 4.12: Pearson's product-moment correlations amongst the total monitoring devices, total multitasking activities, total listening locations, total discovery methods, and *environmental context factor scores* (N = 264).

		Total devices	**Total activities**	**Total locations**	**Total discovery**	**Outdoors FS**	**Indoors & home FS**	**Evenings FS**	**Soundscape & work FS**	**Exercise FS**	**Overall mean FS**
Total devices	PCC	–									
	Sig.	.									
Total activities	PCC	**0.352****	–								
	Sig.	<0.001	**0.138***								
Total locations	PCC	**0.446****	**0.498****	–							
	Sig.	<0.001	<0.001	.							
Total discovery	PCC	**0.303****	**0.318****	**0.308****	–						
	Sig.	<0.001	<0.001	<0.001	.						
Outdoors FS	PCC	-0.074	-0.02	0.115	0.029	–					
	Sig.	0.23	0.751	0.062	0.643	.					
Indoors & home FS	PCC	0.033	**0.138***	0.107	0.081	**0.457****	–				
	Sig.	0.596	0.025	0.084	0.188	<0.001	.				
Evenings FS	PCC	0.01	0.056	0.115	0.115	**0.207****	**0.263****	–			
	Sig.	0.874	0.366	0.061	0.062	<0.001	<0.001	.			
Soundscape & work FS	PCC	0.092	**0.186****	**0.284****	0.052	**0.313****	**0.258****	**0.238****	–		
	Sig.	0.138	0.002	<0.001	0.397	<0.001	<0.001	<0.001	.		
Exercise FS	PCC	0.014	0.028	**0.216****	0.027	**0.244****	**0.205****	**0.146***	**0.334****	–	
	Sig.	0.822	0.655	<0.001	0.666	<0.001	<0.001	0.018	<0.001	.	
Overall mean FS	PCC	0.023	0.121	**0.261****	0.095	**0.691****	**0.679****	**0.577****	**0.667****	**0.600****	–
	Sig.	0.709	0.05	<0.001	0.125	<0.001	<0.001	<0.001	<0.001	<0.001	.

*Correlation is significant at the 0.05 level (2-tailed). **Correlation is significant at the 0.01 level (2-tailed). ***FS = Factor Scores

by visual inspection of Normal Q-Q Plots. Assessment of bivariate scatterplots found there were no outliers.

There were three statistically significant, small positive correlations between the total locations in which podcasts were listened to, and the *soundscape masking & at work*, r(262) = .28, $p < .001$, computed *overall mean*, r(262) = .26, $p < .001$, and *exercise* factor scores, r(262) = .22, $p < .001$, respectively. There were also two statistically significant, small positive correlations between the total number of multitasking activities engaged in and the *soundscape masking & at work*, r(262) = .19, p = .002, and *indoors & at home* factor scores, r(262) = .14, p = .025, respectively.

In summary, there was a total of twelve statistically significant relationships between the measured *aspects of podcast listening engagement* and *environmental context factor scores*, consisting of one moderate positive and six small positive Spearman's rank-order correlations, and five small positive Pearson's product moment correlations. Consequently, as there were several statistically significant positive correlations amongst the measured *aspects of podcast listening engagement* and the proposed *factors of environmental context* the null hypothesis can be rejected and the alternate hypothesis [H3]

accepted, being that there will be positive correlations amongst the *environmental context factor scores* and *aspects of podcast listening engagement*.

4.3 Discussion

4.3.1 Hypothesis 1: Factors of environmental context

Factor interpretations

The EFA conducted in the present study permitted the acceptance of hypothesis [H1] by uncovering the presence of five latent variables representing factors of environmental context that influence attentional engagement in podcast listening experiences. These are shown together with Likert response frequencies in Figure 4.1. Factors are organised in order of most variance explained and items in order of the strongest loadings within each factor. The following section combines results from the factor loadings presented in Table 4.1 with Likert response results shown in Figure 4.1, free text response data from participants collected in the qualitative section of the study, and findings from existing literature, to produce interpretations for the five factors produced by the analysis.

Analysis of the *environmental context factor scores* and factor correlations shown in Table 4.2 indicate that there is considerable variation in participants' scores between the same factors. As such, the interpretations detailed in this analysis are presented with the understanding that the factors they represent influence different sections of the sampled population in different ways.

It is also noted that the results from this EFA should be subject to independent validation via further confirmatory factor analysis studies. Irrespective of this it is also important to consider that the results should not be considered a comprehensive model of environmental context factors that influence attentional engagement. The *factors of environmental context* produced by the EFA were derived from an initial 30-item scale originally presented to participants in the form of an online survey, and as such the findings from this analysis do not preclude additional factors from being identified and proposed by further analysis in future studies.

Outdoors

The first factor was labelled *Outdoors* and explained 12.84% of the total variance amongst all items. All four of the outdoor listening items included in the initial scale loaded together onto the factor, indicating that differences in how outdoor atmospheric conditions were perceived as influencing attentional engagement were not significant enough to cause any of the items not to load. The *outdoors in light*, *cold* and *warm conditions* items all exhibited similarly strong loadings and positive Likert ratings with *light* conditions loading most strongly, whereas the *outdoors in dark conditions* item registered a comparatively weaker loading and less positive Likert scale ratings. This was reflected by some of the free text responses provided by participants who reported experiencing lower attentional engagement when listening in outdoor public spaces in dark conditions, with the primary reason being a concern for their personal safety.

This interpretation is consistent with previous results [79] that investigated the influence of acoustic transparency on auditory mixed reality soundscapes, finding that nearly half of participants believed their safety was compromised when wearing noise cancelling headphones that occluded environmental noise, as opposed to acoustically transparent headphones that enabled enhanced perception of external stimuli in the surrounding soundscape while listening.

The *cold* and *warm conditions* item factor loadings and Likert response data suggested that temperature had minimal influence over attentional engagement. However, research suggests that while the optimum environmental temperature for higher attention changes only slightly over time from cooler to warmer, both extreme hot and cold temperatures are found to significantly lower focused attentional ability [242]. As such further research is required to fully explore the influence of environmental temperature over attentional engagement in the podcast listening experience.

Indoors & at home

The second factor was labelled *Indoors & at home* and explained 12.17% of the total variance across all items. All four of the indoors items included in the initial scale loaded together onto the factor, again indicating that differences in how indoor atmospheric conditions were perceived as influencing

attentional engagement were not significant enough to cause any of the items not to load. The *at home* item also loaded on to this factor, indicating that the factor was more specifically associated with listening at home as opposed to indoor environments more generally.

The *indoors in warm conditions* and *light conditions* exhibited by far the strongest loadings. The *at home* item was the next strongest loading with the *warm conditions*, *light conditions* and *at home* items all receiving similarly highly positive Likert scale ratings. The fourth strongest item loading, *indoors in cold conditions*, received the least positive Likert ratings in the factor, suggesting that cold conditions had more of a negative influence on attentional engagement when listening indoors rather than outdoors.

Conversely, the *indoors in dark conditions* item was the weakest loading on the factor, despite receiving similarly positive Likert ratings to the *indoors in warm*, *light* and *cold conditions* items. The *indoors in dark conditions* item Likert ratings were noticeably more positive compared to the *outdoors in dark conditions* item, further supporting the interpretation of the first factor that some participants experience lower attentional engagement when listening outdoors in dark conditions due to concerns over their personal safety.

Collectively, the *Indoors & at home* items received the most positive Likert ratings of all the factors, suggesting that the *Indoors & at home* factor is the most closely associated with higher attentional engagement. The results from the listening location questions in the present study, and a digital media consumer survey [209], both indicated that the most common location in which podcasts were consumed was the home. When considered alongside the findings of Chan-Olmsted and Wang [215], that characterised podcast listening at home as more active and instrumental [219], these results might suggest that higher attentional engagement is more closely associated with instrumental consumption where the focus of interest is centered around the specific content, than habitual ritualized consumption where the focus is centered on the medium [243].

Evenings

The third factor was labelled *Evenings* and explained 11.30% of the total variance across all items. All four of the *evenings* items included in the initial scale loaded together onto the factor, with all of the

remaining time related *morning* and *daytime* items failing to load.

The *weekend days in the late evening* and *weekdays in the late evening* items represented the strongest loadings on the factor respectively, yet notably received less positive attentional engagement Likert ratings compared to the *weekend days in the early evening* and *weekdays in the early evening* items. Despite receiving the most positive Likert ratings within the factor the *weekdays in the early evening* item registered by far the weakest loading on the factor. Furthermore, an inverse relationship was observed between the strength of loading and Likert ratings for each item, prompting the observation that items most strongly associated with the *evenings* factor were less positively associated with being actively engaged.

These results could potentially be explained, in part, by the free text qualitative responses provided by 12 participants who reported listening to podcasts when trying to fall asleep. Findings from the infinite dial podcast consumer report [209] would also support this interpretation, with statistics showing that 51% of participants surveyed had listened to podcasts when relaxing before going to sleep. This was also mirrored by open ended responses provided to a study [244] exploring young people's engagement with podcasts that indicated some young people use podcasts to help them relax and fall asleep. In fact, the popularity of the use of podcasts as a sleep aid has grown to such an extent that it is now being reflected back by creators and industry writers with podcasts being specifically produced and marketed to satiate the audience's desire for sleep inducing sounds [245].

It is also possible that the natural circadian rhythms that serve to regulate various physical, mental and behavioural changes in an individual over a 24 hour cycle could influence their capacity for higher attentional engagement at different times of the day. Valdez [246] conducted a study investigating the influence of circadian rhythms on the four components of attention outlined in the model proposed by Posner and Rafal [247]. Circadian rhythms were observed in all four components: tonic alertness, phasic alertness, selective attention, and sustained attention. Overall attention was found to increase throughout the day and was at its lowest at night and during the early morning. A review of current work relating to the influence of circadian rhythms on different aspects of auditory research concluded that circadian aspects should be given greater consideration when designing auditory experiments due to the growing breadth of experimental evidence linking circadian variations to the central and peripheral auditory systems [248].

With the exception of the *weekend days in the morning item*, all of the remaining time related items that failed to load on the factor received similarly positive Likert ratings compared to the evening items that did load. This suggests that while most participants reported experiencing largely positive attentional engagement for each time period on an individual basis, collectively it appears that the evening items were the only time related loadings with enough shared covariance to be considered an underlying factor.

Soundscape & at work

The fourth factor was labelled *Soundscape & at work* and explained 9.35% of the total variance amongst all items. The *high levels of background noise* item was the strongest loading, followed by the *moderate levels of background noise*, *environments mainly comprised of mechanical sounds*, *environments mainly comprised of sounds that indicate the presence of humans*, and the *at work* items. The *low levels of background noise* and *environments mainly comprised of natural soundscape* items did not load. Four of the five items that loaded on the factor, the *high background noise*, *mechanical soundscape*, *human soundscape* and *at work* items, received by far the most negative individual Likert ratings across all of the 30 items in the initial scale. However, the *low background* noise item received the single most positive Likert rating and the *natural soundscape* item was also reviewed favourably, despite both not loading on the factor. These results suggest that listeners tend to experience lower attentional engagement when consuming podcasts in soundscapes containing a higher concentration of disturbing stimuli. This is consistent with the results of a study by [249] that investigated how noise affected participants' performance in focused attention and cognitive search tasks, finding that intrusive noise impeded performance in a focused attention task.

However, it is notable that a significant proportion of responses to items that represented disturbing stimuli were positively associated with higher attentional engagement. Therefore, it is argued that the *soundscape & at work* factor could also be indicative of some listeners purposefully using podcasts, together with the occlusion provided by headphones, to mask disturbing elements of their surrounding soundscape and increase their capacity to experience higher attentional engagement within their own personal listening bubble.

This interpretation is consistent with research conducted by Herrmann and Johnsrude [146] which found that, over time, some listeners experience increased levels of absorption and enjoyment when listening to stories masked by multitalker babble. The interpretation could also potentially be explained further by a participant's response to the first qualitative question, who reported listening to familiar voices on their favourite podcasts to make them feel safer when they felt overwhelmed or lonely in public spaces. This response is indicative of a form of parasocial engagement where listeners form deep social bonds with hosts of their favourite podcast shows [223] and personas in the wider broadcast media more generally [250]. The formation of this factor is also consistent with analysis of the monitoring device results, shown in Table 4.5, that found the majority of participants (70.83%) most often used a headphone-based device to consume podcasts.

The results from the present study highlight the need for future work exploring how the acoustic transparency and occlusion of different headphone-based monitoring devices mediates listeners' attentional engagement in mobile listening experiences. Such research would be especially relevant to future studies investigating the soundscape masking as a function of podcast listening engagement interpretation proposed for this factor.

Exercise

The fifth factor was labelled *Exercise* and explained 9.02% of the total variance across all items. The *on a run* item was the strongest loading of any in the EFA and received similarly positive Likert ratings to the *at the gym* and *on a walk in a rural environment* items, both registering successively weaker loadings than the last. It was notable that the other item representing a form of exercise in the initial scale, *on a walk in an urban environment*, failed to load on this factor. This suggests that the factor was more closely associated with forms of exercise engaged in for the purposes of pleasure and wellbeing as opposed to more functional affordances of exercise generally associated with urban environments such as commuting or running errands. This would also support the assertion that the attentional engagement manifested in this factor is associated with the restorative influence of natural environments and stimuli on attentional engagement and cognitive control [178].

The *exercise* factor may also be indicative of the findings of research conducted by Pontifex et al. [251] that investigated the influence of exercise on attentional processes. The study found that a single bout of aerobic physical exercise had the effect of sustaining attentional processing, relative to a prolonged period of sedentary inactivity. It could be hypothesized that a relationship may exist between attentional processing ability and attentional engagement in podcast listening experiences.

4.3.2 Hypothesis 2: Associations amongst aspects of podcast listening engagement

The results of correlation tests that were run to investigate the associations amongst the *aspects of podcast listening engagement*, and permit the testing of Hypothesis [H2], are shown in Tables 4.9 and 4.10. They suggest that individuals who listen for longer periods of time on weekdays and weekend days tend to consume podcasts in a greater number of locations, engage in a larger number of multitasking activities while listening, and use a greater number of methods to discover podcasts, with weekend day listening more strongly correlated with the number of listening locations and multitasking activities compared to weekday listening. The results also suggest that individuals who listen in a larger number of locations tend to engage in more multitasking activities, use a greater number of monitoring devices, and a larger number of discovery methods. The results would also support the assertion that individuals who engage in a greater number of multitasking activities tend to use a larger number of monitoring devices and discovery methods. Finally, the results also suggest that listeners who use a greater number of methods to discover podcasts also tend to use a larger number of devices to listen.

These results support the acceptance of hypothesis [H2], showing that there are positive correlations between most of the *aspects of podcast listening engagement* measured in the present study. They build on the several aspects of podcast listening originally proposed by Tobin and Guadagno[222] to form five additional *aspects of podcast listening engagement* that constitute a congruous suite of engagement metrics that rise and fall together to represent higher or lower levels of listener engagement.

4.3.3 Hypothesis 3: Associations amongst environmental context factor scores and aspects of podcast listening engagement

The results of correlation tests that were run to investigate the associations amongst the *environmental context factor scores* and *aspects of podcast listening engagement* are shown in Tables 4.11 and 4.12. They suggest that listeners who experience higher *overall* attentional engagement tend to listen to podcasts for longer periods of time on both weekday and weekend days and also listen in a greater number of locations. Within this, the results for the constituent factors suggest that listeners who experience higher *soundscape & at work* factor attentional engagement tend to listen for longer periods of time on weekday and weekend days, listen in a greater number of locations and engage in a larger number of activities while listening, with weekend day listening more strongly correlated with *soundscape & at work* factor attentional engagement compared to weekday listening. The results also suggest that individuals who experience higher *exercise* factor attentional engagement tend to listen for longer periods of time on both weekday and weekend days. Finally, the results suggest that listeners who experience higher *evenings* factor attentional engagement tend to listen for longer periods of time on weekend days only.

These results support the acceptance of hypothesis [H3], showing that that there are positive correlations between many of the *aspects of podcast listening engagement* and *environmental context factor scores* surveyed in the present study. The *soundscape & at work* factor was found to be most closely associated with the *aspects of podcast listening engagement*, suggesting that the environmental soundscape that surrounds a listener is an especially important mediating factor for multiple facets of their podcast listening engagement. The results strongly support the assertion that consumers who listen for longer tend to experience higher attentional engagement. The finding that consumers who engage in a greater number of unique multitasking activities while listening tend to experience higher *soundscape & at work* factor attentional engagement is consistent with the *soundscape & at work* factor interpretation, in describing a factor of environmental context born of the listener's desire to audibly mask and move their attention away from disturbing sounds from stimuli in their surrounding environment.

Similarly, the finding that consumers who listen in a larger number of locations tend to experience higher *overall* and *soundscape & at work* factor attentional engagement further supports the *soundscape & at work* factor interpretation. Consumers who listen in a wider variety of locations are more likely

to be exposed to potentially disturbing sounds in their surroundings and are therefore more likely to consume podcasts in their own personal listening bubble as a means of auditory masking and escape.

4.3.4 Implications for related research fields

Podcast studies

The *aspects of podcast listening engagement* metrics were conceptualised by the present study for the primary purpose of investigating how the proposed *factors of environmental context* might be associated with a series of quantitative measures describing different facets of podcast listening engagement. However, the *aspects of podcast listening engagement* results could be highly pertinent to research in the field of podcast studies in their own right, providing new insights in podcast listening engagement, habits, and behaviours.

Greasley and Lamont [252] found that highly engaged listeners were more likely to listen to a greater amount of music in everyday life listening experiences. When considered alongside this finding, analysis of the amount of listening time data in Table 4.3 might suggest that, as listeners tend to listen for shorter amounts of time on weekend days, they are less likely to experience higher attentional engagement on weekend days compared to weekdays. However, further analysis of the correlations between the *environmental context factor scores* and amount of listening time, shown in Table 4.11, found that every factor of environmental context exhibited a stronger positive association with weekend listening compared to weekday listening.

This suggests that while listeners are less likely to engage in longer episodes of podcast listening at the weekend, they tend to experience higher attentional engagement compared to weekday listening experiences. This could potentially be explained by differences in the average length of uninterrupted listening experiences between weekday and weekend listening, as opposed to the overall cumulative amount across a single day. However, as this information was not captured in the present survey, further research is required to test the validity of this hypothesis. This is a distinction highlighted by the results of a study conducted by Herrmann and Johnsrude [146] that found listeners' absorption increased when repeatedly listening to multiple acoustically masked stories sequentially over a longer period of time.

Analysis of the listening location results in Table 4.4 suggests that while listeners possess a keen appetite for consuming podcasts in a wide variety of public locations, most consumers prefer to listen in private spaces that typically contain fewer disturbing environmental stimuli capable of negatively capturing their bottom-up attention and reducing attentional engagement. Further analysis indicates that while the vast majority of individuals consume podcasts in both public and private locations (87.88%), a small proportion of listeners only listen in either private (8.33%) or public (3.79%). Aside from reasons of personal preference, it's possible that listeners who only consume podcasts in private locations could be doing so due to personal circumstances that make it difficult for them to leave their home, or alternatively they might lack access to a headphone-based monitoring device that would allow them to listen in public spaces without disturbing others. In contrast, those who only listen in public spaces might do so as part of a daily routine while commuting or running errands. Further research is required to fully understand the listening behaviours observed in this data.

Analysis of the monitoring device results in Table 4.5 indicates that while participants tend to use a fairly equal number of different headphone- and loudspeaker-based devices to listen, the majority of participants most often consume podcasts using a headphone-based device. This finding is consistent with the assertions of Bull [33] that podcast consumers like to create personal listening bubbles to partially separate themselves from stimuli in their environmental context. The results also support assertions made by Berry [210], Heshmat et al. [253], and Schlutz and Hedder [223] that podcast listening is distinct from other audio-based media, as its strong associations with headphone listening contribute to a greater sense of intimacy in the listening experience.

The multitasking proportion results in Table 4.7 support the findings of Perks et al. [254] that podcasting is a medium closely associated with multitasking, and is both consumed as a secondary activity that competes for the listener's attention alongside an array of other often multi-sensory activities, and also on occasion the sole activity occupying listeners' full attention [215]. Analysis of the multitasking activity results shown in Table 4.6 show that listeners engage in a large number of unique multitasking activities, representing complex combinations of multi-sensory stimuli that push and pull for the listener's attention, alongside external stimuli perceptible in their surrounding environmental context. These findings support those presented by Baumgartner and Wiradhany [255] that explored the shared modalities of media multitasking. The high variance observed amongst the different activities in these results would also support assertions made by Perks and Turner [216] that podcast consumers engage

in a wide variety of different multitasking activities, in part, to satisfy different divergent gratifications that can either command or release attentional resources and consequently engage an array of different attentional and cognitive processes in the listener.

Finally, the podcast discovery methods results shown in Table 4.8 indicate that while methods belonging to the podcasting ecosystem, namely podcast apps, podcasts themselves and other forms of online media, account for a majority of methods used (72.9%), listeners use a wide variety of different methods to discover podcasts, with over 25% of discovery methods being attributed to personal recommendation and offline media. These results highlight how portable listening devices such as smartphones and tablets are not only instrumental in podcast consumption, but also podcast discovery.

Collectively, the results of this present study strongly support those from prior studies that podcast consumption occurs across a wide variety of environmental [215], situational [189], and social contexts [216].

Media personalisation

The results of the EFA in the present study are valuable to the fields of personalised and object-based media, by investigating in which contexts listeners are most focused on the listening experience and, by extension, potentially more receptive to particular types of media personalisation. Understanding how podcast consumers' listening context might relate to their attentional engagement could also be valuable to producers looking to improve content personalisation and consequently maximise the revenue their podcasts are able to generate.

Research has identified an individual's participation with a podcast at the point of consumption, either directly by interacting with the creators on social media or indirectly by researching references made during the podcast, as a key factor in a listener's user-centered engagement [224]. It also highlighted how the environmental, situational, and social context in which a podcast listening experience takes place can be a limiting factor in the listener's ability to engage in these forms of participatory engagement. For example, engaging in multitasking activities that prevent phone or computer use, such as driving, would severely limit opportunities for participatory engagement while listening to podcasts.

Similarly, contextual limitations that inhibit affordances of participatory engagement, could also limit opportunities for listener engagement with various forms of explicit object-based personalisation that can be used to enhance the listening experience by various means, including improving audience accessibility [256], altering the length of media to suit listeners' time constraints [179], or creating interactive fictional stories that are responsive to user input [257]. In this sense, implicit forms of media personalisation, that automatically respond to elements of a user's context, facilitate the delivery of personalised media experiences across a much wider range of environmental and situational listening contexts. As such they are potentially better suited to the ubiquitous nature of podcast listening experiences highlighted by Morris and Patterson [217] and the results of this present study.

Attentional processing

The influence of bottom-up and top-down attention has been shown to extend far beyond auditory attention, with extensive research and debate relating to how bottom-up and top-down attention informs a wide array of multi-sensory and cognitive attentional processes [258]. Results from this present study support assertions from existing research [217] that podcasts are consumed at a variety of times, across different locations, under a range of environmental conditions, and often in parallel with a host of different multitasking activities. As such, podcast listening experiences engage a multitude of sensory and cognitive attentional processes, born of interactions with both the listener's environmental context and the podcast itself. It is therefore important to consider how the results of the present study might relate to bottom-up and top-down attention research.

The *soundscape & at work* factor of environmental context produced by the EFA in the present study can be analysed in terms of the listener's environmental context, acting as either an active or passive influencing factor over their attentional engagement, both of which can potentially lead to higher attentional engagement by either focusing, or distracting and then focusing, the listener's attention.

If external stimuli in the listener's environment were sufficiently unobtrusive, such that it did not either positively or negatively capture their bottom-up attention, then they would best be placed to focus their goal orientated top-down attention on the podcast listening experience and better able to achieve

higher attentional engagement. In this sense, the lack of sufficiently salient stimuli in the environment would passively enable the listener to experience higher attentional engagement.

In contrast, if stimuli in the listener's environment is sufficiently salient so as to negatively capture their bottom-up attention and be perceived as distracting or stress inducing, they may employ their goal orientated top-down attention in actively choosing to mask the external soundscape by listening to podcasts using a partially or fully occlusive headphone device. This would allow them to avoid the stresses of their environmental context by escaping into their own personal listening bubble. Consequently, they would be better placed to experience higher attentional engagement in the listening experience as a result of their active decision to listen to podcasts to counter the intrusive affects of their environmental context. This could also be considered an example of top-down attention being employed to lessen the sensitivity of bottom-up attention to otherwise highly salient stimuli in the environment, and as such is analogous to the shifting interactions between bottom-up and top-down attention, and auditory attention observed by Huang and Elhilali [259], Salmi et al. [260], and Bidet-Caulet et al.[261].

Scenarios presented in this section have described ways in which a listener's bottom-up and top-down attentional processes can passively and actively interact with stimuli in their environment to facilitate higher attentional engagement in podcast listening experiences. However, as shown by the Likert response data in Figure 4.1, a significant proportion of participants also reported experiencing lower attentional engagement for the *soundscape & at work* items. Analysis of the free text qualitative responses indicates that some of the variance could be attributed to differences in the how favourably salient stimuli in the environment are perceived by the listener. It was revealed that some listeners preferred not to listen to podcasts while surrounded by natural stimuli because they preferred to focus their attention on the context of the natural environment and soundscape surrounding them. This suggests that salient environmental stimuli that captures an individual's bottom-up attention in a positive fashion, either because it is perceived as being sufficiently pleasurable or interesting, could lead to a wholesale loss in the listener's desire to focus their top-down directed attention towards engagement with a podcast listening experience.

These findings may be pertinent to research [176] in the field of attention restoration theory [262] that suggests the nature of the stimuli in an individual's surrounding physical environment can influence

their attentional engagement and other aspects of cognitive performance. Urban environments, typically comprised of a high concentration of artificial man made stimuli, often tend to elicit forms of top-down directed attention in a way that can cause fatigue, potentially leading to a decrease in attentional engagement [177]. In contrast, natural environments that contain an abundance of natural stimuli are more likely to invoke automatic bottom-up attentional states that effortlessly capture the individual's attention and have the affect of replenishing individuals' cognitive control and capacity for attentional engagement [178]. While it is important to recognise that the conclusions from these studies are not directly transferable to attentional engagement in podcast listening experiences, they do serve to illustrate how an individual's environmental context, and the nature of the stimuli contained within it, may influence their attentional engagement.

4.4 Conclusion

The results presented in this chapter are the first of their kind to support the hypothesis that a listener's environmental context exerts influence over their attentional engagement in podcast listening experiences. An online survey was used to collect data on the podcast listening engagement habits of a broad global sample. An EFA uncovered five *factors of environmental context* that both positively and negatively influence listeners' attentional engagement in podcast listening experiences. The *soundscape & at work* factor represented an especially insightful finding in its suggestion that podcast consumers actively choose to listen to podcasts as a form of soundscape masking. Separately, five *aspects of podcast listening engagement* were defined and measured across the sample, providing a comprehensive quantitative exploration of contemporary podcast listening experiences. Results together show that the proposed *factors of environmental context* are positively related to the *aspects of podcast listening engagement*, providing further validation and insight to the five defined *factors of environmental context*. Finally, the study considers how different forms of bottom-up and top-down attentional processing might relate to how environmental context influences attentional engagement in podcast listening, and the implications of the study for media personalisation. These results, in particular the identification of the *soundscape & at work* factor and the finding that wireless in-ear headphones were the monitoring device that was most often used by listeners, serve to inform the work

presented in the next chapter, wherein a series of acoustic measurements are conducted to quantify the external sound modification performance of contemporary wireless earbud devices.

Chapter 5

Measurement of Insertion Loss and Gain in True Wireless Earbuds

The results of the study presented in the previous chapter supported the findings of existing research that podcasts are consumed across a wide variety of environmental, situational, and social contexts. The results also revealed that built-in smartphone speakers were the monitoring device that the largest number of respondents (55.3% of the sample) had experience of using to consume podcasts. Wired in-ear headphones were used by the second largest number of respondents (54.9% of the sample), closely followed by wireless in-ear headphones as the device used by the third largest number of respondents (53.4% of the sample). However, wireless in-ear headphones were also found to be the device that the largest number of respondents (58.2% of the sample) reported most often using to consume podcasts. Together, these results showed that wired and wireless in-ear headphones were the most popular forms of auditory headset devices used for the consumption of podcasts. The study also identified five factors of environmental context capable of influencing listeners' attentional engagement in podcast listening experiences, with one of these being the *soundscape & at work* factor. This factor was interpreted as listening while in the presence of high or moderate levels of mechanical or human related sound while engaging in work, leading to both positive and negative attentional engagement in podcast listening experiences.

The study presented in this chapter therefore focuses on the measurement of the external sound modification capabilities of three commercially available true wireless earbud devices, commonly used

with smartphones in the mobile consumption of speech-based media. In doing so the work presented in this chapter addresses the second main aim (A_2) outlined for the present thesis in Chapter 1:

A_2 Quantify how auditory headset devices used in the consumption of speech-based media mediate the extent to which listeners are able to perceive environmental noise in their surroundings.

As discussed in Chapter 3, there have been very few studies that have measured the hear-through performance of headphone devices, while Schlieper et al. [263] conducted the only known study to measure the transfer functions of active noise cancellation and hear-through functionality in headphone devices that contain both features. However, this study only presented results of exponential sweep measurements to evaluate direction-dependent performance. There are no known studies that have used diffuse noise to evaluate active noise cancellation and hear-through performance in devices that feature both modes. There are also no known studies that have measured the performance of adjustable or variable hear-through and ANC modes in headphone devices.

The dual aims of the work presented in this chapter were therefore to quantify both the insertion loss [264] and insertion gain [118] provided by adjustable hear-through and ANC modes in commercially available true wireless earbuds that include both features. The findings of this study represent a set of case studies of devices that include this dual functionality. The consequent data would be valuable to those seeking to improve understanding of how the ability of individuals to audibly perceive their surroundings is mediated by the passive and active attenuation and boosting of external sound that is provided by these devices. The results are also expected to be useful for researchers seeking to model the external sound modification performance of leading contemporary true wireless earbud devices.

5.1 Methods

5.1.1 Inclusion criteria for devices tested

A review of commercially available true wireless earbud devices was undertaken according to the following criteria; devices should have been intended for the mass-market (not high end), released

within the past year (2023) and offer a noise cancellation mode, plus a variable hear-through mode providing different levels of hear-through functionality. This review identified several potentially suitable devices, from which a final three were selected based on positive industry reviews [265–267] of noise cancellation and hear-through performance. These devices were the Apple Airpods Pro (2nd gen.) (firmware 5B58), Bose QuietComfort Earbuds II (firmware 1.3.26+g1226f68), and Sony LinkBuds S (firmware 2.1.3). Photographs of these devices are shown in Fig. 5.1. All devices were new and purchased for use in this study. The devices' firmware and accompanying control apps were updated to the most current releases as of 20/03/23. In the remainder of the chapter these devices are anonymized and referred to as Devices A to C.

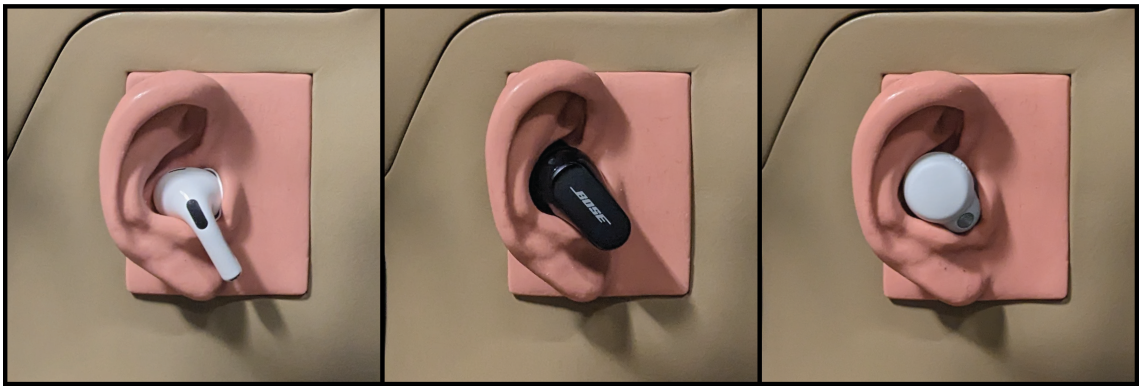


FIGURE 5.1: *Photographs of the devices inserted in KEMAR. From left to right: Apple Airpods Pro (2nd gen.), Bose QuietComfort Earbuds II, and Sony LinkBuds S.*

5.1.2 Device modes tested

The device modes measured in the study are shown in Table 5.1. Devices A and C both featured separate noise cancellation modes, off modes, and variable hear-through modes. The variable hear-through modes allow users to select between multiple stepped levels of hear-through functionality, providing them with a greater or lesser awareness of their surrounding acoustic environment. In contrast, Device B only provided a single stepped continuum of unique variable noise cancellation modes. This continuum performed similarly to the variable hear-through modes in devices A and C, allowing users the same means of achieving a greater or lesser awareness of their surrounding acoustic environment. Device B also featured *Aware* and *Quiet* modes, however measurements carried out in pilot testing, and personal communications with the manufacturer, established that the performance of these modes were identical to the variable noise cancellation modes at either end of the continuum.

TABLE 5.1: The modes tested in each device and study nomenclature used to describe equivalent modes.

Study mode nomenclature	**Device A modes**	**Device B modes**	**Device C modes**
Hear-Through Mode	Variable Hear-Through	Variable Noise Cancellation	Variable Hear-Through
- *HTM 1* - (Greater awareness)	- *VHT 1*	- *VNC 1*	- *VHT 1*
- *HTM 2*	- *VHT 2*	- *VNC 2*	- *VHT 2*
- *HTM 3*	- *VHT 3*	- *VNC 3*	- *VHT 3*
- *HTM 4*	- *VHT 4*	- *VNC 4*	- *VHT 4*
- *HTM 5*	- *VHT 5*	- *VNC 5*	- *VHT 5*
- *HTM 6* - (Lesser awareness)	- *VHT 6*	- *VNC 6**	- *VHT 6*
Noise Cancellation Mode	Noise Cancellation	Variable Noise Cancellation*	Noise Cancellation
- *NCM* - (Least awareness)	- *NC*	- *VNC 6**	- *NC*
Off Mode	Off	Battery depleted (Off)**	Off
- *Off*	- *Off*	- *Battery depleted (Off)***	- *Off*

*The results for Device B's *VNC 6* mode were used to represent both the *Noise Cancellation Mode (NCM)* and *Hear-Through Mode 6 (HTM 6)* for comparative analysis with the other devices. **As Device B did not feature a selectable *Off* mode, the passive Off Mode measurements were recorded after the battery had depleted.

For devices A and C, measurements were taken for the separate noise cancellation and off modes, in addition to six equivalent levels of the variable hear-through modes. The variable hear-through measurements were taken for the levels that provided the greatest and least awareness of surroundings, plus at four levels spaced at approximately equal intervals between these points. As Device B did not offer separate noise cancellation and variable hear-through modes, measurements of this device were taken for the *Off* mode and six of the variable noise cancellation modes. The results for the variable noise cancellation mode that provided the least awareness of acoustic surroundings was used to represent both the *Noise Cancellation Mode* and *Hear-Through Mode 6* for comparative analysis with the other devices. As Device B also did not feature a selectable *Off* mode, the *Off Mode* passive measurement for this device was recorded after the battery had depleted.

5.1.3 Acoustic test fixture

The Real Ear Attenuation at Threshold (REAT) method is the standard technique for measuring the noise attenuation provided by passive hearing protection devices [136]. Test subjects are used to quantify the shift in hearing threshold between conditions with the device fitted and removed, meaning that it also accounts for individual differences in the quality of fit across a sample of participants [268]. Despite this the REAT method is not suitable for measuring the attenuation of ANC devices as the self noise generated by the ANC system will mask the low-level REAT test signals, leading to an inaccurate assessment of attenuation performance. The use of only lower level test signals also means that it is

not possible to measure attenuation performance for higher levels of environmental noise. Rudzyn & Fisher [136] therefore suggested the two main physical measurement methods, of Microphone In Real Ear (MIRE) and Acoustic Test Fixture (ATF), would be better options for measuring the attenuation provided by ANC devices. The MIRE method is most frequently used to measure the performance of hearing aid devices and while it does offer the advantage of accounting for variation of fit within a sample of participants, the ATF method is more suitable for the measurement of ANC devices.

As the ATF method does not require any human subjects it is possible to test the devices up to the limits of their performance, with no risk of injury to participants from exposure to potentially dangerous sound pressure levels. It is also much easier to maintain tight control of test conditions, enabling a higher level of consistency between measurements, thus providing a higher level of accuracy when analysing the performance of different devices. ATF methods have shown to be equally applicable for the measurement of both insertion loss and gain across a range of devices and applications. Crabtree & Behar [269] used an ATF to measure the insertion loss of passive hearing protectors up to ultrasonic frequencies, Cui et al. [264] used ATF to measure the insertion loss of active noise reduction headsets, while Denk et al. [118] used ATF to measure the insertion gain of several hearables in their study on acoustic transparency.

It should be stated however that as ATFs are not able to perfectly replicate all characteristics of the human head and auditory system, aspects of performance such as bone-conduction sound transmission are not accounted for, and as a result the level of attenuation measured using an ATF will be exaggerated compared to what is experienced by real listeners. Despite this, the ATF method was still considered the best of the available options and was therefore selected for use in the present study. The ATF method that was used in a study [136] conducted by Rudzyn & Fisher, on the performance of personal active noise reduction (ANR) devices, consisted of the following steps:

(a) Generate a diffuse sound field in a large reverberation room using pink noise equalised to within ±1 dB from 40 Hz to 10 kHz and within −3 dB at 31.5 Hz.
(b) Record the open ear (OE) response of a Head And Torso Simulator (HATS) in this diffuse field using pink noise with broadband levels of 80–110 dB SPL incremented in 10 dB steps. Record third-octave levels from 31.5 Hz to 10 kHz, linearly averaged over 30 s. Record spectrum for both left and right ears (to calculate average).

(c) Fit the test device on HATS and determine passive response (PR) with ANR off (same recording regime as for step (b)).

(d) Turn ANR system on and determine total response (TR) of the device (same recording regime as for step (b)).

(e) Calculate broadband attenuation:

  (1) Passive attenuation = OE − PR

  (2) Total attenuation = OE − TR

  (3) Active attenuation = PR − TR

(f) Repeat for each device being tested.

5.2 Spectral modification of broadband noise

The term spectral modification is used in this study to describe any attenuation or boosting of external noise introduced by the devices under test. *Passive Spectral Modification* (PSM) is used to describe both the attenuation and resonant peaks caused by the physical occlusion of the devices in *Off Mode*, *Total Spectral Modification* (TSM) describes the insertion loss and gain provided by the combination of passive occlusion and active noise cancellation/hear-through modes, and *Active Spectral Modification* (ASM) is used to describe both the attenuation and boosting that is provided by the active electronics of the noise cancellation/hear-through modes in isolation, without the additional attenuation/resonant peaks from the passive occlusion.

5.2.1 Spectral modification of broadband noise method

The following method for measuring the spectral modification of broadband noise provided by the passive and active modes in the devices was adapted from the methodology proposed by Rudzyn and Fisher [136]. The resulting methodology largely adheres to ETSI TS 103 640 [91], however, limitations with the available materials resulted in some unavoidable deviations. These include exceeding the specified noisefloor by 15 dB, the use of a larger loudspeaker array, and the use of the 30 s analysis duration specified in [136]. The method in the present study comprised of the following steps:

(a) Generate a diffuse sound field in a listening room using uncorrelated pink noise signals (see below for more details).

(b) Record 35 s of the open ear (OE) response of a GRAS 45BC KEMAR ATF in the diffuse field at broadband levels of 65, 70, 80, 90, 100, and 105 dB sound pressure level (SPL).

(c) Fit earbuds in the pinna of KEMAR and record passive response (PR) with *Off* mode selected (same recording process as used in step (b)).

(d) Record total response (TR) of the device for each active mode listed in Table 5.1 (same recording process as used in step (b)).

(e) Remove the earbuds from KEMAR and repeat steps (c) to (e) for a total of five iterations.

(f) Analyse recordings to determine the measurement iteration with the best fit (see below for more details).

(g) Linearly average over the selected measurement from 5 to 35 s to attain stable third-octave root mean square (RMS) levels [270] [271] of the open ear, passive, and total responses for all modes for that iteration.

(h) Calculate broadband noise spectral modification from third-octave RMS level results for all device modes listed in Table 5.1:

  (1) Passive spectral modification = OE $-$ PR

  (2) Total spectral modification = OE $-$ TR

  (3) Active spectral modification = PR $-$ TR

(i) Repeat steps (c) to (h) for all devices being tested.

The broadband noise signal was generated in Matlab as a 50 channel, 35 s stem of uncorrelated pink noise. This was then fed to a 50 point Lebedev grid loudspeaker array using Reaper [272] and two Ferrofish A32 AD/DA converters, located in the listening room (128.7 m^3) at the Audiolab, University of York. The noisefloor was measured at 30 dBA SPL. The array comprised of 40 Genelec 8030A and 10 Genelec 8040A active loudspeakers. An NTi XL2 Analyzer and M4260 measurement microphone were used to calibrate the broadband noise (LAeq) at the centre of the grid for the different presentation levels. The spectrum was found to be flat within $\pm$5 dB for all bands of interest. This was considered an adequate approximation of diffuse field conditions for the purposes of the present study, in accordance with measurement procedures from past studies [118, 137, 273]. The signal was measured with left and right prepolarized ear simulators (RA0045-S1) fitted in a GRAS 45BC KEMAR ATF, the KEMAR

was equipped with large (VA-Style) pinnae [274]. The output from each simulator was then passed through a Presonus DigiMax DP88 analog-to-digital converter and recorded using Reaper.

Earbud fit

Denk et el. [118] and Struck [137] highlighted the importance of conducting multiple iterations of measurements to ensure a tight physical fit and achieve optimal passive attenuation performance. Each device featured a built-in app-based tool designed to check the quality of physical fit for either earbud. These tools were not used in the present study as they were found to provide unreliable measures of the quality of fit in KEMAR. Therefore the medium earbud tips/fins were used for all devices. Fig. 5.1 shows the earbuds inserted in KEMAR. Measurements were repeated for a total of five iterations with the earbuds removed and reinserted between each iteration. The method for determining the measurement iteration with the best fit involved attaining the third-octave RMS levels of the passive responses (*Off* mode) from both ears for each iteration, and then performing an analysis to identify the iteration that demonstrated the broadest and most consistent range of attenuation and boosting of diffuse broadband noise for both ears [118].

Fig. 5.4 shows the PSM for the *Off* modes of the measurement iterations that were selected for further analysis in section 5.2.2. These results illustrate that a broadly similar physical fit and level of passive attenuation was achieved for both left and right ears in all devices, however, there were a few instances where ASM results for the right ear were not consistent with the left ear. It was thought this was due to general inconsistencies in the performance of the devices and differences in the quality of earbud fit between KEMAR's left and right pinna. As such these results were not considered representative of typical TSM performance in the devices and therefore all of the forthcoming results in this chapter were derived from left ear device measurements only unless stated otherwise.

5.2.2 Spectral modification of broadband noise results

The following terminology, illustrated in Figs. 5.2 & 5.3, was adapted from a study by Rudzyn & Fisher [136] and used to describe and analyse properties of the active and total spectral modification results in this study:

- Attenuation amount - the maximum attenuation in dB provided by the device mode.
- Attenuation frequency - the central frequency in Hz (third-octave) at which the attenuation was largest for the given mode.
- Attenuation region - the breadth of the attenuation surrounding the central attenuation frequency in Hz, calculated as the span across which the attenuation was less than -3 dB. When the attenuation is indicative of a shelf rather than a band-stop the region is given as the frequency at which the attenuation first exceeds -3 dB.
- Modification range - the difference in dB between the attenuation and boosting amounts for the device mode.
- Boosting amount (Peak) - the maximum boosting in dB provided by the device mode when boosting is ≥ 0 dB; in cases where there is no boosting in a mode (Peak) this term is used to describe the amount of attenuation in the band with the lowest insertion loss.
- Boosting frequency - the central frequency in Hz (third-octave) at which the boosting was largest for the given mode.
- Boosting region - the breadth of the boosting surrounding the central boosting frequency in Hz, calculated as the span across which the boosting exceeds 3 dB. When the boosting is indicative of a shelf rather than a band-pass the region is given as the frequency at which the attenuation first exceeds 3 dB.

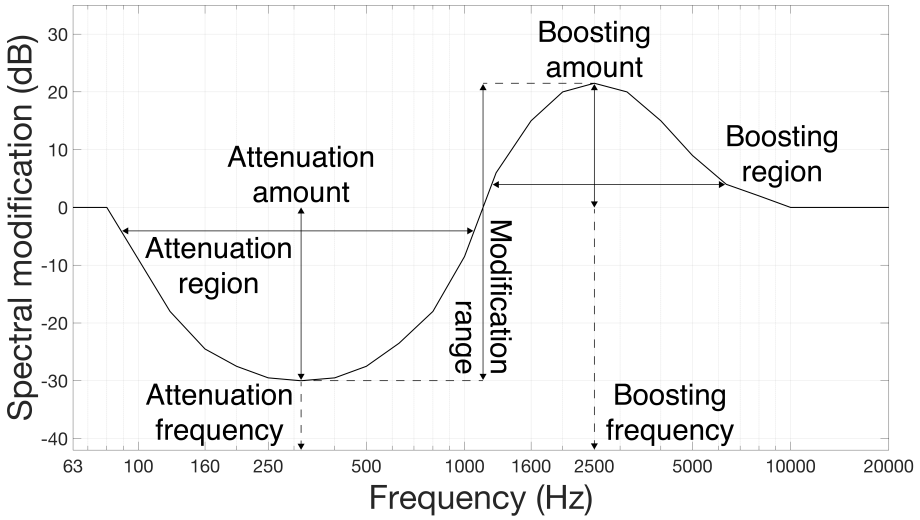


FIGURE 5.2: *Diagram illustrating the spectral modification properties analysis terminology. The line plotted on this graph is not representative of any real data.*

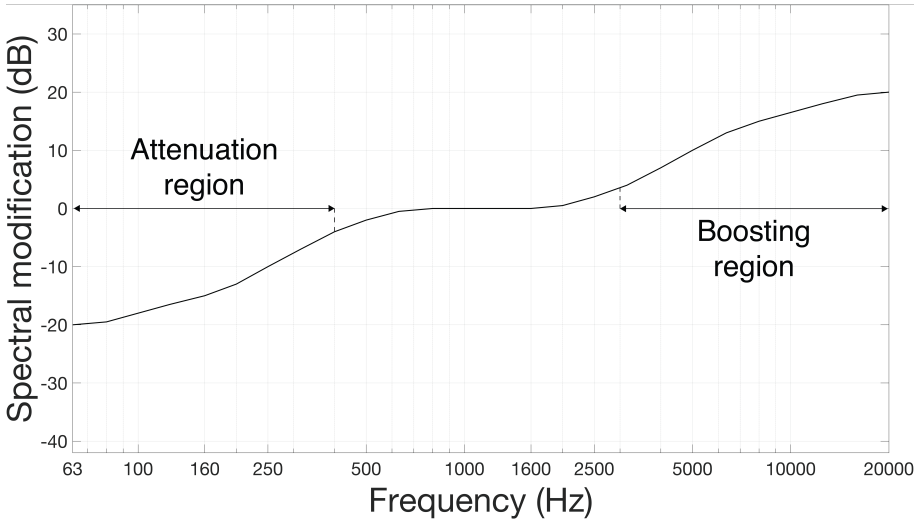


FIGURE 5.3: *Diagram illustrating the spectral modification attenuation and boosting region properties in cases where the modification is indicative of a shelf rather than a band.*

This same terminology is also used to analyse the passive spectral modulation results, with the exception of the term 'resonant peak' being used in place of 'boosting' where applicable. This section will first consider the PSM results to analyse and compare differences in the insertion loss due to the passive occlusion of the devices. The TSM results will then be analysed to establish the nature of the spectral modification as it would be perceived at the ear by a real listener. The ASM results will also be presented to quantify and differentiate between the spectral modification provided by the passive occlusion of the devices and the active electronics in the noise cancellation and variable hear-through modes. No effort was made to speculate on the nature of the active electronics or processing contained within each device as this was not considered relevant to the aims of the study.

Figs. 5.4, 5.5, and 5.6 show the passive, total, and active spectral modification in third-octave bands provided by the devices when subject to 80 dB SPL pink noise in an approximated diffuse field. The frequency range of the measurements was limited to 63-20,000 Hz due to the frequency responses of the loudspeakers. The results attained from the 80 dB SPL measurements are highlighted in the forthcoming analysis as they demonstrated the broadest range of both attenuation and boosting across all of the device modes.

The spectral modification performance of most device modes was found to be dependent on the presentation level of the broadband noise, however, the general characteristics of the spectral modification observed in the 80 dB SPL measurements remained consistent for all hear-through modes, in all devices, across all of the presentation levels tested. The general characteristics of spectral modification

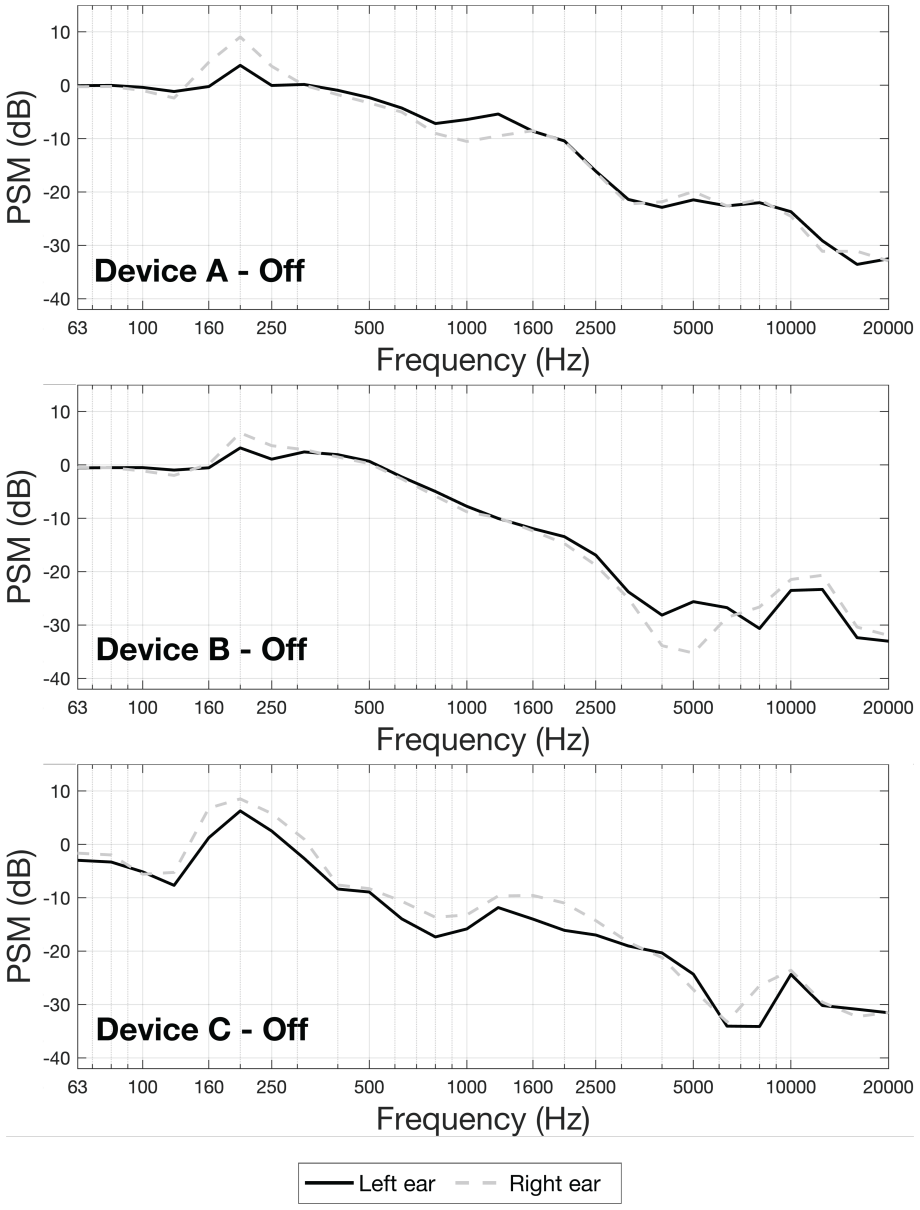


FIGURE 5.4: *Passive spectral modification of diffuse pink noise at 80 dB SPL for Off modes for all devices.*

observed at 80 dB SPL in the noise cancellation mode of Device C also remained consistent across all presentation levels. In contrast, the other devices exhibited significant differences in noise cancellation mode performance at higher presentation levels, with Device B maintaining consistency from 65 to 100 dB SPL, and Device A from 65 to 90 dB SPL. Further details of how the device modes responded at different presentation levels are provided in section 5.2.3.

Passive spectral modification results

PSM results: Off modes

Fig. 5.4 shows the PSM provided by the passive occlusion of the devices at 80 dB SPL in third-octave bands. The results of the accompanying PSM properties analysis are presented in Table 5.2. The results for all three devices showed a notable boosting of frequencies centred around the 200 Hz band. This was the most pronounced in Device C with a boost of 6.27 dB, followed by Device A (3.73 dB), and Device B (3.19 dB). For all devices these regions of boosting only spanned the single band of 200 Hz, according to the analysis criteria of boosting breadth $\geq$ 3 dB adapted from Rudzyn & Fisher [136]. This boosting can be attributed to passive resonances created by the insertion of the devices in the ear canal(s) of the ATF, with the size and position of the passive resonant peak being dictated by the amount of acoustic leakage, volume of the cavity created, and the absorptivity of the earbud tip materials [275]. It should be noted that these results are also influenced by the design of the ear canal extension in the KEMAR ATF that was used in the measurements.

TABLE 5.2: Passive spectral modification properties analysis of results shown in Fig. 5.4.

		Attenuation			Modification	Boosting (Peak)		
Device	**Mode**	**Amount (dB)**	**Freq. (Hz)**	**Region (Hz)**	**Range (dB)**	**Amount (dB)**	**Freq. (Hz)**	**Region (Hz)**
A	Off	-33.57	16000	$\geq$ 630	37.30	3.73	200	200*
B	Off	-33.03	20000	$\geq$ 800	36.22	3.19	200	200*
C	Off	-34.13	8000	$\geq$ 400	40.39	6.27	200	200*

*This region only occupied a single third octave band.

There were similar consistencies between the devices observed in the nature of attenuation across the rest of the frequency spectrum. All devices displayed similar attenuation characteristics whereby the cutoff frequencies of the main attenuation shelves began in the region of 400-800 Hz. The shelves then followed a broadly similar trend with the level of attenuation increasing for successively higher frequency bands, with the exceptions of slight peaks around 1250 Hz in devices A and C. Similarly minimal peaks can be observed at 5 kHz in devices A and B, and more pronounced peaks around 10 kHz in devices B and C. All devices also registered similar maximum levels of passive attenuation,

however, the frequency bands at which these occurred differed, with Device C providing -34.12 dB at 8 kHz, Device A -33.57 dB at 16 kHz, and Device B -33.03 dB at 20 kHz.

Total spectral modification results

The plots in Fig. 5.5 show the insertion loss and gain provided by the noise cancellation and hear-through modes of the devices. The results of the accompanying TSM properties analysis are presented in Table 5.3.

TABLE 5.3: Total spectral modification properties analysis of results shown in Fig. 5.5.

		Attenuation			Modification	Boosting (Peak)		
Device	**Mode**	**Amount (dB)**	**Freq. (Hz)**	**Region (Hz)**	**Range (dB)**	**Amount (dB)**	**Freq. (Hz)**	**Region (Hz)**
A					45.81			
	HTM 1	-19.91	16000	$\geq$ 5000	34.26	14.37	2500	2000-4000
	HTM 2	-20.11	16000	$\geq$ 5000	33.30	13.19	2500	2000-4000
	HTM 3	-21.25	16000	$\geq$ 5000	33.26	12.01	2500	2000-4000
	HTM 4	-22.34	16000	$\geq$ 5000	33.21	10.88	2500	2000-3150
	HTM 5	-23.46	16000	$\geq$ 5000	33.16	9.70	2500	2000-3150
	HTM 6	-24.53	16000	$\geq$ 5000	33.07	8.54	2500	2000-3150
	NCM	-31.46	160	$\geq$ 63	22.80	-8.65 (Peak)	3150	No boosting
B					48.67			
	HTM 1	-28.24	20000	$\geq$ 5000	42.01	13.77	2000	1250-3150
	HTM 2	-29.78	20000	$\geq$ 4000	40.77	10.98	2000	1600-2500
	HTM 3	-31.05	20000	$\geq$ 4000	38.62	7.57	2000	2000-2500
	HTM 4	-31.91	20000	$\geq$ 4000	35.03	3.12	2000	2000*
	HTM 5	-32.34	20000	$\geq$ 63	29.19	-3.15 (Peak)	2000	No boosting
	*HTM 6***	-34.90	125	$\geq$ 63	21.23	-13.67 (Peak)	1250	No boosting
	*NCM***	”	”	”	”	”	”	”
C					44.65			
	HTM 1	-30.11	16000	$\geq$ 5000	42.39	12.28	800	630-3150
	HTM 2	-30.57	20000	$\geq$ 4000	40.07	9.50	800	630-3150
	HTM 3	-31.14	20000	$\geq$ 4000	37.07	5.93	800	800-2500
	HTM 4	-31.34	20000	$\geq$ 4000	34.74	3.39	800	800*
	HTM 5	-31.48	20000	$\geq$ 3150	31.72	0.24	800	800*
	HTM 6	-31.54	20000	$\geq$ 1000	29.04	-2.50 (Peak)	800	No boosting
	NCM	-32.37	8000	$\geq$ 1000	29.61	-2.77 (Peak)	800	No boosting

*This region only occupied a single third octave band. **The results for Device B's *VNC 10* mode were used to represent both the *Noise Cancellation Mode (NCM)* and *Hear-Through Mode 6 (HTM 6)* for comparative analysis with the other devices, as described in Section 5.1.2.

TSM results: Noise cancellation modes

For frequencies above 5 kHz the TSM remained broadly unchanged from the underlying PSM for each device described in section 5.2.2. This was with the exception of slight peaks at 8 kHz and 20 kHz in Device A, and 6350 Hz in devices B and C. The maximum level of attenuation achieved in a

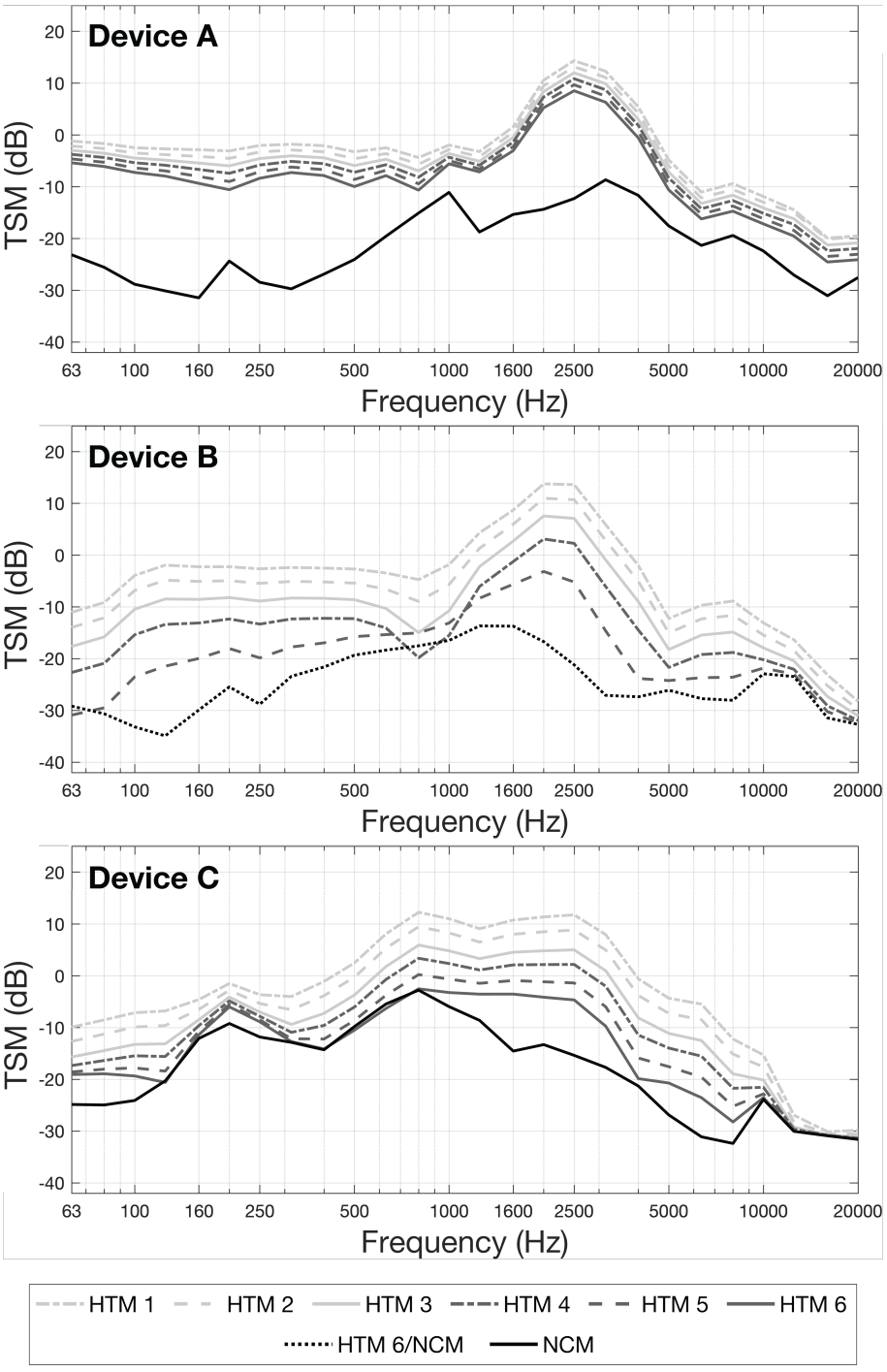


FIGURE 5.5: *Total spectral modification of diffuse pink noise at 80 dB SPL for noise cancellation and hear-through modes for all devices.*

single frequency band was -34.90 dB at 125 Hz in Device B. This compared with -32.37 dB at 8 kHz in Device C and -31.46 dB at 160 Hz in Device A.

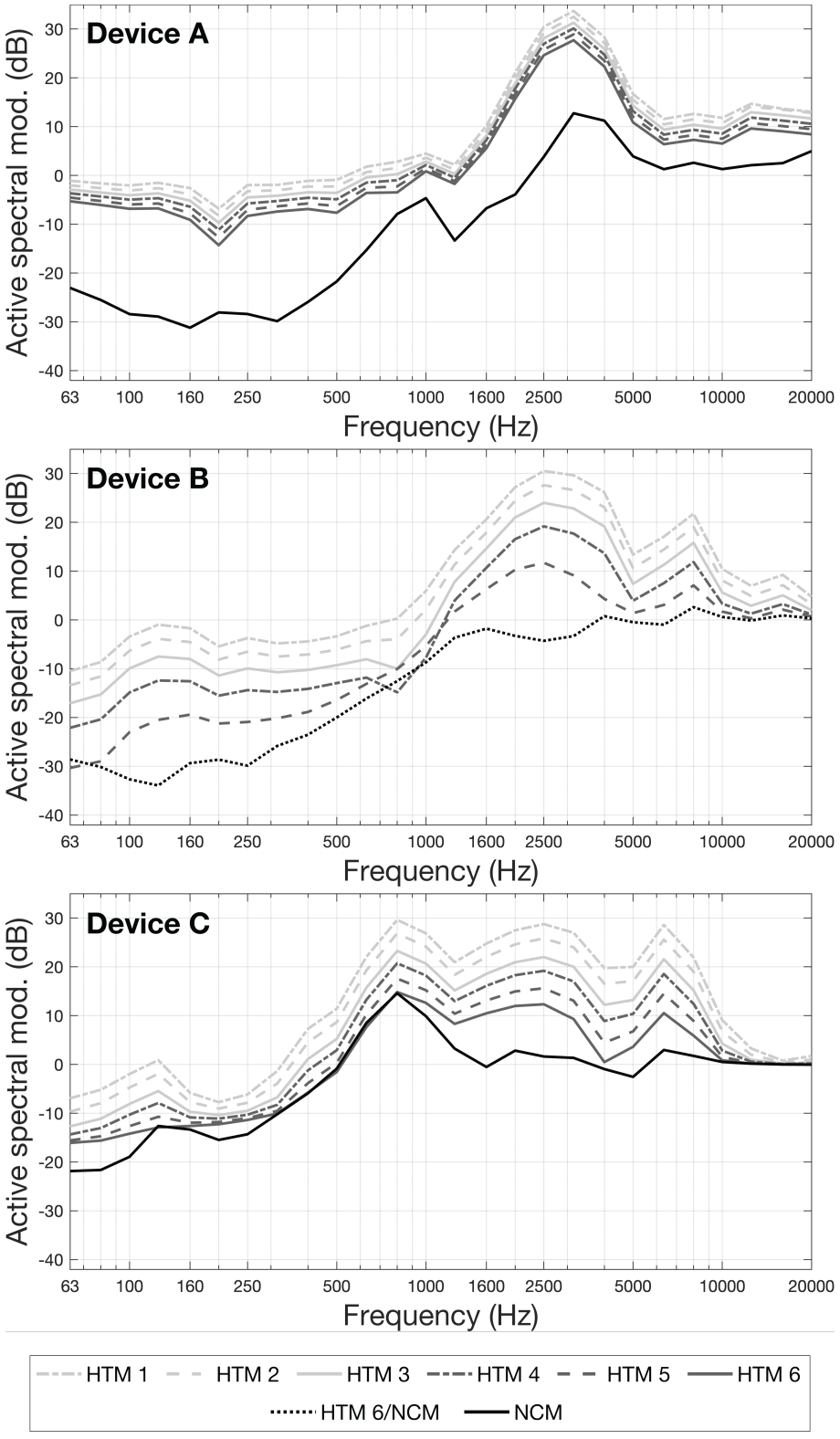


FIGURE 5.6: *Active spectral modification of diffuse pink noise at 80 dB SPL for noise cancellation and hear-through modes for all devices.*

TSM results: Hear-through modes

All six hear-through modes in Device A had singular regions of boosting ranging from a maximum amount of 14.37 dB for *HTM 1* to 8.54 dB for *HTM 6*, with all boosting regions centered around 2500 Hz and ranging from approximately 2000-4000 Hz for *HTM 1-3* and 2000-3150 Hz for *HTM 4-6*. This was contrasted by Device B where only the first four hear-through modes *(HTM 1-4)* featured any boosting, ranging from a maxima of 13.77 dB for *HTM 1* to 3.12 dB for *HTM 4*. All regions were centred around 2 kHz, however, the boosting region was successively smaller for each mode, ranging from 1250-3150 Hz for *HTM 1*, compared with just a single band at 2 kHz for *HTM 4*.

The hear-through modes in Device C also featured singular regions of boosting, however, they were much wider, ranging from 630-3150 Hz for *HTM 1* and centred around two main peaks at 800 and 2500 Hz for *HTM 1-5* with slight dips in the mainly flat plateau between the peaks at 1250 Hz. At 800 Hz this ranged from maximum amounts of 12.28 dB for *HTM 1* to 0.24 dB in *HTM 5*, while the boosting of the peaks centered around 2500 Hz ranged from 11.78 dB for *HTM 1* to 2.20 dB for *HTM 4*.

In the frequencies below the singular regions of boosting for the hear-through modes the boosting from the passive resonant peaks at 200 Hz is countered by broadly equal and opposite amounts of attenuation from the active electronics. This is visible in the PSM *Off* mode plots in Fig. 5.4 and the ASM noise cancellation and hear-through plots in Fig. 5.6. Consequently the regions from 63-1250 Hz in Device A, 125-1000 Hz in Device B, and 160-400 Hz in Device C maintain a relatively flat TSM response for *HTM 1* modes, with the largest attenuation deviations from 0 dB for each device being -4.72 dB and -4.36 dB at 800 Hz in devices B and A respectively, and -4.58 at 160 Hz for Device C. For the frequencies below these regions in devices B and C attenuation then increases with each successively lower frequency band, arriving at 63 Hz with a local minima of -11.03 dB for Device B and -9.92 dB for Device C.

TSM in the frequencies above the singular regions of boosting in the hear-through modes followed a broadly similar pattern for all devices, with the level of attenuation increasing for each successively higher frequency band. The only exceptions to this were slight peaks at 8 kHz for all hear-through modes in Device A and *HTM 1-4* for Device B, and at 10 kHz in *HTM 5* of Device B and *HTM 5* and *6* of Device C. The increases in attenuation then plateau for the *HTM 1* in devices A and C from

16-20 kHz at approximately -20 dB and -30 dB respectively. Device B's hear-through maintained the increase in attenuation for the remainder of the spectrum before reaching a maxima of -28.24 dB at 20 kHz in *HTM 1*.

The general shape of the TSM remained very consistent for all of the hear-through modes in devices A and C, with the only exceptions being a slight widening between the levels of attenuation provided by each mode from 160-800 Hz for Device A, and narrowing between the TSM for the hear-through modes in Device C from 63-100 Hz, 160-250 Hz, and 10-20k Hz. There was an average difference of approximately 1 dB between each hear-through mode in Device A, contrasted by an average difference of 2 dB for the hear-through modes in Device C.

Device B demonstrated a larger degree of variation in the shape of TSM between the hear-through modes. *HTM 1-4* maintained a broadly consistent shape, with the exception of a narrowing between modes from 8-20 kHz, and an average difference of approximately 3 dB. The shape of TSM for *HTM 5* was also broadly consistent with *HTM 1-4* from 63-400 Hz and 1250-20,000 Hz, however, from 400-1250 Hz it deviated, staying much closer to the shape observed in *HTM 6/NCM*, and consequently crossing over with *HTM 4* such that *HTM 4* exhibited 4 dB more attenuation than *HTM 5* at 800 Hz. Moreover the crossover observed in the *HTM 4* measurement at 800 Hz was to such an extent that it also registered 1 dB more attenuation than the *HTM 6/NCM* mode at 800 Hz. However, inspection of the equivalent ASM results in the Device B plot of Fig. 5.6 might suggest that this crossover phenomenon was due to deviations in the shape of attenuation in *HTM 3* and *4*, manifested as two almost identical notches of increased attenuation at 800 Hz. Similar behaviour was also observed at 800 Hz in the results of Device B's *HTM 1* TSM performance for 105 dB SPL, this is shown in Fig. 5.10 and discussed in section 5.2.3.

Active spectral modification results

The plots in Fig. 5.6 show the attenuation and boosting provided by the active electronics in the noise cancellation and hear-through modes of the devices. The accompanying results of the ASM properties analysis are presented in Table 5.4. This data is presented alongside the PSM results to assist

researchers who wish to create models that accurately replicate the spectral modification performance of a typical true wireless earbud device.

TABLE 5.4: Active spectral modification properties analysis of results shown in Fig. 5.6.

		Attenuation			Modification	Boosting (Peak)		
Device	**Mode**	**Amount (dB)**	**Freq. (Hz)**	**Region (Hz)**	**Range (dB)**	**Amount (dB)**	**Freq. (Hz)**	**Region (Hz)**
A					64.92			
	HTM 1	-6.81	200	200*	40.52	33.71	3150	$\geq$ 1000
	HTM 2	-8.25	200	160-315	40.75	32.51	3150	$\geq$ 1000
	HTM 3	-9.70	200	80-500	41.00	31.30	3150	$\geq$ 1000
	HTM 4	-11.13	200	$\leq$ 500	41.26	30.13	3150	$\geq$ 1000
	HTM 5	-12.72	200	$\leq$ 500	41.64	28.91	3150	$\geq$ 1600
	HTM 6	-14.30	200	$\leq$ 800	42.02	27.72	3150	$\geq$ 1600
	NCM	-31.21	160	$\leq$ 2000	43.95	12.74	3150	1250-5000
B					64.46			
	HTM 1	-10.47	63	$\leq$ 100	41.00	30.53	2500	$\geq$ 1000
	HTM 2	-13.40	63	$\leq$ 800	41.02	27.61	2500	$\geq$ 1250
	HTM 3	-17.08	63	$\leq$ 800	41.09	24.00	2500	1250-10000
	HTM 4	-22.11	63	$\leq$ 1000	41.28	19.17	2500	1250-10000
	HTM 5	-30.33	63	$\leq$ 1000	42.05	11.72	2500	1600-4000
	*HTM 6***	-33.94	125	$\leq$ 1250	36.56	2.62	8000	8000*
	*NCM***	”	”	”	”	”	”	”
C					51.49			
	HTM 1	-7.74	200	160-250	37.37	29.63	800	400-12500
	HTM 2	-9.68	63	$\leq$ 100	36.52	26.84	800	400-10000
	HTM 3	-12.69	63	$\leq$ 315	35.97	23.28	800	500-10000
	HTM 4	-14.35	63	$\leq$ 315	35.08	20.74	800	630-8000
	HTM 5	-15.58	63	$\leq$ 400	33.17	17.59	800	630-8000
	HTM 6	-16.08	63	$\leq$ 400	30.92	14.84	800	630-3150
	NCM	-21.86	63	$\leq$ 400	36.44	14.58	800	630-1250

*This region only occupied a single third octave band. **The results for Device B's *VNC 10* mode were used to represent both the *Noise Cancellation Mode (NCM)* and *Hear-Through Mode 6 (HTM 6)* for comparative analysis with the other devices, as described in Section 5.1.2.

5.2.3 Presentation level linearity

Studies that have measured noise cancellation [136] and hear-through [118] performance in headphone devices have conducted measurements using signals at a range of different presentation levels. One of the main reasons for this is to assess how well devices are able to attenuate noise in excess of 80 dB SPL, where prolonged exposure could potentially cause hearing damage [276]. However, as the results in section 5.2.2 of the present study showed that all devices tested provided substantial amounts of both insertion loss and gain relative to the open ear at 80 dB SPL, it is considered especially important to measure how the spectral modification performance might differ depending on the presentation level of the noise source.

Presentation level linearity results

Fig. 5.7 shows the PSM of diffuse pink noise for *Off* modes in all devices at presentation levels of 65-105 dB SPL, while Figs. 5.8 and 5.10 show the TSM of diffuse pink noise for Noise Cancellation and Hear-Through 1 modes in all devices across the same range of presentation levels. As the results in section 5.2.2 showed that the general shape of TSM remained broadly consistent for all of the hear-through modes tested in the devices it was decided that analysis of the linearity results for *HTM 1* alone would be sufficient to assess the presentation level dependency of all hear-through modes in the devices.

Linearity results: Off mode

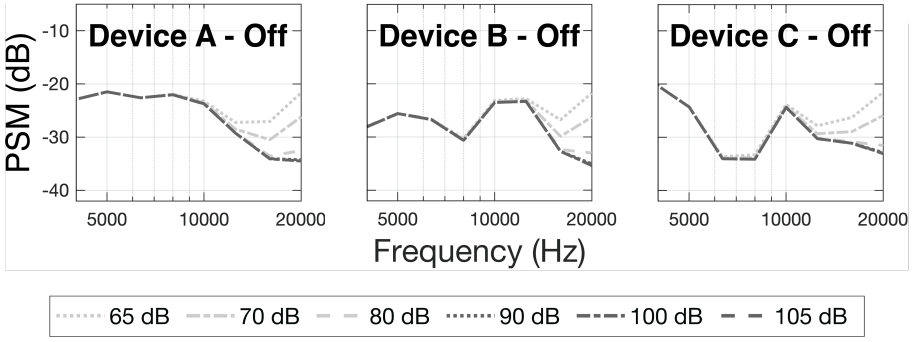


FIGURE 5.7: *Passive spectral modification of diffuse pink noise at presentation levels of 65 to 105 dB SPL for Off modes for all devices.*

The results presented in Fig. 5.7 show that the PSM remains consistent for all presentation levels tested from 63-10,000 Hz in devices A and C, and from 63-12500 Hz for Device B. In frequencies above these regions the level of attenuation appeared to increase with higher presentation levels. This behaviour was found to be similar for all devices across the full range of presentation levels tested. Maximum levels of attenuation were achieved for presentation levels of 90-105 dB SPL, with all the devices demonstrating consistent attenuation from 63-20,000 Hz. At 80 dB SPL the devices provided approximately 2 dB less attenuation at 20 kHz, while at presentation levels of 75 and 65 dB SPL devices A and C both provided successively lower levels of attenuation from 12500-20,000Hz, with Device B demonstrating similar performance for 16-20 kHz. While these results could be indicative of a form of intentional level dependent behaviour [141] where higher frequency noise is subject to greater attenuation at higher levels, it is also possible that they could be a symptom of device self-noise

and/or KEMAR's microphone noise limiting the measurable attenuation of higher frequencies at lower presentation levels.

Linearity results: NCM

The plots in Fig. 5.8 show TSM performance was dependent on presentation level for the *NCM* modes in all devices to varying degrees. Aside from the level dependent PSM behaviour between 12500-20,000Hz, the TSM provided by Device C's *NCM* remained consistent across all presentation levels tested. This was contrasted by Device A's *NCM* where the TSM provided two distinctly different shapes of TSM depending on the presentation level.

For presentation levels of 65 and 70 dB SPL the attenuation in the TSM of Device A's *NCM* followed a very similar shape and featured a singular main peak at 2 kHz, around which there was a region from 1-4 kHz where attenuation was smallest. This was contrasted by the results for presentation levels of 80-105 dB SPL where there were two separate regions where attenuation was least prominent centered around peaks at 1000 and 3150 Hz. The other main difference between the TSM in Device A's *NCM* for presentation levels from 65-70 and 80-105 dB SPL was in the attenuation of frequencies below 200 Hz. In this region, attenuation levels were the highest for the presentation levels of 80-105 dB SPL, while attenuation levels were successively lower for 70 and 65 dB SPL, mirroring the level dependent behaviour observed in the higher frequencies for the PSM and TSM results.

The TSM in the *NCM* of Device B also exhibited level dependent behaviour, however this didn't follow the same trend of greater attenuation for higher presentation levels observed in devices A and C, and therefore may be more likely due to overload rather than by design. The TSM attenuation was consistent for all presentation levels between 6350-10,000 Hz and for frequencies above this followed broadly the same level dependent behaviour as observed in the PSM results. From 200-6350 Hz TSM results in Device B also remained largely consistent for presentation levels of 65-80 and 100 dB SPL, and for frequencies below this, demonstrated similar behaviour to the other devices with more attenuation for higher presentation levels. However, the 90 and 105 dB SPL results exhibited significant differences from 63-6150 Hz that were indicative of two different kinds of inconsistency in their performance.

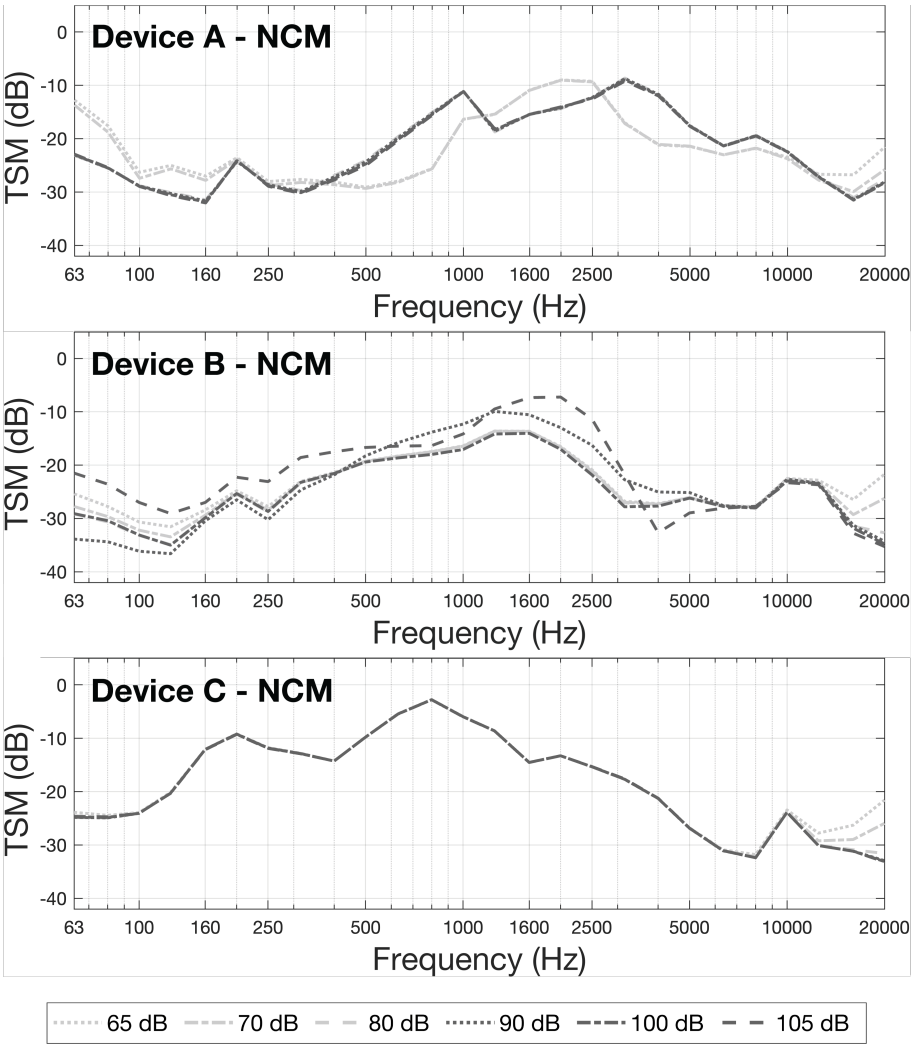


FIGURE 5.8: *Total spectral modification of diffuse pink noise at presentation levels of 65 to 105 dB SPL for Noise cancellation modes for all devices.*

The fact that the results for the 100 dB SPL Device B *NCM* TSM measurement remained broadly consistent with the equivalent results for 65-80 dB, yet the shape of the 90 dB TSM results featured significant deviations, is potentially indicative of temperamental performance in the device. However, it is more likely that the inconsistencies observed in the 105 dB SPL TSM results were due to Device B's *NCM* reaching its overload limit [136]. This interpretation is supported by a spectrogram analysis of the raw pink noise recordings for the *NCM* in Device B, which found temporal discontinuities in both left and right ear devices at 105 dB SPL. This is shown in Fig. 5.9.

Further spectrogram analysis of the raw recordings of *NCM* responses for all devices at all presentation levels found no evidence of similar discontinuities that would be potentially indicative of overload performance.

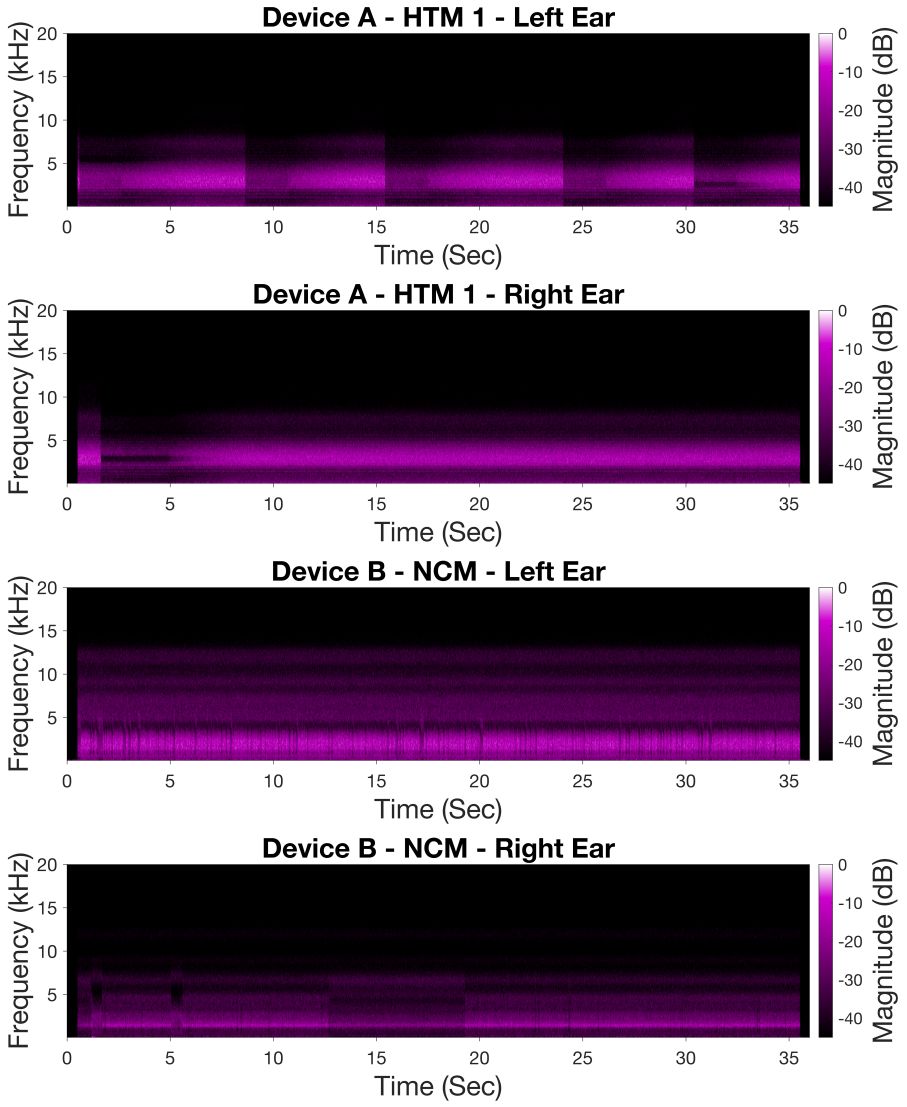


FIGURE 5.9: *Spectrograms showing temporal discontinuities in the performance of devices A and B at a presentation level of 105 dB SPL. Discontinuities are visible approximately every 6-8 s in the left ear of Device A's HTM 1response and from 1-5 s in the equivalent right ear response. For Device B's NCM, discontinuities are visible as vertical lines below 5 kHz throughout the duration of the recording in the left ear response and instantaneous shifts in the frequency spectrum at approximately 1-2, 5-6, and 13-19 s in the right ear response.*

Linearity results: HTM 1

The plots in Fig. 5.10 show that the TSM for the hear-through 1 modes in all devices exhibited level dependent behaviour. In Device A's *HTM 1* the TSM was consistent from 63-16,000 Hz for presentation levels of 65-100 dB SPL.

At 20 kHz, level dependent behaviour consistent with the PSM results was observed for the lower

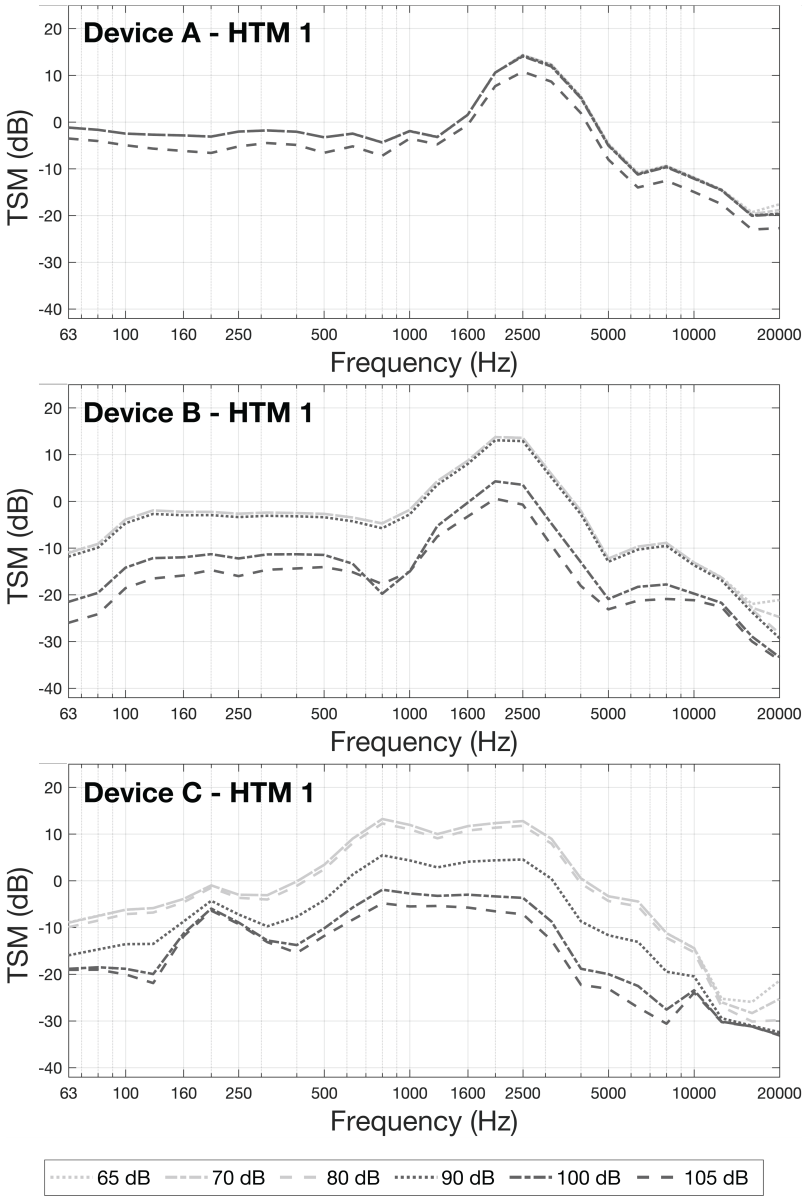


FIGURE 5.10: *Total spectral modification of diffuse pink noise at presentation levels of 65 to 105 dB SPL for Hear-through 1 modes for all devices.*

presentation levels, with successively higher levels of attenuation from 65-80 dB SPL, before remaining consistent from 80-100 dB SPL. For Device A's *HTM 1* response at 105 dB SPL the shape of the TSM remained broadly consistent with the 80-100 dB SPL results, but provided between 2-3 dB more attenuation across all frequency bands. This is potentially indicative of a form of limiting functionality where noise signals above 100 dB SPL are subject to greater levels of attenuation in an attempt to protect the listener from dangerous environmental noise levels. However, it is also possible that this level dependent behaviour could be a symptom of Device A's *HTM 1* reaching their overload limit at 105 dB SPL, as spectrogram analysis of the raw pink noise recordings revealed temporal discontinuities

occurring at intervals of every 6-8 s in both left- and right-ear device responses. This is shown in Fig. 5.11.

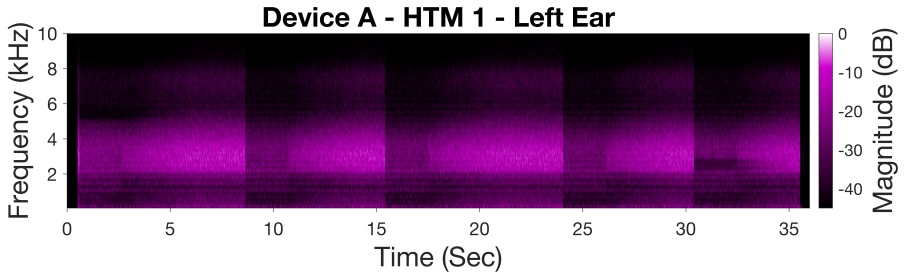


FIGURE 5.11: *Spectrogram showing temporal discontinuities in the spectral modification of 105 dB SPL diffuse pink noise for the Device A's HTM 1. Discontinuities are visible approximately every 6-8 s.*

Further spectrogram analysis of the raw recordings of *HTM 1* responses for all devices at all presentation levels tested found no evidence of similar irregularities that would be potentially indicative of overload performance. These results show that even when subjected to potentially dangerous noise levels of 105 dB SPL Device A's *HTM 1* still applies as much as 10 dB of boosting relative to the open ear, such that a 105 dB SPL noise signal is translated to 115 dB at 2500 Hz.

The TSM performance of Device B's *HTM 1* was also found to be level dependent, with a wider range of variation between the results for different presentation levels compared to Device A. The TSM was perfectly consistent for presentation levels of 65-80 dB SPL between 63-12500 Hz, with level dependent behaviour observed in the 16 and 20 kHz frequency bands broadly consistent with the PSM results discussed in section 5.2.3. At 90 dB SPL the shape of the TSM was identical to the 65-80 dB results from 63-12500 Hz, above this exhibiting similar level dependent behaviour with successive increases in attenuation for 16 and 20 kHz bands, while there was an increase in attenuation of approximately 1 dB across all frequencies compared to the 65-80 dB SPL results.

The shape of the TSM for the 100 dB SPL presentation level was broadly consistent with the results for 80 and 90 dB, however, there was approximately a 10 dB increase in attenuation from 63-8000 Hz compared to the 90 dB SPL TSM, with the difference in attenuation decreasing slightly for frequencies above this region. The TSM for the 105 dB SPL presentation level also exhibited similar level dependent performance with approximately a 3-5 dB further increase in attenuation for most frequency bands compared to the 100 dB SPL results. However, there were two notable deviations from this observed. Firstly, the difference in attenuation between 100 and 105 dB SPL narrowed to approximately 1-2 dB from 10-20 kHz, and then the TSM levels for 100 and 105 dB SPL were also

found to have crossed over around the 630-1000 Hz region, such that the 100 dB SPL measurement provided approximately 3 dB more attenuation than the 105 dB SPL at 800 Hz. Similar behaviour was also observed at 800 Hz in the results of Device B's *HTM 3* and *HTM 4* TSM and ASM performance for 80 dB SPL, this is shown in Figs. 5.5 and 5.6 and discussed in section 5.2.2. These results show that even when presented with potentially dangerous noise levels of 100 dB SPL, Device B's *HTM 1* still applies as much as 4 dB of boosting relative to the open ear, meaning a 100 dB SPL noise signal is received at the ear as 104 dB at 2 kHz. It is only then at a presentation level of 105 dB SPL that the noise is attenuated such that the peak of TSM at 2 kHz is equal to the level that would be received by the unoccluded open ear.

The TSM performance of the *HTM 1* in Device C was also found to be level dependent and exhibited the widest range of difference in TSM across the presentation levels that were tested. The results for the 65-80 dB SPL measurements exhibited consistent shapes of TSM from 63-12500 Hz and above this featured the same level dependent behaviour as observed in the higher frequencies of the PSM and *HTM 1* TSM results for the other devices. The TSM for the 80 dB SPL also exhibited approximately 1 dB more attenuation from 63-12500 Hz compared to the 65 and 70 dB results. At presentation levels of 90-105 dB SPL the TSM in Device C's *HTM 1* exhibited similar level dependent behaviour as observed for 100 and 105 dB SPL results in Device B's *HTM 1*, where the amount of attenuation for noise at presentation levels above 80 dB SPL was broadly equal to the relative increase in presentation level. Despite this the 90 dB SPL was still boosted by as much as 6 dB at 800 Hz and 5 dB at 2500 Hz, and it was only then at presentation levels of 100 and 105 dB SPL that there were no instances of boosting observed in any of the frequency bands for Device C's *HTM 1*.

5.2.4 Time domain performance

Previous studies that have investigated the noise cancellation [137] and hear-through [118] performance of headphone devices have also sought to measure how this performance varies over time. This is carried out to assess both the consistency of performance over time and the nature of the onset of spectral modification following the initial attack of the external noise stimuli. In the present study both of these elements of time domain performance were assessed via spectrogram analysis of the raw diffuse pink noise recordings from the measurement session described in section 5.2.1. In this session

the given device mode under test was set using the devices' control app before the commencement of each measurement recording. The results therefore describe how the already selected and active device modes responded to the consequent onset of the diffuse pink noise signal, rather than the responses of the device modes to being activated during an ongoing signal.

Fig. 5.12 shows the onset of spectral modification for the hear-through modes in all devices. These specific hear-through modes and presentation level measurements give the best visual representations of the onset performance of TSM in hear-through modes for each device. These onset responses were found to be consistent across all hear-through modes and presentation levels for each device with the exception of Device A's hear-through mode responses at 105 dB SPL, where suspected limiting/overload behaviour was observed. This is shown in Fig. 5.11 and discussed further in section 5.2.3.

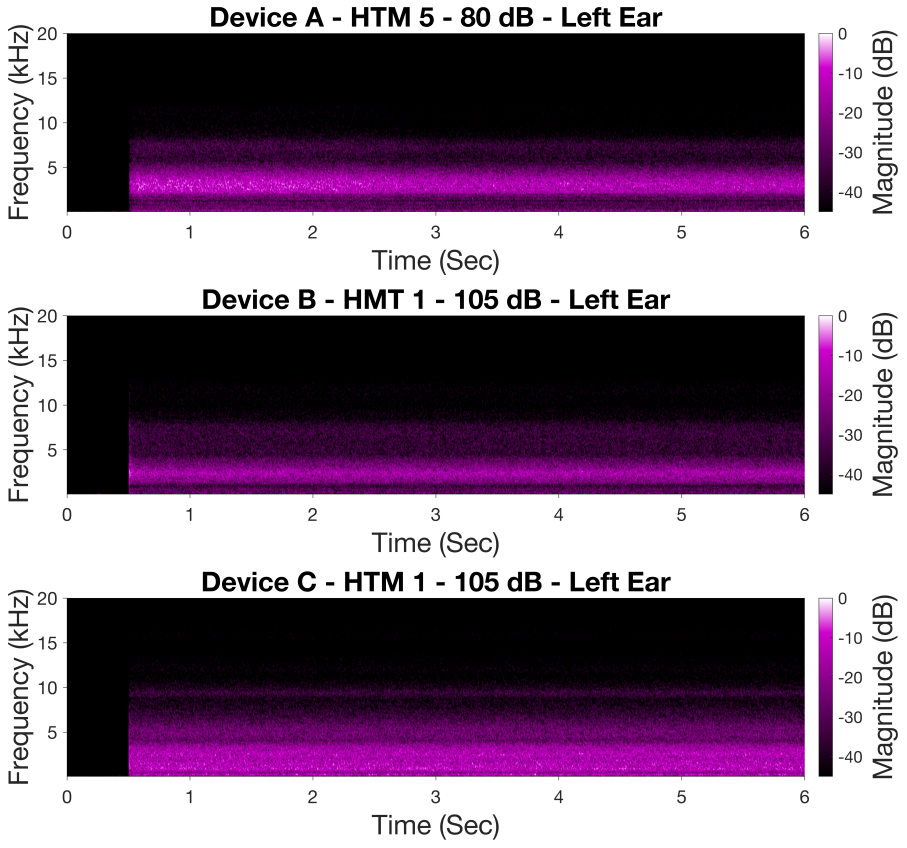


FIGURE 5.12: *Spectrograms showing the onset of spectral modification of diffuse pink noise for Device A's HTM 5, Device B's HTM 1, and Device C's HTM 1. For Device A the onset is visible between 2-3 s and most prominent in the frequencies either side of the 2-4 kHz region, while the onset is instantaneous for devices B and C.*

Inspection of the spectrogram analysis results reveals differences between the onset performance of the hear-through modes in each device. The onset of spectral modification in Device A's hear-through modes exhibited an initial delay of approximately 2 s before the spectral modification was applied between 2 and 3 s after the initial onset of the noise signal. This behaviour was found to be clearly audible when auditioning the measurement recordings during analysis. This was contrasted by the onset performance observed in the hear-through modes of devices B and C where the onset of spectral modification appeared to be almost instantaneous, occurring within a matter of milliseconds. This onset period was found to be audibly imperceptible during the informal auditioning of the measurement recordings and was only detected through visual analysis of the spectrograms. The onset of spectral modification for the noise cancellation modes in all devices also appeared to be instantaneous as visual and audible inspection of the *NCM* response raw recordings failed to reveal any evidence indicating the onset duration.

Further spectrogram analysis of the entire 35 s diffuse pink noise raw recordings found that TSM performance remained consistent and stable over time for all device modes and presentation levels with the exception of Device A's hear-through modes and Device B's *NCM* at 105 dB SPL where discontinuities from suspected limiting/overload were observed. This can be seen in Figs. 5.11 & 5.9. While only the results for the left ear devices are presented in this section, all results discussed were found to be consistent in both left and right ear device recordings.

5.3 Discussion

Analysis of the results shows that all three devices are capable of providing a wide range of both attenuation and boosting of broadband noise signals. The performance of the spectral modification in the devices was found to be highly dependent on the quality of the earbuds' physical fit in the ear canal. Excess acoustic leakage was found to be especially detrimental to the performance of the noise cancellation modes for all devices. The TSM of the noise cancellation modes in all devices was found to provide considerable levels of attenuation across all measured frequency bands, and while the general shape of attenuation varied between devices there were some commonalities observed. For instance, with the least attenuation prominent peaks centred around the lower mid-range region

(800-1600 Hz) were observed, with the greatest attenuation being observed in the regions above 8000 and below 250 Hz.

The general shape of the TSM provided by the hear-through modes remained broadly consistent between the different levels of hear-through for all devices. However, there was a greater degree of variation observed in the TSM shape of the hear-through functionality between the different devices. This is reflected in differences between how the hear-through modes were described in the devices' accompanying app-based control interfaces.

Device A's hear-through mode levels were described as providing a *reduction* of ambient noise. This is perhaps somewhat misleading as, while there was considerable attenuation of noise for frequency bands above 4 kHz, below 1600 Hz there was no greater than 3 dB of attenuation observed for *HTM 1*, increasing to a maximum of 10 dB attenuation in this region for *HTM 6*. In fact 1600-4000 Hz was actually subject to a significant amount of boosting by all of Device A's hear-through modes. It is therefore the case that the reduction of ambient noise perceived by users of Device A's hear-through modes is actually a result of the relative difference between the boosting in the frequencies most important to speech intelligibility (1600-4000 Hz), and the minimal levels of attenuation in frequencies below this region. In contrast, the variable hear-through mode levels in Device C were described as providing the listener with an *awareness* of ambient sound, suggesting that the hear-through functionality in these modes should feature less attenuation and more boosting of frequency bands associated with environmental noise compared to Device A. The results from the present study confirmed that this was the case with much wider regions of boosting observed in four of the six hear-through modes tested in Device C.

As discussed in section 5.1.2, Device B featured separate *Aware* and *Quiet* modes in addition to a variable continuum of variable noise cancellation modes, however, further research established that the performance of the *Aware* and *Quiet* modes was identical to the variable noise cancellation modes at either end of this continuum. Results found that the TSM in Device B's *HTM 1* was very similar to the equivalent measurement in Device A, with the exception of Device B featuring wider breadths of boosting in *HTM 1* and *HTM 2* (1-4 kHz), thus providing a similar hear-through experience to Device A where the singular region of boosting was focused solely around the frequencies that are important for speech intelligibility [143], with frequencies outside of this region were attenuated.

However, the main difference between the hear-through functionality in Device B and the other devices was the range of attenuation observed in the TSM performance across the different levels of hear-through. Device B featured by far the highest levels of attenuation for frequency bands either side of the mid-range in any of the hear-through modes tested across all devices. As with the performance of the other devices, this was also reflected in the use of the term *Noise cancellation* to describe the continuum of hear-through modes in Device B.

The measurement of intermediate levels of hear-through modes in the devices also revealed the extent of differences in the ranges of spectral modification available to the user. Device B provided the largest range of spectral modification, with a span of approximately 15-20 dB from *HTM 1* to *HTM 5* for most of the frequency spectrum. Device C featured a range of approximately 10-15 dB from *HTM 1* to *HTM 6* across the majority of the spectrum. However, Device A provided a much smaller range of approximately 5 dB between *HTM 1* and *HTM 6*. This ability to select a specific level of hear-through allows the user a degree of control over the ease with which they are able to perceive their acoustic surroundings. This therefore allows the user to adjust the volume of the media they are consuming while maintaining the same level of awareness of their surroundings, relative to the media. These results illustrate differences in the extent to which users would be able to adjust the desired volume of media on each device, while still maintaining the same relative awareness of their surroundings.

Overall, the results indicate that there is considerable variation in the nature and intended use cases of insertion loss and gain in contemporary true wireless earbuds, and that these extend far beyond the existing rudimentary definitions of noise cancellation and acoustic transparency. Despite this there were still many commonalities observed in the results.

As no measurements of device performance were made using instationary noise signals, it is not possible to say whether the devices utilised dynamic processing of ambient noise. However, in September 2024 Apple announced the release of a new hearing health experience for the Apple Airpods Pro 2. This included a hearing test and clinical grade hearing aid features, in addition to a loud sound reduction feature which uses a multiband high dynamic range algorithm to actively reduce loud intermittent noise [277].

The results of the linearity analysis revealed that the TSM performance in the devices was highly dependent on the presentation level of the noise signal. This was especially apparent for the performance

of the hear-through modes at higher presentation levels where the amount of boosting was curtailed in all devices to some degree to limit the potentially damaging effects of sustained exposure to high noise levels. However, the nature of this behaviour varied greatly between devices with Device A's *HTM 1* still providing 10 dB of boosting to diffuse noise at presentation levels of up to 100 dB SPL. The results also revealed evidence of temporal discontinuities observed in the results of Device A's *HTM 1* and Device B's *NCM* at 105 dB SPL. These were found to be potentially indicative of overload performance in the devices, however, further measurements at higher presentation levels would be required to fully determine this. Otherwise the TSM performance of the noise cancellation modes in all devices was found to be largely independent of presentation level.

The onset performance of TSM was found to be instantaneous for all device modes tested with the exception of the hear-through modes in Device A where the spectral modification only began to take effect between 2-3 s after the initial onset of the noise signal.

There were also two specific instances of potentially irregular behaviour observed in the performance of Device B's hear-through modes. These occurred at 800 Hz where the level of attenuation was seen to cross over between modes and presentation level measurements counter intuitively to performance found in all other results. It is possible that this performance is representative of the usual behaviour expected in Device B, however, it is also possible that these irregularities were a symptom of issues in the quality of the physical fit of the earbuds in the ATF's ear canal.

The results presented in this chapter are subject to the following limitations:

- The evaluation of a limited number of devices restricts the generalisation of the results.
- Only one pair of each device was tested, meaning it is possible that aspects of the devices' measured behaviour could have been representative of faults rather than normal performance.
- The results are only based on how the devices responded to steady-state broadband noise.
- The relatively low number of measurement iterations conducted increases the likelihood that the results may have been influenced by issues with the quality of physical fit.
- The presentation levels used to assess the linearity of performance in the devices were insufficiently high to assess the overload performance of all devices.
- Various limitations relating to the use of the ATF were also discussed in section 5.1.3.

5.4 Conclusion

The evaluation of the performance of the true wireless earbud device modes tested in this study sought to quantify the extent and nature of insertion loss and gain provided by the devices when subjected to diffuse pink noise, and also measured how the performance of the devices changed with respect to different presentation levels. An analysis of the devices' performance over time was also conducted. Several limitations of the study were also identified and the results and conclusions presented are conditional on these limitations. The results represent the first of their kind to measure the spectral modification of diffuse noise for noise cancellation and hear-through modes in commercially available devices that feature both modes of functionality. The noise cancellation modes in each device were all found to be capable of providing significant amounts of attenuation across the entire frequency spectrum measured. The results also revealed that the hear-through modes in the devices tested exhibited significant amounts of boosting relative to the open ear in frequencies that are considered important to the intelligibility of speech. This result indicates that none of the devices appeared to be attempting to achieve complete acoustic transparency in their hear-through mode functionality. Instead, the levels of boosting around frequencies important for speech intelligibility potentially indicate that some devices are being designed to encourage continual *"always-on audio"* usage [37, 79] throughout the day, allowing verbal communication to occur contemporaneously with the ongoing consumption of media. These results also serve to support the assertions of prior work [37] that devices such as the Apple Airpods are now blurring the boundaries between earphone and *"assistive hearing"* devices [37]. These findings will thereby be used to inform the design of the listening experiment study presented in the next chapter, comprising an investigation of the effects of hear-through augmented environmental noise on listener engagement with speech-based media.

Chapter 6

The Effects of Hear-through Augmented Environmental Noise on Listener Engagement with Speech-based Media

The results of the study presented in the previous chapter showed that hear-through processing in all three of the contemporary commercially available true wireless earbud devices measured imparted significant amounts of frequency dependant boosting to external sound picked up and fed through to the ear. The nature of this boosting was such that, when combined with the passive noise reduction provided by the physical occlusion of the devices, the amplitude of external sound perceptible at a listeners' ear would be boosted (relative to open ear listening) by as much as 15 dB across the band of frequencies considered most important to speech intelligibility (1600-4000 Hz). Prior to this, results of the study presented in Chapter 4 identified five factors of environmental context capable of influencing listeners' attentional engagement in podcast listening experiences. Amongst these, the *soundscape & at work* factor suggested that listening in the presence of high or moderate levels of mechanical or human related sound while also engaging in work related activities could both negatively and positively influence attentional engagement.

The study presented in this chapter builds on the findings of both studies presented in Chapters 4 and 5 to investigate how various aspects of listeners' engagement with speech-based media are affected by different levels of environmental noise, and how the boosting of environmental noise from hear-through

processing in earphone devices further mediates this relationship. This study therefore addresses the third (A_3) and fourth (A_4) main aims as set out for this thesis in Chapter 1:

A_3 Explore the relationship between environmental noise and listener engagement with speech-based media.

A_4 Examine how hear-through modes in auditory headset devices mediate the relationship between environmental noise and listener engagement with speech-based media.

As highlighted in Chapter 3, the results of recent studies [189, 190] have suggested that environmental noise can affect listeners' perception of the media itself, and this was further supported by the findings presented in Chapter 4. However, no studies have been conducted to investigate how boosting of environmental noise from hear-through processing might influence aspects of listeners' engagement with media such as Immersion, Narrative Absorption, and Enjoyment.

6.0.1 Research Aims

To investigate how the boosting of noise from hear-through processing in listening devices might affect listener engagement with media, it is first important to understand how listener engagement is affected when media is consumed with the passive occlusion of a traditional headphone listening experience. This study focuses on speech-based mobile media such as podcasts or on-demand radio, leading to the first research question RQ_1: "How does exposure to noise affect listener engagement with speech-based media over earphones?" with RQ_2 as "How does boosting from hear-through processing in hearables mediate the effects of noise on engagement?". To answer both research questions, an experiment is designed to investigate how different aspects of listeners' engagement with a storytelling podcast is affected when listening under different noise conditions.

6.1 Methods and Materials

6.1.1 Experiment Design

This experiment employs a mixed design with a between-subjects group factor of noise condition and a within-subjects repeated measures factor of stimuli. The use of this design, in addition to a series of other methodological design strategies, was selected in accordance with a series of implications [278] [186] outlined for developing experimental paradigms that aim to investigate immersion in audiovisual experiences. Key implications amongst these are; the use of absolute assessment methods, post-experience assessment of stimuli, sufficient time gaps between consecutive stimuli, monitoring of fatigue, and all responses being collected in a single session. These implications are also highly applicable to experiments investigating similar qualitative measures of listener engagement. As such, a study by Herrmann and Johnsrude [146], that investigated the effect of acoustic masking on story absorption, enjoyment, and listening effort was used as a methodological basis for investigating other aspects of listener engagement, in addition to immersion. The University of York School of Physics, Engineering and Technology ethics committee approved this study (approval number Harrison23112023). This is shown in Figures B.1-B.12 of Appendix B.

6.1.2 Story Stimuli

Four $\sim$5 to 6 minute stories have been selected from the storytelling podcast "The Moth" for use as the stimuli in the experiment: *A Shoulder Bag to Cry On* by Laura Zimmermann (5 min), *Things I Knew For Sure* by Qing Zhao (5 minutes 8 seconds), *A Monkey Meets a Seal Meets a Monkey* by Matthew McArthur (4 minutes 58 seconds), and *I Am Batman* by Paul Davis (5 minutes 43 seconds). They are chosen as they have broad appeal, have been validated as being capable of eliciting absorption and enjoyment [146], and are of sufficient duration for the narrative to develop and allow the participant the opportunity to experience immersion [186]. An additional story (*Oh Brother* by Joey Garfield) from "The Moth" podcast was selected for use as the training story in the experiment. All of the stimuli are recordings of a storyteller recounting a story to a live audience. Audience reactions (laughter, applause, etc) are audible in all stories. Permission has been sought for use of all stories. Stimuli were 16 bit

44.1 kHz wav files normalised to -23 LUFS. In the rest of the chapter the stories are anonymized and referred to as stimuli A to D.

6.1.3 Noise Conditions

There are four noise conditions included in the experiment; *no noise (NN)*, *low noise (LN)*, *high noise (HN)*, and *hear-through (HT)*. The same Ambisonic soundscape recording is used as the noise in all conditions with the duration differing between stimuli to account for the different story lengths. Further details on the soundscape recording are provided in section 6.1.5. For each stimuli, participants in the *NN* group are presented with the story over earphones and no additional soundscape noise. In the *LN* group participants were presented with the story over earphones in addition to noise over a speaker rig (see section 6.1.6) at 52 dBA SPL. Participants in the *HN* group were presented with the story over earphones in addition to noise over the speaker rig at 62 dBA SPL. In the *HT* group participants are presented with the story over earphones, 52 dBA SPL noise over the speaker rig, and a binaural recording of the noise over the earphones, that was filtered and gain adjusted to simulate the boosting of environmental noise that is provided by hear-through processing in hearables. The *HT* condition is therefore a simulation of the noise a listener would perceive if they were listening to the 52 dBA SPL *LN* condition with hear-through activated. The amount of boosting in the hear-through condition is such that the total binaural loudness of the noise in the *HT* condition matches that of the *HN* condition. Further details on the *HT* condition are provided in section 6.1.4. The speaker rig noise levels are selected based on typical sound pressure levels measured in urban open spaces [279]. The transfer function of the passive occlusion provided by the earphones used in the experiment when subjected to diffuse broadband pink noise over the speaker rig is shown in Figure 6.2. A summary of the noise conditions is provided in Figure 6.1. The stimuli was calibrated such that it is 4 dB louder in perceived loudness relative to the 62 dBA SPL noise of the *HN* condition and 14 dB louder relative to the 52 dBA SPL noise of the *LN* condition. The stimuli level is selected based on typical and safe earphone listening levels [280] and was the same for all conditions. It is acknowledged that the absolute SPL levels chosen may affect the nature of the findings, for example it is possible that the results could vary significantly depending on both these absolute levels and the relative levels between conditions and stimuli. However, the minimum required sample size for the mixed design methodology meant that it

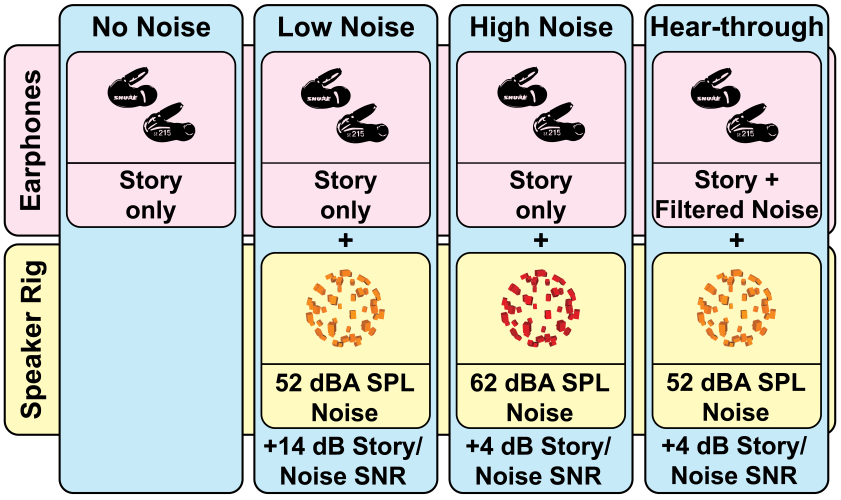


FIGURE 6.1: Graphical summary of the noise conditions; *No Noise (NN)*, *Low Noise (LN)*, *High Noise (HN)*, and *Hear-through (HT)*. The *HN* & *HT* conditions had equal binaural loudness.

was not possible to include more than four noise conditions in the experiment, therefore limiting the number of different SPL levels that it was possible to investigate.

6.1.4 Hear-through Simulation Noise Condition

In the *HT* group participants are presented with the stimuli over earphones, 52 dBA SPL noise over the speaker rig, and a binaural recording of the noise over the earphones, that is filtered and gain adjusted to simulate the boosting of environmental noise that is provided by hear-through processing in hearables as follows.

The design of the hear-through simulation transfer function is based on averaged measurements of the hear-through performance in the Apple Airpods Pro (2nd gen.), Bose QuietComfort Earbuds II, and Sony LinkBuds S, as explored in Chapter 5. The boosting of noise provided by the hear-through simulation resulted in the overall perceived loudness of the *HT* noise condition being equal to that of the *HN* condition, as measured using the method described in section 6.1.6. A target transfer function for the simulation (see Figure 6.2) is derived by averaging a series of diffuse broadband pink noise third-octave RMS measurements of the earbuds' hear-through performance, using the mode that provided the greatest awareness of external noise in each device. The binaural noise recording used in the *HT* condition is produced using the following method: DPA 4560 CORE Binaural Headset Microphones are mounted on the enclosures of the Shure SE215 earphones, fitted in KEMAR's pinnae,

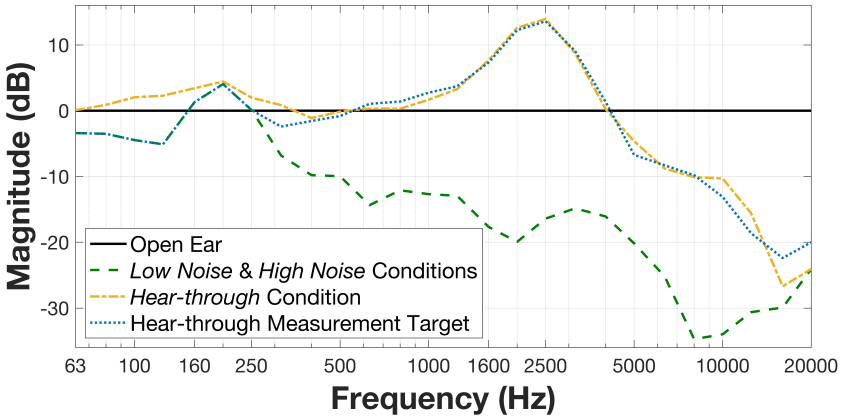


FIGURE 6.2: Graph showing the transfer functions of noise conditions and the target measurement for the *Hear-through (HT)* condition produced using diffuse broadband pink noise.

and used to capture binaural recordings of the soundscape noise and diffuse broadband pink noise played back over the speaker rig. This is shown in Figure 6.4. This is intended to simulate the spatial effects of the built-in microphones used to capture external sound in hear-through equipped devices. A first-order Butterworth bandpass filter bank was designed in Matlab and used to filter and gain adjust the binaural pink noise recording, such that when the processed binaural pink noise recording was played back over the earphones alongside 52 dBA of diffuse broadband pink noise over the speaker rig, the resulting summed signal recorded by KEMAR possessed a similar transfer function to the Hear-through measurement target. This processing chain is illustrated in the block diagram shown in Figure 6.3. The transfer function of the final *HT* condition is shown in Figure 6.2. The same filter bank is then used to process the binaural soundscape noise recording that is played back over the earphones, alongside the synchronised 52 dBA noise over the speaker rig for the *HT* condition. The filter bank code is provided in Listing B.1 of of Appendix B.

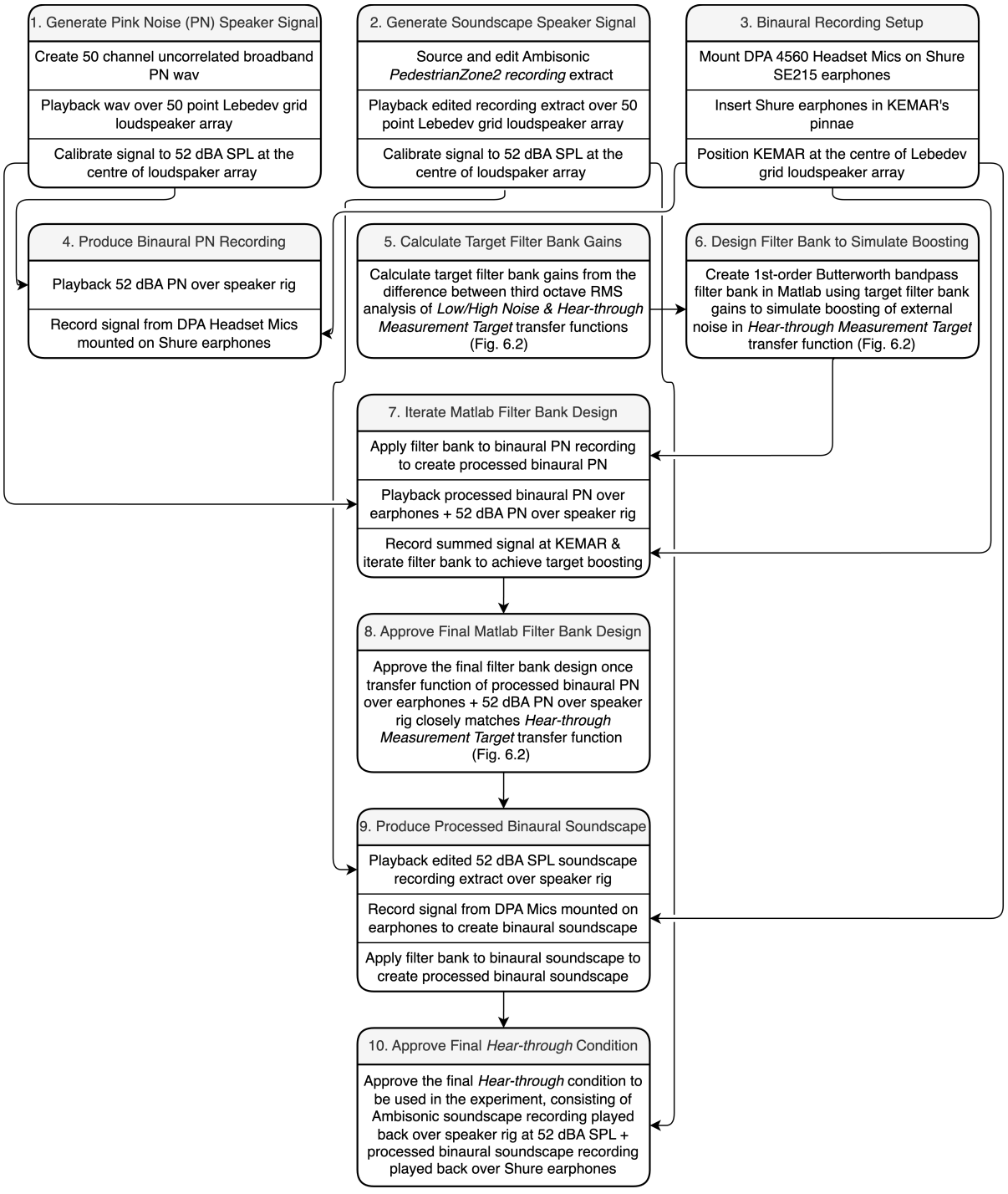


FIGURE 6.3: Block diagram illustrating the processing chain that was used to create the *Hear-through (HT)* condition in the experiment.

FIGURE 6.4: DPA binaural microphone mounted on the Shure SE215 earphones, used for calibration and production of the hear-through simulation in the *hear-through* noise condition.

6.1.5 Noise Condition Soundscape Material

Extracts of the *PedestrianZone2* recording from the Eigenscape database of spatial acoustic scene recordings [281] are selected for use as the environmental noise soundscape material for all the noise conditions in the experiment. The 10 minute fourth-order Ambisonic *PedestrianZone2* recording was captured on Church Street, Liverpool using an em32 EigenMike microphone. The recording represents a typical pedestrianised street scene, mainly consisting of speech and footsteps, with other sounds including intermittent street music, laughter, crying children, and dog barks. All noise condition soundscape extracts began at the same timestamp (47 seconds) in the *PedestrianZone2* recording, however as the story stimuli varied in length from 4 minutes 58 seconds to 5 minutes 43 seconds (see section 6.1.2), the soundscape extracts also varied in length depending on the stimuli. A crossfade edit was applied to remove an atypical loud impulse noise at 6 minutes 22 seconds in the *PedestrianZone2* recording that may otherwise have biased the listening experience for the longest stimuli only.

6.1.6 Apparatus and Level Calibration

A pair of Shure SE215 In-Ear Sound Isolating Earphones are used for playback of stimuli and the hear-through simulation soundscape material as they provide similar passive noise occlusion to contemporary true wireless earbud devices. Earphone audio is played back from Reaper via a Ferrofish A32 AD/DA converter and a Samson S-phone headphone amp. The earphones are fitted with the soft flex sleeves. Participants first tried the fit of the earphones with the medium size sleeves and were asked if the fit felt comfortable and secure. If they answered no the fitting process was repeated with an alternative sleeve size. This resulted in 82.5% of the participants using the medium sleeves and 17.5% using the small. The *PedestrianZone2* Ambisonic soundscape extracts are decoded in Reaper and played back over a 50 point Lebedev grid loudspeaker array via two Ferrofish A32 AD/DA converters. The array comprises of 40 Genelec 8030A and 10 Genelec 8040A active loudspeakers. Participants are seated such that their heads are positioned at the centre of the array, as shown in Figure 6.5. Participants use a keyboard and mouse to interact with a Matlab app that guides them through the experiment, triggers playback of audio with Reaper via OSC, and recorded responses to questions. A pen and five A4 sheets are used to record Immersion ratings.

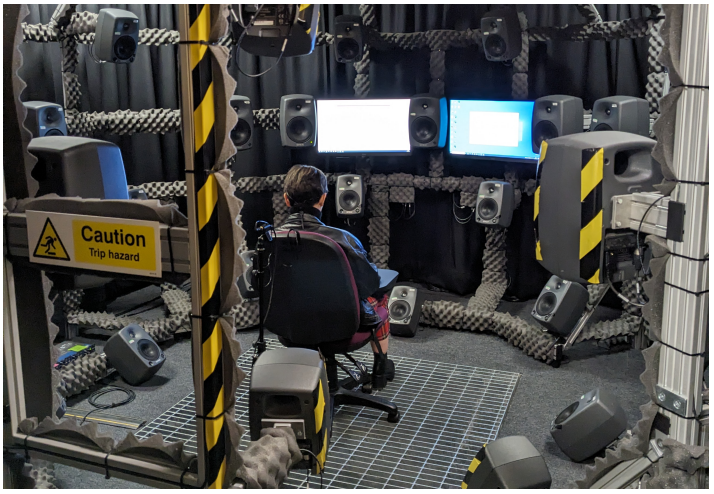


FIGURE 6.5: A participant seated in the listening position.

The speaker rig noise levels are calibrated using a NTi XL2 Analyzer (LAeq) and M4260 measurement microphone placed at the centre point of the speaker rig. In experiments where audio is presented using multiple reproduction systems, such as earphones and loudspeaker array, it is recommended

that a binaural loudness matching method be used for calibration of relative (inter-system) levels [282]. A modified (bypassed head-shadowing filter) ITU-R BS.1770-3 binaural loudness matching method [283] is used for calibration of the signal-to-noise ratio (SNR) of stimuli relative to noise. The binaural recordings are captured with prepolarized ear simulators (RA0045-S1) fitted in a GRAS 45BC KEMAR acoustic test fixture. KEMAR is equipped with large pinnae and positioned at the centre of the speaker rig.

6.1.7 Measures

To assess participants' level of overall Immersion in each story, a method for assessing Immersion in audiovisual experiences was selected for use [278, 284]. This involved participants' rating their Immersion by making a single mark on a 15-cm graphic line scale (shown in Figure 6.6), with reference points of "Not very immersive" and "Very immersive" at either end, and no markers between. Participants are asked to use a text field to describe how their experience influenced their rating of Immersion. To assess participants' Narrative Absorption and Enjoyment for each story, the Narrative Absorption Scale (NAS) and five Enjoyment items from [146] was used. To assess Listening Effort for each story, the three Effort items from [146] are also administered. Participants rated each item statement by selecting an option from a 7-point Likert scale, with 1 referring to "Completely disagree" and 7 to "Completely agree". Questionnaire item statements are shown in Table 6.1. An exit questionnaire was administered upon completion of participants' responses to the main stimuli rating questions. This is intended as a post-experiment measure of participants' awareness of background noise for each condition. They were first asked if they had been aware of any other sound(s) aside from the stories themselves while listening? If they answer yes, they are asked to describe these sounds, and then asked if their awareness of any of the sound(s) influenced their rating of overall Immersion for any of the stories. If they answer yes, they are asked to describe how their Immersion ratings were influenced by their awareness of the sound(s).

TABLE 6.1: Items of the Narrative Absorption Scale (NAS; Capturing Four Dimensions), Enjoyment, and Listening Effort.

Dimension	**Question**
Attention (NAS)	When I finished listening, I was surprised to see that time had gone by so fast.
"	When I was listening, I was focused on what happened in the story.
"	I felt absorbed in the story.
"	The story gripped me in such a way that I could close myself off for things that were happening around me.
"	I was listening in such a concentrated way that I had forgotten the world around me.
Emotional engagement (NAS)	When I listened to the story, I could imagine what it must be like to be in the shoes of the main character(s).
"	I felt sympathy for the main character(s).
"	I felt connected with the main character(s) of the story.
"	I felt how the main character(s) was/were feeling.
"	I felt for what happened in the story.
Mental imagery (NAS)	When I was listening to the story, I could see the situations happening in the story being played out before my eyes.
"	When I was listening to the story, I had an image of the main character(s) in mind.
"	I could imagine what the world in which the story took place looked like.
Transportation (NAS)	When I was listening to the story, it sometimes seemed as if I were in the story world too.
"	When listening to the story, there were moments in which I felt that the story world overlapped with my own world.
"	The world of the story sometimes felt closer to me than the world around me.
"	When I was finished with listening to the story, it felt like I had taken a trip to the world of the story.
"	Because all of my attention went into the story, I sometimes felt as if I could not exist separate from the story.
Enjoyment	I thought it was an exciting story.
"	I thought it was an enthralling story.
"	I listened to the story with great interest.
"	I thought the story was beautiful.
"	I thought the story was presented well.
Listening Effort	I had to invest effort to understand what was said.
"	It was difficult to understand what was said.
"	Understanding the speaker was hard.

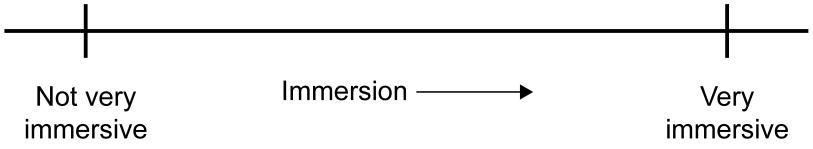


FIGURE 6.6: *The 15-cm printed graphical line scale that participants used to provide their immersion ratings. Participants were provided with a separate scale printed on an A4 sheet of paper for each immersion rating. The A4 sheets were removed from the participant's view before the presentation of the next stimulus.*

6.1.8 Participants

A total of 80 participants[1] took part in the experiment. Participants were each assigned to one of four noise condition groups randomly such that there was an equal balance of genders in each group, as research [172] has suggested differences in engagement between men and women. The mean age of participants was 24.8 years (SD = 6.87). The majority of participants were male (58.75%), 40% were female, and 1.25% non-binary/third gender. Participants were either first language English speakers (72.5%) or having a good level of English language proficiency, all participants rated their English language proficiency as 4 or higher on a 7-point scale. All participants received a £20 voucher as remuneration for their participation in the study. The time taken to complete the experiment ranged from 70 to 95 mins. Experiment sessions took place in the listening room at the AudioLab, University of York. Participants were recruited via University department mail-outs, social media, word of mouth, and posters. All participants were aged 18 or over and did not have any diagnosed hearing loss. Experiments were conducted between the 4th of Jan and the 5th of Feb 2024.

6.1.9 Experiment Procedure

Participants were seated on a chair positioned at the centre of the speaker rig and asked to maintain a forward gaze while listening, this is shown in Figure 6.5. Subjects were provided with an information sheet (see Figures B.13-B.16 in Appendix B) that gave an overview of the study background and research aims, before completing a screening questionnaire, and providing their consent to participate (see Figures B.17-B.18 in Appendix B). Participants were not provided with any information regarding what specific background noise they would experience in experiment sessions, nor were they informed about the between-subjects nature of the study design. Participants then had their earphones fitted before beginning using the Matlab app to provide responses to a series of demographic questions. They next undertook a training section where they listened to the "Oh Brother" story under the same noise conditions as determined by their group, before providing an Immersion rating and responding to a single block of the Likert scale items, selected randomly for each participant. This was to familiarise

[1]As the study is exploratory in nature and no relevant existing literature exists; the intention was to detect small to medium differences (d = 0.2-0.5) [285] among the listener engagement measures between noise condition groups. Using an a priori power analysis in G*Power it was determined that 20 participants in each group could detect an effect size of d = 0.205 for an alpha level of 5% and beta level of 20%, assuming the variation remained constant between the experiments.

participants with listening in this setting and responding to the question response formats they would encounter in the actual experiment. Participants were informed that the training story may not elicit immersion and should not be considered a reference for the main stimuli. Completed Immersion response sheets were removed before presentation of the next stimulus. Participants then undertook the first of four 4 minute interactive distractor tasks (see Figure B.19 of Appendix B) intended to bring participants back to a baseline level of Immersion/Narrative Absorption. The order of tasks in the experiment was randomised. Participants then moved onto the main part of the experiment. The experiment was designed such that participants only experienced each stimulus once and the order of stimuli presentation was randomised. A single sentence narrative synopsis of the story was provided before listening to each stimulus. After listening to a stimulus participants gave their rating of overall Immersion and were asked to describe how their experience influenced their rating. They were provided with the following definition [278] of immersion:

> Immersion, also known as deep mental involvement, can be described as being mentally lost (absorbed) in the experience. Immersion is encountered when the experience is involving and absorbs you mentally by capturing your attention. For example, immersion may be experienced when reading a book, playing video games, watching a movie, etc.

Subjects then responded to six blocks of Likert scale questionnaire items, organised in groupings of the four NAS dimensions, the Enjoyment items, and Listening Effort items. The order in which the blocks appeared was randomised for each stimulus. The participants were then asked if they had heard the story prior to the experiment. Participants then undertook the next distractor task before listening to the next stimuli. After responding to the final set of stimuli questions the participants completed the exit questionnaire, as described in section 6.1.7.

6.2 Data Analysis and Results

All participants reported not having heard any of the stories prior to the experiment. Immersion ratings were translated to scores by measuring from the left limit of the 15-cm scale to within a tolerance of $\pm$1-mm. Narrative Absorption, Enjoyment, and Listening Effort scores were computed by calculating a single mean score of all item responses for each given scale. Statistical analysis was conducted

using a two-way mixed ANOVA model, with analysis being performed in R and SPSS. The following procedures were followed for data analysis. Outliers were detected via examination of studentized residuals for values greater than ± 3. Data normality was assessed via inspection of Normal Q-Q plots and skewness and kurtosis Z scores of studentized residuals. In instances where skewness was found to exceed the recommended threshold of ± 2 [235], it was decided to continue with the analysis using the original data as the cells were all similarly skewed and ANOVAs are considered robust to deviations from normality [286]. Homogeneity of variances and covariances was assessed by Levene's test Box's M test, respectively. Mauchly's test was used to check that the assumption of sphericity, with the Greenhouse-Geisser correction used for interaction and main effects significance results in instances where it was violated. The Bonferroni correction was applied for all pairwise multiple comparisons. Data for pairwise comparison results are mean $\pm$ standard deviation.

6.2.1 Immersion

There were no outliers and the data was approximately normally distributed, with skewness of 2.80 and excess kurtosis of 5.43. There was homogeneity of variances (p = .679) and covariances (p = .907). The assumption of sphericity was met for the two-way interaction, $\chi^2(5) = 10.67$, p = .058. There was no statistically significant interaction between stimuli and noise on Immersion, F(9, 228) = .518 , p = .861, partial $\eta 2$ = .020. The main effect of stimuli showed a statistically significant difference in mean Immersion for different stimuli, F(3, 228) = 2.760, p = .043, partial $\eta 2$ = .035. Pairwise comparisons did not reveal any statistically significant differences between the individual stimuli. The main effect of noise showed that there was no statistically significant difference in mean Immersion between noise conditions F(3, 76) = 2.37, p = .077, partial $\eta 2$ = .085. Estimated marginal means of Immersion for the four stimuli at every noise condition are shown in Figure 6.7.

6.2.2 Narrative Absorption

There were no outliers and the data was approximately normally distributed, with skewness of 3.11 and excess kurtosis of 6.03. There was homogeneity of variances (p = .573) and covariances (p = .671). The assumption of sphericity was violated for the two-way interaction, $\chi^2(5) = 16.16$, p = .006.

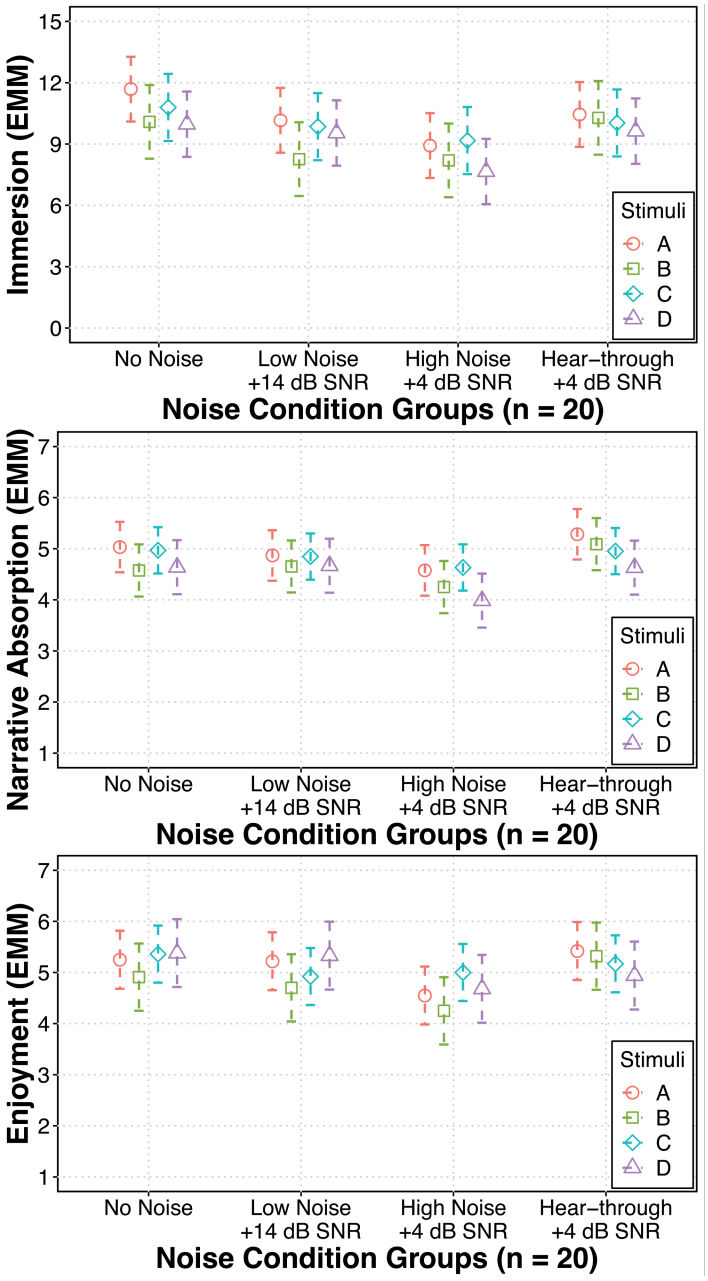


FIGURE 6.7: Estimated marginal means of Immersion, Narrative Absorption, and Enjoyment scores for all stimuli at all noise conditions. Error bars represent 95% confidence intervals.

There was no statistically significant interaction between stimuli and noise on Narrative Absorption, F(7.834, 198.465) = .508 , p = .846, partial $\eta 2$ = .020. The main effect of stimuli showed a statistically significant difference in mean Narrative Absorption for different stimuli, F(2.611, 198.465) = 4.256, p = .009, partial $\eta 2$ = .053. Pairwise comparisons revealed that the mean Narrative Absorption of stimuli A was significantly different from stimuli D. Data are mean $\pm$ standard deviation, unless otherwise stated. There was a decrease in Narrative Absorption from 4.94 $\pm$ 1.12 for stimuli A to 4.48 $\pm$ 1.20 for

stimuli D, a statistically significant decrease of 0.46 (95% CI, 0.13 to 0.907), p = .40. The main effect of noise showed that there was no statistically significant difference in mean Narrative Absorption between noise conditions F(3, 76) = 2.187, p = .0.96, partial $\eta 2$ = .079. Estimated marginal means of Narrative Absorption for the four stimuli at every noise condition are shown in Figure 6.7.

6.2.3 Enjoyment

There was one outlier (-3.24) and the data was approximately normally distributed, with skewness of 2.76 and excess kurtosis of 5.34. There was homogeneity of variances (p = .752) and covariances (p = .686). The assumption of sphericity was met for the two-way interaction, $\chi^2(5)$ = 11.02, p = .051. There was no statistically significant interaction between stimuli and noise on Enjoyment, F(9, 228) = .759 , p = .654, partial $\eta 2$ = .029. The main effect of stimuli showed no statistically significant difference in mean Enjoyment for different stimuli, F(3, 228) = 1.234, p = .298, partial $\eta 2$ = .016. The main effect of noise showed that there was no statistically significant difference in mean Enjoyment between noise conditions F(3, 76) = 2.14, p = .103, partial $\eta 2$ = .078. Estimated marginal means of Enjoyment for the four stimuli at every noise condition are shown in Figure 6.7.

6.2.4 Listening Effort

There were two outliers (3.44, 3.14) and the data was not normally distributed, with skewness of 4.95 and excess kurtosis of 9.60. There was not homogeneity of variances (p = .017) or covariances (p <.001). As the assumptions of normally distributed data and homogeneity of variances were violated, meaning a parametric method of analysis could not be relied on for this data, it was decided to run a robust between-within ANOVA analysis using the *bwtrim* function from the *WRS2* R package [287] with bootstrapped significance tests (1,000 samples) for interaction and main effects with 20% trimmed means. There was no statistically significant interaction between stimuli and noise on Listening Effort (p = .028), the main effect of stimuli showed a statistically significant difference in mean Effort between stimuli (p = .0015), and the main effect of noise showed a statistically significant difference in mean Effort between noise conditions (p = .0016). Pairwise comparisons were conducted for the main effect of stimuli on Effort using the *sppbb* function with an MOM estimator (1,000 samples),

the comparisons did not reveal any statistically significant differences between the individual stimuli (p = .257). Pairwise comparisons were conducted for the main effect of noise on Listening Effort using the *sppba* function with a MOM estimator (1,000 samples), the comparisons did not reveal any statistically significant differences between individual noise conditions (p = .02). The *ezBoot* and *ezPlot2* functions from the *ez* R package were used for visualization of the Listening Effort data, with scores sampled randomly 1,000 times per cell. Cell means of Listening Effort for the four stimuli at every noise condition are shown in Figure 6.8.

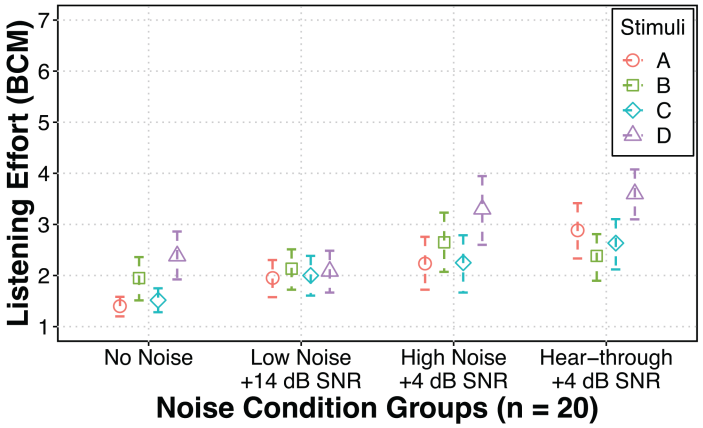


FIGURE 6.8: Bootstrapped cell means (BCM) of Listening Effort scores for all stimuli at all noise conditions. Error bars represent 95% confidence intervals.

6.2.5 Exit Questionnaire: Awareness of Noise

The results of the exit questionnaire, which asked participants about their awareness of other sound(s) aside from the stimuli, are shown in Figure 6.9. They showed that the number of participants who were aware of other sound(s), and those who felt that their awareness of those sounds influenced their Immersion ratings, increased with the amount of noise in each group. In both the *HN* and *HT* groups 15% of participants reported not being aware of any other sound(s), such was their engagement in the stories. While 20% of participants in the *NN* group reported being aware of other sounds, citing an awareness of; "Sounds of my own mouth like swallowing", "My hair moving", and "Slight white noise from earbuds". One participant reported an awareness of: "cars? or the roar of an ocean?". This indicated that at least one participant was aware of faint traffic noise from outside the AudioLab building. This participant reported that their awareness of these sounds negatively influenced their Immersion ratings.

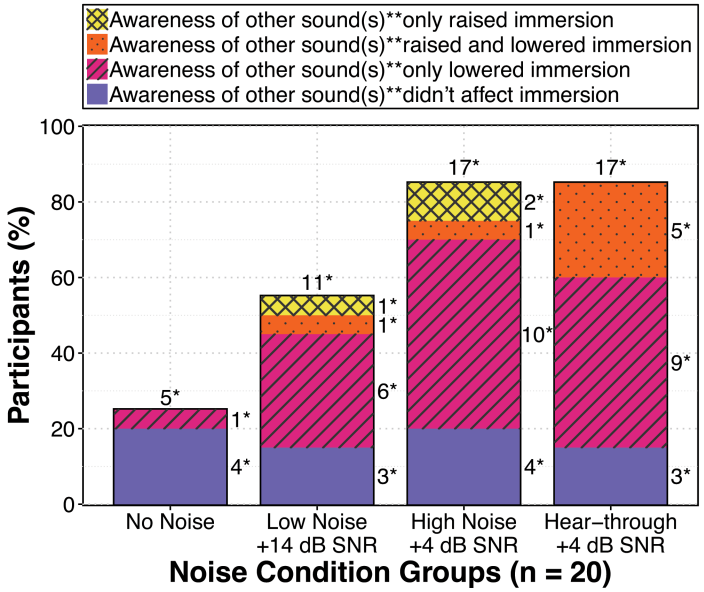


FIGURE 6.9: Participants who were aware of other sounds and how their awareness influenced their Immersion. *Number of participants who reported an awareness of other sound(s) **Other sounds were any sounds aside from the storytellers or audience reactions.

6.3 Discussion

The main aim of this study was to answer two research questions; RQ_1 "How does exposure to noise affect listener engagement with speech-based media over earphones?" and RQ_2 "How does boosting from hear-through processing in hearables mediate the effects of noise on engagement?". The main effect of noise on Immersion, Narrative Absorption, and Enjoyment was not statistically significant, suggesting that there were no substantial differences in mean scores for these measures between noise conditions. However, inspection of the plots in Figure 6.7 suggested a trend of Immersion, Narrative Absorption, and Enjoyment each falling successively as the level of noise exposure increased, from *NN*, to *LN*, and then *HN* groups. The results from the exit questionnaire found that the number of participants who were aware of noise rose from 55% for the *LN* group to 85% for *HN* group. This was reflected in the statistically significant Listening Effort results shown in Figure 6.8, with clear increases in Effort from *NN* to *LN* and then *HN* groups. While most participants who reported awareness in both groups stated that their awareness of other sound(s) lowered Immersion, a proportion reported that their awareness raised Immersion, with others reporting their awareness didn't effect Immersion. The proportion of those who reported that their awareness lowered their Immersion rating increased

slightly from 63.6% to 64.7% between *LN* and *HN* groups, whereas the proportion of those who reported their awareness raised Immersion fell from 18.2% to 17.6%. These results suggest that while exposure to noise results in most listeners experiencing lower engagement with speech-based media over earbuds, a small proportion of listeners experience higher engagement. Further inspection of the plots in Figure 6.7 revealed an interesting relationship between the *LN* and *HT* groups for Immersion, Narrative Absorption, and Enjoyment scores. The *HT* condition had the same 52 dBA SPL of noise over the speaker rig as the *LN* condition, but the additional filtered noise over the earphones in the *HT* condition meant it had a -10 dB SNR relative to the *LN* condition and the same binaural loudness as the 62 dBA SPL *HN* condition. Despite this, the Immersion, Narrative Absorption, and Enjoyment scores for the *HT* group were similar to those of the *LN* group and were in some cases higher. This was particularly notable given that the *HN* group engagement results were notably lower than the *HT* group, despite having equal binaural loudness. The results from the exit questionnaire found that the number of participants who were aware of noise rose from 55% for the *LN* group to 85% for *HT*. This was reflected in the statistically significant Listening Effort results shown in Figure 6.8, with a clear increase in Effort between the *LN* and *HT* groups. The proportion of those who reported that their awareness lowered their Immersion rating increased from 63.6% to 82.4% between *LN* and *HN* groups, while the proportion of those who reported their awareness raised Immersion also increased from 18.2% to 29.4%, however, all participants in the *HT* group reported their awareness only raised Immersion. These results suggest that boosting from hear-through in hearables can provide listeners with an increased awareness of their surroundings, without compromising their engagement with speech-based media. The differences observed between the *HN* and *HT* results are not discussed as this was not considered a valid comparison. The statistically significant results for the main effect of stimuli on Immersion and Narrative Absorption were important as they showed that participants were able to discriminate between the stimuli and illustrated that the experimental method used was effective in capturing differences in Immersion and Narrative Absorption in this context. The results of this study are subject to limitations outlined by Agrawal et al. [284], such as the method not revealing temporal variation in engagement, the possibility that participants may assess immersion based on a limited number of select moments, and post experience assessment being liable to the recency effect and inaccurate recall.

6.4 Conclusion

In this chapter a listening experiment was conducted with 80 participants to investigate how noise affects listener engagement with speech-based media over earphones and how boosting from hear-through processing in hearables mediates the effects of noise on engagement. Measures included Immersion, Narrative Absorption, and Enjoyment, with Listening Effort and awareness of noise also assessed. The results suggest that boosting from hear-through processing provides listeners with a greater awareness of their surroundings without diminishing engagement. Further analysis also suggests that while most listeners' experience lower engagement as a result of noise exposure, some experience higher engagement. There were statistically significant results for the main effect of noise on Effort, with higher noise increasing Listening Effort, and for the main effect of stimuli on Immersion and Narrative Absorption, showing that the methods used to assess these measures were effective in capturing differences in this context.

The work presented in this chapter builds upon findings from Chapter 4, that identified 5 *factors of environmental context*, including the *Soundscape and at Work* factor, that were found to be capable of influencing listeners' attentional engagement in podcast listening experiences. It also expands on the work presented in Chapter 5, that quantified the nature and extent of the external sound modification performance delivered by the passive occlusion and adjustable hear-through features of contemporary true wireless earbuds. These findings will be further reflected upon in the subsequent final chapter, where they will be discussed in a review of all work presented in this thesis, and considered in the broader context of the overall hypothesis for the thesis stated in Chapter 1.

Chapter 7

Conclusion

This thesis constitutes a portfolio of research focused around the central theme of investigating how a listener's surrounding environmental context and the auditory and non-auditory stimuli contained within it can mediate various aspects of their engagement with speech-based media. This has included a survey on listening behaviours and the factors of environmental context that influence attentional engagement in podcast listening experiences, an examination of how true wireless earbuds control the amount of external sound that listeners are able to perceive, and a listening experiment that investigated how these external sound modification capabilities of true wireless earbuds mediate the relationship between environmental noise and listeners' engagement with speech-based media. This chapter will bring this thesis to a conclusion, encompassing a summary of the work presented, and a restating of the hypothesis outlined in Chapter 1. Contributions to the field are also discussed, followed by a discussion of potential applications of future research, and lastly the thesis is brought to a conclusion with final comments from the author.

7.1 Summary

The thesis began with a review of fundamental theory and related work in Chapters 2 and 3 respectively. Chapter 2 opened with a discussion of the necessary concepts required to understand how sound propagates as a wave, how frequency is related to wavelength, how sound intensity decreases with distance, and how the level of a sound can be quantified in SPL units. The fundamental psychoacoustic

theory that underpins the ability of humans' to hear and perceive sound spatially was then discussed, before the chapter was concluded with a review of the concept of environmental noise and how it should be understood within the context of the thesis. Chapter 3 opened with an overview of auditory headsets typically used in the mobile consumption of audio and the different kinds of features inherent in these devices that are capable of boosting and attenuating external sound. Technological and behavioural developments in the mobile consumption of audio by consumers were then discussed. A literature review concerning the aspects and assessment of listener engagement with speech-based media followed, before a further review of literature that detailed research on listener engagement in mobile music, and speech-based media consumption experiences.

Key findings highlighted in Chapter 3 included the trend of some mobile audio consumers using acoustically transparent or adjustable hear-through headphone features for longer periods throughout the day, thus increasing their ability to perceive external sound relative to the media they are consuming. The review of methods for the assessment of listener engagement with narrative-based media experiences found self report scale instruments were most widely used, however, this also revealed a lack of methods specifically devised for the assessment of engagement with spoken narratives. The review also highlighted the complex relationship that exists between the media, the listener, and their surroundings in mobile audio consumption experiences, with a large proportion of research focused on how the mobile consumption of audio shapes a listeners' interactions with and perception of their environmental context. Further to this, a wide array of research was identified on listener engagement in mobile music consumption experiences. This contrasted greatly with the extremely limited number of studies that have focused exclusively on listener engagement in speech-based media consumption experiences. Further research suggested that findings from studies concerning engagement in mobile music listening could not be considered transferable to research on engagement in mobile speech-based media consumption experiences. Therefore, a significant gap in knowledge was identified for research concerning the potential affects of environmental context and noise on listener engagement in mobile speech-based media consumption experiences.

Chapter 4 then presented the first original work of the thesis, wherein a survey of 264 podcast listeners' consumption habits and behaviours was conducted. The results of an exploratory factor analysis identified 5 factors of environmental context that were considered capable of both positively and negatively influencing listeners' attention engagement in podcast listening experiences. These factors

were labelled as; *Outdoors*, *Indoors & at home*, *Evenings*, *Soundscape & at work*, and *Exercise*. The *Soundscape & at work* factor was highlighted as being of particular interest as the factor loadings and consequent interpretation suggested negative and positive changes in attentional engagement as a result of noise masking. In addition, 5 aspects of podcast listening engagement were conceptualised and measured across the sample, providing a comprehensive assessment of contemporary podcast consumption practices. The results of further correlation tests found that the aspects of podcast listening engagement measures were positively related to the factors of environmental context.

Chapter 5 then presented the results of acoustic measurements to determine the insertion loss and gain performance of noise cancellation and variable hear-through modes in three commercially available true wireless earbud devices. The results showed that all devices were capable of a wide range of both insertion loss and gain, providing significant levels of frequency dependant attenuation and amplification to achieve this. This led to a key finding that the hear-through performance in each device was far from transparent, with the modes that provided the listener with the greatest awareness of external sound boosting external sound signals by as much as 12 dB, relative to open ear listening. However, this boosting was largely limited to the 1600-4000 Hz region of frequencies, those considered most important to speech intelligibility. This was highlighted as a finding that supported the notion that the hear-through functionality in these devices encourage longer periods of "always on" usage, while also potentially providing listeners with an increased sense of personal safety when listening in unfamiliar or potentially hazardous public spaces.

Chapter 6 presented the results of a listening experiment that was conducted to investigate how listener engagement with speech-based media over earphones is affected by environmental noise, and how this is further mediated by the boosting of environmental noise that is provided by hear-through modes in true wireless earbuds/hearables, as measured in Chapter 5. The results for the first part of the study found that exposure to noise resulted in lower engagement scores (measured across *Immersion*, *Narrative Absorption*, and *Enjoyment*) for most participants, however, a small proportion reported that their awareness of noise contributed to a raise in the *Immersion* they experienced. The results from the second part of the study revealed that the *Hear-through* condition registered similar, or in some cases higher, engagement scores than both the *Low* and *High Noise* conditions, despite receiving the highest *Listening Effort* scores. Together, these results suggested that boosting of environmental

noise by hear-through modes in true wireless earbud/hearable devices can provide listeners with an increased awareness of their surroundings, without compromising their engagement.

7.2 Contributions to the Field

The work presented in this thesis provides a number of unique contributions to the research fields of applied psychology, mobile audio consumption, augmented reality, headphone technology, and media studies. These are as follows:

- **The study presented in Chapter 4 was the first of its kind to quantitatively investigate the relationship between listeners' environmental context and their attentional engagement in podcast listening experiences.** Prior literature had suggested that such a relationship might exist, however this was only ever based on qualitative interview/free-text response data or anecdotal observations from scholars. In conducting a large scale and wide ranging listener behaviours survey and performing an exploratory factor analysis on the consequent data, this study was the first investigation of this research question using established quantitative methods.
- **The results presented in Chapter 4 were the first to support the hypothesis that a listener's environmental context exerts influence over their attentional engagement in podcast listening experiences.** The exploratory factor analysis methodology that was used revealed how different forms of environmental context related to attentional engagement and led to the formation of five influencing factors of environmental context over attentional engagement in podcast listening experiences.
- **The study presented in Chapter 5 was the first of its kind to measure the insertion loss and gain performance of noise cancellation and adjustable hear-through modes in commercially available devices that feature both modes of functionality.** While prior literature had measured the insertion loss performance of passive noise reduction and active noise cancellation in headphone devices, and other literature had measured hear-through devices that intended to deliver a fully acoustically transparent listening experience, no prior work had quantified the full range of insertion loss and gain performance for devices that featured adjustable hear-through functionality.

- **The results presented in Chapter 5 contained the novel finding that none of the measured true wireless earbuds appeared to be attempting to achieve complete acoustic transparency in their hear-through mode functionality.** Instead, in each device, the hear-through modes that provided the greatest awareness of external sound imparted significant amounts of boosting of external sound, localised around the frequencies commonly considered most important to the intelligibility of speech. These results supported the observations of prior literature that had suggested similar earbuds were being designed with the ability to function as assistive hearing devices. The work presented Chapter 5 is the first to have observed this boosting across multiple devices in a battery of acoustic measurements.

- **The study presented in Chapter 6 was the first of its kind to perform controlled listening experiments investigating the effects of both passively occluded and hear-through boosted environmental noise on listener engagement with speech-based media, measured across Immersion, Narrative Engagement, and Enjoyment.** While prior experimental work [146] has been conducted to investigate the effects of acoustic masking on listener engagement with spoken stories, the motivations and aims of this work are not directly compatible with those contained of this thesis, and aspects of the methodology used mean that valid comparisons cannot be made between the results of either study.

- **The results presented in Chapter 6 were the first to show that for most listeners, in controlled listening experiments, engagement with speech-based media decreases and Listening Effort increases, as a result of exposure to passively occluded environmental noise.** However, a small proportion of listeners reported experiencing increased engagement as a direct result of exposure to noise.

- **Further results presented in Chapter 6 also led to the novel finding that listeners' appear to respond differently to environmental noise that is boosted by hear-through, compared to noise that is perceived only through the passive occlusion of earphones.** This suggests that boosting from hear-through in earbuds can provide listeners with an increased awareness of their surroundings, without compromising their engagement.

7.3 Restatement of Hypothesis

In Chapter 1 the following overall hypothesis was presented as a guiding statement summarising the underlying aims of the work contained within the thesis:

> *The quality of engagement that listeners' experience when consuming speech-based media using auditory headsets is mediated by elements of their surrounding environmental context, the environmental noise that emanates from within it, and the nature of occlusion provided by the auditory headset.*

The conclusions drawn from the listener behaviours survey and listening experiment work presented in Chapters 4 and 6 serve to overwhelmingly support all aspects of the overall hypothesis. These conclusions are drawn from evidence that clearly shows how environmental context, environmental noise, and the nature of environmental noise occlusion in auditory headsets all exert influence over aspects of listeners' engagement with speech-based media in a myriad of different ways. Furthermore, the acoustic measurement work presented in Chapter 5 facilitated the results and conclusions of Chapter 6 and its subsequent support of the overall hypothesis. Specifically, the work contained in this thesis supports the hypothesis in the following ways:

- The results of the listener behaviour survey and exploratory factor analysis presented in Chapter 4 identified five factors of environmental context that both positively and negatively influenced listeners' attentional engagement in podcast listening experiences. These factors were interpreted and labelled as *Outdoors*, *Indoors & at home*, *Evenings*, *Soundscape & at work*, and *Exercise*.
- The results of the listening experiment presented in Chapter 6 showed that exposure to increased levels of environmental noise generally results in decreased listener engagement, measured across Immersion, Narrative Absorption, and Enjoyment, however, results also showed that a small proportion of listeners experience increased engagement as a result of exposure to passively occluded noise.
- Further results presented in Chapter 6 found that the *Hear-through* condition, which actively boosted environmental noise relative to open ear listening, received similar or in some cases

higher engagement scores than both the *Low* and *High Noise* conditions. The deviation in engagement results observed between the *High Noise* and *Hear-through* conditions was a particularly compelling result that supports the nature of occlusion provided by the auditory headset aspect of the hypothesis.

7.4 Future Work

Due to the novelty of the research topics of the original work of this thesis, and the breadth of different research fields they are applicable to, there is extensive scope for future research to be conducted that builds on the findings presented.

Future work is required to provide further validation of the results in Chapter 4, with acceptance of the results of the EFA being conditional on the findings of further confirmatory factor analysis studies. As the data for the present study was collected using a self-report survey, and largely based on participants' typical listening habits, it is also recommended that future studies should use a longitudinal study methodology, such as experience sampling method [288], to further improve the accuracy of results.

The findings presented in Chapter 4 are highly pertinent to the fields of podcast studies, media engagement, and mobile listening experiences. The results provide a basis for future research aiming to explore specific aspects of podcast listening engagement and factors of environmental context as they relate to episodes of listening in different environmental contexts. The research also has potential implications for future research exploring mobile audio listening and environmental context from the perspectives of media personalisation and attentional processing.

The use and application of the survey and accompanying exploratory factor analysis methodology in Chapter 4 provides a methodological basis for future studies that wish to validate the results presented in this thesis. Furthermore, researchers who wish to conduct further studies that investigate how other elements of environmental context, that were not included in the scope of this thesis, might mediate attentional engagement or indeed any other engagement metric of interest. Similarly, this methodology could be applied to the study of how engagement in media consumption is mediated by other categories of context such as situational or social context.

Further work that builds on the findings of Chapter 5 should take the form of perceptual sound quality evaluations of the noise cancellation and hear-through mode performance in the devices tested, in addition to further technical investigation of these and other similar devices to investigate the spectral modification performance when presented with different kinds of environmental noise and non-steady state noise signals. It is expected that the results in this chapter will be useful to those who seeking to produce models that replicate key aspects of insertion loss and gain performance in the noise cancellation and hear-through modes of contemporary true wireless earbud devices.

Future work is also required to provide validation of the results of the listening experiment presented in Chapter 6. Beyond this, a series of further experiments should be performed to investigate how different types of environmental noise soundscapes effect listener engagement with a wider variety of speech-based media stimuli. To build further on the results of Chapter 4 these experiments should focus on investigating how listener engagement differs in soundscapes that mainly comprise of mechanical sounds and typical sounds from the workplace. Further experiments should be conducted to investigate the effects of a wider array of different human sounds and soundscapes. Work should also be conducted to investigate if the key results remain consistent if an experimental methodology is used whereby participants are also presented with a video representation of scenes from the relevant soundscape environment using a screen or virtual reality apparatus.

Finally, a key finding of this thesis is that results in both of the studies presented in Chapters 4 and 6 suggested that some listeners experience increased levels of engagement with speech-based media when exposed to higher levels of environmental noise. These studies are the first to have shown this result using quantitative and mixed methodologies, adding to similar prior findings in the literature that were produced using exclusively qualitative methodologies. It is therefore important that research is now conducted to examine this somewhat counter intuitive phenomenon further.

Firstly this should seek to confirm that the key result holds across a wider body of experimental methodologies. Then investigations should be carried out to understand why some listeners experience this phenomenon and how it sits alongside other contextual considerations of listeners, such as how their awareness of their surroundings impacts their sense of personal safety. Finally, it should then be considered how these findings could be leveraged to provide listeners with improved personalised

mobile listening experiences that accurately reflect and respond to individuals' contextual listening preferences.

7.5 Final Remarks

The purpose of this thesis has been to conduct the first major body of work aimed at understanding how different aspects of listeners' environmental context, and the sound that emanates from this context, relates to their engagement with speech-based media consumed over headphones. The results have shown that aspects of environmental context, noise, and auditory headset technology are all capable of influencing listener engagement in both negative and positive ways. The work has also highlighted how this relationship is often a complex interaction involving many competing factors, with concerns around personal safety in mobile listening experiences also a key factor.

The work demonstrates the demand for further research to be conducted which maps how other aspects of environmental, situational, and social context affect listener engagement. There is also potential for future work to be conducted to investigate how other types of public soundscapes relate to engagement.

An area of particular interest for future work stems from the finding that some listeners do in fact experience increased engagement as a result of exposure to environmental noise. However, there is clearly a great deal of future work required to fully understand this phenomenon. Ultimately, this work could potentially be used to inform ways of intelligently personalising consumers' mobile listening experiences. This could be achieved by using context aware sensing capabilities to control adjustable hear-through and noise cancellation settings and adjust the equalisation of external sound signals in accordance with the preferences of individual listeners.

It should be noted that due to the lack of prior work and the almost infinitely large combinations of possible environmental, situational, and social contexts that a mobile media consumer might encounter, the work contained in this thesis has barely scratched the surface of exploring these relationships between context, sound, and engagement. Nevertheless, the ubiquity and popularity of mobile media consumption shows no signs of dissipating. In September 2024 Apple announced a new hearing health experience [277] for the Apple Airpod Pro 2, comprising of loud sound reduction, a scientifically validated hearing test, and a clinical-grade hearing aid feature. This illustrates the need for future

research that keeps abreast of, and feeds into, the continually evolving technological capabilities of mobile devices and auditory headsets to inform and improve the mobile listening experiences of tomorrow.

The proliferation of ideas and knowledge through the sharing of stories has been, and continues to be, an important factor in the growth and development of all aspects of human civilisation. For much of human history this remained exclusively an oral tradition. However, the inventions and continued development of recorded, broadcast, and social media are just some of the many technological advancements that have now led to dramatic changes in the ways in which stories are able to be communicated and the audiences they are able to reach.

In many ways an individual walking down a busy street while using earbuds to listen to a podcast on their phone is engaging in exactly the same ritualistic tradition that generation after generation of their ancestors would have when sat round a campfire listening to the narrative of a story unfold. Yet in actual fact the listening experience that the individual on the street is engaging in represents a significant deviation from that of the oral tradition, albeit one that the listener may not be consciously aware of at the point of listening.

It is the personal view of the author that as so many of the stories we experience are now consumed on mobile devices via mediatized means, the lack of knowledge around the effects that aspects of environmental and other contextual factors can have on our perception of these stories represents a significant gap in our understanding.

If nothing else it is hoped that the work contained in this thesis will serve as a timely reminder of the value of research that seeks to understand how people interact with and respond to the technologies they use on a daily basis, and the important part this can play in informing future technological advancements.

Appendix A

Supplementary Material for Chapter 4

This appendix contains all supplementary material referenced in Chapter 4. This includes the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire, and the ethical approval form (Harrison101121) for the study that was approved by the University of York Electronic Engineering department ethics committee.

Environmental Context Podcast Listening Habits Pilot Study

Start of Block: Participant Consent

Thank you for your interest in participating in this pilot study. Your contribution is highly valued and will aid research seeking to further understanding of how listening experiences are influenced by the environmental context in which they take place.

Instructions

In order to participate in this study you must have existing experience of listening to podcasts using a smartphone. The questionnaire is comprised of a series of multiple choice, ordinal scale & textual response questions and should take between 10 - 15 minutes to complete, depending on the extent of your listening experience.

Participant Consent

This questionnaire has been designed to adhere to the ethical codes and practices specified by the University of York. The personal data that is provided to this questionnaire by participants is anonymised and treated with the strictest confidentiality. Any elements of participant's personal data that enable identification will only be stored until the 31st of December 2021, after which it will be destroyed. Any anonymised personal data that is kept beyond this point will remain anonymised and will be used for the purposes of analysis and reporting results via internal and external publication.

In agreeing with this statement, you are confirming that you are over the age of 18, consenting to your participation in this questionnaire and the use of the anonymised personal data you provide in subsequent data analysis for the purposes of research conducted by the principal investigator. You are free to withdraw your consent for your participation in the study at any point.

This research is being conducted by the principal investigator, Jay Harrison (jah671@york.ac.uk) as part of an ongoing PhD project based in the AudioLab at the University of York. If you have any queries or wish to withdraw your consent from the study, please contact the principal investigator by email.

I confirm that I have read and understood the above and consent to my participation in the study.

Page 1 of 27

FIGURE A.1: Page 1 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4.

○ Yes

○ No

End of Block: Participant Consent

Start of Block: Participation Qualification

Do you have experience of listening to podcasts using a smartphone?

○ Yes

○ No

End of Block: Participation Qualification

Start of Block: Personal Information

Display This Question:

If Do you have experience of listening to podcasts using a smartphone? = Yes

What is your age?

▼ 18-25 ... 66+

Page 2 of 27

FIGURE A.2: Page 2 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4.

To which gender do you most identify?

- ○ Male
- ○ Female
- ○ Non-binary / third gender
- ○ Prefer not to say
- ○ Other not listed (Please specify)

_____

What is your country of residence?

_____

What is your primary occupation?

- ○ Employed (full time)
- ○ Employed (part time)
- ○ Unemployed
- ○ Homemaker
- ○ Student
- ○ Retired
- ○ None of the above

Page 3 of 27

FIGURE A.3: Page 3 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4.

Display This Question:
If What is your primary occupation? = Employed (full time)
Or What is your primary occupation? = Employed (part time)
Or What is your primary occupation? = Student

When do you usually work?

- ○ Weekdays (daytime)
- ○ Weekdays (nights)
- ○ Weekdays (alternating shifts)
- ○ Weekends (daytime)
- ○ Weekends (nights)
- ○ Weekends (alternating shifts)
- ○ Weekdays & weekends (daytime)
- ○ Weekdays & weekends (nights)
- ○ Weekdays & weekends (alternating shifts)
- ○ Other not listed (Please specify)

_____

Display This Question:
If What is your primary occupation? = Employed (full time)
Or What is your primary occupation? = Employed (part time)
Or What is your primary occupation? = Student

Do you currently commute to your occupation?

- ○ Yes
- ○ No

Page 4 of 27

FIGURE A.4: Page 4 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4.

Display This Question:

If Do you currently commute to your occupation? = Yes

What method(s) of transportation do you usually use to commute to your occupation?

- ☐ Drive my own vehicle
- ☐ Passenger in someone else's vehicle
- ☐ Train
- ☐ Bus
- ☐ Bicycle
- ☐ Walk
- ☐ ⊗ Other not listed (Please specify)

_____

End of Block: Personal Information

Start of Block: Podcast Information

In all of the questions in the next section of the survey, the term podcast is defined as any form of on-demand primarily speech based audio. This also includes on-demand radio programmes that are consumed after live transmission using a streaming service.

End of Block: Podcast Information

Start of Block: Playback p 1

FIGURE A.5: Page 5 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4.

What playback devices have you used with a smartphone to listen to podcasts? (Please select all that apply)

- ☐ Wired over-ear headphones
- ☐ Wired in-ear headphones
- ☐ Wireless over-ear headphones
- ☐ Wireless in-ear headphones
- ☐ Bluetooth speaker(s)
- ☐ Built-in smartphone speakers
- ☐ Built-in vehicle speakers
- ☐ Bone conduction headset
- ☐ Other not listed (Please specify) _____

End of Block: Playback p 1

Start of Block: Playback p 2

Display This Question:

If If What playback devices have you used with a smartphone to listen to podcasts? (Please select all that apply) q://QID39/SelectedChoicesCount Is Greater Than or Equal to 2

Carry Forward Selected Choices - Entered Text from "What playback devices have you used with a smartphone to listen to podcasts? (Please select all that apply)"

FIGURE A.6: Page 6 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4.

Which of the following playback devices do you most often use with a smartphone to listen to podcasts? (Please select one)

- ○ Wired over-ear headphones
- ○ In-ear headphones
- ○ Wireless over-ear headphones
- ○ Bluetooth earbuds (i.e. Airpods
- ○ Bluetooth speaker(s)
- ○ Built-in smartphone speakers
- ○ Built-in vehicle speakers
- ○ Bone conduction headset
- ○ Other not listed (Please specify)

End of Block: Playback p 2

Start of Block: Weekday p

On an average **weekday**, how much time do you typically spend listening to podcasts on a smartphone?

▼ 0 minutes ... More than 10 hours

What percentage of this time is typically spent listening whilst simultaneously engaging in other activities?

▼ 0% (I never engage in other activities whilst listening on weekdays) ... 100% (I always engage in other activities whilst listening on weekdays)

End of Block: Weekday p

Start of Block: Weekend p

Page 7 of 27

FIGURE A.7: Page 7 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4.

On an average **weekend day**, how much time do you typically spend listening to podcasts on a smartphone?

▼ 0 minutes ... More than 10 hours

What percentage of this time is typically spent listening whilst simultaneously engaging in other activities?

▼ 0% (I never engage in other activities whilst listening on weekend days) ... 100% (I always engage in other activities whilst listening on weekend days)

End of Block: Weekend p

Start of Block: Multitasking p 1

Display This Question:

If What percentage of this time is typically spent listening whilst simultaneously engaging in other... != 0% (I never engage in other activities whilst listening on weekdays)

Or What percentage of this time is typically spent listening whilst simultaneously engaging in other... != 0% (I never engage in other activities whilst listening on weekend days)

FIGURE A.8: Page 8 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4.

What activities have you also engaged in while simultaneously listening to podcasts using a smartphone? (Please select all that apply)

- ☐ Watching TV
- ☐ Reading
- ☐ Listening to music
- ☐ Talking on the phone
- ☐ Sending messages via phone or computer
- ☐ Using social media sites
- ☐ Watching films
- ☐ Other computer activities
- ☐ Playing video games
- ☐ Doing housework
- ☐ Preparing food
- ☐ Exercising
- ☐ Other not listed (Please specify) _____

End of Block: Multitasking p 1

Start of Block: Multitasking p 2

Page 9 of 27

FIGURE A.9: Page 9 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4.

Display This Question:

If If What activities have you also engaged in while simultaneously listening to podcasts using a smartphone? (Please select all that apply) q://QID45/SelectedChoicesCount Is Greater Than or Equal to 2

Carry Forward Selected Choices - Entered Text from "What activities have you also engaged in while simultaneously listening to podcasts using a smartphone? (Please select all that apply)"

Which of the following activities do you most often also engage in while simultaneously listening to podcasts using a smartphone? (Please select one)

- ○ Watching TV
- ○ Reading
- ○ Listening to music
- ○ Talking on the phone
- ○ Sending messages via phone or computer
- ○ Using social media sites
- ○ Watching films
- ○ Other computer activities
- ○ Playing video games
- ○ Doing housework
- ○ Preparing food
- ○ Exercising
- ○ Other not listed (Please specify)

End of Block: Multitasking p 2

Start of Block: Location p 1

FIGURE A.10: Page 10 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4.

What locations have you listened to podcasts in using a smartphone? (Please select all that apply)

- ☐ At home
- ☐ At work
- ☐ Driving a vehicle
- ☐ Traveling on public transport
- ☐ On a run
- ☐ At the gym
- ☐ Walking in a rural environment
- ☐ Walking in an urban environment
- ☐ Other not listed (Please specify) _____

End of Block: Location p 1

Start of Block: Location p 2

Display This Question:

If If What locations have you listened to podcasts in using a smartphone? (Please select all that apply) q://QID47/SelectedChoicesCount Is Greater Than or Equal to 2

Carry Forward Selected Choices - Entered Text from "What locations have you listened to podcasts in using a smartphone? (Please select all that apply)"

FIGURE A.11: Page 11 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4.

In which of the following locations do you most often listen to podcasts using a smartphone? (Please select one)

- ○ At home
- ○ At work
- ○ Driving a vehicle
- ○ Traveling on public transport
- ○ On a run
- ○ At the gym
- ○ Walking in a rural environment
- ○ Walking in an urban environment
- ○ Other not listed (Please specify)

End of Block: Location p 2

Start of Block: Discovery p g 1

FIGURE A.12: Page 12 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4.

What methods have you used to discover podcasts? (Please select all that apply)

- ☐ Streaming services (i.e. Spotify, Pandora, etc.)
- ☐ Searching the internet
- ☐ Listening to podcasts
- ☐ Listening to radio
- ☐ Recommendations from friends/family
- ☐ Recommendations on social media
- ☐ Recommendations from YouTube creators
- ☐ Recommendations from YouTube based on your viewing history
- ☐ Other not listed (Please specify) _____

End of Block: Discovery p g 1

Start of Block: Discovery p g 2

Display This Question:

If If What methods have you used to discover podcasts? (Please select all that apply) q://QID49/SelectedChoicesCount Is Greater Than or Equal to 2

Carry Forward Selected Choices - Entered Text from "What methods have you used to discover podcasts? (Please select all that apply)"

FIGURE A.13: Page 13 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4.

Which of the following methods do you most often use to discover podcasts? (Please select one)

- ○ Recommendations from streaming services (i.e. spotify, pandora, etc.)
- ○ Searching the internet
- ○ Curated recommendations from audio based media (i.e. podcasts, radio, etc.)
- ○ Adverts on social media
- ○ Personal Recommendations from friends/family
- ○ Recommendations on social media
- ○ Recommendations from YouTube creators
- ○ Recommendations from YouTube based on your viewing history
- ○ Other not listed (Please specify)

End of Block: Discovery p g 2

Start of Block: Discovery p s 1

Display This Question:

If What methods have you used to discover podcasts? (Please select all that apply) = Streaming services (i.e. Spotify, Pandora, etc.)

FIGURE A.14: Page 14 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4.

What methods have you used to discover podcasts on streaming services? (Please select all that apply)

- ☐ Using the search bar and/or browsing to make self directed selections
- ☐ Personally curated playlists and/or recommendations that follow a particular theme
- ☐ Playlisting and/or recommendations generated from my listening history
- ☐ General editorial recommendations/adverts not related to my listening history or any particular theme
- ☐ Other not listed (Please specify) _____

End of Block: Discovery p s 1

Start of Block: Discovery p s 2

Display This Question:

If If What methods have you used to discover podcasts on streaming services? (Please select all that apply) q://QID51/SelectedChoicesCount Is Greater Than or Equal to 2

Carry Forward Selected Choices - Entered Text from "What methods have you used to discover podcasts on streaming services? (Please select all that apply)"

Which of the following methods do you most often use to discover podcasts on streaming services? (Please select one)

- ○ Using the search bar and/or browsing to make self directed selections
- ○ Personally curated playlists and/or recommendations that follow a particular theme
- ○ Playlisting and/or recommendations generated from my listening history
- ○ General editorial recommendations/adverts not related to my listening history or any particular theme
- ○ Other not listed (Please specify)

Page 15 of 27

FIGURE A.15: Page 15 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4.

End of Block: Discovery p s 2

Start of Block: Subscription p

Do you currently make a regular financial contribution to consume podcasts or support podcast creators (i.e. Spotify or Patreon)? (Please select one)

- ○ Yes, I pay to consume podcasts on a streaming service (i.e. Spotify)
- ○ Yes, I pay to support podcast creators on a subscription service (i.e. Patreon)
- ○ Yes, I pay to consume podcasts on a streaming service (i.e. Spotify) & also pay to support podcast creators on a subscription service (i.e. Patreon)
- ○ No

End of Block: Subscription p

Start of Block: Skipping p

How much of a podcast do you typically listen to on a smartphone before searching for something else that better satisfies your listening preferences?

▼ Less than 5% ... Not applicable

End of Block: Skipping p

Start of Block: Attention p 1

FIGURE A.16: Page 16 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4.

How would you describe the attention you typically afford to the **listening experience** when consuming **podcasts on a smartphone**? (Please select one)

- ○ Focused
- ○ Sustained
- ○ Selective
- ○ Alternating
- ○ Divided
- ○ Background
- ○ Other not listed (Please specify)

_____

End of Block: Attention p 1

Start of Block: Attention p 2

FIGURE A.17: Page 17 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4.

How would you describe the attention you typically afford to the **listening experience** when consuming **podcasts on a smartphone** as the **only activity occupying your attention**? (Please select one)

- ○ Focused
- ○ Sustained
- ○ Selective
- ○ Alternating
- ○ Divided
- ○ Background
- ○ Other not listed (Please specify)
  _____
- ○ Not applicable

End of Block: Attention p 2

Start of Block: Attention p 3

FIGURE A.18: Page 18 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4.

How would you describe the attention you typically afford to the **listening experience** when consuming **podcasts on a smartphone** and **simultaneously engaging in other activities**? (Please select one)

- ○ Focused
- ○ Sustained
- ○ Selective
- ○ Alternating
- ○ Divided
- ○ Background
- ○ Other not listed (Please specify)
  _____
- ○ Not applicable

End of Block: Attention p 3

Start of Block: Attention p 4

FIGURE A.19: Page 19 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4.

How would you describe the attention you typically afford to **your surroundings** when listening to **podcasts on a smartphone**? (Please select one)

- ○ Focused
- ○ Sustained
- ○ Selective
- ○ Alternating
- ○ Divided
- ○ Background
- ○ Other not listed (Please specify)

_____

End of Block: Attention p 4

Start of Block: Actively Engaged Information

In the following section of the survey the term 'actively engaged' is used to describe an attentional state where the listener is fully focused on the listening experience.

End of Block: Actively Engaged Information

Start of Block: Podcasts Time EFA

I feel actively engaged in the listening experience when using a smartphone to consume podcasts...

	Likert Scale					
	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	Not applicable

FIGURE A.20: Page 20 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4.

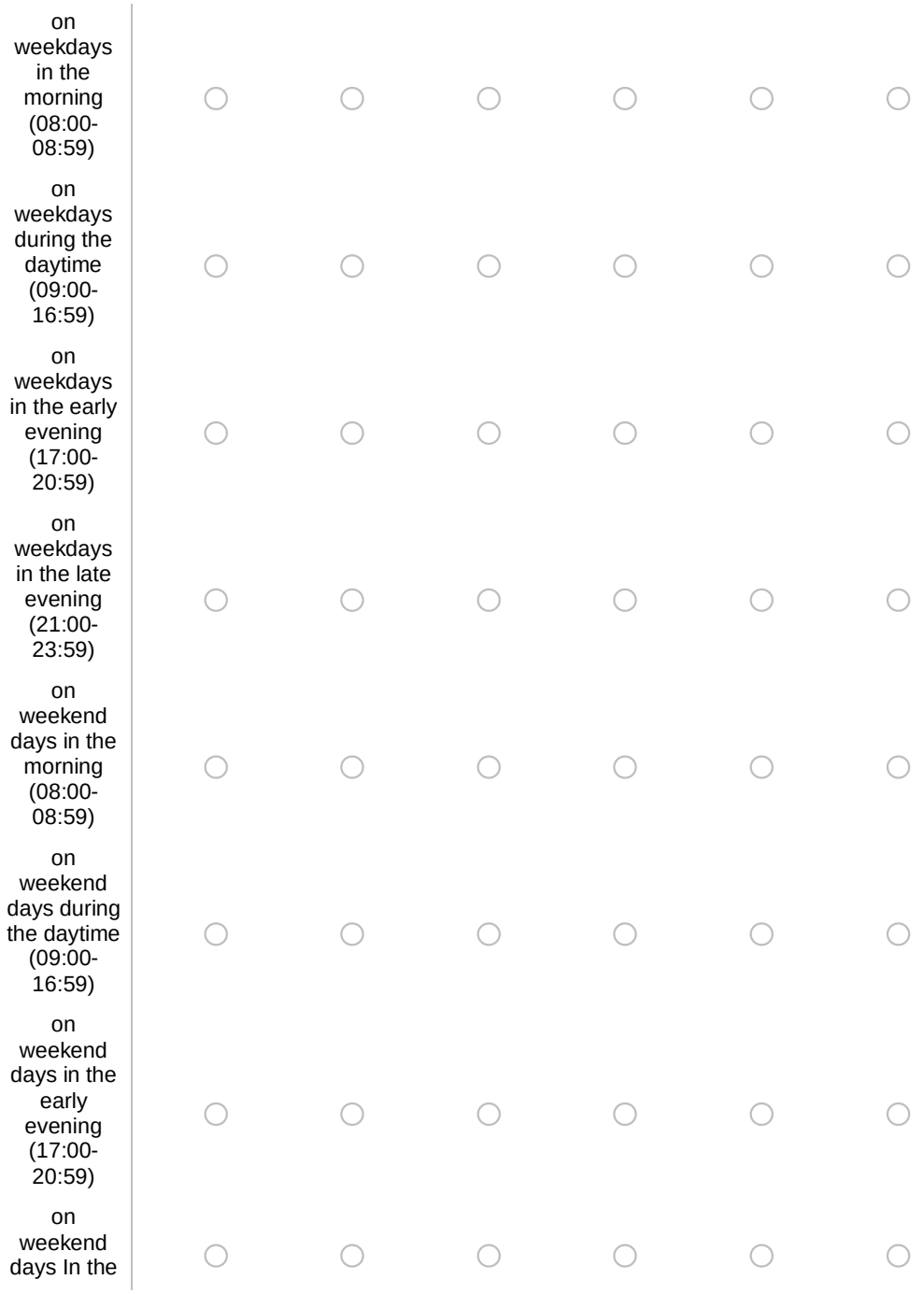


FIGURE A.21: Page 21 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4.

late evening (21:00-23:59)

End of Block: Podcasts Time EFA

Start of Block: Podcasts Location EFA

I feel actively engaged in the listening experience when using a smartphone to consume podcasts...

	Likert Scale					
	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	Not applicable
at home	○	○	○	○	○	○
at work	○	○	○	○	○	○
whilst driving a vehicle	○	○	○	○	○	○
on public transport	○	○	○	○	○	○
on a run	○	○	○	○	○	○
at the gym	○	○	○	○	○	○
on a walk in a rural environment	○	○	○	○	○	○
on a walk in an urban environment	○	○	○	○	○	○

End of Block: Podcasts Location EFA

FIGURE A.22: Page 22 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4.

Start of Block: Podcasts Sound EFA

I feel actively engaged in the listening experience when using a smartphone to consume podcasts...

	Likert Scale					
	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	Not applicable

FIGURE A.23: Page 23 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4.

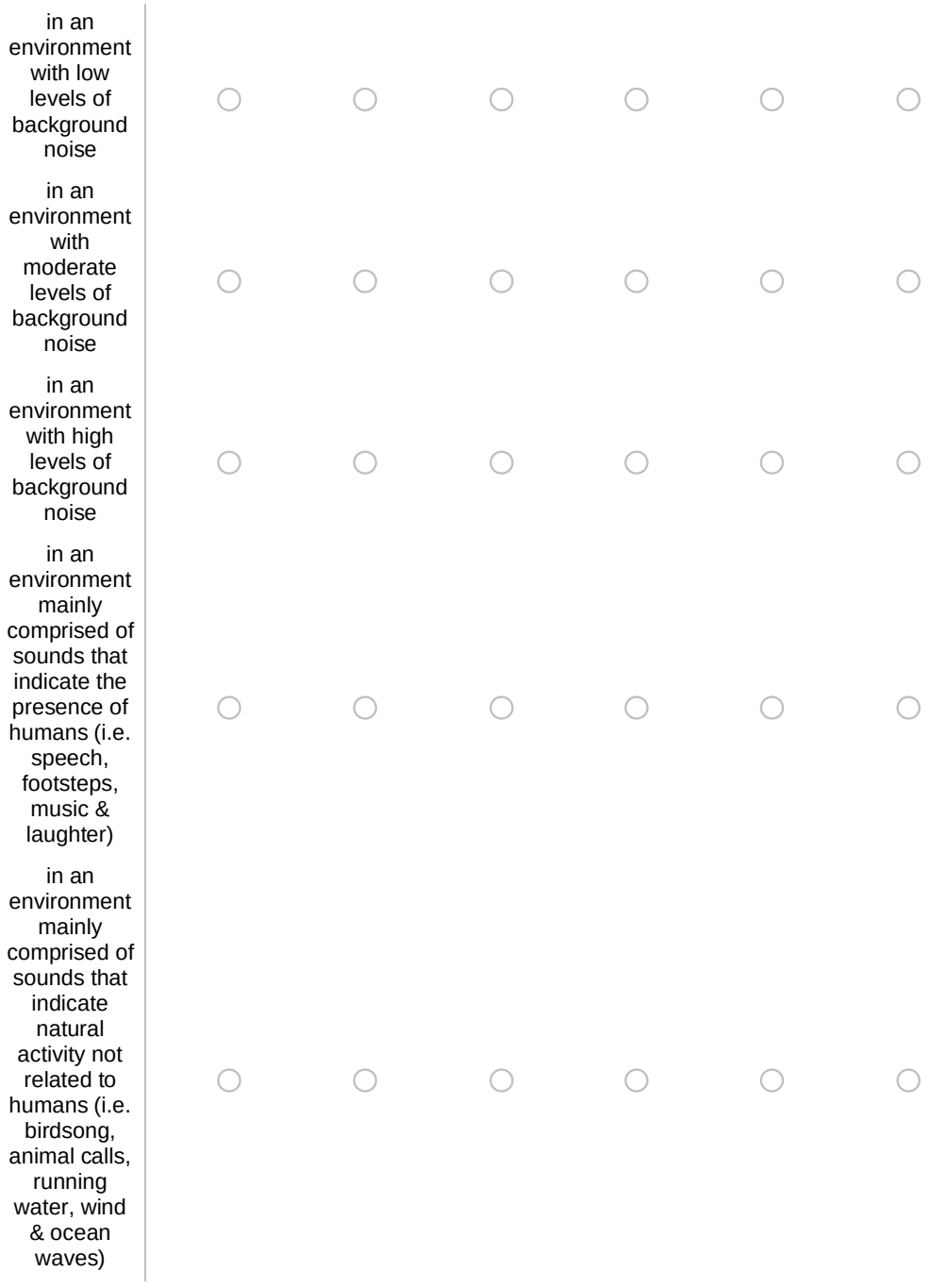


FIGURE A.24: Page 24 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4.

in an environment mainly comprised of mechanical sounds (i.e. machinery, transport & construction)	○	○	○	○	○	○

End of Block: Podcasts Sound EFA

Start of Block: Podcasts Weather EFA

I feel actively engaged in the listening experience when using a smartphone to consume podcasts...

	Likert Scale					
	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	Not applicable

FIGURE A.25: Page 25 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4.

outdoors in warm conditions	○	○	○	○	○	○
outdoors in cold conditions	○	○	○	○	○	○
outdoors in light conditions	○	○	○	○	○	○
outdoors in dark conditions	○	○	○	○	○	○
indoors in warm conditions	○	○	○	○	○	○
indoors in cold conditions	○	○	○	○	○	○
indoors in light conditions	○	○	○	○	○	○
indoors in dark conditions	○	○	○	○	○	○

End of Block: Podcasts Weather EFA

Start of Block: Thematic p

Please enter any observations you may have on how different factors of environmental context influence your engagement level when listening to podcasts using a smartphone:

_____

Please enter any observations you may have on how different factors of environmental context influence your preference for listening to specific types of podcasts using a smartphone:

_____

Page 26 of 27

FIGURE A.26: Page 26 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4.

End of Block: Thematic p

Page 27 of 27

FIGURE A.27: Page 27 of 27 of the consent form, information sheet, survey Instrument, and survey Logic used in the questionnaire described in Chapter 4.

PSEC Application Form V4

THE UNIVERSITY of York

Application Form for Physical Sciences Ethics Committee Approval

Advice for applicants on completing the form

Please ensure that the information provided is:

- *Accurate and concise*
- *Clear and simple and easily understood by a lay person*
- *Free of jargon, technical terms and abbreviations*

Further advice and information can be obtained from your departmental representative on the PSEC and at: http://www.york.ac.uk/admin/aso/ethics/cttee.htm

Please return completed (typed) form to your departmental representative via email to:

elec-ethics@york.ac.uk

Title of project: Audio Based-Media Listening Habits Questionnaire

SECTION 1 DETAILS OF APPLICANTS

Details of principal investigator (name, appointment and qualifications)

Principal investigator: Jay Harrison
Appointment: 1st year PhD student
Qualifications: BSc (Hons), MSc

Names, appointments and qualifications of additional investigators *(student applicants should include their project supervisor(s) here)*

Additional investigator 1: Prof. Damian Murphy
Appointment: University supervisor
Qualifications: BSc (Hons), MSc, DPhil
Additional investigator 2: Dr Jon Francombe
Appointment: Industrial supervisor
Qualifications: BSc (Hons), PhD

7th May 2015

FIGURE A.28: Page 1 of 9 of the ethical approval form (Harrison101121) for the study in Chapter 4, approved by the University of York Electronic Engineering department ethics committee.

Location(s) of project

York, United Kingdom

SECTION 2 FUNDERS

What is the funding source(s) for the project?

XR Stories & BBC R&D

Please answer the following:

(i) Does the express and direct aim of the research or other activity raise ethical issues?

YES [ ] NO [X]

(ii) Is there any obvious or inevitable adaptation of research findings to ethically questionable aims?

YES [ ] NO [X]

(iii) Is the work being funded by organisations tainted by ethically questionable activities?

YES [ ] NO [X]

(iv) Are there any restrictions on academic freedoms – notably, to adapt and withdraw from ongoing research, and to publish findings?

YES [ ] NO [X]

If you answered **Yes** to any of the above, please give details below:

N/A

FIGURE A.29: Page 2 of 9 of the ethical approval form (Harrison101121) for the study in Chapter 4, approved by the University of York Electronic Engineering department ethics committee.

SECTION 3 DETAILS OF PROJECT OR OTHER ACTIVITY

Aims (100 words max)

To conduct a questionnaire seeking to contribute to the various concurrent active to passive listening/interaction continuums in the literature. The outcomes of this contribution would depend on the results of the study but could represent a systematic review & validation of existing concepts or the proposal of a new actively engaged listener 'persona' applicable to audiences that demonstrate consistent consumption habits in both both music & podcast listening.

It is also hoped that this study might indicate the appetite of the general audience for responsive/explicit as opposed to perceptive/implicit interactions with media by investigating how active/passive their existing listening habits are.

Background (250 words max)

Existing research into concepts surrounding active/passive listening and consumption has only considered individual media separately, mainly focusing on its application to music.

This study would be the first step in considering this branch of research in light of the development of new audio-based media (podcasts/audiobooks/interactive audio experience apps) and mobile listening experience device hardware.

To conduct research exploring the user perception of personalised audio experiences as influenced by context, it is first deemed necessary to investigate the listening habits of the general audience. It is intended that this research will provide an insight into the audience's tendency to be actively engaged in the listening experience by examining the extent to which listener's consumption habits are active or passive across both music and podcast listening experiences.

It is expected that gaining an understanding of the active to passive nature of user interactions with media will indicate potential future use case applications for context aware personalised audio experiences in different listening circumstances.

The study will seek to answer the following two primary research questions:

- Are individual's active/passive audio-based media consumption habits consistent between music & podcasts?
- How are individual's audio-based media consumption habits affected by different factors of environmental context?

FIGURE A.30: Page 3 of 9 of the ethical approval form (Harrison101121) for the study in Chapter 4, approved by the University of York Electronic Engineering department ethics committee.

Brief outline of project/activity (250 words max)

A survey primarily based on a series of multiple choice and ordinal Likert scale questions will be distributed online to participants who would be described as representing a general media audience. The questionnaire will also comprise of a secondary component asking participants for textual qualitative responses.

The data obtained from the Likert questionnaire will be processed using exploratory factor analysis to search for the presence of latent variables describing the common variance and possibly indicating the presence of hidden factors informing the participants tendency to be actively engaged with music & podcasts in different environmental contexts.

The textual response data will be processed using a thematic analysis approach to explore any additional interesting ideas or trends raised by the participants in response to the survey questions.

Study design *(if relevant – e.g. randomised control trial; laboratory-based)*

The study will be conducted entirely online using the University of York approved Qualtrics survey platform.

If the study involves participants, how many will be recruited?

At least 200 with an ideal target of 300.

If applicable, what is the statistical power of the study, i.e. what is the justification for the number of participants needed?

There are several highly cited papers available that state a minimum of 300 as being an appropriate number of participants for exploratory factor analysis. Fewer would be feasible if the results of the analysis provided factors with high loading above 0.8, but this cannot be guaranteed. Also, a study [1] that used a questionnaire with EFA to investigate a similar concept in the field had a sample size of 590 participants.

[1] A. E. Krause, S. Mackin, A. Mossman, T. Murray, N. Oliver, and V. Tee, "Conceptualizing Control in Everyday Music Listening: Defining Dominance," *Music Sci.*, vol. 3, p. 205920432093164, 2020, doi: 10.1177/2059204320931643.

FIGURE A.31: Page 4 of 9 of the ethical approval form (Harrison101121) for the study in Chapter 4, approved by the University of York Electronic Engineering department ethics committee.

SECTION 4 RECRUITMENT OF PARTICIPANTS

How will the participants be recruited?

Participants will be recruited via sources in BBC R&D and will be paid an honorarium upon completion of the questionnaire. Any additional participants that are recruited independently of BBC R&D will also receive the same honorarium.

What are the inclusion/exclusion criteria?

Participants must have experience of listening to either music or podcasts using a smartphone.

Participants must also be over the age of 18.

Will participants be paid reimbursement of expenses?	YES [ ]	NO [X]
Will participants be paid?	YES [X]	NO [ ]
If yes, please obtain signed agreement		
Will any of the participants be students?	YES [X]	NO [ ]

FIGURE A.32: Page 5 of 9 of the ethical approval form (Harrison101121) for the study in Chapter 4, approved by the University of York Electronic Engineering department ethics committee.

SECTION 5 DATA STORAGE AND TRANSMISSION

If the research will involve storing personal data, including sensitive data, on any of the following please indicate so and provide further details (answers only required if *personal* data is to be stored).

Manual files	
University computers	
Home or other personal computers	
Laptop computers, tablets	**X**
Website	

Please explain the measures in place to ensure data confidentiality, including whether encryption or other methods of anonymisation will be used.

All files containing participant's personal data will be encrypted and anonymised to ensure confidentially and will be stored on the principal investigator's University of York Google drive account. Any local copies of this data containing elements that enable the identification of participants will be destroyed to avoid a breach of confidentiality in the event of potential theft.

Please detail who will have access to the data generated by the study.

The principal investigator, Jay Harrison.

Please detail who will have control of and act as custodian for, data generated by the study.

The principal investigator, Jay Harrison.

Please explain where, and by whom, data will be analysed.

The data will be analysed on the principal investigator's personal computer by the principal investigator.

Please give details of data storage arrangements, including where data will be stored, how long for, and in what form.

Data will be encrypted and stored on the principal investigator's University of York Google drive account. Any elements of participant's personal data that enable identification (participant email addresses or identifiable information from textual responses) will be destroyed following the conclusion of the data collection section of

FIGURE A.33: Page 6 of 9 of the ethical approval form (Harrison101121) for the study in Chapter 4, approved by the University of York Electronic Engineering department ethics committee.

the study, all identifiable data has been fully anonymised and once all honorariums have been distributed to participants.

SECTION 6 CONSENT

Is written consent to be obtained?

YES	X		NO	

If yes, please attach a copy of the information for participants

If no, please justify

Will any of the participants be from one of the following vulnerable groups?

Children under 18	YES		NO	X
People with learning difficulties	YES		NO	X
People who are unconscious or severely ill	YES		NO	X
People with mental illness	YES		NO	X
NHS patients	YES		NO	X
Other vulnerable groups (if 'yes', please give details)	YES		NO	X

If so, what special arrangements have been made for getting consent?

N/A

SECTION 7 DETAILS OF INTERVENTIONS

Indicate whether the study involves procedures which:

Involve taking bodily samples	YES		NO	X
Are physically invasive	YES		NO	X
Are designed to be challenging/disturbing (physically or psychologically)	YES		NO	X

If so, please list those procedures to which participants will be exposed:

N/A

List any potential hazards: N/A

FIGURE A.34: Page 7 of 9 of the ethical approval form (Harrison101121) for the study in Chapter 4, approved by the University of York Electronic Engineering department ethics committee.

List any discomfort or distress: N/A

What steps will be taken to safeguard

(i) the confidentiality of information

The data will not be shared with any other parties, any data that is published in the results of the study will be completely anonymised and the data will be stored safely and securely to reduce the risk of a data breach.

(ii) the specimens themselves?

N/A

What particular ethical problems or considerations are raised by the proposed study?

There are no specific ethical problems or considerations related to the study, beyond the practices of protecting the confidentiality of participant's personal data.

What do you anticipate will be the output from the study? ***Tick those that apply:***

Peer-reviewed publications	X
Non-peer-reviewed publications	X
Reports for sponsor	X
Confidential reports	
Presentation at meetings	X
Press releases	
Student project	X

Is there a secrecy clause to the research? YES [ ] NO [X]
If yes, please give details below

FIGURE A.35: Page 8 of 9 of the ethical approval form (Harrison101121) for the study in Chapter 4, approved by the University of York Electronic Engineering department ethics committee.

SECTION 8 SIGNATURES

The information in this form is accurate to best of my knowledge and belief and I take full responsibility for it.

I agree to advise of any adverse or unexpected events that may occur during this project, to seek approval for any significant protocol amendments and to provide interim and final reports. I also agree to advise the Ethics Committee if the study is withdrawn or not completed.

Signature of Investigator(s): ……

……

Date: 05/11/21 ……

Responsibilities of the Principal Researcher following approval

- If changes to procedures are proposed, please notify the Ethics Committee
- Report promptly any adverse events involving risk to participants

FIGURE A.36: Page 9 of 9 of the ethical approval form (Harrison101121) for the study in Chapter 4, approved by the University of York Electronic Engineering department ethics committee.

Appendix B

Supplementary Material for Chapter 6

This appendix contains all supplementary material referenced in Chapter 6. This includes the ethical approval form (Harrison23112023) for the study that was approved by the University of York School of Physics, Engineering and Technology ethics committee, the Matlab filter bank code that was used to process the binaural soundscape noise and diffuse broadband pink noise recordings used in the production and design of hear-through condition in the experiment, the study information sheet that was provided to participants during the experiment, the experiment screening and consent google form that participants were required to complete prior to taking part in the experiment, and the four distractor tasks that participants completed during the experiment.

Physical Sciences Ethics Committee
Full Ethical Approval Form

How to Use This Form

You can make a copy of the form for editing through `File > Make a copy`. The copy then can be shared between the main investigator and any additional investigators, and collaboratively filled online. If you prefer to fill the form offline, the file can be downloaded through `File > Download` in a variety of formats including Microsoft Word (`.docx`) and OpenDocument Format (`.odt`). Finalised forms can then be exported as PDF through `File > Download` and submitted following the guidance on submission.

Notes before You Start

You need to consult the University's Code of practice and principles for good ethical governance before filling in this form. Applications that fail to comply with the Code risk being rejected or delayed through multiple rounds of reviewing and amendments.

Please seek guidance from the Departmental Ethics Officer(s) if you are uncertain about any ethical issue arising from this application.

This form is for research students and staff. Taught students should first attempt to see if their application is eligible for a fast-track process and only use this form if they are advised that they need full ethical approval.

The form needs to be filled by the main investigator of the project for which the approval is sought, in collaboration with any additional investigators.

Additional Information

You may find the following codes of ethical practice and conduct relevant to your project:

- British Psychological Society Code of Ethics and Conduct
- IEEE Code of Ethics
- British Computer Society (BCS) Code of Conduct
- Royal Academy of Eng and Eng Council Joint Statement of Ethical Principles
- Royal Society Ethics and Conduct Advice

FIGURE B.1: Page 1 of 12 of the ethical approval form (Harrison23112023) for the study in Chapter 6, approved by the University of York School of Physics, Engineering and Technology ethics committee.

Section 1. Project Information

1.1. Name and email of main investigator:

Jay Harrison (jay.harrison@york.ac.uk)

1.2. Role of main investigator:

	Undergraduate Student		Postgraduate Taught Student
x	Postgraduate Research Student		University Staff
	Other - Please specify:		

1.3. Names and appointments of additional investigators, including any collaborators from other institutions:

Prof Damian Murphy (University of York), Dr Alan Archer-Boyd (BBC R&D)

1.4. Title of the project:

The relationship between external soundscape and engagement in the consumption of speech-based media over headphones

1.5. Funding source, if any:

This research is funded in part by the UK's Arts and Humanities Research Council (XR Stories project, Grant No. AH/S002839/1) and in part by the University of York, in partnership with BBC Research and Development.

1.6. Where is the research taking place? Please specify the countries in which each step of the research, including data collection and data analysis, will take place.

All parts of the research are taking place in the United Kingdom.

1.7. Why are you applying for ethical approval (please indicate all that apply)?

FIGURE B.2: Page 2 of 12 of the ethical approval form (Harrison23112023) for the study in Chapter 6, approved by the University of York School of Physics, Engineering and Technology ethics committee.

x	It involves people.	x	It involves data from or about living human beings, whether directly collected by the project or reused from other sources.
	It involves animals.	x	It involves data protection.
	It is defence / military related.		There is a reputational risk to the University.
	It restricts academic freedom.		It involves collaboration / partnership / funding from organisations tainted by ethically questionable activities.
	Other - Please specify:		

1.8. Brief summary of the project and its key aims:

This project represents the main body of experimental work for my PhD thesis. It is a between subjects listening experiment lasting between 65-75 minutes, wherein participants will listen to audio recordings of four different stories using earphones under a variety of background noise conditions. The background noise will be played back using higher order ambisonics audio files over the spherical loudspeaker rig in the listening room at the Audiolab, with the participant positioned at the centre of the loudspeaker rig.

The key aims of the project are as follows:

- To conduct a listening experiment that will investigate the often cited phenomenon of mobile audio consumers inhabiting a "*personal listening bubble*".
- To answer the research question: how do different signal-to-noise ratios of background soundscape influence immersion and absorption when consuming spoken word based-media over headphones?
- To answer the research question: how do transfer functions of the hear-through modes in wireless earbuds influence the listening experience when consuming spoken word-based media over headphones in the presence of background noise?
- Generate experimental data that will be analysed and the results written up for internal and external research publication outputs.

Data will be collected using a questionnaire presented to participants in the form of a Matlab app with a graphical user interface. Some questions will also use printed hard-copy response sheets. It is not possible/impractical to provide a working copy of the Matlab app, therefore a excel spreadsheet will be provided with this approval form containing all of the questions and the response format the participants will be asked to answer.

FIGURE B.3: Page 3 of 12 of the ethical approval form (Harrison23112023) for the study in Chapter 6, approved by the University of York School of Physics, Engineering and Technology ethics committee.

1.9. External Ethics Approval: Has this work been ethically approved by another external agency or collaborating institution such as NHS? If yes, please provide details.

No (N/A)

1.10. Have you read and understood the University's University's Code of practice and principles for good ethical governance?

Yes

Section 2. Human Participants

If your project doesn't involve people, please skip to Section 3.

You must include an information sheet (containing a privacy notice) and consent form alongside this application form (or give a reason that they aren't needed). If a survey / questionnaire is being conducted please include this alongside the application. If you are conducting an online survey these can be embedded at the front of the survey and need to be included in this application.

These documents can be in combined or separate forms, depending on the complexity of the ethical and legal issues pertaining to the project, at the discretion of the project investigators. In projects with low complexity, the information sheet, privacy notice, and consent form can all be combined in one document, ensuring that it covers the material in the provided guidance checklist. In projects with medium complexity, these can be in two documents: an Information sheet containing the privacy notice for which a template can be used, and a separate consent form. In projects with high complexity, these can be in three separate documents: an information sheet, a separate privacy notice, plus a separate consent form. In this latter case, a generic privacy notice can be used, if deemed appropriate.

2.1. Who will your participants be? Describe the criteria for inclusion / exclusion. Outline the approximate number of participants you are aiming for and explain how the sample size has been determined.

The participants will be any individuals aged 18 or over who do not have any diagnosed hearing loss. There are no other criteria for inclusion/exclusion. We are ultimately aiming to recruit over 60 (ideally up to 80) participants to take part in the study. This sample size has been determined based on references from similar studies in the literature.

FIGURE B.4: Page 4 of 12 of the ethical approval form (Harrison23112023) for the study in Chapter 6, approved by the University of York School of Physics, Engineering and Technology ethics committee.

2.2. Will they be paid? If yes, how much? If your participants are paid, you must obtain a signed receipt of payment.

Each participant will be renumerated with a £20.00 electronic e-commerce voucher.

2.3. Does the research involve participants who are particularly vulnerable or may lack capacity to give informed consent to take part (or to keep taking part) in the research?

No.	**Item**	**Yes**	**No**
a	Children (under 18)		x
b	• If yes, does the investigator have a current DBS check?		
c	People with learning disabilities		x
d	The frail, elderly, those who are unconscious, severely ill, have a mental illness, or are patients		x
e	Anyone who may be easily coerced due to a lack of capacity or an imbalance of power (e.g. UoY students or staff)		x
f	Any other vulnerable groups – specify below:		x
g	If you answered 'yes' to any of the above, explain and justify the procedure and explain the steps taken to safeguard individuals: N/A		

2.4. Recruitment: How will you recruit participants?

Participants will be recruited via multiple means including, online departmental email lists, physical posters, and flyering.

2.5. How will you guarantee anonymity? This includes IP addresses and any identifying information. If anonymity will not be provided, explain why this is necessary.

The last three digits of participants' phone numbers and last three characters of their post codes will be used as anonymised unique identifiers for participants during data

FIGURE B.5: Page 5 of 12 of the ethical approval form (Harrison23112023) for the study in Chapter 6, approved by the University of York School of Physics, Engineering and Technology ethics committee.

collection. No other identifying information will be collected. Participants' email addresses and names will be briefly stored in the principal investigator's secure university email and storage accounts when experiment appointment scheduling responses and consent form responses are received, and voucher codes are sent out. However, this data will be deleted at the earliest possible opportunity once the associated activity has been completed.

2.6. Does any of the following apply?

No.	Item	Yes	No
a	The research is designed to be emotive or aversive.		x
b	The research involves taking bodily samples.		x
c	The research is physically invasive / challenging.		x
d	If you answered 'yes' to any of the above, explain and justify the procedure and explain the steps taken to safeguard individuals: N/A		

Section 3. Data Protection

All personal data must be collected and used in accordance with the UK General Data Protection Regulation (UK GDPR) 2018, the UK Data Protection Act (DPA) 2018, the University's Data Protection Policy and the University's Research Data Management Policy.

Personal data which have undergone pseudonymisation (e.g. replacing names or other identifiers which are easily attributed to individuals with a code) will still remain personal data and within the scope of the UK data protection law (particularly while the code can be tied back to the individual).

3.1. Declarations:

No.	Item	Yes	No
a	I have read and understood the University's Data Protection Policy.	x	
b	I have read and understood the University's Research Data Management Policy.	x	

FIGURE B.6: Page 6 of 12 of the ethical approval form (Harrison23112023) for the study in Chapter 6, approved by the University of York School of Physics, Engineering and Technology ethics committee.

3.2. What types of personal data will you be processing? Indicate all that apply. For further guidance, see University's guide to Data Protection.

x	Personal data		Special category data

3.3. Describe the nature of the personal or special category data you will be collecting or using:

Name, Email address

3.4. Set out the different groups of people whose personal data will be collected or processed (e.g. patient, care-giver, game player, school children, employee, member of the public) and give an approximation of the number of individuals involved:

University students, staff, and members of the public. Between 60 and 80 in total.

3.5. Where is the personal data sourced from (e.g. direct from the participant, from a (named) organisation/source, indirectly from observation, national data registry, from social media, etc.)? If the data is from NHS Digital, a registry (e.g. Eurostat) or organisation, give the identifiers for the datasets and/or reference the sharing agreements.

Direct from the participant.

3.6. Data Protection by Design and Default:

No.	**Item**	**Yes**	**No**
a	Will you be collecting the minimum amount of personal data necessary for the specified research purpose, e.g. gathering anonymised data at source if personally identifiable information is not needed, and ensuring all data to be captured can be justified?	x	
b	Will you use the data only for the purposes of this research project, or if you plan to use the data for additional purposes, will you bring this to the attention of the research participants at point of data collection or, where this is not possible, the University's Data Protection Officer?	x	
c	Will you anonymise personal data wherever and as soon as possible: either at point of data capture, collation, analysis or output?	x	

FIGURE B.7: Page 7 of 12 of the ethical approval form (Harrison23112023) for the study in Chapter 6, approved by the University of York School of Physics, Engineering and Technology ethics committee.

d	Will you use pseudonymised data wherever possible in cases where information cannot be anonymised, e.g. will you separate research participant contact details from the data to be analysed and/or remove identifiers, e.g. specific date of birth and replace with age within a date range?	x	
e	Will you keep any data appropriately secure, in University's secure digital or physical storage facilities, maintaining confidentiality and anonymity (e.g. identifiers will be encoded and the code available on a need to know basis) where possible?	x	
f	If you answered 'no' to any of the above, explain why the deviation is necessary: N/A		

3.7. Please describe the special precautions will you take to ensure anonymity when linking identifiable data to experimental data:

Anonymised unique identifiers are being used for this purpose consisting of the last three digits of the participants' phone numbers and the last three characters of their post codes.

3.8. If your data can be traced to identifiable participants/computers/addresses: a) who will be able to access your data? b) how long will you need to keep it in this identifiable format?

The only data that this applies to in my study (names and email addresses) will not be stored together with the experimental data and will be deleted within one week of the participant completing the study, once appointment scheduling, consent confirmation, and voucher allocation activities have been completed.

3.9. If data subjects are identifiable in the final research output or data publication, e.g. publication of direct quotations from respondents, publication of data that might allow the identification of individuals, please explain why this is necessary and how informed consent from the individuals is ensured.

The data subjects will not be identifiable.

3.10. Explain your personal and special category data retention, disposal, and deposit schedule and plans in accordance with articles 3.5 – 3.9 of the University's Research Data Management Policy. This must include a) how long, in what form, and for what purposes personal data will be kept, b) if, when, by whom, and how any data will be destroyed, c) and

FIGURE B.8: Page 8 of 12 of the ethical approval form (Harrison23112023) for the study in Chapter 6, approved by the University of York School of Physics, Engineering and Technology ethics committee.

if any personal or special category data (i.e. data that is not truly and irrevocably anonymised) be deposited in an archive or external repository.

The only data that this applies to in my study (names and email addresses) will not be stored together with the experimental data and will be stored on a secure university email server and then deleted at the earliest opportunity once appointment scheduling, consent confirmation, and voucher allocation activities have been completed.

3.11. Data Security:

No.	Item	Yes	No
a	Will you use University approved software?	x	
b	Will you use University approved file storage (Google Drive, University networked storage, research computing)?	x	
c	Will you store personal or confidential data on devices with appropriate device encryption and ensure confidential information is encrypted before it is transmitted/shared digitally?	x	
d	If capturing audio, will you use an encrypted device for recording (e.g. an Apple iOS device or encrypted voice recorder)?	x	
e	Where data is held on an encrypted portable device (e.g. laptop, tablet) will you back it up to a University approved service as soon as possible and perform periodic checks to ensure data is being backed up appropriately?	x	
f	Will you ensure the device(s), accounts, or storage area(s) used to store data are not accessible to any unauthorised parties?	x	
g	If you answered 'no' to any of the above, explain why it is not possible to do so and what alternative arrangements will be put in place to provide sufficient data security: N/A		

3.12. Detail what other protections will be used for digital data (e.g. access/edit permissions, procedural safeguards re downloads/making copies, remote access via VDS/VPN, 2 factor authentication) or any hard-copy data (e.g. data on paper stored in lockable filing cabinet or container, and/or a locked room in secure premises)?

All digital data will be kept only on secure university storage, some hard-copy data will

FIGURE B.9: Page 9 of 12 of the ethical approval form (Harrison23112023) for the study in Chapter 6, approved by the University of York School of Physics, Engineering and Technology ethics committee.

also be kept in secure locked storage on University premises. An anonymised unique identifier (described above) will be used to link the digital and hard-copy data before the hard-copy data is digitised. Once it is digitised the hard-copies will then be destroyed.

3.13. Confirm you have reviewed the user commitments under the Policy for the safe use of University information on devices. Detail anything in the user commitments that will pose a challenge in carrying out your proposed research.

A non-university managed device will be used in the scheduling of participants' experiment appointments, consent confirmation, and the allocation of vouchers, for which participants' names and email address will be temporarily stored on secure online university email and storage accounts. As a non-managed device is being used it is therefore especially important to ensure that additional measures are put in place, such as ensuring that auto sign in for these accounts is turned off, and it is especially important that the personal data is deleted at the earliest opportunity once the associated activities (appointment scheduling, consent confirmation, and voucher allocation) have been completed.

3.14. Consult the University's guidance on Data Protection Impact Assessments (DPIA) and explain your decision as to whether a DPIA is required. If a DPIA is deemed to be required, provide the DPIA (or a draft thereof) along with this application and provide details on the finalised or ongoing process of putting the DPIA in place.

I completed the DPIA screening questionnaire and was informed that I do not need to carry out a DPIA for my project or activity.

Section 4. Risk Assessment

4.1. Has a departmental risk assessment been completed for this project, if appropriate? If no, why not?

No departmental risk assessment has not been carried out for this specific project, however the study does not involve any activities (and hence risks) that have not already been assessed and deemed acceptable by existing Audiolab departmental risk assessments that cover routine day to day research activities (such as listening to audio using headphones and loudspeakers) that take place in the Audiolab's facilities. However, as the principal investigator, my personal risk assessment of the project and the activities that the participant will engage in is as follows:

FIGURE B.10: Page 10 of 12 of the ethical approval form (Harrison23112023) for the study in Chapter 6, approved by the University of York School of Physics, Engineering and Technology ethics committee.

Participation in this study is low risk overall with the following specific considerations.

- There is low level risk of noise exposure while listening to the audio during the experiment. To mitigate against this risk the audio levels that the participant will experience in the experiment have been carefully calibrated and are well within the range of safe and comfortable listening levels according to established literature in the field. It should be noted that it is a key requirement of the experimental methodology that the levels used for the playback of audio over headphones and loudspeakers is kept fixed to predetermined levels.
- While wearing headphones, the participant may find it more difficult to hear external sounds, so the participant will be reminded to remain still as much as possible during the playback of each story experience over the earphones.
- To minimise the risk of fatigue, the participant will also be reminded that they are permitted to take a break as needed between each story experience. The participant will also be reminded that their participation in this experiment is voluntary and they may pause the experiment to take a rest or terminate the experiment at any time.

As stated above the audio levels that will be used in the experiment are well within the range of safe and comfortable listening levels stated in the literature, moreover the calibration of these levels has been overseen by the AudioLab technician. Given these considerations it is not thought necessary that a specific departmental risk assessment be carried out for this project.

Section 5. Academic Freedom

5.1. Is there a secrecy clause to the research? If yes, give details:

No.

Section 6. Reputational Risk

6.1. Why is it appropriate for the University to be associated with this project? Please also state what action has been taken to mitigate against potential reputational risk. If any reputational risk is associated with a collaborative partner, please cover in Section 7.

There is no reputational risk associated with this planned study as detailed above in this document, therefore it is appropriate that the University is associated with this project.

FIGURE B.11: Page 11 of 12 of the ethical approval form (Harrison23112023) for the study in Chapter 6, approved by the University of York School of Physics, Engineering and Technology ethics committee.

Section 7. Collaboration with Questionable Ethical Standards or Activities

7.1. Explain the nature of the collaboration / partnership and what about the organisation is ethically questionable:

N/A

7.2. Why is it appropriate for the University to be associated with this organisation? Please also state what action has been taken to mitigate against potential reputational risk.

N/A

Section 8. Declaration

8.1. The project team have read and understood this application:

Signed (Main Investigators) Jay Harrison and Damian Murphy	
Date 21/11/23	

FIGURE B.12: Page 12 of 12 of the ethical approval form (Harrison23112023) for the study in Chapter 6, approved by the University of York School of Physics, Engineering and Technology ethics committee.

LISTING B.1: Matlab filter bank code that was used to process the binaural soundscape noise and diffuse broadband pink noise recordings used in the production and design of *hear-through* condition in Chapter 6. After each filter bank correction was applied a set of three calibration recordings were made with KEMAR wearing the Shure SE215 earphones. In these recordings 52 dB SPL of broadband pink noise was played back over all 50 channels of the speaker rig in addition to a binaural recording of the speaker rig pink noise that was made with the DPA binaural microphones (shown below in Figure 6.4) following the process described in section 6.1 of Chapter 6. The binaural recording was processed through the Matlab filter bank and played back over the Shure SE215 earphones fitted in KEMAR. Third-octave RMS levels were calculated for each of the three calibration recordings and averaged. This averaged set of third octave RMS values were compared against the corresponding set of values for the "Hear-through Measurement Target" transfer function (shown in Figure 6.2 of Chapter 6). Subsequent adjustments were then made to the correction values to get closer to the target transfer function. This process was repeated iteratively until the final *hear-through* condition transfer function (shown in Figure 6.2 of Chapter 6) was achieved. The final iteration of the filter bank was then used to process the binaural soundscape noise recording that was played back over the earphones in the *hear-through* condition.

```
%%%%% Set the directory where the source audio file due to be be filtered is located
source_audio_data_dir = '/Users/example_user/example_source_audio_directory_folder/';

%%%%% Read the source audio file from the directory
[source_audio_data, Fs] = audioread(fullfile(source_audio_data_dir, "example_source_audio_filename.
    wav"));

%%%%% Amount in dB the source audio will be lowered to avoid clipping after boosting
mod = -23;

%%%%%% Adjust the dB level of the source audio by the amount specified
source_audio_data = source_audio_data * 10^(mod/20);

%%%%% Code that was used in the iterative calibration of filter bank gain
%%%%% adjustment values (dB) required to reach the 'Hear-through measurement
%%%%% target' shown in the paper.
Sim_Correction_subset = [0, 0, 0, 0, 0, 0, 0, 4.39605408782001, 8.23163742597313, 9.14249146594283,
    15.3968195040639, 13.4908160408147, 15.4186643891177, 16.6406020531726, 24.9936911365877,
    32.1689229971965, 29.9992467559891, 23.9116978797674, 17.4044683109700, 13.4564661007073,
    16.8829787846110, 24.9110396659511, 20.8879595584894, 11.9874036120996, 7.52162601240115,
    4.08503959828286];

```

```
Sim_Correction_full = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 4.39605408782001, 8.23163742597313,
    9.14249146594283, 15.3968195040639, 13.4908160408147, 15.4186643891177, 16.6406020531726,
    24.9936911365877, 32.1689229971965, 29.9992467559891, 23.9116978797674, 17.4044683109700,
    13.4564661007073, 16.8829787846110, 24.9110396659511, 20.8879595584894, 11.9874036120996,
    7.52162601240115, 4.08503959828286];

Sim_Second_Correction_Factor = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 3.00306196, 7.227672802,
    4.720211224, 7.547893555, 6.609875033, 6.830536758, 6.051910214, 5.032204833, 3.929881634,
    4.082209897, 5.163147686, 8.120509193, 4.977513209, 11.20140072, 14.23617912, 6.535784061,
    2.555981457, 7.840466979, 4.586465439];

Sim_Second_Correction_full = Sim_Correction_full+Sim_Second_Correction_Factor;

Sim_Third_Correction_Factor = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0.280510321, 4.269516773,
    2.256759607, 4.431761194, 3.542129885, 3.384274482, 2.227846339, 0.70243553, 0.054633332,
    0.383113616, 1.988565461, 5.385156937, 0.344698757, 3.274868512, 2.775818707, -2.686891477,
    -4.936839157, 4.980507819, 3.866207393];

Sim_Third_Correction_full = Sim_Second_Correction_full+Sim_Third_Correction_Factor;

Sim_Fourth_Correction_Factor = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, -0.769432205, 2.527470193,
    0.274423423, 2.160636974, 1.932201267, 1.458840879, 0.38840192, -0.89092736, -1.26896596,
    -1.021164955, -0.150629242, 2.159018682, -2.835648071, -0.297841227, -0.639145922, -5.538566797,
     -7.944555923, 3.382162478, 3.557152254];

Sim_Fourth_Correction_full = Sim_Third_Correction_full+Sim_Fourth_Correction_Factor;

Sim_Fifth_Correction_Factor = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 4, 0, 0, 0, 0, 0, 0, 0, 0, 0,
    1.5, -1.5, 0, 0, -8, -11, 5, 5];

Sim_Fifth_Correction_full = Sim_Fourth_Correction_full+Sim_Fifth_Correction_Factor;

Sim_Sixth_Correction_Factor = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 4, 0, 0, 0, 0, 0, 0, 0, 0, 0,
    0, -4, 0, 0, -5, 9.5, 6, 4];

Sim_Sixth_Correction_full = Sim_Fifth_Correction_full+Sim_Sixth_Correction_Factor;

gain_values_db_full = Sim_Sixth_Correction_full;

%%%%% Convert the gain values from dB to a linear scale
gain_values_linear = 10.^(gain_values_db_full / 20);

%%%%% Set the 1/3rd octave frequency values for the filter bank
F = [20 25 31.5 40 50 63 80 100 125 160 200 250 315 400 500 630 800 1000 1250, ...
1600 2000 2500 3150 4000 5000 6350 8000 10000 12500 16000 20000];

third_octave_centers = F;
```

```
50
51 low_freq = [17.8, 22.4, 28.2, 35.5, 44.7, 56.2, 70.8, 89.1, 112, 141, 178, 224, 282, 355, 447, 562,
       708, 891, 1122, 1413, 1778, 2239, 2818, 3548, 4467, 5623, 7079, 8913, 11220, 14130, 17780];
52
53 high_freq = [22.4, 28.2, 35.5, 44.7, 56.2, 70.8, 89.1, 112, 141, 178, 224, 282, 355, 447, 562, 708,
       891, 1122, 1413, 1778, 2239, 2818, 3548, 4467, 5623, 7079, 8913, 11220, 14130, 17780, 22390];
54
55 %%%%% Set order of filter
56 order = 1;
57
58 %%%%% Create the filter bank
59 filter_bank = {};
60
61 for i = 1:length(third_octave_centers)
62     [b, a] = butter(order, [low_freq(i), high_freq(i)]/(Fs/2), 'bandpass');
63     filter_bank{i} = {b, a};
64 end
65
66 %%%%% Apply the filter to the source audio
67 filtered_signals = cell(1, length(filter_bank));
68
69 for i = 1:length(filter_bank)
70     b = filter_bank{i}{1};
71     a = filter_bank{i}{2};
72     filtered_signal = filter(b, a, source_audio_data);
73     boosted_signal = filtered_signal * gain_values_linear(i);
74     filtered_signals{i} = boosted_signal;
75 end
76
77 final_signal = zeros(size(source_audio_data));
78
79 for i = 1:length(filtered_signals)
80     final_signal = final_signal + filtered_signals{i};
81 end
82
83 %%%%%% Set the directory where the filtered audio will be written to
84 filtered_audio_data_dir = '/Users/example_user/example_filtered_audio_directory_folder/';
85
86 %%%% Write the filtered audio to the directory
87 audiowrite(fullfile(filtered_audio_data_dir, 'example_filtered_audio_filename.wav'), final_signal,
       Fs, 'BitsPerSample', 24);
```

Study information sheet

Please read the following information carefully before providing your consent to take part in the study.

This study has been approved by the Physical Sciences Ethics Committee of the University of York. Study approval number: Harrison23112023

Background

Studies have shown that individuals can experience *immersion* when consuming various audio/audio-visual media content. In the context of this study, *immersion*, also known as deep mental involvement, can be described as being mentally lost (absorbed) in the experience. *Immersion* is encountered when the experience is involving and absorbs you mentally by capturing your attention. For example, immersion may be experienced when reading a book, playing video games, watching a movie, etc.

Individuals often use mobile devices and headphones to consume audio-based media content in a variety of environmental, situational, and social contexts. There has been relatively little research carried out to investigate how the level of immersion an individual experiences when listening to audio-based media might relate to their surrounding acoustic conditions.

What is the purpose of the study?

The purpose of this study is to investigate how the level of immersion an individual experiences when listening to audio recordings of stories over earphones might relate to the surrounding acoustic conditions in which the experience is consumed. It will also investigate the influence of an active headphone technology feature as a mediating factor in this relationship. Results from this experiment may be published in the form of a peer-reviewed article.

Why have I been invited to take part?

You have been invited because you satisfy the inclusion criteria of being aged 18 or over and not having any diagnosed hearing loss.

Do I have to take part?

No, participation is optional. If you decide to take part, you may request a copy of this information sheet for your records and you will be asked to complete a participant consent form. If you change your mind at any point during the study, you will be able to withdraw your participation without being required to provide a reason. Any data captured prior to your withdrawal from the study may still be used.

FIGURE B.13: Page 1 of 4 of the study information sheet provided to participants in the experiment described in Chapter 6.

What will I have to do?

You will be asked to take part in a single listening experiment consisting of a training block and four subsequent experiment blocks. Each block will involve listening to a story and then answered a series of questions based on your listening experience. You will also be required to undertake in a distractor task between each block. After the final experiment block there will be a short exit questionnaire. There are no right or wrong answers to any of the questions in the experiment.

The experiment is expected to take between 75-90 minutes to complete, approximately 25 minutes of which will be spent listening to the five stories.

You will receive ***£20.00 in Amazon vouchers*** as renumeration for your participant in the study.

How will I listen to audio during the experiment?

The listening experiment will involve playback of audio over a pair of Shure earphones that are worn inside the ears. The tips of these earphones have been fully sanitised before your use. The listening experiment may also involve playback of audio over loudspeakers.

Are there any risks involved?

Participation in this study is low risk overall with the following specific considerations.

- There is low level risk of noise exposure while listening to the audio during the experiment. To mitigate against this risk the audio levels that you will experience in the experiment have been carefully calibrated and are well within the range of safe and comfortable listening levels according to established literature in the field. While wearing headphones, you may find it more difficult to hear external sounds, so as much as possible please remain still during the playback of each story experience.
- To minimise the risk of fatigue, please take a break as needed between each story experience.

You are reminded that your participation in this experiment is voluntary and you may pause the experiment to take a rest or terminate the experiment at any time.

Data protection (Privacy notice)

On what basis will you process my data?

Under the general data protection regulation (GDPR), the University has to identify a legal basis for processing personal data and, where appropriate, an additional condition for processing special category data.

FIGURE B.14: Page 2 of 4 of the study information sheet provided to participants in the experiment described in Chapter 6.

In line with our charter which states that we advance learning and knowledge by teaching and research, the University processes personal data for research purposes under Article 6 (1) (e) of the GDPR:

Processing is necessary for the performance of a task carried out in the public interest

Special category data is processed under Article (9) (j):

Processing is necessary for archiving purposes in the public interest, or scientific and historical research purposes or statistical purposes

Research will only be undertaken where ethical approval has been obtained, where there is a clear public interest, and where appropriate safeguards have been put in place to protect data.

In line with ethical expectations and in order to comply with common law duty of confidentiality, we will seek your consent to participate where appropriate. This consent will not, however, be our legal basis for processing your data under the GDPR.

How will you use my data?

Listening experiment responses may be used in internal or external research publications and/or to inform the direction of future research in the associated PhD project.

Will you share my data with third parties?

No personal data you provide will be shared with third parties.

Will I be identified in any research outputs?

No, you will not be identified within any research outputs.

How long will you keep my data?

Experimental data including no identifiable information will be retained in line with legal requirements and retention timeframes will be determined in line with the University's Records Retention Schedule. Personal data (the participant's name and email address) will only be retained long enough to facilitate experiment appointment scheduling and voucher allocation, after which this data will be deleted at the earliest opportunity.

What rights do I have in relation to my data?

Under the GDPR, you have a general right of access to your data, a right to rectification, erasure, restriction, objection or portability. You also have a right to withdraw. Please note, not all rights apply where data is processed purely for research processes.

Questions or concerns

FIGURE B.15: Page 3 of 4 of the study information sheet provided to participants in the experiment described in Chapter 6.

If you have any questions or concerns about this participant information sheets or concerns about how your data is being processed, please contact the principal investigator Jay Harrison (jay.harrison@york.ac.uk) in the first instance. If you are still dissatisfied, please contact the University's Data Protection Officer at dataprotection@york.ac.uk.

Right to complain

If you are unhappy with the way in which the University has handled your personal data, you have a right to complain to the information commissioner's office. For information on reporting a concern to the information commissioner's office please visit www.ico.org.uk/concerns.

FIGURE B.16: Page 4 of 4 of the study information sheet provided to participants in the experiment described in Chapter 6.

Listening experiment screening questionnaire and consent form

This is a screening questionnaire and consent form that you are required to complete in order to take part in the associated listening experiment study. Please respond to the following questions. If you have any questions about this form please ask the researcher who will be able to assist you.

This study is being conducted by the principal investigator, Jay Harrison (jay.harrison@york.ac.uk) as part of an ongoing PhD project based in the AudioLab at the University of York. If you have any questions or wish to withdraw your consent from the study, please contact the principal investigator by email.

Thank you

* Indicates required question

Participant information

1. Please provide your full name in the following text box *

2. Please provide your email address in the following text box (this will be used to issue your voucher codes upon completion of the experiment) *

Screening questionnaire

3. Please check the following box to confirm that you are aged 18 or over *

Tick all that apply.

☐ I confirm I am aged 18 or over

FIGURE B.17: Page 1 of 2 of the experiment screening and consent google form that participants were required to complete before taking part in the experiment described in Chapter 6.

4. Please check the following box to confirm you do not have any diagnosed hearing loss *

Tick all that apply.

☐ I confirm that I do not have any diagnosed hearing loss

Consent form

5. Please check the following box to confirm that you have read and understood the accompanying study information sheet *

Tick all that apply.

☐ I have read and understood the accompanying study information sheet

6. Please check the following box to consent to your participation in this study under the terms outlined in the provided information sheet *

Tick all that apply.

☐ I confirm that I consent to my participation in the study

This content is neither created nor endorsed by Google.

Google Forms

FIGURE B.18: Page 2 of 2 of the experiment screening and consent google form that participants were required to complete before taking part in the experiment described in Chapter 6.

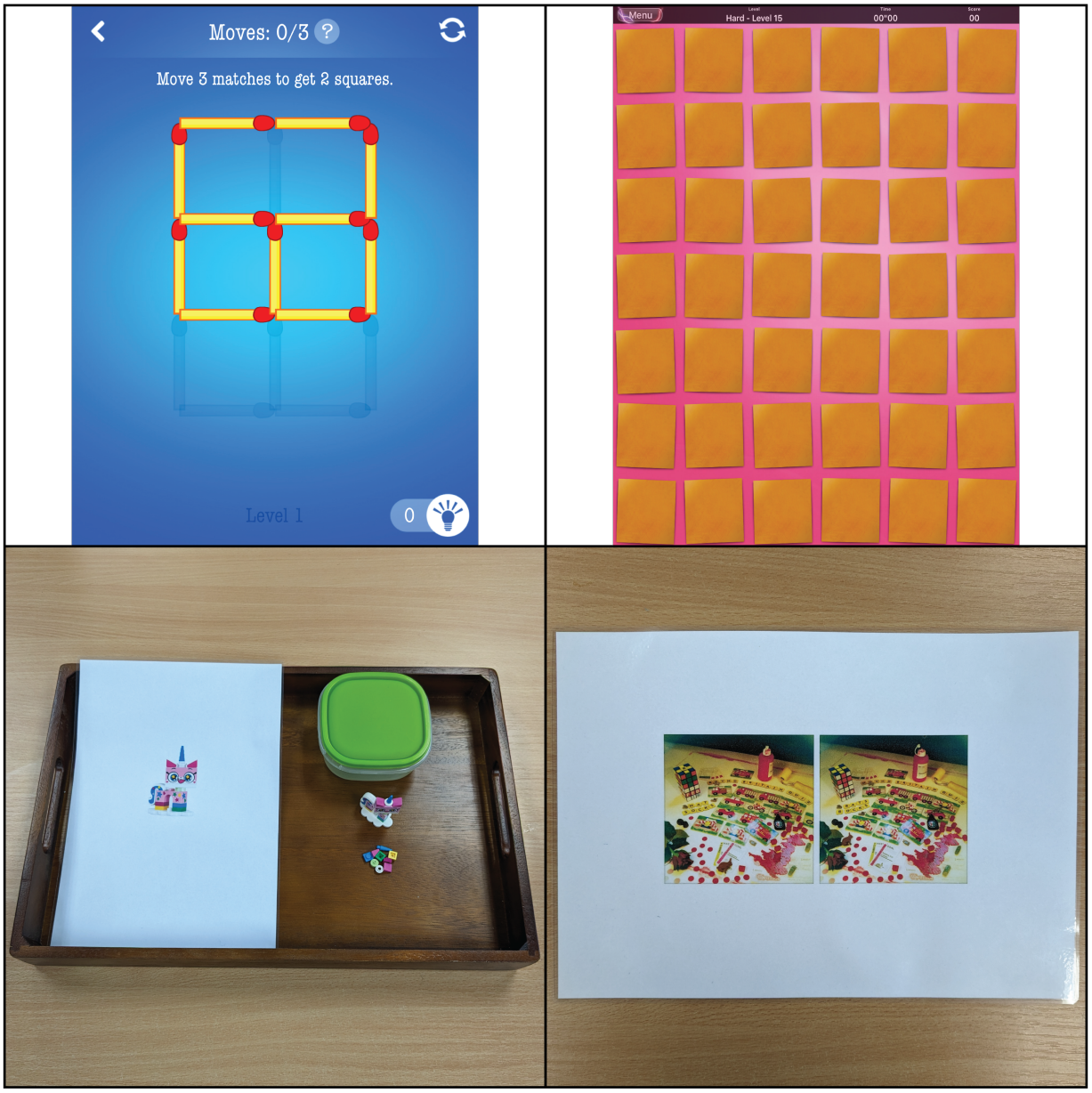


FIGURE B.19: The four distractor tasks participants completed between stimuli in the experiment described in Chapter 6. Participants spent 4 minutes on each distractor task and were informed that their assessment in the tasks was not being assessed. Top left: "Matchsticks ˜ Matches Puzzle" Participants were handed an ipad with level 1 from the second pack of puzzles loaded on the screen and asked to complete as many puzzles as they were able in the time allowed. Top right: "Memorama • Classic" Participants were handed an ipad with level 15 loaded on the screen and asked to complete as many levels as they were able in the time allowed. Bottom left: "LEGO Set 41775-1 Rainbow Unikitty" Participants were handed a wooden tray with a box containing the disassembled parts of the LEGO set. They were shown a printed image of the completed set for 5 seconds and asked to try and build the set from memory in the time allowed. They were told to alert the researcher if they thought they had completed the set, who would then verify if they had built it correctly. Bottom right: "Spot the Difference" Participants were handed a blank A4 notebook and a A4 printout of the creative commons spot the difference image from: https://flic.kr/p/dVQwna. They were asked to find all of the differences between the two images and write down their answers on the notepad. They were told to alert the researcher if they thought they had found all of the differences.

Bibliography

[1] D. Howard and J. Angus, *Acoustics and Psychoacoustics Focal Press*. Amsterdam: Elsevier, 2009.

[2] F. A. Everest, *Master handbook of acoustics*. McGraw-Hill Education, 2022.

[3] J. Granados, M. Hopper, and J. He, "A usability and safety study of Bone-conduction headphones during driving while listening to audiobooks," in *Proceedings of the Human Factors and Ergonomics Society*, vol. 2, no. September 2018, 2018.

[4] M. Vorländer, *Auralization Fundamentals of Acoustic, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. Springer, 2020.

[5] F. Stevens, "Strategies for Environmental Sound Measurement , Modelling , and Evaluation," Ph.D. dissertation, University of York, 2018, [Online] Available at: https://etheses.whiterose.ac.uk/22661/ [Accessed: 21/11/2024].

[6] S. Harriet, "Application of Auralisation and Soundscape Methodologies to Environmental Noise," Ph.D. dissertation, University of York, 2013, [Online] Available at: https://etheses.whiterose.ac.uk/4724/ [Accessed: 21/11/2024].

[7] International Telecommunication Union, "ITU-T. Recommendation P.57, TELEPHONE TRANSMISSION QUALITY, TELEPHONE INSTALLATIONS, LOCAL LINE NETWORKS. Objective measuring apparatus. Artificial ears." Tech. Rep., 2021.

[8] Bukowskis, "a pair of "AKG 1000" headphones, Austria," 2017, [Online] Available at: https://www.bukowskis.com/en/lots/995658-a-pair-of-akg-1000-headphones-austria [Accessed: 28/07/2024].

[9] Dezeen, "Bose launches frames sunglasses that are also headphones," 2018, [Online] Available at: https://www.dezeen.com/2018/12/17/bose-sunglasses-frames-headphones/ [Accessed: 28/07/2024].

[10] R. Gupta, R. Ranjan, J. He, G. W. Seng, and S. Peksi, "Acoustic transparency in hearables for augmented reality audio: Hear-through techniques review and challenges," in *Proceedings of the 3rd AES International Conference on Audio for Virtual and Augmented Reality*, Online, 2020.

[11] Hifi Archiv, "Brandes 1925 76," 2024, [Online] Available at: https://www.hifi-archiv.info/Radio-Werbung/Brandes/index.html [Accessed: 28/07/2024].

[12] Collectors Weekly, "Regency Mandarin Red TR-4 Transistor Radio," 2011, [Online] Available at: https://www.collectorsweekly.com/stories/41941-regency-mandarin-red-tr-4-transistor-rad [Accessed: 28/07/2024].

[13] Smithsonian Institution, "Regency Model TR-1 Transistor Radio," 2024, [Online] Available at: https://americanhistory.si.edu/collections/object/nmah_713528 [Accessed: 21/21/2024].

[14] Reverb, "Sony tps-l2 walkman portable cassette player (1979 - 1980)," 2024, [Online] Available at: https://reverb.com/en-fi/p/sony-tps-l2-portable-cassette-player [Accessed: 29/07/2024].

[15] M. Berwyck, "What was the first CD album you owned?" 2019, [Online] Available at: https://www.quora.com/What-was-the-first-CD-album-you-owned [Accessed: 29/07/2024].

[16] Sony Group Corporation, "Corporate history," 2024, [Online] Available at: https://www.sony.com/en/SonyInfo/CorporateInfo/History/company/ [Accessed: 29/07/2024].

[17] D. Grabham, "The MP3 player is a decade old," 2008, [Online] Available at: https://www.techradar.com/news/portable-devices/mp3-players/the-mp3-player-is-a-decade-old-261622 [Accessed: 29/07/2024].

[18] Elite Obsolete Electronics, "Refurbished Apple iPod Original 1st Generation 2001 5GB 10GB M8541 M85 — eoe.works," 2024, [Online] Available at: https://eoe.works/products/refurbished-apple-ipod-5gb-original-1st-generation-2001-new-battery [Accessed: 29/07/2024].

[19] R. Fernandez, "iPhone (1st generation) - Wikipedia — en.wikipedia.org," 2024, [Online] Available at: https://en.wikipedia.org/wiki/IPhone_%281st_generation%29 [Accessed: 29/07/2024].

[20] RETROSPEKT, "Sony Walkman MDR-3L2 Original TPS-L2 On-Ear Headphones — retrospekt.com," 2024, [Online] Available at: https://retrospekt.com/products/sony-walkman-tps-l2-original-headphones [Accessed: 29/07/2024].

[21] Elite Obsolete Electronics, "Original 2001 apple earbuds headphones for ipod 3.5mm auxiliary 603-6261 new & used," 2024, [Online] Available at: https://eoe.works/products/new-original-2001-apple-earbuds-headphones-for-ipod-3-5mm-auxiliary-603-6261 [Accessed: 29/07/2024].

[22] Consumer Reports, "Apple AirPods (1st Gen) Headphone Review - Consumer Reports," 2024, [Online] Available at: https://www.consumerreports.org/electronics-computers/headphones/apple-airpods-1st-gen/m387635/ [Accessed: 29/07/2024].

[23] Amazon, "Apple airpods pro (1. generation) with magsafe charging case (2021)," 2024, [Online] Available at: https://www.amazon.co.uk/Apple-AirPods-MagSafe-charging-case/dp/B09JQZ5DYM?th=1 [Accessed: 29/07/2024].

[24] Galen Carol Audio, "DECIBEL (LOUDNESS) COMPARISON CHART," 2024, [Online] Available at: https://www.gcaudio.com/tips-tricks/decibel-loudness-comparison-chart/ [Accessed: 24/06/2024].

[25] M. Stojanovic and P. P. J. Beaujean, *Acoustic communication*. New Jersey: Ablex Publishing Corporation, 1984.

[26] the DESIGN MUSEUM, "Sony walkman," 2024, [Online] Available at: https://designmuseum.org/discover-design/all-design-objects/sony-walkman [Accessed: 21/11/2024].

[27] Rajar Audio Measurement, "MIDAS Measurement of Internet Delivered Audio Services," 2023, [Online] Available at: https://www.rajar.co.uk/docs/news/MIDAS_Spring_2023.pdf [Accessed: 21/11/2024].

[28] Edison Research, "The UK Podcast Consumer 2024," Edison Research, Tech. Rep., 2024.

[29] S. Hosokawa, "The walkman effect," *Popular Music*, vol. 4, pp. 165–180, 1984.

[30] C. Hardman, "Walkmanology," *The Drama Review: TDR*, vol. 27, no. 4, pp. 43–46, 1983.

[31] S. Arkette, "Sounds Like City," *Theory, Culture & Society*, vol. 21, no. 1, pp. 159–168, 2004.

[32] D. Beer, "Tune out: Music, soundscapes and the urban mise-en-scéne," *Information Communication and Society*, vol. 10, no. 6, pp. 846–866, 2007.

[33] M. Bull, "iPod: a Personalized Sound World for its Consumers," *Comunicar*, vol. 17, no. 34, pp. 55–63, 2010.

[34] C. S. De Silva, "Private Sound Environments in Urban Public Space: Headphone Use in Parks in New York," in *Routledge Handbook of Urban Public Space: Use, Design, and Management*, T.-S. Franck, Karen A and Huang, Ed. New York: Taylor & Francis, 2023, ch. Private So, pp. 142–159.

[35] E. Schurig, "Measuring the Auditory Bubble: How Mobile Music Listening Affects Personal Space," *Yearbook of Music Psychology*, vol. 32, 2024.

[36] H. Schepker, F. Denk, B. Kollmeier, and K. Doclo, "Acoustic Transparency in Hearables - Perceptual Sound Quality Evaluations," *Journal of the Audio Engineering Society*, vol. 68, no. 7, pp. 495–507, 2020.

[37] J. N. Gilmore, "Design for everyone: Apple AirPods and the mediation of accessibility," *Critical Studies in Media Communication*, vol. 36, no. 5, pp. 482–494, 2019.

[38] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of acoustics*. John wiley & sons, 2000.

[39] A. V. Oppenheim, A. S. Willsky, and S. H. Nawab, *Signals & systems (2nd ed.)*. USA: Prentice-Hall, Inc., 1996.

[40] H. Fletcher and W. A. Munson, "Loudness, its definition, measurement and calculation," *Bell System Technical Journal*, vol. 12, no. 4, pp. 377–430, 1933.

[41] S. A. Gelfand, *Hearing: An Introduction to Psychological and Physiological Acoustics*. M. Dekker, 1990.

[42] R. S. Woodworth, H. Schlosberg, J. W. Kling, and L. A. Riggs, *Woodworth & Schlosberg's Experimental Psychology*. Methuen, 1971.

[43] G. F. Kuhn, "Model for the interaural time differences in the azimuthal plane," *The Journal of the Acoustical Society of America*, vol. 62, no. 1, pp. 157–167, 1977.

[44] V. Benichoux, R. Marc, R. Brette, V. Benichoux, and R. Marc, "On the variations of interaural time differences with frequency," *Journal of the Acoustical Society of America*, vol. 139, no. 4, pp. 1810–1821, 2016.

[45] E. Shaw, "External ear response and sound localization," *Localization of sound: Theory and applications*, pp. 30–41, 1982.

[46] C. I. Cheng and G. H. Wakefield, "Introduction to head-related transfer functions (HRTFs): Representations of HRTFs in time, frequency, and space," *AES: Journal of the Audio Engineering Society*, vol. 49, no. 4, pp. 231–249, 2001.

[47] R. Ranjan and W. S. Gan, "Natural listening over headphones in augmented reality using adaptive filtering techniques," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 23, no. 11, pp. 1988–2002, 2015.

[48] Z. Ben-Hur, D. L. Alon, P. W. Robinson, and R. Mehra, "Localization of virtual sounds in dynamic listening using sparse HRTFs," in *Proceedings of the 3rd AES International Conference on Audio for Virtual and Augmented Reality*, Online, 2020, pp. 1–9.

[49] C. Guezenoc and R. Seguier, "HRTF Individualization : A Survey," in *Proceedings of the 145th Audio Engineering Society Convention*, New York, 2018, pp. 1–10.

[50] C. Armstrong, L. Thresh, D. Murphy, and G. Kearney, "A perceptual evaluation of individual and non-individual HRTFs: A case study of the SADIE II database," *Applied Sciences (Switzerland)*, vol. 8, no. 11, 2018.

[51] M. Schafer, R, *The Soundscape: Our Sonic Environment and the Tuning of the World*. Rochester: Destiny Books, 1977.

[52] K. Bijsterveld, *Mechanical sound: Technology, culture, and public problems of noise in the twentieth century*. MIT press, 2008.

[53] A. Cassidy and A. Einbond, *Noise in and as music*. Huddersfield: University of Huddersfield Press, 2013.

[54] D. H. Schwela, "The world health organization guidelines for environmental noise," *Noise News International*, vol. 8, no. 1, pp. 9–22, 2000.

[55] S. Stansfeld and C. Clark, "Mental Health Effects of Noise," in *Encyclopedia of Environmental Health (Second Edition)*, J. Nriagu, Ed. Oxford: Elsevier, 2019, pp. 287–294.

[56] E. Murphy, E. A. King, and H. J. Rice, "Estimating human exposure to transport noise in central Dublin, Ireland," *Environment International*, vol. 35, no. 2, pp. 298–302, 2009.

[57] A. L. Brown, J. Kang, and T. Gjestland, "Towards standardization in soundscape preference assessment," *Applied acoustics*, vol. 72, no. 6, pp. 387–392, 2011.

[58] K. Paunović, B. Jakovljević, and G. Belojević, "Predictors of noise annoyance in noisy and quiet urban streets," *Science of the Total Environment*, vol. 407, no. 12, pp. 3707–3711, 2009.

[59] D. Ouis, "Annoyance from road traffic noise: a review," *Journal of environmental psychology*, vol. 21, no. 1, pp. 101–120, 2001.

[60] R. Guski, D. Schreckenberg, and R. Schuemer, "WHO environmental noise guidelines for the European region: A systematic review on environmental noise and annoyance," *International Journal of Environmental Research and Public Health*, vol. 14, no. 12, pp. 1–39, 2017.

[61] D. Prasher, "Estimation of hearing damage from noise exposure," in *World Health Organisation and European Centre for Environment and Heath Report on the Technical meeting of exposure-response relationships of noise on health, Bonn, Germany*, 2002, pp. 17–19.

[62] C. Clark and K. Paunovic, "WHO environmental noise guidelines for the european region: A systematic review on environmental noise and cognition," *International Journal of Environmental Research and Public Health*, vol. 15, no. 2, 2018.

[63] M. Basner and S. McGuire, "WHO environmental noise guidelines for the european region: A systematic review on environmental noise and effects on sleep," *International Journal of Environmental Research and Public Health*, vol. 15, no. 3, 2018.

[64] E. van Kempen, M. Casas, G. Pershagen, and M. Foraster, "WHO environmental noise guidelines for the European region: A systematic review on environmental noise and cardiovascular and metabolic effects: A summary," *International Journal of Environmental Research and Public Health*, vol. 15, no. 2, pp. 1–59, 2018.

[65] B. C. Pijanowski, A. Farina, S. H. Gage, S. L. Dumyahn, and B. L. Krause, "What is soundscape ecology? An introduction and overview of an emerging new science," *Landscape Ecology*, vol. 26, no. 9, pp. 1213–1232, 2011.

[66] A. Lex Brown, "A review of progress in soundscapes and an approach to soundscape planning," *International Journal of Acoustics and Vibrations*, vol. 17, no. 2, pp. 73–81, 2012.

[67] S. L. Dumyahn and B. C. Pijanowski, "Soundscape conservation," *Landscape Ecology*, vol. 26, no. 9, pp. 1327–1344, 2011.

[68] Y. Jia, H. Ma, and J. Kang, "Characteristics and evaluation of urban soundscapes worthy of preservation," *Journal of Environmental Management*, vol. 253, no. June 2019, p. 109722, 2020.

[69] F. Aletta and J. Xiao, "What are the Current Priorities and Challenges for (Urban) Soundscape Research?" *Challenges*, vol. 9, no. 1, p. 16, 2018.

[70] S. Gage, P. Ummadi, A. Shortridge, J. Qi, and P. Jella, "Using gis to develop a network of acoustic environmental sensors," in *ESRI international user conference*, 2004, pp. 9–13.

[71] B. L. Giordano, R. de Miranda Azevedo, Y. Plasencia-Calaña, E. Formisano, and M. Dumontier, "What do we mean with sound semantics, exactly? a survey of taxonomies and ontologies of everyday sounds," *Frontiers in Psychology*, vol. 13, p. 964209, 2022.

[72] M. Rane, P. Coleman, R. Mason, and S. Bech, "Survey of User Perspectives on Headphone Technology," *AES Europe Spring 2022 - 152nd Audio Engineering Society Convention 2022*, pp. 106–115, 2022.

[73] C. Schneiderwind, A. Neidhardt, and D. Meyer, "Comparing the effect of different open headphone models on the perception of a real sound source," in *150th Audio Engineering Society Convention, AES 2021*, Online, 2021, pp. 1–10.

[74] V. Erbes, F. Schultz, A. Lindau, and S. Weinzierl, "An extraaural headphone system for optimized binaural reproduction," *German Annual Conference on Acoustics (DAGA)*, pp. 17–18, 2012.

[75] M. McGill, F. Mathis, M. Khamis, and J. Williamson, "Augmenting TV Viewing using Acoustically Transparent Auditory Headsets," in *IMX 2020 - Proceedings of the 2020 ACM International Conference on Interactive Media Experiences*, Barcelona, 2020, pp. 34–44.

[76] T. Bradshaw, "Tech Teardown: How Bose Frames work," 2019, [Online] Available at: https://www.ft.com/content/af0b3ac6-9f46-11e9-9c06-a4640c9feebb [Accessed: 15/04/2024].

[77] A. Barde, R. W. Lindeman, G. Lee, and M. Billinghurst, "Binaural spatialisation over a bone conduction headset: The perception of elevation," in *2019 AES International Conference on Headphone Technology*, San Fransisco, 2019, pp. 1–11.

[78] N. Meyer-Kahlen, D. Rudrich, M. Brandner, S. Wirler, S. Windtner, and M. Frank, "DIY modifications for acoustically transparent headphones," in *148th Audio Engineering Society International Convention*, Vienna, 2020, pp. 2–6.

[79] M. McGill, S. Brewster, D. McGookin, and G. Wilson, "Acoustic Transparency and the Changing Soundscape of Auditory Mixed Reality," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu, 2020, pp. 1–16.

[80] N. Sawhney and C. Schmandt, "Nomadic Radio: Speech and Audio Interaction for Contextual Messaging in Nomadic Environments," *ACM Transactions on Computer-Human Interaction*, vol. 7, no. 3, pp. 353–383, 2000.

[81] V. Bauer, A. Nagele, C. Baume, T. Cowlishaw, H. Cooke, C. Pike, and P. G. T. Healey, "Designing an Interactive and Collaborative Experience in Audio Augmented Reality - Virtual Reality and Augmented Reality," in *Virtual Reality and Augmented Reality*, P. Bourdot, V. Interrante, L. Nedel, N. Magnenat-Thalmann, and G. Zachmann, Eds., vol. 1. Cham: Springer International Publishing, 2019, pp. 305–311.

[82] Meta, "Introducing the New Ray-Ban | Meta Smart Glasses," 2023, [Online] Available at: https://about.fb.com/news/2023/09/new-ray-ban-meta-smart-glasses/ [Accessed: 21/11/2024].

[83] M. F. Simón Gálvez, D. Menzies, and F. M. Fazi, "Dynamic audio reproduction with linear loudspeaker arrays," *AES: Journal of the Audio Engineering Society*, vol. 67, no. 4, pp. 190–200, 2019.

[84] Audioscenic, "The Future of 3D Sound as part of the BBC How To Make Program," 2020, [Online] Available at: https://www.audioscenic.com/news/2020-04-20_bbc_feature [Accessed: 15/04/2024].

[85] T. Betlehem, W. Zhang, M. A. Poletti, and T. D. Abhayapala, "Personal sound zones: Delivering interface-free audio to multiple listeners," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 81–91, 2015.

[86] R. W. Lindeman, H. Noma, and P. G. De Barros, "Hear-through and mic-through augmented reality: Using bone conduction to display spatialized audio," in *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, ISMAR*, Nara, 2007, pp. 173–176.

[87] F. Rumsey, "Headphone technology Hear-through, bone conduction, noise canceling," *Journal of the Audio Engineering Society*, vol. 67, no. 11, pp. 914–919, 2019.

[88] R. Schlieper, L. Song, S. Preihs, and J. Peissig, "Estimation of the Headphone "Openness" based on Measurements of Pressure Division Ratio, Headphone Selection Criterion and Acoustic Impedance Curve," in *145th Audio Engineering Society International Convention, AES 2018*, New York, 2018, pp. 1–9.

[89] D. Satongar, C. Pike, Y. W. Lam, and A. I. Tew, "The Influence of Headphones on the Localization of External Loudspeaker Sources," *Journal of the Audio Engineering Society*, vol. 63, no. 10, pp. 799–810, 2015.

[90] R. Gupta, R. Ranjan, J. He, and W. S. Gan, "Parametric Hear through Equalization for Augmented Reality Audio," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pp. 1587–1591, 2019.

[91] ETSI, "Speech and multimedia Transmission Quality (STQ); Test Methods and Performance Requirements for Active Noise Cancellation Headsets and other Earphones," Tech. Rep., 2022.

[92] J. Rämö and V. Välimäki, "Digital augmented reality audio headset," *Journal of Electrical and Computer Engineering*, pp. 1–13, 2012.

[93] V. Riikonen, M. Tikander, and M. Karjalainen, "An Augmented Reality Audio mixer and equalizer," in *Audio Engineering Society - 124th Audio Engineering Society Convention 2008*, vol. 1, no. January 2008, Amsterdam, 2008, pp. 564–571.

[94] J. Liski, R. Väänänen, S. Vesa, and V. Välimäki, "Adaptive equalization of acoustic transparency in an augmented-reality headset," in *Proceedings of the AES International Conference on Headphone Technology*, 2016, pp. 1–8.

[95] A. Kulkarni and H. S. Colburn, "Variability in the characterization of the headphone transfer-function," *The Journal of the Acoustical Society of America*, vol. 107, no. 2, pp. 1071–1074, 2000.

[96] Qualcomm and SAR Insight & Consulting, "The State of Sound Report 2022 A global analysis of audio consumer behaviors and desires," pp. 1–29, 2022, [Online] Available at: https://assets.qualcomm.com/rs/385-TWS-803/images/The_State_of_Sound_2022.pdf [Accessed: 21/11/2024].

[97] N. Hunn, "Hearables-the new Wearables," 2014, [Online] Available at: https://www.nickhunn.com/hearables-the-new-wearables/ [Accessed: 15/04/2024].

[98] N. Hunn, "The Market for Hearable Devices 2016-2020," 2016, [Online] Available at: https://www.nickhunn.com/wp-content/uploads/downloads/2016/11/The-Market-for-Hearable-Devices-2016-2020.pdf [Accessed: 21/11/2024].

[99] J. Plazak and M. Kersten-Oertel, "A survey on the affordances of "hearables"," *Inventions*, vol. 3, no. 3, pp. 1–16, 2018.

[100] C. Min, A. Mathur, and F. Kawsar, "Exploring audio and kinetic sensing on earable devices," in *WearSys 2018 - Proceedings of the 4th ACM Workshop on Wearable Systems and Applications*, Munich, 2018, pp. 1–6.

[101] C. Min, A. Montanari, A. Mathur, S. Lee, and F. Kawsar, "Cross-modal approach for conversational well-being monitoring with multi-sensory earables," in *UbiComp/ISWC 2018 - Adjunct*

Proceedings of the 2018 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2018 ACM International Symposium on Wearable Computers, Singapore, 2018, pp. 706–709.

[102] F. Kawsar, C. Min, A. Mathur, A. Montanari, O. Amft, and K. Van Laerhoven, "Earables for personal-scale behavior analytics," *IEEE Pervasive Computing*, vol. 17, no. 3, pp. 83–89, 2018.

[103] S. F. Temme, "Testing audio performance of hearables," in *2019 AES International Conference On Headphone Technology*, San Francisco, 2019, pp. 1–9.

[104] F. Kawsar, C. Min, A. Mathur, A. Montanari, U. G. Acer, and M. Van den Broeck, "eSense: Open earable platform for human sensing," in *UbiComp '18: Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, Singapore, 2018, pp. 381–383.

[105] A. Ferlini, C. Mascolo, and R. Harle, "Head Motion Tracking Through in-Ear Wearables," in *EarComp'19*, London, 2018, pp. 1–6.

[106] P. Crum, "Here come the Hearables," *IEEE Spectrum*, vol. 56, no. 5, pp. 38–43, 2019.

[107] E. C. Cherry, "Some experiments on the recognition of speech, with one and with two ears," *The Journal of the Acoustical Society of America*, vol. 25, no. 5, pp. 975–979, 1953.

[108] L. B. Jaloza, "Inside Facebook Reality Labs Research: The future of audio," 2020, [Online] Available at: https://about.fb.com/news/2020/09/facebook-reality-labs-research-future-of-audio/ [Accessed: 15/04/2024].

[109] Y. Jiang, C. Wu, K. Deng, and Y. Wu, "An audio fingerprinting extraction algorithm based on lifting wavelet packet and improved optimal-basis selection," *Multimedia Tools and Applications*, vol. 78, no. 21, pp. 30 011–30 025, 2019.

[110] D. Williams, D. T. Murphy, and B. M. Fazenda, "Towards Bio-responsive Control for Music," in *DMRN+12: DIGITAL MUSIC RESEARCH NETWORK ONE-DAY WORKSHOP 2017*, London, 2017, p. 1.

[111] Apple, "AirPods 1 - Tech Specs," 2024, [Online] Available at: https://support.apple.com/en-gb/111855 [Accessed: 11/11/2024].

[112] V. Välimäki, A. Franck, J. Ramo, H. Gamper, and L. Savioja, "Assisted listening using a headset: Enhancing audio perception in real, augmented, and virtual environments," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 92–99, 2015.

[113] Audio Analytic, "The 2019 Hearables Survey: AI Attitudes and Expectations," Cambridge, pp. 1–43, 2019, [Online] Available at: https://www.audioanalytic.com/hearables-report-2019/ [Accessed: 15/12/2020].

[114] J. Hardley, "Embodied Perceptions of Darkness: Multistable Experiences of Mobile Media in the Urban Night," *M/C Journal*, vol. 24, no. 2, 2021.

[115] Apple, "AirPods Pro (2nd generation) with MagSafe Charging Case (Lightning)," 2024, [Online] Available at: https://support.apple.com/en-gb/docs/airpods/300882 [Accessed: 11/11/2024].

[116] Sony, "LinkBuds S," 2024, [Online] Available at: https://www.sony.co.uk/electronics/truly-wireless/linkbuds-s [Accessed: 11/11/2024].

[117] Bose, "BOSE QUIETCOMFORT® EARBUDS II," 2024, [Online] Available at: https://www.bose.co.uk/en_gb/pxp/quietcomfort-earbuds-ii.html [Accessed: 11/11/2024].

[118] F. Denk, H. Schepker, S. Doclo, and B. Kollmeier, "Acoustic transparency in hearables-technical evaluation," *AES: Journal of the Audio Engineering Society*, vol. 68, no. 7-8, pp. 508–521, 2020.

[119] S. J. Elliott and P. A. Nelson, "Active noise control," *IEEE Signal Processing Magazine*, vol. 10, no. 4, pp. 12–35, 1993.

[120] L. Lu, K. L. Yin, R. C. de Lamare, Z. Zheng, Y. Yu, X. Yang, and B. Chen, "A survey on active noise control in the past decade—Part I: Linear systems," *Signal Processing*, vol. 183, 2021.

[121] L. Lu, K. L. Yin, R. C. de Lamare, Z. Zheng, Y. Yu, X. Yang, and B. Chen, "A survey on active noise control in the past decade–Part II: Nonlinear systems," *Signal Processing*, vol. 181, 2021.

[122] V. Patel, J. Cheer, and S. Fontana, "Design and Implementation of an Active Noise Control Headphone with Directional Hear-Through Capability," *IEEE Transactions on Consumer Electronics*, vol. 66, no. 1, pp. 32–40, 2020.

[123] M. Klatte, K. Bergström, and T. Lachmann, "Does noise affect learning? A short review on noise effects on cognitive performance in children," *Frontiers in Psychology*, vol. 4, no. August, pp. 1–6, 2013.

[124] M. J. Jafari, R. Khosrowabadi, S. Khodakarim, and F. Mohammadian, "The effect of noise exposure on cognitive performance and brain activity patterns," *Open Access Macedonian Journal of Medical Sciences*, vol. 7, no. 17, pp. 2924–2931, 2019.

[125] R. Thompson, R. B. Smith, Y. Bou Karim, C. Shen, K. Drummond, C. Teng, and M. B. Toledano, "Noise pollution and human cognition: An updated systematic review and meta-analysis of recent evidence," *Environment International*, vol. 158, no. October 2021, p. 106905, 2022.

[126] W. K. Persson, R. Rylander, S. Benton, and H. G. Leventhall, "Effects on performance and work quality due to low frequency ventilation noise," *Journal of Sound and Vibration*, vol. 205, no. 4, pp. 467–474, 1997.

[127] W. H. Organization, *Burden of disease from environmental noise: quantification of healthy life years lost in Europe*. World Health Organization. Regional Office for Europe, 2011.

[128] T. Münzel, T. Gori, W. Babisch, and M. Basner, "Cardiovascular effects of environmental noise exposure," *European Heart Journal*, vol. 35, no. 13, pp. 829–836, 2014.

[129] R. Leung, M. M. Cook, M. F. Capra, and K. R. Johnstone, "The contribution of respiratory and hearing protection use to psychological distress in the workplace: a scoping review," *International Archives of Occupational and Environmental Health*, vol. 95, no. 8, pp. 1647–1659, 2022.

[130] J. C. Fernandes, "Effects of hearing protector devices on speech intelligibility," *Applied Acoustics*, vol. 64, no. 6, pp. 581–590, 2003.

[131] M. Hashimoto, M. Kumashiro, and S. Miyake, "Speech perception in noise when wearing hearing protectors with little low-frequency attenuation," *International Journal of Industrial Ergonomics*, vol. 18, no. 2-3, pp. 121–126, 1996.

[132] E. Pekkarinen, V. Viljanen, A. Salmivalli, and J. Suonpää, "Recognition of Low-Pass Filtered Words in a Noisy and Reverberant Environment with and Without Earmuffs," *Scandinavian Audiology*, vol. 19, no. 4, pp. 223–227, 1990.

[133] B. J. Mueller, A. Liebl, N. Herget, D. Kohler, and P. Leistner, "Using active noise-cancelling headphones in open-plan offices: No influence on cognitive performance but improvement of perceived privacy and acoustic environment," *Frontiers in Built Environment*, vol. 8, no. September, pp. 1–15, 2022.

[134] F. Niu, X. Qiu, and D. Zhang, "Effects of active noise cancelling headphones on speech recognition," *Applied Acoustics*, vol. 165, no. 107335, 2020.

[135] Y. Kim and Y. Park, "Effect of active noise control and masking sound on speech intelligibility," *Applied Acoustics*, vol. 123, pp. 152–157, 2017.

[136] B. Rudzyn and M. Fisher, "Performance of personal active noise reduction devices," *Applied Acoustics*, vol. 73, no. 11, pp. 1159–1167, 2012.

[137] C. J. Struck, "Objective measurements of headphone active noise cancelation performance," in *2019 AES International Conference on Headphone Technology*, San Francisco, 2019, pp. 1–6.

[138] H. Y. Seol, S. H. Kim, G. Y. Kim, M. Jo, Y. S. Cho, S. H. Hong, and I. J. Moon, "Influence of the Noise-Canceling Technology on How We Hear Sounds," *Healthcare (Switzerland)*, vol. 10, no. 8, pp. 1–8, 2022.

[139] H. Yu, J. Filloon, D. N. Ramsey, K. K. Ahuja, and L. N. Sankar, "Modelling and Testing Sound Leakage in Ill-fitting Earplugs," in *AIAA SCITECH 2022 Forum*, San Diego, 2022, p. 2562.

[140] A. G. Samelli, R. F. Gomes, T. V. Chammas, B. G. Silva, R. R. Moreira, and A. C. Fiorini, "The study of attenuation levels and the comfort of earplugs." *Noise & health*, vol. 20, no. 94, pp. 112–119, 2018.

[141] E. H. Berger and P. Hamery, "Empirical evaluation using impulse noise of the level-dependency of various passive earplug designs," *Journal of the Acoustical Society of America*, vol. 123, no. 5, pp. 3528–3528, 2008.

[142] S. M. R. Taghavi, G. Mohammadkhani, and H. Jalilvand, "Speech intelligibility index: A literature review," *Auditory and Vestibular Research*, 2022.

[143] K. D. Kryter, "Methods for the calculation and use of the articulation index," *The Journal of the Acoustical Society of America*, vol. 34, no. 11, pp. 1689–1697, 1962.

[144] K. L. Payton and L. D. Braida, "A method to determine the speech transmission index from speech waveforms," *The Journal of the Acoustical Society of America*, vol. 106, no. 6, pp. 3637–3648, 1999.

[145] I. E. Commission, "IEC 60268-16:2020 Sound system equipment - Part 16: Objective rating of speech intelligibility by speech transmission index," 2020, [Online] Available at: https://webstore.iec.ch/en/publication/26771 [Accessed: 09/04/2025].

[146] B. Herrmann and I. S. Johnsrude, "Absorption and Enjoyment During Listening to Acoustically Masked Stories," *Trends in Hearing*, vol. 24, 2020.

[147] B. Herrmann and I. S. Johnsrude, "A model of listening engagement (MoLE)," *Hearing Research*, vol. 397, no. 108016, 2020.

[148] E. C. Westgate and T. D. Wilson, "Boring thoughts and bored minds: The mac model of boredom and cognitive engagement." *Psychological Review*, vol. 125, no. 5, p. 689, 2018.

[149] P. Casagranda, M. L. Sapino, and K. S. Candan, "Context-Aware proactive personalization of linear audio content," in *Advances in Database Technology - EDBT*, vol. 2017-March, 2017, pp. 574–577.

[150] Regency TR-1, "Owners Manual - TR-1G," [Online] Available at: https://www.regencytr1.com/images/OwnersManual-TR-1G.pdf [Accessed: 29/04/2024].

[151] R. Smith, "IBM created the world's first smartphone 25 years ago," 2024, [Online] Available at: https://www.weforum.org/stories/2018/03/remembering-first-smartphone-simon-ibm/ [Accessed: 12/11/2024].

[152] J. Cook, "Smartphone History: From the First Smartphone to Today," 2024, [Online] Available at: https://www.arrow.com/en/research-and-events/articles/smartphone-history-from-the-first-smartphone-to-today [Accessed: 12/11/2024].

[153] Mobile Phone Museum, "ERICSSON R380S," 2024, [Online] Available at: https://www.mobilephonemuseum.com/phone-detail/ericsson-r380s [Accessed: 12/11/2024].

[154] Mobile Phone Museum, "SAMSUNG SPH-M100," 2024, [Online] Available at: https://www.mobilephonemuseum.com/phone-detail/sgh-m100 [Accessed: 12/11/2024].

[155] Edison Research, "The Infinite Dial 2024," 2024, [Online] Available at: https://www.edisonresearch.com/wp-content/uploads/2024/06/Infinite-Dial-2024-Presentation.pdf [Accessed: 21/11/2024].

[156] M. M. Kuijpers, S. Douglas, and K. Bálint, "Narrative Absorption: An Overview," *Handbook of Empirical Literary Studies*, pp. 279–304, 2021.

[157] M. Csikzentmihalyi, *Beyond boredom and anxiety*. San Francisco: Jossey-Bass, 1975.

[158] M. Csikszentmihalyi and M. Csikzentmihaly, *Flow: The psychology of optimal experience*. Harper & Row New York, 1990, vol. 1990.

[159] V. Nell, *Lost in a book: The psychology of reading for pleasure.* Yale University Press, 1988.

[160] M. M. Kuijpers, F. Hakemulder, E. S. Tan, and M. M. Doicaru, "Exploring absorbing reading experiences Developing and validating a self-report scale to measure story world absorption," *Scientific Study of Literature*, vol. 4, no. 1, pp. 89–122, 2014.

[161] M. Lombard, R. Reich, M. Grabe, C. Bracken, and T. Ditton, "Presence and television.." *Human Communication Research*, vol. 26, no. 1, pp. 75–98, 2000.

[162] B. G. Witmer and M. J. Singer, "Measuring Presence in Virtual Environments: A Presence Questionnaire," *Presence: Teleoperators and Virtual Environments*, vol. 7, no. 3, pp. 225–240, 1998.

[163] R. Agarwal and E. Karahanna, "Time Flies When You're Having Fun: Cognitive Absorption and Beliefs about Information Technology Usage," *MIS Quarterly*, vol. 24, no. 4, pp. 53–54, 2000.

[164] R. Gerrig, *Experiencing Narrative Worlds: On the Psychological Activities of Reading*. Yale University Press, 1993.

[165] M. C. Green and T. C. Brock, "The role of transportation in the persuasiveness of public narratives." *Journal of personality and social psychology*, vol. 79, no. 5, p. 701, 2000.

[166] S. Raoul and J. Arthur, *Gehirn und Gedicht. Wie wir unsere Wirklichkeiten konstruieren (Brain and Poetry: How We Construct Our Realities)*. Carl Hanser Verlag, München [Niemcy], 2011.

[167] M.-L. Ryan, *Narrative as virtual reality 2: Revisiting immersion and interactivity in literature and electronic media*. JHU press, 2015.

[168] R. Busselle and H. Bilandzic, "Measuring Narrative Engagement," *Media Psychology*, vol. 12, no. 4, pp. 321–347, 2009.

[169] B. A. Thissen, W. Menninghaus, and W. Schlotz, "Measuring optimal reading experiences: The reading flow short scale," *Frontiers in Psychology*, vol. 9, no. DEC, pp. 1–12, 2018.

[170] D. Kuiken and S. Douglas, "Forms of absorption that facilitate the aesthetic and explanatory effects of literary reading," in *Narrative Absorption*, F. Hakemulder, M. M. Kuijpers, E. S. Tan, K. Bálint, and M. M. Doicaru, Eds. Amsterdam: John Benjamins Publishing Company, 2017, pp. 217–249.

[171] J. Cohen, "Defining Identification: A Theoretical Look at the Identification of Audiences With Media Characters," *Mass Communication and Society*, vol. 4, no. 3, pp. 245–264, 2001.

[172] M. C. Green, T. C. Brock, and G. F. Kaufman, "Understanding media enjoyment: The role of transportation into narrative worlds," *Communication Theory*, vol. 14, no. 4, pp. 311–327, 2004.

[173] S. Hartmeyer, R. Grzeschik, T. Wolbers, and J. M. Wiener, "The effects of attentional engagement on route learning performance in a virtual environment: An aging study," *Frontiers in Aging Neuroscience*, vol. 9, no. JUL, 2017.

[174] H. Song, E. S. Finn, and M. D. Rosenberg, "Neural signatures of attentional engagement during narratives and its consequences for event memory," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 118, no. 33, 2021.

[175] E. M. Kaya and M. Elhilali, "Modelling auditory attention," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 372, no. 1714, 2017.

[176] H. White and P. Shah, "Attention in urban and natural environments," *Yale Journal of Biology and Medicine*, vol. 92, no. 1, pp. 115–120, 2019.

[177] K. J. Linnell, S. Caparos, J. W. de Fockert, and J. Davidoff, "Urbanization decreases attentional engagement," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 39, no. 5, pp. 1232–1247, 2013.

[178] M. G. Berman, J. Jonides, and S. Kaplan, "The cognitive benefits of interacting with nature," *Psychological Science*, vol. 19, no. 12, pp. 1207–1212, 2008.

[179] M. Armstrong, M. Brooks, A. Churnside, M. Evans, F. Melchior, and M. Shotton, "Object-based broadcasting - curation, responsiveness and user experience," *IET Conference Proceedings*, 2014.

[180] A. Gradinar, D. Burnett, P. Coulton, I. Forrester, M. Watkins, T. Scutt, and E. Murphy, "Perceptive Media – Adaptive Storytelling for Digital Broadcast," in *Human-Computer Interaction - INTERACT 2015: 15th IFIP TC 13 International Conference*, vol. 9299, Bamberg, 2015, pp. 586–589.

[181] OFCOM, "Object-based media report," 2021, [Online] Available at: https://www.ofcom.org.uk/siteassets/resources/documents/research-and-data/technology-research/2021/object-based-media-report.pdf?v=322207 [Accessed: 21/11/2024].

[182] R. Stagg, "Adaptive podcasting - the future of personalised podcast experiences." 2022, [Online] Available at: https://www.bbc.co.uk/rd/blog/2022-09-adaptive-podcasting [Accessed: 6/12/2022].

[183] K. M. Lee, "Presence, Explicated," *Communication Theory*, pp. 27–50, 2004.

[184] C. S. Oh, J. N. Bailenson, and G. F. Welch, "A systematic review of social presence: Definition, antecedents, and implications," *Frontiers Robotics AI*, vol. 5, no. OCT, pp. 1–35, 2018.

[185] M. Slater and S. Wilbur, "A framework for immersive virtual environments five: Speculations on the role of presence in virtual environments," *Presence: Teleoperators and Virtual Environments*, vol. 6, no. 6, pp. 603–616, 2007.

[186] S. Agrawal, A. Simon, S. Bech, K. BÆrentsen, and S. Forchhammer, "Defining immersion: Literature review and implications for research on audiovisual experiences," *AES: Journal of the Audio Engineering Society*, vol. 68, no. 6, pp. 404–417, 2020.

[187] G. C. Stecker, T. M. Moore, M. Folkerts, D. Zotkin, and R. Duraiswami, "Toward objective measures of auditory co-immersion in virtual and augmented reality," in *Proceedings of the AES International Conference on Audio for Virtual and Augmented Reality*, Redmond, 2018, pp. 1–6.

[188] M. Bull, "Investigating the Culture of Mobile Listening: From Walkman to iPod," *Consuming Music Together*, pp. 131–149, 2006.

[189] L. Nyre, "Urban Headphone Listening and the Situational Fit of Music, Radio and Podcasting," *Journal of Radio and Audio Media*, vol. 22, no. 2, pp. 279–298, 2015.

[190] P. Wash and S. Dance, "MP3 listening levels on London underground for music and speech," *Applied Acoustics*, vol. 74, no. 6, pp. 850–855, 2013.

[191] A. E. Krause, A. C. North, and L. Y. Hewitt, "Music-listening in everyday life: Devices and choice," *Psychology of Music*, vol. 43, no. 2, pp. 155–170, 2015.

[192] A. E. Krause, A. C. North, and L. Y. Hewitt, "The Role of Location in Everyday Experiences of Music," *Psychology of Popular Media Culture*, vol. 5, no. 3, pp. 232–257, 2016.

[193] A. E. Krause and A. C. North, "Pleasure, arousal, dominance, and judgments about music in everyday life," in *Psychology of Music*. SAGE Publications Ltd, sep 2016, vol. 45, no. 3, pp. 355–374.

[194] A. E. Krause and A. C. North, "How do location and control over the music influence listeners' responses?" *Scandinavian Journal of Psychology*, vol. 58, no. 2, pp. 114–122, 2017.

[195] A. E. Krause and S. C. Brown, "A uses and gratifications approach to considering the music formats that people use most often," *Psychology of Music*, vol. 49, no. 3, pp. 547–566, 2021.

[196] A. Heye and A. Lamont, "Mobile listening situations in everyday life: The use of MP3 players while traveling," *Musicae Scientiae*, vol. XIV, no. 1, pp. 95–120, 2010.

[197] L. Detry, "Mobile Music Listening: The Users and their Art of Making Do," *Networking Knowledge: Journal of the MeCCSA Postgraduate Network*, vol. 9, no. 2, pp. 1–13, 2016.

[198] S. Behbehani and J. Steffens, "Musical DIAMONDS: The influence of situational classes and characteristics on music listening behavior," *Psychology of Music*, 2020.

[199] M. Kuch and C. Wöllner, "On the Move: Principal Components of the Functions and Experiences of Mobile Music Listening," *Music & Science*, vol. 4, pp. 1–14, 2021.

[200] A. Li, A. Wang, Z. Nazari, P. Chandar, and B. Carterette, "Do podcasts and music compete with one another? Understanding users' audio streaming habits," *The Web Conference 2020 - Proceedings of the World Wide Web Conference, WWW 2020*, pp. 1920–1931, 2020.

[201] E. Jones and J. Morris, "Competing sounds? Podcasting and popular music," *Radio Journal:International Studies in Broadcast & Audio Media*, vol. 20, no. 1, 2022.

[202] M. Bimm, ""Let's See If We've Been Missing Out !" Affect, Reactivity, and the Middlebrow in Popular Music Podcasting," *Journal of Popular Music Studies*, vol. 35, no. 4, pp. 18–33, 2023.

[203] M. Rane, P. Coleman, R. Mason, and S. Bech, "Quantifying headphone listening experience in virtual sound environments using distraction," *Eurasip Journal on Audio, Speech, and Music Processing*, no. 1, 2022.

[204] S. Bosshard, E. Rodero, I. Rodríguez-de Dios, and J. Brickner, "Radio, Podcasts, and Music Streaming—An Electroencephalography and Physiological Analysis of Listeners' Attitude, Attention, Memory, and Engagement," *brain sciences*, vol. 14, no. 4, 2024.

[205] J. Coens, E. Degryse, M. P. Senecaut, J. Cottyn, and G. Clarebout, "Listening to an Educational Podcast While Walking or Jogging: Can Students Really Multitask?" in *IADIS International Conference Mobile Learning 2011*, Rome, 2011, pp. 70–76.

[206] M. T. Soto-sanfiel, "Impact of Mobility on Spatial Presence during Audio Narrative Reception," in *The Eighth International Conference on Digital Society*, no. c, Barcelona, 2014, pp. 152–156.

[207] P. Vorderer, W. Wirth, F. Gouveia, F. Biocca, T. Saari, F. Jäncke, S. Böcking, H. Schramm, A. Gysbers, T. Hartmann *et al.*, "Mec spatial presence questionnaire (mec-spq): Short documentation and instructions for application. report to the european community, project presence: Mec," *MEC (IST-2001-37661)*, 2004.

[208] M. T. Soto-Sanfiel, "Engagement and mobile listening," *International Journal of Mobile Communications*, vol. 13, no. 1, pp. 29–50, 2015.

[209] Edison Research, "The Podcast Consumer 2019," Edison Research, Tech. Rep., 2019, [Online] Available at: https://www.edisonresearch.com/the-podcast-consumer2019/ [Accessed: 16/05/2024].

[210] R. Berry, "Podcasting: Considering the evolution of the medium and its association with the word 'radio'," *Radio Journal*, vol. 14, no. 1, pp. 7–22, 2016.

[211] K. M. Markman, "Considerations - Reflections and Future Research. Everything Old is New Again: Podcasting as Radio's Revival," *Journal of Radio and Audio Media*, vol. 22, no. 2, pp. 240–243, 2015.

[212] M. Edmond, "All platforms considered: Contemporary radio and transmedia engagement," *New Media and Society*, vol. 17, no. 9, pp. 1566–1582, 2015.

[213] R. Berry, "Radio, music, podcasts – BBC Sounds: Public service radio and podcasts in a platform world," *Radio Journal*, vol. 18, no. 1, pp. 63–78, 2020.

[214] Radio Research Consortium, "Radio Listening Location Percentages In CDM Markets In June (Apr-Jun) 2022," 2022, [Online] Available at: https://rrconline.org/radio-listening-location-percentages-in-cdm-markets-in-june-apr-jun-2022/ [Accessed: 21/11/2024].

[215] S. Chan-Olmsted and R. Wang, "Understanding podcast users: Consumption motives and behaviors," *New Media and Society*, pp. 1–21, 2020.

[216] L. G. Perks and J. S. Turner, "Podcasts and Productivity: A Qualitative Uses and Gratifications Study," *Mass Communication and Society*, vol. 22, no. 1, pp. 96–116, 2019.

[217] J. W. Morris and E. Patterson, "Podcasting and its Apps: Software, Sound, and the Interfaces of Digital Audio," *Journal of Radio and Audio Media*, vol. 22, no. 2, pp. 220–230, 2015.

[218] M. Spinelli and L. Dann, "Don't Look Back: The New Possibilities of Podcast Drama," in *Podcasting: The Audio Media Revolution*. London: Bloomsbury Academic, 2019, ch. Don't Look, pp. 116–120.

[219] A. M. Rubin, "Ritualized and instrumental television viewing." *Journal of Communication*, vol. 34, no. 3, pp. 67–77, 1984.

[220] A. E. Greasley and A. Lamont, "Exploring engagement with music in everyday life using experience sampling methodology," *Musicae Scientiae*, vol. 15, no. 1, pp. 45–71, 2011.

[221] d. S. Gabriel Tassinari, R. Simoni F, and C. F. S. Netto, "Listen as Much as You Want: Antecedents of the Engagement of Podcast Consumers," in *XLIV ENCONTRO DA ANPAD - EnANPAD*, Online, 2020, pp. 1–10.

[222] S. J. Tobin and R. E. Guadagno, "Why people listen: Motivations and outcomes of podcast listening," *Plos One*, vol. 17, no. 4, 2022.

[223] D. Schlütz and I. Hedder, "Aural Parasocial Relations: Host–Listener Relationships in Podcasts," *Journal of Radio and Audio Media*, vol. 00, no. 00, pp. 1–18, 2021.

[224] D. García-Marín, "Mapping the factors that determine engagement in podcasting: Design from the users and podcasters'experience," *Communication and Society*, vol. 33, no. 2, pp. 49–63, 2020.

[225] L. L. Thurstone, *Multiple factor analysis.* University of Chicago Press: Chicago, 1947.

[226] M. Susino, "ICMPC-ESCOM 2021 Emotional responses to music: the need to account for contextual factors," *16th International Conference on Music Perception and Cognition & 11th triennial conference of ESCOM*, 2021, [Online video] Available at: https://www.youtube.com/watch?v=k8UqGq5ouO8 [Accessed: 12/04/2022].

[227] H. F. Kaiser, "An index of factorial simplicity," *Psychometrika*, vol. 39, no. 1, pp. 31–36, 1974.

[228] V. Nassiri, A. Lovik, G. Molenberghs, G. Verbeke, and T. Busch, "Multiple Imputation for Exploratory Factor Analysis [R package mifa version 0.2.0]." 2021, [Online] Available at: https://cran.r-project.org/package=mifa [Accessed: 12/04/2022].

[229] V. Nassiri, A. Lovik, G. Molenberghs, and G. Verbeke, "On using multiple imputation for exploratory factor analysis," *Behavior Research Methods*, vol. 50, pp. 501–517, 2018.

[230] S. van Buuren, "Multiple imputation of discrete and continuous data by fully conditional, specification." *Statistical methods in medical research*, vol. 16, no. 3, pp. 219–242, jun 2007.

[231] D. B. Rubin, *Multiple imputation for nonresponse in surveys*. John Wiley & Sons, 2004, vol. 81.

[232] T. O. Gibson Jr., J. A. Morrow, and L. M. Rocconi, "A Modernized Heuristic Approach to Robust Exploratory Factor Analysis," *The Quantitative Methods for Psychology*, vol. 16, no. 4, pp. 295–307, 2020.

[233] W. Revelle, "psych: Procedures for Psychological, Psychometric, and Personality Research [R package psych 2.2.3]." 2022, [Online] Available at: https://cran.r-project.org/package=psych [Accessed: 12/04/2022].

[234] T. Busch and V. Nassiri, "MIFA Github Repository." 2021, [Online] Available at: https://teebusch.github.io/mifa/ [Accessed: 12/04/2022].

[235] J. Hair, R. Anderson, B. Babin, and W. Black, *Multivariate Data Analysis*, 7th ed. New Jersey: Pearson Prentice Hall, 2010.

[236] D. Goretzko, T. T. H. Pham, and M. Bühner, "Exploratory factor analysis: Current use, methodological developments and recommendations for good practice," *Current Psychology*, vol. 40, no. 7, pp. 3510–3521, 2021.

[237] M. C. Howard, "A Review of Exploratory Factor Analysis Decisions and Overview of Current Practices: What We Are Doing and How Can We Improve?" *International Journal of Human-Computer Interaction*, vol. 32, no. 1, pp. 51–62, 2016.

[238] D. Goretzko, "Factor Retention in Exploratory Factor Analysis With Missing Data," *Educational and Psychological Measurement*, pp. 1–21, 2021.

[239] R. B. Cattell, "The Scree Test For The Number Of Factors," *Multivariate Behavioral Research*, vol. 1, no. 2, pp. 245–276, 1966.

[240] Signal Hill Insights, "The Canadian Podcast Listener 2021: Summary Report," 2021, [Online] Available at: https://static1.squarespace.com/static/5d96a5747b6ebf10df78f632/t/61b1282a42edb11bde4b1c92/1639000108131/CPL+21+Summary+Report+%28Dec8%29.pdf [Accessed: 21/11/2024].

[241] P. Schober, C. Boer, and L. A. Schwarte, "Correlation Coefficients: Appropriate Use and Interpretation." *Anesthesia and analgesia*, vol. 126, no. 5, pp. 1763–1768, may 2018.

[242] Y. Choi, M. Kim, and C. Chun, "Effect of temperature on attention ability based on electroencephalogram measurements," *Building and Environment*, vol. 147, no. October 2018, pp. 299–304, 2019.

[243] A. M. Rubin and E. M. Perse, "Audience Activity and Television News Gratifications," *Communication Research*, vol. 14, no. 1, pp. 58–84, 1987.

[244] E. Best and A. Cole, "A National Literacy Trust research report Young people's engagement with podcasts in 2021," pp. 1–18, 2022, [Online] Available at: https://cdn.literacytrust.org.uk/media/documents/Young_peoples_engagement_with_podcasts_in_2021.pdf [Accessed: 21/11/2024].

[245] E. Hunt, "Five of the best podcasts: sleep-inducing sounds and stories." 2021, [Online] Available at: https://www.theguardian.com/culture/2021/nov/02/five-of-the-best-podcasts-sleep-inducing-sounds-and-stories [Accessed: 9/12/2022].

[246] P. Valdez, "Circadian rhythms in attention," *Yale Journal of Biology and Medicine*, vol. 92, no. 1, pp. 81–92, 2019.

[247] M. I. Posner and R. D. Rafal, "Cognitive theories of attention and the rehabilitation of attentional deficits." in *Neuropsychological rehabilitation.* New York, NY, US: The Guilford Press, 1987, pp. 182–201.

[248] C. Cederroth, F. Gachon, and B. Canlon, "Time to listen: circadian impact on auditory research," *Current Opinion in Physiology*, vol. 18, no. Ic, pp. 95–99, 2020.

[249] A. P. Smith, "Noise and aspects of attention," *British Journal of Psychology*, vol. 82, no. 3, pp. 313–324, 1991.

[250] A. J. Vickery and S. Ventrano, "Listening Goals and Parasocial Relationships: How Listening Styles Impact the Development of Parasocial Relationships with Media Personas," *International Journal of Listening*, vol. 35, no. 3, pp. 149–165, 2020.

[251] M. B. Pontifex, A. C. Parks, D. A. Henning, and K. Kamijo, "Single bouts of exercise selectively sustain attentional processes," *Psychophysiology*, vol. 52, no. 5, pp. 618–625, 2015.

[252] A. E. Greasley and A. M. Lamont, "Music preference in adulthood: Why do we like the music we do?" in *Proceedings of the 9th International Conference on Music Perception and Cognition*, Bologna, 2006, pp. 960–966.

[253] Y. Heshmat, L. Yang, and C. Neustaedter, "Quality 'Alone' time through conversations and storytelling: Podcast listening behaviors and routines," in *Graphics Interface Conference 2018*, Toronto, 2018, pp. 67–74.

[254] L. G. Perks, J. S. Turner, and A. C. Tollison, "Podcast Uses and Gratifications Scale Development," *Journal of Broadcasting and Electronic Media*, vol. 63, no. 4, pp. 617–634, 2019.

[255] S. E. Baumgartner and W. Wiradhany, "Not all media multitasking is the same: The frequency of media multitasking depends on cognitive and affective characteristics of media combinations." *Psychology of Popular Media*, 2021.

[256] L. Ward, M. Paradis, B. Shirley, L. Russon, R. Moore, and R. Davies, "Casualty accessible and enhanced (A&E) audio: Trialling object-based accessible TV audio," in *147th Audio Engineering Society International Convention 2019*, New York, 2019, pp. 1–5.

[257] M. Ursu, D. Smith, J. Hook, S. Concannon, and J. Gray, "Authoring Interactive Fictional Stories in Object-Based Media (OBM)," in *IMX 2020 - Proceedings of the 2020 ACM International Conference on Interactive Media Experiences*, Barcelona, 2020, pp. 127–137.

[258] J. Hartcher-O'Brien, S. Soto-Faraco, and R. Adam, "Editorial: A matter of bottom-up or top-down processes: The role of attention in multisensory integration," *Frontiers in Integrative Neuroscience*, vol. 11, no. February, pp. 10–12, 2017.

[259] N. Huang and M. Elhilali, "Push-pull competition between bottom-up and top-down auditory attention to natural soundscapes," *eLife*, vol. 9, pp. 1–22, 2020.

[260] J. Salmi, T. Rinne, S. Koistinen, O. Salonen, and K. Alho, "Brain networks of bottom-up triggered and top-down controlled shifting of auditory attention," *Brain Research*, vol. 1286, pp. 155–164, 2009.

[261] A. Bidet-Caulet, L. Bottemanne, C. Fonteneau, M. H. Giard, and O. Bertrand, "Brain Dynamics of Distractibility: Interaction Between Top-Down and Bottom-Up Mechanisms of Auditory Attention," *Brain Topography*, vol. 28, no. 3, pp. 423–436, 2015.

[262] S. Kaplan, "The restorative benefits of nature: Toward an integrative framework," *Journal of Environmental Psychology*, vol. 15, no. 3, pp. 169–182, 1995.

[263] R. Schlieper, S. Preihs, and J. Peissig, "An open dataset of measured HRTFs perturbed by headphones," in *Audio Engineering Society - 152nd Convention*, The Hague, 2022.

[264] J. Cui, A. Behar, W. Wong, and H. Kunov, "Insertion loss testing of active noise reduction headsets using acoustic fixture," *Applied Acoustics*, vol. 64, no. 10, pp. 1011–1031, 2003.

[265] rtings, "Apple AirPods Pro (2nd generation) Headphones Review," 2022, [Online] Available at: https://www.rtings.com/headphones/reviews/apple/airpods-pro-2nd-generation-truly-wireless [Accessed 19/12/2022].

[266] rtings, "Bose QuietComfort Earbuds II Truly Wireless Headphones Review," 2022, [Online] Available at: https://www.rtings.com/headphones/reviews/bose/quietcomfort-earbuds-ii-truly-wireless [Accessed 19/12/2022].

[267] rtings, "Sony LinkBuds S Truly Wireless Headphones Review," 2022, [Online] Available at: https://www.rtings.com/headphones/reviews/sony/linkbuds-s-truly-wireless [Accessed 19/12/2022].

[268] J. Lancaster and J. Casali, "Reat and mire insertion loss comparison for eight headphones of various passive and anr designs, inclusive of nrr and spectral attenuation," *Grado Dept. of Industrial and Systems Engineering, Virginia Polytechnic Institute and State University*, 2004.

[269] R. B. Crabtree and A. Behar, "Measurement of hearing protector insertion loss at ultrasonic frequencies," *Applied Acoustics*, vol. 59, no. 3, pp. 287–299, 2000.

[270] C. Couvreur and I. C. Couvreur, "Implementation of a one-third-octave filter bank in Matlab," *Applied Acoustics*, pp. 0–12, 1997.

[271] J. J. Jamil, "One-third-octave filter bank, including additional inputs and outputs," 2011, [Online] Available at: https://www.mathworks.com/matlabcentral/fileexchange/30942-one-third-octave-filter-bank-including-additional-inputs-and-outputs [Accessed: 04/07/2023].

[272] REAPER, "REAPER Digital Audio Workstation," 2024, [Online] Available at: https://www.reaper.fm/ [Accessed: 13/11/2024].

[273] I. Veit and H. Sander, "Production of Spatially Limited "Diffuse" Sound Field in an Anechoic Room," *Journal of the Audio Engineering Society*, vol. 35, no. 3, pp. 138–143, 1987.

[274] I. E. Commission, "BS EN IEC 60318-7:2022 Electroacoustics — Simulators of human head and ear," 2022, [Online] Available at: https://knowledge.bsigroup.com/products/electroacoustics-simulators-of-human-head-and-ear-head-and-torso-simulator-for-the-measurement-of-sound-sources-close-to-the-ear [Accessed: 17/03/2025].

[275] M. R. Paurobally and J. Pan, "The mechanisms of passive ear defenders," *Applied Acoustics*, vol. 60, no. 3, pp. 293–311, 2000.

[276] A. Lie, M. Skogstad, H. A. Johannessen, T. Tynes, I. S. Mehlum, K. C. Nordby, B. Engdahl, and K. Tambs, "Occupational noise exposure and hearing: a systematic review," *International Archives of Occupational and Environmental Health*, vol. 89, no. 3, pp. 351–372, 2016.

[277] Apple, "Apple introduces AirPods 4 and the world's first all-in-one hearing health experience with AirPods Pro 2," 2024, [Online] Available at: https://www.apple.com/uk/newsroom/2024/09/apple-introduces-airpods-4-and-a-hearing-health-experience-with-airpods-pro-2/ [Accessed: 21/11/2024].

[278] S. Agrawal, S. Bech, K. de Moor, and S. Forchhammer, "Influence of Changes in Audio Spatialization on Immersion in Audiovisual Experiences," *AES: Journal of the Audio Engineering Society*, vol. 70, no. 10, pp. 810–823, 2022.

[279] W. Yang and J. Kang, "Acoustic comfort evaluation in urban open public spaces," *Applied Acoustics*, vol. 66, no. 2, pp. 211–229, 2005.

[280] S. E. Widen, S. Båsjö, C. Möller, and K. Kähäri, "Headphone listening habits and hearing thresholds in swedish adolescents," *Noise and Health*, vol. 19, no. 88, pp. 125–132, 2017.

[281] M. C. Green and D. Murphy, "EigenScape: A database of spatial acoustic scene recordings," *Applied Sciences*, vol. 7, no. 1204, pp. 1–14, 2017, [Online] Available at: http://www.mdpi.com/2076-3417/7/11/1204 [Accessed: 21/11/2024].

[282] S. Bech and N. Zacharov, *Perceptual Audio Evaluation - Theory, Method and Application*. New Jersey: Wiley, 2006.

[283] J. Francombe, T. Brookes, R. Mason, and F. Melchior, "Loudness matching multichannel audio programme material with listeners and predictive models," *139th Audio Engineering Society International Convention*, pp. 1–10, 2015.

[284] S. Agrawal, S. Bech, K. Berentsen, K. De Moor, and S. Forchhammer, "Method for subjective assessment of immersion in audiovisual experiences," *AES: Journal of the Audio Engineering Society*, vol. 69, no. 9, pp. 656–671, 2021.

[285] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*. L. Erlbaum Associates, 1988.

[286] Laerd Statistics, "Two-way mixed ANOVA using SPSS Statistics." 2015, [Online] Available at: https://statistics.laerd.com [Accessed: 17/03/24].

[287] P. Mair, R. Wilcox, and I. Patil, "WRS2: A Collection of Robust Statistical Methods," 2023, [Online] Available at: https://cran.r-project.org/package=WRS2 [Accessed: 17/03/24].

[288] N. Van Berkel, D. Ferreira, and V. Kostakos, "The experience sampling method on mobile devices," *ACM Computing Surveys*, vol. 50, no. 6, 2017.