Where is the listener? Changing conceptions of the audience in immersive radio and podcasting

Dr Abigail Wincott, Falmouth University.

Working Group MARS (Music, Audio, Radio and Sound)

ExOrdo submission ID number 1349

Abigail.wincott@falmouth.ac.uk Twitter @abigailwincott

Project website https://spatialaudiojournalism.home.blog/

Where is the listener? Changing conceptions of the audience in immersive radio and podcasting

<u>Introduction:</u>

In this paper I want to share my work in progress on a piece of research I've been doing on immersive sound for audio journalism and documentary making. In my earlier work (Wincott, Martin and Richards 2020, 2021) on this topic I focused on the people who make the content, but in this paper I want to talk about the audience or the listener. How are they imagined? How is their listening anticipated and directed by programme makers and platforms, and what can we learn from this about responses to changing technology?

First of all, for those who don't know, immersive audio – which is also called spatial audio, or sometimes 3D audio – is a term used to describe any audio format that allows the listener to experience sound in 360 degrees around them. When you listen to immersive audio, you should feel a sense that the sound scene you are hearing is real, vivid, and that you are in there somewhere, that it's around you. The most common way to deliver this is via a format called binaural, which suits listening on headphones.

Binaural plays a trick on your brain, which, if you have two hearing ears, has trained itself to work out where sounds are coming from, depending on how they hit one ear before the other. People's heads are roughly the same size, their ears are similar in placement and shape – and so if I place small microphones in my ears or in a dummy head and record, your brain should be able to hear the recording as I heard it at the time, giving you a much more detailed sound picture, in three dimensions and a sense of sound all round you.

This technology isn't new at all – it's older than radio programmes. It was certainly tested out in 1881 using telephone lines (Pike, 2019). Speech audio, factual audio, as opposed to music – really took off from the 1920s with radio broadcasting, and fairly early on they developed speakers, so that radio became something that came out of a speaker. It stayed that way for many decades. You can't deliver immersive audio under those conditions, you need to control how it reaches the person's ears for the trick to work. If the sound is coming out of a radio to one side of you and bouncing round the room or the car, then the spatial information is largely lost. For this reason, even though at various times people made

content in immersive formats like binaural, Dolby Surround and quadrophonic sound (Wincott, Martin and Richards 2020), immersive sound never took off.

That's changed in recent years perhaps partly because of its use in cinema and in gaming (Melchior, Churnside, & Spors, 2012) but probably more importantly the relevant microphones have become very, very small and cheap, and people have started to listen on-demand, on headphones. The numbers listening on headphones are still increasing, and that's true all round the world (Mathivanan and Bassari 2022; Stefani 2020), because even in low income countries a mobile phone will come with a pair of headphones. So you have a big renaissance in the use of immersive audio with podcasters, producers and journalists using it and investment at big European broadcasters including BBC, Radio France, RFI and Bayerischer Rundfunk. It's used now for a range of output including music concerts, dramas and – of interest to me – it is used for audio journalism. There is some news reporting, studio discussions, but most often it's used for features and documentaries, on radio and in podcasts.

This field of production, audio journalism, doesn't have the status of cinema or literature. It doesn't even get the attention that TV does or immersive journalism, which is always thought of as virtual reality or 360 video. Yet radiophonic speech of this kind is a hugely important cultural form that has a long history and its own conventions and grammar (see, eg. Crisell 1994: 45; Hendy, 2010; Madsen 2010). It must employ many thousands of people worldwide, and it has large and growing audiences globally via radio and on-demand listening (Pew Research Center, 2021; Statista, 2022).

I want my research to enrich podcast and radio journalism and documentary making; to help people develop the conceptual tools to talk about immersive production in-depth, to articulate issues, share knowledge and to experiment productively with it and make the most of immersive audio as one extra tool in an array of radiophonic techniques.

The problem of the immersive listener

In the first phase of my research in 2019-20 I talked to journalists and producers who were working with immersive audio (Wincott, Martin and Richards 2020;2021). And I found that the presence of the listener in immersive audio is a key concern in the industry. It's the

subject of enthusiasm: the listener is described as a witness to unfolding events, part of the action, stepping onto the stage.

Producers are also positive about the idea that they can use immersive sound to transport listeners to an exciting or exotic location – among a flock of butterflies in the Amazon rainforest, on a boat on a river in Benin going to watch a Vodun ceremony. A boat on the Congo river. Inside caves in France where prehistoric people painted now extinct animals. Inside a nuclear power station. Both these concepts, of witnessing and of being transported, are also common in VR journalism.

My interviewees were also concerned that the immersive listener presents problems and that mishandling the listener results in failed attempts at immersion, for which they need better training. Or even any training at all as most people receive no training in immersive production, those who use it often bought a microphone and taught themselves.

So now in my second phase of research, I'm turning my attention to the 'problem' of the immersive listener and adding to my interviews a content analysis of around 10 hours of immersive factual audio content from the UK, France and Germany (because these are active sites of immersive audio production, because I can speak these languages and the content is often labelled as immersive online – without this it's harder to locate texts to study), and the promotional materials that are with them online – text and images. These all contain information about how the listener is being imagined, and how they are directed in their listening.

So what is the nature, first of all, of the problem of the immersive listener? Well what was clearly articulated by everyone I interviewed was the fact that when you work in mono or stereo, you can consider the listener in terms of informational content: Will they be interested, will they understand, will they interact with the programme by messaging us? But suddenly in immersive audio, the listener is expected to feel as if they are 'there', in the sound scene, because of the nature of 360 degree sound, it is in 360 degrees *around the listener.* The listener must therefore be planned for and the journalist or producer has to plan a scene or picture when they record.

We don't normally plan an audio scene like you might plan a sequence of shots for video journalism, considering framing, perhaps moving items to produce a good shot. We don't do

that because we tend to think about the speech of the people who we will record – what they will say and what we will ask them. We then might record other sounds and we layer it over as we wish in post production. So for example, if I were to interview a football coach, I'd plan to attend a training session and record the sounds of playing and being coached, I might record the sounds of the players calling goodbye as they leave, and I'd record an interview with the coach. I would take all those files back and layer then one over the other, fading them up and down as I wished when telling the story. In other words I would create a generalised sense of place and atmosphere, I wouldn't plan an entire scene spatially, I wouldn't think about where people would be in relation to each other and other things. And I would certainly never, ever have to think about where the listener is, because they're not there in mono and conventional stereo.

Now if you use record in binaural, you need to think what will I do with the listener? Where will they be in relation to interviewees, presenters, and other sounds like the footballers playing or calling goodbye? This issue spatial planning and the place of the listener was articulated to me as a potential problem, illustrated by anecdotes about mistakes or failed attempts at immersion.

What are these failed attempts? The first thing to remember is that when you record in binaural, wherever the microphones are, that's where the listener will be.

Attempt 1) I listened to a binaural investigative podcast, where a reporter went to a café, looking for witnesses to the case in question. As they enter the café I had an amazingly vivid sense of place and space – it felt very 3D and I could distinguish different voices and sounds. But then the reporter, who is probably wearing in-ear binaural microphones – starts walking around and turning as they take us on a tour of the cafe. Now my brain has no visual cues here, and no bodily cues I would have if I was really there and walking around the space. So suddenly the café seems like an unfathomably vast maze of different rooms packed with unknown numbers of people. At times it seems like the café customers are moving past me. I can't work out where I am and what's happening.

Attempt 2) A reporter wears in-ear binaural microphones (so the listener takes their place in the scene) and a guide explains the situation to them. Then the reporter asks them questions, and then finally addresses the audience to summarise what they have seen.

Listening back the reporter has a sense they have confused the placement of the listener by having someone else's voice emerge from the listener's position.

Attempt 3) An interviewee told me about a radio news programme team who used a dummy head (like this one here https://en-de.neumann.com/ku-100) to do record a studio discussion and put the head onto the round table, with people all around. The listener is where the mics are of course and so listening back the team found this unacceptable, they felt they were a head on a table with people talking behind them, which was deeply uncomfortable.

Attempt 4) A documentary maker wanted to record a discussion at a meal in a restaurant. They 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.

What do these failed attempts have in common? The French café presents a technical issue, as the reporter broke the auditory illusion by moving too many variables, so the listener's brain (my brain) couldn't decode the space. The reporter who speaks while wearing binaural microphones probably also breaks the spatial illusion, by presenting conflicting information to the listener about who and where they are.

But 3 and 4 I would say are to do with social scripts. Newton and Soukup (2016) found VR users were more comfortable with immersive journalism and experienced a sense of 'being there' when they had a familiar social script that helped explain 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. I can't conceive of a familiar social script where I am a head on a table while people discuss local politics round me! And people speaking all round me would be very likely to evoke a sense of vulnerability, unless you could provide more cues that would make it okay.

The producer at the meal table rejected the first possibility as 'eavesdropping' an auditory equivalent of being a 'voyeur'. In this case, there is a social script, but it offers only a negative role for me as a listener.

So thinking about social scripts can help us understand why some kinds of listening might feel problematic. And it could help producers anticipate problems and avoid them by asking how the listener know what they're doing or know what's going on. It is possible to offer some direction to the listener. The most common framework or social script is the traveller – sometimes a leisure tourist, often more of an explorer. This is signposted in various ways including being invited to pack your suitcase and go on a sonic voyage ('Si Loin Si Proche', RFI) and Congo: A River Journey (BBC) will 'take you on an epic adventure in sound in the Democratic Republic of Congo'. The reporters and presenters here are framed as guides on the journey who set the itinerary and explain what you are experiencing. Here both social codes (traveller and guide) and media codes (reporter as guide to events) overlap to produce a comfortable and familiar script for the listener.

It is one that is common in mono and stereo too of course, though we have no physical sensation of being in a scene and are not acknowledged directly. That traditional position of the listener as an invisible but all-hearing traveller is supremely comfortable. All sounds and voices, people's stories and struggles are laid out for me as the listener. I hear everything but am asked to feel responsible for nothing, and am unlikely to feel a sense of eavesdropping or awkwardness. Therefore the 'problematic' or failed attempts are seen as failed partly when they challenge familiar media scripts to make me uncomfortable. I think this is problematic.

If we take, for example, the journey on the River Congo, the implied listener is not someone living or working on the Congo. The listener is invited to find the journey exotic and exciting, and the trope of the exploring listener conforms to colonialist and or touristic gaze – in this case perhaps a touristic or colonial *ear* – always consuming sounds of the 'Other' without obligation.

I'd like to mention next a programme made for the BBC called 'Joujouka', because there is evidence these programme makers have tried to avoid invoking this colonising ear, or perhaps to unsettle that relationship a little. The programme is about traditional musicians in a village in Morocco, whose music is meant to invoke the god Pan and bring about a trance like state that can be healing. Westerners have long travelled to hear them. There are a few ways the production avoids the colonising tourist ear – for example by naming all the musicians who greet the presenter and the listener. Perhaps the most striking thing is

the way the presenter addresses the listener directly throughout, and refers to their 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!' He also refers to the way the musicians have stopped chatting because they saw the recording equipment – calling on the listener to be aware of the difference their listening makes.

I love the way they are trying to enhance my sense of 'being there', of a special experience, and at the same time challenge me to be responsible for my listening, to be aware of myself from the perspective of the 'other' in the programme. And yet this programme also made me cringe. The direct address and reference to my bodily presence, responding to imagined responses ('don't be shy') seemed undignified. It reminded me of kids TV programmes where they ask the children to shout out answers and pretend they've heard them. So one reason this is, for me, a partially 'failed' immersion is that I am drawing this time not on a 'real world' social code but a media code to interpret this programme – and the only one I have is kids TV, which I find inappropriate.

Conclusion

As new media technologies are adopted, they present opportunities and trigger anxieties over the role of the listener, and the kind of listening that is desirable. This is seen in strategies by current affairs podcasters to create a sense of live co-audiencing (Euritt, 2019) without broadcast. And here and in VR journalism research we can see some anxiety over the use of immersion and the role of the immersive listening/user.

There are idealised roles for the listener as witness to events, as an explorer on an exciting journey, but perhaps more interestingly, if we look at things framed as mistakes or problems, we can see some interesting things about how the immersive listener is imagined and how their listening is being directed. Failures often result from a lack of a social or media script for the listener to draw on, or a negative or inappropriate script.

I want to argue that this moment is the perfect one to question and experiment – before production and editorial norms are established in immersive audio. We should listen to these judgements about successes and 'mistakes', because can tell us something interesting about changing ideas of the audience and listening to audio journalism and how

technological change might shift these ideas. Once patterns have been established for a long time, it becomes difficult to see them, especially in a field like audio journalism that doesn't received as much critical and academic attention as, say, film.

And I would also like to propose that we might use an awareness of discomfort and comfort to reconsider some of expectations about listening that developed in the colonial age – expectation the listener is entitled to hear everything, the whole truth about the world, without any doubt or confusion. That they should have the world and its people laid out for their enjoyment and education, without themselves feeling seen and heard, obligated or intimately connected to the people whose lives they witness. Clearly audio journalism already sometimes addresses these issues, but not often enough, and causing discomfort is a difficult thing to do when we are concerned to attract big audiences, in a competitive media marketplace.

One of the claims often made about immersive journalism is that we can elicit a meaningful sense of connection between the audience and other people, empathy, seeing life from someone else's perspective. These claims are common in VR journalism (de la Peña, 2016; Elmezeny, Edenhofer, & Wimmer, 2018; Van Damme, All, De Marez, & Van Leuven, 2019). The same claims were made by interviewees for this research and they come up again and again in web texts and presenter talk in the immersive audio documentaries I've listened to. It seems to me that one way to do this is to deliberately harness discomfort of the listener, rather than always avoiding it, to deliberately play with distance and closeness, confusion and clarity to really open up the potential of this journalistic tool to trigger a stronger sense of the connectedness of all of us on the planet.

References

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

Elmezeny, A., Edenhofer, N., & Wimmer, J. (2018). Immersive storytelling in 360-degree videos: An analysis of interplay between narrative and technical immersion. *Journal of Virtual Worlds Research, 11*(1), 1-13. doi:https://doi.org/10.4101/jvwr.v11i1.7298.

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, 20*(14), 2053-2076. doi:10.1080/1461670x.2018.1561208

Crisell, Andrew (1994), *Understanding Radio,* 2nd ed., London: Routledge.

Euritt, Alyn. 2019. 'Public Circulation in the NPR Politics Podcast'. *Popular communication 17*(4), 348–59.

Hendy, D. (2010). Listening in the dark: Night-time radio and a 'deep history' of media. *Media History, 16*(2), 215-232. doi:10.1080/13688801003656249

Madsen, V. M. (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. doi:10.1080/01439685.2010.505039

Mathivanan, S. and Bassari, A. (2022) 'Futuresource Consumer headphones market report', 14 Apr 2022. https://www.futuresource-consulting.com/reports/futuresource-consumer-headphones-tracker/?locale=en#:~:text=The%20headphones%20market%20grew%20by,are%20reported%20at%20sell%2Din.

Melchior, F., Churnside, A., & Spors, S. (2012). Emerging technology trends in spatial audio. SMPTE Motion *Imaging Journal, 12*1(6).

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

Pew Research Centre (2021) Audio and Podcasting Fact Sheet, June 29. https://www.pewresearch.org/journalism/fact-sheet/audio-and-podcasting/

Pike, C. W. (2019). *Evaluating the perceived quality of binaural technology.* Retrieved from http://etheses.whiterose.ac.uk/24022/

Statista (2022) Number of podcast listeners worldwide from 2019 to 2024, Feb 18. https://www.statista.com/forecasts/1147560/podcast-reach-uk#:~:text=As%20of%202021%2C%20there%20were,28%20million%20listeners%20by%202026.

Stefani, B. (2020), 'Tournage: RFI: Le Son 3D Dans Tous Ses États', *Mediakwest,* January, pp. 16–18.

Wincott, A., Martin, J. and Richards, I. (2021) Telling stories in soundspace: placement, embodiment and authority in immersive audio journalism. *Radio Journal: International Studies in Broadcast & Audio Media 19*(2).

Wincott, A., Martin, J. and Richards, I. (2020) *Spatial awareness: State of the art and future needs of spatial audio journalism*. https://spatialaudiojournalism.home.blog/