Un bon voyage sonore: Avoiding listener discomfort in immersive audio documentary

Abigail Wincott, Falmouth University, UK

Abstract

This article interrogates claims made in the emerging discourse of immersive audio documentary that spatial sound is more real, allowing the listener to step into another space, and understand the world better. However my analysis shows makers are failing to make good on these claims. Use of the technical affordances of spatial audio is limited and producers enrol concepts of the real and of transportation in a colonial discourse of exploration and adventure, reproducing a disengaged mode of listening, while avoiding discomfort at all costs.

Imagining the listener in radiophonic factual programmes

While in mono or stereo speech programmes the listener is outside the scenes they hear, the presence of the listener *in* the scene is one of the defining factors of immersive audio (Agrawal et al., 2019). In interviews producers told me that learning to use spatial audio requires a shift to conceiving of storytelling as a series of spatial sound scenes, with detailed thought given to the placement of interviewees, presenter and ambient sounds, but most of all to the listener (Wincott, Martin and Richards 2020, 2021) and they perceived managing the listener's experience as key to successful use of the technology. Listening – how much, what type – is always important to producers, for whom success is tied in various ways to the reception of the work. Technological changes affect this relationship and engender hopes and anxieties about how to manage listener engagement (Euritt, 2019; Tzlil & Nicholas, 2019). In this article I interrogate the emerging discourse of immersive or spatial listening (for more on the uses of these terms see below) in audio journalism and documentary making, which borrow immersive journalism concepts to make claims for immersive listening as radically different and capable of profoundly influencing the way listeners engage with content.

The analysis draws on 18 interviews with programme-makers and sound recordists in six different countries, who use immersive sound for audio journalism and documentary, and extensive listening to factual programmes that use immersive audio. I searched online for content using the terms spatial, binaural, 3D and so on and my sample does not include news content, which is not usually labelled in this way, though interviewees tell me it is being used in news production. The programmes were otherwise varied in subject matter and style, to allow me to gauge the range of ways immersive listening was being structured, rather than as a representative cross section of output. They were almost all from public service broadcasters, who have invested in immersive audio production in recent years (Wincott, Martin and Richards 2020) in France (Radio France and RFI), the UK (BBC) and Germany (WDR, Bayerischer Rundfunk), though I also found a short series from the Guardian newspaper – Oceans of Noise – and a podcast series produced by French independent, Nouvelles Ecoutes – Le canon sur la tempe – and included one of each in my listening. In total I listened to 17 programmes, just under 11 hours of content, noting tactics that situated the listener in relation to the content both in terms of spatial placement (technical immersion) and linguistic and other discursive techniques. To explain the affordances of the former, I will briefly explain the terms used and the differences between spatial and conventional audio.

Conventional features-making uses mono or stereo to mix atmospheric sounds, music and speech, which can be recorded separately and layered to produce a generalised atmospheric sound collage, using technologies and techniques developed across the early to mid-twentieth century (Madsen, 2010; Madsen, 2013). Sound recordings contain acoustic information about the size of a space, the materials in it and the distance of voices or other sounds from the microphone. In this sense, sound is always spatial. Sound is also diffuse and enveloping, and listening does not require directed attention in the same way watching a screen does, so defining 'immersive' sound is difficult (Agrawal et al., 2019). Nonetheless, formats like binaural offer something new: an envelopment in sound that we perceive as emanating from the sides and behind, with more precise localisation cues that can give a very vivid sense of three dimensional space (Romigh & Simpson, 2014). In binaural recording a pair of microphones is placed where human ears would be, either in a dummy head or worn by a sound recordist. The listener's audioposition (Verma, 2012) or point of audition (Chion, 1994) is therefore that of the binaural microphone pair. Binaural is the most common format used to produce this effect, though there are others in use in both recording and post-production. The sound produced by these formats is described variously as immersive, spatial or sometimes 3D, and although my interviewees had personal preferences, there was no overall consensus on which terms to use.

In VR and 360 video, which have received much greater scholarly and industry attention, authors such as Owen et al (2015), Dominguez (2017) and Maschio and Baumann (2017) write about the potential for technical immersion (the sense of being sensorily enveloped by the medium [Dominguez 2017]) to produce a sense of 'being there' or 'presence'. In immersive journalism literature, the immersed subject is seen in a highly spatialised framework as 'entering' a story, because the fourth wall is broken (Dominguez, 2017). According to another metaphor, the user is 'transported' to another place or time (Chainon & Chainon, 2018; de la Peña et al., 2010; Dominguez, 2017) – a concept which was used to describe the wonder of radio listening in its earliest days (Douglas, 2004, p. 162; Madsen, 2013). In industry and academic discourses of immersive (VR) journalism the viewer becomes a 'user', someone who, as an individual, passes through the fourth wall, becoming witnesses to events first hand (Dominguez, 2017, p. 2; Owen et al., 2015) or even stepping into someone else's shoes (Chainon & Chainon, 2018, p. 4) and all of this is expected to increase their engagement with and understanding of other people's realities, and

thus heighten their sense of connection and empathy – a proposition much discussed and disputed in the VR journalism literature (see eg Goutier et al., 2021 for an overview).

The factors that engender immersion or undermine it have been explored in the literature on telepresence, such as vividness and realism (Bracken & Skalski, 2010) and the illusion of non-mediation (Bracken & Botta, 2010). Many of the conventions of video journalism are problematic because they break this illusion and undermine 'virtual body ownership' (De la Pena et al 2010). These include the presence of a reporter-narrator (Murray, 2016; Watson, 2017), text overlays and music (Murray 2016), video montages, and the use of different camera angles and framing (Owen et al., 2015). As Murray instructs VR journalists, 'Everything the interactor sees or hears must be part of the space VR creates for them to inhabit.' Newton and Soukup (2016) also found VR users' sense of presence was supported if they could draw on a familiar 'social script' to explain – or excuse – their role. For example being in a classroom was familiar and acceptable, but being in someone's bedroom left them with a feeling of voyeurism or vulnerability that undermined their sense of presence.

As my interviewees talked about the need to carefully structure spatial scenes, and their examples of what they felt were successful and unsuccessful examples referred to the placement of the listener, I aimed to conduct an in-depth analysis of a scene from each of the programmes in my sample, thinking of spatial audioposition in this very technical way. I listened out for sounds that might imply embodiment of the listener, including footsteps or rustling clothes, reference to me being there in person, any other auditory cues as to the height from which I was listening, and the distance between me and the people, places and events. Through this analysis, I learnt that the use of spatial placement in scenes was very limited, and the main ways I was positioned were rhetorical not technical, through claims made in presenter narration and programme web text. These worked to suggest a very narrow range of ways I might relate to the stories, people and places in the programmes. In the next section I analyse the most common themes in these claims, and their implications for the emerging discourse of spatial audio listening.

Spatial listening as experiencing reality

One of the dominant tropes of this immersive listening discourse is the idea that spatial audio is simply more 'real', with consequences for how listening is imagined. Two of my

interviewees went so far as to refer to stereo as 'fake', while they and others used the words 'real', 'natural' and 'accurate' to describe binaural. Binaural has been associated with vivid realism since its popularisation in the 1970s (Krebs, 2017, p. 114), placing it in a long line of technological developments in audio heralded for their perceived fidelity (Lacey, 2013).

The promise that spatial sound is more real is reflected in interviewees' use of the word 'capture' to describe recording using spatial formats. Programme promotional texts echo this framing. One example that illustrates this is the Bayerische Rundfunk (BR) 2 special edition in binaural to mark the 50th anniversary of the Bavarian Forest National Park, where the accompanying web text frames the recording process as simple capture of 'real' woodland sound, which is therefore 'pure', implying that listening enables one to simply accompany the team into the forest:

> For the 50th anniversary of the Bavarian Forest National Park, something special: An hour of pure woodland sounds. You'll accompany us to Bodenmais, in the middle of the luxuriant woodlands on the southern slopes of the Großer Arber [mountain]. A BR team went there at the start of spring and recorded, with a so-called dummy head microphone, one hour of real woodland life. For daydreaming, relaxing, contemplating or, worst of all, working. (translated from German)

The podcast thumbnail is an image of a floral folding camping chair, which sits empty in the middle of woodland, an invitation to step into the scene that further plays down the constructed nature of all mediated sound, even one hour of non-narrative woodland atmosphere.

In interviews, the word 'capture' was used in conjunction with claims spatial audio is better than conventional stereo or mono at transporting the listener straight into the scene without the need for a reporter or presenter's explanations. For example, one French producer told me,

> 'The journalist doesn't have to describe, doesn't have to say "okay we're in a train, which is making a lot of noise, there are not many people around us", it will just be delivered direct, the people will just have the sensation, perceive it directly when they listen to the binaural ambience.' (translated from French)

Others described this ideal of listening as direct experience, using visual metaphor:

> 'If I am in mono, I'm a witness. I watch. If I am in 5.1 or immersive sound, I am no longer a witness, I almost become an actor.' (Sound engineer, France. Translated from French)

> 'You're suddenly part of the story as opposed to watching the story.' (Producer, UK)

There's something interesting about you as a listener stepping onto the stage with the action. That's happening around you rather than you're looking at it like a theatre.' (Producer, UK)

These quotes conform to VR discourse's concern with maintaining the 'illusion of non-mediation'. The trope of stepping into the story echoes that of VR journalism discourse too, even borrowing the distinction from video between watching (not listening) from outside and stepping inside an immersive storyworld, as if breaking the fourth wall of the screen – a wall which has, of course, never existed in radio and podcasting.

Of key importance for those I interviewed was the belief that, in journalism and documentary, that presence at the scene will help the listener understand important issues better, including the gilets jaunes protests in France, the workings of parliament, or the experience of civil war on children. The BBC webpage for binaural documentary *Hearing Homelessness* reads:

The programme is made using binaural recording, which allows the listener to hear the world through someone else's ears, as though the sound is coming from beyond their head. They're placed in a location through sound, and through this immersion they come away with a very personal understanding of what life is like for someone living a very different life to them, right under their nose.

This goes beyond transportation in space where the listener can 'witness' things, and equates listening with being transported to *another person's experience* of the world. The text suggests transportation should foster a different mode of knowing about or understanding a subject, an experiential knowing that renders a subject like homelessness what Nonny de la Peña (2016) calls 'viscerally comprehensible'. This experiential understanding appears to be contrasted with (or perhaps in addition to) rational and objective knowing (Lacey, 2013), commonly associated with the visual in Western culture (Jay, 1988) but facilitated in audio journalism through the balanced presentation of information and differing viewpoints on a topic like homelessness.

VR journalism has often used immersion to encourage audiences to understand what Van Damme et al. (2019) call 'distant suffering' by stepping into the spaces they occupy. De la Peña's work includes documentary sequences about people queuing for food aid (*Hunger in Los Angeles*, 2012), and the life of a prisoner in Guantanamo Bay (*Gone Gitmo*, 2007 with Peggy Weil). However, in spatial audio programmes, most claims to let the listener step inside someone else's reality relate to interesting professions. The BBC's short series *In My Head* (2018) 'allows the listener to step into the world of a compelling character with an extraordinary job', including paparazzo, boxing trainer and bomb disposal officer and Arte

Radio did the same with their series *Dans la Tête (In the head)* (2014-2018) which offers the listener 'Trois minutes en binaural dans la tête d'un dentiste, une skieuse, un aveugle, un batteur' (three minutes in binaural in the head of a dentist, a skier, a blind person, a drummer). Professions, recreational activities and living with a disability are all treated the same, and stepping inside the body of another person seems more a matter of satisfying the curiosity of the listener than cultivating visceral understanding of social issues. It has much in common with the many other examples of the transportation metaphor for spatial listening used to offer the listener a kind of aural holiday.

Un bon voyage sonore: listening as travel

My interviewees tended to say spatial audio should be reserved for extraordinary locations, rather than recording in a home or a studio (though there are certainly plenty of people who use spatial sound for studio work and others who feel it should simply be the default format for all output in future). Many programme promotional texts and presenter introductions propose spatial listening as relaxing and escapist. I have already mentioned the 'meditative' woodland sounds on BR2, which was a special 'bonus episode' of travel series *radioReisen* (radio travels). They produced another in 2021, called 'Fleißige Bienchen: Meditative Naturaufnahmen' (Busy bees: meditative nature recordings), while the BBC iPlayer offers BBC Wales's 'A Celtic Dawn Chorus' (2021): 'Put your headphones on to experience the full beauty of the dawn chorus at Newport Wetlands National Nature Reserve'. Many binaural enthusiasts in the audio production sector will be familiar with the work of soundscape composers such as Hildegard Westerkamp (see https://www.hildegardwesterkamp.ca/sound/ for a list of works), who used binaural in impressionistic ways, often nature sounds. And even now, a quick search online for 'binaural' will reveal a strong association between so called binaural beats and relaxation, which perhaps supports the commonsense understanding that spatial sound is ideal for these kind of non-narrative and meditative applications.

Like *radioReisen* on BR2 (tagline: 'Urlaub für den Kopf' (a holiday for the head), RFI's travel series *Si Loin Si Proche* (So Far, So Near) has produced several episodes in binaural, but rather than being meditative specials, these are full programmes, based on a story about a location. Its series blurb invites the listener to 'faire sa malle' (pack your suitcase) and wishes us 'un bon voyage sonore' (a nice sonic trip). Both series already frame

listening as travel in their original stereo format, but what is interesting is the way that binaural is seen as a good fit for that travel content, presumably for its association with transportation and telepresence.

Even more common than claims that immersive listening is a relaxing mode of leisure are promises of adventure, as the programme takes the listener on a journey somewhere extraordinary. These are persuasive techniques that my interviewees told me also helped to secure the extra funding or time to produce a programme in spatial sound, or to win a broadcast commission. They appear in web and podcast texts and in the programmes themselves, where they aim to secure the attention of the listener, who is faced with an overabundance of content, necessitating competitive strategies on the part of media produces, such as promises of immersion (Jones, 2017). *Congo: A River Journey* (BBC World Service, 2018) offers to 'take you on an epic adventure in sound in the Democratic Republic of Congo'. The programme begins:

> To make this programme we've used special binaural recording equipment to bring you with us on an immersive journey in sound. For the full effect put on your headphones shut your eyes and join our expedition up the Congo river. [sounds of forest in background] it's where we're starting our epic adventure up one of the world's epic rivers.

Presenters of two binaural episodes of *Si Loin Si Proche* on the 'mysterious network' of caves in Ardèche tell listeners to expect 'a unique journey, spectacular… that's it, you're ready. We're going deep into the caves, deep into the ages' (translated from French) and repeatedly emphasise the difficulty of accessing the spaces, now closed to tourists.

The discourse of listening as exploration of exotic and difficult places presupposes certain things about the listener of course. *Congo: A River Journey*'s promotional text addresses a listener who is not Congolese, though it is a World Service programme destined for a global audience, and millions of people live and work along the river, home to two capital cities. The programme itself is a guide to how the history and politics of DRC shapes the spaces and places along the river. The presenter interviews Congolese people on location, who speak as experts on their country and its history, including colonial oppression. Still, in reaching out to its audience in its web text and narration, it appeals to them through the code of 'exploration'. It addresses the listener as a Western explorer of the exotic, non-Western world, one dominated by the forces of nature. The history of exploration links it to the colonial project, where knowledge about 'natural' resources, including local people, was

directed to the conquest, exploitation and extraction of their value for colonising countries (Hall, 1992; Hall, 1997).

In *Congo: A River Journey* the binaural recordings provide beautifully rich and atmospheric ambiences, yet the format adheres to radiophonic speech programme conventions in all other respects. Binaural atmospheres are quickly overlain with close miked speech from presenter or interviewees, recorded close to the microphone and mixing it front and centre, thus shrinking the auditory space and bringing it inside the head of the listener, despatialising it (Pullki & Karjalainen, 2015). Presenter speech is often scripted 'links', recorded in studio, which speak about the trip in the past tense. Thus the presenter can travel back and forth, in time and space, enjoying a superior level of agency over the story space, while the listener receives glimpses of spaces, always framed by the speech of the presenter. The spatial grammar is further overwritten by other non-spatial production conventions, such as the fading down of non-English language interviewee voices, so that the voice of a translator can be played in over the top. Non-diegetic music is also introduced and fades up and down.

When the presenter describes where he is and what he is doing and can be heard interacting with his environment – panting when climbing the steps of ruined buildings, for example – we may become aware of his embodiment. In contrast there is no encouragement for the listener to feel embodied in the scene, with one exception. In a helicopter the presenter announces the need to wear ear protectors, and because the producer is wearing the in-ear binaural microphones and puts on ear protectors, what we hear becomes muffled. This is an insight into the ways that an immersive, spatial programme grammar – described as vital by interviewees – could have been deployed consistently, to encourage a sense of embodiment and envelopment in the scene.

Almost all the spatial audio programmes I have listened to for this study undermine technical immersion by using non-spatial format conventions in these ways. In that sense, though fascinating documentaries and often richly radiophonic, they are spatial failures, by the terms set by spatial audio producers I interviewed, and against the claims of the programmes themselves. In the next section, I interrogate a series of attempts at spatialisation that interviewees themselves articulated as failures, to understand more about the values of spatial listening implicit in their judgements.

Spatial failures

Attempt 1. A producer describes listening back to a recording for a documentary. The reporter is wearing in-ear microphones in a busy streetscape. The sound scene is immersive and three-dimensional, but when the reporter speaks, their voice seems to emanate from the listener's (in this case our producer's) location, yet the listener knows this is not their own voice. The producer finds this too 'odd': 'you are hearing what the reporter is hearing… the guy is walking through the streets he's kind of like a conventional reporter… and yet, he's kind of coming from inside your head.' Another producer referred to hearing a presenter speaking from the listener's position as 'disconcerting', causing confusion about where the voice is coming from, and causing the listener to question whether their point of audition is that of the presenter or somewhere else. This level of uncertainty and self-consciousness, they worry, 'that's gonna pull you out of the scene'.

Attempt 2. A sound engineer told me about a radio news programme team who used a binaural dummy head to record a studio discussion. They put the head onto the round table, with contributors seated around it. Listening back the team found this unacceptable, because they felt as listeners they were in the middle of a circle of strangers, who were talking behind them as well as to the sides and front, and this was deeply uncomfortable: 'Basically everyone hated it.' Afterwards the sound engineer advised them to put the dummy head to one side in future, and avoid placing voices behind the listener.

Attempt 3. A documentary maker wanted to record a conversation between documentary subjects, over a meal in a restaurant. She considered walking round behind the guests seated at the table but rejected this because it would seem like 'eavesdropping'. In the end the documentary maker settled on a seat at the table for the binaural microphones/listener and found this satisfactory.

Are there any patterns to these failed attempts? The first example presents an incoherence in the spatial schema, by providing conflicting information to the listener about who and where they are. In that sense it is a technical failure of spatiality, because of the impossibility in the real world of being in the same place as a presenter's body. However, there are plenty of conventions in radiophonic speech programmes that are not 'possible', are what we might call non-naturalistic, such as the 'voice of god' narrating outside time and space in a story, the fading of one voice speaking a foreign language, under its translated counter-part, and the frequent appearance and disappearance of sound effects and music. We

accept these elements in factual programmes, and surely we could learn to accept the binaural reporter, who speaks from where we are. But like these other non-naturalistic techniques of features making, attempt 1 is believed by professionals to undermine the sense of embodiment, of 'being there' in sonic space, by drawing our attention to mediation and by eliciting feelings of uncertainty and discomfort.

Attempts 2 and 3 are not technical issues that interfere with the reproduction of spatial sound. Or more accurately, they are not *only* technical – the studio manager explained one likely response to the discomfort of hearing voices from all around is to turn your head to the person behind. As the voice then also moves (because the headphones you are wearing turn with you), this tends to destroy the binaural effect, as the listener's brain recalculates what is happening (Hacıhabiboğlu et al., 2017). These failures can further be explained as problems to do with social scripts, as observed by Newton and Soukup (2016). I can't conceive of a familiar social script where I am a head on a table while people discuss local politics round me. People speaking to each all around me would be very likely to evoke a sense of extreme vulnerability as well difficulty in interpreting my role. The producer at the meal table rejected listening from behind someone as 'eavesdropping', the auditory equivalent of Newton and Soukup's 'voyeurism'. In this case, there is a social script that explains the listener's position, but the role it offers is socially unacceptable and therefore judged undesirable by the documentary-maker.

What is clear from their anecdotes about what they consider to be failed attempts at immersion and from listening to the programmes in my sample, is that producers are quite content to break the illusion of non-mediation, but they reject what might make listening uncomfortable – voyeurism, vulnerability, unfamiliar placements that disorient. Meanwhile they are repeatedly applying social scripts in content choice, promotional texts and presenter talk that are both familiar and pleasurable for the listener, meditation in nature, tourism and exploration of exotic places. These scripts are problematic in that they cultivate listenership as an entitlement to hear everything without being perceived oneself – an inequality of power that makes this the auditory counterpart to the colonial gaze (Spurr, 1993), something like a colonising ear.

The Western colonial imagination is informed by an entitlement to consume the world and this power has often been closely associated with the privilege of looking. Mapping and surveying offer the colonial explorer a 'commanding view' (Spurr 1993) through which they assert ownership of resources. The disengaged, spectating position (disparaged by

interviewees) is associated with judgement and categorically separates the coloniser from the colonised (Said, 1979, p. 109). This privilege of looking, to consume and to judge, has in recent times manifested in the practices and structures of tourism (Osagie & Buzinde, 2011) as places and people are subject to the 'tourist gaze' (Urry, 1990). But as Spurr (1993, p. 15) writes, journalism is also an inheritor of the colonial gaze. Journalists often take or recreate in their work a literal or metaphorical 'commanding view' of the world and its events that separates them from those who experience events at close quarters and therefore cannot see as clearly. The journalistic view is offered to the audience by proxy and provides:

> 'aesthetic pleasure on one hand, information and authority on the other … it conveys a sense of mastery over the unknown and over what is often perceived by the Western writer as strange and bizarre.' (Spurr, 1993, p. 15).

The journalistic gaze that consumes and judges also operates in a 'disproportionate economy of sight' where 'the answering gaze of the other' is suppressed (Spurr, 1993, p. 17-18), and the audience is neither seen nor judged.

Less has been said about the politics of auditory regimes (Lacey, 2000), relations of listening or listening positions. It is interesting that interviewees used visual metaphors to talk about what they felt were the special values of audio immersion: witnessing, watching and looking at events were used to describe non-immersive sound, and associated with distance. They were contrasted not with words for listening but with the idea of entering the storyspace. We seem, in English at any rate, not to have the same range of vocabulary that implies subtle differences of connection, closeness, intention or power that gaze, spectatorship and watching have. Listening may superficially seem to be the counterpart to watching, but while watching implies both detachment and power, listening to someone does not. Looking has been said to be more active that listening (Lacey, 2013, p. 118) and 'being listened to' implies a position of agency that 'being watched' does not have. But the mediated listener-listened to relationship is contingent on the ways these encounters are structured through radiophonic formats.

For all the talk of moving in close and stepping onto the stage or into someone's head, the spatial audio documentaries in my study replicate Spurr's 'disproportionate economy' offering a pleasurable auditory 'gaze' from afar, for which we have no common word. The listener is not called on to interpret the scene and their physical placement or social role in it, for example, or to consider the ways they might be indirectly implicated in the economic, social or environmental problems experienced others. Nor is the listener often clearly placed

in the position of others within a scene, to apprehend things from their perspective, despite rhetorical claims made in interviews and media texts. But I found one programme that I'd like to spend a little time considering next, because it does break with some of these emerging conventions.

Uncomfortable journey: Joujouka

One episode of the BBC's *Wireless Nights* series is made in binaural. Called 'Joujouka' (BBC Radio 4, 2015), the programme is about traditional musicians in a village in Morocco, whose music is said to invoke the god Pan and bring about a trance-like state that can heal troubled minds. The strangeness of the music, the remoteness and difficulty of access are certainly repeatedly referenced, but there are a few ways the production avoids invoking the colonising ear. The presenter, some distance from us in the binaural space of the performance room, names all the musicians in turn, who each reply with a greeting. He also refers to the way the musicians have stopped chatting to each other because they have seen the recording equipment. This calls on the listener to be aware of the difference their listening makes. He adds (in an aside, close-miked, studio-recorded and thus disembodied and associated with an 'interior' monologue (Wincott, Martin and Richards 2021) how difficult it is going to be to remember the names of the musicians, which may remind the listener of the way meeting new people is a mutual interaction that entails certain expectations on the part of the incomer. These features are the auditory equivalent of Spurr's 'answering gaze' and may prompt a slight shift in the disproportionate economy of radio documentary listening.

Perhaps the most striking difference between this and the other immersive audio programmes under study is the way the presenter addresses the listener directly throughout, referring to our imagined embodied presence at the scene:

> 'Look out for the cactus… Let's go and say hello, come on, don't be shy… come on look they're very friendly… They're inviting us to sit in the middle of them – come on!'

The presenter invites us to see feel an embodied sense of 'being there', yet refers to the mediated nature of the experience even so, by noting the effect of recording equipment on the behaviour of the musicians, and by assuming we are also lying in bed with headphones on:

> 'you might want to put your headphones on if you have any handy. Bit awkward when you put your head on the pillow? Okay you listen in whatever manner you feel comfortable.'

This is in keeping with the knowing, arch humour that characterises this series, it both wills the listener to feel they are there, and laughs at the pretension of the exercise.

Joujouka has long been a destination for Western travellers seeking the mystical Other (Campion, 1995) and the presenter reproduces this exploration discourse by offering to transport the listener physically and spiritually to a remote and difficult to reach location, for a transformative and healing experience. At the same time, by recognising the listener from the perspectives of the local people and the presenter, the production denies the listener the comfortable anonymity of most radiophonic listening. It fails to adhere to the rule that immersion requires the illusion of non-mediation by referring to microphones and headphones.

The direct address and reference to the listener's bodily presence, the way the presenter answers our imagined responses ('bit awkward when you put your head on the pillow?') place it outside the conventions of audio documentary and recall children's TV programmes, where the audience is invited to shout out answers and the characters or presenters pretend they have heard them. This programme plays with discomfort in more than one way then, by using real-world social scripts in ways that invoke social awkwardness, and by breaking with some of the conventions of adult factual audio. It is impossible to say if the awkwardness is a strategic choice. It might also undermine the presenter's use of the immersive claims of journeying and experience. His final words show this tension in the form of anticipation of rejection:

> We've witnessed a magical rite as we've been here… I hope that by coming on this trip with me you've also got that same thing and that the music of Joujouka has stirred something within you… So what did you get from this evening? Don't go hiking In Cuban heels? Very funny. It's ok you can think about it for a while longer.

Conclusion

'Joujouka' is interesting because of the ways it is different to much recent immersive audio documentary work. It still draws on immersive journalism frames of immersive audio as real and listening as going on a journey to an exotic and hard-to-reach location, so that there is a tension in the programme, which both denies its own mediation and draws frequent attention to it, exhibiting what Bolter and Grusin (2000) call the 'double logic of remediation' in contemporary media technology discourses.

There is another tension in the claims made elsewhere in immersive audio discourse explored in this article: that the listener might not only step inside the reality of a story, but also step into someone else's shoes or be inside their head. There are explicit and implicit claims made for subjectivity and empathy here, yet at other times the frame of the real is enrolled in the service of another central journalistic ideal at odds with empathy and involvement: the ideal of objectivity. By obfuscating mediation and editorial intervention, producers and programme texts are framing the immersive documentary as an objective reality where the listener can see things for themselves.

These contradictory ideas of objective reality and subjective experience both rest on sometimes unrealistic claims about the technical distinctiveness of spatial or immersive audio. In fact, producers make inconsistent use of listener placement within scenes as well as mixing in non-diegetic music and speech, which undermine the immersive production ideals stated in interviews. Where their work produces strong spatial effects that brings them up short when they listen back, producers reject these too, for fear they will engender listener discomfort. Instead, the immersive trope of transportation to another reality is exploited extensively in the less challenging promise of listening as travel and exploration of the exotic.

Objectivity and travel come together in an immersive audio discourse that draws on colonial ways of knowing the world, offering the listener the aural equivalent of the distanced, uninvolved 'commanding gaze' or colonial ear. While it might seem less risky to offer listeners pleasure rather than discomfort, it fails to follow through on the potential for visceral engagement with people and events that is central to claims made about the value of immersive listening for journalism.

Colonial relations structure journalism and documentary discourse, beyond the connotations of this language of exploration and adventure. Spatial audio technology development and use is dominated by public service broadcasters in three former colonising nations: the UK, Germany and France. The novel technology adds to the tools the Global North to provide a privileged listening position from which to know the world, as it cannot be known to those on the ground (Spurr, 1993). Like all journalism in the Global North, spatial audio journalism extracts value from the lives, actions and knowledge of people in the Global South, telling their stories through the recording technologies, cultural conventions and distribution platforms they maintain privileged access to.

Despite claims about their unique affordances for involvement, for transportation or for objectivity, the technological characteristics of spatial audio technologies are not enough to explain their uses and effects (Théberge, 2004, p. 760). Its capacity for more realistic sound has afforded hyperbolic claims about realness that have mainly been deployed to envelop listeners in comfortable and pleasurable travel to exotic places. Producers talk about allowing listeners an active role, but they also link successful immersion with comfort, and equate discomfort with failed immersion. But Joujouka points to some ways that producers have more choices available than they are currently using. They can acknowledge or even contrast the parallel realities of listening at home, alone, with headphones on, and a virtual presence in the mediated storyworld. Drawing attention to mediation by referring to the recording process could prompt recognition of the documentary subjects' awareness of the listener. Such self-consciousness or discomfort might undermine immersion understood as envelopment (Agrawal et al., 2019), but could be highly affecting, evoking a wider range of feelings including negatives ones such as embarrassment, shame, uncertainty. These could potentially recall more vividly real-world experiences of travelling and meeting people apparently offered to the listener in the claims about epic adventures. More importantly, these could help unsettle the disproportionate economy of documentary listening and work towards delivering on the promise to engage listeners with the lives of others.

Biographical note

Abigail Wincott is Associate Professor of Audio Journalism at Falmouth University. Her research explores all forms of factual audio storytelling including news, documentary and heritage, with a special interest in what happens when new media technology or techniques meet existing professional cultures and contexts.

References

References

Agrawal, S., Simon, A., Bech, S., Bærentsen, K., & Forchhammer, S. (2019). Defining
   Immersion: Literature Review and Implications for Research on Immersive Audiovisual
   Experiences. Paper presented at the 147th AES Pro Audio International Convention.
   Audio Engineering Society.

BBC. (2018). *In My Head* *https://www.bbc.co.uk/programmes/b0bj6sk6*

BBC Radio 4. (2015). *Joujouka.* *https://www.bbc.co.uk/programmes/b08hmd23*

BBC World Service. (2018). *Congo: A River Journey.*
   https://www.bbc.co.uk/programmes/p06w66lm

Bracken, C. C., & Skalski, P. D. (2010). *Immersed in media telepresence in everyday life*.
   Routledge.

Bracken, C. C., & Botta, R. A. (2010). Telepresence and Television. In C. Bracken, & D.
   Skalski (Eds.), *Immersed in Media:Telepresence in everyday life*. Routledge.

Campion, C. (1995, Master Musicians Of Joujouka: Chris Campion Follows Brian Jones and
   William Burroughs into the Heart of Sonic Darkness. *The Wire 138*, 26-30.

Chainon, J., & Chainon, C. (2018). *A Pocket Field Guide: 360 Video Storytelling for Editors
   and Newsroom Managers*. The Lookout Station/Global Editors Network.
   https://drive.google.com/file/d/1i8cn54PavfNqRCPIOpe7rOJEqW50u1An/view

Chion, M. (1994). *Audio-Vision: Sound On Screen* (C. Gorbman Trans.). New York:
   Columbia University Press.

de la Peña, N. (2016). Step inside the future: virtual reality will let us understand the world in
   ways photography never could.(MOBILE). *MIT Technology Review, 119*(4), 10.

de la Peña, N. (2012) 'Hunger in Los Angeles'. VR documentary. Available

at http://emblematicgroup.com/experiences/hunger-in-la/

de la Peña, N., Weil, P., Llobera, J., Giannopoulos, E., Pomés Freixa, A., Spanlang, B.,

Friedman, D., Sánchez - Vives, M. V., & Slater, M. (2010). Immersive journalism:

immersive virtual reality for the first-person experience of news. *Presence:*

Teleoperators and Virtual Environments, , rators and Virtual Environments, 2010, vol.

19, num. 4, p. 291-301.

de la Peña, N. and Weil, P. (2007) 'Gone Gitmo' VR documentary in the game Second-Life, no

longer running. Details of the documentary at https://docubase.mit.edu/project/gone-

gitmo/

Dominguez, E. (2017). Going Beyond the Classic News Narrative Convention: The

Background to and Challenges of Immersion in Journalism. *Frontiers in Digital*

*Humanties, 4*10.3389/fdigh.2017.00010

Douglas, S. J. (2004). *Listening in: Radio and the American Imagination*. Minneapolis:

University of Minnesota Press.

Euritt, A. (2019). Public circulation in the NPR Politics Podcast. *Popular Communication,*

17(4), 348-359. 10.1080/15405702.2019.1637522

Goutier, N., de Haan, Y., de Bruin, K., Lecheler, S., & Kruikemeier, S. (2021). From "Cool

Observer" to "Emotional Participant": The Practice of Immersive Journalism. *Journalism*

Studies (London, England), 22(12), 1648-1664. 10.1080/1461670X.2021.1956364

Hacıhabiboğlu, H., De Sena, E., Cvetkovic, Z., Johnston, J., & Smith III, J. O. (2017).

Perceptual spatial audio recording, simulation, and rendering: An overview of spatial-

audio techniques based on psychoacoustics. *IEEE Signal Processing Magazine, 34*(3), 36-54. 10.1109/MSP.2017.2666081

Hall, S. (1992). The West and the Rest: Discourse and Power. In B. Gieben, & S. Hall (Eds.), *Formations of Modernity: Introduction to sociology* (pp. 275-332). Oxford: Polity Press.

Hall, S. (1997). The Spectacle of the 'Other'. In S. Hall (Ed.), *Representation: Cultural Representations and signifying practices* (pp. 223-290). London: Sage.

Jay, M. (1988). Scopic Regimes of Modernity. In H. Foster (Ed.), *Vision and Visuality: Discussions in contemporary culture* (pp. 3-28). Seattle: Bay Press.

Jones, S. (2017). Disrupting the narrative: immersive journalism in virtual reality. *Journal of Media Practice, 18*(2-3), 171-185. 10.1080/14682753.2017.1374677

Krebs, S. (2017). The Failure of Binaural Stereo: German Sound Engineers and the Introduction of Artificial Head Microphones. *ICON, the Journal of the International Committee for the History of Technology, 23*, 113-144.

Lacey, K. (2000). Towards a periodization of listening: Radio and modern life. *International Journal of Cultural Studies, 3*(2), 279-288. 10.1177/136787790000300216

Lacey, K. (2013). *Listening Publics: The Politics and Experience of Listening in the Media Age*. Cambridge and London: Polity.

Madsen, V. (2010). A Call to Listen: The 'New' Documentary in Radio-Encountering 'Wild Sound' and The 'Filme Sonore'. *Historical Journal of Film, Radio and Television, 30*(3), 391-410. 10.1080/01439685.2010.505039

Madsen, V. (2013). Your ears are a portal to another world': The new radio documentary imagination and the digital domain. In J. Loviglio, & M. Hilmes (Eds.), *Radio's new wave: Global sound in the digital era* (pp. 126-144). New York: Routledge.

Maschio, T., & Baumann, K. (2017). *Storyliving: An ethnographic study of how audiences experience VR and what that means for journalists.* Google News Lab. https://newslab.withgoogle.com/assets/docs/storyliving-a-study-of-vr-in-journalism.pdf

Murray, J. H. (2016). *Not a film and not an empathy machine: How necessary failures will help VR designers invent new storyforms.* Immerse. https://immerse.news/not-a-film-and-not-an-empathy-machine-48b63b0eda93#.xrpd2397t

Newton, K., & Soukup, K. (2016). *The storyteller's guide to the virtual reality audience.* The Standford D.School, Medium. https://medium. com/stanford-d-school/the-storyteller-s-guide-to-the-virtual-realityaudience-19e92da57497#.cgxvjzo2h

Osagie, I., & Buzinde, C. N. (2011). Culture and postcolonial resistance. *Annals of Tourism Research, 38*(1), 210-230. 10.1016/j.annals.2010.08.004

Owen, T., Pitt, F., Aronson-Rath, R., & Milward, J. (2015). *Virtual Reality Journalism.* ().Columbia Journalism Review/Tow Center. https://www.cjr.org/tow_center_reports/virtual_reality_journalism.php/

Pullki, V., & Karjalainen, M. (2015). *Communication Acoustics : An Introduction to Speech, Audio and Psychoacoustics*. John Wiley and Sons.

Romigh, G. D., & Simpson, B. D. (2014). Do you hear where I hear?: isolating the individualized sound localization cues. *Frontiers in Neuroscience; Front Neurosci, 8*, 370. 10.3389/fnins.2014.00370

Said, E. (1979). *Orientalism*. New York: Vintage Books.

Spurr, D. (1993). *The Rhetoric of Empire: Colonial Discourse in Journalism, Travel Writing, and Imperial Administration*. Duke University Press.

Théberge, P. (2004). The Network Studio: Historical and Technological Paths to a New Ideal in Music Making. *Social Studies of Science, 34*(5), 759-781.

Tzlil, S., & Nicholas, J. A. (2019). Imagining An Ideal Podcast Listener. *Popular Communication, 17*(4), 333-347. 10.1080/15405702.2019.1610175

Urry, J. (1990). *The tourist gaze : leisure and travel in contemporary societies*. London: Sage.

Van Damme, K., All, A., De Marez, L., & Van Leuven, S. (2019). 360° Video Journalism: Experimental Study on the Effect of Immersion on News Experience and Distant Suffering. *Journalism Studies (London, England), 20*(14), 2053-2076. 10.1080/1461670x.2018.1561208

Verma, N. (2012). *Theater of the Mind: Imagination, Aesthetics and American Radio Drama*. Chicago.

Watson, Z. (2017). *VR for News: The New Reality?* (). http://reutersinstitute.politics.ox.ac.uk/our-research/vr-news-new-reality