Title: Cuts, fades and layers: Audio production interfaces and mental schemas for radiophonic storytelling.

Author: Dr Abigail Wincott, Falmouth University, UK   abigail.wincott@falmouth.ac.uk

Music, Audio, Radio and Sound Working Group

ExOrdo submission ID: 3719

Introduction: A mismatch between what features makes say and what their editing software does

This paper is about a new piece of research for me, into the way that audio programme makers think through the production tools they use. My focus of is on the use of spatial audio to make radiophonic documentaries or features.

There has been a growth in the use of spatial audio by journalists and documentary-makers in recent years. Large broadcasters in France, Germany and the UK have invested in research and development, and are encouraging its use by more of their programme-making teams, from news journalists to long-form documentary teams.

Over the past few years, I've interviewed some of those people who use spatial audio. They claim it required them to approach their work in a radically different way: avoiding conventional techniques such as layering sounds in the final mix, making frequent cuts and using narration. Instead, they say, they have to plan a series of sound scenes in three dimensions, thinking about how to place the various people and sounds in the scene, including the listener. (Wincott, Martin and Richards 2020, 2021). This is more akin to how people work in video journalism or in audio drama, so it's a departure for audio feature makers.

However, in course of my research I've also analysed many hours of spatial features and documentaries and have found little evidence that programme-makers are practising these spatial grammar 'rules'. They are in fact cutting, narrating and layering sound and in these and other ways, reducing the spatiality of their audio work (Wincott 2023).

In this paper I want to begin to interpret this apparent contradiction between what they say and what they do as a product of the relationship between the way programme-makers approach the process of storytelling and the editing software they use. I am going to argue that editing software is metaphorical and so are the conceptual frameworks that programme-makers use to describe their work: like when they talk about the spatial sound scene. We think and make audio work through these conceptual metaphors, which have affordances and limitations. And ultimately I think there is a mismatch between the conceptual framework of the spatial scene, and editing software metaphors. Because this is developing research, I'll point to some questions I now want to answer, as I spend some time learning more about alternative editing and mixing software.

Deconstructing the radiophonic feature

It is useful deconstruct what a radiophonic feature is, in order to explain its meaning for those who are unfamiliar with it but also because deconstruction and redescription are a useful way to see afresh the details and peculiarities of familiar cultural forms.

So **features are factual**, not fiction. They vary greatly in length and can be as short as two minutes (usually as part of a longer programme) or as long as an hour. In English, they are sometimes referred to as 'speech' programmes, opposed to 'talk radio', which is, on the face of it, a confusing distinction. Talk radio and conversational podcasts are characterised by improvised talk, largely or entirely unedited, live or 'as live'. They cover an issue in a non-narrative form, while features are a **narrative audio form**.

Like talk radio and conversational podcasts, features are not a single-voice narrative, so they differ from other narrative audio forms like audio books, lectures, or reporters' dispatches from the field. Features are what some media researchers call '**polyphonic**' (Charron and Charlton 2019; Krieg 2000), in that they are woven together from many voices. They have this in common with most journalistic texts but in radiophonic features the polyvocality and the narrative structure are created through a particular post-production process. Features are composed of many sound recorded elements that are **edited and mixed together** to form a whole.

The elements can be categorised as follows: there is usually a **narrating voice** that addresses the listener directly (though some features have no narrating voice). All features include several other voices in the form of excerpts of **recorded interviews** with people who have a connection to the subject matter. Those excerpts might also feature the questions and responses of the interviewer-presenter, so a lot of talking forms the explicit informational content of features.

The inclusion of background sounds or '**atmospheres**' is also characteristic of the feature form. Atmospheric sound provides information too, for example about the size of a space, its geographical location, the weather, the time of day, number of people present and their mood. Features also include more foregrounded sounds of things happening, which is sometimes called '**actuality**'. For example, in a feature is about police repression of protest, we would expect to hear an altercation between police and protestors, or the sounds of protest. Features also often use **music**, **archive clips** and sometimes **sound effects**.

These sound elements are then *woven together* in the mixing process by the programme-maker. Or you might say that the programme-maker *orchestrates* these multiple voices and sounds. And because radiophonic features are narrative, despite their polyvocality, they are not as fragmentary or open in meaning as sound art or field recordings and sound walks. Though they are complex and rich, their makers offer multiple voices and atmosphere as a way to provide depth and editorial balance. The makers of radio features generally work hard to close down the possibility of multiple meanings, and guide the listener's interpretation of the sound elements.

They do this through para-texts: on air announcements, and podcast or web text descriptions. They do it through the presenter's speech, which structures the story, explains and instructs the listener on how to evaluate speakers and their relationship to the subject (Clayman, 1991). But programme makers also direct listeners in how to interpret the meaning of the voices and sounds through the order in which they introduce those voices and sounds, through cutting out some content but leaving other content in, and through choices in the way they juxtapose content, by cutting away from one voice or sound, fading it down below another or raise one above another to signal where the listener's attention should be directed.

It is a characteristic of the feature format that the speakers (interviewees) do not interact directly with each other, they are juxtaposed in response to each other's words only through the choices made in editing, mixing and narration (Clayman, 1991).

I hope this shows how important is to pay attention to those **rules of the grammar of audio features making**. The listener's access to the apparent polyphony or openness of meaning of radiophonic features is fairly tightly directed, by this narration and *by the very process of editing and mixing.*

Let's deconstruct editing and mixing. If you listen to a feature, you will hear that sometimes one sound ends before another starts. There may be a gap, a silence, between the two. Sometimes more than one of these sound elements plays at once – the **layering** mentioned above. Any sound can be made louder or quieter over time as well. We refer to this in English as '**fading up'** or '**fading down'**. The practice of arranging sounds that layer and overlap, making them louder and quieter, is called '**mixing'** as if they were ingredients in a cake or paint colours.

I have tried to redescribe what a feature is and how it is put together in the most factual and descriptive terms, but *mixing, weaving, fading up and down are all metaphors*: they describe one thing as though it were another and in doing so suggest how we ought to think about the first thing (Blackwell 2006). It is not possible to talk about the art of features making without choosing metaphors and these metaphors are not only linguistic, they take the form of the audio editing software we use and the visual representations that software provides on a screen (García-Crespo, Ramahí-García and García-Mirón 2021).

Typical computer editing and mixing interfaces

Although the media form of the feature is not at all visual, the process of editing now involves engagement with a visual display. Let's have a look at a typical screen that people will engage with as they mix a feature. Below (fig. 1) is an image from Adobe Audition, there are many other programs, but they follow a similar schema. They have these representations on screen of audio tracks, layered one above the other. They have a timeline that runs left to right (like text written in the Latin alphabet used in English, the language that dominates software design). There are visual representations of sounds themselves on screen too. When using software like this you will also interact with several windows, click on icons and store your content in a folder structure, which will be familiar to users of other software.

This interface affords and enables the production of a feature as we are familiar with it. Sound recordings (presenter links, interviewees, actuality, music) can be layered on these tracks, cut and moved around in time. You can fade up and fade down. When you are happy you can 'mix down' or export the final mixed version.
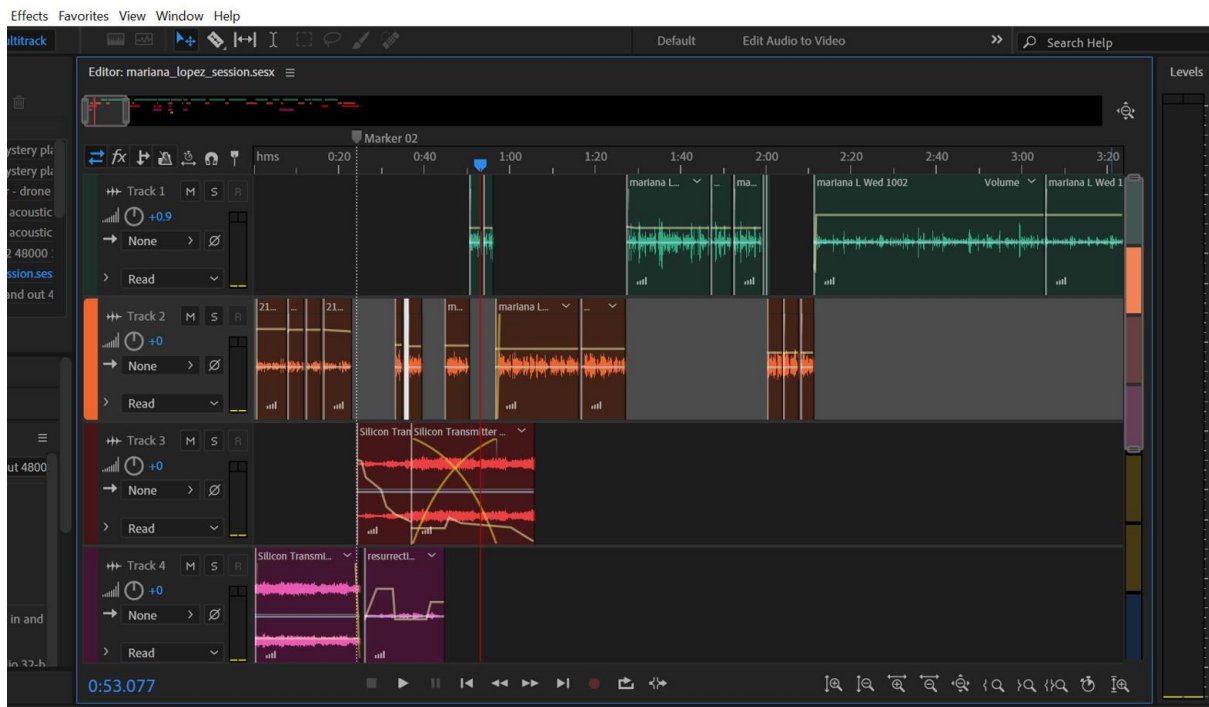
Fog. 1: A screenshot of a multitrack mixing window in Adobe Audition.

This process produces a richly layered, impressionistic sound cloud, where the programme maker directs your attention by fading down one thing and fading up another as well as by addressing you directly in their narrating speech. *But this is not what spatial audio producers claim is the way they should be producing their features, which is conceiving of them as a series of carefully arrange three-dimensional sound scenes.*

Spatial audio metaphors

Spatial audio is produced using a range of technical formats but the most common is binaural stereo, which can be mixed in any of these standard editing programs. What these formats have in common is the way they can make the listener feel both that the sound is external to them, and all around them in pretty much 360 degrees. They are often referred to as immersive too, and discourses of spatial audio production have a lot in common with other immersive media discourses (Wincott 2023)

As I mentioned at the start of this paper, one of the things that characterises the discourse of spatial audio storytelling is the claims that a story should be a series of spatial scenes, and that the listener will not just be told about what happens by a reporter or presenter – they will enter the scene, to **experience a reality** unfolding, at first hand. According to the producers, you plan and capture these scenes but you don't need to explain them, because this 3D scene is so real, the **listener can hear for themselves**.

Wincott, A. 'Cuts, fades and layers: Audio production interfaces and mental schemas for radiophonic storytelling', IAMCR 2023.

According to this discourse of spatial storytelling, not only should you not over-explain, you shouldn't layer sound at all, because if you work in binaural, the spatial effect is destroyed if you layer sounds.

However, I analysed 11 hours of documentaries and I found that they were all pretty much exactly the same as non spatial documentaries in their grammar or format. The presenter and interviewees spoke close to the microphone, often either in mono or mixed stereo but very front and centre so they go inside the head. Non-diegetic music and sound effects are layered over the top of spatial recordings. The presenters record links in studio, which explain the story in the past tense. *Producers are not adhering to their own rules of spatial audio feature grammar*.

So I want to ask the question – why not? And it occurs to me that one of the reasons might be that producers' ideals of spatial production – that it's just a scene that 'speaks for itself', where the listener steps in and understands for themselves – that might not be achievable.

But perhaps it is also a product of the software they use to make the content in post production.

To talk about the feature as a scene that the listener steps into is a metaphor, one that has a long tradition that goes back to the early days of radio, when it was frequently heralded as a 'portal' to another time and place (Madsen 2013). And to talk about the feature as woven or layered – these are also metaphors. But they are not compatible.

Metaphors of audio production

Much metaphorical thinking is grounded in bodily experience of moving in space (Lakoff and Johnson 1980, 1999) so one thing 'follows' another, sound goes 'up' and 'down', a clip is 'brought to the front' or 'on top', and so on.

Our bodies continue to engage with metaphors when we move a fader on a virtual or physical mixing desk up or down, the linguistic metaphor – fade up, fade down – is influenced by and presumably influenced the material metaphor of the desk design.

Metaphors might be motivated by material realities but they aren't inevitable. When I first worked at the BBC they still had some studio mixing desks produced by the BBC themselves, where you moved the faders *down* to make sound louder and *up* to make sound quieter, and people referred to 'opening' and 'closing' the faders.

Metaphors are often cultural and historically contingent, building on earlier formats and technologies. Computer interfaces borrow from mid-twentieth century offices with desks, rubbish bins, documents and folders, for example (Blackwell, 2006). In the early days of audio features making, people borrowed terms and techniques from photography and film: cross-fading and mixing, layering, cutting or montage, speaking of sound pictures or sound films (Madsen, 2010) and radio production may well also have influenced film-making (Hendy 2013). The word fade itself is a visual metaphor. In French the word for a fade is 'fondu' – melted, so metaphor is clearly culturally contingent as well. We have already observed this in the way sound runs from left to right on editing program timelines, like writing in the Latin script, which opens up questions about global power

inequalities at the point in time when technologies are developed too, and whose work and interests their metaphors represent.

Lance Sieveking, often credited as the pioneer of the radio feature in the 1920s and 30s, had a 'control panel' built that he could use to coordinate the live mixing of performances in separate studios (Hendy 2013) . He wore white tie and tails like a conductor of an orchestra to conduct the feature – live classical music performances were of course a logical analogy to draw on, when embarking on this new media form. And though other metaphors have come to dominate, the feature is still conceived sometimes through **the metaphor of musical composition** or the conducting of musical polyphony. For example the NHS Symphony (BBC, 2018) produced in binaural by radio documentary maker Laurence Grissell. The metaphor of radio feature as musical composition or performance is associated with content that has no presenter-narrator and is more open to interpretation. Perhaps it also reflects the background of the individual features-maker, and whether they are musicians or music producers, or if they have worked in news journalism, or theatre, or television documentary, for example.

Metaphors can become significant across many media at certain points in time in a culture. David Hendy (2013) writes about Sieveking's metaphor of the '**kaleidoscope'** – the name given to his feature of 1928. The kaleidoscope as a metaphor implies an element of chance, each time you turn the kaleidoscope, a different picture forms, it also has a circular symmetry at odds with the linear narrative that has come to dominate the radiophonic feature. It is hard to imagine thinking kaleidoscopically while using the editing window pictured in fig.1.

Neil Verma (2012) writes about the way early radio drama producers **blocked scenes** for performances (planning and rehearsing actors movements in the space of the studio, a kind of **choreography** of actors bodies as they speak their lines), taking theatre as their analogy and skills base. Actors' voices moved around therefore in three-dimensional space in relation to the microphone and therefore the listener at home. This means that, even though the recording technology was not what we would refer to as spatial, their productions were in some ways much more spatial than modern audio, where everything is recorded 'flush against the microphone' (Verma 2012).

Sound engineers' audio-technical discourse is likewise characterised by metaphors such as sound recordings as **faithful**, as bringing a concert hall into the living room, or transporting the listener to a concert hall, that frame not only the sound but the role of sound engineers themselves and have thus been the site of contestation in the industry (Krebs 2017).

The digital audio workstations or DAWs – editing software people use today - give us an interface on a screen that represent the tracks of the **reel to reel tape** that the music industry used for so long to record separate vocals and instruments which would then be mixed together onto the master reel. Radio was using the same ¼ inch tape technology until around 20 years ago. It works for both features and music mixing.

This layout – this graphical user interface – facilitates the production of linear narrative and the layering of sounds for a rich acoustic montage. I can record binaural stereo and add it in to these tracks. But there is nothing here to encourage me to conceive of my programme as a series of spatial

scenes someone can enter and experience. For one thing, there is no representation of space on the screen – there is loudness and time order, but the detailed spatial information in a binaural recording is not represented and cannot be manipulated.

There is no representation either of the boundaries around a scene, within a story, as we might have, for example, available to us on a Word document where we can create section headings and chapter headings, if we were telling stories in print.

Fig. 2: From top, clockwise : B360 ambisonics encoder, Adobe Audition 5.1 surround sound mix window and the Harpix plug-in.

What about some of the software that has been designed specially for other spatial sound formats than binaural stereo (see fig. 2) ? There is software that does allow for the movement of sound recordings in space. They still use screens – so they are trying to find ways to represent in two dimensions, three dimensions of space, plus the other dimension of time passing. Fig. 2 shows some screenshots of software representations of three dimensional sound mixing.

Some of the European public service broadcasters are already developing and beginning to use **object based audio** (https://www.bbc.co.uk/rd/projects/orpheus). This technology has the potential to allow listeners to experience optimized spatial sound wherever they listen and on whatever devices, but it requires programme makers to record all elements as separate sound objects, which

are transmitted separately, rather than mixing the programme to their satisfaction and saving that final, fixed mix. This seems likely to be a challenge for features makers, given what we have seen about their current ways of working.

Some of the people I have worked with in my research tell me some of their colleagues resist using new spatial and object based audio technologies, seeing them as inconvenient, irrelevant, or imposed from outside. And as I indicated at the start, even those who do use them, do not use their full potential, or always achieve what they say they wish to achieve in their work.

New research

The 'problem' of spatial audio producers not following their own advice can be understood in terms of conceptual metaphors – linguistic and material. Some of the existing metaphors for features making – composing, storytelling, layering, mixing – are supported and echoed by the metaphors of the most common DAWs that producers and journalists know how to use and have access to. But other metaphors, more spatially oriented, are not. The materials of the screen, keyboard and mouse used to interact with the software and sound must surely play a role in the shape these metaphors take.

Metaphors matter because they structure the process of producing features in ways that impact on their meaning, purpose, craft and content. Metaphors also say something about the imagined user (Blackwell, 2006) and can be gendered and classed. People who use software tend to make do and adapt, or they may adapt their work until such a point comes where the divergence from what they want to do becomes too great when they become aware – as music composers have about music composition software (Duignan 2008).

The fact that spatial innovation rhetoric doesn't match reality, that there are reports of tensions and unwillingness to use new production tools in some contexts means, I would argue, that it is time to examine the production tools available to features makers.

In my new research I want to understand more about the range of conceptual metaphors that people already use to structure the way they create audio features, including orchestrating, choreographing, telling stories, weaving and so on. I want to think about the relationship between these metaphors and the values or priorities they entail – the difference the metaphors make to the content.

I then want to consider the role played by individual backgrounds of features-makers, and the common connections between a career making features and doing other things – for example working in a radio newsroom, writing poetry, directing theatre, composing music or sound engineering.

The questions to be asked of special spatial editing interfaces like those discussed above are

– what kind of metaphors are deployed in these interfaces and processes?

- How do they affirm or conflict with the metaphors common to features making and the values and conventions of factual programme making and journalism, such as highly directed meaning-making or clarity of meaning, linear narrative, balance and polyvocality?
- How do producers of features engage with these interfaces, and how do they adapt their work and their understanding of that work to fit the software? How do they adapt that software for their own features-making habits and needs?

I want to use a method for this work that takes into account the fact that metaphors are not only linguistic artefacts – words like 'scene' or 'layer' or 'compose'. They are also processual and material. And we make programmes with our bodies, moving and working in real-world spaces, interacting with other people, with materials as well as with software (Fauré 2013). Metaphorical frameworks can involve all of these things as well as the language we use to talk about our work.

As features makers are encouraged to learn to use spatial and object based production technology, that research and development is structured largely by sound engineering knowledge frameworks. The clash between expectation and reality that began this paper, and the resistance of some programme-makers to taking up these new tools and technologies shows there is a need for cultural and social research into audio production, that can bring different professions and their knowledge paradigms into dialogue.

References

BBC (2018) The NHS Symphony. broadcast 30 June 2018, BBC Radio 3. https://www.bbc.co.uk/programmes/b0b86s0j

Blackwell, A. (2006). 'The Reification of Metaphor as a Design Tool'. *ACM transactions on computer-human interaction 13*(4), 490–530.

Charron, J. and Charlton, S. (2019) 'La Polyphonie Du Journal Télévisé'. *Sur le journalisme 8*(2), 138–53.

Clayman, S. (1991) 'News Interview Openings: Aspects of Sequential Organization'. In P. Scannell (Ed.), *Broadcast Talk* (pp. 48-75). Sage.

Duignan, M., Noble, J. and Biddle, R. (2010) 'Abstraction and Activity in Computer-Mediated Music Production'. *Computer music journal 34*(4), 22–33.

Fauré, L. (2013) 'Analyser les pratiques discursives radiophoniques: nouveaux enjeux et perspectives'. *Les cahiers de praxématique* 61, 5-17 https://doi.org/10.4000/praxematique.2429

García-Crespo, O., Ramahí-García, D. and García-Mirón, S. (2021) 'From the Tool to the Spatial Metaphor: Graphical User Interface and Audiovisual Creation in New Media'. *Convergence 27*(6), 1696–713.

Hendy, D. (2013) 'Representing the Fragmented Mind: Reinterpreting a Classic Radio Feature as "Sonic Psychology"'. *Radio Journal: International Studies in Broadcast & Audio Media 11*(1), 29–45. doi: 10.1386/rajo.11.1.29_1

Krebs, S. (2017) 'The Failure of Binaural Stereo: German Sound Engineers and the Introduction of Artificial Head Microphones'. *Icon* 23, 113–44.

Krieg, A. (2000). 'Analyser le discours de presse'. *Communication. Information Médias Théories Pratiques 20*(1), 75-97. 10.4000/communication.6432

Lakoff, G. and Johnson, M. (1980) *Metaphors We Live By*. Chicago: University of Chicago Press.

Lakoff, G. and Johnson, M. (1999) Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought. Basic Books/Perseus.

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). Routledge.

Wincott, A. (2023) 'Un bon voyage sonore: Avoiding listener discomfort in immersive audio documentary'. In peer review.

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/