Audio Engineering Society

# Conference Paper 1

Presented at the AES 5th International Conference on Audio for Virtual and Augmented Reality
2024 August 19–21, Redmond, WA, USA



# The effects of hear-through augmented environmental noise on listener engagement with speech-based media

Jay Harrison[1], Alan W Archer-Boyd[2], and Damian T Murphy[1]

[1]*Audiolab, School of Physics, Engineering and Technology, University of York, Genesis 6, Church Lane, Innovation Way, Heslington, York, YO10 5DQ, North Yorkshire, United Kingdom*
[2]*BBC Research and Development, The Lighthouse, White City Place, 201 Wood Lane, London, W12 7TQ, Greater London, United Kingdom*

Correspondence should be addressed to Jay Harrison (`jay.harrison@york.ac.uk`)

**ABSTRACT**

Adjustable hear-through in devices now affords users control over the extent to which their surroundings are augmented over media. However, it is not known how listeners' engagement with media is affected by environmental noise that is boosted by hear-through. This paper presents results of a listening experiment that investigates how noise influences engagement with speech-based media presented over earphones. The engagement measures used were; Immersion, Narrative Absorption, and Enjoyment, with Listening Effort and awareness of noise also assessed. The results suggested that boosting from hear-through provides listeners with a greater awareness of their surroundings without diminishing their engagement with speech-based media. Results also suggested that while most listeners experience lower engagement due to noise, some experience higher engagement.

## 1 Introduction

Since the advent of the first personal stereo device in 1979 (Sony Walkman TPS-L2) scholars have studied how listening to audio over headphones in public spaces shapes how individuals perceive and interact with their surroundings. The personal stereo was in many ways the first device that allowed immersion and an augmentation of reality using a portable device. In fact, the designers of the first Walkman were so concerned that listeners would be isolated from their surroundings that they included two headphone jacks 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. Fast forward to today and many mobile consumers of audio listen using devices that offer features such as head-tracked spatial audio, active noise cancellation (ANC) and hear-through that further alter the ways in which mobile media listeners' manage and perceive the relationship between inner and outer soundworlds. While some hear-through users seek full acoustic transparency, where perception of external sounds is ideally identical to open ear listening [1], other users desire an enhanced awareness of external sounds where parts of the frequency spectrum are boosted relative to the open ear. Many hearable devices now provide adjustable hear-through functionality [2], 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 blurring the boundaries between earphone and assistive hearing devices [3]. 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 [3, 4]. Recent research has also suggested that environmental noise can also affect listeners' perception of the media itself [5, 6, 7], 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.

### 1.1 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 within the passive occlusion of a traditional headphone listening experience. In the present study the authors were interested in speech-based mobile media such as podcasts or on-demand radio, therefore the first research question $RQ_1$ was "How does exposure to noise affect listener engagement with speech-based media over earphones?" $RQ_2$ was "How does boosting from hear-through processing in hearables mediate the effects of noise on engagement?". To answer both research questions, an experiment was designed to investigate how different aspects of listeners' engagement with a storytelling podcast was affected when listening under different noise conditions.

## 2 Methods & Materials

### 2.1 Experiment Design

The experiment employed 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 [8] [9] 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 [10], 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).

### 2.2 Story Stimuli

The following four $\sim$5 to 6 min stories were 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 min 8 s), *A Monkey Meets a Seal Meets a Monkey* by Matthew McArthur (4 min 58 s), and *I Am Batman* by Paul Davis (5 min 43 s). They were chosen as they had broad appeal, had been validated as being capable of eliciting absorption and enjoyment [10], and were of sufficient duration for the narrative to develop and allow the participant the opportunity to experience immersion [9]. 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 were recordings of a storyteller recounting a story to a live audience. Audience reactions (laughter, applause, etc) were audible in all stories. Permission was sought for use of all stories. Stimuli were 16 bit wav files normalised to -23 LUFS. In the rest of the paper the stories are anonymized and referred to as stimuli A to D.

### 2.3 Noise Conditions

There were 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 was 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 Sec. 2.5. For

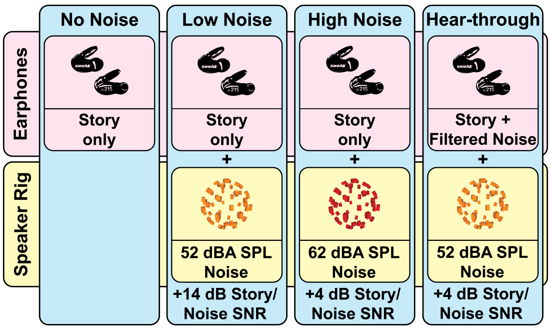


**Fig. 1:** Graphical summary of the noise conditions; *No Noise (NN)*, *Low Noise (LN)*, *High Noise (HN)*, and *Hear-through (HT)*. The *HN* & *HT* conditions had equal binaural loudness.

each stimuli, participants in the *NN* group were 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 Sec. 2.6) 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 were 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 was 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 was such that the total binaural loudness of the noise in the *HT* condition matched that of the *HN* condition. Further details on the *HT* condition are provided in Sec. 2.4. The speaker rig noise levels were selected based on typical sound pressure levels measured in urban open spaces [11]. 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 Fig. 2. A summary of the noise conditions is provided in Fig. 1. The stimuli was calibrated such that it was 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 was selected based on typical and safe earphone listening levels [12] and was the same for all conditions.

### 2.4 Hear-through Simulation Noise Condition

As mentioned in Sec. 2.3, participants in the *HT* group were simultaneously presented with noise over the speaker rig and earphones to simulate the boosting of environmental noise provided by hear-through processing in hearables. The design of the hear-through simulation transfer function was based on averaged measurements of the hear-through performance in the Apple Airpods Pro (2nd gen.), Bose QuietComfort Earbuds II, and Sony LinkBuds S. 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 Sec. 2.6. A target transfer function for the simulation (see Fig. 2) was 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 was produced using the following method: DPA 4560 CORE Binaural Headset Microphones were mounted on the enclosures of the Shure SE215 earphones, fitted in KEMAR's pinnae, 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 Fig. B.1 of the supplementary materials, accessible via the appendix. This was 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 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. The transfer function of the final *HT* condition is shown in Fig. 2. The same filter bank was then used to process the binaural soundscape noise recording that was 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 the supplementary materials.

### 2.5 Noise Condition Soundscape Material

Extracts of the *PedestrianZone2* recording from the Eigenscape database of spatial acoustic scene record-

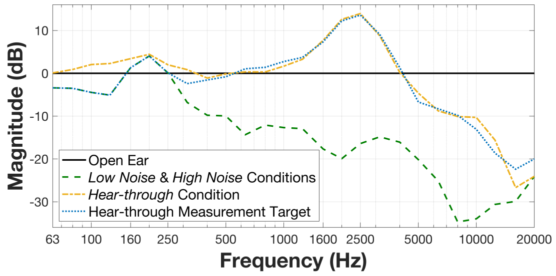


**Fig. 2:** Graph showing the transfer functions of noise conditions and the target measurement for the *Hear-through (HT)* condition produced using diffuse broadband pink noise.

ings [13] were selected for use as the environmental noise soundscape material for all the noise conditions in the experiment. The 10 min fourth-order Ambisonic *PedestrianZone2* recording was captured on Church Street, Liverpool using an em32 Eigen-Mike 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 s) in the *PedestrianZone2* recording, however as the story stimuli varied in length from 4 mins 58 s to 5 mins 43 s (see Sec 2.2), 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 min 22 s in the *PedestrianZone2* recording that may otherwise have biased the listening experience for the longest stimuli only.

### 2.6 Apparatus & Level Calibration

A pair of Shure SE215 In-Ear Sound Isolating Earphones were used for playback of stimuli and the hear-through simulation soundscape material as they provided similar passive noise occlusion to contemporary true wireless earbud devices. Earphone audio was played back from Reaper via a Ferrofish A32 AD/DA converter and a Samson S-phone headphone amp. The earphones were 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 were decoded in Reaper and played back over a 50 point Lebedev grid loudspeaker array via two Ferrofish A32 AD/DA converters. The array comprised of 40 Genelec 8030A and 10 Genelec 8040A active loudspeakers. Participants were seated such that their heads were positioned at the centre of the array, this is shown in Fig. C.1 of the supplementary materials. Participants used a keyboard and mouse to interact with a Matlab app that guided them through the experiment, triggered playback of audio with Reaper via OSC, and recorded responses to questions. A pen and five A4 sheets were used to record Immersion ratings. The speaker rig noise levels were 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 [14]. A modified (bypassed head-shadowing filter) ITU-R BS.1770-3 binaural loudness matching method [15] was used for calibration of the signal-to-noise ratio (SNR) of stimuli relative to noise. The binaural recordings were captured with prepolarized ear simulators (RA0045-S1) fitted in a GRAS 45BC KEMAR acoustic test fixture. KEMAR was equipped with large pinnae and was positioned at the centre of the speaker rig.

### 2.7 Measures

To assess participants' level of overall Immersion in each story, a method for assessing Immersion in audio-visual experiences was selected for use [8, 16]. This involved participants' rating their Immersion by making a single mark on a 15-cm graphic line scale (shown in Fig. A.1 of the supplementary materials), with reference points of "Not very immersive" and "Very immersive" at either end, and no markers between. Participants were 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 Herrmann and Johnsrude [10] was used. To assess Listening Effort for each story, the three Effort items from Herrmann and Johnsrude [10] were 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 1. An exit questionnaire was administered upon completion of participants' responses to the main stimuli rating questions. This was 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 answered yes, they were asked to describe these sounds, and were then asked if their awareness of any of the sound(s) influenced their rating of overall Immersion for any of the stories. If they answered yes, they were asked to describe how their Immersion ratings were influenced by their awareness of the sound(s).

### 2.8 Participants

A total of 80 participants took part in the experiment. Participants were each assigned to one of four noise condition groups randomly and such that there was an equal balance of genders in each group, as research [17] 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.

### 2.9 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 Fig. C.1 of the supplementary materials. Subjects were provided with an information sheet (see Figs. C.2-5 of the supplementary materials) that gave an overview of the study background and research aims, before providing their consent to participate. 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 a Immersion rating and responding to a single block of the Likert scale items, selected randomly for each participant. This was to familiarise 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 min interactive distractor tasks (see Fig. C.6 of the supplementary materials) 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 [8] 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 Sec. 2.7.

**Table 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. |

## 3 Data Analysis & 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 $\pm 3$. 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 $\pm 2$ [18], 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 [19]. 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.

### 3.1 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 Fig. 3.

### 3.2 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. 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 Fig. 3.

### 3.3 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

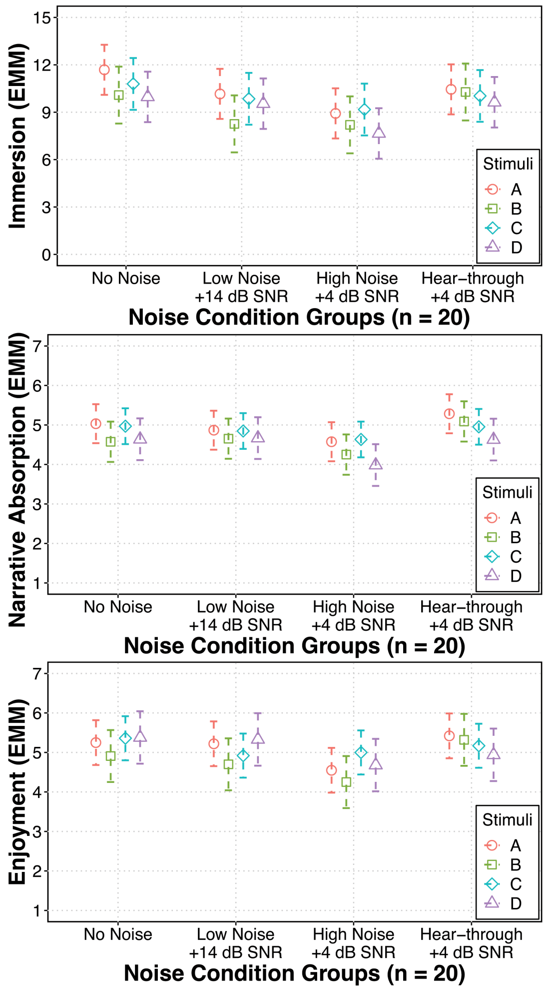


**Fig. 3:** Estimated marginal means of Immersion, Narrative Absorption, and Enjoyment scores for all stimuli at all noise conditions. Error bars represent 95% confidence intervals.

= .103, partial $\eta 2$ = .078. Estimated marginal means of Enjoyment for the four stimuli at every noise condition are shown in Fig. 3.

### 3.4 Listening Effort

There were two outliers (3.44, 3.14) and the data was not normally distributed, with skewness of 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

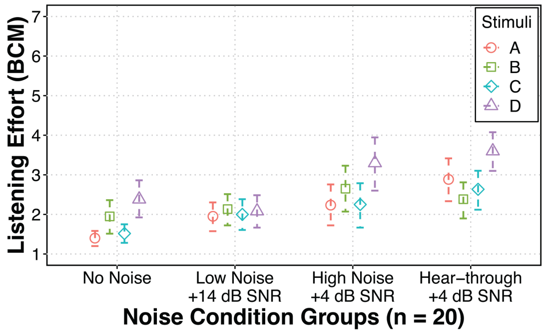


**Fig. 4:** Bootstrapped cell means of Listening Effort scores for all stimuli at all noise conditions. Error bars represent 95% confidence intervals.

*WRS2* R package [20] 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 Fig. 4.

### 3.5 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 Fig. 5. 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

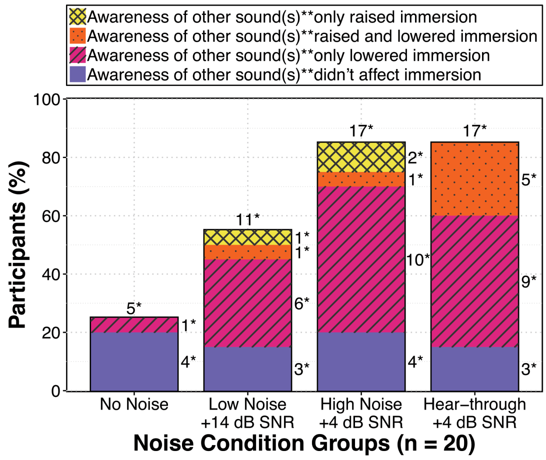


**Fig. 5:** Participants who were aware of other sounds and how their awareness influenced their Immersion. *Number of participants who reported an awareness of other sound(s) **Other sounds were any sounds aside from the storytellers or audience reactions.

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.

## 4 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 Fig. 3 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 Fig. 4, 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 Fig. 3 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 Fig. 4, 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 earbuds can provide listeners with an increased awareness of their surroundings, without compromising their engagement. These results suggest that boosting from hear-through processing in hearables provides listeners with a greater awareness of their surroundings without diminishing 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. [16], 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. Additionally, a wider variety of speech-based media stimuli and soundscape scenes should be assessed to increase generalizability, the ecological validity could be improved by providing subjects with a virtual reality representation of the soundscape scene, and $H_2$ is only relevant for hear-through processing that provides similar boosting to the measured devices.

## 5 Conclusion

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.

## Acknowledgments

This research was funded in part by the UK's Arts and Humanities Research Council (XR Stories project, Grant No. AH/S002839/1) and the University of York, in partnership with BBC Research and Development. The authors would like to thank Andrew Chadwick.

## Appendix

The supplementary materials for this paper can be found online at: https://t.ly/FdP30

## References

[1] Schepker, H., Denk, F., Kollmeier, B., and Doclo, K., “Acoustic Transparency in Hearables - Perceptual Sound Quality Evaluations,” *Journal of the Audio Engineering Society*, 68(7), pp. 495–507, 2020.

[2] Liski, J., Väänänen, R., Vesa, S., and Välimäki, V., “Adaptive equalization of acoustic transparency in an augmented-reality headset,” in *Proceedings of the AES International Conference on Headphone Technology*, pp. 1–8, 2016, ISBN 9781942220091.

[3] Gilmore, J. N., “Design for everyone: Apple AirPods and the mediation of accessibility,” *Critical Studies in Media Communication*, 36(5), pp. 482–494, 2019, ISSN 14795809, doi:10.1080/15295036.2019.1658885.

[4] McGill, M., Brewster, S., McGookin, D., and Wilson, G., “Acoustic Transparency and the Changing Soundscape of Auditory Mixed Reality,” in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–16, Honolulu, 2020, ISBN 9781450367080, doi:10.1145/3313831.3376702.

[5] Harrison, J., Archer-boyd, A. W., Francombe, J., Pike, C., and Murphy, D. T., “The relationship between environmental context and attentional engagement in podcast listening experiences,” *Frontiers in Psychology*, 2023.

[6] Nyre, L., “Urban Headphone Listening and the Situational Fit of Music, Radio and Podcasting,” *Journal of Radio and Audio Media*, 22(2), pp. 279–298, 2015, ISSN 19376537, doi:10.1080/19376529.2015.1083155.

[7] Wash, P. and Dance, S., “MP3 listening levels on London underground for music and speech,” *Applied Acoustics*, 74(6), pp. 850–855, 2013, ISSN 0003682X, doi:10.1016/j.apacoust.2012.12.008.

[8] Agrawal, S., Bech, S., de Moor, K., and Forchhammer, S., “Influence of Changes in Audio Spatialization on Immersion in Audiovisual Experiences,” *AES: Journal of the Audio Engineering Society*, 70(10), pp. 810–823, 2022, ISSN 15494950, doi:10.17743/jaes.2022.0034.

[9] Agrawal, S., Simon, A., Bech, S., BÆrentsen, K., and Forchhammer, S., “Defining immersion: Literature review and implications for research on audiovisual experiences,” *AES: Journal of the Audio Engineering Society*, 68(6), pp. 404–417, 2020, ISSN 15494950, doi:10.17743/jaes.2020.0039.

[10] Herrmann, B. and Johnsrude, I. S., “Absorption and Enjoyment During Listening to Acoustically Masked Stories,” *Trends in Hearing*, 24, 2020, ISSN 23312165, doi:10.1177/2331216520967850.

[11] Yang, W. and Kang, J., “Acoustic comfort evaluation in urban open public spaces,” *Applied Acoustics*, 66(2), pp. 211–229, 2005, ISSN 0003682X, doi:10.1016/j.apacoust.2004.07.011.

[12] Widen, S. E., Båsjö, S., Möller, C., and Kähäri, K., “Headphone listening habits and hearing thresholds in swedish adolescents,” *Noise and Health*, 19(88), pp. 125–132, 2017, ISSN 19984030, doi:10.4103/nah.NAH_65_16.

[13] Green, M. C. and Murphy, D., “EigenScape: A database of spatial acoustic scene recordings,” *Applied Sciences*, 7(1204), pp. 1–14, 2017, ISSN 20763417, doi:10.3390/app7111204.

[14] Bech, S. and Zacharov, N., *Perceptual Audio Evaluation - Theory, Method and Application*, Wiley, New Jersey, 2006, ISBN 978-0-470-86923-9, doi:10.1002/9780470869253.ch1.

[15] Francombe, J., Brookes, T., Mason, R., and Melchior, F., “Loudness matching multichannel audio programme material with listeners and predictive models,” *139th Audio Engineering Society International Convention, AES 2015*, pp. 1–10, 2015.

[16] Agrawal, S., Bech, S., Berentsen, K., De Moor, K., and Forchhammer, S., “Method for subjective assessment of immersion in audiovisual experiences,” *AES: Journal of the Audio Engineering Society*, 69(9), pp. 656–671, 2021, ISSN 15494950, doi:10.17743/jaes.2021.0013.

[17] Green, M. C., “Transportation Into Narrative Worlds: The Role of Prior Knowledge and Perceived Realism,” *Discourse Processes*, 38(2), pp. 247–266, 2004, ISSN 0163-853X, doi:10.1207/s15326950dp3802_5.

[18] Hair, J. F., Black, W. C., and Babin, B. J., *Multivariate Data Analysis: A Global Perspective*, Global Edition, Pearson Education, 2010, ISBN 9780135153093.

[19] Statistics, L., “Two-way mixed ANOVA using SPSS Statistics.” 2015, `http:https://statistics.laerd.com` [Accessed: 17/03/24].

[20] Mair, P., Wilcox, R., and Patil, I., “WRS2: A Collection of Robust Statistical Methods,” 2023, `https://cran.r-project.org/package=WRS2` [Accessed: 17/03/24].