EURASIP Journal on Audio, Speech, and Music Processing

EMPIRICAL RESEARCH **Open Access**

Quantifying headphone listening experience in virtual sound environments using distraction

Milap Rane[1*], Philip Coleman[1], Russell Mason[1] and Søren Bech[2,3]

Abstract

Headphones are commonly used in various environments including at home, outside and on public transport. However, the perception and modelling of the interaction of headphone audio and noisy environments is relatively unresearched. This work investigates the headphone listening experience in noisy environments using the perceptual attributes of distraction and quality of listening experience. A virtual sound environment was created to simulate real-world headphone listening, with variations in foreground sounds, background contexts and busyness, headphone media content and simulated active noise control. Listening tests were performed, where 15 listeners rated both distraction and quality of listening experience across 144 stimuli using a multiple-stimulus presentation. Listener scores were analysed and compared to a computational model of listener distraction. The distraction model was found to be a good predictor of the perceptual distraction rating, with a correlation of 0.888 and an RMSE of 13.4%, despite being developed to predict distraction in the context of audio-on-audio interference in sound zones. In addition, perceived distraction and quality of listening experience had a strong negative correlation of — 0.953. Furthermore, the busyness and type of the environment, headphone media, loudness of the foreground sound and active noise control on/off were significant factors in determining the distraction and quality of listening experience scores.

Keywords: Headphone rendering, Quality of listening experience, Distraction

1 Introduction

Headphones are a significant part of the daily life experience of most individuals; consumers in the USA and UK own 2.4 pairs of headphones on average [1]. It is a common sight to observe users wearing headphones in various contexts such as on the street, inside a bus, or in a cafe. The history of the sonic interaction of the user and the environment notes that listening technologies over the years have become more individualised [2], giving the user greater ability to adapt their auditory experience — the possibility of *personal soundscape curation* [3, 4]. This can be exhibited in a process as simple as choosing to wear or not wear headphones in a noisy environment, or being able to turn active noise cancellation (ANC) on or off.

Headphones are often used in noisy environments, which means there are two auditory components: the headphone media (the content being listened to on headphones) and the environment around the headphone user. These components interact and can impact the listening experience. Research is needed to better understand the perceptual effect of this interaction. Rämo et al. [5] used a model based on auditory masking to simulate how the ambient environmental noise could mask musical signals. This model was then used to improve equalisation for headphone listening in noisy environments [6]. Haas et al. [3] conducted a focus group with ten people to determine the main effects of this interaction, and used these in an online survey with a larger participant pool, identifying the need for users to be able to modify their perceived soundscape. Furthermore, both the headphone media and the environment sound are dynamic in the real world [3], which means the interaction and its perception is complex. A

*Correspondence: m.rane@surrey.ac.uk

[1] Institute of Sound Recording, University of Surrey, Guildford, UK
Full list of author information is available at the end of the article

real-time computational model of perceptually relevant factors that can track these varying conditions could assist in improving the overall listening experience.

This interaction of headphone media and environment could be considered to be a target-interferer paradigm, similar to the audio-on-audio interference experienced in sound zones (where there is target audio in the presence of interferer audio). Francombe et al. [7] elicited attributes to evaluate audio-on-audio interference in sound zones, discovering four most salient attributes: distraction, annoyance, balance and blend, and confusion. Using principal component analysis it was found that distraction accounted for the majority of the variance in describing the effect of audio-on-audio interference. Further work by Francombe et al. [8, 9] resulted in a perceptual model to predict this perceived distraction. Features selected to contribute to this model included the root mean square (RMS) level of the left ear of the target, the interference-related perceptual score (IPS) from the PEASS (Perceptual Evaluation for Audio Source Separation [10]) toolbox, and the loudness ratio of target and interferer (TIR). A real-time model based on this work was subsequently created by Rämo et al. [11, 12]. These perceptual models were created to predict distraction in audio-on-audio applications as opposed to the audio-on-noise situation of headphone listening in noisy environments. Whilst these can be considered similar in terms of the target-interferer paradigm and the possibility to separately measure the levels of each of these using headphones with externally mounted microphones (as used for ANC), research is needed to determine whether the existing models can be directly applied to the differing context and signal types (noise instead of audio as the interferer), or whether modification is required.

It is also essential to understand how the headphone-environment interaction also affects the overall quality of the listening experience (QoLE), an attribute that incorporates all factors critical to the individual assessor [13]. A significant factor in listeners' dissatisfaction with current headphones is insufficient noise cancellation [14]. Hence it is likely that the distraction caused by the headphone-environment interaction affects the QoLE. Research is therefore needed to determine the magnitude of this effect, and the relationship between perceived distraction and QoLE. If these attributes are closely related, then it is also possible that a perceptual model of distraction could meaningfully contribute to predictions of QoLE. A perceptual model that assists in the prediction of QoLE could be used as a design tool or to assist with real-time optimisation of headphone audio.

Therefore, this research aims to answer the following questions.

- How do the interactions of various environmental factors and headphone media properties affect the perceived distraction and quality of listening experience (QoLE) of participants?
- How does the perceived distraction relate to QoLE for headphone-environment interference?
- How well do distraction models developed for audio-on-audio interference predict the perceived distraction and QoLE for headphone-environment interference?

By answering these questions, this research contributes new understanding of the user experience of headphone listening in noise and the extent to which the attribute and model of distraction can describe this experience. With this understanding, a system that can gauge and respond to the perceptual effect of changes in environmental audio to the headphone listening experience can be envisaged.

The paper is organised as follows. Section 2 describes the experimental stimulus parameters, virtual sound environment, headphone media, and listening test design used to elicit ratings of perceived distraction and QoLE for a range of simulated headphone-environment stimuli. The results obtained from the listening test are then analysed and discussed in Section 3, before conclusions and potential future work directions are given in Section 4.

2 Methodology

In order to investigate the effect of the properties of the environment sound and headphone media on perceived distraction and quality of listening experience (QoLE), the experiment needed a method to create and reproduce environmental sound so that it, and its synchronisation with the headphone audio, was identical for each trial. The parameters used for the experiment are discussed in Section 2.1, the method of recording and reproducing the environmental sound is discussed in Section 2.2, the headphone media selected is discussed in Section 2.3, and the listening test design is discussed in Section 2.4.

2.1 Environmental sound and headphone media parameters

The first research question asks how the interactions of various environmental factors and headphone media properties affect the perceived effect. There are potentially an infinite number of combinations that could be included in the experiment, so previous research was used to determine the most important categories of environment sound and headphone audio that could affect the perceived result.

Research into urban sound quality by Maffiolo [15] indicated that environment sounds can be considered to be event sequences (isolated sounds that are processed based on their meaning or source), or textural/amorphous sequences (processed as a whole as they cannot be separated into individual events). Similarly, Guastavino [16] categorised environmental sound into source events and background noise. This separation of foreground events from background noise has subsequently been widely used (e.g. [17–20]).

This foreground-background categorisation was used as the basis for the construction of the experiment stimuli, together with an intended context to help with the plausibility of the simulated environment sound, as well as the headphone media selection.

- *Context*: the overall scene that dictates the audio content of the simulated environment that surrounds the headphone wearer [21], for example a street would have a different set of encountered sounds compared to a classroom.
- *External foreground elements (EF)*: the salient audio events that can affect the user's interaction with their surroundings, e.g. nearby talking or car/train announcements [15, 16]. Parameters of these that could affect the headphone listening experience include direction of arrival, loudness or sound pressure level (SPL), and spectrum and dynamic range [22].
- *External background elements (EB)*: the relatively continuous and less separately-identifiable components of the scene, that are likely to be more textural in nature [15, 16]. Relevant parameters include the average background noise levels and spectrum [22].
- *Headphone media (HM)*: the media playing in the headphones. The relevant parameters include: the loudness of the headphone media [23]; spectrum and dynamic range (which to a degree can be conflated with the genre or type of the content) [22]; and the stereo or spatial nature of the headphone media content.

The parameters chosen for the experiment were as follows. These are summarised in Table 1.

Table 1 List of parameters for the experiment along with the number of choices and the level names

Parameter	Number	Level names
Contexts	3	Home, street, public transport
EF object	1 (per context)	**Home**: Vaccum cleaner **Street**: Jackhammer **Public transport**: announcement
EB busyness	2	Busy, and not busy
EF distance	2	Close and distant
EF spatial positions	2	0° and 90°
H media program	3	Pop, Classical, and Radio Drama
H ANC	2	ANC on and ANC off

1 *Context and foreground elements*: Three contexts were selected (home, street, and public transport), as these are the most common environments for headphone usage [1, 3]. A single foreground element was selected for each context that would plausibly be a loud and close element in that scene (vacuum cleaner for home; jackhammer for street; and announcement for public transport).
2 *Busyness of background element*: It is known that the loudness of the environment has a significant effect on the headphone listening experience (e.g. [24, 25]). To manipulate the environment loudness in a plausible manner, the busyness of the background element was changed in each context as follows: home – empty vs busy kitchen; street — quiet vs busy; public transport — quiet train vs busy underground train.
3 *Environmental foreground loudness*: The foreground element will also have a significant effect on the loudness of the simulated environment, so this was changed independently of the busyness of the background element. To maintain the plausibility of the resulting scene, the foreground element varied in both loudness and distance, with the further being 6dB lower in level.
4 *Environmental foreground position*: It is known that the relative position of target and interferer sounds can have a significant effect on masking and intelligibility [26, 27]. To investigate the effect of this on headphone-environment interference, two foreground positions were used: front; and left.
5 *Active noise cancellation (ANC)*: ANC is highly ubiquitous today and significantly affects the headphone user's perception of the loudness of the environment [1, 3]. As noted above, the loudness of the environment has a significant effect, so this was included as an additional variable in the experiment (ANC on or off).
6 *Headphone media program*: Previous experiments have indicated that the properties of the headphone media also affect the listening experience, due to factors such as the loudness and dynamic range [24, 25]. Therefore three program items were selected with a range of loudness and dynamic range: pop, classical, and radio drama.

2.2 Virtual sound environment

For this experiment the environment sound needed to be consistent between identical trials, which required the capture and reproduction of sound scenes using a virtual sound environment (VSE). The goal was to generate realistic sounding environments to closely mimic the perception of real world environments for each context. Loudspeaker-based VSEs have been used previously for testing hearing aids [28, 29], and benefit from clear externalisation unlike some binaural simulations [30].

Based on Oreinos et al. [29], the created VSE combined features of higher order Ambisonics (HOA) and nearest loudspeaker (NLS) rendering. HOA was used to record and reproduce the background elements to give a homogeneous rendering of a real environment. NLS was used to reproduce the foreground elements, allowing clear variation of distance, perceived location and locatedness. Based on [29], the created VSE combined desirable features of higher order Ambisonics (HOA) and nearest loudspeaker (NLS) rendering. HOA (third-order) was used to record and reproduce the background elements. This allowed real-world spatial recordings to be used, giving a realistic background that directly included the spatial direction, distance and source width cues from the recorded scene. To increase the flexibility of the background rendering [32], NLS was used to reproduce the foreground elements. This ensured that the foreground elements could be varied independently of the HOA background content, while avoiding any timbral quality and localisation issues that may have arisen from spatial processing (to achieve panning and distance perception, for example).

2.2.1 Stimulus recording

The background elements were sourced from HOA recordings. The quiet and busy street recordings were taken from the Eigenscape dataset [31] (10 s excerpts from the files, QuietStreet1.wav, and BusyStreet8.wav in the dataset). The home recordings were made using a Zylia ZM-1 microphone at one of the co-author's homes. For the quiet scenario a quiet afternoon was chosen, and for the busy scenario the recording was made in the kitchen with the dishwasher, washing machine, as well as the sink on, indicative of a busy home atmosphere. The train recordings were made using a Zylia ZM-1 microphone: for the quiet scenario, recordings were made on a national train between two nearby stations; for the busy scenario, recordings were during the daytime on the Bakerloo line (one of the noisiest lines of the London Underground [32]).

The foreground element sounds (vacuum cleaner, jackhammer and railway announcement) were sourced from *freesound.org*. To make sure that these sounds acoustically blended into the scene, the foreground object signals were convolved with ambisonic reverberation of a similar background element: these reverberation impulse responses were sourced from the ARTE dataset [33]. The foreground element was reproduced from a single loudspeaker (using NLS), and the reverberation was added to all loudspeakers (using HOA).

The ANC was simulated by filtering the foreground and background elements to simulate the ANC of state-of-the-art headphones (Sony WH-1000XM4), using the target frequency dependent attenuations sourced from *rtings.com*[1]. The target response was matched visually using the 10-band graphical EQ provided in the Digital Audio Workstation Reaper[2], and used to filter all external environment sounds, i.e. both the environmental foreground and background. Even though the headphones used in the experiment had their own ANC, the aforementioned technique was used to ensure a streamlined and fully double-blind experimental protocol, avoiding participants changing hardware ANC settings during the experiment. While this approach is likely to outperform practical ANC implementations, which cannot converge fast enough to apply to transient sounds as effectively as steady state ones, our main goal was to gauge the overall effect of ANC on distraction and quality of listening experience independently of the ANC implementation on a specific product.

The VSE was created in a listening room that conforms to ITU-R BS 1116 containing a 22.2 loudspeaker setup, using 22 Genelec 8330A loudspeakers and 2 Genelec 7350A subwoofers [34]. Two additional loudspeakers were added for the close EF object positions, one in front, and one on the left side, 1 m from the listener's position. Acoustically transparent but visually opaque curtains surrounded the participant so that they were not influenced by being able to see the loudspeaker locations [35, 36].

2.2.2 Calibration and reproduction

To maintain the plausibility of the stimuli and the external validity of the experiment, the environmental sounds needed to be reproduced at a realistic sound pressure level (SPL). The literature on urban soundscapes indicates that the SPL of street scenes can vary from 55.8 to 95.0 dBA SPL [37–39]. Research into overground and underground trains indicate that the former are generally quieter, with a range of between 54.0 and 67.8 dBA SPL for overground trains [40–42] and between 75.1 and 104.5 dBA SPL for underground trains in London [32]. Home environments are often quieter, but can cover a range of between 35.0 and 85.0 dBA SPL, including the noisiest conditions with multiple appliances running in the kitchen [43–45].

[1] https://www.rtings.com/headphones/1-5/graph#16490/7981

[2] https://www.reaper.fm/

Table 2 Average SPL (in dBA) measurements of the VSE stimuli. The foreground element conditions are F - front position; S - side position; C - close and louder; D - distant and quieter

	dBA					
	Quiet street	**Busy street**	**Quiet train**	**Busy train**	**Quiet home**	**Busy home**
ANC off + FC	79.8	82.8	64.2	87.2	61.7	79.8
ANC off + SC	79.9	82.7	64.3	87.3	61.8	79.9
ANC off + FD	66.1	78.9	58.8	82.1	57.4	75.4
ANC off + SD	65.9	79.1	58.7	82.1	57.2	75.6
ANC on + FC	75.3	76.8	56.1	76.6	55.3	73.2
ANC on + SC	75.3	76.8	56.1	76.8	55.4	73.4
ANC on + FD	60.8	70.5	54.8	73.4	53	70.1
ANC on + SD	60.4	70.6	54.7	73.5	53.1	70.1

Table 3 Headphone media chosen to be used for the experiment, along with the exact timestamps

Genre	Artist	Media chosen	Time stamps	Sound level at ear (dBA)
Pop	Billie Eilish	Bad Guy	1:29–1:39	80
Classical	Gustav Mahler, Berlin Philharmonic and Claudio Abbado	Mahler: Symphony No. 8 in E-flat major "Alles Vergangliche"	2:45–2:55	80
Radio Drama	Eloise Whitmore, BBC	The Turning Forest	0:12–0:22	80

Table 2 presents the SPL measurements of the VSE (in dBA), averaged over the duration of each stimulus. These measurements were made using an NTI Acoustilyzer AL1 and miniSPL microphone placed at the approximate ear height of a participant at the centre of the listening room. It can be seen that the values for each condition without ANC applied fall with the target ranges outlined above.

2.3 Headphone media

The headphone media being listened to is an important part of the listening experience. The literature shows that the presence of environmental sound usually degrades the headphone listening experience, and users will often increase the loudness of the headphone media to compensate [24, 25]. However, the characteristics of the headphone media are likely to have an effect on the result.

Shimokura and Soeta [24] compared vocal and instrumental music, highlighting that extracts that were quieter or had larger dynamic range resulted in the participants increasing the headphone level more when listening in the presence of train noise. Wash and Dance [25] found that the level was increased more for speech than music when in the presence of underground train noise. Therefore, the headphone media programs selected covered both music and speech, and with varied dynamic range. The extracts chosen were pop (a wide frequency range and a small dynamic range); classical (densely orchestrated choral and orchestral performance with a wide frequency range and a large dynamic range); and radio drama (both speech and musical elements with a large dynamic range). Table 3 lists the excerpts chosen for the listening experiment. These excerpts were presented to the listeners at an average SPL of 80 dBA at each ear (measured using a KEMAR head simulator), based on a comfortable listening level and guidelines for safe listening of personal devices [46].

2.4 Listening test design

The stimuli presented to listeners consisted of simultaneous reproduction of the VSE (using a 22.2 arrangement plus 2 additional loudspeakers for the close foreground elements) and the headphone media (using Bang & Olufsen BeoPlay HX headphones). A full-factorial experiment was conducted using the factors shown in Table 1: 3 contexts, 2 states of EB busyness, 2 states of EF loudness, 2 EF spatial positions, 2 states of ANC, and 3 headphone media programs. This led to 144 combinations in total. Each stimulus was 10 s long, giving a minimum time to audition all stimuli of 1440 s or 24 min.

The listening tests were conducted using a multiple-stimulus grading test, with 10 stimuli per page. An explicit reference was included that was the pop track headphone media without added environmental noise. A hidden reference was included on each page, as was an anchor stimulus that was selected for having the highest

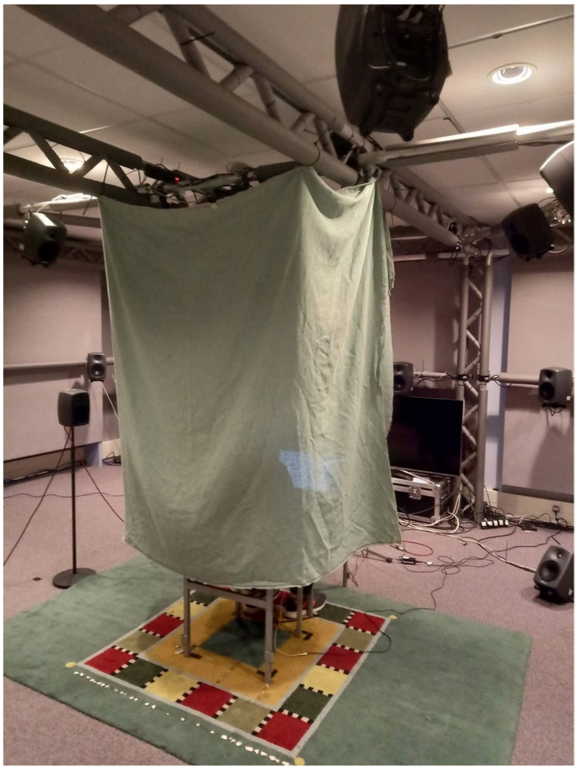


Fig. 1 Presentation of the VSE stimuli to the listener in the experiment

distraction model score. The other stimuli on each page were randomised across sliders and pages for each participant to reduce systematic bias effects caused by consistent co-location of stimuli (Fig. 1).

The participants were asked to rate distraction and QoLE in separate sessions completed on separate days; the order of these sessions was randomised. The participants were asked to imagine that they want to listen to the headphone media in the context evoked by the environmental sound. For distraction, the participants were asked: *how much does the external environmental audio distract you from your headphone listening experience?* The answer was scaled from 0 to 100, where 0 = *not at all distracting* and 100 = *overpowered*. For QoLE, the participants were asked: *how do the various environmental sounds involved impact the overall quality of the listening experience?* The answer was scaled from 0 to 100, where 0 = *low quality* and 100 = *high quality*.

The listening test interface was developed using HULTI-GEN version 2 [47, 48], and is shown in Fig. 2. Before each test the participants undertook a familiarisation stage. This allowed the participants to listen to all the headphone media in the listening test with a range of environmental sound from none to the loudest in the test. The participants were informed that if they found the familiarisation stage to be uncomfortably loud, they could withdraw from the rest of the test. The familiarisation also allowed the participants to become familiar with the user interface and the rating scale used in that session of the experiment.

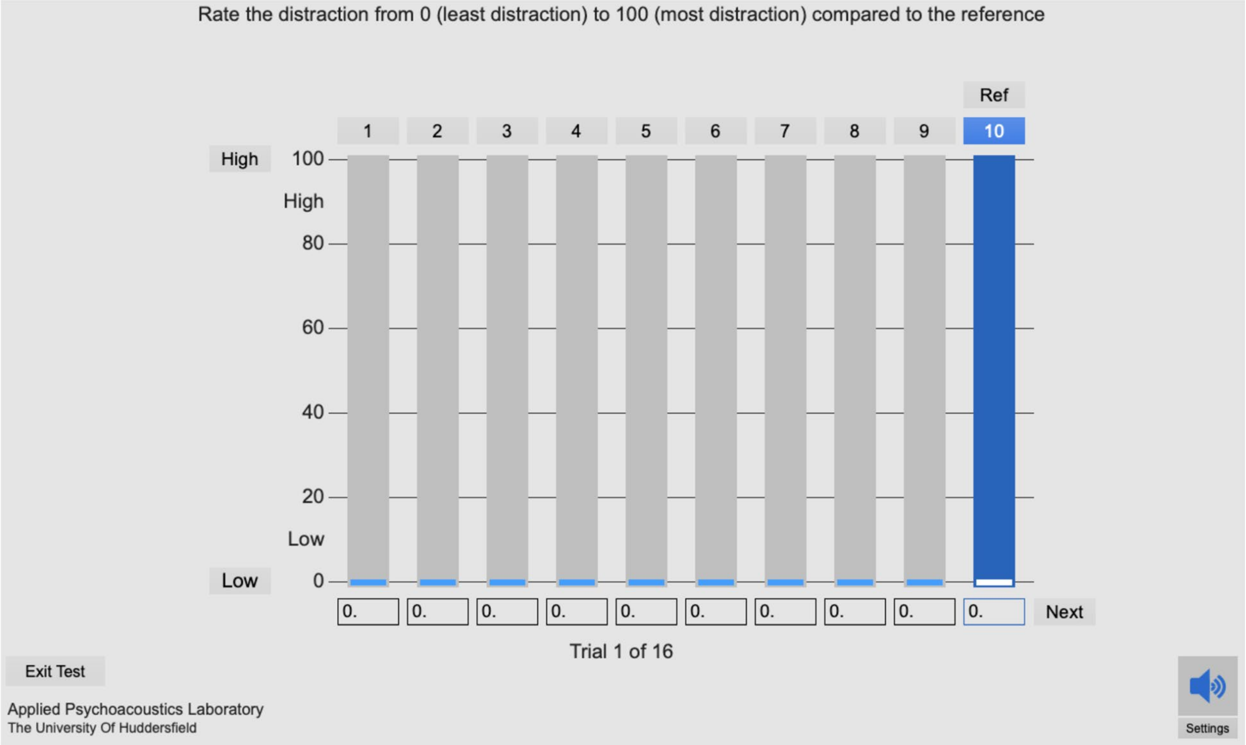
Fifteen participants undertook the experiment, and on average they took between 30 and 45 min to complete each session of the listening test. These participants were trained listeners who reported no significant hearing loss, aged between 18 and 30, mostly from the Institute of Sound Recording, University of Surrey.

3 Results

Using the methodology mentioned in the above section, 15 participants undertook the listening tests to gather both distraction and QoLE ratings. Informal conversations with the participants indicated that they found the stimuli to be realistic and that they had previously experienced similar headphone-environment interactions, especially the busier environments such as the busy train and street. They also felt that the different tasks — rating distraction or QoLE — made them focus on different parts of the listening experience. When they were grading distraction, they reported that they focused more on the source of distraction (the external environment), whereas when QoLE was being graded they reported that they focused on the headphone media itself and the overall experience.

3.1 Data pre-processing

Along with the distraction and QoLE ratings for the 144 stimuli with varying parameters, the listening test acquired multiple ratings for the reference and anchor stimuli, which could be used to judge the accuracy of the participants' understanding of the task presented to them. If there are errors in judging the reference by a participant, this could indicate that the participant has misunderstood the task. For QoLE, one subject was eliminated because they had an average error of more than 10 points from the expected value for the reference. For distraction, all participants correctly identified the reference; however, three participants were eliminated because they had an average error of more than 10 points across multiple ratings of the anchor stimulus. Based on these thresholds, the distraction ratings have 12 valid participants, and the QoLE ratings have 14 valid participants. The results for each stimulus were examined for normality of distribution using Kolmogrov-Smirnov tests, and it was found that all results apart from the reference and anchor had a normal distribution.



Rate the distraction from 0 (least distraction) to 100 (most distraction) compared to the reference

Trial 1 of 16

Applied Psychoacoustics Laboratory
The University Of Huddersfield

Fig. 2 Interface for the MUSHRA based grading of stimuli presented to listeners [47, 48]

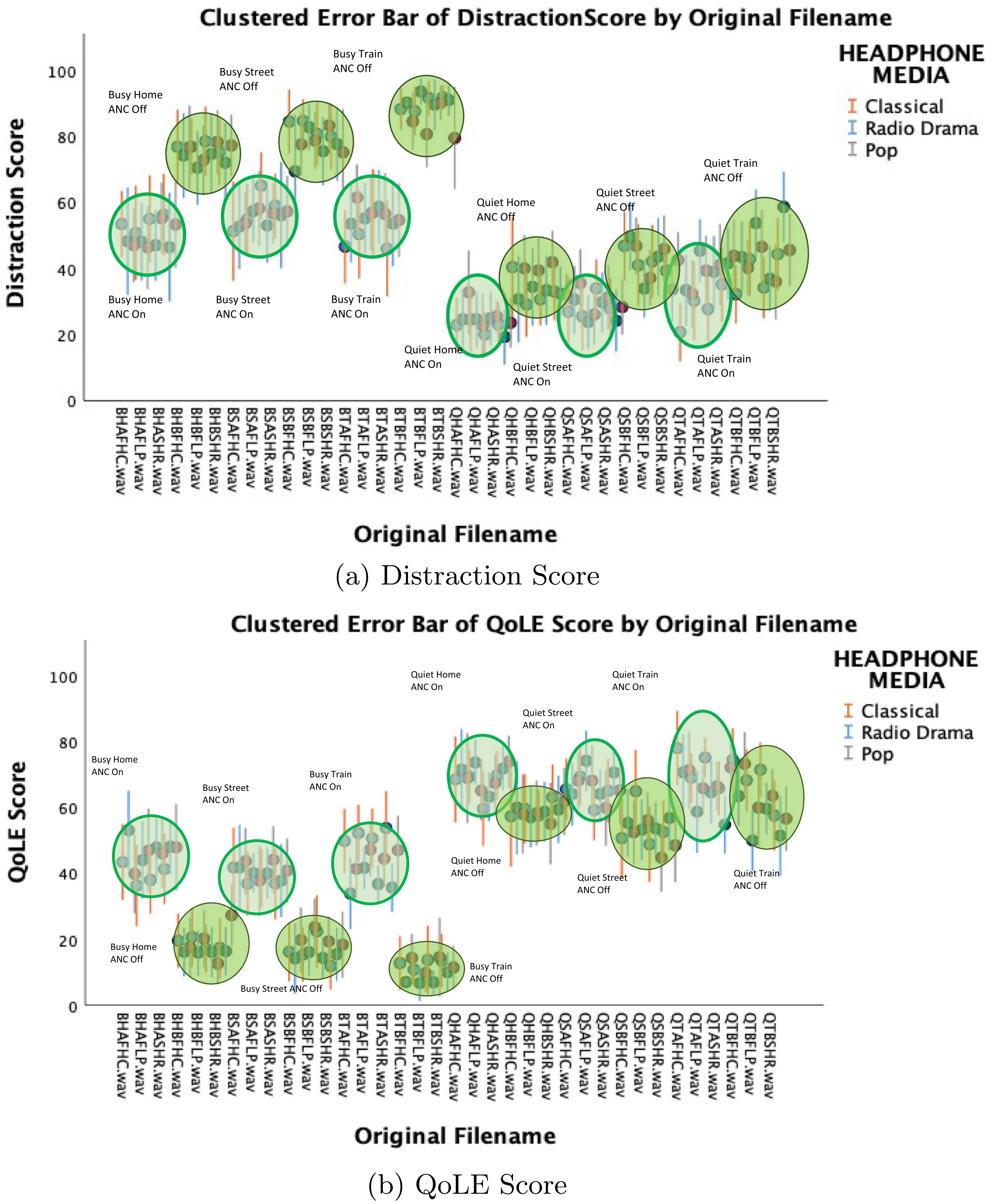
Table 4 The statistical significance and partial η^2 values resulting from the analysis of variance for the distraction scores and QoLE, where the statistical significance was <0.001

Source	Distraction Score Sig.	Distraction Score Partial η^2	QoLE Score Sig.	QoLE Score Partial η^2
Busyness	<0.001	0.454	<0.001	0.493
Context	<0.001	0.026	<0.001	0.008
ANC	<0.001	0.214	<0.001	0.217
Busyness * ANC	<0.001	0.052	<0.001	0.062
Context * HeadphoneMedia			<0.001	0.017
Busyness * Context * ANC	<0.001	0.011	<0.001	0.013
Busyness * Context * EF Loudness	<0.001	0.012	<0.001	0.010

3.2 Parameters affecting the distraction and quality of listening experience

The experiment examined the effect of parameters of the environmental audio and headphone media on the perceived distraction and QoLE. The parameters included context, busyness, foreground sound event spatial position and distance/loudness, ANC On/Off, and the headphone media program. A multivariate analysis of variance (ANOVA) was undertaken to determine which parameters had a significant effect. The factors and interactions with statistical significance <0.001 are shown in Table 4.

It can be seen that the parameters that had the largest effect on the results (based on the partial η^2 values) were the busyness of the environment, and the ANC being on or off. This effect can be observed in Fig. 3a and b. However, it can be seen from these figures, and the statistically significant interaction between busyness and ANC, that the ANC has less effect on the results for the quieter contexts. When the environment is quieter the ANC provides less advantage. The context had a much smaller effect, with the train stimuli being rated more distracting than the street and home stimuli in turn.



(a) Distraction Score

(b) QoLE Score

Fig. 3 Distraction and QoLE ratings across stimuli. This plot provides the variation of the distraction and QoLE ratings through all the stimuli, and labels have been provided to show the variations of the ratings across various stimuli and parameters. The green circle signifies ANC Off, and blue circle signifies ANC On

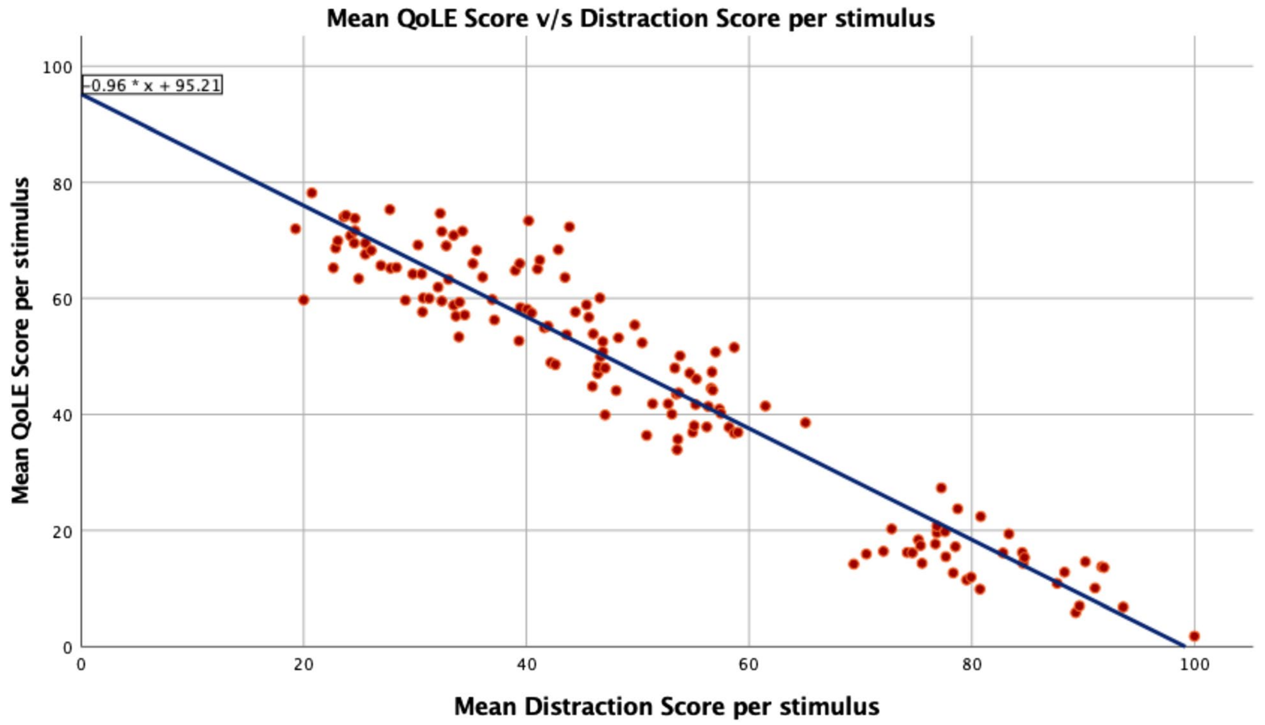


Fig. 4 Scatter plot of the QoLE Score with the distraction rating. The fit line is also visible, indicating that there is a negative, but linear correlation between the QoLE ratings and distraction ratings

There are two three-way interactions shown in Table 4. The interaction between busyness, context and ANC shows more detail about the two-way interaction mentioned above; investigation of post hoc tests indicated that the ANC had no statistically significant effect for the quiet train stimuli, again highlighting that ANC has less benefit in quieter situations. Examination of the post-hoc tests of the interaction between busyness, context and the foreground element loudness indicated that the foreground element had minimal effect for the busy scenes, and a much greater effect for the quiet scenes. This suggests that the characteristics of the foreground elements are more important when they are in relative isolation — in a busy scene their perceptual importance is reduced by the presence of loud background elements.

The headphone media only had a significant effect on the QoLE results as an interaction with the context. Examination of the post-hoc tests indicated that this predominantly occurred with the train context; the stimuli containing the pop program were rated as having lower distraction than the stimuli with either the classical or radio data programs. This may be explained by the pop track masking the environmental sound more effectively due to its lower dynamic range and wide frequency range.

Finally, the only parameter not to appear in Table 4 is the spatial position of the foreground element; it appears that this did not significantly affect either the distraction or QoLE ratings.

3.3 QoLE vs perceived distraction

The second research question asked about the relationship between the perceived distraction and QoLE, to give an indication of the importance of distraction on the overall listening experience of headphone media in environmental noise.

A scatter plot of the means of the distraction ratings against the QoLE ratings for each stimulus is shown in Fig. 4. It can be seen that these are highly negatively correlated in this experiment. Calculation of the Pearson correlation coefficient between the means of the ratings for each stimulus gave a result of $r = -0.953$. This result demonstrates a best-case scenario, where the parameters in the experiment were mostly selected because they were expected to affect the perceived distraction. It is not anticipated that there will be such a close relationship for all headphone-environment interactions, particularly with a wider range of headphone media. However, this result demonstrates that factors that affect perceived distraction can also have a significant effect on QoLE.

3.4 Distraction model vs distraction and QoLE scores

Each stimulus was recorded using a KEMAR head and torso simulator — the headphone media and the VSE both

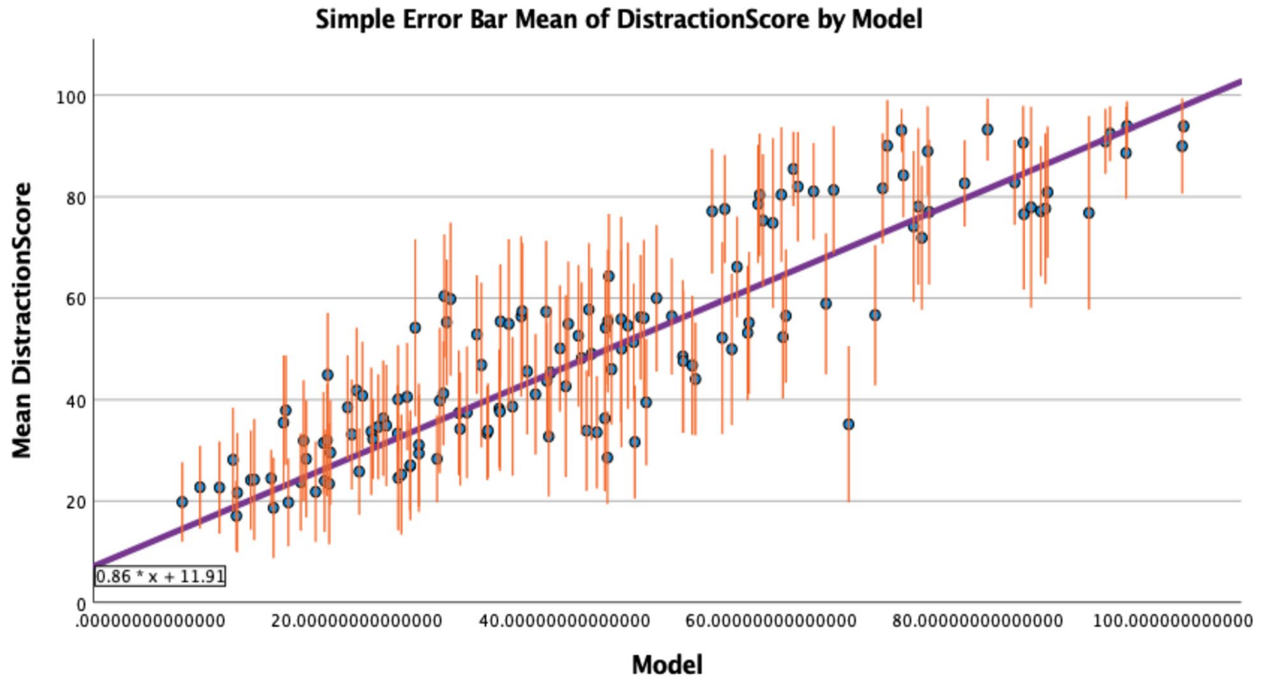


Fig. 5 Error Bar of the Distraction Score with the Distraction model. The fit line is also visible, indicating that there is a linear correlation between the distraction ratings and distraction model

separately and combined — as inputs to the real-time distraction model [11]. In this case the headphone media is viewed as the target and the VSE as the interferer; these were used to calculate the modelled distraction against which the distraction and QoLE ratings were compared.

Figure 5 shows the perceived distraction ratings (Distraction Score) against the modelled distraction (Distraction Model). It can be seen that there is a positive correlation between the distraction model and the obtained distraction ratings. A Pearson correlation undertaken between the model results and the means of the distraction ratings for each stimulus gave a result of $r = 0.888$.

The figure also contains a fit line, obtained using linear regression over all the distraction rating values to find a fit between the distraction model and the distraction rating. This shows that the distraction model results can be fitted to the distraction ratings using Eq. 1, where D is the mean distraction rating and M is the distraction model output:

$$D = 0.86 * M + 11.91 \quad (1)$$

Even without this fitting equation, the distraction model shows a good fit to the means of the distraction ratings in this experiment. The original distraction model achieved a root-mean-square error (RMSE) between the model results and the means of the subjective distraction ratings of 9.92% for the training stage, and for the three validation stages: 13.98%, 12.84%, and 19.27% [9]. The real-time version of the model achieved an RMSE of 10.2% and 12.6% for the two validation zones [11]. In comparison, for this experiment the real-time distraction model achieved an RMSE of 13.4% for the means of the distraction rating; this compares favourably to the validation RMSE scores of the original model. In addition, if the results are scaled using Eq. 1, the RMSE reduces to 10.12%, comparing favourably to both studies.

This indicates that the distraction model developed for perceptually optimised sound zones by Francombe et al. [8, 9] and modified for real-time calculation by Rämo et al. [11], gives useful predictions of perceived distraction for headphone listening in noisy conditions.

The distraction model results versus the QoLE ratings is shown in Fig. 6. As may be expected due to a high correlation between perceived distraction and QoLE, and between the distraction ratings and modelled distraction, the Pearson correlation coefficient between modelled distraction and the means of the QoLE ratings is also relatively high: $r = -0.875$. Once the QoLE ratings were transformed by subtracting each QoLE rating from 100, the RMSE between the modelled distraction and means of the QoLE ratings for each stimulus was 15.27%.

As mentioned above, this is a best-case scenario, with stimuli mostly selected because they are expected to

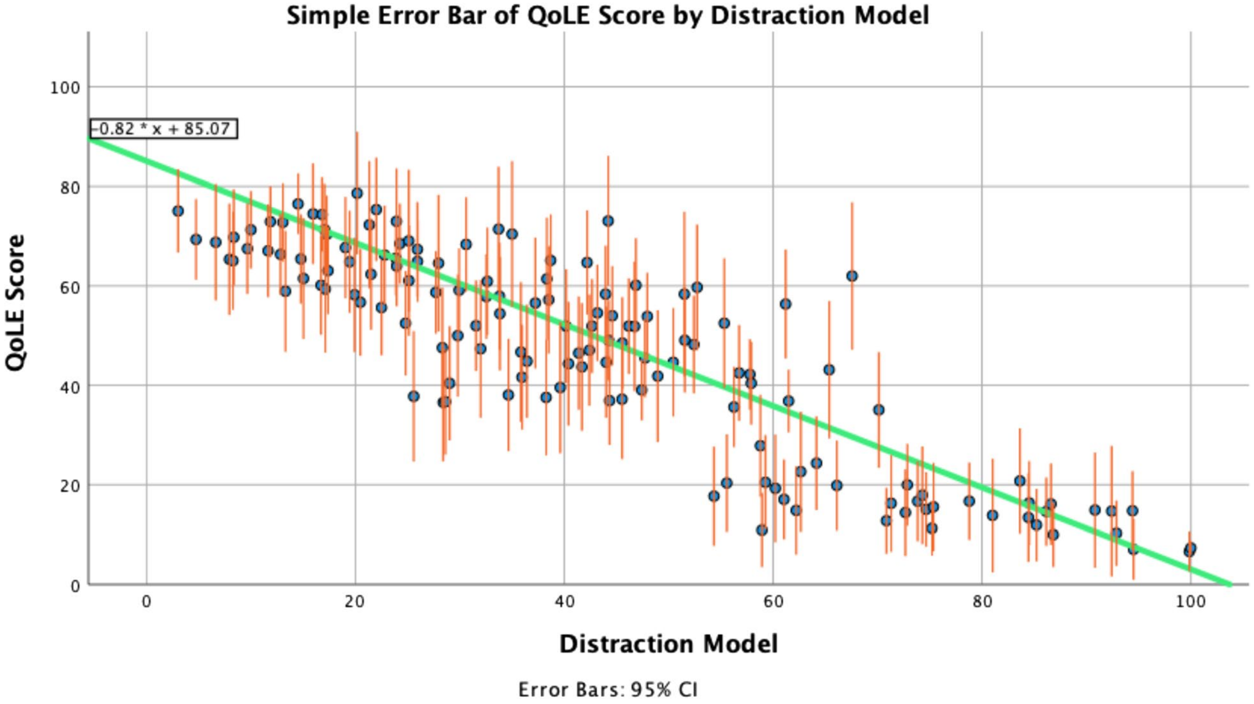


Fig. 6 Error Bar of the QoLE Score with the distraction model. The fit line is also visible, indicating that there is a negative, but linear correlation between the QoLE ratings and distraction model

affect the perceived distraction, hence there is a strong relationship between distraction and QoLE in this experiment. However, this result indicates that the distraction model could be used to predict the distraction-related elements that contribute to QoLE, and hence could contribute to a larger model of QoLE or be used to help optimise the distraction-related elements that could negatively affect the perceived quality of headphone media in the presence of environmental noise.

4 Conclusions and future work

The experiment described in this paper set out to investigate: the effect of headphone-environment audio interactions on perceived distraction and quality of listener experience (QoLE); the relationship between perceived distraction and QoLE; and the ability of distraction models to predict perceived distraction and QoLE. The results have demonstrated the following.

- Of the various parameters affecting the headphone-environment interaction and the listening experience, the busyness of the environment and the ANC on/off conditions have the largest effect. The context, headphone media, and foreground element loudness also had an effect in combination with the busyness and ANC settings. The spatial location of the foreground element was the only parameter that did not affect the distraction or QoLE ratings.
- Distraction and QoLE ratings in this experiment were highly negatively correlated, with a Pearson correlation coefficient between the means of these for each stimulus of $r = -0.953$. This is a best-case scenario given that the stimuli varied predominantly in a manner that affected the perceived distraction, but indicates that stimulus properties that affect perceived distraction can have a significant effect on the perceived overall quality.
- The distraction model results were fairly well correlated with the means of the distraction ratings ($r = 0.888$) and QoLE ($r = -0.875$). In addition, the RMSE between the modelled distraction and distraction ratings was 13.4%, similar to that of the validation experiments in the original distraction model literature. This indicates that even though the model was developed for audio-on-audio interference, it works well for headphone-environment audio interference. The RMSE between the modelled distraction and QoLE was higher at 15.27%, and as above this is a best-case scenario for QoLE but indicates that the distraction model could contribute to a larger model of QoLE or be used to help to optimise the distraction-related elements that affect perceived audio quality.

The future implications of this work indicate that the distraction model can be used to predict the perceived distraction and distraction-related elements of the quality of headphone-environment audio interaction, and could contribute to automated optimisation of these parameters. It would be possible to test and validate techniques to minimise distraction (e.g. equalisation [6]) using the distraction model, and hence use this to drive optimisation of the headphone listening experience in noisy conditions.

The results also indicate that the position of the foreground environment sources may not be important in influencing the distraction or QoLE. However, only two positions were used in this experiment; it would be pertinent to examine whether this is true for a wider range of positions. Furthermore, the directional nature of the ANC is another area that can be explored with regards to bringing the overall experience closer to real-life stimuli.

This test involves the headphone media being used as the target listening task. However, in a lot of cases, the environment might be an important target for the headphone user to focus on, especially in traffic/train situations, where certain stimuli (like vehicles or announcements) might require a user's attention. Previous work done on such situations have indicated how headphone listening can impact the user's perception of the environment, potentially even causing catastrophic results [49–51]. These results can be used to develop systems and models to gauge and improve the quality of headphone listening experience based on changes in the external environment. This area of research would be useful to investigate to understand how to optimise the overall experience of headphone listening in noisy environments.

Abbreviations

QoLE: Quality of listening experience; ANC: Active noise cancellation; BAQ: Basic audio quality; EB: Environmental background; EF: Environmental foreground; VSE: Virtual sound environment; HOA: Higher order Ambisonics; NLS: Nearest Loudspeaker; ARTE: Ambisonics Recordings of Typical Environments; MUSHRA: MUltiple Stimuli with Hidden Reference and Anchor; HULTI-GEN: Huddersfield Universal Listening Test Interface Generator; PEASS: Perceptual Evaluation for Audio Source Separation; IPS: Interference-related Perceptual Score; TIR: Target to Interferer Loudness Ratio; ANOVA: Analysis of variance; RMSE: Root mean square error.

Acknowledgements

The authors are grateful for discussions with the InHEAR project team in the development of the listening contexts/use-cases.

Authors' contributions

MR conducted the literature review, and prepared, designed and conducted the listening tests. MR also analysed the results from the listening tests and wrote the initial versions of the manuscripts. RM, PC and SB assisted in the design of the experiments, supported the analysis and reviewed early versions of the manuscript. RM edited the manuscript for publication. MR and RM conducted field recordings for the virtual sound environments. The authors read and approved the final manuscript.

Funding

This research forms part of the InHEAR project (https://iosr.uk/projects/InHEAR/), supported by the University of Surrey Doctoral College, Bang & Olufsen A/S, and the Danish Innovation Fund.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

This research was conducted in accordance with the University of Surrey's research and governance procedures.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Institute of Sound Recording, University of Surrey, Guildford, UK. [2]Department of Electronic Systems, Aalborg University, Aalborg, Denmark. [3]Bang & Olufsen A/S, 7600 Struer, Denmark.

Received: 17 April 2022 Accepted: 8 November 2022
Published online: 09 December 2022

References

1. N. Cooper, Hearables Report 2019 (White Paper, Audio Analytic, 2019). https://www.audioanalytic.com/hearables-report-thank-you/. Accessed 09 July 2020
2. V. Ris, The Environmentalization of space and listening. SoundEffects - Interdiscip. J. Sound Sound Experience 10(1), 158–172 (2021). https://doi.org/10.7146/se.v10i1.124204. Accessed 12 Mar 2021
3. G. Haas, E. Stemasov, E. Rukzio, in *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia* (ACM, Cairo Egypt, 2018), pp. 59–69. https://doi.org/10.1145/3282894.3282897. https://dl.acm.org/doi/10.1145/3282894.3282897. Accessed 03 Apr 2021
4. G. Haas, E. Stemasov, M. Rietzler, E. Rukzio, in *Proceedings of the 2020 ACM Designing Interactive Systems Conference*. Interactive auditory mediated reality: towards user-defined personal soundscapes (ACM, Eindhoven Netherlands, 2020), pp. 2035–2050. https://doi.org/10.1145/3357236.3395493. https://dl.acm.org/doi/10.1145/3357236.3395493. Accessed 02 Jan 2021
5. J. Rämö, V. Välimäki, M. Alanko, M. Tikander, in *Audio Engineering Society Conference: 45th International Conference: Applications of Time-Frequency Processing in Audio*. Perceptual frequency response simulator for music in noisy environments (Audio Engineering Society, Helsinki, 2012)
6. J. Rämö, V. Välimäki, M. Tikander, in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. Perceptual headphone equalization for mitigation of ambient noise (2013), pp. 724–728. https://doi.org/10.1109/ICASSP.2013.6637743
7. J. Francombe, R. Mason, M. Dewhirst, S. Bech, Elicitation of attributes for the evaluation of audio-on-audio interference. J. Acoust. Soc. Am. **136**(5), 2630–2641 (2014). https://doi.org/10.1121/1.4898053. Accessed 24 May 2021
8. J. Francombe, R. Mason, M. Dewhirst, S. Bech, Modelling listener distraction resulting from audio-on-audio interference. Proc. Meet. Acoust. **19**(1), 035036 (2013). https://doi.org/10.1121/1.4799636. Accessed 03 Feb 2022
9. J. Francombe, R. Mason, M. Dewhirst, S. Bech, A Model of Distraction in an Audio-on-Audio Interference Situation with Music Program Material. J. Audio Eng. Soc. **63**(1/2), 63–77 (2015). https://doi.org/10.17743/jaes.2015.0006. Accessed 02 Mar 2021
10. V. Emiya, E. Vincent, N. Harlander, V. Hohmann, Subjective and Objective Quality Assessment of Audio Source Separation. IEEE Trans. Audio Speech Lang. Process. **19**(7), 2046–2057 (2011). https://doi.org/10.1109/TASL.2011.2109381. Accessed 28 Mar 2022

11. J. Rämö, S. Bech, S.H. Jensen, Real-Time Perceptual Model for Distraction in Interfering Audio-on-Audio Scenarios. IEEE Sig. Process. Lett. **24**(10), 1448–1452 (2017). https://doi.org/10.1109/LSP.2017.2733084
12. J. Rämö, S. Bech, S.H. Jensen, Validating a real-time perceptual model predicting distraction caused by audio-on-audio interference. J. Acoust. Soc. Am. **144**(1), 153–163 (2018). https://doi.org/10.1121/1.5045321. Accessed 09 Feb 2022
13. M. Schoeffler, J. Herre, The relationship between basic audio quality and overall listening experience. J. Acoust. Soc. Am. **140**(3), 2101–2112 (2016). https://doi.org/10.1121/1.4963078. Accessed 12 Sep 2021
14. M. Rane, P. Coleman, R. Mason, S. Bech, in *Proceedings of the 152nd Convention of the Audio Engineering Society.* Survey of User Perspectives on Headphone Technology (Audio Engineering Society, Amsterdam, 2022). https://www.aes.org/e-lib/browse.cfm?elib=21669. Accessed 05 May 2022
15. V. Maffiolo, De la caractérisation sémantique et acoustique de la qualité sonore de l'environnement urbain. Semantic and acoustic characterization of urban environmental sound quality Ph. D. dissertation (Université du Maine, France, 1999)
16. C. Guastavino, The ideal urban soundscape: Investigating the sound quality of French cities. Acta Acustica U. Acustica **92**(6), 945–951 (2006)
17. G. Lafay, M. Lagrange, M. Rossignol, E. Benetos, A. Roebel, A Morphological Model for Simulating Acoustic Scenes and Its Application to Sound Event Detection. IEEE/ACM Trans. Audio Speech Lang. Process. **24**(10), 1854–1864 (2016). https://doi.org/10.1109/TASLP.2016.2587218
18. J. Salamon, D. MacConnell, M. Cartwright, P. Li, J.P. Bello, in *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA).* Scaper: A library for soundscape synthesis and augmentation (IEEE, New Paltz, 2017), pp. 344–348. https://doi.org/10.1109/WASPAA.2017.8170052. http://ieeexplore.ieee.org/document/8170052/. Accessed 28 Mar 2022
19. A. Taghipour, E. Pelizzari, Effects of Background Sounds on Annoyance Reaction to Foreground Sounds in Psychoacoustic Experiments in the Laboratory: Limits and Consequences. Appl. Sci. **9**(9), 1872 (2019). https://doi.org/10.3390/app9091872. Accessed 03 May 2021
20. M. Olvera, E. Vincent, R. Serizel, G. Gasso, in *EUSIPCO 2020 - 28th European Signal Processing Conference.* Foreground-Background Ambient Sound Scene Separation (Amsterdam / Virtual, Netherlands, 2021). https://hal.archives-ouvertes.fr/hal-02567542. Accessed 03 May 2021
21. A. Schmidt, K.A. Aidoo, A. Takaluoma, U. Tuomela, K. Van Laerhoven, W. Van de Velde, in *Handheld and Ubiquitous Computing*, vol. 1707, ed. by G. Goos, J. Hartmanis, J. van Leeuwen, H.W. Gellersen. Advanced Interaction in Context (Springer, Berlin Heidelberg, Berlin, Heidelberg, 1999), pp.89–101
22. T. Walton, M. Evans, D. Kirk, F. Melchior, in *Proceedings of the 141st Convention of the Audio Engineering Society.* Does Environmental Noise Influence Preference of Background-Foreground Audio Balance? (Audio Engineering Society, Los Angeles, 2016). https://www.aes.org/e-lib/inst/browse.cfm?elib=18441. Accessed 15 July 2022
23. G. Ruedl, E. Pocecco, M. Kopp, M. Burtscher, P. Zorowka, J. Seebacher, Impact of listening to music while wearing a ski helmet on sound source localization. J. Sci. Med. Sport. **22**, S7–S11 (2019). https://doi.org/10.1016/j.jsams.2018.09.234. Accessed 03 Mar 2021
24. R. Shimokura, Y. Soeta, Listening level of music through headphones in train car noise environments. J. Acoust. Soc. Am. **132**(3), 1407–1416 (2012). https://doi.org/10.1121/1.4740472. Accessed 07 Mar 2022
25. P. Wash, S. Dance, MP3 listening levels on London underground for music and speech. Appl. Acoust. **74**(6), 850–855 (2013). https://doi.org/10.1016/j.apacoust.2012.12.008. Accessed 04 Oct 2021
26. A. Bronkhorst, The Cocktail Party Phenomenon: A Review of Research on Speech Intelligibility in Multiple-Talker Conditions. Acta Acustica U. Acustica. **86**, 117–128 (2000)
27. M.L. Hawley, R.Y. Litovsky, J.F. Culling, The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer. J. Acoust. Soc. Am. **115**(2), 833–843 (2004). https://doi.org/10.1121/1.1639908. Accessed 11 July 2022
28. M. Nilsson, R.M. Ghent, V. Bray, R. Harris, Development of a Test Environment to Evaluate Performance of Modern Hearing Aid Features. J. Am. Acad. Audiol. **16**(1), 27–41 (2005). https://doi.org/10.3766/jaaa.16.1.4. Accessed 16 June 2020
29. C. Oreinos, J.M. Buchholz, Evaluation of Loudspeaker-Based Virtual Sound Environments for Testing Directional Hearing Aids. J. Am. Acad. Audiol. **27**(7), 541–556 (2016). https://doi.org/10.3766/jaaa.15094. Accessed 08 Mar 2020
30. P. Minnaar, S.F. Albeck, C.S. Simonsen, B. Søndersted, S.A.D. Oakley, J. Bennedbæk, in *Proceedings of the 135th Audio Engineering Society Convention.* Reproducing Real-Life Listening Situations in the Laboratory for Testing Hearing Aids (Audio Engineering Society, New York, 2013). https://www.aes.org/e-lib/inst/browse.cfm?elib=17001. Accessed 13 July 2022
31. Marc Green, Damian Murphy, EigenScape: A Database of Spatial Acoustic Scene Recordings. Appl. Sci. **7**(11), 1204 (2017). https://doi.org/10.3390/app7111204. Accessed 14 Dec 2020
32. T. Singh, T. Biggs, E. Crossley, M. Faoury, A. Mahmood, A. Salamat, T. Patterson, N. Jayakody, A. Dando, F. Sipaul, K. Marinakis, H. Sudhoff, P. Brown, Noise Exposure on the London Underground, an Observational Study over a Decade. Laryngoscope **130**(12), 2891–2895 (2020). https://doi.org/10.1002/lary.28547. Accessed 19 July 2021
33. A. Weisser, J.M. Buchholz, C. Oreinos, J. Badajoz-Davila, J. Galloway, T. Beechey, G. Keidser, The Ambisonic Recordings of Typical Environments (ARTE) Database. Acta Acustica U. Acustica. **105**(4), 695–713 (2019). https://doi.org/10.3813/AAA.919349. Accessed 15 June 2020
34. R. Mason, Installation of a Flexible 3D Audio Reproduction System into a Standardized Listening Room, in: Proceedings of the 140th Audio Engineering Society Convention. Presented at the Audio Engineering Society Convention 140, Audio Engineering Society, Paris, France (2016).
35. K.R. May, B.N. Walker, The effects of distractor sounds presented through bone conduction headphones on the localization of critical environmental sounds. Appl. Ergon. **61**, 144–158 (2017). https://doi.org/10.1016/j.apergo.2017.01.009. Accessed 02 Dec 2021
36. T.R. Letowski, S.T. Letowski, Auditory Spatial Perception: Auditory Localization. Final Report ARL-TR-6016, US Army Research Laboratory (Aberdeen Proving Ground, MD 2012)
37. M.R. Ismail, Sound preferences of the dense urban environment: Soundscape of Cairo. Front. Archit. Res. **3**(1), 55–68 (2014). https://doi.org/10.1016/j.foar.2013.10.002. Accessed 06 Feb 2021
38. T.P. McAlexander, R.R. Gershon, R.L. Neitzel, Street-level noise in an urban setting: assessment and contribution to personal exposure. Environ. Health **14**(1), 18 (2015). https://doi.org/10.1186/s12940-015-0006-y. Accessed 19 July 2021
39. C.S. De Silva, Private Sound Environments in Public Space: Use of Headphones in Public Parks and Public Transit. Ph.D. (New Jersey Institute of Technology, United States – New Jersey, 2021). ISBN: 9798516963162. https://www.proquest.com/docview/2556377960/abstract/F027EE7AF54D4D86PQ/1. Accessed 30 Sep 2021
40. J.Y. Hong, Y. Cha, J.Y. Jeon, Noise in the passenger cars of high-speed trains. J. Acoust. Soc. Am. **138**(6), 3513–3521 (2015). https://doi.org/10.1121/1.4936900. Accessed 30 July 2021
41. P.H. Trombetta Zannin, F. Bunn, Noise annoyance through railway traffic - a case study. J. Environ. Health Sci. Eng. **12**, 14 (2014). https://doi.org/10.1186/2052-336X-12-14. Accessed 19 July 2021
42. M. Němec, A. Danihelová, T. Gergeľ, M. Gejdoš, V. Ondrejka, Z. Danihelová, Measurement and Prediction of Railway Noise Case Study from Slovakia. Int. J. Environ. Res. Public Health. **17**(10), 3616 (2020). https://doi.org/10.3390/ijerph17103616. Accessed 19 July 2021
43. G. Jackson, H. Leventhall, Household appliance noise. Appl. Acoust. **8**(2), 101–118 (1975). https://doi.org/10.1016/0003-682X(75)90028-6. Accessed 30 July 2021
44. M. Fischer, B. Spessert, E. Emmerich, in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings.* Noise reduction measures of noisy kitchen devices and evidence of their improvement by an objective analysis of spontaneous EEG measurements, vol. 249 (Institute of Noise Control Engineering, Melbourne, 2014), pp. 340–347. Issue: 8
45. H.J. Lee, I.S. Jeong, Personal Listening Device Use Habits, Listening Belief, and Perceived Change in Hearing Among Adolescents. Asian Nurs. Res. 1976131721000025 (2021). https://doi.org/10.1016/j.anr.2021.01.001. Accessed 08 May 2021
46. World Health Organization, International Telecommunication Union, Safe listening devices and systems: a WHO-ITU standard (World Health Organization, Geneva, 2019). https://apps.who.int/iris/handle/10665/280085. Accessed 05 Oct 2021

47. C. Gribben, H. Lee, (Warsaw, Poland, 2015). http://www.aes.org/e-lib/browse.cfm?elib=17622. Accessed 28 Feb 2022
48. D. Johnson, H. Lee, in *Proceedings of the 149th Convention of the Audio Engineering Society*. Huddersfield universal listening test interface generator (HULTI-GEN) version 2 (Audio Engineering Society, Online, 2020)
49. R. Lichenstein, D.C. Smith, J.L. Ambrose, L.A. Moody, Headphone use and pedestrian injury and death in the United States: 2004–2011. Inj Prev. **18**(5), 287–290 (2012). https://doi.org/10.1136/injuryprev-2011-040161. Accessed 26 Feb 2021
50. J. Wachnicka, K. Kulesza, Does the Use of Cell Phones and Headphones at the Signalised Pedestrian Crossings Increase the Risk of Accident? Balt. J. Road Bridg. Eng. **15**(4), 96–108 (2020). https://doi.org/10.7250/bjrbe.2020-15.496. Accessed 26 Feb 2021
51. H.M. Lee, Z. Bai, Y.S. Ho, J.X. Soh, H.P. Lee, Effect of music from headphone on pedestrians. Appl. Acoust. **169**, 107,485 (2020). https://doi.org/10.1016/j.apacoust.2020.107485. Accessed 26 Feb 2021

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.