Acoustic Transparency and the Changing Soundscape of Auditory Mixed Reality

Mark McGill,
Stephen Brewster
University of Glasgow
Glasgow, Scotland, UK
first.last@glasgow.ac.uk

David McGookin
Aalto University
Espoo, Finland
davidmcgookin@gmail.com

Graham Wilson
Strathclyde University
Glasgow, Scotland, UK
graham.a.wilson@strath.ac.uk

ABSTRACT
Auditory headsets capable of actively or passively intermixing both real and virtual sounds are in-part *acoustically transparent*. This paper explores the consequences of acoustic transparency, both on the perception of virtual audio content, given the presence of a real-world auditory backdrop, and more broadly in facilitating a wearable, personal, private, always-available soundspace. We experimentally compare passive acoustically transparent, and active noise cancelling, orientation-tracked auditory headsets across a range of content types, both indoors and outdoors for validity. Our results show differences in terms of presence, realness and externalization for select content types. Via interviews and a survey, we discuss attitudes toward acoustic transparency (e.g. being perceived as safer), the potential shifts in audio usage that might be precipitated by adoption, and reflect on how such headsets and experiences fit within the area of Mixed Reality.

Author Keywords
Acoustic Transparency; Mixed Reality; Audio

CCS Concepts
•**Human-centered computing** → **Human computer interaction (HCI);**

INTRODUCTION
Headphones and earphones are a common presence in everyday life - when mobile, at work, or in the home, they allow for personal, private, discreet audio experiences. Until recently, their basic form had remained stable - one or two in-ear, on the ear (supra-aural) or over the ear (circumaural) speakers, capable of reproducing monaural and stereo sound. They are the gatekeepers to auditory awareness of our surrounding reality. Predominantly, headphones have been used to isolate us from reality, providing an auditory cocoon [38] by being either passively or actively (in the case of noise cancellation headphones) occlusive. Though acoustic isolation is not necessarily a desirable trait: we impair our ability to hear ourselves (e.g. footsteps, our own voice), as well as other necessary sounds, for example the train conductor asking for a ticket; being aware of cars when crossing the road.

It is however becoming increasingly evident that mobile consumer audio may be on the precipice of significant changes in form, function and capability. Headphones, Hearables [83, 24] and Earables [49, 110] capable of intermixing virtual audio with the sounds of reality are typically referred to as being *acoustically transparent*, with research such as Nomadic Radio [91] first exploring this concept through the use of wearable directional speakers. Consumer audio has rapidly begun to catch up with these research visions. For example, consider *passive* acoustic transparency, where the ear canal is sufficiently open to hear both virtual and real sounds concurrently. Bone conductance headsets are now consumer items and there are wearable headsets which rely on directional non-contact speakers to the same end, at high fidelity [19, 18]. Active noise cancellation headphones now commonly use their occlusive nature to their advantage by selectively filtering reality based on user preference or need, e.g. enhancing speech whilst dampening the surrounding environment [17, 103] - capable of selective *active* acoustic transparency of a sort. Some of these devices now support orientation sensing using IMUs, and consequently head-tracked spatialized auditory experiences, such as supporting the perception of externalised speech [43] to immersing listeners in spatialized content [33]. In the near future, users might find themselves surrounded by a *personal*, *private* (though shareable through software), *acoustically transparent*, *high fidelity* and *always-available* audio space. However, little is understood regarding the potential impact of acoustic transparency, both on the perception of auditory experiences, and on attitudes toward adoption and use. Now, virtual sounds have the potential to appear grounded spatially *and* acoustically in reality. What impact might this have on the perceived realism of sounds? And how might such wearable technologies change our day-to-day relationship with virtual audio?

This paper examines the impact of acoustic transparency on the perception of auditory experiences and attitudes toward use. A two-session, within subjects study (n=15) across both laboratory and in-the-wild settings compared acoustic transparency with occlusive noise cancellation headphones. We explored the perception of four types of audio content: a spatialized multi-speaker podcast; a spatialized immersive drama; a location-based audio tour guide and an affective

CHI '20, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6708-0/20/04 ...$15.00.
DOI: https://doi.org/10.1145/3313831.3376702

ambient soundscape. We address whether the intermixing of real and virtual sounds influence the user's perception of the virtual audio, and report on the impact on perceived realism/presence/safety of the experiences, user attitudes toward acoustic transparency, and the potential shifts in audio usage that adoption of acoustically transparent headsets might bring about. Finally, we discuss how we might better define and describe auditory Augmented Reality (AR) experiences, and reflect on how wearable, acoustically transparent, always-available audio spaces might impact the future of computing.

RELATED RESEARCH

(H)earables and Acoustically Transparent Headwear

From *Hearables* [83, 24] (ear-mounted wearable computer interfaces) to *Earables* [49, 110] (augmented earbuds) to *Intelligent Headsets* [114], wearable personal audio is undergoing significant changes, now capable of head orientation tracking, physiological sensing, contextual awareness, etc. [47, 72, 48]. Of particular note are *Acoustically transparent displays* which should "not cause audible modification to the surrounding sound" [84]. Bone conductance headphones have enabled open-ear acoustically transparent audio for years, however their fidelity has impeded uptake [108]. However, recent advances have led to affordable, consumer-oriented auditory headwear that can deliver audio with a high degree of fidelity in either a passive or actively acoustically transparent manner. In the case of passive examples, they typically use directional speakers integrated into glasses frames (e.g. Bose Frames [19], Amazon Echo Frames [36], Vue [109]) or wearables (e.g. Bose SoundWear [18], Amazon Echo Loop [36]). Alternatively, noise cancellation headphones/earphones, typically used for "acoustic cocooning" [14, 38], can also actively mimic acoustic transparency by using microphone arrays in the headset, or in the environment [95], to selectively incorporate real-world audio [101, 17, 27] in what has previously been referred to as mic-through augmented reality [56] using 'hear-through' [61] displays (e.g. AirPod Pro transparency mode [112]).

Prior research has typically examined acoustic transparency from the perspective of spatial audio perception e.g. accuracy discerning moving targets [57] (increased relative to standard headphones) and (poorer) localization accuracy [8, 9], or specific applications such as navigation [3]. However, little is known regarding the impact of acoustic transparency on the perception of virtual audio (e.g. in terms of realism), given the underlying visual *and* auditory grounding in reality. The consequence of acoustic transparency is that we can begin to envisage scenarios predicted in Ubiquitous Computing, for example a glasses wearer might find that with their next purchase, they gain not only corrected vision, but also a personal soundspace, without occluding the sounds of reality.

Advancements in Mobile Spatialization

Formats such as ambisonic audio [26, 39] define virtual spatialized audio experiences that are capable of *"simulating the filtering effect performed by the human head"* using head-related transfer functions (HRTFs [59, 86]), to create sounds *"so realistic many listeners struggle to differentiate between a real loudspeaker and a virtual loudspeaker"* [70]. Advances in spatialized audio and binaural rendering (see [94]) have been transposed to mobile platforms, driven by increasing realism in VR [94]. Spatial audio libraries have been developed by Google (Resonance audio) [87], Oculus [81] and Valve/Steam [105], complete with support for environmental reverberation [50], occlusion and directivity [52]. Combined with gaming engines such as Unity [102] and libraries such as Mapbox [60], mobile devices have all the tools necessary to render spatialized experiences based on any real-world outdoor location.

Auditory Augmented Reality: Intermixing with Reality

Mariette [63] described audio AR as *"simply the introduction of artificial sound material into the real world"*, with a range of characteristics defining different types of audio AR, such as the predominant sensory modality, the spatial characteristics of the sound, the means of presentation, and user tracking/mobility. Such a broad definition encapsulates many other terms and auditory experiences, from different forms of Spatial and Location Based Audio [63, 101, 12, 62, 89, 85, 68, 21] to Mediated Reality and Augmented Perception [99, 24, 107, 104, 100, 111]. Indeed, even the use of "mobile transistor radios or early portable tape players with headphones could both be understood as presenting an augmented reality" [63] as real and virtual content is combined in both cases.

Lindeman and Noma [55] classified AR experiences based on the "axis of mixing location". They talked about two audio AR techniques in particular: "microphone-hear-through" AR and "acoustic hear-through AR", similar to active/passive acoustic transparency, suggesting that "the appropriateness of a particular display will vary based on the application, the usage environment, and/or cost". They noted, in particular, the convergence of technology to a point where virtual audio could appear indistinguishable from, and interleaved with, the real world. Larsson *et al.* [53] discussed this in terms of the Mixed Reality (MR) continuum, noting that each modality could independently exist at a point along this continuum (each permutation of {Reality, AR, Augmented Virtuality (AV), Virtual Reality (VR)} for audio and visuals), hypothetically allowing for combinations such as auditory AR with visual reality, or auditory VR with visual AR, etc. Murray *et al.* suggested that open headphones allowing for a "normal perception of the naturally occurring background sounds would thus be extremely important to retain user's presence in the real world [and]... allow for simple mixing of real and virtual acoustic elements which of course may be a highly desirable functional demand of AR/MR applications" [76, 53]. Conversely, the use of earplugs can result in a heightened awareness of self and a sensation of disconnectedness from reality. If perceptual differences between noise cancelling acoustically opaque, and acoustically transparent, experiences were demonstrated, this would have implications for how we describe auditory experiences that may or may not be grounded in reality. The catch-all description of audio AR may not be sufficiently descriptive regarding the different types of experiences now possible. Such a reliance may also inhibit alignment of auditory research with visually-oriented research expanding upon MR [97].

Auditory Experiences

Presence and Immersive Audio Content

Presence refers to *"the psychological experience of 'being there"'* [25] (see Cummings and Bailenson [25, 80] for a summary), effectively the *"perceptual illusion of non-mediation"* [58]. It has long been understood that head tracking (and subsequently positional tracking) plays a significant role in the formulation of a sense of presence in VR. So too does the experience of sound, albeit to a lesser degree, with the presence of sound, and sound spatialization, significantly increasing presence in visually-oriented VR experiences [25]. Larsson *et al.* [53] noted key differences between the auditory and visual modalities in terms of presence, with the auditory system being "less accurate than the visual one in terms of spatial resolution" but "on the other hand [it] has the ability of providing us with spatial cues from the entire surrounding space at the same time... thus, sound should be able to induce both object presence ('the feeling of something being located at a certain place in relation to myself') and spatial presence ('the feeling of being inside or enveloped by a particular space')".

Larsson *et al.* discussed some of the parameters that contribute to auditory-induced presence, noting the impact of externalization [76] (that the sounds appear to emanate from the world, rather than from inside the listener's head) and localization [82] (that the sounds appear grounded in reality), spaciousness (having appropriate reverberation matching the real-world space), sound quality (in particular, bass content and sound pressure level [34]), and even prior expectations regarding the visual environment, with sounds that mis-match with expectations leading to less presence [22]. Audio is an important part of making a reality seem alive - an absence of an aural environment has been reported by deaf observers with acquired hearing loss as *"dead and lacking movement... [with] the world [taking] on a strange and unreal quality"* [37]. If we consider Lombard's definition of presence being a perceptual illusion that there is no mediating technology, newer auditory displays are arguably capable of doing just that - with head tracking and acoustic transparency, we may not be able to distinguish between real and virtual sounds.

Speech Audio

Speech audio has been assessed in terms of the clear benefits of spatialization [2, 43] and head tracking [28, 40, 15] in audio multi-party conferencing, particularly with respect to facilitating the *"cocktail party effect"* whereby a listener can focus on an individual conversation amongst multiple conversations. Notably, however, these benefits have never been transposed to podcast content, aside from the subtle stereo spatialization of voices employed by commercial productions e.g. in radio.

Location-Based Audio (LBA)

Given location sensing, we can also form *"relationships between the physical and digital"* by triggering different audio based on location [21], for example creating an auditory tour guide [10, 16]. Microsoft, in particular, have revisited this concept with their *"Soundscapes"* smartphone app [71], intended to help *"those with blindness or low vision to build a richer awareness of their surroundsings"* through Global Navigation Satellite System (GNSS)-based delivery of 3D audio cues tied to features in the real-world environment. Notably for 3D positional audio, the capability to deliver accurate GNSS-based experiences will improve significantly with new mobile phone chipsets capable of decimetre accuracy [7, 29].

Soundscapes and Affective Audio

Nonverbal, ambient, or environmental audio is a key contributor to experiencing a sense of place: *"birdsong might evoke early morning and with it a sense of calm; the sounds of cattle lowing create a mental picture of an open space"* [69]. Natural environments [42] and natural sounds have been examined in particular, often considered a pleasant component of an aural soundscape [20, 54]. Virtual natural soundscapes have similarly been shown as particularly positively affective [117, 116], beneficial for coping with stress [4], inducing calm [116], making unpleasant environments more acceptable [106], or urban environments more appealing [115]. Recent research has used sounds such as forests [116], birdsong and rain [13], fountains with moving water [4], and with wind and forest noises being particularly relaxing [117, 116]. The impact that acoustic transparency, and a consequent grounding in a real-world audio landscape, has on the perception of affective ambient content remains unknown.

STUDY: PERCEPTION OF ACOUSTIC TRANSPARENCY

This study investigates the perceptual differences between acoustically transparent and occlusive auditory headsets, and the issues around their adoption and usage for personal audio. Regarding perceptual differences, we explore the impact of indoor and outdoor locations on the headset types across a variety of content. Two experimental factors were defined:

Headset: The acoustically transparent Bose **Frames** [19]; and the acoustically opaque Bose NC700 **Headphones** [17], see Figure 1. Both had orientation tracking to allow for ego/exocentric spatial rendering, and represent the two ends of a continuum of the intermixing of reality;

Location: Indoors seated in a quiet lab, with no changing visual stimuli and little background noise; **Outdoors** where participants walked a ~400m outdoor route covering green space, a road with pavement, and a cobbled pedestrian street filled with bars and restaurants.

Bose headsets were used as they were both state-of-the-art in terms of fidelity, and provided a common API for accessing headset orientation - this study was not funded by Bose in any way. We do not address comparisons to other forms of wearable headset (e.g. neck worn [91], in-earphones that allow a degree of reality in) or speakers. Whilst these would be interesting comparisons, we chose to examine the extremes of this space to better understand the impact that transparency/noise cancellation has. For each combination of these two factors, we assessed four different audio types, chosen based on prior research and their suitability for being spatialized, such that audio sources would appear fixed in egocentric (or exocentric, for location based audio) positions around the listener.

Podcast A multi-speaker podcast. We used an episode of the BBC's"Infinite Monkey Cage" [44] which had five presenters, an ecologically valid scenario of the cocktail party effect. Presenters were hand-labelled in Audacity, and these labels used to move an AudioSource in our Unity scene to five equidistant spatial points (a pentagon, with each speaker placed approximately 1.5m away) arranged around the listener based on the current speaker;

Drama "The Vostok-K Incident" [33], an immersive spatialized audio drama by the BBC and others [41];

Ambient Two ambisonic clips selected based on prior research into sounds that positively alter the emotional response to a GNSS location: *Birdsong* [73] and *Waves on a beach* [74]. For *Indoors*, these were played sequentially. For *Outdoors* these were attached to specific locations in the world, see Figure 1;

Tour Audio tours are a common use of GNSS location-based audio. Three speech audio clips were recorded, describing Wikipedia entries for features along the cobbled pedestrian street in Figure 1. For *Indoors* these clips were played sequentially and placed spatially in front of the listener. For *Outdoors* playback was triggered based on proximity and with sounds placed on the landmarks they referred to. See the accompanying video for more details.

This led to 16 combinations of the factors: *Headset*, *Location* and *Content* ($2*2*4$). No analysis was planned across the *Content* factor as differences between content types was not a research question of interest. We treated this study as a two-factor analysis on *Headset* and *Location*, defining two research questions:

RQ1 Are audio experiences perceived differently when rendered on acoustically transparent versus occlusive headsets?

RQ2 To what extent does the real world backdrop influence the perception of the virtual sounds across the headsets?

Demographics and Measures

For practical reasons the study was split into two sessions within subjects: indoor then outdoor. Within each session, *Content* order was counter-balanced, with alternating *Headset* order. For the outdoor conditions, we combined the Ambient/Tour content into one lap per headset to reduce walking. We recruited 15 participants (10 female, 4 male, 1 non-binary, average age=26.8, Std.Dev=7.6) from University mailing lists, each paid £20 for taking part. The study was approved by our University ethics committee. After each condition, participants answered the following Likert-type questions:

Figure 1. ***Left:*** **The Bose Headphones (top) and Frames (bottom).** ***Right:*** **Map of the outdoor route. For the Ambient/Tour lap blue indicates** ***Ambient Birdsong*****; orange indicates** ***Ambient waves on a beach*****; yellow circles indicate** ***Tour*** **clips.**

Enjoyment: "I enjoyed listening to the audio using the given headwear"

Ease of Listening: "It was easy to pay attention to the audio content"

Audio Part of World: "The audio felt like it was part of the real world"

From the *Igroup Presence Questionnaire (IPQ)* [93]: We used a subset of the Involvement subscale and the Real subscale:
Awareness of Reality - INV1 "How aware were you of the real world surroundings";
Attention to Reality - INV3 "I still paid attention to the real environment";
Captivation - INV4 "I was completely captivated by the audio";
Realness - REAL1 "How real did the audio seem to you".

From Murray *et al.* [76] We used:
Presence in Reality "How was your feeling of being present in the real world environment affected by the audio/headwear";
Social Presence "How was your sense of being present amongst other people affected by the audio/headwear";
Active Environment "How was your sense of being present in an active, changing environment affected by the audio/headwear".

Externalization: "Sound appeared as coming from outside my head, rather than from inside my head", based on [11].

Spatial Realism: From Begault [11], "Please rate the realism of the spatial reproduction / your sensation of sounds coming from different directions."

Localization: "The sounds felt like they had an associated real-world location they were connected to/placed at" (for the Tour and Ambient content types only)

Relaxation: From Labbe [51], "Please rate your level of relaxation as a result of the ambient soundscapes" (for the Ambient content only)

For these questions, a repeated measures two-way *anovaBF* on $Headset * Location$ was performed using the *BayesFactor* package [75] following [79], with Bayes Factors reported, see [45] for interpretation. Bayes Factors between 3 to 20 constitute weak evidence between levels of the independent variables; 20+ constitutes strong evidence, a conservative interpretation as suggested by Kass and Raftery [46]. At the end of the study, participants were interviewed regarding their experience and whether they noted any differences between the Headsets. They then filled out an exit survey which asked about perceived **safety/security**, concerns regarding **overhearing** content, attitudes toward **adoption** and **consumption**, and preferences regarding the Headset used to **deliver** the audio.

RESULTS

Within Subjects Questionnaires

See Table 1 for a breakdown of results across factors.

Enjoyment: There was at-best weak evidence of an interaction effect that the Frames were more enjoyable indoors, and the headphones outdoors, across the Podcast, Drama and Tour content.

Ease of Attention: There was weak evidence that the headphones made it easier to attend to the Podcast, Ambient and Tour content. For the Drama content in particular, the interaction effect weakly suggests that the frames made it harder to attend to the content outdoors, and conversely that the headphones made it easier. The interactions for the Ambient/Tour content types appeared to suggest the frames were impacted outdoors in the same way.

	95% Confidence Intervals				Bayes Factor		
	Indoor		Outdoor				
Question	Frames	Headphones	Frames	Headphones	Headset	Location	Interaction
Podcast Content							
Enjoyment	[4.8, 5.8]	[4.4, 5.9]	[4.4, 5.7]	[4.9, 6.1]	1	1	5
Ease of Attention	[4.4, 5.8]	[5, 6]	[3.5, 4.9]	[4.8, 5.9]	14	2	1
Captivation	[3.6, 5]	[3.9, 5.4]	[2.6, 4.3]	[4, 5.4]	23	1	1
Audio Part of World	[3.3, 5.5]	[2.8, 5.2]	[2.2, 4.2]	[1.9, 4]	1	102	7
Spatial Realism	[4.6, 6]	[3.5, 5.3]	[3.3, 5.2]	[3.5, 5.5]	1	2	2
Externalization	[3.9, 5.4]	[2.5, 4.9]	[3.6, 5.4]	[2.5, 4.9]	448	1	11
Realness	[0.46, 2.7]	[0.76, 3]	[1.3, 2.8]	[0.98, 2.4]	1	1	7
Presence in Reality ∗	[1.7, 3.3]	[1.2, 3.2]	[2.1, 3.9]	[1.7, 3.3]	1	1	6
Social Presence ∗	[1.8,3.4]	[1.5,3.3]	[1.8, 3.6]	[1.3, 3.2]	2	1	11
Active Environment ∗	[2.3,4]	[1.7,3.6]	[2.1, 3.9]	[0.88, 2.7]	5	1	3
Awareness of Reality	[1.5, 3.1]	[2.2, 4.5]	[0.7, 2.8]	[2, 3.8]	68	1	4
Attention to Reality	[3.4, 4.9]	[2.4, 4.4]	[3.6, 5.4]	[2.8, 4.9]	3	1	6
Drama Content							
Enjoyment	[5, 6]	[5.2, 5.8]	[4.9, 5.7]	[4.8, 6]	1	1	11
Ease of Attention	[4.5, 5.8]	[4.7, 5.8]	[4.1, 5.4]	[4.4, 6.1]	2	1	7
Captivation	[3.7, 5.2]	[4, 5.2]	[3.9, 5.5]	[3.9, 5.3]	1	1	10
Audio Part of World	[3.6, 5.6]	[2.5, 4.9]	[3.5, 5.3]	[2.4, 4.6]	20	1	10
Spatial Realism	[4.1, 5.6]	[3.8, 5.5]	[4.6, 5.6]	[4, 5.5]	2	1	9
Externalization	[3.2, 5.2]	[2.6, 4.8]	[3.8, 5.6]	[2, 4.7]	10	1	6
Realness	[0.74, 2.7]	[1, 3.1]	[0.82, 2.0]	[0.91, 2.8]	2	1	7
Presence in Reality ∗	[1.9, 3.6]	[1.4, 3.2]	[2.2, 4.3]	[1.2, 3.5]	3	1	8
Social Presence ∗	[1.9, 3.7]	[0.89, 2.8]	[2.1, 4.1]	[1.4, 3.7]	2	1	5
Active Environment ∗	[2.5, 4.3]	[1.3, 3.2]	[2.3, 4.5]	[1.3, 3.7]	27	1	10
Awareness of Reality	[2.1, 3.5]	[2.6, 4.4]	[1.2, 2.7]	[2.2, 4.1]	8	1	3
Attention to Reality	[2.8, 4.4]	[2.2, 4.2]	[3.2, 5.0]	[2.9, 4.6]	1	2	6
Ambient Content							
Enjoyment	[4.7, 5.7]	[5.3, 5.9]	[5.3, 6.1]	[4.7, 5.8]	1	1	2
Ease of Attention	[4.6, 5.6]	[5.2, 5.9]	[4.4, 6]	[5.4, 6.1]	8	1	8
Captivation	[2.6, 4.3]	[3.7, 5.1]	[3.4, 4.8]	[4, 5.5]	7	1	3
Audio Part of World	[3.4, 5.4]	[2.7, 5.1]	[4.1, 5.8]	[3.4, 5.5]	1	1	5
Spatial Realism	[3.5, 5.3]	[3.7, 5.1]	[4.8, 5.9]	[4.2, 5.8]	1	27	8
Externalization	[3.8, 5.4]	[2.6, 5]	[4.2, 5.8]	[2.6, 4.9]	53	1	8
Realness	[0.92, 2.9]	[0.9, 3]	[0.44, 1.8]	[0.91, 2.6]	1	3	3
Presence in Reality ∗	[2.9, 4]	[1.6, 3.7]	[3.1, 4.6]	[1.6, 3.6]	46	1	9
Social Presence ∗	[2.9, 3.7]	[1.7, 3.5]	[2.4, 3.9]	[1.3, 3]	44	1	5
Active Environment ∗	[2.6, 3.9]	[1.9, 3.7]	[3.2, 4.5]	[1.9, 3.7]	16	1	4
Awareness of Reality	[1.5, 3.1]	[2.8, 4.5]	[0.92, 2.4]	[1.7, 3.7]	43	3	3
Attention to Reality	[3.5, 4.9]	[2.3, 4.3]	[4.6, 5.5]	[2.8, 5.2]	27	5	2
Localized	[2.4, 5]	[2.5, 4.8]	[3.4, 5.4]	[3.2, 5.5]	1	5	11
Relaxation	[4.1, 5.2]	[4.2, 5.2]	[4.5, 5.6]	[4.2, 5.6]	1	2	9
Tour Content							
Enjoyment	[3.9, 5.4]	[3.7, 5.4]	[3.9, 5.6]	[4.3, 5.6]	1	1	10
Ease of Attention	[3.2, 4.8]	[3.7, 5.5]	[3.3, 5.1]	[4.3, 5.6]	5	1	8
Captivation	[2, 3.9]	[2.6, 4.2]	[2.7, 4.5]	[2.9, 4.9]	1	2	6
Audio Part of World	[2.4, 4.1]	[1.6, 3.9]	[3.3, 4.9]	[3.1, 4.8]	1	32	6
Spatial Realism	[3.6, 5.4]	[3, 5]	[3.2, 4.8]	[2.9, 4.7]	1	1	6
Externalization	[3.5, 5.2]	[2.3, 4.8]	[3.4, 5.4]	[3, 4.9]	4	1	8
Realness	[1.9, 3.6]	[2.1, 4.1]	[1.5, 3]	[1.7, 3.4]	1	2	7
Presence in Reality ∗	[2.4, 4.1]	[1.9, 3.5]	[3, 4.6]	[1.8, 3.7]	9	1	6
Social Presence ∗	[2.5, 3.7]	[2.1, 3.6]	[2.8, 4.3]	[1.5, 3.2]	15	1	3
Active Environment ∗	[2.3, 3.6]	[1.9, 3.6]	[3.1, 4.4]	[1.8, 3.5]	3	1	3
Awareness of Reality	[1.2, 2.6]	[1.4, 3.2]	[0.58, 1.4]	[1.2, 2.9]	2	1	2
Attention to Reality	[4, 5.2]	[3.6, 4.9]	[4.1, 5.9]	[3, 5.2]	3	1	7
Localized	[2.9, 4.7]	[2.3, 4.4]	[3.4, 4.8]	[3.7, 5.1]	1	6	6

Table 1. Within-subjects questionnaire. Light green indicates $3 < Bayes factor < 20$ (weak evidence). Dark green indicates a $Bayes factor > 20$ (strong evidence). All questions were on a 7-point scale (0 to 6). For the majority, higher constitutes "more" of the measure e.g. more presence/realness/relaxing. For the Murray *et al.* questions (indicated by ∗) the scales ranged from much lower to much higher, meaning that a score of 2.5 was a neutral response.

Captivation: There was strong evidence for the Podcast content, and weak evidence for the Ambient content, that the headphones made the experience more captivating. There was weak evidence regarding interactions that suggest the frames were more captivating outdoors than indoors (Drama, Ambient), with the converse being true of the headphones.

Audio Part of World: There was strong evidence in favour of this for Podcast content indoors, with a weak interaction effect (with both headsets being perceived as less real outdoors, with the headphones decreasing to a greater degree); strong evidence in favour of acoustic transparency for the drama content, with a weak interaction effect (presented feeling less part of the world outdoors, again with the headphones affected to a greater degree); an interaction effect on the Ambient content, with the frames being perceived as more real outdoors; and strong evidence in favour of outdoor location for the Tour content along with a weak interaction effect with the headphones being less real indoors.

Spatial Realism (IPQ): For the Podcast and Tour content there was an interaction effect, with the frames exhibiting the highest realism indoors, and dropping off to the level of the headphones when outdoors. Conversely, for the Drama and Ambient content, there was an interaction effect suggesting the frames exhibited higher spatial realism outdoors. Regarding main effects, there was strong evidence in favour of the Ambient content when listened to outdoors. This is affirmed by the responses to the *Localized* question, which provides weak evidence to the same effect. Whilst tour content did not have the same spatial realism, it also had weak evidence in favour of location for localization. However, perceived realness had no meaningful evidence for main effects.

Externalization (IPQ): There was strong evidence across the majority of the content types that acoustic transparency positively affected the sense that sound was coming from outside the head, rather than from inside, particularly for the Podcast.

Realness (IPQ): There was an interaction effect for podcast content, with the frames increasing to a degree outdoors, but conversely weak evidence that the realness of the Tour content for the Frames decreased when outdoors. For Drama content outdoors, the frames marginally increased, with the headphones marginally decreased.

Presence in Reality: The Frames typically had increased presence outdoors, with the headphones featuring decreased presence outdoors. The headphones featured a diminished sense of being present in the real world environment in the case of the Ambient content (strong evidence) and Tour content (weak evidence).These effects were mirrored for *Social Presence*.

Active Environment: In contrast, when asked about their sense of being in an active, changing environment, there was strong evidence in favour of the acoustic transparency for the Drama content leading to a heightened sense, with weak evidence in favour of the frames for Podcast/Ambient content.

Awareness of Reality: Awareness of reality was strongest for the frames Podcast/Ambient content, with weak evidence for the frames in the Drama content, and none for the Tour content.

Attention to Reality: There was strong evidence for a main effect on headset, as well as weak evidence for interactions on many content types, with the frames typically exhibiting higher attention to reality outdoors.

Relaxation: There was a weak interaction for the Ambient content suggesting that the frames were more relaxing outdoors.

Exit Interviews

Loosely guided interviews were conducted at the end, regarding whether the experiences were different between the Frames / Headphones; whether their perception of the real world changed based on the headset type; whether their perception of the virtual sounds changed based on the headset type; and in what ways they would/would not want to use the acoustically transparent frames. Interviews were coded using Initial Coding, then grouped and sorted based on frequency and interest (see [90]), with representative excerpts quoted.

Perceived Realism of Content: Ten participants made comments to the effect that the Frames made some content appear more real, with different interpretations across content types. Seven participants (P1/2/4/6/10/12/13) commented on the bird song from the Ambient content type specifically:

> *P10:* The frames made it seem more like real and nice to listen to it whereas with the headphones it was more just intense birdsong.

For the voice tour/podcast, five participants (P4/5/8/9/15) commented that voices appeared either more real, or as if emanating from the environment:

> *P4:* But with the frames it did sound more like it was someone standing beside you talking rather than just on a headphone.

Five participants (P1/2/3/5/15) commented on environmental aspects of the Drama as appearing real using the Frames, e.g.:

> *P1:* It was a little bit disturbing when I thought I could hear thunder because the audio was so good... When I had the noise cancelling headset on I was very aware that the bird song sounded excellent and it was appropriate to the environment. But it wasn't real, it was just relaxing. Whereas when I was walking through there with the spectacles on I felt like it could be (real). I knew it wasn't real. But I didn't know the thunder wasn't real so it changed my perception of what was happening.

It should be noted that opinions were not clear cut in this aspect, and could even vary for the same participant across content types, with some appearing equally real across frames/headphones (P7/8/12/14).

Privacy / Impact on Others: A repeated concern regarding acoustic transparency (P1/3/4/5/10/15) was the potential for others to overhear the content, specifically provoked by the Frames' directional audio. Issues were raised regarding disturbing others on public transport (P1/4), use for conversations/private messaging that might be overheard (P3/5/15), and use for more sensitive media that others might find offensive (P10) or might reveal personal information e.g. listening to self-development podcasts (P15).

Awareness of Surroundings and Safety: It could be expected that noise cancellation would lead to an improved ability to focus by removing real-world auditory distractions and increasing isolation, and that was indeed noted for some (P2/4/9/10/12). However, the "in-the-wild" usage did reveal a subset of users for whom the opposite was the case (P1/2/4/13/15). When walking, the auditory isolation meant they felt they had to pay *more* visual attention to the real world to compensate for the lack of auditory cues:

> *P15:* I felt that with the noise cancelling headphones it was a lot more difficult to become aware of the surroundings so I noticed that instead of focusing on the audio I was having to exert extra effort towards the outside which was more difficult to do. So it made the audio experience less enjoyable because I remember it was kind of trying to suck me in, but I had to resist because as you as you saw there are a few instances with all these cars around. You don't want to bump into anyone or get run over.

Personal safety was repeatedly mentioned as a significant benefit of acoustic transparency either directly (P6/13) or as an implicit benefit in comparison to noise cancellation (P1/6/10/11/13/15), with pedestrian safety around cars, awareness of others (e.g. P11 noting they had been startled by someone they hadn't heard), and personal safety all mentioned as positive points for acoustic transparency:

> *P13:* I really needed to watch when I was walking because I can't really hear the cars passing through. It's really kind of dangerous for me walking on a road. *P5:* If I don't hear someone's behind me (that) is a bit scary

Exit Questionnaire

Attitudes to Acoustic Transparency as a Feature

As Figure 2 shows, participants almost universally agreed that acoustic transparency would be a key feature to consider in future headwear purchases, making it more likely that they would wear an audio device throughout the day and use audio-based applications. Using a headset such as the frames was also seen as preferable to smartphone speakers with respect to auditory notifications. However, based on the experience wearing the frames, just under half of participants were concerned about other's being able to overhear their audio.

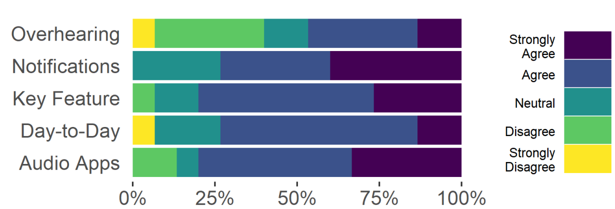

Figure 2. Responses to questions specific to acoustic transparency. *Overhearing:* "I would be concerned that others could hear what I was listening to"; *Notifications:* "It would be more socially acceptable to hear audio notifications over acoustically transparent headsets instead of from my smartphone"; *Key Feature:* "Acoustic transparency would be a key feature I would look for in future audio purchases"; *Day-to-Day:* "Acoustic transparency would make it likely that I would wear such an audio device throughout my day-to-day life"; '*Audio Apps:* "I would be more likely to use audio-based applications in my day-to-day life given an acoustically transparent headset"

Personal Safety

As can be seen in Figure 3, there was a difference in perceived personal safety when wearing the frames compared to the headphones, with just under half of participants believing their safety was compromised by the noise cancelling headphones.

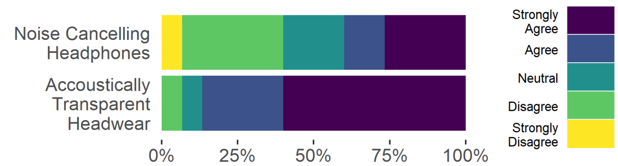

Figure 3. Responses to "I felt safe/secure listening to audio in public using the Headphones / Acoustically Transparent Headwear"

Changes in Usage Given Acoustically Transparent Headwear

We asked participants to consider whether they might be more or less likely to use a given audio type if they owned acoustically transparent headwear (see Figure 4). Media content (particularly ambient audio) were rated as more likely to be used by over half of participants, with just under half more likely to use voice assistants / engage in conversations.

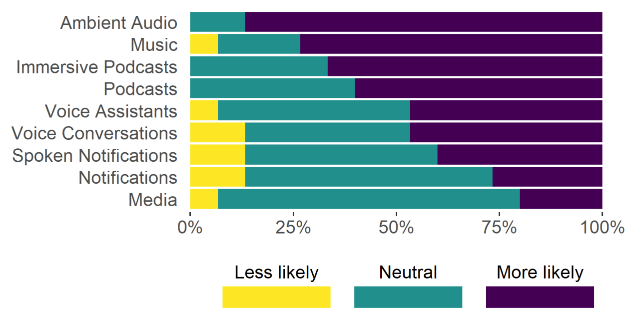

Figure 4. For each audio content type listed, participants were asked "whether owning and wearing an acoustically transparent headset throughout the day would change your likelihood of listening".

Acoustically Transparent Headwear versus Existing Devices

We explored the potential adoption of acoustically transparent headwear across public, workplace and private contexts, and content types (see Figure 5). Our aim was to examine if, based on this short exposure, participants might anticipate their habits or patterns of audio usage changing. Regarding the media/entertainment content, the noise cancelling headphones dominated responses, with participants favouring isolation and auditory fidelity. Regarding more functional audio, voice conversations and voice assistants showed the most anticipated adoption for acoustically transparent headwear, particularly in private settings. Given the previous responses regarding overhearing, we might anticipate that if user concerns in this regard were assuaged, the anticipated adoption for voice conversations and voice assistants in private (noted by approximately half of participants) might be transposed to more public settings. Interestingly, for notifications, over half of participants typically either wanted no audio, or to use speakers to deliver audio based on the context. There was opposition to general UI sounds, regardless of the headset type, perhaps indicative of the current landscape of mobile computing where smartphones are often on vibrate only, and UI audio is less utilized.

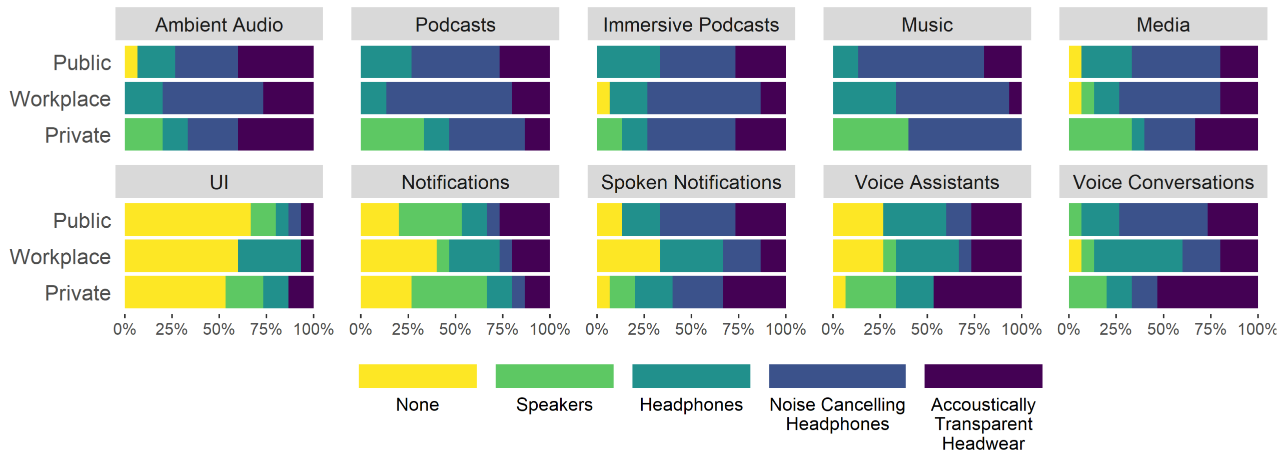

Figure 5. For each audio content type , participants were asked to indicate how they would prefer to listen to it in private (e.g. in the home or a car), in public (e.g. public transport, parks, restaurants) and in the workplace (e.g. office, library).

DISCUSSION

It is important to note that our interpretation should be considered against the following limitations. Firstly, we cannot separate visual/auditory influences on perception, nor can we account for additional multi-sensory, transient aspects that occurred during our study, for example changes in weather, environment, etc. Secondly, it is known that differences in audio quality (e.g. bass response) can influence perception (e.g. presence). Whilst the Bose Frames are high fidelity, they have less bass compared to the headphones. Whilst we considered constraining the rendering of the headphones to match (e.g. using filters), such an approach would be impossible for us to validate. Thirdly, the perception of audio may have been influenced by the latency of the headset tracking (in the region of 200+ms by our measure for both headsets). Fourthly, as content was repeated across both sessions, participants had prior knowledge of what was real and what was virtual, and this may have influenced responses. Finally, regarding the Ambient content in particular, we tested two ambisonic clips, one coherent with the existing physical setting (birdsong heard when walking by trees) and one contrary to the existing physical setting (crashing waves on a beach heard when walking along a city road), which in retrospect may have confounded responses to the questions regarding perceived realness, which covered both audio types in one set of 'Ambient' questions.

Acoustic Transparency...

...Can Alter Perception For Specific Content

Regarding RQ1, in select instances, participants noted that audio delivered via the acoustically transparent frames could appear indistinguishable from reality, particularly aided by the increase in perceiving audio as being externalized. For example, the Drama/Tour content was considered more part of the real world. However, regarding RQ2, these effects were often moderated by the real world auditory backdrop. For example, the Podcast content indoors was perceived as more part of reality than when outdoors, intermixed with a real world soundscape. Broadly, it would appear that for speech content in particular, the acoustic transparency helped in making speech appear more located in reality.

...Improves Awareness/Presence of Reality (Sometimes)

This point was perhaps expected, with increased awareness of reality when using the frames, and diminished presence in reality with the headphones. However, attention to reality featured at-best weak evidence of an effect on headset type - while participants were aware, they could seemingly still actively choose not to attend to reality when listening to immersive content. For the Drama content in particular, frames led to an increased sense of being in an active, changing environment, suggesting that the sounds of the Drama were in-part interpreted as being the real auditory environment.

...Appears Safer than Acoustically Opaque Headsets

Where awareness was facilitated, participants sense of safety was clearly improved. Noise cancellation headphones are an extreme point of comparison (actively occluding reality), but nonetheless they posed dangers in pedestrian settings (somewhat contrary with [3], which did not note any significant safety impediments in using music for pedestrian and cyclist navigation), with participants sometimes appearing oblivious to the dangers around them. Anecdotally, we can corroborate these responses, with the experimenter that accompanied participants throughout the study confirming there were a handful of instances where an intervention had to take place (i.e. stopping the participant or catching their attention) during the noise cancellation conditions to ensure participant safety.

...Could Encourage Voice Assistant Usage

Participants appeared broadly favourable to adoption of acoustically transparent headwear. Whilst these results reflect only exposure to Bose Frames (a glasses-centric device), it is clear that such an audio headset would impact audio consumption habits on a day-to-day basis, with ambient audio, music and podcasts all more likely to be consumed by the majority of participants. Interestingly, there also appears some promise for further adoption of voice-based usage, in part validating visions such as [91]. The scope of personal audio assistants might expand to operate in an always-available capacity, capable of privately interleaving feedback into real-world conversations/contexts [32]. This poses new challenges for audio assistant designs in the future e.g. at what point (and how) should agents be shared within a multi-user, multi-agent group context? Acer *et al.* recently suggested city-wide spatially constrained conversational agents [1] - the combination of head tracking, location-based audio and perceived externalization would appear ideally suited to such visions. However, whilst the speech of the assistant may be private, the voice interactions currently required are often not, which still presents a significant social acceptability barrier [88] to be overcome e.g. through non-voice or ingressive speech input [35].

...Will Require Careful Design Not To Overload Users
Whilst participants were broadly against usage for notifications and UI sounds, we would argue that this is a case where user expectations might be subverted, if suitably useful interactions can be designed. For UI sounds in particular, the *sound of silence* has dominated mobile and personal computing, given devices are often muted or set to vibrate. However, for notifications, we could envisage Audio Aura-esque [77] spatialized notifications that take into account the auditory context (being delivered only when no speech is detected) or physical context, such as encoding an association with a person (hearing the ping of the IM coming from a particular office) or action (hearing the calendar alert coming from the door to the office, to indicate you have to leave). Fundamentally, personal, private, wearable, always-on audio presents designers and practitioners with a powerful spatial audio canvas that was previously only infrequently available. *If* its merits can be demonstrated and encroachment into our day-to-day auditory experiences managed, users might yet be persuaded.

...Could Integrate Into A Variety of Day-to-Day Activities
Wearable acoustically transparent audio has significant potential to support and augment our day-to-day lives in a number of ways: quick "goto" auditory UIs; re-imagining interactive audio such as "filtears" [23]; extended perception; a supportive voice when stressed; private speech subtitling for those with hearing impairments; private speech support for those with visual impairments; personal audio for multi-view TV [67]; or even additional immersive spatial audio effects supplementing any display with surround sound capabilities.

...Has Form Factor-Specific Impediments to Adoption
This point is largely specific to the Bose Frames. However, participants were particularly concerned regarding the ability of bystanders to overhear the audio from the Frames. This impacted their willingness to use the Frames in public and workplace settings, with interviews suggesting that they would not use the Frames for sensitive or personal content (e.g. voice calls, self-development podcasts). Some also noted the glasses form factor as being a particular impediment, however this is of less concern given the variety of forms that might be envisaged e.g. the around-the-neck Bose SoundWear [18].

Acoustic Transparency and Auditory Mixed Reality
Our investigation into the effects of acoustic transparency suggests that the descriptor of *"auditory augmented reality"* does not fully encapsulate the differences in interpretation of content across contexts and headset types. We now have headworn auditory displays which have functional equivalency with audio-visual counterparts on the MR continuum in both rendering transparency/occlusion and orientation sensing. For *acoustic transparency*, i.e. a combination of real and virtual sound interleaved, see *AR headsets* such as the Microsoft Hololens. For *noise cancelling headwear*, i.e. excluding reality *in lieu* of virtual content, see *VR headsets* such as those from Oculus or Valve. We can even describe *augmented virtuality*, where a VR experience integrates reality virtually, such as using video pass-through cameras to view a keyboard [66], in this way - the auditory equivalent being the microphone pass-through of noise cancelling headphones.

However, when considering the MR continuum, the constant visual grounding in reality complicates the placement of auditory experiences, with the resulting interpretation of audio a function of headset type and real-world context, something predominantly absent in traditional visual-audio VR. For some users, hearing thunder on the acoustically transparent Frames in a dark, cloudy context lead to an AR-type experience, where the virtual thunder was interleaved with reality and perceived as real. However, equally, there were users that noted that the location-based birdsong appeared real on the noise cancelling headphones, despite the prior knowledge that they were blocking out the sounds of reality. This too could be described as an augmented reality experience, despite being rendered on the equivalent of an auditory VR headset. The visual grounding in reality, in conjunction with the audio, can play perceptual tricks across headset types. Equally, it is difficult to discern whether presence in an auditory virtual environment is possible given such visual grounding. Usage of frames *and* headphones could exhibit low awareness of reality, depending on the content being consumed, whilst suitably engaging content can exhibit attentional tunelling [113] regardless of headset type. Whether this constitutes an auditory presence of sorts however remains to be established, requiring further research exploring ego/exocentric rendering of content specifically, and some reformulation of existing presence questionnaires, which often focus on the perception of the underlying reality in a visually-oriented virtual experience. We would suggest the Vostok-K BBC drama [33] would be ideal for such explorations.

As well as experiential differences, these headsets also have notable functional differences, going beyond whether reality can be heard or not. Auditory headsets have the capability to alter reality, potentially in ways that might leave the user uncertain as to what is real or virtual, or even what is absent. Acoustically transparent headsets can effectively render on-top of the existing auditory landscape, interleaving virtual sounds with an unaltered reality. For example, we could detect real birdsong in a location, and introduce an additional virtual sound of a non-native bird which, for all intents and purposes, might be indistinguishable to the real birds. However, in contrast, a noise cancelling headset can perhaps more readily capture the real birdsong and alter, remove or entirely replace it, in effect, a distant relative of the 'Black Mirror' vision of erasing elements of reality [5, 65]. Here, even labels such as "auditory augmented reality" or "augmented virtuality" become insufficiently descriptive. Yet the prevailing interpretation in research has been to ascribe the label of *auditory augmented reality* to *all* such experiences - an attribution that appears incompatible with the new capabilities of auditory wearables. As such, for clarity we appropriate and re-define existing terms used for MR, and provide working definitions that might better assist in describing auditory MR experiences:

Auditory Headset A Hearable [83, 24] / Earable [49, 110] with orientation/position/location-based tracking, with content capable of being registered in 3D/rendered egocentrically (i.e. spatialized) and exocentrically (i.e. localized in the world). For clarity, we suggest that Hearable apply to all auditory headset wearables with integrated sensing, whilst Earable applies exclusively to those worn in the ear canal.

Auditory Mixed Reality As with the MR continuum, this encapsulates any auditory VR and AR experiences. Terms such as *Locative Audio* [12], *Personal, Location-Aware Spatial Audio* [62] and *Location Based Audio* [89, 85, 68, 21, 63] can apply anywhere within this continuum.

Auditory Virtual Reality Auditory experiences that are intended to supplant the existing reality or feature no meaningful contextual connection with reality, typically (but not exclusively) rendered on noise cancelling, {occlusive, opaque-to-reality, closed-ear} headsets.

Auditory Augmented Reality Auditory headset experiences intended to co-exist with/supplement reality or exploit spatial congruence with real-world elements, typically rendered on {acoustically transparent, hear-through [61], mic-through [56], open-ear} headsets. This encapsulates *Spatial Audio AR* [63] and *Augmented Audio Reality* [101], and is similar to Azuma's conceptualization of visually-oriented AR [6].

We also see close parallels to the AR taxonomy of Schraffenberger [92], allowing for *extended reality* (virtual supplementing real), *diminished reality* (virtual removing real), *altered reality* (virtual transforming real), *hybrid reality* (virtual completes real) and *extended perception*. In particular:

Auditory Extended Perception MR audio experiences intended to enhance reality. Also known as Superhearing [24], Auditory Contrast Enhancement [111], Enhanced Audio Perception [104] and Augmented Perception [107].

Auditory Altered Reality MR audio experiences intended to modify or overwrite reality. Also known as *Augmented Audio Reality*, and includes Auditory Mediated Reality [99], Warped Reality [100] and auditory equivalents of Substitutional Reality [96]. A notable example is Spence and Shankar's work altering the perception of food with sound [98], the auditory equivalent to MetaCookie+ [78].

These capabilities can utilize existing spatial information encoded in the captured real-world sounds, e.g. binaural pass-through microphones, and could be actively moderated based on user interaction or context (e.g. activating transparency when talking, and noise-cancelled diminished reality when listening to music). Our working definitions are more descriptive than previous ones [31], assisting us in describing experiences on a new generation of audio headsets that incorporate orientation/position tracking, and perhaps SLAM sensing [30] in the future. There are, however, edge cases that are not directly considered here. It could be suggested that all auditory AR constitutes *altered reality* insofar as virtual audio interleaved with real audio might alter the perception of reality. However, we differentiate between the implicit and explicit, with the intent of the application/designer guiding usage of descriptors.

A Personal, Private, Always-Available Soundspace

Auditory MR headsets have the capacity to infiltrate our everyday lives in ways that visually-oriented AR and VR headsets have yet to fully accomplish, be it for reasons of cost, technological capability or social acceptability. Take the Bose Frames: the auditory modality has been integrated into an affordable form factor that could be worn throughout the day, at a higher fidelity than prior bone conductance headphones, without occluding the sounds of reality. In effect, a glasses wearer might find that with their next purchase they gain not only corrected vision, but also a personal, private, spatialized, always available soundspace. Bose were the first to market, but others have since followed suit on this concept [109, 36]. Consequently, auditory MR headsets are likely find their way into our lives well before more costly, bulky, socially unacceptable visually-oriented AR headsets do. For auditory experiences more generally, an IMU represents a low-cost addition with potential to transform existing audio experiences such as podcasts in terms of perceived externalization, in much the same way that VR 360 content provided a new, immersive, engaging way of consuming visual media. Mark Weiser discussed the concepts of invisible computing and *embodied virtuality* [64], with the aim of having technology woven into the fabric of everyday life. Auditory MR headsets have properties which suggest they could be an endpoint for such a vision. From embedding a conversational virtual assistant into your ear, changing your perception of an environment, providing spatial auditory prompts and reminders, to enhancing externalization in auditory experiences such as conversations or podcasts, an always available, private, orientation tracked soundspace provides many new possibilities for HCI and personal computing.

CONCLUSION

This paper has explored the potential impact acoustic transparency might have on personal audio experiences. In a comparison of acoustically transparent and noise cancelling auditory headsets, across both outdoor and indoor settings and a variety of content types, we found differences in terms of presence, realness and externalization for select content. Via interviews and a survey, we then discussed attitudes toward acoustic transparency (e.g. being perceived as safer), and the potential shifts in personal audio consumption that might be precipitated by their adoption. More broadly, this paper has reflected on how we describe auditory mixed reality experiences and provided a richer set of definitions and examples to capture the mixed reality audio space. With wearable, acoustically transparent, personal, private, head-tracked, always-available soundspaces now a reality, this presents designers and practitioners with a powerful spatial audio canvas, and may represent the first successful push toward wearable, everyday mixed reality headsets that see consumer adoption.

ACKNOWLEDGEMENTS

This research received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (#835197, *ViAjeRo*).

REFERENCES

[1] Utku Günay Acer, Marc Van den Broeck, and Fahim Kawsar. 2019. The city as a personal assistant. In *Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers - UbiComp/ISWC '19*. ACM Press, New York, New York, USA, 1102–1106. DOI: http://dx.doi.org/10.1145/3341162.3350847

[2] Emanuel Aguilera, Jose J. Lopez, Pablo Gutierrez, and Maximo Cobos. 2014. An immersive multi-party conferencing system for mobile devices using binaural audio. In *Proceedings of the AES International Conference*, Vol. 2014-Janua. https://pdfs.semanticscholar.org/b7c9/842d3a32550f449dbbec55e65f3744b78d47.pdf

[3] Robert Albrecht, Riitta Väänänen, and Tapio Lokki. 2016. Guided by music: pedestrian and cyclist navigation with route and beacon guidance. *Personal and Ubiquitous Computing* 20, 1 (2016), 121–145. DOI: http://dx.doi.org/10.1007/s00779-016-0906-z

[4] JJ Alvarsson, S Wiens, and Mats Nilsson. 2010. Stress recovery during exposure to nature sound and environmental noise. *International Journal of Environmental Research and Public Health* (2010). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2872309/

[5] Jesse Armstrong, Brian Welsh, and Charlie Brooker. 2011. The entire history of you. (2011).

[6] Ronald T. Azuma. 1997. A survey of augmented reality. *Presence: Teleoperators and Virtual Environments* 6, 4 (aug 1997), 355–385. DOI: http://dx.doi.org/10.1162/pres.1997.6.4.355

[7] Sean Barbeau. 2018. Dual-frequency GNSS on Android devices. (2018). https://medium.com/@sjbarbeau/dual-frequency-gnss-on-android-devices-152b8826e1c

[8] Amit Barde, William S. Helton, Gun Lee, and Mark Billinghurst. 2016. Binaural spatialisation over a bone conduction headset: Minimum discernable angular difference. In *140th Audio Engineering Society International Convention 2016, AES 2016*. Audio Engineering Society. http://www.aes.org/e-lib/online/browse.cfm?elib=18254

[9] Amit Barde, Gun Lee, William S. Helton, and Mark Billinghurst. 2017. Binaural Spatialisation Over a Bone Conduction Headset: Elevation Perception. In *Proceedings of the 22nd International Conference on Auditory Display - ICAD 2016*. The International Community for Auditory Display, Arlington, Virginia, 173–176. DOI: http://dx.doi.org/10.21785/icad2016.013

[10] Benjamin B. Bederson. 1995. Audio augmented reality: a prototype automated tour guide.. In *CHI '95 Conference Companion on Human Factors in Computing Systems*. 210–211. DOI: http://dx.doi.org/10.1145/223355.223526

[11] D. R. Begault, E. M. Wenzel, and M. R. Anderson. 2001. Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. *AES: Journal of the Audio Engineering Society* 49, 10 (2001), 904–916. http://www.aes.org/e-lib/browse.cfm?elib=10175

[12] Frauke Behrendt. 2015. Locative Media as Sonic Interaction Design: Walking Through Placed Sounds. *Wi: Journal of Mobile Media* 9, 2 (2015), 25. http://wi.mobilities.ca/frauke-behrendt-locative-media-as-sonic-interaction-designwalking-through-placed-sounds

[13] Jacob A. Benfield, B. Derrick Taff, Peter Newman, and Joshua Smyth. 2014. Natural sound facilitates mood recovery. *Ecopsychology* 6, 3 (2014), 183–188. DOI: http://dx.doi.org/10.1089/eco.2014.0028

[14] Karin Bijsterveld. 2010. Acoustic Cocooning. *The Senses and Society* 5, 2 (jul 2010), 189–211. DOI: http://dx.doi.org/10.2752/174589210X12668381452809

[15] Mark Billinghurst and Shaleen Deo. 2007. Motion-tracking in spatial mobile audio-conferencing. *Workshop on spatial audio for mobile devices (SAMD 2007) at Mobile HCI* (2007), 3–6. https://www.researchgate.net/profile/Mark

[16] Costas Boletsis and Dimitra Chasanidou. 2018. Smart tourism in cities: Exploring urban destinations with audio augmented reality. In *ACM International Conference Proceeding Series*. 515–521. DOI: http://dx.doi.org/10.1145/3197768.3201549

[17] Bose. 2019a. Smart Noise Cancelling Headphones 700. (2019). https://www.bose.co.uk/en_gb/products/headphones/noise_cancelling_headphones/noise-cancelling-headphones-700.html

[18] Bose. 2019b. SoundWear Companion Wearable Speaker. (2019). https://www.bose.co.uk/en_gb/products/speakers/portable_speakers/soundwear-companion.html

[19] Bose. 2019c. Wearables by Bose - AR Audio Sunglasses. (2019). https://www.bose.co.uk/en_gb/products/frames.html

[20] A. L. Brown and Andreas Muhar. 2004. An approach to the acoustic design of outdoor space. *Journal of Environmental Planning and Management* 47, 6 (nov 2004), 827–842. DOI: http://dx.doi.org/10.1080/0964056042000284857

[21] Alan Chamberlain, Mads Bødker, Adrian Hazzard, David McGookin, David De Roure, Pip Willcox, and Konstantinos Papangelis. 2017. Audio technology and mobile human computer interaction: From space and place, to social media, music, composition and creation. *International Journal of Mobile Human Computer Interaction* 9, 4 (2017), 25–40. DOI: http://dx.doi.org/10.4018/IJMHCI.2017100103

[22] Priscilla Chueng. 2002. Designing Auditory Spaces to Support Sense of Place: The Role of Expectation. *The Role of Place in On-line Communities (Workshop)* November (2002). http://scom.hud.ac.uk/scompc2/research.htmhttp://pnainby.businesscatalyst.com/pdf/pchueng

[23] Michael Cohen. 1993. Throwing, pitching and catching sound: audio windowing models and modes. *International Journal of Man-Machine Studies* 39, 2 (1993), 269 – 304. DOI:http://dx.doi.org/https://doi.org/10.1006/imms.1993.1062

[24] Poppy Crum. 2019. Hearables: Here come the: Technology tucked inside your ears will augment your daily life. *IEEE Spectrum* 56, 5 (may 2019), 38–43. DOI:http://dx.doi.org/10.1109/mspec.2019.8701198

[25] James J. Cummings and Jeremy N. Bailenson. 2015. How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence. *Media Psychology* (may 2015). http://www.tandfonline.com/doi/abs/10.1080/15213269.2015.1015740

[26] Jérome Daniel, J B Rault, and J D Polack. 1998. Ambisonics encoding of other audio formats for multiple listening conditions. *105th AES Convention* 4795 (sep 1998), Preprint 4795. http://www.aes.org/e-lib/browse.cfm?elib=8385

[27] Florian Denk, Marko Hiipakka, Birger Kollmeier, and Stephan M.A. Ernst. 2018. An individualised acoustically transparent earpiece for hearing devices. *International Journal of Audiology* 57, sup3 (may 2018), S62–S70. DOI:http://dx.doi.org/10.1080/14992027.2017.1294768

[28] Shaleen Deo, Mark Billinghurst, Nathan Adams, and Juha Lehikoinen. 2007. Experiments in spatial mobile audio-conferencing. In *Mobility Conference 2007 - The 4th Int. Conf. Mobile Technology, Applications and Systems, Mobility 2007, Incorporating the 1st Int. Symp. Computer Human Interaction in Mobile Technology, IS-CHI 2007*. 447–451. DOI:http://dx.doi.org/10.1145/1378063.1378133

[29] Frank Van Diggelen, Roy Want, and Wei Wang. 2018. How to achieve 1-meter accuracy in Android. (2018). https://www.gpsworld.com/how-to-achieve-1-meter-accuracy-in-android/

[30] H. Durrant-Whyte and T. Bailey. 2006. Simultaneous localization and mapping: part I. *IEEE Robotics & Automation Magazine* 13, 2 (jun 2006), 99–110. DOI:http://dx.doi.org/10.1109/MRA.2006.1638022

[31] Andreas Floros, Nicolas Alexander Tatlas, and Stylianos Potirakis. 2011. Sonic perceptual crossings: A tic-tac-toe audio game. In *ACM International Conference Proceeding Series*. 88–94. DOI:http://dx.doi.org/10.1145/2095667.2095680

[32] Claudio Forlivesi, Utku Günay Acer, Marc van den Broeck, and Fahim Kawsar. 2018. Mindful interruptions: a lightweight system for managing interruptibility on wearables. In *Proceedings of the 4th ACM Workshop on Wearable Systems and Applications - WearSys '18*. ACM Press, New York, New York, USA, 27–32. DOI:http://dx.doi.org/10.1145/3211960.3211974

[33] Jon Francombe. 2018. Vostok-K Incident: Immersive Audio Drama on Personal Devices. (2018). https://www.bbc.co.uk/rd/blog/2018-10-multi-speaker-immersive-audio-metadata

[34] J Freeman and J Lessiter. 2001. Here, there and everywhere: The effects of multichannel audio on presence. (2001). https://smartech.gatech.edu/handle/1853/50621

[35] Masaaki Fukumoto and Masaaki. 2018. SilentVoice: Unnoticeable Voice Input by Ingressive Speech. In *The 31st Annual ACM Symposium on User Interface Software and Technology - UIST '18*. ACM Press, New York, New York, USA, 237–246. DOI:http://dx.doi.org/10.1145/3242587.3242603

[36] Samuel Gibbs. 2020. Amazon launches Alexa smart ring, smart glasses and earbuds. (2020). https://www.theguardian.com/technology/2019/sep/26/amazon-launches-alexa-smart-ring-smart-glasses-and-earbuds

[37] Robert H. Gilkey and Janet M. Weisenberger. 1995. The Sense of Presence for the Suddenly Deafened Adult:Implications for Virtual Environments. *Presence: Teleoperators and Virtual Environments* 4, 4 (jan 1995), 357–363. DOI:http://dx.doi.org/10.1162/pres.1995.4.4.357

[38] Stephen Groening. 2016. 'No One Likes to Be a Captive Audience': Headphones and In-Flight Cinema. *Film History: An International Journal* (2016). https://muse.jhu.edu/article/640056/summary

[39] Catherine Guastavino, Véronique Larcher, Guillaume Catusseau, and Patrick Boussard. 2007. Spatial Audio Quality Evaluation: Comparing Transaural , Ambisonics and Stereo. In *Proceedings of the 13th International Conference on Auditory Display*. 53–58. DOI:http://dx.doi.org/10.1121/1.3652887

[40] Etienne Hendrickx, Peter Stitt, Jean Christophe Messonnier, Jean Marc Lyzwa, Brian Fg Katz, and Catherine de Boishéraud. 2017. Influence of head tracking on the externalization of speech stimuli for non-individualized binaural synthesis. *The Journal of the Acoustical Society of America* 141, 3 (2017), 2011. DOI:http://dx.doi.org/10.1121/1.4978612

[41] Kristian Hentschel and Jon Francombe. 2019. Framework for Web Delivery of Immersive Audio Experiences Using Device Orchestration. In *Adjunct Proceedings of ACM TVX 2019*. ACM.

[42] Thomas R. Herzog and Patrick J. Bosley. 1992. Tranquility and preference as affective qualities of natural environments. *Journal of Environmental Psychology* 12, 2 (jun 1992), 115–127. DOI:http://dx.doi.org/10.1016/S0272-4944(05)80064-7

[43] Mansoor Hyder, Michael Haun, and Christian Hoene. 2010. Placing the participants of a spatial audio conference call. In *2010 7th IEEE Consumer Communications and Networking Conference, CCNC 2010*. IEEE, 1–7. DOI: http://dx.doi.org/10.1109/CCNC.2010.5421622

[44] Robin Ince, Brian Cox, Jo Brand, Jo Dunkley, and Adam Masters. 2019. The Infinite Monkey Cage, Series 19, How We Measure the Universe. (2019). https://www.bbc.co.uk/programmes/m0002g8k

[45] Andrew F. Jarosz and Jennifer Wiley. 2014. What Are the Odds? A Practical Guide to Computing and Reporting Bayes Factors. *The Journal of Problem Solving* (2014). DOI: http://dx.doi.org/10.7771/1932-6246.1167

[46] Robert E. Kass and Adrian E. Raftery. 1995. Bayes Factors. *J. Amer. Statist. Assoc.* 90, 430 (jun 1995), 773–795. DOI: http://dx.doi.org/10.1080/01621459.1995.10476572

[47] Fahim Kawsar and Alastair Beresford. 2019. EarComp 2019 - 1st International Workshop on Earable Computing. (2019). www.esense.io/earcomp

[48] Fahim Kawsar, Chulhong Min, Akhil Mathur, and Allesandro Montanari. 2018a. Earables for Personal-Scale Behavior Analytics. *IEEE Pervasive Computing* 17, 3 (jul 2018), 83–89. DOI: http://dx.doi.org/10.1109/MPRV.2018.03367740

[49] Fahim Kawsar, Chulhong Min, Akhil Mathur, Alessandro Montanari, Utku Günay Acer, and Marc Van den Broeck. 2018b. eSense: Open Earable Platform for Human Sensing. In *Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems - SenSys '18*. ACM Press, New York, New York, USA, 371–372. DOI: http://dx.doi.org/10.1145/3274783.3275188

[50] Richard King, Brett Leonard, Will Howie, and Jack Kelly. 2017. Real or Illusion? A Comparative Study of Captured Ambiance vs. Artificial Reverberation in Immersive Audio Applications. (may 2017). http://www.aes.org/e-lib/browse.cfm?elib=18621

[51] Elise Labbé, Nicholas Schmidt, Jonathan Babin, and Martha Pharr. 2007. Coping with stress: The effectiveness of different types of music. *Applied Psychophysiology Biofeedback* 32, 3-4 (2007), 163–168. DOI: http://dx.doi.org/10.1007/s10484-007-9043-9

[52] Chris Lane and Richard Gould. 2018. Let's Test: 3D Audio Spatialization Plugins. (2018). http://designingsound.org/2018/03/29/lets-test-3d-audio-spatialization-plugins/

[53] Pontus Larsson, Aleksander Väljamäe, Daniel Västfjäll, Ana Tajadura-Jiménez, and Mendel Kleiner. 2010. Auditory-Induced Presence in Mixed Reality Environments and Related Technology. Springer, London, 143–163. DOI: http://dx.doi.org/10.1007/978-1-84882-733-2_8

[54] Catherine Lavandier and Boris Defréville. 2006. The contribution of sound source characteristics in the assessment of urban soundscapes. *Acta Acustica united with Acustica* 92, 6 (2006), 912–921. https://www.ingentaconnect.com/content/dav/aaua/2006/00000092/00000006/art00009

[55] Robert W. Lindeman and Haruo Noma. 2007. A classification scheme for multi-sensory augmented reality. In *Proceedings of the 2007 ACM symposium on Virtual reality software and technology - VRST '07*. ACM Press, New York, New York, USA, 175. DOI: http://dx.doi.org/10.1145/1315184.1315216

[56] Robert W. Lindeman, Haruo Noma, and Paulo Gonçalves De Barros. 2007. Hear-through and mic-through augmented reality: Using bone conduction to display spatialized audio. In *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, ISMAR*. DOI: http://dx.doi.org/10.1109/ISMAR.2007.4538843

[57] Robert W. Lindeman, Haruo Noma, and Paulo Goncalves de Barros. 2008. An Empirical Study of Hear-Through Augmented Reality: Using Bone Conduction to Deliver Spatialized Audio. In *2008 IEEE Virtual Reality Conference*. IEEE, 35–42. DOI: http://dx.doi.org/10.1109/VR.2008.4480747

[58] Matthew Lombard and Theresa Ditton. 2006. At the Heart of It All: The Concept of Presence. *Journal of Computer-Mediated Communication* 3, 2 (jun 2006), 0–0. DOI:http://dx.doi.org/10.1111/j.1083-6101.1997.tb00072.x

[59] Dave Madole and Durand Begault. 1995. 3-D Sound for Virtual Reality and Multimedia. *Computer Music Journal* 19, 4 (1995), 99. DOI: http://dx.doi.org/10.2307/3680997

[60] Mapbox. 2019. Mapbox Live Location Platform. (2019). https://www.mapbox.com/

[61] Georgios Marentakis and Rudolfs Liepins. 2014. Evaluation of hear-through sound localization. In *Conference on Human Factors in Computing Systems - Proceedings*. 267–270. DOI: http://dx.doi.org/10.1145/2556288.2557168

[62] Nick Mariette. 2007. From Backpack to Handheld: The Recent Trajectory of Personal Location Aware Spatial Audio. *academia.edu* (2007). http://www.academia.edu/download/698511/Nick

[63] Nicholas Mariette. 2013. Human Factors Research in Audio Augmented Reality. In *Human Factors in Augmented Reality Environments*. Springer New York, New York, NY, 11–32. DOI: http://dx.doi.org/10.1007/978-1-4614-4205-9_2

[64] Mark Weiser. 1991. The Computer for the 21st Century. *Scientific American* 265, 3 (1991), 94–104. https://ieeexplore.ieee.org/abstract/document/993141/http://www.cse.nd.edu/

[65] Grzegorz Maziarczyk. 2018. Transhumanist Dreams and/as Posthuman Nightmares in Black Mirror. *Roczniki Humanistyczne* 66, 11 Zeszyt specjalny (2018), 125–136. DOI: http://dx.doi.org/10.18290/rh.2018.66.11s-10

[66] Mark McGill, Daniel Boland, Roderick Murray-Smith, and Stephen Brewster. 2015a. A Dose of Reality: Overcoming Usability Challenges in VR Head-Mounted Displays. In *Proc. of CHI '15*. ACM Press, New York, New York, USA. DOI: http://dx.doi.org/10.1145/2702123.2702382

[67] Mark McGill, John Williamson, and Stephen A. Brewster. 2015b. It Takes Two (To Co-View). In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video - TVX '15*. ACM Press, 23–32. DOI: http://dx.doi.org/10.1145/2745197.2745199

[68] David McGookin and David. 2016. Towards ubiquitous location-based audio. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct - MobileHCI '16*. ACM Press, New York, New York, USA, 1064–1068. DOI: http://dx.doi.org/10.1145/2957265.2964196

[69] Siobhán McHugh. 2012. The affective power of sound: Oral history on radio. (2012). DOI: http://dx.doi.org/10.1093/ohr/ohs092

[70] Frank Melchior and Darius Satongar. 2019. Spatial Audio for Broadcast. (2019). https://www.bbc.co.uk/rd/projects/periphony-for-broadcast

[71] Microsoft. 2019. Microsoft Soundscape. (2019). https://www.microsoft.com/en-us/research/product/soundscape/

[72] Chulhong Min, Akhil Mathur, and Fahim Kawsar. 2018. Exploring audio and kinetic sensing on earable devices. In *Proceedings of the 4th ACM Workshop on Wearable Systems and Applications - WearSys '18*. ACM Press, New York, New York, USA, 5–10. DOI: http://dx.doi.org/10.1145/3211960.3211970

[73] Miscellaneous. 2019a. Countryside birds 1 - 'GN00421-Birds_chirp_during_the_early_spring-wxyz.wav'. (2019). http://www.spheric-collection.com/

[74] Miscellaneous. 2019b. VR Ambisonic Sound Library - 'Waves, Mediterranean Sea, Rocky Beach, Crushing, Birds, Seagulls, Splashing, Windy, Edro 3 Shipwreck, Cyprus, Zoom H3VR,9624, 02.wav'. (2019). https://freetousesounds.com/product/vr-ambisonic-sound-library/

[75] Richard D. Morey. 2019. Computation of Bayes Factors for Common Designs [R package BayesFactor version 0.9.12-4.2]. *cran* (2019). https://cran.r-project.org/web/packages/BayesFactor/index.html

[76] Craig D. Murray, Paul Arnold, and Ben Thornton. 2000. Presence accompanying induced hearing loss: Implications for immersive virtual environments. *Presence: Teleoperators and Virtual Environments* 9, 2 (apr 2000), 137–148. DOI: http://dx.doi.org/10.1162/105474600566682

[77] Elizabeth D. Mynatt, Maribeth Back, Roy Want, and Ron Frederick. 1997. Audio Aura: Light-weight audio augmented reality. In *UIST (User Interface Software and Technology): Proceedings of the ACM Symposium*. 211–212. http://www.roywant.com/cv/papers/1997/1997-10(UIST97)AudioAura.pdf

[78] Takuji Narumi, Shinya Nishizaka, Takashi Kajinami, Tomohiro Tanikawa, and Michitaka Hirose. 2011. MetaCookie+. In *2011 IEEE Virtual Reality Conference*. IEEE, 265–266. DOI: http://dx.doi.org/10.1109/VR.2011.5759500

[79] Danielle Navarro. 2018. Learning statistics with R: A tutorial for psychology students and other beginners. (Version 0.6.1). In *University of New South Wales*. Chapter Chapter 17. https://learningstatisticswithr.com/book/bayes.html

[80] Rolf Nordahl and Niels Christian Nilsson. 2014. The Sound of Being There: Presence and Interactive Audio in Immersive Virtual Reality. *The Oxford Handbook of Interactive Audio* March 2016 (2014), 213–233. DOI: http://dx.doi.org/10.1093/oxfordhb/9780199797226.013.013

[81] Oculus. 2019. Oculus Audio SDK. (2019). https://developer.oculus.com/documentation/audiosdk/latest/concepts/book-audiosdk/

[82] Kenji Ozawa, Yoshihiro Chujo, Yôiti Suzuki, and Toshio Sone. 2003. Psychological factors involved in auditory presence. *Acoustical Science and Technology* 24, 1 (2003), 42–44. DOI: http://dx.doi.org/10.1250/ast.24.42

[83] Joseph Plazak and Marta Kersten-Oertel. 2018. A survey on the affordances of 'hearables'. (jul 2018). DOI: http://dx.doi.org/10.3390/inventions3030048

[84] Jussi Rämö and Vesa Välimäki. 2012. Digital Augmented Reality Audio Headset. *Journal of Electrical and Computer Engineering* 2012 (oct 2012), 1–13. DOI: http://dx.doi.org/10.1155/2012/457374

[85] Josephine Reid, Erik Geelhoed, Richard Hull, Kirsten Cater, and Ben Clayton. 2005. Parallel Worlds: Immersion in Location-Based Experiences. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems (CHI EA '05)*. Association for Computing Machinery, 1733–1736. DOI: http://dx.doi.org/10.1145/1056808.1057009

[86] Resonance Audio. 2019a. Fundamental Concepts. (2019). https://resonance-audio.github.io/resonance-audio/discover/concepts.html

[87] Resonance Audio. 2019b. Resonance Audio - Spatial Audio Toolkit. (2019). https://resonance-audio.github.io/resonance-audio/

[88] Julie Rico and Stephen Brewster. 2010. Gesture and voice prototyping for early evaluations of social acceptability in multimodal interfaces. In *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction on - ICMI-MLMI '10*. ACM Press, New York, New York, USA, 1. DOI: http://dx.doi.org/10.1145/1891903.1891925

[89] Chiara Rossitto, Louise Barkhuus, and Arvid Engström. 2016. Interweaving place and story in a location-based audio drama. *Personal and Ubiquitous Computing* 20, 2 (2016), 245–260. DOI: http://dx.doi.org/10.1007/s00779-016-0908-x

[90] J Saldaña. 2015. *The coding manual for qualitative researchers*. Sage.

[91] Nitin Sawhney and Chris Schmandt. 2000. Nomadic Radio: Speech and Audio Interaction for Contextual Messaging in Nomadic Environments. *ACM Transactions on Computer-Human Interaction* 7, 3 (2000), 353–383. DOI: http://dx.doi.org/10.1145/355324.355327

[92] Hanna Kathrin Schraffenberger. *Arguably augmented reality: relationships between the virtual and the real*. Ph.D. Dissertation. https://openaccess.leidenuniv.nl/handle/1887/67292

[93] T Schubert, F Friedmann, and H Regenbrecht. 2001. The experience of presence: Factor analytic insights. *Presence* (2001). http://ieeexplore.ieee.org/xpls/abs

[94] Stefania Serafin, Michele Geronazzo, Cumhur Erkut, Niels C. Nilsson, and Rolf Nordahl. 2018. Sonic Interactions in Virtual Reality: State of the Art, Current Challenges, and Future Directions. *IEEE Computer Graphics and Applications* 38, 2 (mar 2018), 31–43. DOI:http://dx.doi.org/10.1109/MCG.2018.193142628

[95] Sheng Shen, Nirupam Roy, Junfeng Guan, Haitham Hassanieh, and Romit Roy Choudhury. 2018. MUTE: bringing IoT to noise cancellation. In *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication - SIGCOMM '18*. ACM Press, New York, New York, USA, 282–296. DOI:http://dx.doi.org/10.1145/3230543.3230550

[96] Adalberto L Simeone, Eduardo Velloso, and Hans Gellersen. 2015. Substitutional Reality: Using the Physical Environment to Design Virtual Reality Experiences. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15* (2015). DOI: http://dx.doi.org/10.1145/2702123.2702389

[97] Maximilian Speicher, Brian D. Hall, and Michael Nebeling. 2019. What is Mixed Reality?. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. ACM Press, New York, New York, USA, 1–15. DOI: http://dx.doi.org/10.1145/3290605.3300767

[98] Charles Spence and Maya U. Shankar. 2010. The influence of auditory cues on the perception of, and responses to, food and drink. *Journal of Sensory Studies* 25, 3 (mar 2010), 406–430. DOI: http://dx.doi.org/10.1111/j.1745-459X.2009.00267.x

[99] Evgeny Stemasov, Gabriel Haas, Michael Rietzler, and Enrico Rukzio. 2018. Augmenting Human Hearing Through Interactive Auditory Mediated Reality. (2018). DOI:http://dx.doi.org/10.1145/3266037.3266104

[100] D Storek, J Stuchlik, and F Rund. 2015. Modifications of the surrounding auditory space by augmented reality audio: Introduction to warped acoustic reality. (2015). https://smartech.gatech.edu/handle/1853/54143

[101] Miikka Tikander, Matti Karjalainen, and Ville Riikonen. 2008. An Augmented Reality Audio Headset. In *Proc. of the 11th International Conference on Digital Audio Effects (DAFx-08)*. https://core.ac.uk/download/pdf/21721632.pdf

[102] Unity Technologies. 2019. Unity Game Engine. (2019). https://unity3d.com/

[103] Vesa Valimaki, Andreas Franck, Jussi Ramo, Hannes Gamper, and Lauri Savioja. 2015a. Assisted listening using a headset: Enhancing audio perception in real, augmented, and virtual environments. *IEEE Signal Processing Magazine* 32, 2 (mar 2015), 92–99. DOI: http://dx.doi.org/10.1109/MSP.2014.2369191

[104] Vesa Valimaki, Andreas Franck, Jussi Ramo, Hannes Gamper, and Lauri Savioja. 2015b. Assisted Listening Using a Headset: Enhancing audio perception in real, augmented, and virtual environments. *IEEE Signal Processing Magazine* 32, 2 (mar 2015), 92–99. DOI: http://dx.doi.org/10.1109/MSP.2014.2369191

[105] Valve. 2018. Steam Audio. (2018). https://valvesoftware.github.io/steam-audio/

[106] Daniel Västfjäll, Mendel Kleiner, and Tommy Görling. 2003. Affective Reactions to and Preference for Combinations of Interior Aircraft Sound and Vibration. *The International Journal of Aviation Psychology* 13, 1 (2003), 33–47. DOI: http://dx.doi.org/10.1207/S15327108IJAP1301_3

[107] Javier Vilanova. 2017. *Extended Reality and Abstract Objects: A pragmalinguistic approach*. DOI: http://dx.doi.org/10.1515/9783110497656-003

[108] James Vincent. 2016. Are bone conduction headphones good enough yet? (2016). https://www.theverge.com/2016/10/24/13383616/bone-conduction-headphones-best-pair-aftershokz

[109] Vue. 2020. Meet Vue - Smart Glasses. (2020). https://www.enjoyvue.com/

[110] Jack Wallen. 2015. Earables: The next big thing. (2015). https://www.techrepublic.com/article/earables-the-next-big-thing/

[111] Marian Weger, Thomas Hermann, and Robert Höldrich. 2019. Real-time Auditory Contrast Enhancement; Real-time Auditory Contrast Enhancement. (2019), 23–27. DOI:http://dx.doi.org/10.4119/unibi/2935786

[112] Chris Welch. 2020. Apple AirPods Pro hands-on: the noise cancellation really works. (2020). https://www.theverge.com/2019/10/29/20938740/apple-airpods-pro-hands-on-noise-cancellation-photos-features

[113] CD Wickens, R Martin-Emerson, and I Larish. 1993. Attentional tunneling and the head-up display. (1993). https://ntrs.nasa.gov/search.jsp?R=19950063577

[114] Kris Wouk. 2014. Intelligent Headset - the first big thing in wearable tech? (2014). https://www.soundguys.com/meet-intelligent-headset-headphones-first-big-thing-wearable-tech-676/

[115] Wei Yang and Jian Kang. 2005. Soundscape and Sound Preferences in Urban Squares: A Case Study in Sheffield. *Journal of Urban Design* 10, 1 (feb 2005), 61–80. DOI: http://dx.doi.org/10.1080/13574800500062395

[116] Bin Yu, Jun Hu, Mathias Funk, and Loe Feijs. 2016. A Study on User Acceptance of Different Auditory Content for Relaxation. (2016). DOI: http://dx.doi.org/10.1145/2986416.2986418

[117] Bin Yu, Jun Hu, Mathias Funk, and Loe Feijs. 2017. A Model of Nature Soundscape for Calm Information Display. *Interacting with Computers* 29, 6 (nov 2017), 813–823. DOI:http://dx.doi.org/10.1093/iwc/iwx007