Immersive and Personalised Podcasting Using AI-driven Audio Production Tools

Jemily Rime

PhD

University of York,
School of Arts and Creative Technology

July 2024

Abstract

From the first podcasts to the current diverse content landscape, there has been a drastic expansion of what we consider a podcast. Genres have emerged and died, survived, and thrived, and the podcast landscape in 2024 is nothing like the one from 2004. So what will podcasts in 2044 sound like, and what does Next-Generation Podcasting entail? After highlighting the gaps within podcasting innovation literature, this thesis proposes definitions and frameworks to advance podcast research, informed not only by context, but by involving a group of podcasters whose opinions and expectations of podcasting are gathered through interviews and workshops. This participatory methodology is not only valuable to map the current practices and production habits of professionals, but also to substantiate the development of new tools for immersive and personalised podcasting. The peculiarities and requirements linked to such tools are explored, and the iterative development process that occurs results in the creation and evaluation of a web-app for modular podcasting, Podulr, and automatic chapterisation algorithm, pod-CLIPR (Podcast Chapter Localisation through Intelligent Pattern Recognition).

The contributions of this thesis are: **1/** A definition of podcasting alongside a framework for podcasting innovation; **2/** A contemporary archetypal workflow for podcasting; **3/** A summary of expectations of producers for Next-Generation Podcasting, views on new technologies, and a reflection on the systems already in place and how they'll need to adapt to enable it; **4/** A system for automatic podcast audio chapterisation, pod-CLIPR, comprising of a sound recognition model combined with a rule-based algorithm, and its evaluation; and **5/** A reflection on participatory design for developing media tools and a practical application in the form of the modular podcasting web-app Podulr. This work has interdisciplinary impact, in podcasting, audio production, interactive media, and participatory design for new media tools.

Acknowledgements

Although a PhD is oftentimes seen as a solitary process, no part of this research would have been possible without the help and involvement of many – from those whose research influenced me, to those who proofread with assiduity, to those who cheered with metaphorical pompoms; their contributions to my work were immense. Without them, there would be no literature review, no experiments, no software – there would be a musician and audio nerd who thinks of research but fills her days teaching piano, rehashing *Für Elise* and *Let It Go* for the hundredth time with students.

There are countless who were instrumental to this project, but I want to shine a light on a few who made this PhD possible.

First, a necessary nod of deep gratitude to funders, institutions, and staff that have permitted this research: the University of York - especially, those in the Music department, and across the campus, in the Audio Lab; XR Stories - and particularly there, Damian Murphy for signing on this project proposal in the first place; the BBC and BBC R&D – it is thanks to the ARP that this PhD was able to extend beyond the pages of a dissertation, and reach real producers and professionals.

There would be no thesis without the excellent team of supervisors who have helped me and guided me along the way:

Tom Collins, thank you for taking a chance on a random musician from France with a Physics Bachelors. I wouldn't have believed I could make it if it weren't for your encouragements and your trust. Thank you for teaching me what being a good researcher entails, and for your attention not only to the science, but to the words, and most importantly, to the people in your lab. I am so grateful to have had you as a supervisor, and so thankful that you carried on working with me even after leaving York.

Alan Archer-Boyd, thank you for joining this project with unmatched enthusiasm. Your input and expertise were invaluable, and the height of

your determination to ensure this work not only contributed to knowledge but also had a legacy, was only matched by that of your camaraderie. Slack might not allow me to download our conversation thread, but I will remember every exchange.

Jude Brereton, thank you for your guidance, enlightened opinions, and seeing me through the last year of this work. Getting your feedback on this thesis allowed me to take a step back from the pages that I had been staring at for years.

Catherine Robinson, thank you for sharing your knowledge and skills since the genesis of this project.

Additionally, I would like to thank Chris Pike and Jon Francombe who wore the "industry supervisor" mantle in the past, for seeing me through the beginning of this PhD.

Throughout the past four years, I have been a part of several groups, although I found my research-home primarily within the Music Computing and Psychology Lab, and the BBC R&D Audio Team.

To the past and present members of the Music Computing and Psychology Lab (Kyle, Alex, Adrián, Chenyu), thank you for sharing ideas, work, and laughs even with thousands of miles sometimes separating us. I will remember our meetings fondly.

Thank you to everyone in the BBC R&D Audio Team (and especially those of the former PAAE group, Kristian, Lawrence, Matt, David, and Jay, who endured my daily stand-ups for a year), for being fantastic colleagues, inspiring researchers, and for letting me join your meetings for almost four years, instead of six months (a testament to how much I enjoyed myself there).

Somewhat outside these groups, but still completely entwined with daily life, thank you Kim – it is my absolute honour to be your friend and research buddy, to share and receive academia memes to combat existential crises, and to never know when I walk in a room you are in if we are going to be dressed the same.

Thanks to my participants for making this research what it is – a co-created, participatory project. Thank you Jenn for being there at every step of the way, and Jana for taking a chance on new software.

Finally, I would like to acknowledge all those in my personal life, family and friends, who made it possible for me to pursue this degree.

To my parents Audren and Christophe for always nurturing and facilitating curiosity. I wouldn't have been able to do this work without the example of jusqu'au-boutisme you've set.

To Sydney for being a perfect neighbour, duet partner, friend, and sister through the ups and downs of research.

To Julia, for all the shared rants, and staying close by in spirit, even though we lived far away from one another these past years. Your friendship means the world to me.

To Chris, for your endless kindness.

Merci à Kertanguy, un havre de paix pendant ma première année de thèse.

Pour Percy, tu ne sais pas lire, mais tu mérites tout un roman de remerciements.

Pour Annie, l'académicienne éclairée, et Marc, l'artiste curieux, qui n'ont jamais cessé de m'inspirer avec leur amour d'apprendre.

And, to Piers; your love and support were instrumental to my well-being, and I'm forever grateful you managed to remain genuinely interested in this research for four years even though you "never really vibed with podcasts" (there has never been a truer proof of love).

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for a degree or other qualification at this University or elsewhere. All sources are acknowledged as references.

Although the material presented in this thesis is the result of personal work, collaboration was a key part of this endeavour:

All producers and colleagues involved in conversations, interviews, and workshops, helped shape this project and its outputs.

The sound recognition model used within pod-CLIPR was created by Kong et al.

Chris Baume created an end-point and interface to showcase this model easily. In the first stages of development of Podulr, this endpoint was used as a way to visually represent the chapterisation process and its code was used to trial pod-CLIPR, before being replaced by bespoke scripts.

Related publications

Chapter 2 is based on the following journal paper.

> Rime, Jemily, Chris Pike, and Tom Collins. "What is a podcast? Considering innovations in podcasting through the six-tensions framework." Convergence 28.5 (2022): 1260-1282

Chapter 3 and 5 are based in part on the following conference paper.

> Rime, Jemily, Jon Francombe, and Tom Collins. "How Do You Pod? A Study Revealing the Archetypal Podcast Production Workflow." In ACM International Conference on Interactive Media Experiences (IMX), pp. 11-18. 2022.

Chapter 3, 4, and 5 are based in part on the following journal paper.

> Rime, Jemily, Alan Archer-Boyd, and Tom Collins. "How Will You Pod? Implications of Creators' Perspectives for Designing Innovative Podcasting Tools." ACM Transactions on Multimedia Computing, Communications and Applications (2023).

Chapter 4 is based in part on the following conference paper.

> Rime, Jemily "And, Scene! Role-Playing With ChatGPT-Generated Personas To Inform Design Decisions." In International Conference on Computer-Human Interaction Research and Applications (CHIRA), 2024.

Chapter 6 and Chapter 7 are based in part on the following journal paper

> Rime, Jemily, Alan Archer-Boyd, and Tom Collins, "Podcasting, the next chapter: Podcast Chapterisation through Intelligent Pattern Recognition (pod-CLIPR)" Under Review for IEEE/ACM Transactions on Audio, Speech, and Language Processing.

Although the following publication is not represented in this thesis, the project it describes was carried out to refine my web audio programming skills:

> Rime, Jemily, Tom Collins. (2022, June 28). Resonance Choir: The Renaissance madrigal meets spatial audio. Web Audio Conference 2022 (WAC 2022), Cannes, France.

Contents

List of Tables

List of Figures

1 Introduction

1.1 Preamble

When I moved to the UK for my undergraduate degree, I downloaded the BBC iPlayer Radio app (since then, rebranded as "Sounds" app) on my phone to immerse myself in British content as I transitioned from being a Parisian lycéene into a first-year Physics student. I would listen to podcasts on my commute into central London, and dull out the noises of a busy dorm room with the back catalogue of the podcast version of Desert Island Discs [1]. I grew into the habit of turning to podcasts for education, entertainment, and sometimes more simply, companionship. After this, podcasts became a reflex, an expectation of long drives, or spring cleaning afternoons, as well as quiet moments dedicated to a favourite show in a comfortable chair.

This duality of listening modes was to me the most appealing feature of the medium. I enjoyed making time to listen to a program I was particularly looking forward to, but also being able to couple listening with another activity. Sharon and John (2019) speak of these two approaches as an introduction to their article on engaging with ideal podcast listeners. Moreover, scientific literature is abound with motivations for podcast listening. Chan Olmsted

[1] https://www.bbc.co.uk/programmes/b006qnmr

and Wang (2020)(p.691) for instance sees seven possible incentives: *Audio platform superiority, Social interaction, Entertainment, Information, Personal/communal identification, Companion/connection, Escapism/pastime.* As this doctoral project evolved, it was interesting to see myself reflected in this and other data collected about podcast listeners in the Western world. Chapter 2 investigates these "typical" podcast listeners in depth – if you are also an avid podcast listener, I invite you to compare yourself to the average listener, as variations from this epitome of the English-speaking podcast audience influences our relationship to and expectations of the medium – and therefore of this research

This thesis lies at the intersection of several disciplines: audio production, computer science, music technology, and media studies. Its primary goal is to investigate the impact of new technologies on podcasting, and how they might shape the future of this medium. It covers a span of three and half years of research, combining essays on digital communications and media, interviews and workshops with podcasters, as well as details of the development process and implementation of a new podcasting web-app which enables automatic chapterisation (segmentation into chapters) of audio files, and an evaluation of its underlying algorithms.

Since the title of this thesis is almost entirely composed of media and technology buzzwords, I would like to begin by examining each constituting term, and how they will be employed throughout these chapters. As Van Den Eede (2020) does in their chapter within "*Relating to things: Design Technology and the Artificial*" (Wiltse, 2020), this will not only set the scope of the research, but also detach these concepts from their "buzz-wordiness",

grounding the work in reality, rather than relying on the first, inescapably biased, impression they can leave us with. Each of these concepts will be defined in more rigorous depth in Chapters 2 and 3, but this introductory chapter should provide enough context so that the research aims (Chapter 1.2) and questions (Chapter 1.3) are both comprehensive and comprehensible.

Immersive - From the push to transition to a Metaverse office space (Orel, 2022b), to investigating the uses of Audio Augmented Reality (Yang et al., 2022), "immersion"[2] seems to be almost synonymous with the word "progress" in the tech world. For some innovative Virtual Reality (VR), Extended Reality (XR), and Augmented Reality (AR) projects, immersion is a goal in itself - whether that is to improve the underlying technology or showcase some new technical advancement. This thesis takes the approach of seeing immersion as a tool rather than a goal in itself.

What is the point of creating immersive media? Is it to captivate audiences? Is it to create unmissable narratives? Is it to engage someone in a particular point of view? Is it to create or strengthen communities? Is it to make a profit? Is it to build something new? Is it to drive innovation forward? Rather than categorically answering these questions, this research asks a large group of media creators (around fifty different podcasters have contributed to the work described here across three and a half years) to share their own expectations and goals for immersion. By using participatory de-

[2] "*Immersion is the experience of losing oneself in a fictional world. It's what happens when people are not merely informed or entertained but actually slip into a manufactured reality*" - (Rose, 2015, p.3)

sign (PD) [3] and co-creation [4], the issues of bias that would naturally come with a subjective answer to these interrogations are avoided. Chapter 3 will cover the many benefits (and the few limitations) of using PD and co-creation in innovative creative technologies, further justifying why these concepts are at the forefront of our methodology and overall approach.

Personalised - In the early stages of this research, this word was "interactive" [5], but "interaction" was found by Mütterlein (2018) to contribute significantly to immersion, and is not necessarily indicative of the individualised and user-specific meaning behind personalisation. Broadly, personalisation is the process of catering something to someone. It can be as seemingly trivial as changing the interface of a website when a user is logged in, as fine-tuned as Spotify playlisting (and more recently, the evolved form of automatically generated user-centred Daylists, with evocative titles such as "*soul crushing relatable Monday afternoon*", "*rainy day napping Wednesday night*" or "*enlightened millennial Sunday evening*"[6]), or as engaging as the interactive storylines in Netflix's Bandersnatch [7]. Of course, these examples are all within the realm of technology and media, but, more broadly, personalisation can be applied to any one thing meant for a person to consume – a meal

[3] "*Participatory design is a democratic process for design (social and technological) of systems involving human work, based on the argument that users should be involved in designs they will be using, and that all stakeholders, including and especially users, have equal input into interaction design.*"(Hartson and Pyla, 2019, (p.355)

[4] "*Consumer co-creation refers to research methods that involve end users in developing ideas and concepts for the client to commercialise. These include using social media, online communities, workshops, discussion groups or in-depth interviews*" (Association for Qualitative Research (AQR), 2022)

[5] Interactivity "*refers to the degree to which users of a medium can influence the form or content of the mediated environment*" (Steuer, 1995, p.11)

[6] https://mashable.com/article/spotify-daylist

[7] https://www.nytimes.com/2019/01/04/arts/television/bandersnatch-black-mirror-netflix.html

can be customised (e.g. *"Could I please have a Ceasar salad, but replace the chicken with cheese, the anchovy dressing with béarnaise, hold the bacon"*), a jean jacket can be covered in patches accumulated over several years to reflect the eclectic musical tastes of the wearer, a museum tour can be articulated around a particular topic to fit its audience and so on...

> *"Personalisation is a pervasive phenomenon in all human activity, encompassing decoration, re-configuration, modification, customisation, and tailoring of human-made objects like cars, jewellery, clothes, houses, workplaces, tools, software and so forth. People have created whole cultures of personalisation – like wine-tasting, fashion, gastronomy, and car-customisation – where choices express the individual tastes and personalities of its members" - (Oulasvirta and Blom, 2008, p.1)*

As Oulasvirta and Blom (2008) hints towards, there is an infinite number of things that can be personalised, and in turn, infinite ways for them to be customised. This makes the term "personalised" versatile and particularly tricky to define. From this point onward, I use the definition given by Blom and Monk (2003): "*personalisation is a process that changes the functionality, interface, information content, or distinctiveness of a system to increase its personal relevance to the individual*". This concept is covered in more detail in Chapter 3. Combined with the highly specific context of this PhD, we have to wonder, what does personalisation look like (or perhaps more accurately, sounds like) for podcasts? There is already personalisation at play when a listener browses through a catalogue and picks out a programme for instance, but there can also be a level of content personalisation, with for example dynamic ad insertion, that targets each listener with customised ads within their episodes. Accordingly, there are two facets of personalisation explored through this research: on the one hand, personalisation of content,

with changes in the podcast itself, and on the other hand, personalisation of interface – that is modifications at the platform level. These same facets are explored in web personalisation research (Burns et al., 2013; Frias-Martinez et al., 2006). This comparison also serves as a disclaimer that a lot of the methods and ideas investigated in and extended through this work are taken from the field of software engineering, particularly web application development, as podcasts are first and foremost online objects, that rely almost completely on digital tools and processes.

Podcasting – "A podcast is a piece of episodical, downloadable or streamable, primarily spoken audio content, distributed via the internet, playable anywhere, at any time, produced by anyone who so wishes."

This is the definition Chapter 2 will conclude with when answering the question "What is a podcast?". It is collated through a thorough review of literature, informed by historical and analytical material. Since the publication of this review, some podcasting scholars have wondered why the need to still seek to define a well-known media (Sharon, 2023). To this, I will say that the nature of podcasting, as further explored in Chapter 2, is metamorphic and that therefore, it is necessary to go through these ontological hoops systematically when engaging in podcasting research. A Podcast in March 2024 is wildly different from a Podcast in March 2004. At a smaller scale, a podcast this year follows completely different trends and conventions than it did last year. The fast-paced, innovative nature of the medium, combined with its growth and the global interest it gathers, means that there is never a way to crystalise the term into a definition. The one presented above is valid at the time of writing, even if some of its aspects can already be dis-

cussed (e.g. does a podcast need to be episodical? There have been several standalone podcasts in the past year, that are to podcasts what films are to TV shows).

The act of creating a podcast – "podcasting" - although sometimes linked to the performative act of presenting, most often refers to being involved in one or more parts of the production process as a whole, from booking guests, to recording, to editing, to distributing and marketing a show. In other words, a sound engineer, whose role is to edit two-hour-long discussions into succinct thirty-minute shows is " podcasting" just as much as would be the guest (or host) who rambled on about their favourite cheese for that time. Whether that rambling takes place in a living room or a professional radio studio also makes little difference to that appellation now. And just like podcast listeners have an archetype, so do podcast-makers, and their attributes will be covered in Chapter 2. Henceforth, I will use the terms "podcaster", "podcast creator" and "podcast producer" interchangeably, not just to make the prose more agreeable to read, but also because podcast making includes a variety of roles (as will be explored in Chapter 5) that make these umbrella terms necessary to podcasting research.

Using – this transitive gerund underlines the direction of the work: podcasting aided by new technologies, as opposed to new technologies applied to podcasting.

AI-driven – Technically, two words, although, this portmanteau works well in overviewing our approach to artificial intelligence (AI, for the rest of this thesis) in the context of media production.

The term AI carries both weight and a variety of meanings depending on

who uses it or hears it. Where the notion refers to the borderline aspirational concept of machines able to replicate and act upon human intelligence and understanding (Moritz, 2024; Born et al., 2021), the term is currently used as fact by researchers, industry, and public alike. This separates the aim of perfectly reproducing human intelligence within a machine, and the reality of approximating human responses more or less successfully based on available datasets of examples of a specific task (Birtchnell, 2018).

Sub-categories of AI, like machine learning, deep learning, or reinforcement learning have been responsible for many recent milestone technological innovations (Bieda and Panchenko, 2022). These statistical models can be used as ways to achieve some specific complex task, attempting to replicate human understanding not overall, but looking at particular domains of expertise. This is why Born et al. reflects that AI is still best used as a way to solve narrow issues, rather than give a broad overview and solution to an issue that a human would easily access.

The AIs explored in this thesis are always encapsulated within another system or interface; indeed although some creators have the technical background to pick a model and implement it, it is a rare occurrence (Chapter 5), and thus these AIs must be wrapped in a more user-friendly coat, like plugins (e.g. Waves' Clarity Vx audio restoration plugin) or apps (e.g. Riverside FM or Descript for smart podcast recording and editing tools). Reading this thesis does not require a deep knowledge of machine learning (ML, for the rest of the thesis), however Chapters 3 and 7 cover some of the basics behind the algorithms and models mentioned.

This thesis did not begin with the intention of restricting itself to AI

applications to podcasting tools, indeed, the focus was on new technologies as a whole, as will be apparent in Chapter 3. But, as the PD process evolved, some particular AI audio tools (notably, sound recognition via audio event tagging) stood out as potential candidates for a deeper investigation.

AI for audio usually lands in one of three categories: analysis (Davies and Plumbley, 2007; Ellis, 2007), generation (Kreuk et al., 2023; Mehri et al., 2017), and processing (Défossez et al., 2019). For non-musical audio, the work is often done in the audio domain (the realm of waves, Fast Fourier Transforms, and spectrograms). Analysis comprises the tasks that break down and replicate understanding of sound, answering questions such as "What is the pitch of this door creek?", "What frequencies in this spectrogram make up a female voice?", and "What is being said in this conversation?". This category of AI-audio can yield results of its own, for instance, sound event detection (SED) or speech-to-text, but is often combined with audio generation. Indeed, a key feature of ML, is the "learning" aspect, meaning large training datasets are first analysed before a model is able to replicate the datasets' features. Typical candidates for non-musical audio generation are speech (e.g. Lyrebird[8]) and sound effects (e.g. Nemesindo[9]). Although this thesis touches upon the ethical considerations of making and using such models, other academics' work focusing solely on the issue will be able to cover more ground in more depth (Barnett, 2023). Still, the thought of ethical AI, if such a thing exists, remains at the forefront of this research.

By the time my work began and conversations around the topic of new technologies and automations within the field of audio production were a

[8]https://www.descript.com/lyrebird
[9]https://nemisindo.com/

frequent occurrence, the term AI was already firmly defined in the cultural zeitgeist. Oftentimes, it carried with it creators' fears and apprehension towards the eventual replacement of their human labour by that of machines (Birtchnell, 2018). There was no way to change how the term AI was perceived by stakeholders, but a technological pragmaticism was applied, breaking down AI into the current realities of a wide concept formed by its many components. AI tools were presented as such, but they were not assumed, in fact, to be intelligent. The term "machine learning" or specifics regarding the type of architectures and models were used to explain the functioning of these tools or solutions. This ensured that the term AI, although still loaded with personal assumptions, represented more accurately the scope of this project: an assemblage of learning, analysis, generation, and processing tasks, that can culminate in tools or methods that solve a finite subset of problems (Birtchnell, 2018).

In the rest of this thesis, when the term AI is used, it refers to this particular outlook, and whenever discussed with participants or stakeholder, it was this viewpoint that was defended.

The current state of AI at the point of writing means there is still a need for human intervention in most AI-assisted or AI-driven tasks – but more and more, systems attempt to remove human expertise from these workflows. As this doctorate focuses on creators and their work, I cannot accept, condone, or encourage the possible removal of humans from AI-assisted tasks through technological innovation. This thesis and its composing studies highlight the need for machines helping but not replacing human producers, augmenting their work without removing creative agency (Birtchnell, 2018; Born et al.,

2021).

Audio Production Tools – to the modern producer, this is possibly synonymous to digital audio workstations (DAW), like ProTools, Adobe Audition, Reaper, or Logic; but the breadth of the term "audio production tool" captures more, whether these are analogue or digital, highly specific plug-ins or swiss army knives online audio editors. It also brings about the methodological decision of a "creator first" approach (more on this Chapter 3) – where innovation is there to aid the producer, and not facilitate the desires of a broad listener demographic. Even if the latter would be a valid approach, the yields would be different: by focusing on the creator, one can not only draw theoretical conclusions on the role of new technologies in what is sometimes called in the industry "Next-Generation Podcasting" (NGP), but also produce concrete, co-created, evidence of new tools for podcasting in the form of functioning, formally evaluated, R&D software.

This lexical overview has hopefully introduced the key concepts at play in this thesis. Before carrying on, I would like to address two intertwined matters: audience, and tone. This thesis is intended for academics, researchers, podcasters, and more generally, interested parties of the podcasting industry and public. I hope that the language used contributes to a smooth reading; I will do my best to define and cover all necessary material so there is no pre-requisite knowledge required to understand the conclusions drawn from this research, although, some prior familiarity with media production, especially, audio production, will surely make the writing more approachable. This work is, in nature, academic. Some of this thesis has been presented at conferences, or published in journals, mostly to scientific publications or

proceedings in the fields of engineering, computing, and technology. This entails a specific tone, relying on passive voice and a vague yet omnipotent first-person plural "we". For the sake of keeping with these disciplines' written and grammatical conventions, Chapter 1.2 - Chapter 9 will keep to some of these rules, but "I" will be used throughout to reflect the personal nature of the research, and the necessity to acknowledge the role and impact of the researcher in participatory processes (Frauenberger et al., 2015). When "we" is used, please read "I and the readers" or "we as a society", as relevant to the context.

This PhD was carried out in close partnership with industry, specifically, the BBC. By associating with the BBC and BBC R&D, I was able to apply my research not only to independent podcasting, but to the vastly different exigencies of a major broadcasting company. There is an unavoidable bias that comes with working with a set industry partner. Beyond restricting the research in geographical scope, it pushes the researcher to adopt the hyper-specific expectations and language that come with stakeholder involvement. PD using not just BBC producers, but independent producers and producers from other media companies alike contributes to tackling this bias, and this will be covered in more depth in Chapter 3. The benefits of having an industry partner were numerous, but to name a few: the industry connections with producers, the free access to a wide library of content, and the experience and know-how of engineers who have been designing similar production tools for decades. The research outputs and tools put together during this doctoral project will be available to use by the BBC, but, as much as possible, will be made available to the public to prevent a form of corporate logo-phagism [10],

[10]Neologsim as in, consuming, amalgamating, appropriating, of knowledge

which could be considered unscientific, if the goal of science is to advance knowledge everywhere, for everyone.

1.2 Research Aims

This section will detail the aims of this research, breaking down "immersive and personalised podcasting using AI-driven audio production tools" into three main axes:

Research Aim 1. Mapping the habits and expectations of podcasters

Understanding more about a group and its expectations is key when building new tools and solutions for its members. As we've touched upon earlier, "podcaster" is a wide term, encompassing a multitude of individuals – but as a group, they constitute a special demographic, that has not undergone much academic scrutiny. Harman (2018) would call this group an "object" as it cannot be broadened out to concepts it's included within, or broken down into its integral parts without losing its essence. This object, "podcasters", stands alone, and is defined by its relationships with other objects. As persons (ourselves, other individual objects, as we cannot be broken down into smaller components of ourselves), we perceive this group through our eyes and the preconceptions they carry. Podcasters can be "white men in attics who have trouble sticking to their production schedules", or "a diverse team of young journalists in a media network", depending on your own background, views, and prior experience of "podcasters". But there are some attributes to an object that are undeniable – these are called "real attributes" (Harman, 2018; Van Den Eede, 2020) – and being able to know them eliminates some

(but not all) of the individual variations in how we think of "podcasters".

Quite like a cartographer's job is invaluable for sailors to recognise shorelines, a researcher can map boundaries for concepts and objects to facilitate further exploration in the field. Research aim (RA) 1 focuses on building such boundaries, discovering, defining and refining the real attributes of podcasters to ground further podcasting research in an evidence-informed baseline.

Research Aim 2. Exploring the peculiarities of immersive and personalised podcasting

The breadth of the terms "immersive" and "personalised" has been mentioned in Section 1.1, but this thesis will be particularly interested in their direct application to podcasting: What is an immersive podcast? How do we personalise audio content? Can we feasibly create and distribute such experiences? What are the risks and costs of exploiting AI technologies for media production?

The specificity of these terms in the context of podcasting is interesting for several reasons: first, it gives us insights into the trends and evolution of the medium; second, it showcases the adaptability of new media to innovative technologies; third, it reflects on the possible applications and reach of these innovative technologies to our cultural landscape.

Research Aim 3. Investigating and documenting an application of participatory design to the development of an AI-driven Next-Generation Podcasting tool, all the way from conception, to functional software

There has been significant research into PD in the field of software development; from its applications to User-Centred Design (UCD) (Jones, 2018),

to its integration within other development and design techniques (Ferrario et al., 2014), the literature abounds with successful examples of PD and its applications to different areas, such as medicine (De Croon et al., 2014), education (Danielsson and Wiberg, 2006), or in social sciences (Freire et al., 2011). Reporting on the application of a tried and tested method like PD to a nascent medium and research field will yield interesting results; both because it'll push for this doctoral project to have some concrete, usable, outputs in the form of software, but also because it'll serve as an example of small-to-medium scale UCD tasks for next-generation media, a type of content and art with hyper-specific requirements in terms of planning, production, and distribution.

Through PD and co-creation, this tool will focus on the process of chapterisation to fulfil two separate stakeholder wishes:

- The simplification of the chapterisation process and easy annotation of segments within a new programme or existing library of shows

- The facilitation of "modular" podcast production - that is programmes that are arranged into different versions of themselves to fit different listening scenarios or preferences.

Although AI ends up being the main focus of this thesis, new technologies that enable immersive and personalised experiences regardless of their architectures were first considered, as will be detailed in Chapter 3.

1.3 Research Questions

While investigating these three aims, I will answer the following research questions (RQs):

- **RQ 1:** What is Next-Generation Podcasting?
- **RQ 2:** How can we feasibly create and distribute immersive and personalised podcasts?
- **RQ 3:** Why is modular podcasting attractive to creators, and how can it be facilitated without making the production process more complex?
- **RQ 4:** What are the perceived benefits, risks, and costs of exploiting AI technologies for podcast production?

1.4 Overview of Chapters

Before diving into the contents of this research, please find a short summary of the work presented.

Chapter 2 - Literature review: Contextualising podcasts This chapter answers the question: "What is a podcast?", via a review of the literature, investigating podcasting history and its evolution. This question is necessary to examine, as it'll set a basis for the rest of this research, and is required to answer RQ 1. The definition of podcasting arising from this analysis – centring on episodic audio, convenient both to produce and experience – takes into account recent changes, providing an up-to-date description of the term, useful for further research on the topic. It also addresses the question: "How do we design new ways to produce and listen to podcasts

without denaturing the medium?", a key component of RQ 2 (in the rest of this section, the relevant research questions to specific material covered will be given in brackets). By reflecting on the essential features of podcasting and the necessity for innovation in this interdisciplinary medium, a framework of six tensions is proposed as a means of grounding and potentially boosting innovation.

Finally, this chapter introduces the term "Next-Generation Podcast" (NGP), exploring the different meanings of a term that has been under little academic scrutiny, by looking at our understanding of what "next-generation" means for other media (RQ 1). Understanding the implications of immersion, personalisation and interaction on the advancements made in other innovative media allows us to get an impression of what NGP will sound like.

Chapter 3 - Literature review: Innovative production tools for new media This chapter looks at the production workflows of professionals in various media and explores how being aware of such practices enables new habits to be seamlessly integrated within existing structures (RQ 2 and RQ 3). The gap in academic knowledge regarding podcast production will be highlighted, particularly when compared to the prolific documentation of production methods for radio, TV, movies, and video games. This chapter also maps the various technologies that are changing the way we make, share, and listen to audio online (RQ 1). How AI and new technologies can be harnessed to facilitate podcast production and improve listener experience is detailed (RQ 4). Looking at the capabilities of these technologies, it is assessed how feasibly they could be implemented within podcast production or distribution tools. The various formats and systems in place to deliver

and consume interactive audio media are compared and reviewed, looking at how other media have transitioned from static to interactive products, specifically focusing on enabling technologies and standards. Because there is no consensus on what format interactive online audio should take, potential avenues and examples are investigated, as well as how and if these could be feasibly applied today (RQ 2).

Chapter 4 - Methodology In this chapter, the methods undertaken in this research are examined, both in terms of overall approaches to new media and audio engineering research, but also the specific experimental and analytical procedures used. Audio-visual tools design techniques, and the different actors invested in the development of such products are detailed. Other possible methods are acknowledged, and why this thesis takes a "creator-centric" approach is justified. The motives and uses of PD and iterative software development (ISD) are delved into. By investigating these practices, an understanding of how agile software development can be used to swiftly implement new media tools in existing production workflows is gained (RQ 2). Beyond defending this method of software design, best practices of such methods are investigated, alongside how to transfer these recommendations to the specifications of designing a podcast production tool. The methodology used for gathering creative practitioners' opinions, impressions and thoughts on their work is detailed(all RQs), presenting the advantages of relying on a mixture of quantitative and qualitative research to assess the context of media creation. Data analysis practices that will be used in this thesis are highlighted, including the pros and cons of iterative development, and the bias and error mitigation conducted throughout the research.

Chapter 5 - Podulr: designing a next-generation podcast production tool This chapter introduces the process and flow of a modular podcasting app (Podulr), giving an overview of each step undertaken for its development and their tangible results (RQ 3). It details the results of an exploratory study of potential innovation in the field of podcasting (all RQs). Sixteen podcast creators were interviewed about their work and what they wanted from next-generation podcasts, in order to understand the requirements and expectations of tools built to create new forms of audio-based programming. These interviews shed light on what podcast creators envision as NGP (RQ 1), and reveal the archetypal podcast production workflow. Combining these findings, how the workflow could be modified to include new steps that will help to realise podcast creators' visions is identified (RQ 2). Through a combination of qualitative and quantitative analysis, the technologies that podcast creators associate with Next-Generation Podcasting are detailed, and the dual need of both listener-centric and creator-centric innovations, to improve listener experience but also to unleash new creative possibilities nascent in the production workflow, is highlighted (RQ 1). This chapter also details the workshops carried out with seven creators, to further refine their expectations for NGP, and influence the design of a podcasting tool. The concept of "modular podcasting" is introduced, as a way for the same program to be presented in different ways to users depending on preferences.

Chapter 6 - Insights on Chapters and Modular Podcasting This chapter deals with the design process, implementation and evaluation of the NGP tool (Podulr) built as part of this doctoral project within the scope of

RA 3. The features of this personalised podcasting app are detailed and its limitations are underlined (RQ 2 and RQ 3). The tool enables the production of modular podcasts through automatic segmentation of audio files using sound recognition and audio tagging. Listeners can be oriented towards the version that suits them most. The tool is built in collaboration with podcast creators through ISD, as set up in Chapter 5.

In order to set a benchmark and scope for such automatic segmentation, the nature of chapters is investigated through a study asking 10 BBC podcast producers to annotate 49 5-minute podcast excerpts. This creates a dataset of chapterised podcast audio, POD 49, that can be used to evaluate segmentation solutions.

Chapter 7 - Podcast Chapter Localisation through Intelligent Pattern Recognition (pod-CLIPR) This chapter details the maths and development of pod-CLIPR, a system that uses the natural changes in soundscapes present within a podcast to determine its chapters, and conveniently allows creators to re-order these chapters into different versions of the same programme.

This chapter includes the results of an online questionnaire that enabled us to filter a list of tags used to match the specific context of podcasts, as well as a thorough description of the rule-based approach used. The evaluation of pod-CLIPR is also presented. The accuracy and usefulness of pod-CLIPR are evaluated in a study conducted as part of a BBC placement (RQ 3 and RQ 4). From this comparison, pod-CLIPR is found to produce plausible segmentation, on par with expert producers.

Chapter 8 - Podulr in practice This chapter describes different appli-

cations of the modular podcasting tool described in Chapter 5, 6, and 7. It gives an overview of the final form of Podulr and its functionalities, specifically focusing on how pod-CLIPR is integrated to the tool (RQ 3). It covers the final explorations and remarks of a group of producers via case studies (RQ 3, RQ 4). The case studies take the form of podcast projects produced using this tool in collaboration with creators. The process through which the podcasts are produced is detailed, and the creators' assessment of the tool is discussed. Relying on this data, the best use cases for this tool are discussed and necessary improvements and future work are pointed out. (RQ 3, RQ 4)

Chapter 9 - Conclusions and future work A discussion of the conclusions drawn across the previous chapters. Including, a more in-depth look at the contributions made through this research, as follows:

1.5 Overview of Contributions

- A definition of podcasting alongside a framework for podcasting innovation.
- A contemporary workflow for podcasting
- A summary of expectations of producers for Next-Generation Podcasting, views on new technologies, and a reflection on the systems already in place and how they'll need to adapt to enable it
- A pipeline for automatic podcast audio chapterisation, pod-CLIPR (Podcast Chapter Localisation through Intelligent Pattern Recognition) comprising of a sound recognition model combined with a rule-based algorithm, and its evaluation

- A reflection on participatory design for developing immersive media tools and a practical application in the form of the modular podcasting web-app Podulr

2 Literature Review: Contextualising Podcasts

2.1 Introduction

Do you have a podcast? You could be the creator of one of the 473,870 active shows on Apple Podcast in August 2023 (Lewis, 2023), or have published one of the 85,047,441 episodes made available there since 2005 (CNET, 2011). With an exponentially growing library and listenership since the first podcasts in 2003 (Wallick, 2003), and 1 out of 10 UK adults planning to start a podcast in 2022 (Podnews.net, 2022), it is clear that podcasting has become a key feature of our media landscape. If the present for podcast creators around the world is a seemingly boundless space, filled with encouraging promises of things to come, the future of this medium is still unclear.

In 2004, a Google search on "podcast" gave 6,000 search results; in 2005, 60 million (Berry, 2006); and in 2024, more than 6.5 billion. This is paralleled by an ever-growing amount of listeners and content created around the world (RAJAR, 2020). Whether the podcast medium is set for further growth, a plateau, or decline, the future will bring opportunities for podcasting to evolve and respond to new trends and changing expectations, as well as to leverage the development of new state-of-the-art audio technology and tools (Benito et al., 2018; Uhlich et al., 2017; Forrester, 2013), which could

alter the means and outcomes of podcast production and experience.

For a person selected at random in the US in 2023, the probability of them having listened to at least one podcast in their life was 0.64 (Beniamini, 2023). Although most of us will therefore be familiar with the foundational ideas behind podcasting, there is a grey area when it comes to drawing a definite line separating one piece of audio content from another. For instance, does an audiobook count as a podcast? It matches the Oxford Dictionary's definition of a podcast: it is a digital audio file available for download on any portable device (*Oxford English Dictionary*, 2015, podcast entry). Yet, audiobooks are not counted as podcasts in most market research surrounding podcasting (RAJAR, 2020; Beniamini, 2023) and on most on-demand platforms.

This conceptual fuzziness becomes even more unclear when thinking of the future. Our first research question "*What is Next-Generation Podcasting?*" attempts to pinpoint an industry-specific and elusive term that relates to the future of podcasting. The abundance of new technology will presumably modernise the podcast format (Berry, 2016), raising an important question around ensuring the integrity of its development: how do we design new ways to produce and listen to podcasts without denaturing the medium? Of course, this question, together with RQ 1, can only be answered once a working, current definition of "podcast" is established, more detailed than the one provided by most dictionaries, so that its nature is understood before proposing a framework for podcasting innovation.

I begin this exploration into the definition and limitations of podcasting by assuming those limits are real and fixed, and as objective as possible,

even though they will necessarily be subject to discussion, as each person – be they listener, academic, or creator – might have differing opinions on the topic, stemming from their individual relationship with the medium.

Looking at the past and present of podcasting should provide sufficient perspective to postulate what a podcast is, beyond simple technical requirements. As will be demonstrated, the "nature of podcasting" will need to be able to withstand the many technological innovations that will change how podcasts are made and consumed, and hence should abstain from relying on specific, easily dated technology as a means to define it. For instance, the definition of podcast as a "*a downloadable digital audio file distributed over the internet using RSS, designed to be played back on a computer or personal MP3 player*" as given by Markman (2012) (p.552) is now dated. People no longer only use computers or MP3 players to listen to podcasts, and downloadability, although still of considerable importance within the podcasting community, could be questioned when looking at the statistics of downloaded versus streamed podcasts for the major podcast providers in 2022, for instance, on Apple Podcast, 13.7% of downloads to 86.3% of streams (The Simplecast Blog, 2019).

This investigation into the nature of podcasts will introduce the important concept of "tension" in podcasting, where the medium is pulled between two competing concepts (e.g., "universality" and "uniqueness"). This is inspired by relational ontology theories, like Actor-Network Theory [1] or Object-Oriented Ontology[2], both somewhat constructivist approaches to re-

[1] "*Actor–network theory (ANT) is rooted in science and technology studies. As a method for in-depth research it has now been used in other areas of science as well. ANT focuses on the connections that are being made and remade between human and non-human entities that are part of the issue at stake.*"(Dankert, 2012)(p.46)

[2] "*Object-oriented ontology (OOO) is an intellectual movement in the arts and humani-*

search and ontology that supersede essentialist explanations of phenomena or innovations.

These tensions have been highlighted following a reflection on what podcasting has been and became as well as informal conversations with industry professionals. The aim of establishing a theoretical framework from a review of literature is a way to adequately justify and encase this thesis. I will hypothesize the different aspects of podcasting which are in tension with one another, and later, confirm them through a historical, contextual, and ontological analysis. Figure 2.1 illustrates these relationships in a network, showing the concept of podcasting at the centre of a series of tensions, pulling from either side. The six pairs are:

1. Personalisation and Automation
2. Independent and Mainstream Production
3. Unique and Universal content
4. Current Audience and Possible Demographic
5. Immersion and Interactivity
6. Art and Technology

I will demonstrate that podcasting is, above all, a medium that has relied on these tensions to define itself. It will become apparent that thinking of these as a framework when designing new ways to produce or listen to

ties sharing certain affinities with both phenomenology and Actor-Network Theory (ANT). It is a philosophically realist position often at odds with existing currents in postmodernism and critical theory. The best-known idea of OOO is that objects "withdraw" from all direct human and non-human contact, so that relations between things are always indirect and must be accounted for rather than taken for granted." (Harman, 2019) (p.1)

podcasts is essential, as they are immutable attributes that the medium has entailed since its inception. Delving into these various trade-offs will bring to light the areas of the literature surrounding podcasts that would benefit from further research, lacking details and studies to provide a complete, evidence-led picture of what podcasting is today.

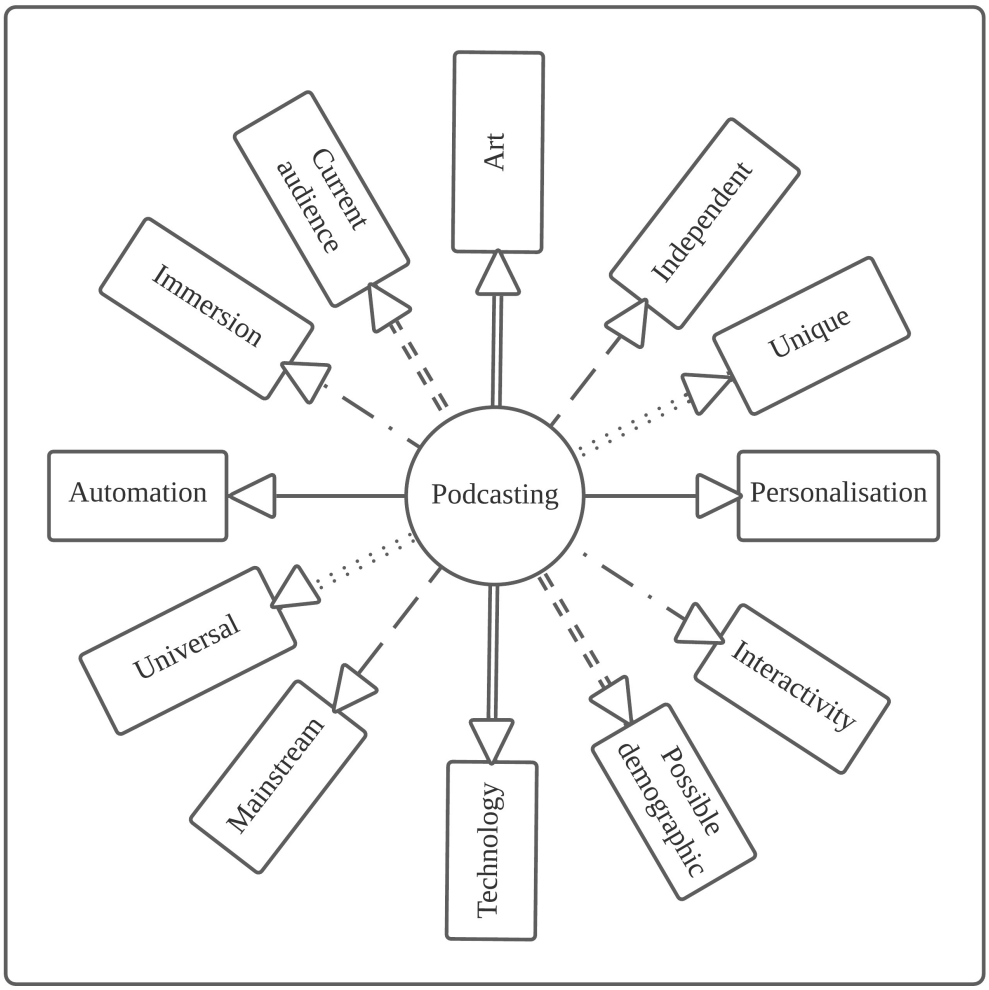


Figure 2.1: The Six Tensions Framework for podcasting innovation, with the concept of podcasting at the centre of six pairs of concepts essential to the medium. The hypothesis presented is that the balance within the pairs must be kept throughout any innovation for the nature of podcasting to be preserved.

In order to validate this proposed framework for podcast innovation, "podcast" must be defined by looking at the origin of the medium and its evolution, including how other media have influenced its form. I will therefore be contextualising podcasting by first looking back at its creation and development. The literature from the time will be reviewed and put into perspective by analysing it through a contemporary lens, informed by the reality of what podcasts became and what current studies say about the content of podcasts and the people who listen to them. This will demonstrate that the tensions in Figure 2.1 have been evident throughout the history of podcasting and justify their inclusion in the framework.

Following on from this historical analysis, a working definition of podcasting will be inferred, representing what podcasts have been and are, which will inform our reflection on the metamorphosis podcasting is undergoing at present: what is changing, why it is necessary for the medium to evolve, and how future innovations may or may not denature the nature of podcasting. This latter step will showcase the use of the six-tensions framework (although more examples will be provided in Chapter 3), and enable an answer to my research questions.

The literature review presented here adheres to Popper's Theory of Falsification (Popper, 1992), where scientific hypotheses are provisional, confirmed over time by empirical validation, or eventually disproved through falsification. This review provides the necessary grounds of justification for hypothesising a system for podcasting innovation, so that it can be tested in the future throughout this thesis and other research projects.

2.2 What Is a Podcast?

2.2.1 A Brief History of Podcasting: From the Radio to the Portable, On-Demand Format

The Radio: A Parent/Sibling Medium

The radio's influence over podcasting will be explored first, as radio is widely acknowledged to be the conceptual predecessor of the podcast (Madsen, 2009; Murray, 2009; Edmond, 2015). The radio was first an experiment to broadcast music and talks to a wide audience, but, although it was intended as a way to distribute audio content to the public without distinction, it quickly grew as a "uniquely personal medium" (Encyclopedia Britannica, 2010). From being able to choose your programme, to the growth of talk-in (or phone-in) shows in the 1990s, radio listeners were motivated not only by access to information and entertainment but also companionship (Perse and Butler, 2005). The "personal" nature of radio only bolstered this last aspect, building a certain intimacy between the listener and their radio set. In the early 2000s, as people started getting used to ultimate musical sovereignty with the rise of the MP3 format and portable media players, the then-apogee of playlist making and track discovery, this expectation of personalisation drew people away from radio. A small but significant shift in practices can be observed in data from the time: in 2007, radio lost 3.1 percentage points of 15-24-year-old UK listeners (OFCOM, 2007). This was the beginning of a trend of radio appealing to fewer young people every year. This showcases a change in media consumption choices and expectations for young people. Some turned to podcasts as an alternative to bring them "information", "di-

version" and "companionship", which were the main motives for listening to the radio at the time (Perse and Butler, 2005), efficiently replacing the traditional roles the radio had, while maintaining the personalisation, freedom, and convenience that they enjoyed from using a portable audio device (like an iPod) to access their media (Albarran et al., 2007).

We often think of the radio as an ancestor of podcasts, or as a product of "radio's cultural renaissance" (Edmond, 2015). The auditory nature of both media, as well as the temporal precedence of radio over podcast, seem to confirm this filial relationship, but I am inclined to characterise them here rather as siblings, because, beyond their intrinsically different vessels of transmission (FM and the internet, respectively), podcasts seem to have emerged from the circumscription of the radio in a changing media landscape, leading to 1) a sort of competition between them, where the audience of one is not necessarily the audience of the other (Albarran et al., 2007) and 2) a cooperation in the introduction of listeners to audio entertainment, where audiences of either can be shared by association (Berry, 2016). This ambivalent relationship was an important concern when podcasting was first developed, as many wondered how its growth would impact the radio's future, and whether the two would eventually merge into one.

What Could Have Been "Audio-Blogging"...

The first podcasts and initial media coverage give valuable insight as to what podcasting was intended to be. What could have been "audio-blogging" (Hammersley, 2004), was first mentioned in 2004 in a Guardian article, after Mark Curry, now dubbed the "Podfather" (Berry, 2006), and Dave Winer came up

with the primitive form of what would become on-demand audio entertainment, using RSS feeds to automate the downloading of sound files from the web (Mclung and Johnson, 2010b). This introduces the initial tension which led to the development of podcasting: **personalisation based in automation**. The listener's input is prevalent, but, overall, machines are responsible for delivering the chosen content and maintaining the listener's attention and auditorship. It took very little time for this new technology to gain popularity: in 2005, it was named "Word of the Year" by the New American Dictionary (Podcast., 2008; Durrani et al., 2015). Although I have mentioned the appeal podcasts could have had over traditional radio for young people, the fast expansion of what started out as a niche for technology and audio aficionados cannot be explained by this alone, as will be explored below.

What Are the Features of Podcasts?

Answering this question will highlight which features have been consistent since podcasting's inception, and therefore isolate the essential characteristics of podcasts to establish a current definition of the phenomenon. The novel features of podcasting largely contributed to its rapid gain in popularity. In addition to the ability to have complete control over the type of programme one could listen to, the possibility to "time-shift" and "place-shift" (Mclung and Johnson, 2010b) – that is, to play the content whenever and wherever, but also being able to fast forward, rewind, or listen to a programme again – made the medium incredibly convenient. Combining this convenience with the idea of free subscriptions, which according to Berry (2006) was a key concept of podcasting, makes for an enticing package, where one's favourite

shows would be easily accessible, with no cost or time limitation. It should be noted that until the introduction of Apple's paying podcast subscriptions in 2021, podcasts were mostly free. Despite offering an interesting monetisation and remuneration option for creators, it is still unclear how this decision will affect the podcasting community in the future (Apple Newsroom, 2021).

2.2.2 The Recent Podcasting Landscape

The Importance of "Seriality"

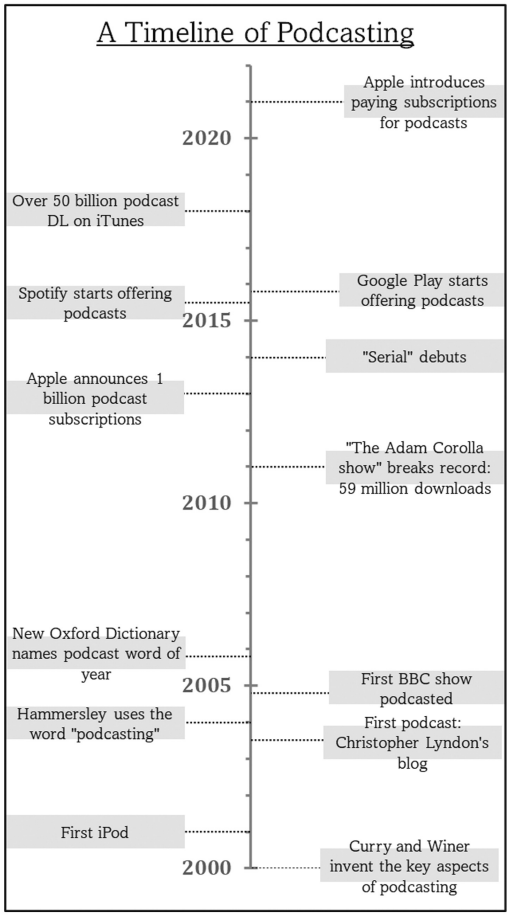


Figure 2.2: Timeline of podcasting, showing the usually recognised key dates in podcasting history.

The episodical predisposition of podcast creators and the influence of this predisposition on the medium shows how podcasting technology affected its content, and vice versa. It demonstrates a form of podcasting Darwinism, where the prevalent and popular features are highlighted and enhanced, making them more important to the medium than other less fashionable aspects. There is somewhat of a convention amongst scholars to highlight several key dates in the history of podcasting (cf. Figure 2.2.): The first use of the RSS feed to distribute audio blogs, the release of the iPod (CNET, 2011), the first podcast (Wallick, 2003), the first use of the word podcasting (Hammersley, 2004), and, flash forward a decade, the release of *Serial*[3], a true crime podcast, which, during its first months, broke all previous downloading records for podcasts. These dates are all seen as turning points, but why is a singular podcast often deemed as important as the invention of the medium itself within podcasting history? To understand the impact of *Serial*, we need to become acquainted with the programme and its context.

Serial was a podcast show spun off of *This American Life*[4], a popular radio show mixing documentary-style stories and audio experiments. Where *This American Life* had no preference of genre, *Serial* only focused on true crime and investigative journalism, where one case was developed over a season of episodes. It was first released in 2014, with new episodes posted each week for 12 weeks, distributed via RSS feeds through the podcast's website. It quickly reached over 5 million iTunes downloads, a record at the time. In his paper, Berry (2015) argues that *Serial* "moved [podcasting] from a niche activity to a mainstream media platform" (p.171), "raised the production

[3] *Serial*, 2014-Present serialpodcast.org
[4] *This American Life*, 2007-Present www.thisamericanlife.org

quality bar" (p.176) and "presented podcasting as a viable alternative for creators and storytellers" (p.176); which are three major achievements to attribute to one program.

Like Berry, many others have commented that Serial ushered in a new era for podcasting (McHugh, 2016; Hancock and McMurtry, 2018; Sharon and John, 2019; Sherrill, 2022); one where "seriality" became a prominent feature of podcasts. Arguably, this episodic structure is inherited from the tendency of TV and radio shows to use cliffhangers to retain listeners from one episode to the next. This led to a "renaissance" (McHugh, 2016) of fictional and non-fictional storytelling formats, inspiring creators to try to replicate *Serial*'s success, producing fiction or non-fiction crime thrillers featuring an investigative team leading interviews and discussions to solve mysteries. Such podcasts are still extremely popular today, with some of the most downloaded shows ever amongst this category (e.g., *Limetown* (2015)[5], *Up and Vanished* (2016)[6], *S-Town* (2017)[7], *Atlanta Monster* (2018)[8], *Faerie* (2020)[9], *Welcome to Your Fantasy* (2021)[10],...). This demonstrates that podcasting features evolved to highlight certain initial aspects of the medium, like the importance of the episodic format. Conversely, although RSS feeds were part of the "building blocks" of podcasting, they are no longer necessary to distribute podcasts, as streaming has gotten more and more popular and erased the need to download content directly.

[5] *Limetown*, 2015-2018 www.limetownstories.com
[6] *Up and Vanished*, 2016-2018 upandvanished.com
[7] *S-Town*, 2017 stownpodcast.org
[8] *Atlanta Monster*, 2018-Present atlantamonster.com
[9] *Faerie*, 2020 www.parcast.com/faerie
[10] *Welcome to Your Fantasy*, 2021-Present gimletmedia.com/shows/welcome-to-your-fantasy

A "Programme-Led" Medium

In order to further chart the changes in the podcast format, the current podcast landscape needs to be compared to that of the 2000s. If the first podcasts were technology talks (The Conversations Network, 2013), weblogs (Wallick, 2003) and radio shows offering their programmes on replay (e.g. NPR in 2005), there is now a plethora of genres and subgenres in the "pod-verse". In 2020, Spotify broke down its most popular genres as Society and Culture, Comedy, Lifestyle and Health, Arts and Entertainment, Education (Spotify Newsroom, 2020b), with less downloaded/streamed genres being Stories, Music, Games, Business and Technology, Sports and Recreation, News and Politics, Comedy, Kids and Family, and True Crime still abounding with content. This variety is part of the reason why podcasts are now so popular: they offer a **unique experience, with a universality of content**, covering the same genres one could expect from other traditional media (radio, television, books/magazines, ...). This second tension is partly why podcasts have been described as "programme-led" (Berry, 2006), where a person is completely in charge of choosing their content, as opposed to "format-led", where a person can only choose when to tune in for scheduled content following strict formats, like the radio. There has been an increase in the diversity of presentation formats also:

- Interviews or Conversations (e.g. *Table Manners with Jessie Ware*, 2018[11]);
- Monologues (e.g. *Have You Heard George's Podcast?*, 2018[12]) ;

[11] *Table Manners with Jessie Ware*, 2018-Present www.tablemannerspodcast.com
[12] *Have You Heard George's Podcast?*, 2018-Present www.georgethepoet.com

- Repurposed Media (e.g. *The Skewer*, 2019[13]);
- Panel Discussions (e.g. *The Infinite Monkey Cage*, 2009[14]);
- Fictional Storytelling (e.g. *Limetown*, 2016);
- Non fictional Storytelling (e.g. *Lore*, 2015[15]) ...

In the beginning, most podcasts fell into the first three categories, but now, styles and genres are barely guidelines, with programmes mixing and matching presentation formats to make unique content. Wyld (2021) highlights the effectiveness of podcasting as a storytelling tool, and showcases how the "traditional" podcasting genres can be bent to create new and engaging audio experiences.

The Independent/Mainstream Podcast Antithesis

Markman (2012) described the "typical podcasters" as "older educated professional males" (p.547). Making an inference from recent studies on the development of the medium (Beniamini, 2020; RAJAR, 2020; Edison Research, 2019), over the last decade, the accessibility of production and popularity of the medium seem to have opened the doors to a more diverse podcasting landscape. Podcast diversity has clearly contributed to the growth of the medium, and vice versa, has greatly benefited from these new audiences and creators bringing their varied backgrounds and experiences to podcasting.

[13] *The Skewer*, 2018-Present www.bbc.co.uk/programmes/m000czyb/episodes/downloads

[14] *The Infinite Monkey Cage*, 2018-Present /www.bbc.co.uk/programmes/b00snr0w/episodes/downloads

[15] *Lore*, 2015-Present www.lorepodcast.com

These new voices should be examined to establish a comprehensive anatomy of podcasting.

"Who doesn't have a podcast?" has become somewhat of a sarcastic *topos* in the recent years. There is a proliferation of tools (free ones like Spotify for podcasters, Spreaker and Podbean, or paying ones like Podomatic, Cast, and many more) making it easy for anyone with a vague understanding of the technology to create a podcast. It was dubbed "amateur" podcasting, but many of these independent productions have nothing "amateur" about them (e.g. *Nightvale Presents*), even when compared to podcasts produced by major companies (e.g. *Tracks*, a BBC Radio 4 programme) (Markman, 2012).

This calls to mind what Berry (2006) termed the "podcast problem" – that a medium originally intended as a means of *independent distribution* for audio media is heavily associated with *one global corporation* (Apple). Today, major providers like BBC Sounds, Spotify, Global Player or Castbox (RAJAR, 2020) are sharing the podcast industry with "attic" producers able to publish their content on personal websites as well as on large platforms like Spotify, Google or Apple podcast. This tension between **independent and mainstream production** generates a heterogeneity in content, which feeds off the uniqueness and universality equilibrium, but also underlines one of the fundamental traits of podcasting: its means of production should be within reach of any aspiring creator.

2.2.3 A Medium Defined by Its Audience

A Global Format

Because podcast listening is not fixed to specific times, they are more accessible to different schedules but also locations. The "global" nature of podcasting refers primarily to a lack of geographical boundaries but also to types of listening locations (e.g., train carriage or kitchen) and listening devices (Berry, 2006). This global nature, only reinforced by the variety of programmes offered, leads to more potential engagement, itself leading to a wider reach in audience. This global nature is also amplified by increased smartphone and tablet ownership, in that these are commonly owned devices affording access to podcasts. According to RAJAR (2019), 79% of podcast listening is done via smartphone. This can be explained by the surge of smartphone ownership in the past decades: over 80% of adults own a smartphone in the UK (OFCOM, 2020), making it simple and accessible to tune in to one's favourite podcasts.

Smartphone use for podcast listening decreased slightly during the COVID-19 crisis of 2020, −9% in the first quarter of 2020 (RAJAR, 2020), but this is likely due to access to other home devices such as tablets and smart speakers that would have been available to listen to podcasts instead. As expected, statistics from 2020 were skewed and reflected the extra time spent indoors. Looking at the data collected by RAJAR (2023), it appears that the listening habits of podcast listeners seem to be changing.

RAJAR (2019) reports that 48% of podcasts are listened to at home, 37% while travelling, and 11% while working. RAJAR (2023) reports that 61% of podcasts are listened to at home, 25% while travelling, and 8% while

working. It appears that sedentary podcast listening has grown in popularity - underlining that the versatility of podcast listening is not only a reason why podcasts are popular but also a factor in its growth.

The Evolution of Podcast Elitism

Back in 2009, most podcast listeners were tech-savvy, college graduates earning over $75,000 per year (Mclung and Johnson, 2010b). This narrow demographic lent the medium an elitist aura (Sharon and John, 2019), very different from the utopian view of podcasts as a "global" medium. To confirm or disprove this impression, large studies gathering personal data like social class, income range, education, ethnic background etc. from podcast consumers should be looked at. However, if these studies exist, the results do not appear to be publicly available. The diversity factors of published studies are rarely the same from one year to another on each report, preventing a temporal synthesis. Furthermore, there are scant overlaps between the variables looked at by different research groups, which makes analysis and discussion difficult. As seen before, there is no shortage of data on podcast consumption and general information on the typical listener, but the lack of more detailed studies prevents us from truly understanding the shifts and evolutions in podcast audience habits, preferences and expectations.

Some useful conclusions can still be drawn from the data at hand. For instance, according to RAJAR (2020), podcasts are still more listened to by men (54% of the consumers), with 64% of monthly podcast listeners being in the 25-54 range. Compared to the 2009 statistics (Mclung and Johnson, 2010b), where 15-24 men listened to podcasts the most, this indicates that the

demographic is slowly ageing, perhaps following these initial listeners through different life stages. In 2019, 53% of new podcast listeners were women, which is a substantial shift compared to the gender distribution in "veteran" listeners, which had a 63:37 men:women ratio. These new female listeners are slowly tipping the scale, and an even split in the men-to-women ratio of podcast listeners in the coming years can be projected (Edison Research, 2019).

Beyond gender, a trend has emerged in the last years that the growth in overall podcast listening has caused a gradual realignment of the distribution characteristics of podcast listeners with those of the population at large. This extends to distinctions of gender, age, ethnic background, education level and income range, and could lead to believe that "podcasting elitism" is slowly receding (Beniamini, 2020)

The Expectations of the Listeners

Using the profile of both current and potential listeners to infer their common expectations allows executives to imagine a future of podcasting that would better cater to both groups, capitalising on the opinions of an already existing fanbase to keep the medium relevant and more engaging. This introduces the third tension, created by any rift in expectations between the **current and potential listeners**, and the importance of marketing for the producers of the podcasts of tomorrow. How can we construct and so understand the profile of these future listeners?

Let us place the expectations of current listeners at the centre of this reflection to build the profile of the "future" podcast listener. If we know

why people listen to podcasts, we can better understand what type of people would be keen to listen.

In Table 2.1, I have grouped thematically the different reasons people listened to podcasts as described by four studies (Mclung and Johnson, 2010b; Beniamini, 2020; Glebatis Perks et al., 2019; Chan Olmsted and Wang, 2020). Through this synthesis, I collapse a long list of terms into a set of categories that encapsulate all the motives previously highlighted in the literature. These terms are categorised as follows: Divertissement (entertainment, inspiration, relaxation and escapism), Social belonging (social activity, and support), Education (news and learning), Companionship, and Medium attributes (convenience and quality). Entertainment is deemed the most important motive in each of the studies mentioned. However, expectations of entertainment vary depending on who is queried: a young person's expectations differ from those of older generations. In the most prevalent age group for podcast listeners (15-35 according to RAJAR spring 2020), there seems to be an expectation of personalisation (e.g. pre-made playlists, recommendations) and social connection (shareability, personal or global relevance, cultural phenomena).

Table 2.1: Why do people listen to podcasts? A thematic synthesis of the motives behind podcast listening according to four studies: Mclung and Johnson, 2010; Edison Research, 2019; Perks et al 2019; Chan Olmsted and Wang, 2020.

Themes	Key Idea	Mclung and Johnson, 2010	Edison Research, 2019	ChaPerks et al 2019	Chan Olmsted and Wang, 2020
Divertissement	Entertainment	Entertainment	To be entertained	Edutainment	Entertainment
	Inspiration	Building library	To feel inspired		
	Escapism		To escape	Storytelling	Escapism/pastime
	Relaxation		To relax		
Social Belonging	Social activity	Social aspect		Engagement	Personal/communal identification
	Support	Advertising			
Education	News	Information	To stay up with latest topics	Edutainment	
	Learning		To learn new things		
Companionship	Company		For companionship		Companionship/connection
Medium Attribute	Convenience	Timeshifting		Multitasking	Audio-platform superiority
	Quality				

As marketing is a key part of podcast production, specifically for the major podcast providers (BBC Sounds, Spotify, Global Player, Apple Music, iHeartRadio, etc.), the expectations of the listeners are always at the forefront of a producer's mind, and content is also created to maximise each podcast's reach and increase its audience. The relevance of each programme is key in its publication, which encourages creators to consider innovation and ideas to maximise engagement.

2.2.4 A Medium Shaped By Its Producers

A creative utopia

In the early days, podcasts were niche ventures from tech enthusiasts or journalists. Once the medium became more popular, and iPods or other portable audio devices were readily available to listeners, the practice extended to independent creators. These independent producers contributed to the medium's expansion, and, as expressed in 2.2.2, to the independent/mainstream podcast antithesis.

Surveying producers in 2014, from independent shows listed on iTunes, Markman and Sawyer enquire about the driving factors for podcasting, through a study with 120 independent podcasters. This dataset is not representative of the industry or this specific demographic as a whole but can illustrate some of the reasons that people podcast for.

When looking at why this group started podcasting, Markman and Sawyer see four distinct themes: *public creativity* (demonstrating expertise), *joining the podcast movement*, self-expression, enhance podcasting skills), *performance/promotion* (performing, seeking attention, promoting people or con-

tent, or wanting to do radio), *long tail* (convenience of the medium, freedom of the medium, niche market, interest in new technology), and *entertainment* (personal enjoyment). This mix of intrinsic and extrinsic rewards is a little reminiscent of Table 2.1, with some reason for listening (e.g. convenience of the medium, enjoyment, advertising) overlapping with reasons for production.

A 2022 Podcast Host study with 537 respondents also supports these broad themes, reporting people podcast - as a hobby – to build a personal brand – to grow a business – for a cause or activism – for an employer (listed in descending order of predominance in answers) (Friel, 2022a).

Both of these studies speak to a more ideal, utopian version of podcasting; including a producer driven by a wide range of factors, and not mainly by monetary gain.

Confronted to the reality of industry

However, motivations for starting podcasts have indubitably shifted with the growth of the industry. In their blogpost "*15 Benefits of Podcasting — Why You Must Start a Podcast*"[16], Riverside FM (a podcasting editor/host) lists reasons for people to venture in podcasting. Monetisation and business opportunities represent over half of all the reasons listed.

In the 2022 Podcast Host survey (Friel, 2022a), over 85% of respondents said they were interested in monetisation; now that podcasts reach so many people, producing them can be a lucrative business.

[16]https://riverside.fm/blog/benefits-of-podcasts

Where are the women?

This initial question extends to other diversity criteria, examining gender balance in production can be seen as an example marker across the field. In Markman and Sawyer's study, respondents had a mean age of 41, and the majority (82.5%) were male. Respondents were also highly educated with 40% having earned a bachelor's degree. If one attempts to answer the question: "why do women podcast?", they will find there is a lack of reported scientific data that might lead to accurate conclusions.

In their commentary, Werner et al. highlight the social ramifications of this lack of diversity, and calls for listeners to amplify women's voices by actively seeking out podcasts created by women. Although that would help visibility, like in many other media, industry leaders have a strong hand in deciding which programmes are successful or not (for example via marketing, distribution, increased reach, larger production budget leading to better quality content etc.), and therefore the load of responsibility could fall to systems and institution rather than on personal listener responsibility, just as other aspects of diversity and inclusion can be addressed in other fields.

2.2.5 Arriving at a Definition of "Podcast"

Before looking at ways to bring innovation to podcasting without denaturing the medium, and looking at Next-Generation Podcasting, we have to look back at the discoveries highlighted in this literature review, in order to answer the question: "What is a podcast?". Through their evolution, podcasts have laid at the intersection of a set of four tensions, which will be essential when subsequently establishing the "six-tensions" framework: Person-

alisation and Automation, Unique and Universal content, Independent and Mainstream production, and Current Audience and Possible Demographics (see also Figure 2.1). How a subtle balance is achieved between each of these pairs, and how this equilibrium has defined the podcast since its inception was paramount to building this framework.

Using these tensions in combination with a more technical or feature-centric definition, as seen in Section 2.2.1, the following overall definition is deduced:

> *A podcast is a piece of episodical, downloadable or streamable, primarily spoken audio content, distributed via the internet, playable anywhere, at any time, produced by anyone who so wishes.*

Going back to the question "is an Audiobook a podcast?", I would argue, using this definition, that it is: virtually anyone can record themselves reading a book and, hopefully with appropriate permissions granted, publish it in an episodic format, which ticks all the figurative boxes the definition offers. But even though this definition was informed by the evolution of podcasting up until now, it overlooks the ability for the medium to change from now on.

So far, podcasts have been shown to be incredibly metamorphic, evolving from something quite simple to a whole world of possibilities, even requiring academics to define its nature in order to pursue research around it. If the boundaries of podcasting appear to lie within the confines of these above-mentioned tensions, how far can we expect them to move, shift or change in the coming years? How will this affect this definition of podcasts? And how can these boundaries be pushed while preserving the nature of podcasting, to which they are integral?

These questions can seem like an over-complication of the seemingly sim-

ple interrogation "What is Next-Generation Podcasting" (RQ 1), but it is necessary to look at the past and present to inform the future of a medium, not only to be able to be able to forecast adequately what the media landscape might look like, but also to inform and lead creation and innovation within the field.

2.3 The Six Tensions Framework for Podcasting Innovation

2.3.1 A Glimpse at How Podcasting Is Already Pushing Its Boundaries

Transcending the Limitations of File Formats

Podcasting is already undergoing a slow but steady metamorphosis, with file format being a telling example. When podcasts were first produced, .mp3 offered acceptable quality, a file size reduction of approximately a factor of twelve, and space for metadata, which made it an automatic favourite to export and share podcasts. The .mp3 format was chosen out of convenience, but as requirements evolved and creators began thinking of more creative uses of podcasts, its limitations became apparent.

The first signs of change were the adopting of other audio file formats like M4A or .MP4, which use the AAC codec (BBC Sounds, 2021), or .ogg (Spotify for Artists, 2021), which allow for tighter compression maintaining a similar bitrate as MP3, translating into higher quality and still relatively small file sizes. Even though this opens the door to better audio quality, the

podcast delivered is still a fixed product: an immutable audio file over which the listener has very minimal control.

The metadata carried alongside the audio files have introduced a little more flexibility to the listener. For instance, by adding information on chapters, the user can skip from one chapter to another at the touch of a button, provided they are using a compatible player. Yet, the technology behind these metadata formats has not evolved as drastically in the last 20 years, and currently caps the potential for personalisation inherent to the MP3/M4A/OGG formats in use for podcasting. The addition of transcripts, illustrative accompaniments, or new navigation methods rely on the podcast provider's decisions, which restricts the potential for customisation of a programme by its producers.

Some podcast creators have therefore decided to publish their content independently of a traditional podcasting host like Podbean, Buzzsprout, or Spotify for podcasters, and rather chose to create standalone web pages (e.g. *The Garden*, 2020[17]) or apps (e.g *This American Life*), to gain complete freedom over the components in their programmes. However, this process only redefines the format on a per-podcast basis, with each programme or company creating a format that matches their need, without using a more universal podcast format. This lack of consensus on the format to use for more personalisable or responsive podcasts leaves a considerable gap in the industry, which stunts the growth of many innovative projects looking to the podcast format to host their new forms of content (Blind, 2013), but also highlights the need for innovation in this area.

[17] *The Garden*, 2020 www.bbc.co.uk/taster/pilots/the-garden

Immersion in Podcasts

The podcast metamorphosis has also changed features expressed more subtly until now, but that have grown more notable and popular over the past few years. The evolution of these "hidden" features is now taking a central place in leading podcasting innovation. The audio properties of podcasts are amongst these features. For a while, podcast innovation was bound to the content, rather than exploiting the creative opportunities offered by the auditory nature of the medium. But, as seen in Table 5.1., the success of podcasting is not based on the content's entertainment values or convenience alone, but also on its audio properties, which make it possible to immerse oneself in an acoustic environment.

Witmer and Singer (1998) defined immersion as "*a psychological state characterized by perceiving oneself to be enveloped by, included in, and interacting with an environment that provides a continuous stream of stimuli and experiences*"(p.227). Zhang et al. (2017) distinguishes two main types of immersion: a) *Embodied*, which for audio encompasses both quality (e.g., better headphones, file formats, sound systems) and spatialisation (e.g., stereo panning, ambisonics, binaural audio, 3D realism or illusion using virtual, augmented or extended reality) and b) *Empathetic*, where the substance or subject of the content is relatable, interactive or generally captivating.

From the above definition, immersivity is not only a function of form but also of content. Beyond making relatable and interesting programmes, another option creators are starting to consider is interactivity, shifting traditional podcasting to a more personal medium, with examples of interac-

tive content (*Responsive Radio*, 2015[18], *The Mermaid's Tears*, 2017[19], *Solve*, 2019[20], ...) or personalised advertisement (Radio Works, 2018).

Interactivity as a Tool

Although a definition of personalisation was provided in the preamble (Chapter 1), when applied to podcasting and audio more globally, personalisation can take several forms and occur on multiple levels: personalisation of the listener experience (e.g. on-demand audio, playlist curation, automatic recommendation etc.), personalisation of interface (e.g. changing appearance of podcast app based on preferences, in-car interface, automatic downloads, etc.), and personalisation of content (e.g. news with different headlines depending on listener's location, choose your own adventure content, etc.). These can be combined or applied separately. In this thesis, the focus is brought to personalisation of interfaces and content - as personalisation of the listener experience depends entirely on systems and platforms already prevalent in the industry. Recommendation systems are a broad topic of research on their own, and the scope of the thesis could not investigate this level of personalisation conjointly.

Personalisation has always been part of the appeal for podcast users, where their listening habits would reflect choices and preferences, as opposed to radio where channels would dictate content to its users. So do personalisation and immersion combine in this respect, and if so how? Interactivity is already a feature of modern podcasts: the user has to make a series of choices before accessing their content, which differentiates it from the ra-

[18] https://www.bbc.co.uk/taster/pilots/responsive-radio
[19] *The Mermaid's Tears*, 2017 mermaidstears.ch.bbc.co.uk
[20] *Solve*, 2019-Present solvehq.com/podcast

dio. Recommendation systems attempt to simplify this process, expressing the "**personalisation and automation**" tension in another way, through automatic personalisation. The extent to which recommendation systems "succeed" in any sense has been challenged recently (Born et al., 2021; Born, 2020), but interactivity is important beyond the initial decision of what to listen to, and introduces another pair of concepts in tension with one another: **immersion and interactivity**.

If immersion is a goal, interactivity can either be seen as a way to achieve it or as a hindrance (Ryan, 1999). There is a subtle balance to achieve, to not have the listener interact so much that they will lose their sense of immersion within the content, but to still engage enough with the audience that the programme offers a "*continuous stream of stimuli and experiences*" (Witmer and Singer, 1998, p.227). These concepts are not opposed, but rather in competition, as both can be described as integral to the other.

Lately, interaction with the podcasts' content has been prioritised in productions, as it is seen to boost engagement, driving ratings and popularity. For instance, Spotify introduced a new "poll" feature in 2020 (Spotify Newsroom, 2020a), which allows presenters to survey their audience at specific moments of their program. This was preceded by a myriad of amateur "choose your own adventure" podcasts, that offered nonlinear narratives to their listeners, and succeeded by the creation of new apps, like Stereo[21], which is the "talk-in radio" equivalent of the podcasting world, where users can send in voice snippets to podcasters during their live shows, or Clubhouse[22], En-

[21]Stereo, `stereo.com`
[22]https://www.clubhouse.com

tale[23], Adori[24], Hypercatcher[25], or the Spotify tool Spotlight, which enable the user to experience additional data, like visuals, links, descriptions, without leaving their media player. More audio-based interaction has also been investigated, and BBC Taster has put out a range of audio experiences making use of various forms of interactivity to alter the audio of podcasts (e.g. *Pick a Part*, 2020, *Monster*, 2020).

If this trend continues, it is natural to ask how much interactivity crosses the line between "passive" and "active" entertainment, effectively turning a podcast into a game. Interactivity of the right quality and quantity has the potential to make programmes more personalised to the listener. It reinforces the **universal yet unique** aspect of podcasts and could be used as a tool to maximise immersion rather than to gamify the listening experience.

2.3.2 Innovation for Podcasting

The Fundamental Transformation of Art and Technology

In 2006, podcasts were seen by many as a "revolution" (Berry, 2006); a medium grounded in innovation from the beginning, using computers to help deliver media and more broadly, art. In order to look at how we can innovate while respecting the medium's essence, we have to look at why podcasts need innovation to exist in the first place, and why the metamorphosis mentioned above has been occurring and is already pushing boundaries.

Creation (*poiesis*) and technology (*technè*) are often seen as opposite or heterogeneous, but Coeckelbergh (2018) (and countless others) argues rather

[23]Entale, www.entale.co
[24]Adori, www.adorilabs.com
[25]Hypercatcher, hypercatcher.com

that they are very much intertwined. Art's essence is connected to human creativity, which changes constantly. Technology is also driven by perpetual reinvention. Podcasting is an example of this interconnection between **Art and Technology**, where innovation lies at the centre of a new tension.

Podcasts, like art or technology, are intrinsically linked to innovation. As with any other artistic medium or technological endeavour, it is important to embrace innovative drift so the medium can flourish. Podcasting is not apart from "more traditional" media in this respect. It is driven by the same need for reinvention as film, TV, and radio, and therefore should not be expected to remain the same forever, in the same way that these other media are given space to grow and change while still maintaining their appeal and audience.

The Past, Present, and Future of Podcasting

If innovation is intrinsic to podcasts, how can the definition of "podcasts" given earlier be valid, when the medium is expected to change? Thomasson (2010) believes the ontology of art is determined by "human intentions and practices" and that the boundaries of a work of art are defined by the "beliefs and practices of those who ground and reground the references of these general terms"(p. 128). Looking at podcasting as media, and by extension, as a form of art, Thomasson's postulate infers that this analysis and exploration of what podcasts are is at least momentarily valid because it is grounded in the human experience of podcasts and based on factual evidence. We know what podcasts were, we witness what podcasts are, and our imagination of what they will be only influences their future. The definition given here is crystallising as it is being written.

As an example of this cultural impact over podcasts' definition and nature, let us consider France's history with the word "podcast". In 2010, "podcasteurs" referred to comedians who talked directly to their camera and posted short comedy sketch videos on YouTube (Beuscart and Mellet, 2015). Everyone in France called their content "podcasts", even though they had virtually nothing in common with what podcasts had been in the US or the UK so far. "Podcast" took on a new meaning in France, associated with humour and a specific type of mostly visual entertainment. This meaning was replaced only when "actual podcasts" grew in popularity in France, and these YouTubers' notoriety eventually decreased. The notion of what podcasts were is completely different to what podcasts are now, and yet, podcasts were always called the same thing. We choose what a podcast is, and what a podcast will be. Our preferences, our colloquialisms, our culture, give the word "podcasting" its meaning.

For the time being, this definition is accurate. However, the hope is that further research and innovative endeavours, perhaps informed by the remarks made here, will transform what podcasts are, so that necessarily this definition will have to be revised in the future; that would mean the medium is evolving, which would be a positive outcome if the goal is for podcasts to maintain, or grow in, popularity.

How to Go About Innovating for a Chameleon Medium?

Over the years, the podcasting landscape has changed drastically: from a few shows focused on technology in 2004 (Hammersley, 2004), to the current pod-verse, with over 48 million podcast episodes (Podcast Insights, 2021)

of all types and genres, and from iPod and MP3 players to smartphones and smart speakers, there is no shortage of changes that have made podcasts more enjoyable and accessible in the last decade. Somehow, through all these modifications, podcasts could all fall under the same umbrella of audio-based, downloadable or streamable content, which is what motivated the definition of podcasts presented here.

The versatility of podcasts is a double-edged sword for producers: it can help create a wide variety of content, but also quickly transform their podcasts into another type of media or entertainment entirely. The potential "gamification" of podcasts demonstrated this, but this can also be applied to podcasts that rely on so much visual information they lose their audio focus and become predominantly visual. Still, these adaptive properties should not be tossed aside. Indeed, it is innovative drift that permitted podcasts in the first place. When Curry and Winer put together the system which would allow for podcasting to develop, they could not tell that it would draw characteristics from the world of TV (dramatisation convention), radio (talk-in, panel shows), literature (audio books, transcripts) and many more.

This only corroborates the fact that the definition of "podcast" is not fixed; on the contrary, I hope it will change. In other words, this definition is too contemporary to be considered the "nature" of podcasts. So, is there an alternative way to define the essence of podcasting which would encompass possible evolutions of the format, encouraging and not restraining innovation, without losing sight of the most important features of podcasting?

2.4 Next-Generation Podcasting

2.4.1 Grounding the Endless Possibilities of a New Medium

Finding the Balance to Retain Podcasting Integrity

Limitations can help innovation in research and development environments (Rosso, 2014). It is crucial to set boundaries to restrict research, particularly in a field where the projects can take such a wide scope, to narrow down possible goals and help frame creative endeavours. The boundaries I have chosen to respect are, from previous reflections, the ones which allow for the most creative freedom, while still highlighting what I consider has been and will be essential to podcasting. I propose that the nature of podcasting is in fact its boundaries, and that its boundaries are this set of tensions, opposing forces striving for balance, representing a summation of equilibria that have characterised podcasting since its inception. The tensions are as given in Figure 2.1.

Developing podcasting with these at the forefront of a reflection will enable to conserve podcasting's essence through its evolution, without constraining the medium to a strict checklist, or attempting to match the definition I have provided. Instead, they will act as guidelines, concepts to acknowledge while trying to bring new ideas to the world of podcasting. These boundaries should overrule the definition of podcasts given when thinking of the "nature" of podcasting. If nature is the immovable essence of an idea, these boundaries are more adapted to take on this role, rather than the more restrictive definition I have set out. How might these guidelines be used to

help channel innovative ideas in the field of podcasting? Simply, by verifying that these ideas do not break the equilibrium in place. If there is a conceptual pair in tension (see Figure 2.1.), there should be an approximately equal and opposite move in the other.

What Comes Next?

To contextualize this framework, let us look at some examples of innovation which are already on their way to modify the way we make and listen to podcasts. Any method or interface which reduces the hindrance of interactivity to immersion, for instance interactivity through personalisation rather than intrusive query for choices (e.g. *Instagrammification* [26] or other similar programmes which personalise the format, story or soundscape) would fall within the confines of the Six-Tensions Framework. So would any other innovation facilitating interactivity without interrupting immersion, might it be a reimagination of the user/podcast interface using AI to recognise user behaviours (voice, sounds, motion) or the content itself, offering interactive narratives or variable podcasts. In both cases, interactivity and immersion are both "pushed", preserving the balance established between them.

Frank et al. (2015); Francombe et al. (2017); Pardoe et al. (2020); Shirley et al. (2019) and many others have been researching ways to make better use of the new audio systems and their variety, adapting the listening experience to a user's device, or improving quality overall. In this case, personalisation and automation are pushing and pulling one another to make audio experiences more responsive.

Podcasts could also adapt to each user's listening preferences, increasing

[26] *Instagrammification*, 2020, bbc.co.uk/taster/pilots/instagramification

accessibility by for instance allowing for the volume of elements less important to a narrative to be turned down to maximise comprehension (Shirley et al., 2017), which would concurrently increase uniqueness and universality for the podcasts involved.

New technologies are proving a great source of inspiration to create next-generation podcasts, e.g., object-based audio, voice or sound synthesis using AI, augmented reality, voice or text recognition, the use of metadata for adaptive audio (Churnside, 2015b,a; Ward, 2020). These can all be considered with the six-tensions framework in mind, so as not to amalgamate podcasts with another, already-established medium

So how will the definition of podcasts provided change in the coming years? How will these innovations shape the podcast format? What will be "Next-Generation Podcasting"? If the boundaries set are right, the issue of podcasting becoming something entirely different could be avoided, as hypothesised by Berry (2016). The tension system set out will give innovation the leeway to contribute to podcasting metamorphosis while preserving the fundamental aspects of the medium, no matter what a podcast ends up being.

2.4.2 Introducing Next-Generation podcasting (NGP)

A Useful, Although Catch-All, Term

NGP is primarily audio and broadcast researcher lingo – although less catchy in the cultural and academic zeitgeist than some of its "next-generation" siblings (e.g. Next-Generation Audio[27]), and just as broad as "Object-based

[27]https://tech.ebu.ch/nga

media" [28]- NGP is used to refer to innovative podcasts and the way they can be made. As seen in the previous Section 2.3, the metamorphic nature of podcasting makes innovation unavoidable, and with all these innovations come inevitable crossovers with other forms of media. If a podcast is accompanied by a video, is it still a podcast? If a podcast is an interactive experience, with clear stakes and rewards, is it now a game instead (Rowe, 1992)? And if a podcast is only broadcast at certain times, does it become radio? These questions highlight the shape-shifting nature of podcasting, and although conserving its essence should be at the forefront of reflection when innovating for the medium, the endless possibilities it offers for creators and audiences alike should also be emphasised. The term NGP reflects this openness. Nevertheless, we've seen that a definition for "podcast" is perpetually changing, so why even have a term to determine what comes next, as that will be assimilated in a new definition of podcasting?

Mostly, for our present purposes, NGP stands for all the innovative podcasting techniques and ideas that are yet to take off. This includes the beta software, the R&D shows and episodes, the vague ideas, and the lengthy brainstorming sessions. For such work, it is helpful to differentiate the "now" from the "future" – which is why this thesis will make use of the definition given in this chapter for "podcast" (c.f. 2.2.5), and will attempt to specify the term NGP, while still taking into account the development framework presented above.

[28]https://www.media.mit.edu/groups/object-based-media/overview/

New Technologies for NGP

If NGP is a category, how do we crystallise what NGP means to the podcasting industry? The exact components of NGP aren't set. One of the first steps of this thesis will be to look into and evaluate the predominance of new technologies within the concept. This will include (but not be limited to) audio and speech generation, non-linear narratives, responsive mixing, and interface changes. Different implementations of these technologies will be evaluated within the Six Tensions Framework presented in this chapter. In Chapter and 2, I ask podcast producers about their impressions of NGP, in order to be able to answer the first research question.

2.5 Summary

How the boundaries of podcasting are defined and redefined by innovations, past and future, is of wide and deep interest to the podcast industry, creators and listeners. In 2006, Matthews (2006) was already interested in the future of podcasting, and he theorised that the two areas where podcasts would have the biggest impact would be education and business. His predictions on the "capitalisation" of podcasting proved right – still today, advertising has an important place in the podcasting industry.

Although universities and schools have used podcasts more and more in the past decade as a modern way to teach and interact with their students, the application of this theoretical enthusiasm for engaging with students through audio has proved somewhat underwhelming, with doubts being raised around the effects of using podcasts as a pedagogical tool on physical attendance and

engagement with the teaching material (Drew, 2017). However, beyond the classroom, podcasts have indeed become a popular communication tool for academic research (Turner et al., 2020; Fox et al., 2021a) and more generally to communicate knowledge (MacKenzie, 2019), which was not foreseen by Matthews. The shortcomings of podcasting he identifies mainly revolve around the lack of transcripts, which were unavailable or too expensive at the time. Today, this problem has been solved by widely accessible, if error-prone, AI-driven transcription methods.

Looking back at Matthews (2006) hypothesis on the future of podcasting reminds us that context should be weighed carefully when making such projections. In this case, the slight push-back from lecturers and teachers to move more of their materials to the podcast format, as well as the growth in interest in *learning* podcasts which exist outside of traditional educational structures, and the overall affordability of AI transcription, all had a big impact on the evolution of podcasting.

Berry (2016) questioned whether this evolution would end up causing the term "podcast" to be replaced by another neologism. The imperfections in Matthews' prognoses, combined with Berry 's interrogations, support the reasoning for building a *framework* for podcasting innovation, as opposed to trying to be more specific: the future of a medium cannot be predicted; it can merely be anticipated. This justifies the need for a "future" specific term for the sake of communicating this research effectively: "Next-Generation Podcasting" encompasses all the new and promising technologies and ideas that have not yet been assimilated into the concept of "podcasting".

The six tensions framework anticipates the future needs of NGP, while

still allowing for the technological and sociological context to influence the evolution of the medium. It is likely that AI and other technological developments will provide other unexpected solutions and creative affordances for podcasting, and that, as a society, our expectation of podcasts will shift in ways we cannot yet imagine.

I have described how podcasting has changed dramatically since its inception, and the future of the medium and how to innovate for it should be considered as a key aspect for the medium's overall advancement, as more changes are undoubtedly already in the making. This chapter looked back at what podcasting was and what it is now, to bring together a set of six tensions (Figure 2.1) that have been consistent since the medium's inception, and reflects on the implementation of this set of tensions as a frame of reflection to bolster innovation in the field.

An analysis of the origins of podcasting and what it became brings to focus many changes that have occurred already (genre, format, mode of consumption, listener's expectations, and audience) but also highlights the areas of research which would benefit from further attention or transparency in data from major podcast providers, to give an unbiased picture of what podcasting looks like today. This reflection also allows me to propose a definition of what podcasts are currently: a piece of episodic, downloadable or streamable, primarily spoken audio content, distributed via the internet, playable anywhere, at any time, produced by anyone who so wishes.

This chapter establishes 1) that innovation is fundamental to podcasting and 2) courses of action to podcast innovation that do not lose sight of the nature of the medium. Considering the podcasts that have been made to

date, I attempt to pinpoint what the nature of podcasting is by identifying and tracing the features that have been present consistently throughout the medium's evolution. This process reveals a set of tensions (cf. Figure 2.1), which I name the six-tension framework, and which I hypothesise could be used as a framework for innovation that would allow the preservation of the nature of podcasting through the unavoidable changes already in motion.

The definition of "podcast" will be subject to change, but provided it does so while respecting the set of tensions revealed in this Chapter, the essence of podcasting should be preserved throughout its changes. And if "Next-Generation Podcasting" represents the unattainable, work-in-progress, future of the medium, the boundaries set in the Six Tensions Framework will help focus innovation and ground new research and development of ideas and concepts for podcasting.

The perspective expressed in this chapter is the one of a technologist, interested in how technical developments influence the definition of a medium or its audience. There is a parallel reflection stemming from a creative-editorial point of view, which focuses on content and tone rather than technology, which would be interesting to explore in order to challenge or confirm the thoughts presented here.

The data used to determine what a podcast is and was is limited. There has been relatively little research on the evolution of the medium, particularly when looking at the evolution of the typical podcast listeners and their habits.

Even though more broadcasting companies and research groups are now looking into podcast listeners' profiles and consumption habits, there are still gaps surrounding key analytical factors, like income, education, ethnicity and

social class, and the influence of genre over its demographic. Having access to this data would be constructive for both research and business, helping build a more accurate listener profile, informing current trends and, consequently, what the trends of the future might be.

This chapter involves a very Western-centric view of the evolution of podcasts. It focuses on the English-speaking podcasting industry and community, because there is little worldwide data or country-by-country reports on podcasting listenership available. The set of tensions presented is a framework to boost and constrain innovation in the field of podcasting. The presumption of innovation for the future of podcasting is necessary for this research. Although some mediums can reach an equilibrium where innovation is not necessary to ensure prosperity, development of new ideas, tools, and projects orbiting these mediums can lead to significant technological and artistic developments which can be argued to be a goal in itself. The framework is theoretical in nature, but its use is justified by being based on a review of literature and survey data. The validity of the Six Tensions Framework can only be confirmed over time by looking at new podcasting projects and tools and their impact on these pairs of tensions.

The framework proposed will be useful to researchers in academia and industry, for producers, podcasting platforms, listeners and all other stakeholders in the field of podcasting. It will ground new ways to consume, interact with or make podcasts in relation to existing material, and bridge the gap between research and new media.

The definition of "podcast" provided will act as a basis for further reflection, as well as a "time-stamp" of what podcasts are today for future

researchers to look back on when thinking about the evolution of podcasting.

Knowing which technologies will change the face of podcasting is nearly impossible, but some of the current trends can give us hints of what podcasting could become. The use of AI, outside of its use for transcription, can be applied to the audio production process to create podcasts which would push the boundaries of podcasting while preserving the equilibrium presented in this chapter. This could translate into more adaptive content, that would follow the users' preferences, different forms of responsive audio, which would allow for more interaction between the user and the podcasts, sound generation, which would create tailored content for the listeners, or even new interfaces between the listener and the podcasts. All of these possible modifications included within the idea of NGP can be investigated under this framework, and their outcomes will certainly alter our current answer to the question: "What is a podcast?".

3

Literature Review: Innovative Production Tools For New Media

3.1 Introduction

There are multiple new technologies currently in the process of "revolutionising" production methods and listener experience – to list only a few: new transcription solutions using AI to generate subtitles for episodes (Matthews, 2006; Trivedi et al., 2018), semantic audio editing (Baume et al., 2018), spatial audio capabilities in listening devices (like the Apple AirPods Pro) and programs (Hyperradio Radio France, 2021; BBC, 2020), the development of tools allowing for new types of spatialised audio experiences, such as Audio Orchestrator (a BBC Makerbox tool responsible for immersive podcasts like *Monster* (BBC Taster, 2020) and *Spectrum Sounds* (BBC Taster, 2022)), and the growing interest in object-based media and its potential applications to podcasts, through adaptive podcasting (Dwornik, 2021) or non-linear programs (The Orpheus Project, 2017).

These projects are all akin to forms of personalisation, where personalisation serves the overall goal of immersion (Kalpokas, 2021). However, relatively little is known about podcast creators' perspectives on these technologies, how they are integrating them into their workflows, and what they

consider the technologies' impact to be on listeners. That is because the data available focusing on podcast creators are often limited to demographic information rather than their opinions on their field.

For other traditional media, like film and radio, perspectives of industry professionals have been thoroughly documented, in dedicated academic publications (Burgess, 2013; Broth, 2008; Mamer, 2013; Vonderhaar, 1983). Comparatively little is known about the corresponding aspect of podcasting.

As listeners, our enthralment with on-demand audio content can be linked to several facets of the medium, as shown in Table 2.1. Its versatility of genres and styles widens with every passing year. Shows explore novel formats and transcend expectations, reaching new audiences (McHugh, 2016). The episodic nature of podcasts fosters the loyalty of listeners (Petitjean, 2008), and boosts engagement, by encouraging them to experience the episodes by, for instance, organising listening parties or releasing complementary content (Sharon and John, 2019). Therefore, motives for listening to podcasts are varied, but the ideas of divertissement and social belonging appear in several studies looking at reasons behind podcast consumption (see Chapter 2).

To media producers, podcasting's appeal is threefold:

1. Podcasts reach over 41% of people over 12 years old in the US every month (Beniamini, 2022), a percentage that has grown yearly. This wide and increasing audience constitutes an incentive for both larger broadcasting companies (like the BBC, NPR, Megaphone, iHeartRadio) and independent creators to invest resources and time into the production of podcasts.

2. These investments have a good chance of turning a profit. Indeed, the podcasting industry was worth $11.46 billion in 2020 (Grand View Research, 2021), thanks to advertisements, sponsored content, and more recently, paying subscriptions (Apple Newsroom, 2021).

3. The creative freedom the medium offers allows for projects to find unique spaces in which to develop.

To answer RQ 2: *"How can we feasibly create and distribute immersive and personalised podcasts?"*, we need to better understand the current practices, behavior, and perspectives of podcast creators, as well as the various existing methods to develop production tools for new media. Intersecting these two facets of innovative audiovisual technology research will not only paint a picture of the primary user of these new podcasting tools, but also enable us to explore solutions for NGP tool development. This should provide designers and researchers the grounds to justify future design decisions on the basis of empirical data.

After giving an overview of production habits in other media, and subsequently in podcasts, I will describe how AI tools and new technologies can be used within the concept of NGP. This includes an overview of "personalised media", followed by a non-exhaustive list of technologies that could be used in personalised podcasting and are particularly relevant to this project. Finally, a review of prior work on formats for personalised media provides the context required to think of practical applications of the topics discussed throughout the chapter.

3.2 Production Habits and Workflows in Other Media

3.2.1 Production Workflows

According to Aalas and Jablonski (2000), "*workflow process definitions (workflow schemas) are defined to specify which tasks need to be executed and in what order*" (p. 267). Mathematically, a workflow abstraction is a directed graph, where each node represents a stage or a process, and the vertices linking these nodes represent a path to take from one stage to the next (Bondy and Murty, 2002). Traditionally, workflows are represented as diagrams, showing a series of events, sometimes grouped together, linked by arrows representing the common "path" to get from an idea to a finished product. Baume (2018) investigates radio production workflows in different settings through operational sequence diagrams. Comparably, Murdoch (2016) describes four phases for a progressive animation pipeline through a simple diagram, and Meixner et al. (2017) speaks of common workflows in TV production in prose rather than by using visual aids.

A production workflow is an invaluable tool when trying to integrate new tools into established media practices (Ward et al., 2020), as was the case for digital music production (Ramshaw, 2006) and 4K digital production for movies (Ion and Humphrey, 2004). This is one of the many benefits of having an established production workflow for a medium. In gaming for instance, McAllister and White (2010) describes how to use typical production phases to evaluate user experience.

In parallel creative domains, production workflows have been used for

decades to reveal the inner workings of creative processes. In music production, countless educational books and articles cover the typical pipeline of the creative and production process, such as Hepworth-Sawyer and Golding (2011), who see two main phases – one of preparation and one of action, subdivided into shorter, manageable events. Music production workflows have evolved over time with, for example, the normalisation of digital mixing over analogue practices (De Man et al., 2017) and will most likely change again with the inclusion of new tools such as AI assistants or digital production software.

This evolution not only demonstrates that the standardisation of some technical practices influences workflows, but also that having an identifiable production workflow can act as a springboard for innovative practices. For podcasting, a medium intrinsically linked to technology, it seems unavoidable that workflows will change with the introduction of new podcasting tools. If we are able to identify an archetypal podcast production workflow, not only will it act as a snapshot of the nature of podcast production in the early 2020s, but also as a potential basis for podcasting innovation.

3.2.2 Producing Innovative Media

The available detail within other media production workflows allows for complex innovative tools and processes to be easily integrated into existing habits, such as new storyboarding tools (Bartindale et al., 2012; Ford, 2016), interactive TV narrative software (Ursu et al., 2008, 2020b), 3D cinema production tools (Bailer et al., 2020), virtual videography or intelligent video

direction (Heck et al., 2007; Pan et al., 2021), and semantic audio editing. [1]

Baume (2018) lays the theoretical foundation necessary for using semantic audio tools in radio production, by detailing traditional workflow patterns in different areas of radio production. This thorough investigation enables researchers to understand the motives and habits of producers, and ensures that any new creative tool would fit the expectations of professionals in the field. This is in line with the findings of Ward et al. (2020) around the key principles for building new media tools: "*Designing tools requires an understanding of what the desired functionality is, what workflows the tool will be integrated into and what value production staff see in the tool*"(p.5).

3.2.3 Detailing the Specificities of Audiovisual Tool Design

Becker et al. (2017) puts together a spectrum of actors that influence a product, from the passive listener (audience) to the content creator (producer). At the extrema are listener-centric and creator-centric approaches to designing innovative audiovisual tools. By "listener-centric", I mean audience-focused reflections based around the listener's needs, meanwhile "creator-centric" pertains to features relevant to the production and delivery of podcasts.

In the former, listeners could be queried on what direction they would like the medium to take, and their answers would be influenced by their individual preferences. Researchers drawing conclusions from any such study would therefore need a large number of participants to paint an accurate picture of the expectations of the audience, and to accept that non-professionals'

[1] https://www.descript.com/

answers would not be rooted in knowledge of the technical ramifications and possible burdens of their expectations.

In the latter, creators could steer innovation towards products they would be interested in using and that inform their design from practices already in place. Combining these approaches would require a better understanding of creators' perspectives, to match the breadth of information available on listeners.

3.3 Podcast Production

3.3.1 The Downside of Podcasting as a Cultural Phenomenon

The focus of podcasting as a cultural phenomenon (Fox et al., 2021b; Sherrill, 2022; Durrani et al., 2015; Hancock and McMurtry, 2018; Markman, 2012) seems to have obscured an equally important aspect of the medium: its production. As evidenced by Chapter 2, the emergence of podcasting has inspired researchers to conduct very thorough user-based studies in the past. Academics have painted a detailed, nuanced picture of the technical, social, and cultural landscape that has led to the emergence of podcasts. This focus on the listener and the social impact of podcasting, although highly relevant and by itself necessary, overshadows another key, seemingly basic, sides of podcasting research. Podcast production, whereas that represents techniques, tasks, or actors, has not undergone the same level of inquiry as the rest of the field. In the following paragraphs, we will collate the available information, and highlight the various missing pieces of the podcast

production puzzle.

3.3.2 The Un-Formalised Workflows for Podcasting

This a good place to remind the reader that this thesis focuses more specifically on"native podcasts" rather than on-demand radio (Novăceanu, 2020). Podcast production processes have been widely documented on various websites, blogs and how-to guides intended for budding podcasters (Podcast Insights, 2018; The Podcast Host, 2021; Riverside.fm, 2022; Geoghegan and Klass, 2005). There has been some academic scrutiny into podcast production, specifically focusing on advising professors and educators on how to produce podcasts to accompany their teaching (O'Donoghue et al., 2008), but if O'Donoghue et al. (2008) and Strickland et al. (2021) address how to produce podcasts for educational purposes, the creators' production methods, workflows, habits, and preferences have not been detailed beyond the many "DIY guides" published so far (Buzzsprout.com, 2022).

Guidelines for podcast production therefore exist for particular genres, such as education (NPR, 2022b) or news (Lindgren, 2021; Frary, 2017), but there is scant evidence of commonalities in the production workflows, habits and goals across genres and production networks. These are gaps that are paramount to fill in order for researchers to develop the tools for NGP.

Cohen (2021) talks about a podcast production workflow more specifically, but the material this article is based on is unavailable. Cohen (2021) identifies 4 stages in podcasting: conception and development, raw content curation, post production, distribution. However, this is more akin to a guide or tutorial rather than the discussion of the results of a scientific study. In-

deed, there is a gap between existing knowledge of podcast production and the reality as it is being practised today, as a function of individual and organisational differences, target genres, and budgetary allowances.

3.3.3 Pod-actors

In the introduction, we defined the term "podcaster", as a blanket term for a person involved in any part of the podcast making process. This comprehensive word is invaluable for the industry, as a lot of actors in the podcasting industry wear several hats at a time: a host is sometimes also a producer and an editor; an editor can be a guest; an executive producer can be a host... But, understanding the different roles that one can have in the podcast production process is a helpful glossary for a podcast enthusiast and researcher. Let us examine these actors and how their work can be described -

Host: Perhaps the most synonymous with "podcaster" for audiences, it is almost impossible to disassociate making a podcast from presenting it. Although not all types or genres of podcast require a host (e.g. fiction and repurposed media), many are centred around this monolithic figure. For documentaries, a host is an entry point into the topic at hand, for comedy, a host is a relatable and likeable master of ceremony, for business, they are a knowledgeable, charismatic leader to look up to... Whatever performative role the host acts out, it is with them that the listeners create parasocial bonds, and oftentimes, for them, that audiences carry on listening week after week (Schlütz and Hedder, 2022). Their importance in the podcasting industry should not be neglected, as it's been reported that a podcast host can influence their listener with ease (Brinson and Lemon, 2023).

There is also now some form of social capital associated to hosting a podcast. NBC reported on the rise of "fake podcasts" (influencers pretending to host podcasts on video to appear more knowledgeable or trustworthy on a topic) on social media platform like TikTok [2]. This trend can be explained by the perceived knowledge and importance of hosts: a host is an expert, a friend, someone you can trust.

Editor: There is little audio intended for mass consumption online that isn't edited before it's published. Comping [3] speech in podcasts is almost as commonplace as comping vocals in pop music is. For non-fiction, an editor's job might be to turn a two-hour conversation into a 25 minutes highlight, to choose snippets of archival material, or to delete most of the "erms" and "ums" in a guest's recording. For fiction and repurposed media, the editor's role grows and adapts, taking on the role of editing not only speech, but sound effects, music, and more.

Sound designer: Said sound effects and music are sometimes procured by a dedicated actor: the sound designer. Their job might be particularly relevant to larger productions, requiring dedicated soundscapes.

Sound engineer: Sometimes encompassing the job of the editor and sound designer, it mostly refers to recording and mixing/mastering the audio. The standards for recording have grown alongside the industry, and it's now commonplace to record quality audio through a microphone and in a sound-proofed studio. A sound engineer not only ensures a smooth recording process (recording engineer), but can also be tasked with creating a coherent final file respecting genres and distribution conventions (mixing and mastering

[2] https://www.nbcnews.com/video/that-tiktok-podcast-may-not-be-real-182704197597

[3] Comping is an editing technique that combines the best portions of multiple takes into a single track.

engineer).

Producer: Perhaps the most evasive role, as it's sometimes (including in this thesis) used interchangeably with "podcaster" or "podcast creator". Notwithstanding, a producer has actor-specific duties, mostly related to coordinating the podcast-making process. This includes booking guests, planning with and hiring other actors (like an editor), and publishing (or in any other way, finalising) episodes.

It is worth mentioning here that producers sometimes become hosts over time, turning in-show acknowledgements of an invisible production team into dedicated live roles for them, serving to enhance engagement with audiences and fortifying the parasocial relationships that might develop with members of the team.

Executive Producer: Slightly different to a producer, an executive producer is often times found in larger media networks or corporations. They are at the head of several production teams, and often are tasked with coming up with concepts, commissioning, "greenlighting" pilots, and approving final versions;

Researcher / Script writer: For the programs that require such preparation, a researcher or script writer will put together information necessary for a podcast. They will work on the structure of the episodes and overall structure of the show.

3.4 Leveraging AI Tools and New Technologies for Next-Generation Podcasts

3.4.1 The Plural Meanings of "Personalised Media"

In Chapter 2, we described the reasons for and ways in which podcasts can use interactivity, mentioning a palette of examples, from recommendations systems to Spotify polls, to "choose your own adventure podcast", and collaborative podcasting through apps like Stereo. Throughout these paragraphs, we skirt the concept of "personalisation", preferring the somewhat wider idea of "interactivity" as a key to understanding how podcasts are consumed nowadays. If Postma and Brokke (2002) define personalisation as "a *segmented form of communication that sends (groups of) different recipients different messages tailored to their individual preferences*" (p.137), Blom and Monk (2003) rather see personalisation as a process, one "*that changes the functionality, interface, information content, or distinctiveness of a system to increase its personal relevance to the individual*" (p.193).

While personalisation as a feature centers on the content and its modifications, personalisation as a process centers on the user, as the tools and technologies in this case continuously adapt to best meet the individual needs, whether that is to increase accessibility, or to create new forms of experiences. Like Oulasvirta and Blom (2008), who comments and reviews the work of Blom and Monk, I agree that personalisation as a process ensures that the user is integrated within the concept, where personalisation as a feature takes out the users from the definition, having them merely as recipient rather than actors.

These articles focus especially on personalisation in technology, but what of media? And even more specifically, what of podcasts? Vázquez-Herrero and López-García (2019) says: "*our media behavior always seems to involve some level of participation, co-creation and collaboration, depending on the degree of openness or closedness of the media involved*"(p.261). As demonstrated in Chapter 2, podcasting is a particularly "open" media (in the sense that it relies on its users, communication, and innovation, to thrive). It is therefore expected that our "podcast behaviour" would also involve high level of "participation, co-creation and collaboration". Indeed, we see plenty of examples of user involvement and personalisation in the podcast world. From podcast playlisting on Apple podcast to reposting relevant segments on a personal page – podcasting is so intrinsically linked with online behaviour that we can hardly tell the difference between the personalisation offered by podcast distributor website (for instance, Pocket Casts) to a logged-in user, from the one offered by a social media platform (for instance, Instagram). In March 2024, the website of Pocket Casts advertises:

> *"Take your podcasting experience to the next level with exclusive access to features and customization options."* - *Casts (2023)*

This is a great example of the perceived value of personalisation in podcasting. For a medium that is already primed for being personal, user-focused, and intimate, any additional customisation caters to an audience already craving such features.

When I discussed "interactivity" in podcasting, I included in this personalisation. In this way, enhanced media, adaptive media, and flexible media, can all be akin to forms of personalisation. The minute nuances between

those terms are often linked to their modes of delivery and the actors involved:

- Enhanced media (sometimes linked to "hypermedia") refers to additional content latched onto existing media
- Adaptive media refers to content that changes based on user preferences
- Flexible media is an umbrella term for any media that can be easily modified by the user or the media producer

And, within all these terms, the idea of personalisation.

These ideas are often times facilitated through Object-Based Media (OBM). OBM refers to the process of breaking down content into customisable parts. For instance, OBM can be used in adaptive media, relying on objects to customise content for the consumer.

3.4.2 The Two Facets of Personalisation

Burns et al. (2013) sees two facets of personalisation that work hand in hand to facilitate usage and improve the quality of smartphone-based help-on-demand services: "personalisation and adaptation of both content and user interface"(p.2). Similarly, Frias-Martinez et al. (2006) hypothesises that services provided by personalised digital libraries can be categorised into three groups: "mechanisms for the personalisation of content", "mechanisms to help in the process of navigation", and "information filtering and information retrieval mechanisms". Corresponding themes emerge in literature focused on personalisation, particularly web-based (Gao et al., 2010; Murugesan and

Ramanathan, 2001) or smartphone-based technology. (Chen et al., 2014; Tossell et al., 2012)

Along this line of thought, both podcast content and user interfaces for experiencing them can be personalised. Currently, podcasts interfaces are visual, accessed primarily through a smartphone screen (Beniamini, 2022). This leaves many possibilities for new listener-audio interactions, perhaps through voice (Lim et al., 2000; Trivedi et al., 2018), sound (Kong et al., 2019b; Kumar and Raj, 2016b), gesture (Kellogg et al., 2014; Lee et al., 2013), motion recognition (Huynh et al., 2018), or using implicit preferences and habits through user metadata (Rousseau et al., 2005, 2004; Abriella Kazai and Pearmain, 2018).[4]

The benefits of exploring different interfaces would lie in two aspects: immersion and accessibility. By allowing producers to integrate new modes of interaction with their programs, they could ensure that the interface would suit the listener's habits, navigate the content with more ease, or simply enable their audience to better concentrate on the programme. Alternative methods of interaction could be an achievable solution not only for people listening to podcasts while engaging in another activity, when touching a screen is not practical, but also for those with visual or motor disabilities.

The fixed nature of the typical podcast's content is a consequence of the immutable nature of the MP3 format. Other "fixed" media, like TV, are becoming more flexible, with accessible soundscapes (Pardoe et al., 2020; Shirley et al., 2017), or non-linear content (e.g. Bandersnatch, You vs. wild),

[4]The term metadata has a wide range of definitions – more precisely, over 46 definitions are recognised by Furner (2020). Here, it is meant the information stored about the user and the podcasts, particularly the ones that can help personalisation (preferences, habits, categorisations). All subsequent uses of the term "metadata" refer to this specific definition.

and yet still rely on the traditional "fixed" video formats. Therefore, although file format should be considered when developing ways to personalise content, it should not be seen as a barrier to innovation, as exemplified by the growing interest in enhanced podcasts, chapters, and adaptive podcasting, which use a markup language to bring together personalised content for its users (Dwornik, 2021).

Besides non-linear narratives and soundscapes made more accessible via levels control, other modifications of content could be applied to podcasting, such as responsive spatialisation (Pike, 2019; Malham, 1998), voice (Bendel, 2019) and sound synthesis (Parham et al., 2018), reverse engineering tracks to stems (Colonel and Reiss, 2021), and server communication supporting saving of user-generated data and near-real-time collaboration (Chaniotis et al., 2015).

3.5 Mapping Out the Technological Landscape for NGP

Many new technologies could facilitate immersive and personalised podcasting - in fact, there are so many innovative solutions and systems, that it is almost impossible to list them all. Nonetheless, this PhD explores in more depth a selection of such technologies, picked on the basis of usability, ease of implementation, and clear applications to podcasting. Although I am most interested in tools that enable the creation of new experiences for a user, an attractive offshoot of this exploration is that these tools could alternatively be used for production purposes, with no repercussions on the user

experience. For each concept, the advantageous and disadvantageous factors of implementation on both the listener and producer side will be detailed in a table. Finally, the Six-tensions Framework (c.f. 2.1) will be applied systematically to each technology. This exploration into new technologies' possible applications to podcasting will be done in three parts. Section 3.5.1 will give an overview of content personalisation on the user side, while Section 3.5.2 will look at possible changes in interfacing systems, looking at the different types of interactions laid out by Chao (2009) (*Data, Visual, Voice, Intelligent,* updated to be more specific to podcasting applications as: *Data, Visual, Audio, Gesture*).

This will be done through tables (Table 3.4 - 3.9) , examining the advantages and disadvantages of each concept for the listener and producer.

Finally, section 3.5.3 will detail existing frameworks for deploying such technologies within existing tools or structures.

3.5.1 User-Side Personalisation

Voice synthesis: Generation of speech using AI. These voices can replicate an existing person's voice (e.g. deepfakes (Müller et al., 2022)) or emulate a more neutral non-specific voice (Aylett et al., 2021). Synthetic voices can be trained directly on text-audio pairs, or created through a combination of processes: a text analysis framework, an acoustic model, and an audio synthesis module (Wang et al., 2017). At the time of writing, voice synthesis is already a well-adopted technology. There are many tools that enable the public to make use of these complex models, like Amazon's Polly [5], Descript [6],

[5] https://aws.amazon.com/polly/
[6] https://www.descript.com/lyrebird

Table 3.1: Summary of the advantageous and disadvantageous factors of using voice synthesis in personalised podcasting

	Listener	**Producer**
Advantages	- A potentially seamless way to receive customised content - Immediate and clear reward to interaction	- Limits the amount of recording and editing necessary for non-linear programs ("combinatorial explosion" (Bruckman, 1990) problem partially solved) - Overdubbing for corrections/technical issues is no longer necessary
Disadvantages	- Depending on the interface, making the purpose of the required interaction to trigger the personalisation might be complicated - Could hinder a more "passive" mode of listening	- Ethical and moral considerations

or ElevenLabs [7]. Voice synthesis could be a feature implemented on the producer's side (as a production assistant, for example, to avoid dubbing) or on the user side, generating content in real time depending on user behaviour.

In the context of the Six Tensions Framework, voice synthesis would primarily involve the pair "*automation and personalisation*". By automating a portion of the production process, the resulting content could offer a deeper sense of customisation to the user (e.g. using the user's name, or more broadly, creating a narration specific to their preferences).

Soundscape synthesis: Generation of sounds using AI. By sounds we mean any non-speech, non-musical elements that can form the sonic atmosphere of a podcast. Particularly, we think of SFX generation, using tools

[7] https://elevenlabs.io/

Table 3.2: Summary of the advantageous and disadvantageous factors of using soundscape synthesis in personalised podcasting

	Listener	**Producer**
Advantages	- Audio elements could be picked, created, or customised to a user	- Simplifies a creator's workflow through library-free sound designing
Disadvantages	- Overwhelming the user with choice - Possible lag and processing problems	- Sound-designing agency taken away from producers - Quality of sound

like Nemesindo [8], NoiseBandNet [9], or AudioLDM [10]. For the user, it could be used as a way to generate soundscape components in real-time (for instance, to adapt to their surroundings); for the producers, it could be used as a way to circumvent the need for hyper-specific Foley [11] recordings.

In the context of the Six Tensions Framework, sound synthesis would primarily involve the pair "*interactivity and immersion*". By enabling the hyper-customisation of sounds, a podcast can become more realistic, and/or more anchored in the listener's environment, both of which would contribute to immersion.

Responsive mixing: This idea includes responsive spatialisation (Agrawal et al., 2022), adaptive EQ (Gentet et al., 2020; Chanda and Park, 2007), object-based levels control (Ward et al., 2019; Shirley et al., 2017) and source separation into remixable elements (Makino, 2018; Pardo et al., 2018). It would enable sonic objects, or tracks, to be mixed in real-time, adapting

[8] https://nemisindo.com/
[9] https://www.adrianbarahonarios.com/noisebandnet/
[10] https://huggingface.co/spaces/haoheliu/audioldm-text-to-audio-generation
[11] "Foley sound effects are custom sounds made in post-production." ado (2024)

to a user. If the producer maintains control over these changes (by picking settings or parameters for instance), responsive mixing could in turn make programs more accessible. For instance, the Narrative Importance plugin described by Ward et al. (2019) enables producers to create different versions of audio tracks associated to different levels of comprehension that the user could require.

On the spatialisation front, native Web Audio packages like Resonance SDK [12] or Tone JS [13] allow for the spatial rendering of audio files. This spatial immersion can also enhance an experience by adapting to a user's hardware (Oldfield et al., 2015; Niamut et al., 2013), customising audio to a particular reproduction system without the need to manually create different versions to fit different systems.

Source separation can be used independently or as a complement to these techniques, enabling an audio file to be broken down into various tracks that could undergo some form of personalised processing. There are some well-known models for source separation that can be used for unmixing audio into its components (Nugraha et al., 2016; Vincent et al., 2018) and user-friendly apps exist, such as of Moises [14] or Demucs [15].

In the context of the Six Tensions Framework, responsive mixing would primarily involve the pair "*unique and universal*". By enabling a podcast to be re-mixed to follow user interactions or preferences, each listen is different and unique, but could also reach wider audiences, by offering more accessible experiences.

[12] https://resonance-audio.github.io/resonance-audio/
[13] https://tonejs.github.io/
[14] https://moises.ai/
[15] https://demucs.danielfrg.com/

Table 3.3: Summary of the advantageous and disadvantageous factors of using responsive mixing in personalised podcasting

	Listener	**Producer**
Advantages	- Immediate and clear reward for interaction - Improves immersion - Improves accessibility	- Many possibilities within one umbrella concept - Allows for complex acoustic environments to be rendered
Disadvantages	- Overwhelming the user with choice - Complex interface - Possible lag and processing problems	- Sound-designing agency taken away from producers - Quality of sound

Non-linear storytelling: This broad category includes both the more traditional "choose your own adventure" style content, and more nuanced approaches like variable length programs with BBC's Squeezebox [16]. In all cases, it equates to changing how content unfolds to respond to user preferences or choices. It already exists in many forms, but is yet to be fully explored for podcasts, mostly due to distribution or implementation issues. Some tools such as StoryKit [17], Cutting Room [18], and charisma.ai [19], have been designed to facilitate the production of such experiences, although none focus primarily on audio.

In the context of the Six Tensions Framework, responsive mixing would for instance involve the pair "*interactivity and immersion*". By allowing a listener to interact with the content and choose the narrative thread they

[16]https://www.bbc.co.uk/rd/projects/squeezebox
[17]https://storykit.io/
[18]https://audienceofthefuture.live/cutting-room/
[19]https://charisma.ai/

Table 3.4: Summary of the advantageous and disadvantageous factors of using non-linear storytelling in personalised podcasting

	Listener	**Producer**
Advantages	- Immediate and clear reward to interaction - Improves engagement - Gamification of podcasting	- Can present various points of view and experiences from a single show
Disadvantages	- Overwhelming the user with choice - Gamification of podcasting - Can render a passive experience impossible	- Issue of "combinatorial explosion" - Editorially driven relevance (not all projects would be suitable) - Extra planning required to cater for branching narrative structure - Script agency is taken away from producers

follow, they can become more engaged with the program, which would contribute to overall immersion. But, similarly, by forcing interactivity, one might disrupt immersion.

Participatory podcasting: This idea refers to communicating changes to a programme and reactions between listeners, possibly in real time Chaniotis et al. (2015). This could be achieved via a number of web-based systems, including a simple Node.js web-app, or a custom-built integrated interface within a podcasting platform.

In the context of the Six Tensions Framework, responsive mixing would primarily involve the pair "*immersion and interactivity*". Encouraging audiences to interact with a podcast and other listeners would increase feelings of social belonging and overall engagement, but in doing so, might shift the

Table 3.5: Summary of the advantageous and disadvantageous factors of using participatory systems in personalised podcasting

	Listener	**Producer**
Advantages	- Immediate and clear reward for interaction - Improves engagement	- Many possibilities within one umbrella concept
Disadvantages	- Overwhelming the user with choice - Complex interface - Possible lag and processing problems	- Project specific, not applicable to all podcast

focus from the contents of the programme to the interactivity in itself.

3.5.2 Interfaces

Visual interface: Visual interfaces are already the main way we interact with podcasts -via a screen, whether it's a phone screen or a laptop screen. However, some of the content personalisation technologies mentioned in the previous section might require more complex interfaces that would call for even more specialised visual interfaces.

In the context of the Six Tensions Framework, pushing visual interfaces further would involve the pair "*current audience and possible demographic*". Basing deeper interactions on an already mastered mode of interaction could help cater to both existing audiences and appeal to new ones.

Metadata: The term metadata has a wide range of definitions – more precisely, over 46 definitions are recognised by Furner (2020). Here, we mean the information stored about the user and the podcasts, particularly the ones that can help personalisation (preferences, habits, profile information etc ...).

Table 3.6: Summary of the advantageous and disadvantageous factors of relying on visual interfaces in personalised podcasting

	Listener	**Producer**
Advantages	- Convenient - Already integrated and accepted	- Can be based on/extend existing software and platforms
Disadvantages	- Requires a visual interaction for an audio-based medium (eg. Visually impaired listeners, or listeners who cannot look at or interact with a screen) - Some of the personalisation options mentioned would require very complex interfaces which might be overwhelming for smaller devices	- Requires new visual content

Table 3.7: Summary of the advantageous and disadvantageous factors of relying on metadata in personalised podcasting

	Listener	**Producer**
Advantages	- Convenient - Already integrated and accepted - Avoids interruptions	- Allows for the content to be generated or modified before the podcast starts, therefore preserving creative agency
Disadvantages	- Users might not wish to share their data	- Requires a system for handling user data

Personalisation through metadata could enable for more "passive" customisations, basing modifications on previously fetched data like user preferences and habits. A few visual programs make use of this type of interface, like "Instagramification" [20] and "Brooke leave home" [21].

In the context of the Six Tensions Framework, pushing the role of metadata in personalisation further involves the pair "*unique and universal*", as more user-specific content can be delivered, therefore reaching a broader listenership.

Audio interfaces: We are already used to some audio interfaces, with Alexa, Siri, Google Assistant, and other virtual assistants that can be triggered by recording speech or sound. This type of interface includes both speech detection (natural language processing), and sound recognition (audio event tagging). It would provide an audio-only way to interact with content. For speech, it could rely on recognising sentences like "more bass", "I feel happy", or simply, "tell me more" for the listener, or recognising set

[20] https://www.bbc.co.uk/taster/pilots/instagramification
[21] https://brookeleavehome.github.io/

Table 3.8: Summary of the advantageous and disadvantageous factors of relying on audio interfaces in personalised podcasting

	Listener	**Producer**
Advantages	- Information shared is volunteered - Already integrated and accepted - Convenient (hands free) - Improves immersion	- Creative and practical applications
Disadvantages	- Some environments are not optimal for speaking aloud - Could be unsettling for certain use cases	- Complicated implementation within existing workflows

phrases in the content, like filler words "erm" or "um", for the producer. For sound recognition, a wide range of processing and reactions could be triggered by recording the listener's environment. On the production side, sound recognition could be used to facilitate production, by improving tagging or chapterisation (this particular application of sound recognition will be explored in more depth across the next chapters).

In the context of the Six Tensions Framework, pushing the role of audio interfaces for the producer involves the pair "*automation and personalisation*", by automating some tagging tasks, it'll be easier to create personalised experiences, like non-linear narratives.

Motion interfaces: This includes gesture recognition (hand/head tracking), device motion (tilt recognition) and touch recognition (on-screen mouse motion). These techniques are already popular in AR/VR/XR, and would require some form of motion capture (either via ML, or simply a rule-based system associating specific gestures to actions).

Table 3.9: Summary of the advantageous and disadvantageous factors of relying on motion interfaces in personalised podcasting

	Listener	**Producer**
Advantages	- Information shared is volunteered - Intuitive - Convenient	- Creative and practical applications
Disadvantages	- Requires free range of motion	- Requires some changes to the user interfaces in place to include additional processing

Alternatively, motion recognition hardware can also be attached to portable devices, as Google's project Soli offered [22].

In the context of the Six Tensions Framework, pushing the role of audio interfaces for the producer involves the pair "*technology and art*", by linking some more functional new technologies to creative applications.

3.5.3 Producer-side Implementation

Some content personalisation tools for producers already exist, and although none are specifically linked to podcasting, it was important to review the ones available through this doctoral project: a BBC tool - StoryKit[23] - and an XR Stories [24] tool - Cutting Room[25] - in order to determine whether they could be extended/refined into an NGP application.

Cutting Room: Cutting Room is an Object-Based Media editor, primarily intended for film. It is a tool developed with funding support from

[22]https://www.theverge.com/2019/10/15/20908083/google-pixel-4-project-soli-radar-motion-sense-explainer
[23]https://www.bbc.co.uk/makerbox/tools/storyformer%20
[24]https://xrstories.co.uk/
[25]https://audienceofthefuture.live/cutting-room/

XR Stories that ran as a web tool (2018) before being moved to Unity (2019). It enables producers to create non-linear and interactive stories, but at the time of writing requires a software specialist to accompany the producers in the production process (Manni et al., 2019). This in turn has the effect of making it a project-by-project solution rather than a production-agnostic software (Ursu et al., 2020b). This makes the software complicated to scale up for larger productions or production networks, but very relevant for highly specific research projects.

Storykit: StoryKit is a BBC R&D OBM tool freely available through BBC Makerbox [26]. It makes it possible to arrange media elements into different narrative threads and customise the portals between each narrative branches via explicit or implicit interactions (Armstrong et al., 2020). It authors[27] and delivers programs in a dedicated player. It can have a wide range of applications, from varying the length, depth, narration type or format, depending on the elements dragged. Storykit is a mostly visual tool, that only caters for two audio tracks (background and foreground), and does not offer the same range of transition options in the audio domain than in the visual one. Some examples of StoryKit projects can be found online, such as Click 100[28], Dance Passion: NOISE[29], or Instagrammification [30]

Web based : The solution of creating a custom-built web interface is still available to circumvent some of the compatibility issues mentioned with the above tools, podcasters and their workflows. A web-based R&D tool

[26]https://www.bbc.co.uk/makerbox

[27]"Authoring is the creation of documents, especially for the internet."- Collins online dictionary

[28]https://storyplayer.pilots.bbcconnectedstudio.co.uk/experience/click1000

[29]https://www.bbc.co.uk/taster/pilots/noise

[30]https://www.bbc.co.uk/taster/pilots/instagramification

could be translated into a local application, like a plugin, or desktop app.

3.6 Personalised Audio: Formats and Systems

3.6.1 Designing Interactive Media Tools

Interactive media tools have to take into account the summation of all the involved actors and their intrinsic relationships and expectations. When designing interactive media tools, an engineer must consider not only an editorial drive, but also how the content will be received and modified by audiences, and how the distribution platforms it will rely on will alter the experience. Becker et al. (2017) defines a framework for audiovisual design, seeing each actor involved as either Audience (uses the content), Synthesizer (engages and shares the content), Modifier (improves or re-mixes content), Player (interacts with the content) or Producer (creates the content). The relationships inherent to these actors form a network across four axes: Identity, Motivation, Experience, and Content. This is also the case for interactive media tools more specifically. Therefore, when designing new tools for interactive media, a complex framework of interplaying actors must be kept in mind. There are many ways to design interactive media tools, using any of the actors, from Audience to Producer, as a starting point.

When looking at the interdependence between the Audience and the Content particularly, it is important to raise the question of value (Vázquez-Herrero and López-García, 2019). Is the interactivity worth the effort required of the actors? Is there a reward to using such tool? What motivates

the interactivity? In the context of a study looking at the effects of varying levels of different audio objects in media excerpts with hearing impaired people, Shirley et al. (2017) speak of "acceptable overhead" for personalisation, framing the phrase as a reasonable demand of user-interaction to trigger adaptive features.

Beyond the issue of "acceptable overhead" for personalisation, an important query regarding of interactive media is the issue of sharing and preserving the content created. Becker et al. (2007) says that "*preserving the inherent complexities of interactive multimedia is a very difficult task, particularly because formats used in multimedia art are ephemeral and unstable*"(p.259). Indeed, this ephemerality and instability, combined with a lack of standards and pre-existing pipelines makes designing interactive media tools all the more challenging.

3.6.2 Formats for Interactive Media

Interactive media requires much information to be distributed to audiences. This data is carried in files oftentimes hidden from audiences, but it is there, under the surface, ensuring for example the proper functioning of the latest interactive Netflix program [31]. Each file therefore carries a number of layers, each with different categories of information. There are four different layers for interactive podcast files: Content, Context, Structure, and Behaviour (adapted from Becker et al. (2007)). We will detail these layers taking the example of a non-linear narrative podcast :

- Content is the media (most likely, audio) data of the podcast – here,

[31] Although, they are sunsetting the trend, we cannot forget Netflix's forays into interactive media in the recent years

the section that is played after the listener makes a certain choice.

- Context is what surrounds the content - here, a reference to the narrative as whole, including authorship information and other relevant information.

- Structure is the place of the Content within the Context – here, the place of a section within the narrative.

- Behaviour is the way the Content will play – here, the transitions (fades) information, the volume, the logical gate it needs to trigger after it plays to trigger the following section, etc.

In terms of format, the information within these layers can be shared by combining two types of data: a multimedia/audio format (e.g. WAV, MP3, or M4A) and its accompanying metadata (either as a file header, or as a separate, but attached, file).

The evolution of audio file formats for podcasts was mentioned in Chapter 2. Nowadays, there are still mainly four options for sharing podcast audio [32]:

- MP3 (small file size, low quality, compatible with all the main podcast distributors)

- M4A/AAC (small file size, good audio quality, possibility to add bookmarks/chapters, (almost always) compatible with podcast distributors)

- WAV (large file size, lossless quality, usually not compatible with distributors)

[32]https://www.acast.com/blog/podcaster-resources/best-audio-file-formats-for-podcasts

- FLAC (small file size, good audio quality, usually not compatible with distributors)

There are several options for the implementation and distribution of interactive media.(Meixner et al., 2017, Figure 1) tracks the timeline/evolution of formats for interactive multimedia.The adaptive podcasting app described by Dwornik (2021) makes use of SMIL [33], a markup language derived from XML [34]. It is a relatively simple and straightforward way to annotate, compose, and author object-based media, although heavily reliant on being read by an appropriate system to be delivered to a listener. Nested Context Language (NCL), another application language of XML, is a declarative authoring language sometimes used for hypermedia documents (Xavier Leitão et al., 2020; Meixner et al., 2017). These are current examples of metadata formats for interactive media, but as we think of applications for NGP, we can look to the coming trends in order to postulate what the future of metadata for interactive media could be.

In keeping with an XML based solution, García and Celma proposes a mapping system architecture, that uses MPEG-7 mapped to OWL (Web Ontology Language). MPEG-7 is a comprehensive, although large, file format defining a wide range of elements, attributes and types [35]. A practical application of MPEG-7 in the context of personalised TV broadcasting is described in Niamut et al. (2013), using MPEG-7 Audiovisual Description Profile (Sano et al., 2013) to describe tracked elements (e.g. persons, or regions), within a football match visual recording.

[33]https://www.w3.org/AudioVideo/RA-examples.html
[34]https://developer.mozilla.org/en-US/docs/Web/XML/XML$_{i}ntroduction$
[35]https://www.mpeg.org/standards/MPEG-7/

OWL [36] is a semantic web language designed to represent rich and complex knowledge about things, groups of things, and relations between things.García and Celma's complex proposal of an encoding system demonstrates the complexity of concisely storing and sharing interactive multimedia data.

3.6.3 Personalised Audio: A Heterogenous Landscape

Ward (2020) describes the pipelines for creating and delivering object-based audio for a few case-study projects. In the case study looking at accessible episodes of the TV program Casualty, using a system of Narrative Importance, Ward (2020) describes the trials and tribulations of having a real-world application for an R&D tool that uses unstandardised formats. Particularly, the beginning of their project included three different toolsets to go from production to testing out the user interface. We can assume more intermediaries will be necessary when such projects are taken all the way to audience's living rooms.

In the case of podcasts, although the visual aspect is null (except in the case of video podcasting, but as per our definition of podcasting, Chapter 2, this research focuses on audio-mostly content), there is still the issue beyond producing and sharing, of reading whichever format is chosen to share the personalised components of the program. In the case of adaptive podcasting (Dwornik, 2021) a bespoke user app with its own interface was created so that the SMIL files could be read by the listener's phone. This highlights the lack of standardisation present in this field. The delivery systems, just as the formats, depend entirely on the stakeholders involved. But do

[36]https://www.w3.org/OWL/

we need standardisation? As when radio went from one standard (FM) to many (DAB/DMB, DAB+, DRM, DVB) (Jedrzejewski, 2015), can podcasting withstand the format boom that comes with the endless innovation with which it is associated?

If we return to our example of non-linear podcasts, used in the previous section, what could be a safe and appealing standard so that these experiences could be shared across platforms regardless of who produced them? Do the platforms have to come up with such standards, or will they emerge naturally as new formats are tried and tested in smaller test environments? BBC taster[37] is a great example of a tendency to silo R&D projects and limit their reach to research outputs. BBC Taster was a platform intended to highlight new media content and processes, as well as serve as a research-highlight of BBC R&D. Although the platform is no longer maintained, when it was active, it offered a great overview of tools that could possibly enter the market for general usage. However, more times then none, the ideas presented there didn't make it to public usage.

3.7 Summary

This chapter investigates the literature and technologies pertaining to the development and implementation of new production tools for NGP, specifically focusing on the idea of personalised audio. By highlighting the literature on other new media, in comparison to what can be found in academic research on podcasts, it becomes apparent that some more formal investigation into the processes of podcasting is necessary to properly justify and implement

[37] https://www.bbc.co.uk/taster

the development of NGP tools.

To further our understanding of what NGP could entail, a description of how AI-tools and new technologies can be used within the context of podcasting is given. This includes an overview of "personalised media", followed by a non-exhaustive list of technologies that could be used in personalised podcasting and are particularly relevant to this project. Finally, a review of prior work on formats for personalised media provides the context required to think of practical applications of the topics discussed throughout the chapter. This work highlights the gaps in literature necessary to be filled conduct this research, fleshing out the justification of the methodology subsequently used, as well as an answer to RQ 2.

4 Methodology

4.1 Introduction

Method sections of PhD thesis can be particularly complex, as describing the unravelling of over three years of work, while simultaneously detailing the particulars of specific analysis processes can prove challenging. This chapter will provide the bibliography and justification necessary to understand the choices made in the methodology roadmap presented in Section 4.6. It will describe the various analysis processes undergone, although more detail about the set-ups, participants, and data analysis techniques used in each of the individual studies conducted will be given in their associated chapter.

4.2 Approaching Podcast Research

4.2.1 A Note on Philosophy of Science

In this section, let us give some attention to philosophy of science and technology – as a way to establish an author's point of views and motivations. This isn't necessarily common practice in the field of computer science or engineering, but it is an expected part of many humanities research theses. Because this PhD lands in the murky (but wonderful) waters of interdis-

ciplinarity, I want to take the time to discuss this research's approach to scientific knowledge. As such a topic can hardly be covered within a few paragraphs, please consider this brief summary of a researcher's point of view simply as a point-measure of an otherwise continuous process.

In Chapter 1, I briefly mentioned Harman (2019)'s work discussing Object Oriented Ontology (OOO). I believe it has much to bring to academically driven innovation and new technology research as a whole [1]. As a quick reminder, in OOO, everything is an "object", whether it is alive, artificial, or even conceptual. An object can be real (its undeniable essence), or sensual (how it is perceived), and is defined by real and sensual attributes that can be experienced. Van Den Eede (2020) examines OOO in the context of Philosophy of Technology, highlighting some of the theory's most thought-provoking applications to this field. Indulging in some vulgarisation and context-specific analogies, these applications are as follows:

- *Uncertainty* - we can only approach reality indirectly ; "all podcasts" is an object. When trying to define this wide concept, as was done in Chapter 2, we can only attempt this by looking at its relationships with other objects, whether these are literature, or our personal opinion.

- *Objects come together to form new objects* - separate objects can meld into "compound objects"; we can create the object "personalised podcast" from the "personalised media" object and the "podcast" object.

- *Creation of New Objects instead of Discovery of the Withdrawn* - finding an object's real attributes does not simply add to the already existing

[1] Like any philosophical current, one will find adopters and retractors, this is simply an opinion.

object's figurative "list" of properties, but rather creates a more transparent, known, object; for instance, answering "RQ 1: What is NGP" means we create a new object "NGP that has received a certain level of academic scrutiny" different from "NGP that is being briefly defined in the introduction of this thesis for clarity's sake".

- *Objects have life trajectory* - objects contribute to other objects in a timeline; continuing this analogy, these two different NGP objects have a temporal relationship, where one influences the next.

In these few examples, it becomes apparent that OOO is not only a greatly potent ontological framework, but also can form a driving contributor of a researcher's philosophy of science. It is particularly well suited for new technology research, as eloquently put by Van Den Eede (2020): "*We must investigate further how technology at the same time appears and disappears*"(p.210). Embracing the constant changes in technology and multimedia inherently encourages discovery and curiosity, and an attitude that values all objects (knowledge, output, software, ideas etc.) created along the way as they can contribute to other objects' "life trajectory".

Indeed, this idea matches our reference to Popper (1992)'s work mentioned in Chapter 2; their Theory of Falsification (Popper, 1992), where scientific hypotheses are provisional, confirmed over time by empirical validation, or eventually disproved through falsification is at the center of much investigative and exploratory work necessary to the advancement of science.

4.2.2 "Creator-Centric" vs. "Listener Centric" Podcasting Innovations

In Chapter 3, we looked at Becker et al. (2017)'s framework for audiovisual software design in the context of podcasting and interactive media, and concluded there were as many points of entry into innovative podcast software design as there were actors involved (Audience, Synthesier, Modifier, Player, Producer) – and that more broadly, there were two directions the process could take, a creator first direction (or, a "creator-centric" focus), where the designer focuses on the needs and expectations of the creators, and a listener-first direction (or, a "listener centric" focus). This is represented in Figure 4.1

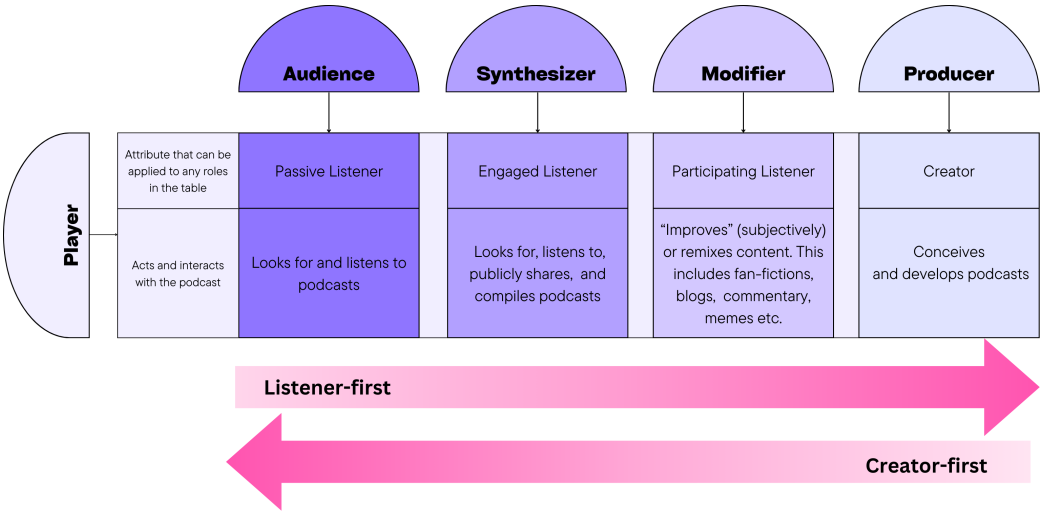


Figure 4.1: Diagram translating Becker et al.'s actors in the Audiovisual Desgin process and their roles to podcasting.

The arrow "listener first" means that the reflection for tool design is centred around the expectations and experiences of the audience. Chapter 2 examines the flourishing listener-centric literature available. It brings

us valuable information into the current habits of podcast audiences, often across a large pool of participants. Indeed, the only way to draw conclusions from such a varied and large demographic ("listens to podcast") is to query varied and large subsets of this population. This would require resources that a single PhD student, even with the help of institutions such as the BBC and XR Stories, would struggle to bring together. Additionally, a keyword of this thesis is "audio production tool" – which clearly states the intended recipient of the research: people who would use such tools. Why begin with Audiences to mould a product for Producers?

The "Creator-first" arrow follows such logic. It focuses on establishing the needs and requirements of Producers to distribute new content for Audiences. Working with a creator-centric approach means that our development process will come from the fields of HCI rather than Media and Social studies (Becker et al., 2017). It enables us to base design decisions on smaller groups of individuals, forming focus or testing groups to feedback on the development. It also ensures that the outputs of these tools will be embraced and anticipated by producers. As an exaggerated analogy, say that through a thorough analysis of a large group of podcast listener, we find audiences are interested in consuming more clown-related podcasts, should producers be expected to produce such content even if they are afraid of clowns?

This issue is also true in a creator-first direction. The enthrallment for "choose your own adventure" style content across the 2015-2020s is a great example of this, and more specifically, how the lack of equivalently enthusiastic audiences meant that many projects were side-tracked or simply put to rest Moore (2024). The issue of conflicting expectations of creator and

listener-centric reflections is complicated to tackle. A solution could be to choose a direction (either creator or listener first), and refine the product by evaluating it role by role; for example, producers want a non-linear storytelling tool, modifiers are happy with the breadth of content to re-mix, synthesisers aren't very engaged with the end show, and passive audiences report only making decisions because they had to.

This layered evaluation approach is a great way to fully examine a media technology and its impact on all actors – but, as mentioned, would require resources that a PhD student does not have access to. The decision was made to only focus on one role, the one of the "Producer" in a creator-centric approach. This means that the first layer of this overall evaluation will be conducted in this thesis, leaving evaluating its application to future work.

4.3 Using Participatory Design To Build Creative Tools

4.3.1 User-Centred Design

In this "top-down" (or, more accurately in Figure 4.1, right to left) approach to designing tools for NGP, we can use UCD techniques to involve the target user (the podcaster) in the development and help steer the project. The benefits of user involvement for software development have been outlined by many researchers. Kujala (2003) shows that user involvement has positive effects on system success and user satisfaction, and Kuhn (2000) says collecting target users' design requirements can guarantee a user-informed

design process, and therefore contribute to the end product. This has solidified the practice of User-centred design (UCD) in the past decades (Abras et al., 2004).

UCD is "*a broad term to describe design processes in which end-users influence how a design takes shape.*"(Abras et al., 2004). Software developers are routinely taught to consider the end-user when designing projects (Dennis et al., 2015). It can make use of and borrow from many different design and development techniques, such as for instance:

- Agile Software Development (ASD) [2]
- Participatory Design (PD) [3]
- Iterative Software Development (ISD) [4]

Multiple methods aim to connect designer and user, and hybrid solutions that mix and match approaches are becoming more common. For example, Ferrario et al. (Ferrario et al., 2014) describes a framework that melds together elements of ASD and ISD, where ASD itself comprises different

[2] "*What makes a development method an agile one? This is the case when software development is incremental (small software releases, with rapid cycles), cooperative (customer and developers working constantly together with close communication), straightforward (the method itself is easy to learn and to modify, well documented), and adaptive (able to make last moment changes).*"(Abrahamsson et al., 2017, p.19)

[3] "*As the name implies, the approach is just as much about design, producing artifacts, systems, work organizations, and practical or tacit knowledge—as it is about research. In this methodology, design is research. That is, although participatory design draws on various research methods (such as ethnographic observations, interviews, analysis of artifacts, and sometimes protocol analysis), these methods are always used to iteratively construct the emerging design, which itself simultaneously constitutes and elicits the research results as co-interpreted by the designer-researchers and the participants who will use the design*" (Spinuzzi, 2005, p.164)

[4] "*With an iterative and incremental approach, [the software development] process is completed little by little, step by step, by splitting the overall project into several mini-projects, each of which is called an iteration*"(Bittner and Spence, 2006, p.6)

methods that have in common the "unforgiving honesty of working code and the effectiveness of people working together with goodwill" (Highsmith and Cockburn, 2001), and ISD fundamentally includes conversation and feedback (Basil and Turner, 1975)).

The process of requirements gathering (RG) is a key aspect of many of these methods. It is a pertinent way to ensure a software developer is not faced with an endless list of requests from their intended users once a project is already finished, but also to go beyond habits and bias in design. Involving users in software design can take various forms: Lane et al. (1995) distinguishes three main ways to agglomerate people's perspectives in this context: *surveys*, *prototyping*, and *open-ended interviews*. Regardless of the way feedback is gathered, there is an emphasis on establishing and maintaining an open dialogue between the designer and user.

Kautz (2011) lists the benefits of using Participatory Design (PD) as follows:

> "*(1) Improving the knowledge on which information systems are built. (2) Enabling people to develop realistic expectations, and reducing resistance to change. (3) Increasing workplace democracy by giving the members of an organization the right to participate in decisions that are likely to affect their work*" (p.216)

PD is not often criticised; its implementation almost falls under common sense for many projects. However, it can sometimes be complicated to implement, because of time or monetary constraints, as it is a complex process involving large groups of stakeholders. Similarly, ASD requires planning and complex organisation strategies to avoid delays (Kula et al., 2022). The next

section outlines some of the limited criticism RG receives.

A parallel idea to PD is co creation: "*Co-creation involves the joint creation of value by the firm and its network of various entities (such as customers, suppliers and distributors) termed here actors. Innovations are thus the outcomes of behaviours and interactions between individuals and organizations*" (Perks et al., 2012, p.935)

From a literature review, and empirical studies with companies and businesses, Frow et al. (2015) identifies nine motives for co-creation: *access to resources, enhance customer experience, create customer commitment, enable self-service; create more competitive offerings; decrease cost; faster time to market; emergent strategy; build brand awareness.* From similar resources, Frow et al. (2015) hypothesises a design framework of 12 concepts – here, our project particularly calls upon ideas of (1) "co-conception of ideas" and (2) "co-design".

One of the driving factors for this research to integrate forms of PD and co-creation, aside from the clear motivations stated above, is to mitigate bias. Technology is a particularly biased environment – AIs especially are inherently biased, as they are developed, trained on, and used by a really small subsection of the population (Crawford, 2021). These biases colour research and without being addressed, will carry on skewing research trajectories in related fields. Similarly, podcasting and audio-engineering are known for lacking diversity (Chapter 2). By integrating end-users, we attempt to re-introduce some inclusivity in the use cases and applications of these inherently biased systems.

4.3.2 Requirements and Feedback Gathering

Gathering the opinions of stakeholders is a key aspect of all the design techniques that fall under UCD (Moore and Shipman III, 2000). Design requirements are necessary to inform the various methods that utilise PD. By gathering requirements, a designer can "*capture information through the use of multidisciplinary views. Such views express what is to be built.'* (Christel and Kang, 1992, p. 34, l.19). This design roadmap can be created using *ethnographic anecdotes*, *expert verification*, *usability testing* or*semi-structured interviews* (Salminen et al., 2022). The common denominator of these procedures is communication (Lane et al., 1995). Because RG does not come with strong methodological constraints, Kuhn (2000) argues that one of its main drawbacks is that there can be no scientific analysis of the data gathered, and hence no scientific grounding for design decisions made from these requirements. This can be mitigated by applying qualitative analysis methods such as thematic analysis (Clarke et al., 2015) to transcripts of conversations or tests, offering a justification for otherwise informal impressions. This analysis can sometimes be aided by automated, procedural, or AI-driven tools (Moore and Shipman III, 2000).

Kautz (2011) describes an integrated framework for user participation, and when defining "user", includes *individual*, *average* and *fictive* users. "Fictive users" refers to the use of personas. Personas are fictitious, specific, concrete representations of target users (Pruitt and Grudin, 2003, p.11). Personas are often based on data collected through ethnographic study and observation, (see "Cooperian Personas" (Floyd and Twidale, 2008)), and help designers visualise their archetypal users and their needs. Personas can be

utilised in role-playing, focusing on issues, meeting maintenance, empathy, clarification or approximation (Friess, 2012). When used for role-playing, designers will often "put-on" these characters, pretending to interact with a prototype or give feedback on an idea. This reinforces the storytelling aspect of persona use, as described by Pruitt and Grudin (2003). Personas are not just typical users, they are also characters, Pruitt and Grudin (2003) even uses the term "method acting". Role-playing was particularly useful for this project, as a way to test out features and how users would react to them. By "putting on" end user personas informed by the literature review and the interviews carried out throughout the development process, it becomes easier to pick out flaws in the design.

The purpose of using personas for design is summarised by (Salminen et al., 2022, p.5) as a way to elicit "*user preferences and requirements necessary for designing key software components*". The usefulness of using personas for PD was evaluated by Grudin and Pruitt (2002), who compared the effect of PD with real and fictional people. Although the use of personas is supported by their findings, the author concludes that "*personas are not a panacea. They should augment and enhance – augment existing design processes and enhance user focus.*"

4.3.3 Iterative Software Development

Iterative software development is a framework for software design that often includes forms of PD. One of its foundational aspects is the ability to "make progress in the face of change, or perhaps in spite of change" (Bittner and Spence, 2006, p.23).

It is built on the basis of "loops" where a developer's idea is refined via end-user involvement, then modified based on the feedback received, and finally, the process is repeated again. Rauterberg et al. (1995) calls these loops "optimisation cycles", and divides each into two phases "action" and "test", where the action reacts to a test phase, itself bringing "intereferences and constraints" to prior iterations of the project. Figure 4.2 gives a general example of a more modern iterative development model.

Figure 4.2: Iterative software development generalised process

Iterative models have the advantage of allowing a development team to show results of prior iterations and gain feedback from target users of the system (Orel, 2022a). However, because repetitive models often require this user engagement throughout the entire process, each new iteration will likely require testing and feedback from users to evaluate the necessary changes, making it a costly, time-consuming, and inconvenient endeavour (Orel, 2022a).

4.4 Understanding Practitioners: Mapping Out Current Practices and Creative Intentions

4.4.1 Questionnaires

In amongst the many methods of gathering opinions, questionnaires are a convenient, reliable, easily deployable way to collect data (Fricker and Schonlau, 2002). Questionnaires can be distributed physically or online. Because this PhD began in November 2020, Covid-19 restrictions and considerations were still largely applied, and for most of this research, no contact – online interactions are therefore prioritised (Benson et al., 2021).

Online questionnaires enable a researcher to reach a diverse population in order to create a broad dataset (Nayak and K A, 2019). They are relatively low cost, can be deployed over a short period of time, in a convenient fashion for the participant (Nayak and K A, 2019; Fricker and Schonlau, 2002), and informed consent online does not substantially differ from obtaining it physically (Varnhagen et al., 2005). Add to this the fact that data does not need to be digitised, and can be anonymised and stored seamlessly, online questionnaires become an attractive solution for collecting data from large pools of respondents.

When looking at the advantages and disadvantages of online surveys, the literature mentions a few important aspects. First, the practical problem often reported of "poor participation" or incomplete questionnaires (Nayak and K A, 2019; Heiervang and Goodman, 2011); indeed the nature of online

questionnaire, both impersonal and completely user managed, means it's simple to skip questions, give up, or answer to things quickly, much more so than in situations where a researcher physically encourages a participant to fill out a questionnaire. Secondly, a system with purely digital take-up and implementation comes with some ethical considerations – including something as specific as "use of email addresses" (Lefever et al., 2007), to a broader issue like data storage. Finally, and perhaps most importantly, online questionnaires can be linked to "non-probabilistic" sampling, that is that the results of the queried set will not be representative of a wider group, due to biases linked to "self-selection, under-coverage, non-response, and sampling errors" (Nayak and K A, 2019).

Overall, Lefever et al. (2007) summarises the consensus on using online questionnaires:

> *"Despite their limitations, web-based surveys provide researchers with unique opportunities for collecting data through the Internet. They can be particularly useful for collecting preliminary data and for pretesting research design and question comprehension." - (Lefever et al., 2007, p.581).*

4.4.2 Interviews

> *"Qualitative interviews exist on a continuum, ranging from free-ranging, exploratory discussions to highly structured interview" -(Magaldi and Berler, 2020, p.4825).*

Unstructured interviews are free-flowing conversations with no *à priori* or pre-prepared topics Mueller and Segal (2014), where structured interviews follow a specific pre-established framework.

If the goal is to investigate broad concepts, and to allow participants to

elaborate on specific ideas and contribute beyond the scope of the question set, semi-structured interviews are a highly appropriate data collection solution (Gillham, 2000). Adams (2015) describes semi-structured interviews as "*a blend of closed and open-ended questions, often accompanied by follow-up* ***why*** *or* ***how*** *questions*" (p.493). Thus, by nature, semi-structured interviews are well-suited for exploratory investigations and studies. They break down the steps involved in semi-structured interviews as follows: selecting and recruiting the respondents, drafting the questions and interview guide, techniques for this type of interviewing, and analysing the information gathered. When preparing for semi-structured interviews, a researcher must establish an "interview guide" (Adams, 2015, p.496). These guides can be brought together by a researcher, or if the topic requires it, by using a focus group.

However practical semi-structured interviews are, they come with some drawbacks: they can be "*time-consuming, labor intensive, and require interviewers' sophistication*"' (Adams, 2015, p.496). They are also restricted in terms of reach, so therefore do not warrant precise outputs.

4.4.3 Workshops

The term "Workshops" has taken a lot of meanings over the past decade, from university practicals, to corporate meetings. The Cambridge dictionary defines it as "*a meeting of people to discuss and/or perform practical work in a subject or activity*". Workshops can be set in person or online to accommodate for restrictions, and can consist of synchronous or asynchronous tasks (Benson et al., 2021). Jones (2018) sees stakeholder workshops as integral parts of co-creation within a PD framework. The designer then acts

as a facilitator who guides the users through a design and idea-generation process (Sanders and Stappers, 2008). A common process to facilitate innovation is the double diamond [5]. Books like Gray et al. (2010) detail useful techniques that gameify brainstorming tasks to maximise involvement throughout a workshop [6].

4.5 Analysis Methods

4.5.1 Qualitative Analysis

Much of the data gathered from questionnaires and interviews requires formal analysis. Before analysis, interview recordings need to be transcribed. For this, we use a combination of speech-to-text software (Descript) and checking the generated transcripts by hand while listening to the recordings. This allows the researcher to get an in-depth familiarisation with the data, while still being aided by contemporary tools.

A common method for analysing the responses of participants in an agile and fluid fashion is Thematic analysis. The method for thematic analysis used in this research follows the "phases of thematic analysis" as described by Braun and Clarke (2006): *Familiarizing yourself with the data; generating initial codes; searching for themes; reviewing themes; defining and naming themes; producing the report.* The exact nature of the thematic analysis is highly dependent on the nature and content of the data (Maguire and

[5] "*The Double Diamond is a visual representation of the design and innovation process. It's a simple way to describe the steps taken in any design and innovation project, irrespective of methods and tools used*" - https://www.designcouncil.org.uk/our-resources/the-double-diamond/

[6] https://gamestorming.com/

Delahunt, 2017). There are a few tools that can be used for coding and attributing themes, the most widely used being Excel and NVivo.

By its very nature, qualitative research deals with subjectivity. The issue of bias therefore arises organically. We try to minimise the skewed aspect of thematic analysis by adhering to current best practices best practices (Braun and Clarke, 2006; Castleberry and Nolen, 2018; Guest et al., 2012), through acknowledging bias, and having discussions surrounding the codes and results (Malterud, 2012; Chenail, 2011; HU and CHANG, 2017).

4.5.2 Quantitative Analysis

Although this research features a lot of qualitative analysis, quantitative elements are a key portion of two of the studies conducted. Mainly, statistical analysis was necessary to analyse Likert scale responses, and to gauge the efficiency of an algorithm within a complex evaluation process.

Where there are Likert scales, non-parametric, inferential statistics were used to analyse the results [7]. Specifically, a Friedman test can be performed to determine integral differences between ratings, followed by a Wilcoxon signed-rank test to compare the ratings to one another. Where possible, the experimental processes used repeated measures design, meaning that ratings were available across all data points.

The algorithmic evaluation was based on Inter Annotator metrics (IAA)[8], rather than Mean Opinion Scores (MOS). Where MOS are greatly useful to

[7]Non-parametric tests relate to data that cannot be assumed to fit normal distributions. Inferential statistics make predictions about a population from a sample from that population. For more detail and maths, please hold for Chapter 5

[8]IAA is a metric that represents how well several annotators agree on their annotations. For more detail and maths, please hold for Chapter 6

compare an algorithm to other similar ones on the market, IAA enables a direct comparison of the performance of a human group (most likely, a group of experts), and the performance of an algorithm.

4.5.3 Ethics and Data Privacy

All studies carried out in this thesis received ethical approval from the University of York Arts and Humanities Ethics Council. These studies were also compliant with BBC ethical and legal guidelines.

Participants' data were handled in accordance to this approval. All required data were stored on secure, university-approved sites, and were anonymised in the records. Participants' ID were encrypted, and throughout the thesis, participants were attributed random labels, different to their participant IDs, for an added level of protection. The participants that chose to wave their anonymity are still given labels, until Chapter 8, which sees a few real producers use the tool created. It is necessary to address who these creators were in order to convey the full breadth of the case studies.

4.6 Research Methodology Roadmap

I have addressed and justified the main techniques that will be used to explore the three aims presented in Chapter 1:

- **RA 1.** Mapping the habits and expectations of podcasters
- **RA 2.** Exploring the peculiarities of immersive and personalised podcasting

- **RA 3.** Investigating and documenting an application of participatory design to the development of an AI-driven next generation podcasting tool, all the way from conception, to functional software

In order to investigate RA 3. we used both ISD principles in a creator-first approach, and PD techniques. Through this process, we gathered the opinions of creators on present and future podcasting (RA 1), and were able to comment on the nature of immersive and personalised podcasting (RA 2) through practical projects and empirical studies with real producers. Using ISD, and by extension, RA 3, as a framework for our investigation ensured the project amounted not only to functional software, but also to more theoretical contributions regarding podcasting and the integration of new technologies. The research questions 1-4 [9] will be answered as a synthesis of the process undergone.

With an ISD framework at the centre of this research, we can represent the process through a simple diagram, as seen in Figure 4.3. We start our journey on the left-hand side (light blue arrow), and loop our way through towards the right (black arrow). The keen observer will realise we have, without their knowledge, already embarked into the first loop – as reviewing the literature and mapping the technologies available are the first steps of the blue-to-purple loop.

[9] RQ 1: What is next-generation podcasting?
RQ 2: How can we feasibly create and distribute immersive and personalised podcasts?
RQ 3: Why is modular podcasting attractive to creators, and how can it be facilitated without making the production process more complex?
RQ 4: What are the benefits, risks, and costs of exploiting AI technologies for podcast production?

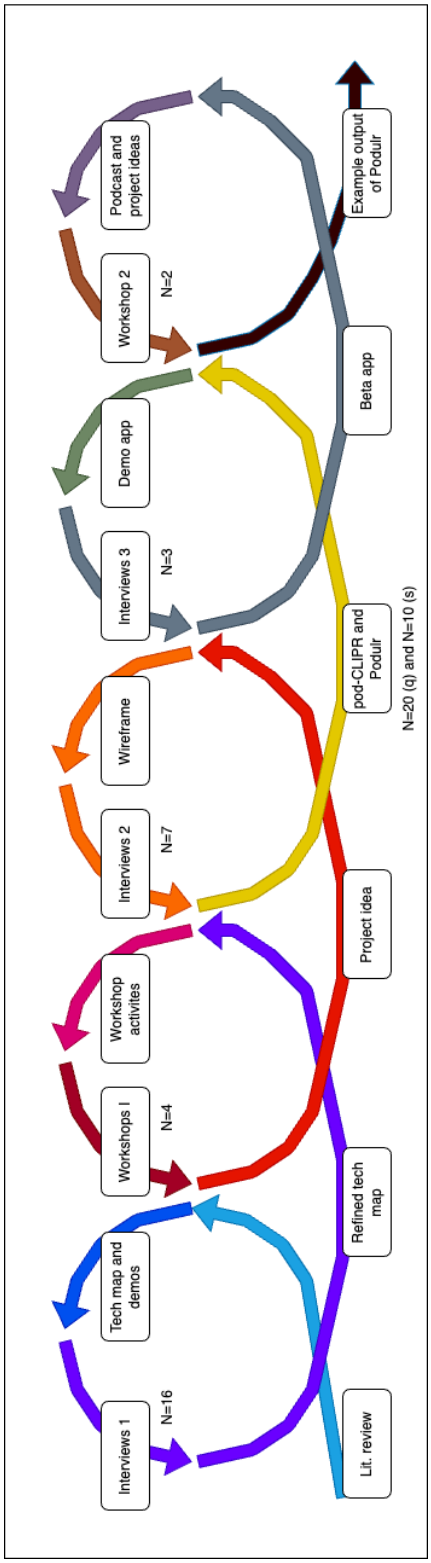


Figure 4.3: ISD process applied to this doctoral project. N is the number of participants involved at each step.

Each loop is annotated with "N=" ; that N represents the number of participants involved on this iteration of design. For instance, the first set of interviews were carried out with 16 podcasters, and the last workshops were joined by 3 producers. Overall, this process included around fifty different participants.

For ease of reference, let's further detail the content of each of these loops, called by the colour of their top right arrow, referencing the chapters and sections of this thesis that cover them.

Blue Loop: (1 – Chapter 2) Starting with a review of existing knowledge, in the field of podcasting research and interactive media tool development, (2 – Chapter 3) we mapped the applicable technologies and ideas that could contribute to the concept of NGP. These were selected on the basis of relevance, applicability, and feasibility (within short-medium time frame and non-existent budget). (3 – Chapter 5) We presented these technologies to 16 independent and BBC podcasters, to gather immediate reactions and thoughts from a varied group of creators. We took the opportunity of these semi-structured interviews to also make some progress on RA 1.

Pink Loop: (1 – Chapter 5) The combined quantitative and qualitative analyses of these interviews enabled the refinement of a new technologies map, narrowing the number of concepts in play and their potential applications. (2 – Chapter 5) We designed a workshop using this refined map as initial input, focusing on tangible applications and use cases. (3 - Chapter 5) We conducted these workshops with four participants.

Orange Loop: (1 – Chapter 5)From these workshops, we defined the bounds of the tool, based on the most popular and common proposed NGP

ideas. This is where the idea of "modular podcasting" (RQ 3) is introduced [10]. (2 – Chapter 6) Based on this goal, we created a wireframe for a web-app, (3 – Chapter 6) that we presented to seven producers in a set of more informal interviews and questionnaires.

Green Loop: (1 – Chapter 7) The feedback from these conversations not only provided us with wireframes modifications, but also informs the creation of a back-end system (pod-CLIPR) that enables automatic chapterisation using a combination of sound recognition and a rule-based system. The rule-based system was informed by a questionnaire with 20 producers, and the finished system was formally evaluated with a group of 10 participants. (2 – Chapter 8) We integrated this system to a demo web app, Podulr. (3 – Chapter 8) Three producers reviewed this app and provided some feedback.

Grey loop: (1- Chapter 8) The reviews given by the producers were integrated within a functioning beta version of the app. (2- Chapter 8) Concrete applications to this app were conceived and producers were contacted, so that (3 – Chapter **??**) three different use-cases could be workshopped.

To a certain extent, Figure 4.3 is misleading because it makes these cycles appear of equal length and effort – they were not. The first set of semi-structured interviews with 20 people took much more planning and analysis than the informal interviews that occurred in the green loop. Similarly, building and evaluating pod-CLIPR took months, including the better part of a six-month placement with BBC R&D.

[10]The possibility create multiple versions of a same programme to cater to different listeners

4.7 Summary

Through an examination of the existing methodological processes, this chapter details the general method this research follows, including some more specific information regarding analysis techniques. The flow of the research is detailed in a diagram (Figure 4.3). More particulars regarding statistical analysis and concrete examples of thematic analysis can be found in the dedicated results chapters of this thesis (Chapter 5 - 8).

5 What a Podcaster Wants, What a Podcaster Needs

5.1 Introduction

This chapter introduces the first steps in the participatory design process, refining the literature review and mapping the new technology, presented in Chapter 2 and 3, through interviews and workshops with podcast producers. The structure of this chapter follows the chronological order through which this iterative process takes place. Section 5.2 describes and discusses the results of semi-structured interviews with 16 podcast producers with the goal of discussing their current practices, and their views of NGP. Section 5.3 follows the preparation and outcomes of the first creator-workshops, to further converge the focus of the research. The interviews lead to some initial answers for *"RQ 1: What is NGP*?", as the initial investigation of the literature review can be combined with the opinions of professionals in the field. The generalisation of their roles, habits, and workflows (Figure 5.2) also provides a basis to answer *"RQ 2: How can we feasibly create and distribute immersive and personalised podcasts?"* as it highlights the ways in which new tools could be integrated within existing production pipelines. This chapter also introduces the concept of modular podcasting, key to both *RQ 3* and *RQ 4*.

Most importantly, the results discussed in the following sections are highly relevant to the industry outside of the specific context of this research, as the investigation conducted is exploratory in nature and therefore broad in its scope and application.

5.2 Podcast Creators' Perspectives

The literature review in Chapter 2 and 3 has accentuated the necessity to gather more information on podcasters and their outlook on the future of podcasting. This knowledge is instrumental to the development of new software, tools, and, more generally, products for podcasting. Knowing a target user is paramount to UCD (Abras et al., 2004). Re-iterating a few items from Chapter 4 relevant to this study, there are various methods for interacting with target users and gathering requirements, including *ethnographic anecdotes*, *expert verification*, *usability testing*, and *semi-structured interviews* (Salminen et al., 2022). Semi-structured interviews are particularly useful when attempting to investigate large concepts and topics, allowing participants to elaborate on initial answers and contribute beyond the scope of the question set (Gillham, 2000), which was the purpose of this first set of interviews. The interview guide (Adams, 2015, p.496) reflected this broad scope of investigation, relying not only on open-ended questions, but also on Likert scales, and reactions to short demo videos highlighting the technologies described in Chapter 3.

After a detailed review of the study design process, I will present and analyse the results of these interviews, before highlighting some of the limitations specific to this study as well as a discussion of the analysis. Short discussion

sub-sections are included following the chronological order in which they occurred throughout the results chapters, but a full review of results - including for each study, as well as the broader research axes – will be conducted in Chapter 9.

5.2.1 Study design

Participants

Sixteen creators (independent and BBC podcast creators) took part in an exploratory study. They were recruited through a combination of word-of-mouth, BBC internal communication channels, and media publication advertisements. A gender balance was not achieved, with only 3/16 (= 19%) participants identifying as female. This distribution is supported by a recent study from Sounds Profitable and Edison Research looking at 617 active podcast creators in the US, finding that only 29% of podcasters identified as female, and 2% as Non-Binary or Other (Devlin, 2022). Sixty-two per cent of participants were independent creators, while the remainder were affiliated with the BBC. Twelve percent were 25-35 years old, 7/16 (= 44%) were 26-50, and 5/16 (= 31%) were 51–65. 12/16 (= 75%) had over five years of experience in the field, with a few reporting having been involved with podcasting since the early days of the medium (2005-2010).

"Producer" was the most common occupation (with 12/16 (= 75%) of participants describing producing as one of their main roles when making a podcast). The term "jack of all trades" was mentioned freely, without cue, in 5/16 (= 31%) interviews. 4/16 (= 25%) described themselves as "hosts", 4/16 (= 25%) as "sound engineers", 3/16 (= 19%) as "advisors",

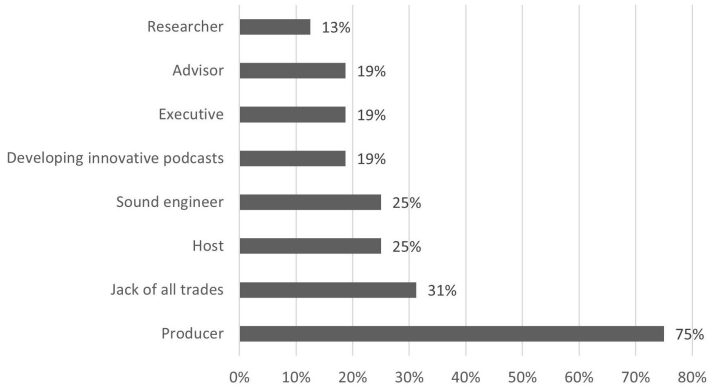


Figure 5.1: Current self-reported professional roles of participants within podcast productions.

3/16 (= 19%) as "developing innovative podcasts", 3/16 (= 19%) as "executive producers" and 2/16 (= 13%) as "researchers". These proportions were represented in 5.1, and justify the usage of the word "creators", as the roles of a podcaster can be varied, even when working for a large media organisation.

The genres of podcasts in which participants were involved are varied, according to Spotify's genre classification in 2022 (Spotify, 2022), with recent work in "Lifestyle" (6/16 (= 38%)), "Stories" (4/16 (= 25%)), "Business and Technology" (3/16 (= 19%)), "Educational" (3/16 (= 19%)), "True Crime" (2/16 (= 13%)), "News and Politics" (2/16 (= 13%)), Sport (1/16 (= 6%)), "Comedy" (1/16 (= 6%)), and "Music" (1/16 (= 6%)) reported.

Procedure

Due to the nature of the research questions, an exploratory study was designed, taking the form of approximately 45-minute long, semi-structured interviews (Gillham, 2000; Adams, 2015), conducted over Zoom with individual participants. The interviews covered participants' views on their

current work and the future of podcasting. The participants were contacted one week following the interview with a questionnaire to gather additional thoughts that might form after the interview.

Participants were first asked a series of questions about their work and creative process. These questions were designed to detail the inner workings of podcast production, and are set by a focus group of researchers and podcast producers from the BBC. Participants were also asked to react to 12 short videos presenting new technologies that could be applied to podcasting (Table 5.1), divided into two categories ("personalisation of content" and "personalisation of interface") to maintain focus and coherence in the discussion. They were asked to rate how interested they would be in using them on a scale of 1 (strongly disinterested) to 5 (strongly interested) and prompted for more information on particularly high or particularly low scores.

In the follow-up questionnaire designed conjointly with the interview guide, the same videos were used as prompts once again to gather any additional thoughts. The participants were also asked what most interested them in the study and given a space to provide feedback.

Materials

To build this interview guide, a series of conversations were held with a group of expert podcast producers from the BBC, supervisors, members of the BBC R&D Audio team, and R&D members who were interested in the project and could potentially facilitate future workshops.

These meetings functioned as focus groups: "*a technique involving the use of in-depth group interviews in which participants are selected because*

they are a purposive, although not necessarily representative, sampling of a specific population, this group being 'focused' on a given topic". (Thomas et al., 1995)

The expertise of the attendants of these meetings was invaluable, as they had experience leading such interviews, but also, first-hand knowledge of either podcast production processes, or audio engineering. The conversations around the interview guides spanned two separate meetings, in-between which I was able to refine a draft of questions and expectations.

In the first hour-long meeting, I proposed an initial proposal for conversation points to have with producers, stemming from the literature review presented in Chapter 2. The questions answered by this meeting were as follows: **A/** How can this interview encourage creative, out-of-the-box thinking from participants?; **B/** How can this interview cater for executive producers, sound engineers, and every other possible podcast-related profession without restricting the scope of the questions presented?; **C/** What is the appropriate format for demonstrating tools and technologies to participants?

To address **A/** , a new question was added to the draft guide to begin the interviews with systematically, asking participants to engage in "blue skies thinking" about the future of podcasting, to kick-start a creative reflection. To address **B/**, the focus group agreed that having different "scripts" for different profile types was an adequate way to deal with the wide range of jobs and responsibilities of participants. These scripts were thought of in terms of modular conversational flows and guidelines, rather than strictly restricting a participant to a specific set of questions. For **C/**, a few options were explored, including live and pre-recorded demonstrations. Live demonstrations

were judged to introduce too much variability (what if a tool didn't work properly on the day? How would this influence the participants' response?), so the focus group concluded that short pre-recorded videos would showcase the tools, without the possible pitfalls of sporadic errors and overall lack of accurate of repeatability.

Before the follow-up meeting, a new proposal for an interview guide and associated material (demonstration videos) were put together. These were presented to the focus group, inviting final feedback and reactions. Most changes stemming from this meeting concerned phrasing and pacing, highlighting the need to allow participants to be able to drive the conversation towards their own opinions and preferences. From these conversations, a final interview questions (italics) and rationale for asking the questions (non-italics) were set as follows:

1. *If anything was possible, what's a podcast you would like to make or hear that would transcend the current format?* To get the conversation going, creators were asked to project themselves into a boundless future, where any podcast project could be carried out.

2. *What tools are necessary for your work?* Knowing what equipment or software producers rely on would allow us to determine how to make a new tool compatible with their current setups.

3. *What attributes make for good podcasting tools?* Participants were asked to provide some adjectives that justified their choice of software to infer some requirements for a podcast production tool.

4. *Do you have a particular workflow when creating a podcast?* Workflow

processes or schemas are defined as "specify[ing] which tasks need to be executed and in what order" (Aalas and Jablonski, 2000, p. 267). Here, I mean the sequence of events that begins with the idea for a programme and ends with a final product available to audiences. Understanding the production habits of practitioners would enable any new podcasting tool to find its right place within an established process.

5. *Do you have any experience with coding? And would the need for coding deter you from using a tool?* New media tools can sometimes require users to code in order to achieve a particular feature. This question sought to evaluate whether this requirement is reasonable from podcast creators' perspectives.

6. *What do your listeners seek in your programs?* This question was answered primarily from a producer's perspective, therefore ended up interpreted as "Why do you want your listeners to tune in?", rather than "Why do they listen to your programs?". Knowing the motives of producers could help contextualise their answers and understand how to help them achieve their goals.

Following these questions, the creators were queried on some concepts that were deemed particularly relevant by: the focus group used to bring together the interview questions; and the review of literature and technological capabilities conducted upstream of the interviews (Chapter 3). These concepts are detailed in Table 5.1. For each concept presented, a technology, existing tool, or software was showcased, to demonstrate the possible applications of the concept to podcasting. These technologies were selected not

necessarily because they were the best in their respective fields, but because they could easily be deployed and showcase the features in focus adequately.

As decided during the focus group meetings, so that all the participants would respond to the same stimuli, a series of 40-second video presentations was put together to explain these concepts. Applications were not described in detail, so that the participants did not feel compelled to restrict their imagination to particular potential uses.[1]

Data Analysis

Interviews were transcribed using Descript, and an inductive thematic analysis was conducted per interview question using NVivo. The rationale for conducting the analysis per question was to obtain clear answers to individual questions, and that the codes for each question would -by design - divide the themes per question in turn. To make sure no overarching theme was overlooked by this particular method, the transcripts were investigated in their totality after a question-by-question analysis, to make sure all cross-question thematic occurrences were noted. Although the pool of participants was relatively small, the percentage of participants mentioning specific themes in their answers is calculated, in order to get a representation of how common ideas and opinions are across this group. This type of numerical data analysis strategies can complement a qualitative analysis (Guest et al., 2012; Cavanagh, 1997, p.112).

It is important to acknowledge that the group interviewed might not be perfectly representative of podcasters as a whole, but they represent a group of users interested in using NGP tools – which is the important factor of the

[1] These videos can be accessed in the supplementary material provided.

Table 5.1: Summary of the contents of the video demonstrations presented to the participants of this study. No particular technology was presented for Metadata, as it is a wide topic that can be addressed differently depending on the desired outcome. These demo videos are provided in the supplementary material.

Type of Personalistion	Concept	Example Technology Used For Demonstration
Interface	Gesture Recognition	MediaPipe (JS)
	Touch/Tilt Recognition	NexusUI (JS)
	Voice Recognition	Descript
	Sound Recognition	Audioset Tagging CNN
	Metadata	Not Applicable
Content	Non-linear Narratives	StoryKit
	Reverse Engineering Music	Moises
	Sound Synthesis	Nemesindo
	Voice Synthesis	Lyrebird
	Server Communication	Node (JS)
	Responsive Spatialisation	Resonance Audio SDK (JS)

ISD process.

Quantitative data were gathered also, via a Likert scale, and analysed using non-parametric, inferential statistics. A Friedman test was performed to determine differences between concept ratings, followed by a Wilcoxon signed-rank test to compare the ratings of each concept to one another. A Friedman test was performed rather than a Kruskal–Wallis test because this design took a repeated measures approach for each topic. This quantitative analysis complemented the qualitative analysis, and offered insight into whether some opinions were widely shared among participants. This enabled us to construct a global overview of the creators' opinions on a wide range of subject matters.

The results of this interview are separated into two sections: "How do you pod?..."(Section 5.2.2), which focuses on current practices, and "How will you pod?..." (Section 5.2.3), which looks at the views of the interviewees on NGP.

5.2.2 How Do You Pod? Revealing the Archetypal Podcast Production Workflow

For the sake of transparency, and to demonstrate how the thematic analysis was performed throughout this study, the complete table of codes is provided in Table 5.2. The first column corresponds to the codes recorded from the transcripts at first examination (step 2 in the phases of thematic analysis according to Braun and Clarke (2006)). The ratio of participants mentioning this code, q, is noted in the second column. After going through the first instances of codes, these were grouped by theme, as indicated in the third

column. The ratio of participants mentioning such a group, with each participant only counted once in each group of codes, is presented in the fourth column, again labelled q.

In response to interview question 1 (p. 131), participants were keen to envision new ways of making or delivering podcasts. 7/16 (= 44%) of participants expressed an interest in increasing or facilitating listener engagement. Participant A shared their interest in forms of social audio:

> *"The whole area of social audio is really interesting, and I think I would like to do more that combines social listening and on-demand audio... Probably the next thing for us in terms of innovation, aside from extra insight [on our audience], and aside from producing more and better podcasts, would be to engage more deeply with our audiences. And I think possibly social audio is one way of doing that. That would be quite interesting to explore." - Participant A*

This is the first example encountered of creators focusing on "listener-centric" innovations – ways to improve or change the listener experience. In a similar mindset, 6/16 (= 38%) of participants expressed an interest in personalised podcasting. Participant B brought up the concept of "hyper-personalisation", meaning content is modified based on elements of the listener's environment or context.

> *"Hyper-personalisation, that could be discretely slipped into shows to make things really interesting. The easiest example I have of what's available right now is you can have a show and then the host is like: 'It's 6:59 PM', and if you look at your clock, 6:59 PM too! If you were to listen again, it would [say]: 'It's three in the afternoon'. All these types of things can really make for an engaging experience. Something as simple as when you start the show, it says, 'good morning' if it's in the morning and 'good afternoon' in the afternoon." - Participant B*

Table 5.2: Detail of the thematic analysis process for interview question 1 *If anything was possible, what is a podcast you would like to make or hear that would transcend the current format?*, with q the ratio of participants mentioning the themes in their interviews.

Codes recorded from transcripts	q	Thematic groups	q
Connected audience	5/16 (= 31%)	Listener engagement	7/16 (= 44%)
Learning more about audiences	1/16 (= 6%)		
Reaching a global audience	1/16 (= 6%)		
Universal story	3/16 (= 19%)		
Accessibility	1/16 (= 6%)	Personalised podcasts	6/16 (= 38%)
Adaptive podcasts	2/16 (= 13%)		
Flexibility within personalisation	1/16 (= 6%)		
Interactivity	3/16 (= 19%)		
Chose your own adventure podcasts	2/16 (= 13%)		
Easier interaction	1/16 (= 6%)		
Non-fixed podcast	1/16 (= 6%)		
Celebrity interviews	1/16 (= 6%)	Pushing or exploring other genres	6/16 (= 38%)
Reality podcasts	1/16 (= 6%)		
Pushing fiction podcasts further	3/16 (= 19%)		
Pushing storytelling further	2/16 (= 13%)		
Immersion	5/16 (= 31%)	Immersion	5/16 (= 31%)
Passivity	1/16 (= 6%)		
Spatial audio	2/16 (= 13%)		
Better audio quality	1/16 (= 6%)	Technical ameliorations	5/16 (= 31%)
Recording easier, in better quality	3/16 (= 19%)		
Lower entry to production	1/16 (= 6%)		
More efficient editing	1/16 (= 6%)		
Questions the form of podcast	2/16 (= 13%)	Questions the form of podcast	2/16 (= 13%)
No changes necessary	1/16 (= 6%)	No changes necessary	1/16 (= 6%)
Prioritizing audio during production	1/16 (= 6%)	Prioritizing audio during production	1/16 (= 6%)

This was echoed throughout the interviews by 5 participants, who were keen to be able to integrate these "hyper-personalised" features into their productions.

6/16 (= 38%) were interested in expanding their own work via exploring new genres. One participant thought podcasts do not need to be changed, but amended his answer later during the interview, talking about immersive audio and accessibility for disabled and international audiences. Overall, the following idea was shared by most:

> *"I think that we're at such an early stage in the podcast industry, we've barely scratched the surface." - Participant G*

In response to interview question 2 (p. 131), participants focused on several aspects of their work. Table 5.3 represents the grouped codes emerging from participants' answers regarding editing software. Although Adobe Audition was used by 7/16 (= 44%) of participants, 14 other types of editing software were mentioned. Participants also detailed their preference in recording tools: Zoom is used as a recording tool by 5/16 (= 31%) participants, but 6 other pieces of software (like Riverside or Zencaster) were mentioned also.

In response to interview question 3 (*What attributes make for good podcasting tools?*), efficiency was mentioned by 10/16 (= 63%) of participants, and so was compatibility (with software, but also with team workers). Utility was brought up in 7/16 (= 44%) of interviews, while a tool being comfortable was important to 5/16 (= 31%) of participants. 2/16 (= 13%) of creators wanted their tools to be good value for money. Participant K went into detail regarding their choice of software and why the most important features

Editing Software	q
Adobe Audition	7/16 (= 44%)
Protools	4/16 (= 25%)
Sadie	3/16 (= 19%)
Hindenburg	2/16 (= 13%)
Logic	2/16 (= 13%)
Audacity	2/16 (= 13%)
Powair	1/16 (= 6%)
Ableton	1/16 (= 6%)
Descript	1/16 (= 6%)
Reaper	1/16 (= 6%)
Wavelab	1/16 (= 6%)
RX Advanced	1/16 (= 6%)
Garage Band	1/16 (= 6%)
Sony Vegas	1/16 (= 6%)
Levelator	1/16 (= 6%)

Table 5.3: Detail of the range of responses to the subsection of question 2 *What tools are necessary for your work?* focusing on editing software.

of podcasting software are the ability to easily collaborate on projects and follow industry conventions:

> *"[I use] Pro Tools because the clients I work with are using it, and it really comes down to collaboration. I think that if a format were to come around that would improve on the AAF (Advanced Authoring Format) for the OMF (Open Media Framework) file exchange formats, you might see people being a little more agnostic when it comes to their audio editors. But because the predominant number of projects I use are in Pro Tools and it's very, very hard to get information out of one audio editor and into another in a seamless way, so I'm going to use what everyone else is. If I woke up tomorrow morning and everyone was in Logic for some reason, or Reaper, I'll learn that and use that, but that's not the case." - Participant K*

The idea of setting an industry standard for podcasts to facilitate work across teams and platforms is a diverging evolution from the independent

and free nature of podcast production, but as the medium evolves and attracts larger more mainstream stakeholders, the need for uniformity in format develops too.

All participants spoke of following a specific routine for podcast production, although it might vary slightly depending on the requirements of the project. Overall, some similarities emerged between participants' answers. To interpret them, a thematic analysis is performed on the transcripts of the interviews. The thematic analysis revealed the codes and thematic groups shown in Table 5.4.

Initial Codes	q	Thematic Group	q
Advising	3/16 (= 19%)	Advising	3/16 (= 19%)
Casting or Booking	4/16 (= 25%)	Pre-production	15/16(=94%)
Conceptualisation	8/16 (= 56%)		
Organisation	7/16 (= 44%)		
Research	3/16 (= 19%)		
Scripting	8/16 (= 50%)		
Work-shopping	1/16 (= 6%)		
Recording	15/16 (= 94%)	Production	15/16 (= 94%)
Sound-designing	15/16 (= 94%)		
Edition	15/16 (= 94%)		
Publication	8/16 (= 56%)	Post-production	13/16(=81%)
Revisions	8/16 (= 50%)		

Table 5.4: Codes and themes from the analysis of participants' transcripts when asked to describe their workflows

The first column represents the initial key codes found when looking at the typical workflow descriptions. The ratio of participants q who mentioned the code is noted in the second column. The deduced thematic groups and the ratio of participants who mentioned them appear in columns three and

four, respectively.

Although q gives valuable insight into the prevalence of each process, this result also seems to correlate with participants' jobs. Comparing column two of Table 5.4 with Figure 5.1, 19% of participants reported being "Advisors", which corresponds to $q = 3/16$ for the code "Advising". Similarly, 25% of participants reported being "Sound engineers" and the same participants mentioned "sound design", together with another participant describing themselves as a "jack of all trades". The value of q allows us to contextualise and justify this generalisation, but should not be thought of as a true representation of the amount of industry professionals undertaking a particular action within the workflow produced.

Excluding "Advising", which was mentioned as part of various stages of the production process, three groups were identified, using widely accepted media production terminology:

a) Pre-production

Pre-production comprises the concepts of conceptualisation, organisation, research, scripting, work-shopping and casting/booking. Participants for whom conceptualisation is part of their pre-production process (9 out of 16) remarked their projects often began with questions:

> *"Often it begins with research. It starts with a story idea, with a theme and with exploring that theme and understanding how much depth there is in a particular story, then you decide the format: is this a one-off? is this a documentary? [...] Has it got enough juice and story in it to be a podcast series? If it does, then what format does that take?" - Participant A*

> *"How can we sustain that? What are the threads that need to be involved? [...] What are the voices that we need? We just*

> *throw everything onto the table, and then [ask] okay, how does this [story] break down? Where is our episode? Where do we start? What are the characters we need to introduce first?" - Participant B*

> *"We first were trying to decide: is this something we pitch to one of our employers? [...] The next question was format, literally, what is it going to sound like? So, we went through a couple of different iterations of [versions]: what if we started it out this way? How do we want to organize? How much do we want to cover? Do we want to have guests? Do we not want to have guests? How do we want to integrate them into the show?" - Participant C*

Organisation was portrayed as a crucial matter, particularly for large-scale productions and weekly shows, where creators spend a lot of time planning episodes in advance, often relying on planners and productivity or organisation software. It is apparent that a lot of thought goes into planning episodes, booking guests or actors, organising recording sessions, etc. All but one of the four of participants reporting being involved in "stories" also talked about "scripting". But, as seen in Table 5.4, 50% of participants overall mentioned scripting as being a part of their workflows. So, what other types of programmes are these creators involved with? After querying the data, it can be observed that scripting is part of the workflow of those working within the following genres: stories, lifestyle, music, news and politics, sport, true crime, and comedy.

Only one BBC creator used the term "workshopping", seemingly as a practice that could be equivalent to both "conceptualisation" and "scripting" occurring at the same time.

b) **Production**

All but one participant mentioned being involved in the production and recording phase. This participant is a sound engineer and advisor, who has more impact at the beginning and end of the production workflow. Recording was reported to be done in studio, over the internet, in a closet or a bedroom, with several creators mentioning the COVID-19 pandemic as a factor in the adoption of more "creative" recording set-ups.

Although sound design and editing would be considered post-production processes in other media such as film or game design Swartz (2004), I argue that for podcasts, these steps play an integral part of the final product, where recording, sound designing and editing function like a constructive loop rather than a linear process.

Most participants follow an iterative production workflow, with 50% of interviews mentioning revisions occurring at one point or another of their projects. For creators of scripted programmes, it was described as common practice to write a script, record audio, then to go back to modify the script until it was deemed satisfactory.

> *"We had a full script, and a full draft of audio. [...] It sounded awful, but you have to get that first can of grumpy draft out. And from there, I went back to the script myself and with the story editor re-worked and re-worked that for two other months."* - *Participant B*

c) **Post-production**

Similarly, revisions are triggered in the post-production phase by superiors, advisors, colleagues, or the creators themselves, deciding to go back to the production phase to amend their programme until they are satisfied with the result.

> *"It goes through a rough cut and a final cut, and editorial approvals, and then rounds of feedback and improvements, kind of like an iterative process in that sense." - Participant A*

This was common for BBC and independent creators alike. The post-production phase often involves these kinds of revisions, and then a publication phase follows wherein the podcasts are distributed and promoted. The following quote summarises the post-production phase quite succinctly:

> *"This is good enough, set a release date. Get it out there, and get it to the people" - Participant D*

Figure 5.2 represents the podcast creator's archetypal workflow, inspired by the operational sequence diagrams graphs appearing in Baume (2018). This generalisation into an archetype stems from the observations made in Section 5.2.2. A gradient (shading) is applied to each concept and group introduced in Table 5.4 as they appear in Figure 5.2, to represent the frequency of participants mentioning the term in their interviews (c.f. Table 5.4, column two), from green (1 participant or 6%) to red (16 participants or 100%). The flow between categories was inferred from the sequential order arising from the interviews.

A particular project might call for certain steps to be skipped or repeated, with the most common iterative production patterns represented in Figure 5.2, going from recording/sound-editing to scripting/work-shopping, and from editing to recording/sound-editing. Looking at narrative podcasts, from the point of view of both an independent (participant D) and a BBC creator (participant E), there is a clear, common process:

> *"Let's say this is a scripted piece. You get the concept going, [...] recording the scripts' content in order. You go through various*

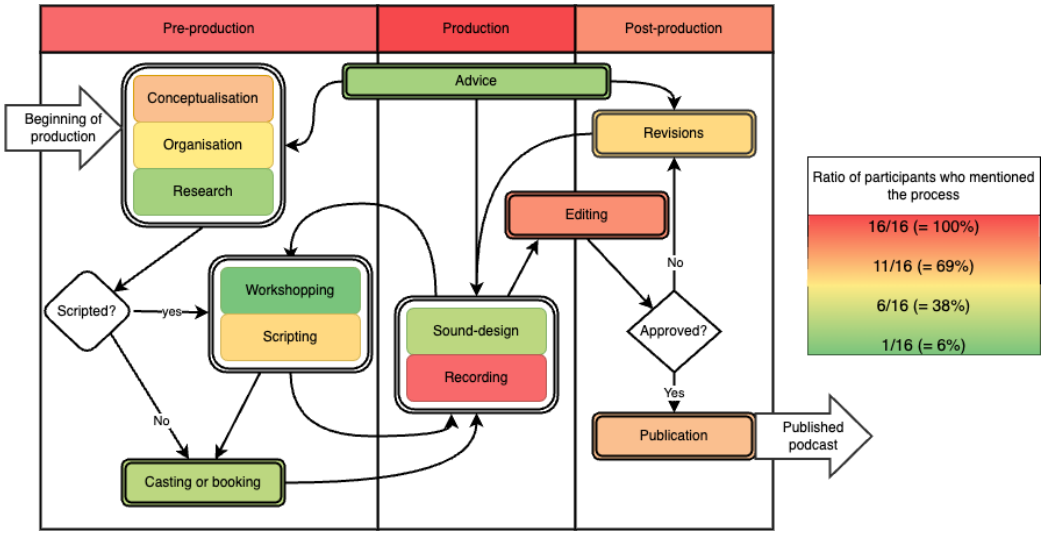


Figure 5.2: Diagram of the archetypal podcast production workflow, using the codes and themes presented in Table 5.4. A gradient corresponding to the frequency of participants mentioning the concept is associated with each idea. Editing is represented at the threshold between production and post-production, as it could be included in both groups.

> *edits, reviews, you might record a couple of those scripts and then just edit the audio and then listen back to it and [...] add this, maybe add that [...]. It depends on the constraints of the project, it might be once, or twice, or three, or four times, depending on how much time you've got. Then you get to the final products and review it, so that the higher-up people listen to it, or in my case, I'm the person calling the shots." – Participant D.*

> *"You've got the script ready; you've gone through four or five drafts with the writer, and you're more or less happy with it. You format it [and] you send that out to the actors; you get your cast together; you have your recording day, or two days, or three days, or however long it takes. And then at that point, I would send notes to [a] supervisor and they would cut it together and send me a speech edit, which might be a bit too long. [...] I will then cut it down again so that it's closer to the time and get rid of the bits I didn't like and play around with it for a bit. Send it back to them. Usually with, if I can, some of the key elements of music on. Then they'll cut it together with the music and sound design elements. [...] And then we kind of have a system of sending*

> *it back and forth probably about three more times with notes." – Participant E*

In response to interview question 5 (p. 131), an equal number of participants 8/16 (= 50%) than not said they would be deterred from using a tool if it required some coding. People who do not program but still said they would not be bothered by the necessity of programming mentioned "ease" and "value" as key conditions to their decision.

In response to interview question 6 (p. 131), Participant D explains:

> *"[We want the audience] to be tuning in because it's really great content, because [the listeners] are really interested in what we're saying, and that we're helping them understand the world in a particular way or giving them a new view or perspective on it." - Participant D*

Concurrently, 11/16 (= 69%) participants mentioned forms of edutainment, 9/16 (= 56%) mentioned connecting with the content, 4/16 (= 25%) said reasons could vary depending on the programme or listener, and 3/16 (= 19%) talk about quality. This seemed to highlight the desire from creators to create unique, valued programs, that can compete with other forms of media. Participant E explicitly mentioned this in their answer:

> *"We're not just competing against audio, right. We're not competing or just looking at the audio landscape in isolation; we are competing with incredibly immersive experiences such as gaming and social media." - Participant E*

5.2.3 How Will You Pod? Implications of Podcast Creators' Perspectives for Designing Innovative Podcasting Tools

Demonstrations

To facilitate the interviews and follow the personalisation subcategories explored in Chapter 3, the demonstrations were divided into two categories: "personalisation of content" and "personalisation of interface". Within "personalisation of content", I presented the concepts of Levels Control (responsive mixing), Server Communication (participatory podcasting), Reverse Engineering Music (unmixing), Non-linear Narratives,(non-linear storytelling) Responsive Spatialisation (responsive mixing), Sound Synthesis, and Voice Synthesis. Within "personalisation of interface": Touch/Tilt Recognition, Gesture Recognition, Metadata, Sound Recognition, Voice Recognition. The contents of these videos are detailed in Table 5.1, and can be accessed in the supplementary material.

Review Of Demonstrations

Personalisation of Interface

The graph in Figure 5.3 represents the interest of participants for different ways to interact with audio content. Participants were asked to rate on a scale 1–5 how interested they were in concepts of: Touch/Tilt Recognition (Median = 3, IQR = 1.25), Gesture Recognition (Median = 2, IQR = 2), Metadata (Median = 5, IQR = 1), Sound Recognition (Median = 4, IQR = 2) and Voice Recognition (Median = 4, IQR = 2). They are represented as a

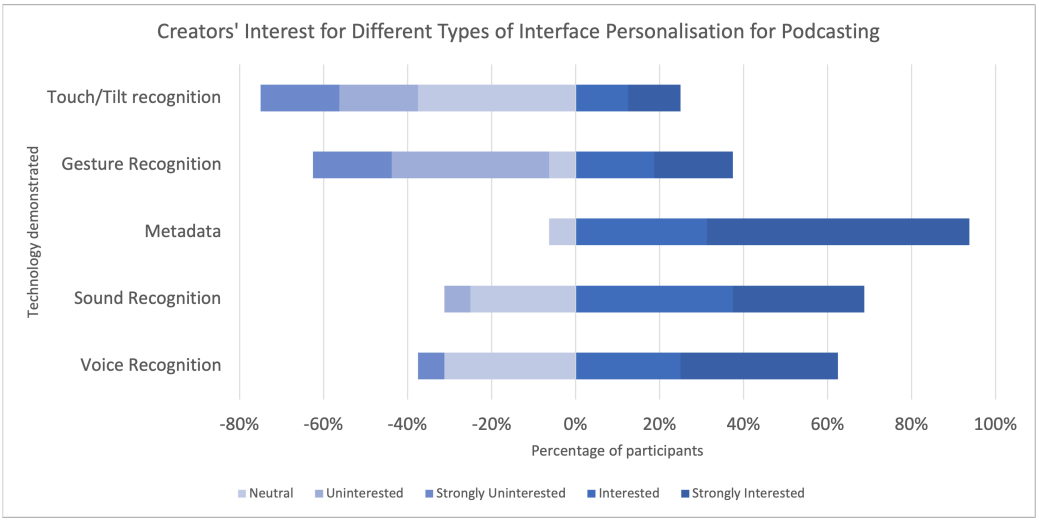


Figure 5.3: Diverging stacked bar chart showing the interest of participants in different interface personalisation technologies for podcasting, as recorded on a 1–5 Likert scale. The percentages correspond to the number of participants having given each answer on the Likert scale.

diverging stacked bar chart, where each bar is divided into stacked segments around a baseline, of length proportional to the percentage of participants having given each rating on the Likert scale.

A Friedman test was conducted to determine whether interest levels differ across the types of interface personalisation. The results showed a significant difference ($\chi^2 = 21.99$, $p < .001$). Post-hoc tests using a Wilcoxon signed-rank test with Bonferroni-adjusted α-level of .05 suggests that Metadata was preferred overall to Touch/Tilt Recognition and Gesture Recognition.

In a follow-up questionnaire, participants were asked to choose their "favourite" demonstration. Out of 11 respondents and 12 demonstrations, Metadata was the preferred answer of 5/11 (= 45%) participants. Although great interest was shown for this concept, a majority of participants were adamant that all data from listeners should be gathered ethically and stored

safely, such as Participant F, who rated the concept a 4 on the Likert scale (4: interested), on the condition the data were collected *"within the bounds of privacy, and not making people feel like they were surveilled"*.

Overall, creators imagined ways to use metadata to personalise the listener's experience on several levels: on a platform level, Participant G noted how understanding and knowing your "niche audience" could help you "find the right audience" and "connect" with more listeners; on a business level, Participant A highlighted the importance of getting more "robust commercial models"; and on a content level, Participant H shared this specific example of how they would use the technology:

> *"Football fans have an extremely high level of interest in the team that they follow and a negligible interest in every single other team. If we know who someone supports – or if not that kind of metadata, then at least where they are in the world – we might be able to infer what the local stories are that are of interest to them. I would use that kind of metadata in my programme now to give people in Manchester stories about Manchester clubs and players." - Participant H*

This is an example of how the technologies demonstrated were envisioned by participants as tools that could help adapt podcasts to a listener's context. Other examples of listener-centric innovations were described by Participant B when they talked of how they would use Voice Recognition in their program: *"You can see that it's 11:00 PM at night, that person sounds like they're tired, maybe you put up a different audio where the host is actually speaking a lot more relaxed and quietly, and maybe the theme song, all the guitars and drums doesn't play"*, thus bridging the concept of adapting to data contained in the listener's voice and other contextual information.

Although the listener experience was prioritised in the phrasing of the

interview questions, creators mentioned ways they see these technologies facilitating their workflows and production processes. For example, Participants A and C saw a clear value in using sound recognition to isolate and delete unwanted sounds from recordings, while Participant I would like to see voice recognition technology evolve to include tone of voice to supplement podcast transcripts.

Personalisation of Content

Figure 5.4 represents the scores participants gave on a Likert scale to different types of audio content personalisation for podcasting: Voice Synthesis (Median = 3, IQR = 2.5), Sound Synthesis (Median = 4, IQR = 1.5), Responsive Spatialisation (Median = 3.5, IQR = 3), Non-linear Narratives (Median = 5, IQR = 1), Reverse-engineering Music (Median = 3.5, IQR = 3), Server Communication (Median = 4, IQR = 2), Levels Control (Median = 4, IQR = 2.25);

A Friedman test was conducted to determine whether interest levels differ across the types of content personalisation. The results showed no significant difference ($\chi^2 = 7.25$, $p = .290$), indicating there was no particular variance within levels of interest for these concepts. 12/16 (= 75%) of respondents rated Non-linear narratives as a "5: strongly interested". When talking about possible applications for the concept, the idea of enabling the user to easily jump around chapters was predominant. Participants J and H expanded on this:

> *"I like this idea of being able to jump around inside of a podcast, especially the way that I do mine: discussion, interview, interview, however many interviews, and then discussion again... I know some people don't give a s— about the discussion, they just want to hear the interviews. So if they were able to jump*

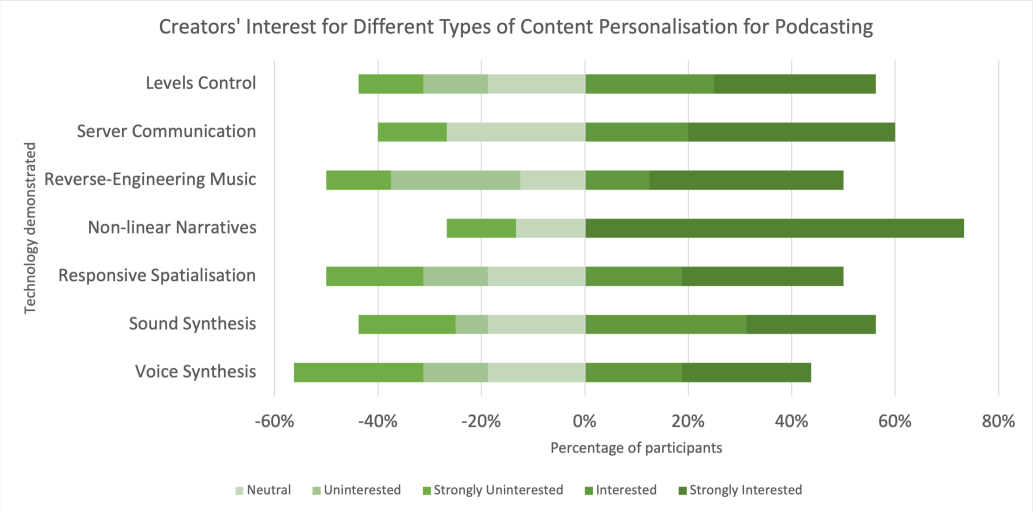


Figure 5.4: Diverging stacked bar chart representing the interest of participants in different content personalisation technologies for podcasting, as recorded on a 1–5 Likert scale. The percentages correspond to the number of participants having given each answer on the Likert scale

> *around that way or tailor the interviews to themselves, or rearrange things based on that... I think that would be kind of a neat thing." - Participants J*

> *"With sports then think having that kind of branching narrative would be really useful to let the audience decide essentially the duration of the content. The podcast that I produce is short form – it's under 10 minutes long, but there's no reason for it to be that way." - Participants H*

On the topic of chapters, participants shared the opinion that the current system for chapter tagging and navigating is impractical and imperfect. Overall, offering podcasts adapting to a listener's environment was a favoured idea, being talked about unprompted by four different participants.

Although the median interest for Voice Synthesis on the Likert scale is 3, 10/16 (= 63%) participants had reservations relating to ethics, and it was rated most uninteresting by participants. Participant K detailed their

concerns surrounding this technology:

> *"This is really concerning technology. I'm very worried about this technology hitting the newsroom. What's going to happen when that first debate sparks about? 'Well, all we needed was an 'S' so I just synthesised an 'S' to make the noun plural' – there's some real, real implications there." - Particpant K*

Six participants expressed concerns surrounding authenticity and quality of Sound Synthesis, and 6/16 (= 38%) participants had overall reservations about Reverse-engineering Music. Sound Synthesis and Reverse-engineering Music were envisioned as production tools by a majority of participants (9/16 (= 56%)). The enhanced accessibility offered by Levels Control was noted by most participants, with Participant L saying:

> *"I had this problem with my son the other day. He was trying to listen to his podcast on a long car journey. He is trying to listen to speech content and he can't hear any dialogue because the hum of the car is too much. And you know, sometimes we might try and mix a podcast [taking these situations into account], to make sure that you can get that clarity, but not always. So an adaptive feature that makes listening more accessible is interesting." - Participant L*

Participant L underlined the importance of not falling into over-personalisation – or baseless personalisation, saying:

> *"It has to have a payoff... The producers have to pay the audience back for that engagement rather than just be about contact for the sake of it." - Participant L*

The reasoning behind personalisation was something considered by most of the interviewees. Although it was widely accepted that personalisation plays a part in how we make and consume podcasts in the near future,

creators were aware of the caveats of producing personalised content for the sake of personalisation.

5.2.4 Discussion

How podcast creators envision innovation and then explore and produce with innovative techniques are matters that will affect the experience of millions of listeners worldwide. This study indicated how innovation might be incorporated into production workflows, and formulates some requirements and expectations of tools built to create new forms of audio-centric programming. In this final section, I interpret the results of the qualitative and quantitative analyses presented, draw some conclusions, and finish by identifying some limitations and ideas for future work, laying down the foundations necessary to make advances in the world of podcasting, particularly in terms of production tools and listener experience.

What Podcast Creators Envision as "Next-Generation Podcasting"?

Creators interviewed in this study expressed two separate goals that appear to contradict each other. The first was to improve listener experience, through a combination of new formats, higher quality audio, or finding ways for content to be more engaging for audiences; the second was to simplify and streamline their production process, by using faster, smarter, more efficient tools. However, practices that could simplify the creator's work, like synthesising voices, sound effects, or un-mixing music to separate tracks, could have the adverse effect of worsening the listener experience overall. Conversely,

adding features to podcasts in order to improve listener experience could greatly complicate an already-convoluted workflow.

And thus, this duality in expectations, that matches the duality of approaches in audiovisual tool design highlighted in Chapter 3 and 4, is brought to light. Participants agreed that next-generation podcasting should involve a form of improvement of the listener experience (a "listener-centric" vision of next-generation podcasting), but they also showed an interest in using the technologies presented as tools to simplify production and reduce their workload (a "creator-centric" reaction to the demonstrations). Aligning these two approaches to podcast innovations could be paramount to improving both the listener's and creator's experience of podcasting. For all interviewees but one cautious independent producer, this combined improvement is synonymous with "next-generation" podcasting. Regardless of its application (listener or creator-centric), purpose-driven innovation prevailed in participants' reasoning, with the aim of easily producing better quality, more entertaining, informative, engaging and immersive content at the centre of "next-generation podcasting".

The nuance this qualitative analysis brings to the quantitative results presented helps us decipher participants' answers and bring into focus technologies that appear plausible candidates for "next-generation podcasting", while discarding more problematic ideas. For instance, the idea of motion-based recognition can be discarded, as both Touch/Tilt and Gesture Recognition raises concerns over accessibility and disability, and generally goes against the idea of podcasts being a "hands-free" medium. Reverse-engineering music to separate stems and synthesising sound effects or voices could very well

be an asset for creators, but participants interviewed express reservations concerning ethics, authenticity, and quality, which could potentially hinder the listener experience more than the creator's process would benefit from the implementation of these tools.

It is also important to remark that the technologies presented in the video demonstrations evoked similar ideas in participants. Four creators were interested in creating podcasts that adapt to the listener's environment; three were keen to explore location-based personalisation; two wanted to create podcasts that vary depending on the listener's time of day. Participant K said:

> *"I am really looking forward to the day when a mobile device can respond to the environment that the listener was in, and automatically change the dynamics of the content, or change the loudness of the content presented to the listeners, so if you're on a subway and it's very loud, it will decrease the dynamic range of the content and perhaps turn it up just a little bit for you to make it easier to listen to." - Participant K*

These notions of modularity, and adaptivity to attributes on the user side are consistent with the concept of "perceptive media" (media that perceives one's actions and then adapts to them), as coined by Ian Forrester, and his goal to create podcasts that adapt to the listeners Dwornik (2021).

Across the board, participants stressed that they do not want to overwhelm the listeners with decisions, like Participant D who explained: *"Trying to get listeners to interact or do anything... It's non-existent."* Any interactivity should therefore work in a non-intrusive fashion, hand in hand with immersion, as a means to achieve it rather than as a distraction from it, and have a clear purpose.

What Tools Do Podcast Creators Use and Why?

Podcast creators value easy-to-use, highly compatible, "no-code" software. Due to the lack of standardisation within podcast production practices, both independent and BBC-affiliated creators use a variety of tools to record, edit, and distribute their podcasts. But, within this multitude, the corollaries of what is usually a collaborative process prevail, with creators favouring highly compatible, simple-to-use tools. Via the question, "What tools do podcast creators use and why?", the habits and expectations of practitioners pertaining to their software and equipment are uncovered, which could give us insights into the requirements a podcasting tool should aim to fulfil.

According to answers to interview question 3 (*What attributes make for good podcasting tools?*), a podcasting tool should be efficient, compatible, useful, comfortable, and good value for money (in order of importance, from most important to least important to the group of participants). This should be read in the context of participants' current practices. For instance, the six BBC creators agreed that their choice of software was influenced by the habits of people they worked with, yet, they mention using four different DAWs (Question 2: *What tools are necessary for your work?*). Although compatibility seems high on their list of priorities, personal preferences and background appear to play a bigger role in their choice of editing software, which speaks to the conflicting expectations of seeking universality, but lacking conformity.

This lack of conformity – but need for universality – means any new podcasting tool should aim to offer widespread support across different work tools. The need for simplicity and lack of coding expertise from the partici-

pants (question 5: *Do you have any experience with coding? And would the need for coding deter you from using a tool?*) informs us that any podcasting software should be very easy to use and not require any programming skills.

Understanding the desired functionalities and attributes of a new media tool is fundamental to its development (Ward et al., 2020), and, by studying the requirements and expectations of podcast producers, a foundation on which innovative podcasting tools could be built is presented.

How Would New Tools and Habits Be Integrated to Podcasters' Established Production Workflows?

Integration of innovation will come in pre- or post-production phases. Podcast production is a complicated process, which, for the sake of producers, should be simplified rather than complexified further. Some apps such as Spotify for Podcasters[2] take this approach of drastically simplifying the podcast production process, with all the steps required for basic podcast production (Figure 5.2) contained within one single web app. But, if the purpose of these new tools is to add features or improve substantially on existing ones, it can be expected that a minimal modification to the archetypal workflow presented would need to occur.

I asked about the specifics of each participant's workflow (interview question 4: *Do you have a particular workflow when creating a podcast?*). A remarkable finding of these interviews and subsequent thematic analysis is the high level of consistency in the production workflow described by podcast creators, and distilled in the archetypal workflow shown in Figure 5.2. As well as the assembling of the workflow itself being a contribution of this chapter,

[2] https://podcasters.spotify.com/

the consistency with which it is used constitutes an important finding: the analysis indicates that the archetypal workflow does not vary considerably with genre; nor does it vary much between independent creators and those working at the BBC. Rather, the small differences in production processes comes from the specific needs of particular projects. These differences do not much alter the overall appearance of the archetypal workflow. The most common variations entail a skipped or added node, or one creator iterating through a loop of existing nodes more often than another.

Indeed, podcast production was found to be a highly iterative process, and that therefore, we should respect the loops already in place (like writing $\leftrightarrow$ recording $\leftrightarrow$ editing), or take precautions to preserve them, but also not shy away from introducing another step that a creator could loop into their existing workflow. This analysis suggests a new step could be embedded as part of the pre-production phase, before or in tandem with booking, or in the post-production phase, after editing but before distribution.

These iterations (which most often constitute revisions) can be triggered by a formal process, like a supervisor or editorial board requesting changes to the latest version of the podcast, or take place organically throughout the project. The "constructive loop" (editing $\rightarrow$ revisions $\rightarrow$ recording and sound design $\rightarrow$ editing etc.) that is identified via the interview analysis underlines the importance of iteration in podcast creation and production, and contrasts with rather more linear post-production processes in other media. While some of the creators I talked to worked alone, the importance of collaboration was made apparent in all conversations, with the role of "advisor" being mentioned on several occasions as a key personality in the

podcast production process.

The preliminary analysis of the participants' occupations revealed the multifaceted nature of the work of individuals in the podcasting industry. Although predominant when asking the participants to describe their role, "producer" is often too reductive to encompass the variety of tasks required to make a podcast. Instead, the "creator" role more accurately represents the contributions of these individuals, further justifying the interchangeable nature of these terms throughout this thesis.

Many independent creators reported thinking of publication from the beginning of their production process, with prepared images and social media templates to help promote their podcasts, and BBC creators enjoyed the help of dedicated teams/team members working on publishing, marketing and promoting their releases. The publication phase appeared more fundamental to the podcast production workflow than it is in radio. I postulate that this is because, unlike radio shows, podcasts have to stand out to be consumed. Choosing a podcast is an active decision, where a listener has to browse through a library or catalogue to pick a programme.

A Reflection on Accessibility

Prince (2020) acknowledges that podcasts are "unusually accessible", referencing ease of use, low cost, and the flexibility that transcriptions offer to deaf or hard-of-hearing listeners. However, this last feature relies on the assumption that most podcasts would use transcripts, and that those would be of good quality. Seven participants discussed accessibility, just under half of the total number of creators interviewed. It seemed widely agreed upon that pod-

casts are not the most accessible in their current form, often lacking proper transcripts or simplified/audio-described interfaces. Often, transcripts for podcasts are not available, and a complicated feature for creators to include in their programmes. Although some tools already exist that facilitate this process (AI transcription tools, distribution platforms that specifically query for transcripts, etc.), these solutions often come at a cost for the creators. The multitude of distribution options and hosts, each with their own upload platforms and requirements, makes it harder for creators to expect and rely on the same accessibility features from one project to the other. This lack of consistency might in turn discourage some potential listeners. There was a clear reflection on these accessibility shortcomings in the medium as a whole by the aforementioned participants.

Limitations

The exploratory nature of this study required choices to be made in preparation for the interviews. For instance, although Chapter 3 justifies the inclusion of the 12 demonstration videos presented, they do not represent an exhaustive list of technologies that could be used for next-generation podcasting, but rather a selection of technologies that could be implemented within a time frame appropriate to my overall research project and aims. The demonstrations presented may therefore be perceived as a subjective collection of potential technologies, with their inclusion (and the exclusion of others) justified by the aim of this research.

Overall, the average interest in the technologies demonstrated is itself above average. This could be explained by participant self-selection – the

recruitment process may have appealed to people who were particularly passionate about the application of new technology to audio and related media. Overall, the content and personalisation categories are rated as interesting as one another, with a median interest in these two groups of technologies of 4 on a 1–5 Likert scale.

Potential bias that some creators may have had due to prior familiarity with certain technologies also needs addressing. This might lead them to have a more favourable impression of the technologies of which they were already aware of, and in turn skew the data towards these concepts, like non-linear narratives, where all participants were familiar with various existing incarnations. It is unclear whether the high interest registered for this concept was due to a general, mainstream knowledge of the technology compared to other demos, or to a real preference.

The creators interviewed, although representing a variety of genres and principal occupation, were not as diverse as one might hope for a study aiming to "generalise" a concept. Participants are based in the United Kingdom, United States, Canada and Columbia, with a good knowledge of English (the language in which the research is conducted). This excludes a significant portion of the international podcast industry, however, where production workflows may be different.

Data regarding the size of the teams in which the creators worked were not systematically collected, so conclusions regarding the impact of team size of affiliation cannot be made.

The comparison drawn between independent and mainstream broadcasting companies is made on the basis that the BBC is an accurate representa-

tion of the latter. However, there is a possibility that other large broadcasters do not share similar production patterns. Therefore, it would be valuable to confirm this framework with creators affiliated with other production companies and networks (e.g. other broadcasting companies, and online platform in-house production teams, like Gimlet for Spotify, or Futuro Studio for Apple Podcasts), as well as with freelancers.

This study focused solely on exploring the creators' current production habits and outlook on the future of podcasting, in order to bridge a gap noted in the literature regarding the role and expectations of the professional podcaster. Other actors' points of view, like those of the listeners, advertisers, or platforms, could be explored to better contextualise the research presented in this study.

5.2.5 Takeaways

The study presented in the previous section delves into the intricacies of podcast production and the concept of NGP. It explores the current practices of podcast producers, revealing their archetypal production workflows and habits, and postulates that these preferences could form the basis for podcasting innovation and research in the future. It investigates the perspectives of independent and mainstream creators on next-generation podcasting, bringing to light their expectations for tools that enable better listener experiences, but also tools that facilitate their work, and their view of how a selection of new technologies could be leveraged within their production process.

The amalgamation of these findings allows us to hypothesise how a new

podcasting tool could be implemented within existing production habits, through seven key takeaways: Podcast creators -

- are interested in delivering better, more immersive and engaging experiences to their listeners.
- have an already-complex workflow comprised of a wide range of tasks and skills.
- are looking for ways to simplify this complex production process.
- want their production tools to be efficient, compatible, useful, comfortable, good value for money, and no-code.
- are looking for ways to adapt their podcasts to their listeners.
- are concerned with accessibility and reaching as wide an audience as possible.
- are wary of unethical uses of AI in media.

These points enable us to refine the list of technologies and ideas explored within this research – indeed, many of the concepts described in Chapter 3 cannot coexist with the requirements listed above. For instance, although voice or sound synthesis would facilitate some podcasters' workflows, it is a highly contentious technology in the field, ethically. More broadly, technologies that gathered less interest, like participatory podcasting, or motion recognition-based interfaces can also be discarded. Following this meticulous pruning, a smaller series of "promising" technologies for NGP remains: responsive mixing; non-linear storytelling; implicit interactions via metadata;

and intuitive audio/visual interactions. Combined with some key personalisation concepts that came out of the interviews, the rest of this thesis explores how podcasts could adapt to listener's preferences, listening environment or context, and choices – as this was a recurrent comment amongst producers when thinking of NGP.

5.3 Creator Workshops To Lead and Refine Design Decisions

This refined set of ideas was presented to a smaller set of four creators in a workshop - to put together ideas for practical applications and further direct the tool development at the centre of this thesis. This chapter section details the planning, content, delivery and outcomes of the workshops and conversations held to fulfil this aim.

5.3.1 Workshop Design

The goal of these workshops was to gather the thoughts of podcast creators on the NGP they would like to produce. I was interested in finding out what they would create using the technologies and concepts highlighted in the exploratory study discussed, and what they would need to make these projects happen. This workshop initiated a creative collaboration with the participants, and to start designing one or more tools that would enable next-generation podcast production. The same participants were invited to other workshops or discussions to support the development of the tool and start concretising accompanying podcast projects. Table 5.5 was provided

to the two facilitators (BBC R&D Audio Team members) to describe the particulars of the workshops. I would also act as a facilitator, although more so to explain concepts than to lead activities. The workshop was conducted on Zoom, and participants' work was collected via collaborative Jamboards [3].

Important outcomes were highlighted for facilitators so they know what to prioritise throughout and can respond to changes on the day with agility. Two extra participants canceled on the day because of unforeseen circumstances.

Activities Detail

Set up (10 minutes) - Making sure the session is prepared and the facilitators have no questions.

Overview presentation (10 minutes) - Introduction to the project and expectations for the day.

Warm up exercise : 3 words pitch (10 minutes) - Each participant generates three words with a random word generator[4] and has 5 minutes for writing a short story/pitch of a few sentences maximum. They are then encouraged to read out their pitches, or to post the story in the chat if they are more comfortable with one of the facilitators reading their work. These pitches are not recorded to ensure participants are comfortable an do not feel pressured when engaging in creative activities.

Dot vote (10 minutes) - Participants are asked to vote for a maximum of three of their preferred modes of adaptation (Time or place of listening, Preferences and/or listening habits, Listener's choices, Listener's acoustic

[3]Soon to be obsolete: https://jamboard.google.com/
[4]https://randomwordgenerator.com/

Table 5.5: Overview of the workshop provided to facilitators.

Who (Participants)	• 4 participants • Podcast creators from the UK and US • BBC and independent creators, working on a variety of genres of podcasts.
Motivation, Research Questions, Deliverables	• Validate previous research and focus on NGP • Initiate a creative collaboration between the researcher and the creators • How might concepts and technologies that seem conducive to more immersive/personalised podcasts manifest in real/practical applications? • Gather requirements for a possible tool design: What do creators need to make their vision come to be? What problems do they anticipate?
Who (Team)	• Lead (Jemily) • Two BBC R&D Audio team members to facilitate in groups and handle logistics
When and where (Team)	• May 2021 • Online (Zoom) • 2h30 (two 10 minutes breaks at the 40 minutes and 1h30 marks)

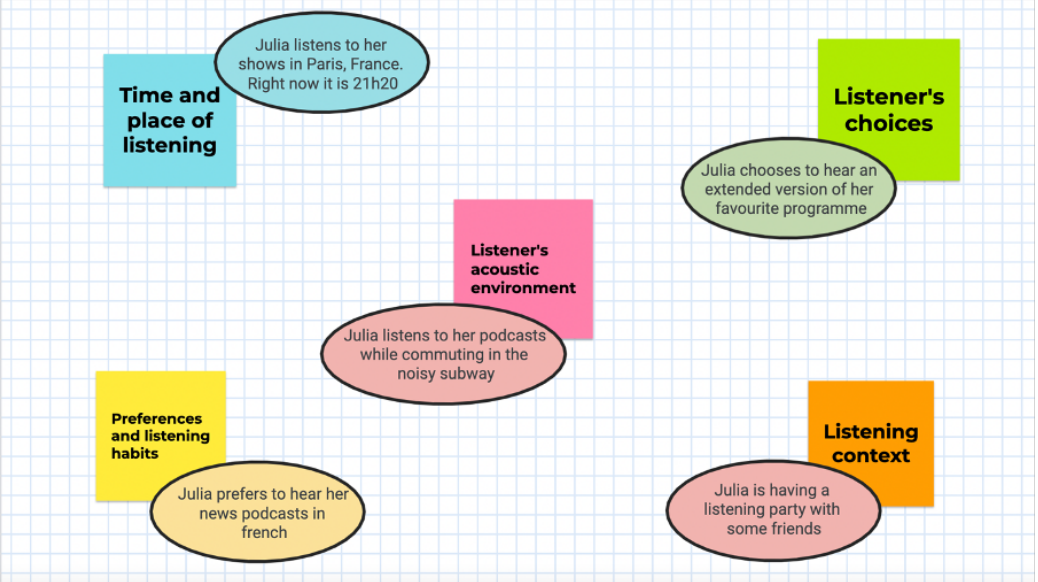


Figure 5.5: Empty dot vote Jamboard page. Each post-it represents a concept participants can register interest in through voting. Alongside is an oval that includes an example application of the concept.

environment, Listening context (social)) on a shared digital board, which can be seen in Figure 5.5.

Post Up ideation exercise (10 minutes) - On a personal Jamboard, participants write on digital post-it notes ideas for possible outcomes of podcasts adapted to their chosen concept. This includes a few minutes to share ideas.

Break (10 minutes)

Mutation game (20 minutes) - Participants choose one "effect" and highlight the ideas of their Post-up board that would still work with this mutation. Their ideas are discussed as a group. The effects are as follows: Modifying one or more tracks; Offering modular (i.e. different, or à la carte) version of the same programme, Changing the stereo positions of elements; Adding an extra feature (purposely broad to allow for participants to bring

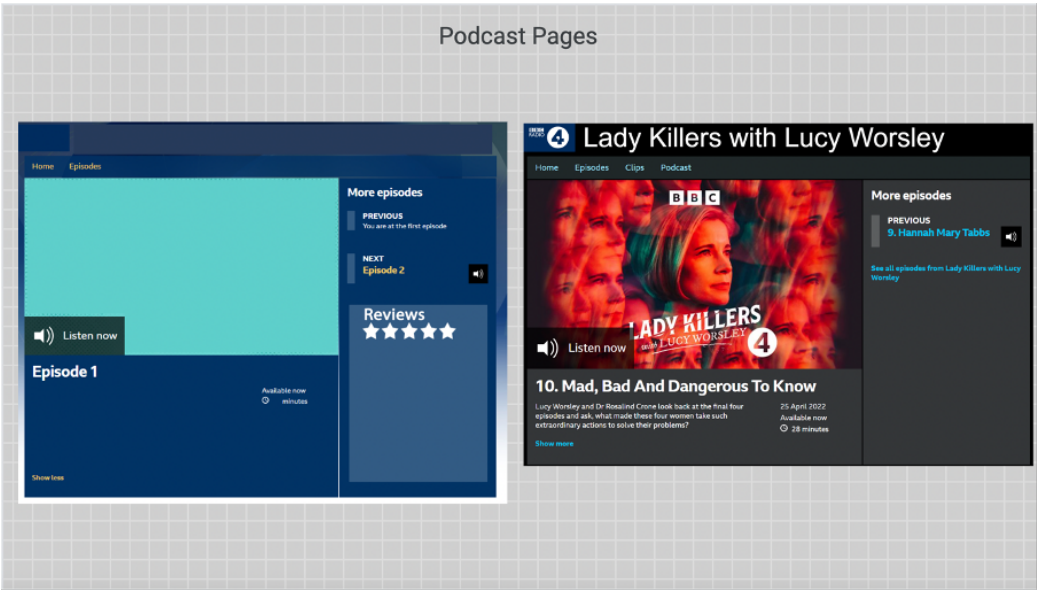


Figure 5.6: Template for creating an artefact podcast page, with an empty one on the left, and an example one from the BBC show "Lady Killers" on the right

in highly specific ideas). The participants are given an opportunity to add ideas to their post-up board and expand on concepts they like.

Podcast pages (30 minutes) – Using a template (Figure 5.6), participants create the description page of a podcast (Title, image, description, number of episodes, review from listener etc.) based on one of their ideas and present their podcast page to each other. Creating digital artefacts of their project enables them to further expand on their initial project idea.

Break (10 minutes)

Speedboats (15 minutes) - on a personal template (Figure 5.7), participants use the anchor/icebergs/sun metaphors to identify problems and goals of podcast project ideas. They are asked to think, focusing on one item at a time, of goals/outcomes (sun), immediate issues or things preventing them from getting the project done (anchors), and potential issues (iceberg).

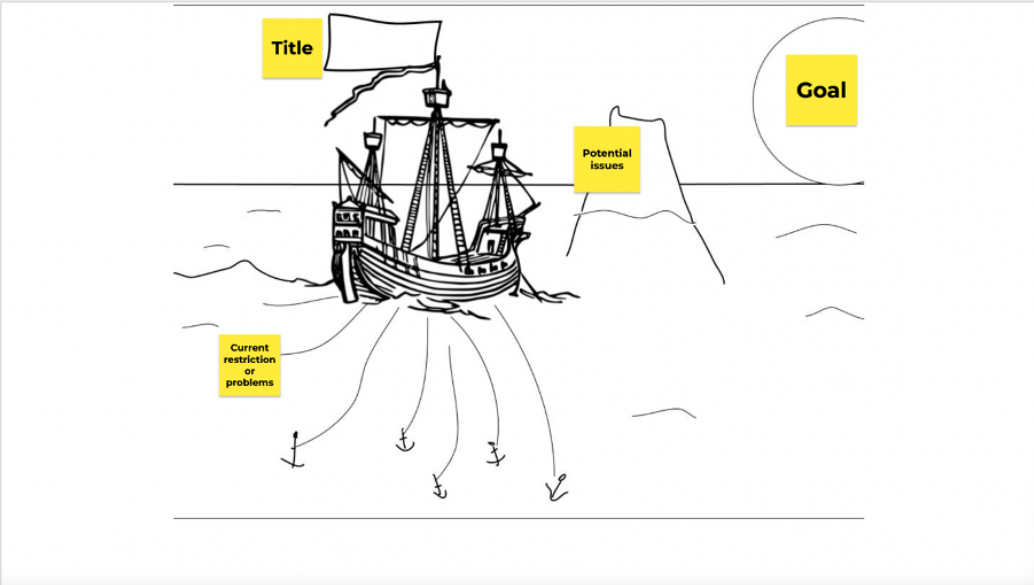


Figure 5.7: Empty speedboat Jamboard page

Discussion (15 minutes) – A review of the work produced in the prior Speedboat exercise, focused on the following questions: 1/What problems were common across different projects? 2/ Are there prevalent issues for participants having chosen the same "mutation" or mode to adapt to? 3/How to lift a few "anchors" or avoid some "icebergs"?

Close up remarks/wrap up (10 minutes)

5.3.2 Outcomes

Figure 5.8 displays some of the results of the dot voting, post-up, and mutation game. Dot voting is a collaborative activity, each small white shape (square, circle, diamond, and triangle) represents the vote of a participant. Interest was spread out across the various concepts, but adapting to the listeners' choices was picked by all participants. A participant came in with prepared project ideas, so their thought process was slightly different when

voting as they are trying to fit the concepts to their idea rather than the opposite.

Participant C's post-up page (Figure 5.8) mentioned various ideas that fit with the idea of adapting to a listener's choices – these were then refined and reorganised during the mutation game, according to those that would fit the concept of "modular podcasting". In total, 3 participants chose "modular versions", one chose "modifying the content of a track", and one chose "adding a button" (specifically, an extra button to unlock some new features, like accessibility or social features). A participant chose two mutations ("new feature" and "modular versions").

Participant C chose to focus specifically on the idea of a modular book club, *The Armchair Book Club*, a podcast that could adapt to a listener's preferences. The artefact created during the podcast page exercise can be seen in figure 5.9. It features variable length (duration), and a made-up review highlighting the value of customisable podcasts. Other podcast ideas that were created at this stage are *The infinite wormhole of wonder* ("A comedian takes us on a journey of nerdy delight, which sucks you in with custard creams and spits you out with asexual reproduction." – an infinite podcast where the listener can move from one topic to the next); *The Jigsaw* ("Three versions of the same story but you can only choose one - which witness will you choose? and what are they hiding? once you choose one - the others will disappear - but are they telling the truth? - only you can decide. And you will need to find other listeners to find out what really happened." – a multiple version crime story); and *Philadelphia Museum of Art Podcast Tour* ("Tune in as you walk into the historic Philadelphia

Museum of Art! Learn where to buy tickets, find maps, and check out our latest displays! Episode 2 will appear in your feed as you exit the atrium." – a museum podcast).

Finally, the Speedboats exercise reflected on current and potential issues this project could face, including the need for highly specific technology, distribution issues due to lack of standardisation, and possible problems with advertising. During the discussion, the technologies required to make different versions of the same programme are discussed, and the idea of using AI to automatically make segments to be rearranged was discussed by two participants. Participant D, whose idea was the multiple versions whodunnit, commented on the idea that offering different versions to different audiences might create an impression of scarcity, which might annoy or intrigue the listeners :

> *"Annoyed listeners, intrigued listeners... it's all good publicity in the end." - Participant D*

Overall, most of the ideas discussed could be facilitated by a well-integrated tool that automatically segments a podcast and offers modular flows for different listener experiences that could adapt to implicit or explicit decisions by the user. From then on, I refer to these different-versions-podcasts as **modular**. The question of accessibility is raised, with issues regarding captions when preparing such programmes.

5.4 Summary

Through these conversations with podcast creators, the initial investigation of the literature presented was refined and provided partial answers to the

research questions listed in Chapter 1. These partial answers are as follows:

RQ 1: What is NGP?

Earlier, in Chapter 2, I described NGP as a conglomerate of ideas for the future of podcasting: "NGP stands for all the innovative podcasting techniques and ideas that are yet to take off"

I also highlighted the need to gather the opinions of producers to be able to understand what this term means across the industry. In the first study presented in this Chapter, NGP appears to be combining two facets of innovation, reflecting the two ways that Chapter 4 determines innovative podcast production tools could be developed: listener-centric (improving the listener experience), or creator-centric (simplifying or streamlining their production process). Purpose-driven innovation prevailed throughout the interviews, with the main goal of NGP being to easily produce better quality, more entertaining, informative, engaging and immersive content. The idea of adaptivity and modularity were discussed by a large portion of participants.

This is further confirmed by this first workshop, which saw all four participants interested in adapting to a listener's choices, and three participants further exploring the concept of "modular versions of similar programmes".

RQ 2: How can we feasibly create and distribute immersive and personalised podcasts?

The interviews filled some of the gaps mentioned in Chapter 2 and 3, formalising the podcast production (Figure 5.2) process and the roles and habits of podcasters. Through this highly iterative, complex, media production process, an NGP tool could be integrated within the pre-production phase, before or in tandem with booking, or in the post-production phase,

after editing but before distribution. It was also highlighted that producers want their production tools to be efficient, compatible, useful, comfortable, good value for money, and no-code.

RQ 3: Why is modular podcasting attractive to creators, and how can it be facilitated without making the production process more complex?

The term modular came out of the interviews, specifically from the conversation points surrounding adaptivity and the concrete NGP project ideas brainstormed by the participants. This term is further refined during the workshop, with its initial appeal seemingly being linked to the high level of customisation it would enable. Through this customisation, it could make projects more appealing, allow for new creative outputs (e.g. the idea of a multiple point-of-view murder mystery of Participant D), create social bonds (comparative or collaborative listening), or even drive engagement (by curating "perfect" versions of programmes).

RQ 4. What are the benefits, risks, and costs of exploiting AI technologies for podcast production?

Podcasters seemed wary about using AI in some cases but nonetheless show interest in the technologies provided they provide a clear benefit to their workflows. Their main concerns were related to generative AI specifically, but also models that could hinder the quality of the product they release.

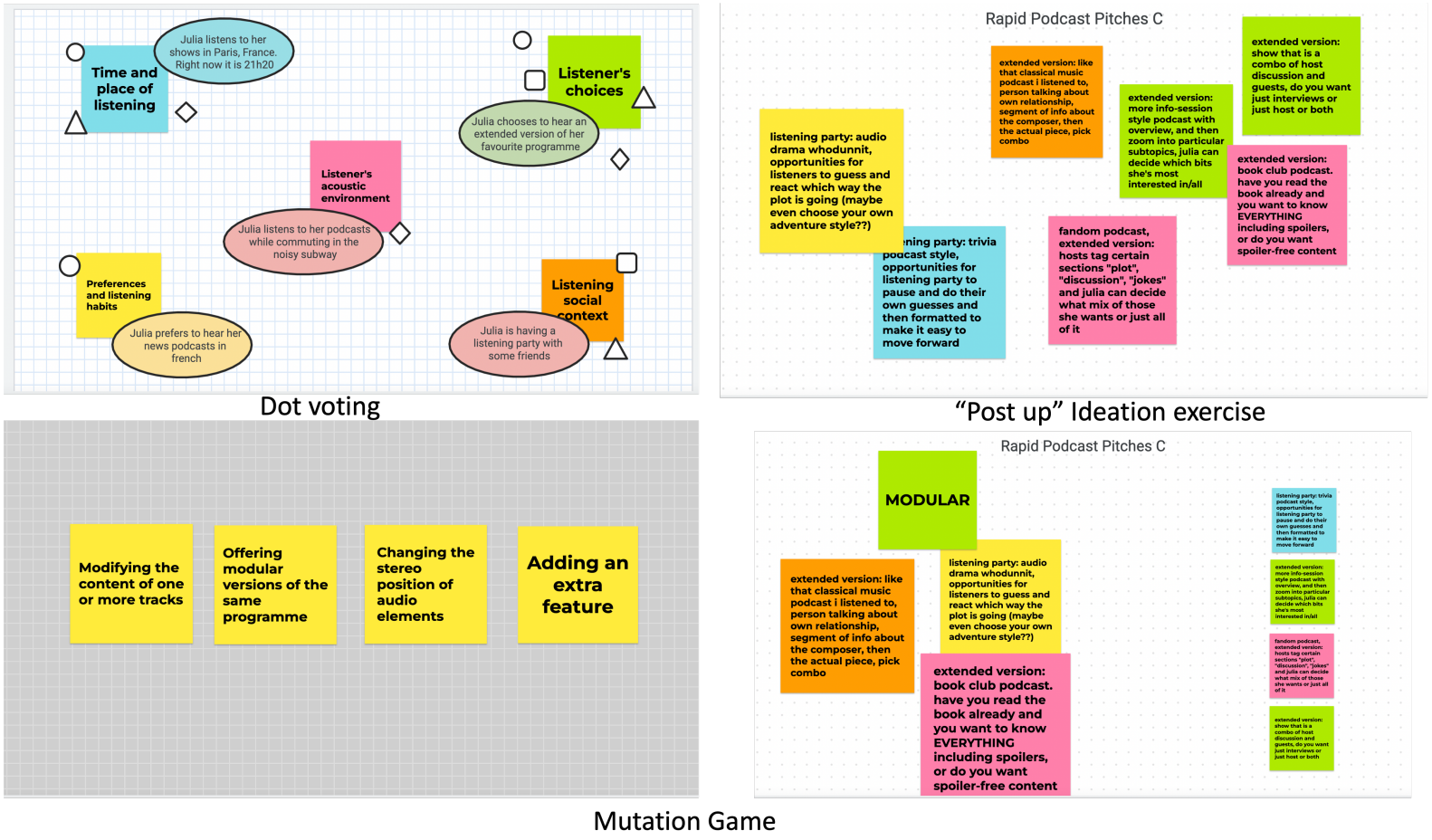


Figure 5.8: Results of the dot voting (top left), initial post up from participant C (top right), effects and updated post up board from mutation game (bottom right and left)

Figure 5.9: Podcast page created by Participant C for the *The Armchair Book Club*

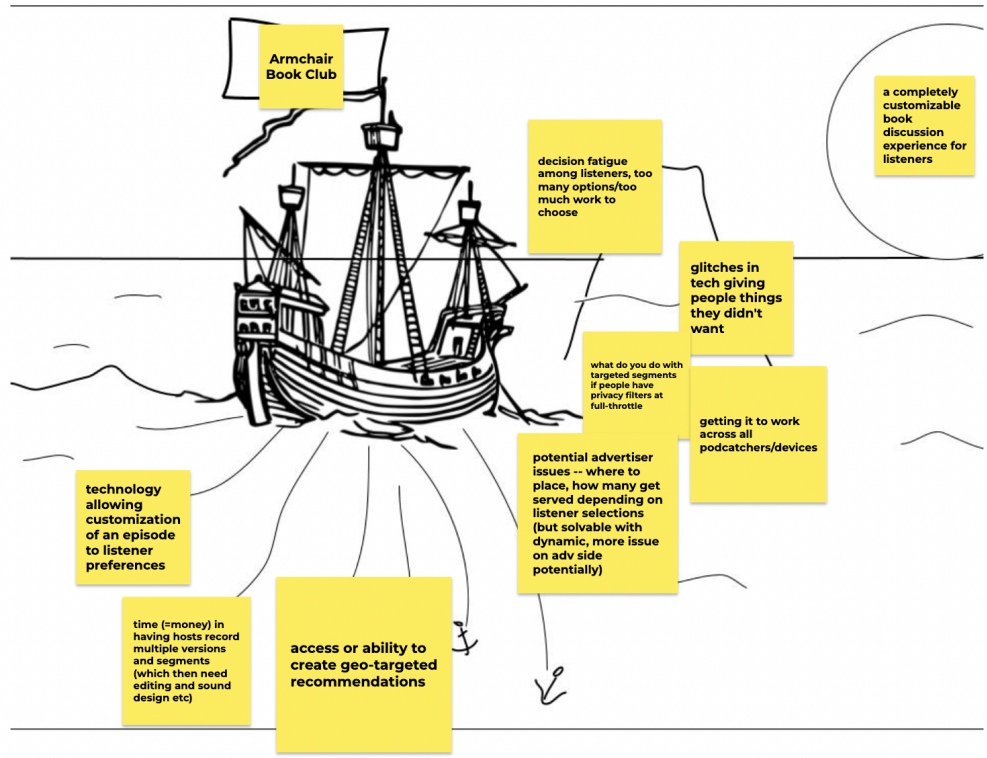


Figure 5.10: Participant C's Speedboat exercise for the *The Armchair Book Club*

6 Insights on Chapters and Modular Podcasting

6.1 Introduction

In this Chapter, the first steps in developing a modular podcasting tool are described. A modular podcasting app concept, Podulr, is introduced and explored with seven producers through 30-minute Zoom semi-structured interviews or equivalent online surveys. The opinions gathered are reported upon, so they can be taken into account when building further iterations of this proposed product (as will be described in Chapter 8). The interviews also contribute to a partial answer to *RQ 2: How can we feasibly create and distribute immersive and personalised podcasts?*

Following these interactions, and with the goal in mind to create a benchmark of chapterised podcasts, 10 expert producers are invited to annotate a dataset of 49 5-minute excerpts of podcasts from the Apple Podcast Top 40 (UK) with chapter markers in order to determine whether podcasters can agree upon the practical definition of "chapters". The resulting dataset, POD 49, shows that producers "moderately agree" on what a chapter is, and this, alongside prior interview answers regarding the usefulness of chapterisation and creative applications of modular podcasting help answer *RQ3: What is the appeal of modular podcasting, and how can it be facilitated with-*

out complexifying the production process?

After describing Podulr, the interview process through which creators' points of view are gathered is described, before reporting on a study aiming at creating a dataset of podcasts annotated with chapter markers.

6.2 Podulr, the Modular Podcasting App

6.2.1 Motivations

From prior conversations reported in Chapter 5 and an investigation into available technologies, as presented in Chapter 3, I proposed an app that would facilitate modular podcasting using automatic sound-based segmentation. This tool leverages sound recognition and allows creators to easily produce several versions of a single podcast, to offer listeners personalised experiences. As seen in Chapter 3, the available systems developed by partnered groups (StoryKit and CuttingRoom), would not provide the required framework to develop such a tool – therefore, I chose to design a web app that could contain all the necessary features and requirements outlined in Chapter 5.

An initial investigation into sound recognition systems was carried out to validate the feasibility of the project. In 1999, Martin details the way that humans recognise sounds and translate this process into a computational one. This work focuses primarily on recognising instruments from audio data, but applications of sound recognition go beyond music. Moreover, Martin mentions "media annotation" as the first in their list of potential applications of sound-source recognition. Sharan and Moir (2016) gives an overview of

the evolution of automatic sound recognition (ASR) since Martin's work. They detail the three key steps in ASR systems: signal preprocessing (the sound signal is divided into smaller chunks to be analysed), feature extraction (features are inferred from the signals so the input signal is represented by a vector) and classification (based on training data, these representations are assigned to one of the learned classes). A signal can be analysed in both time and frequency domains (Chachada and Kuo, 2014), where a temporal analysis can lead to "*exact measurable representation of signal*" (p.4), whereas frequency domain methods can be used to describe "*the nature of the physical phenomenon constituting the signal*" (p.4).

Sound Event Detection (SED) can be used for detecting audio events (Mesaros et al., 2021). This can be done using machine learning, more specifically, supervised learning using a training dataset of annotated sounds to form an acoustic model. These annotations can be strong labels (contain temporal information about the events) or weak labels (only registering active/inactive sound class) (Morfi and Stowell, 2018). Gaussian Mixture Models (GMMs) (Mohanapriya et al., 2014; Elizalde et al.) Hidden Markov Models (HMMs) (Chandrakala and Jayalakshmi, 2019; Ozkan and Barkana, 2019) and Multiple Instance Learning (MIL) (Kumar and Raj, 2016b) have been used in the past, but recently DNNs (Çakır et al., 2017; Zhang et al., 2017; Adavanne et al., 2017; Kong et al., 2019a) have led to state-of-the art results in terms of enabling multilabel classification (Mesaros et al., 2021).

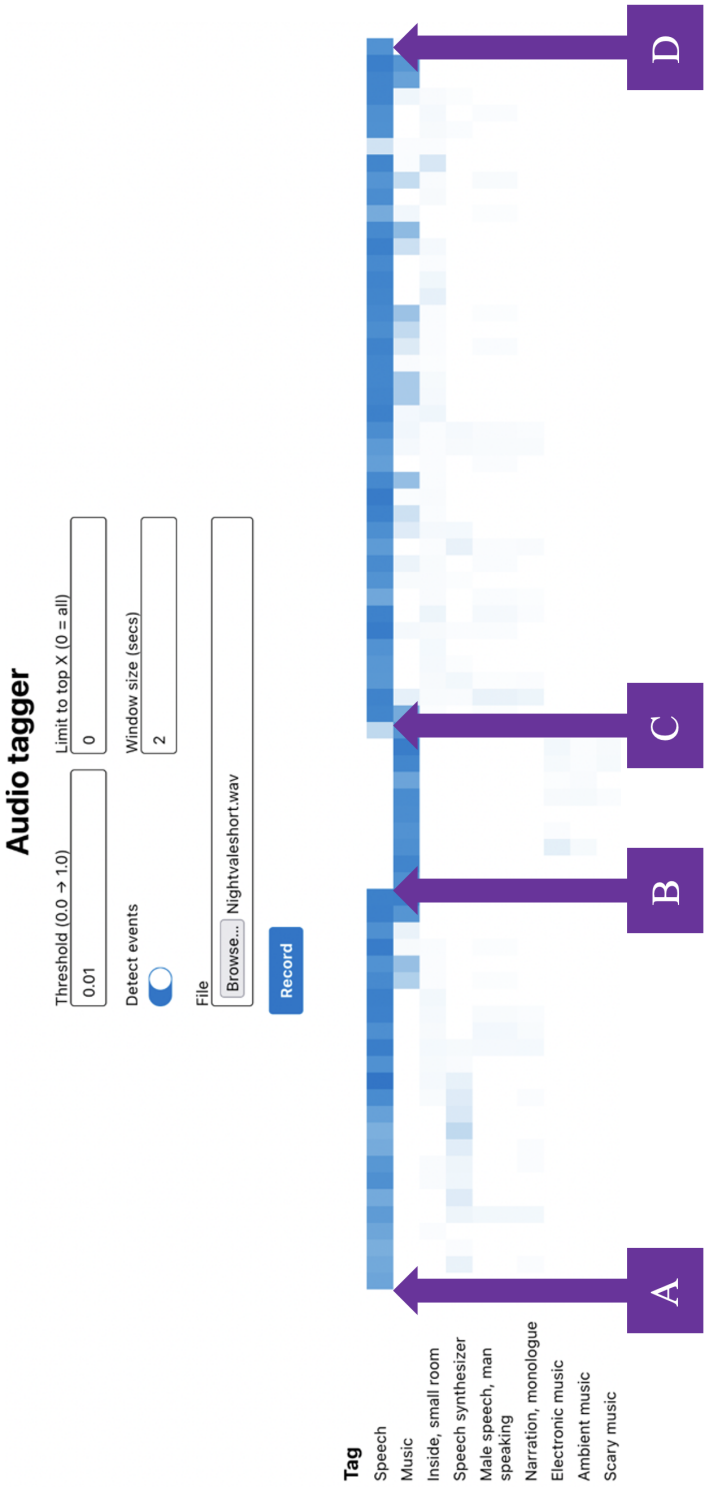


Figure 6.1: Screen Capture of C.Baume's implementation of the SED model described by Kong et al. (2020)

As evidenced by the references provided, there is a wide variety of models for SED. This PhD took the direction of expanding on the work of Kong et al. (2020), as the training datasets (Gemmeke et al., 2017) were transparent, the performance was state-of-the-art and the work was replicable.

Figure 6.1 is a screenshot of a GUI developed by Chris Baume from BBC R&D implementing the SED model described by Kong et al. (2020). Baume put together an end-point to demonstrate how this particular model worked for a research presentation. I used his interface as an easy way to communicate the nature of sound recognition and how it was envisioned to work within a segmentation algorithm. It was particularly helpful in meetings with stakeholders, but also when presenting ideas to producers during workshops or interviews. It was also used as a way to test whether the algorithm's pipeline could be successful (PANN => rule-based system => candidate chapter boundaries). By the time the algorithm was integrated within Podulr, Baume's interface was entirely replaced by bespoke back-end scripts, that triggered and formatted the output of Kong et al. (2020)'s model.

Each blue rectangle is a frame of 2 seconds, and the tags on the left-hand side are the possible classifications for the sonic elements comprised within that frame – the deeper the blue of the rectangle, the more likely its associated tag is to occur within that frame. The example used is a short extract of "*Welcome to Nightvale*", featuring an introduction (A-B), jingle (B-C), and a monologue (C-D). This figure by itself showcases the potential applications for sound-based segmentation: there is a clear distinction between the fingerprint of the music (B-C) and introduction (A-B), and even between the

introduction (A-B) and monologue (C-D). This model could be integrated within a system that analyses the fingerprint of audio on a frame-by-frame basis and compares each with its neighbours to infer possible boundaries. The details of this system will be provided in Chapter 7.

6.2.2 User Interface

For producers to be able to interact with this system, a GUI needed to be created. Figure 6.2 is a flowchart that illustrates the process modular podcasting production could follow. The creator's agency would be preserved by offering the possibility of circumventing the AI process altogether. Once the audio is segmented, whether that is by a user or software, the chapters could be re-ordered into different version tracks, associated with version-specific metadata.

6.2.3 Envisioned Uses

There are many uses of chapterised audio beyond both listener-centric and creator-centric. Like transcripts, chapters can improve the accessibility and overall reach of programmes, mainly by enabling easy navigation. However, adding this additional information to a programme is often deemed too time-consuming by producers: transcripts are still not systematically provided (Chelsey, 2021), and the process of chapterisation is not well supported by systems in place, on both the producer and the listener's sides (Chapter 5).

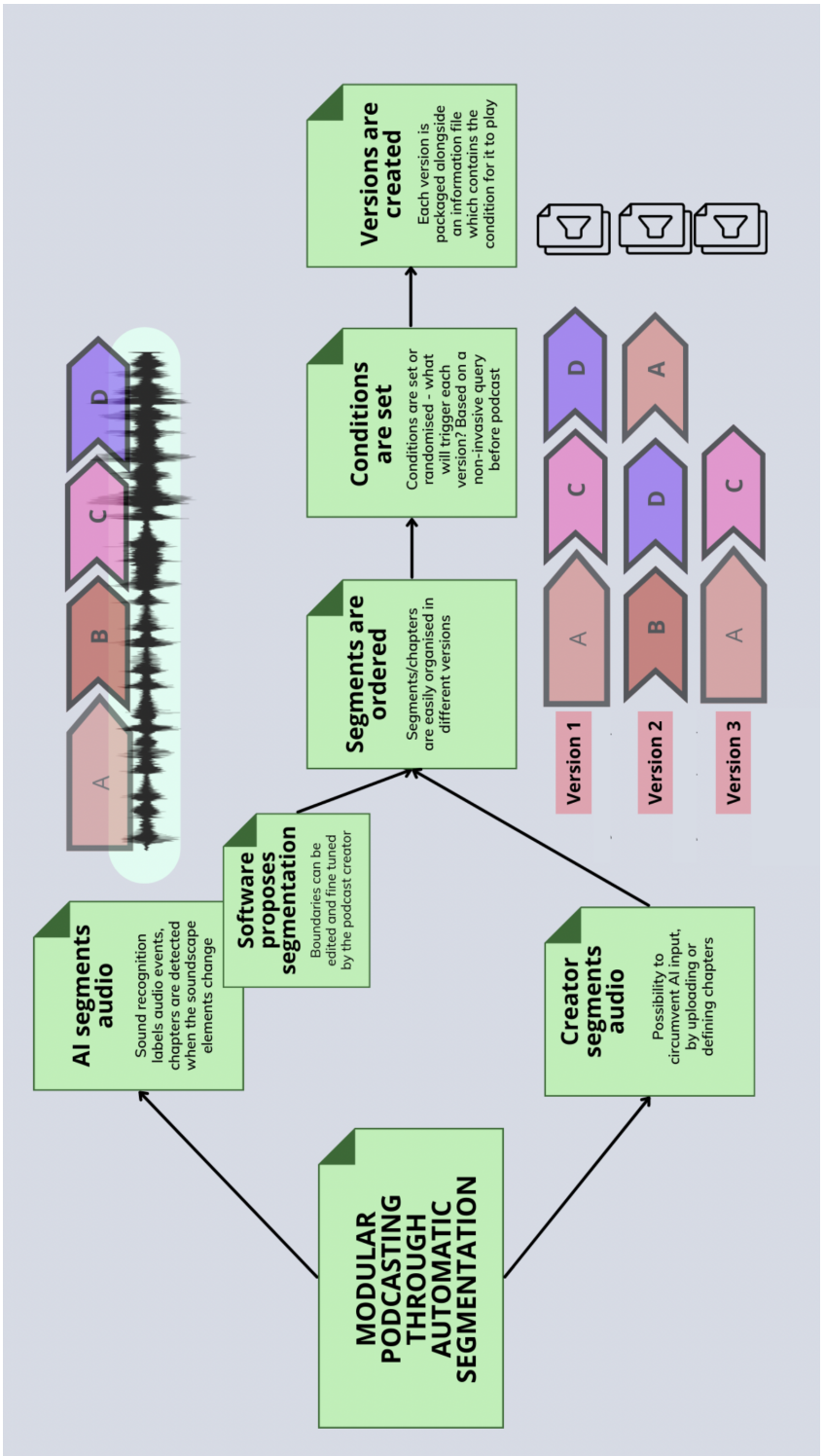


Figure 6.2: Flow diagram for an implementation of the idea of modular podcasting

Additionally, just as podcasting encourages audiences to personalise their listening experiences (by choosing when, where, and how to listen to a show), developments in display and usage of timelines in radio/audio engender another development in interactivity and personalisation, going from "*passive listening to active choosing*" (Jedrzejewski, 2015).

For Podulr specifically, the three possible use cases envisioned at this stage are:

- As a creative tool: different versions of new or existing programmes can be created easily to fit different listeners. The versions created can be hosted on different podcast pages, and the user can simply be pointed towards the right one to go around the lack of standardisation for distribution of interactive audio formats on podcast platforms (Chapter 3).

- As a catalogue manager: to gather and annotate chapters as they are recorded, edited, or published from existing work so that they can be used in the future in compilations, other formats, or other shows altogether

- As a production assistant - easily including chapter information in a podcast's metadata, which would simplify creators' jobs and improve listener experience.

Taking these use cases into account, as well as the requirements gathered in the previous iterative design tasks, a first wireframe was put together (Figure 6.5). It showcased the main point of interactions between the software and the user, having three main prongs: an upload/authentication page (a),

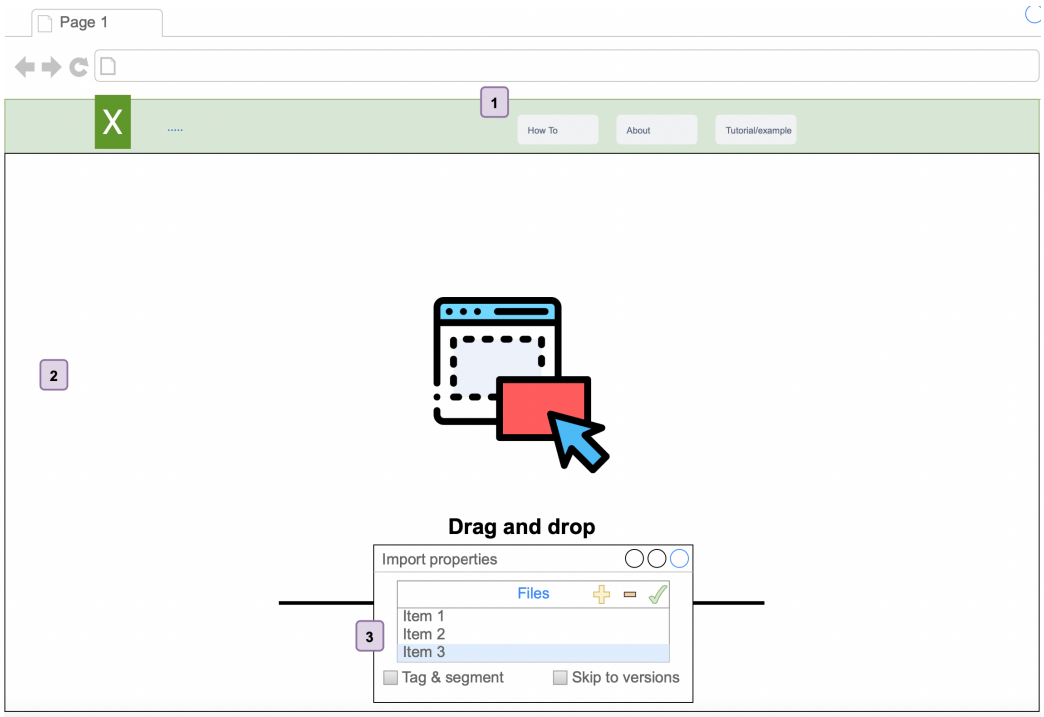


Figure 6.3: Landing Page : (1) Navigation bar: quick access to help and tutorials. (2) Drag and drop: Upload audio files in WAV format. (3) Pop-up screen: Recapitulation of uploads, possibility to add or delete files. Opt-in checkbox for automatic tagging and segmentation of user files.

an editing page (b), and an export page (c).

Although the app's name on the wireframe is "X", possible names for this project at this stage included: Podflow; Flowcast; Flexpod; Chopcast; Podflux; Podular; and StitchCast.

6.3 Interviews About Podulr

6.3.1 Process

Seven participants from prior interactions related to this project were contacted. They were a mix of BBC (3) and Independent (4) creators, and these

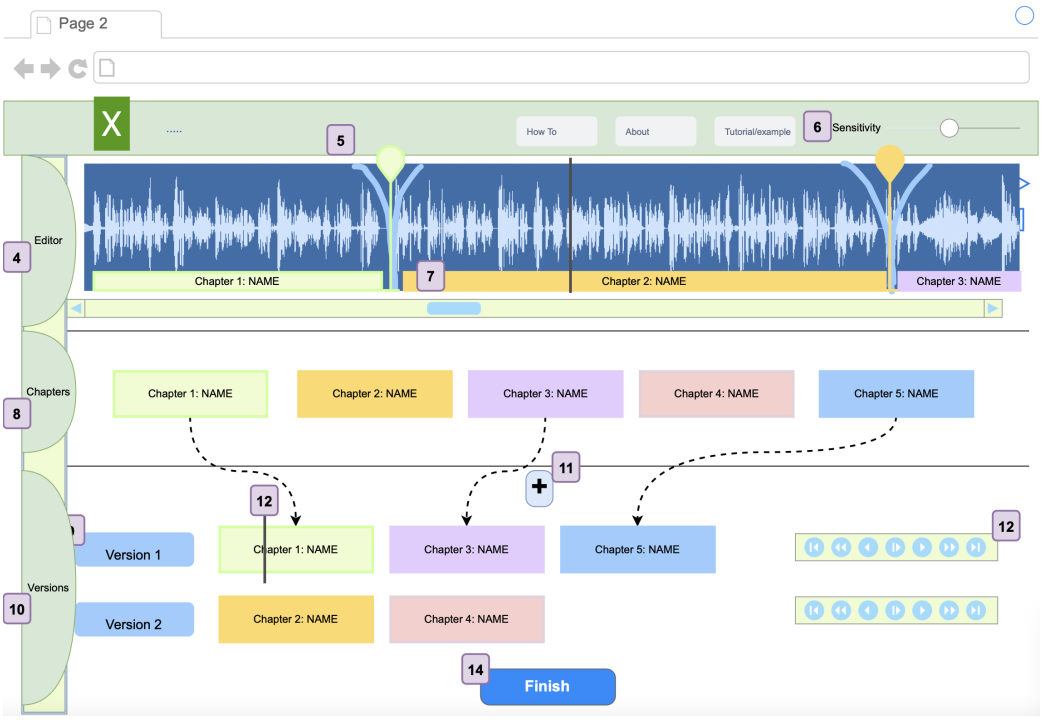
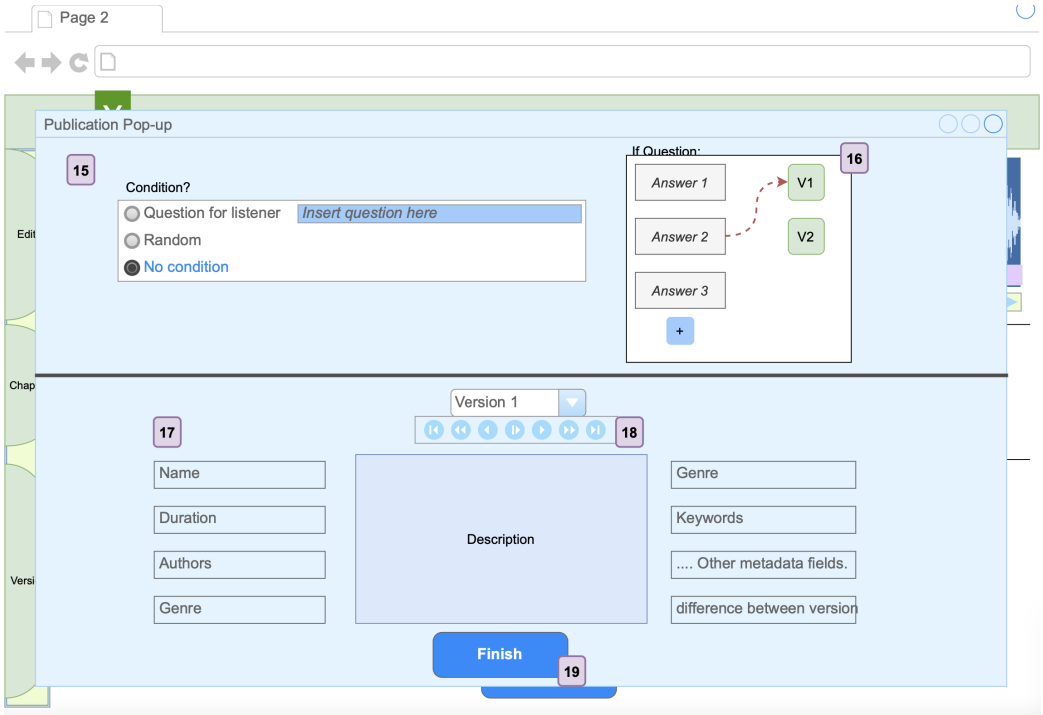


Figure 6.4: App page: (4) Editor: Visualisation of audio file(s) uploaded. Moveable pins (5) and fades are applied at the intersection between segments. Fade duration can be fixed using the slider (6). Chapter names (7) can be edited. (8) Chapters bank: Chapters are represented as boxes that can be dragged onto different versions (9). (10) Versions maker: Make different versions using the chapter bank. Create new versions (11), listen back to the versions (12), and play-head (13). (14) Finish button: opens publication pop-up.



Publication pop-up: (15) Condition check: The user decides whether the versions should be listened to only if some condition is set, randomly, or exported without associated conditions. If a condition is set, the user fills out the question/cue that will be asked to listeners as a pop-up before accessing their podcast. The user associates answers to his set questions to specific versions using a simple flow diagram (16). (17) Metadata portal: Opportunity to correct/add any additional information on the different versions before they are exported—possibility to replay the versions (18). (19) Finish: Download the appropriate file packages.

interviews took place both online via a 30-minute Zoom conversation, or via an open-response questionnaire for two participants who could not attend a call. The app was presented via a summary of the material given in the previous section. This presentation was recorded for the questionnaire and can be accessed in the supplementary material provided.

Following this presentation, participants were asked the following:

- What would you make with this tool?
- Would you use it "on the go" (smartphone/tablet) as well as/instead of on a computer?
- How important is it for your work to be saved so you can return to it later?
- To comment on the wireframe presented, focusing on different sections at a time so the task is not too overwhelming.
- Would you prefer a simple interface with few functionalities or a more complex interface with many functionalities?
- How would you see programmes made with this tool integrated within your podcast host apps on the listener's side?
- To give opinions on names.
- To provide general comments/feedback.

Interviews were not recorded to keep the conversation informal, but notes were taken throughout. The same level of thematic scrutiny demonstrated in Chapter 5 was not replicated, as the questions were quite narrow and fewer

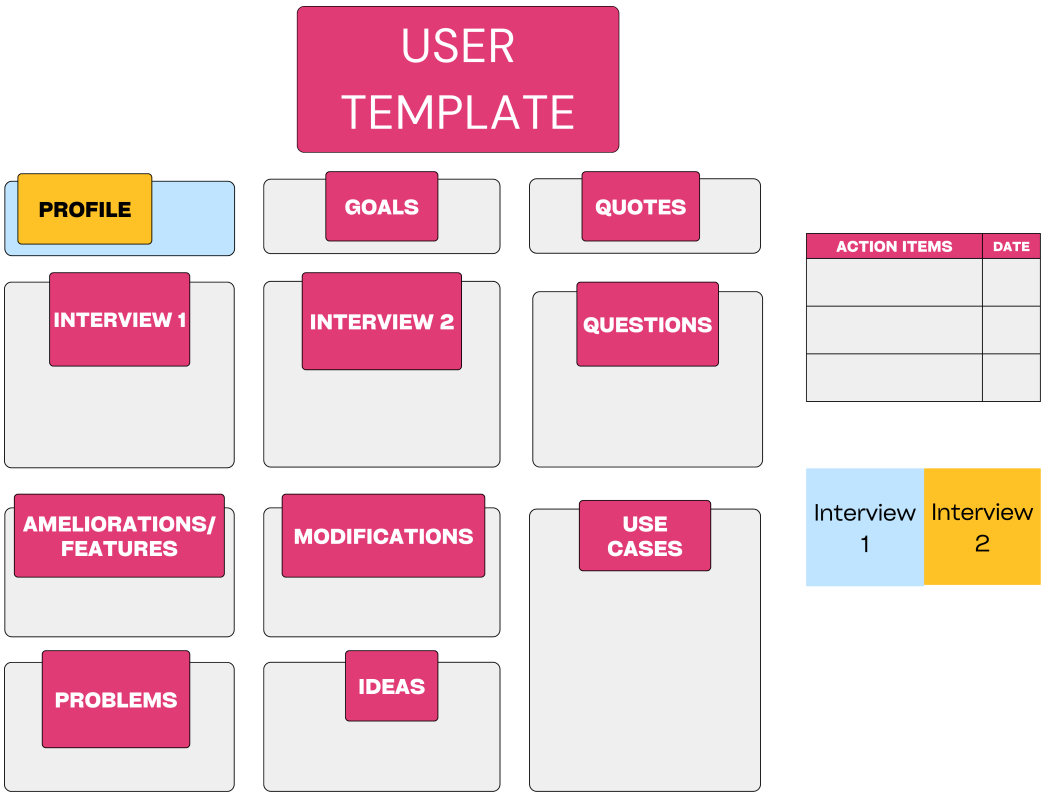


Figure 6.6: Diagram representing the key components of the initial discovery phase

participants were involved. Instead, a deductive analysis was conducted, where overall themes were hypothesised before analysis. The notes were ordered within a table, as can be seen in Figure 6.6. This table includes space for follow-up interviews, the full results will be detailed in Chapter 8 and raw data can be consulted in the supplementary material.

6.3.2 Results

Overall, the participants showed great interest in the app, whether to create new content or as a way to facilitate production. Participants easily projected themselves into use cases, and came up with thought-provoking questions

regarding the app concept and design choices. One BBC participant noted that this type of tool would be well received by their organisation: "*The BBC is going towards modular content [...]. There is also a strong appetite for library building tools for producers*" - Participant B. I grouped the comments received into 4 themes: *Things to add; Things to change; Things to consider; Things to aim for.*

In *Things to add*, the need to preserve the creator's agency as much as possible was highlighted. Five of seven participants mentioned the need to add, remove or change boundaries returned by the software, mostly because they did not trust AI to perform a job to match their exact expectations.

> *"The most obvious issue I can think of is the AIs incorrectly identifying the segments within an episode, resulting in choppy audio or even 'dead air'. I would want to manually review all of the transitions, just in case."-Participant F*

Additionally, creators wanted to retain control of fade times, and playback speed, as well as to keep track of the backend progress of the chapterisation of their files via a progress bar or email updates.

In *Things to change*, the use of authentication was validated, but the possibility of accessing one's work beyond a single browser session was mentioned. The compatibility of the tool with mobile devices also seemed important to the participants, as it would enable users to annotate audio as it is being recorded or if it is being edited "on-the-go".

In *Things to consider*, participants raised some very useful concerns regarding the incorporation of existing transcripts and metadata, as well as compatibility with existing systems. A BBC producer wondered about the internal compatibility of such tools, and independent Participant D wondered

if this tool would only be useful for people hosting their own websites/RSS feeds because of the lack of standards in distribution formats. A fiction podcaster (Participant C) also reinforced the idea that these kinds of personalised experiences should be driven by an editorial motivation:

> *"Choice is everywhere but needs motivation."- Participant C*

Some more specific technological queries were brought up, like how the model would differentiate between music used as "transition" and used as "glue" (Participant B), or how well this would perform on simpler podcasts or different types of speech content. A BBC participant (Participant B) mentioned that offering many different versions of a programme could come with unexpected legal consequences for the BBC

> *"The BBC can only produce a certain amount of audio. Because of competition law, we can't create infinite audio. It's measured in hours, and is strict and rigid at the editorial level. If a podcast changes more than 20% then it is classified as a new version. A new version bumps up content output in quota." - Participant B*

Finally, some other tools were referred to as possible crossovers or points of inspiration, including Cleanvoice.ai [1], the Adaptive podcasting app[2], IDX from BBC News Lab [3], or Starfruit [4].

In *Things to aim for*, participants had clear ideas for goals and outputs. The main ideas discussed are as follows:

> *Creating something easily shareable with their community - Participant A*
>
> *Helping local radios create modular content - Participant B*

[1]https://cleanvoice.ai/
[2]https://www.bbc.co.uk/makerbox/tools/adaptive-podcasting
[3]https://www.bbc.co.uk/rdnewslabs/projects/idx
[4]https://starfruit.virt.ch.bbc.co.uk/

Creating an ad/ad-free version, and using the tool as a production assistant - Participant C

Creating an ad/ad-free version - Participant D

As a publication assistant - Participant E

For episode editing - Participant F

Participant F expanded on more practical applications of a chapterisation tool like the one presented:

> *"This seems like an incredible and easy-to-use tool. Since I produce comedy interview podcasts, narrative segments aren't too important to me, but I would definitely use this for organization and episode editing! Being able to have different versions of episodes would be very helpful, and potentially something that on-air talent would interested in too." - Participant F*

The importance of keeping the UI simple and offering fewer services over a complex tool with an equally complex interface was shared by all participants.

Participants were also asked to vote on their preferred names for this tool; Podular was picked by all participants. The name was adopted and changed to Podulr in all further work.

6.4 What Is a Chapter?

Podulr warrants the implementation of an automatic chapterisation system. But can a machine replicate a process if the process is not universal? Can an algorithm segment audio into chapters in a satisfactory manner for a wide array of users, if users disagree on how audio should be segmented?

From the literature, we can glean at a general definition of "chapter". Carpenter (2024) says:

> *"A chapter is a distinct section of a book that is typically numbered and serves as a division of the overall narrative. Chapters are used to organise the content of a book into manageable segments, allowing readers to easily navigate through the text and follow the progression of the story or information being presented."*

Beyond the literary world, "chapter" is commonly used to refer to "a part of a larger amount of time during which something happens" (Cambridge Dictionary, 2024). For podcasts, just as for books, the expectation and exact definition of chapters can vary depending on who produces the content. Where some authors choose to break down their work in short sections or in large ones, producers may have similar approaches to how they break down their files into its composite chapters. For fiction, the chapters may inherently be equivalent to scenes, like those of dramatic plays, but for non-fiction, the choice falls upon the producer to put down the chapter markers in their content. More practically, chapters are usually set on host platforms, adding time-codes and descriptions of each section just before publishing an episode.

How much producers agree upon what chapters are is essential to building and evaluating the underlying chapterisation system of Podulr. In order to build such a reference frame, a study was conducted with 10 BBC producers tasked with annotating a corpus of 49 podcast excerpts across genres. These annotations were used to determine whether and to what extent producers agree with one another. Since moderate agreement was reported, the annotations were collated into the POD 49 dataset; podcast audio extracts segmented into chapters by two participants each.

This PhD acknowledges that "chapter" can mean different thing to different producers, and through the different studies carried out, attempts to

Table 6.1: Confusion matrix for TP, FP, FN, TN, with l a label given to an object, z an assignment, and + / - the object's relevance of irrelevance respectively

		Assignment z	
		+	-
Label l	+	TP	FN
	-	FP	TN

highlight the ways in which podcast creators agree upon the real properties of a chapter (see 4.2.1).

6.4.1 Putting Together the POD 49 Dataset

In order to measure agreement between producers, Inter-annotator agreement (IAA) can be observed in multiple ways (Kim and Park, 2023; Artstein, 2017), relying on metrics based on counts of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) - where the status of a label l, given to an object, is determined in comparison to an assignment z. Both l and z can denote the object relevant (+), or non-relevant (-) to the hypothesis at hand. Table 6.1 is a confusion matrix representing these terms from Goutte and Gaussier (2005).

Using these terms, IAA can be estimated through:

- **Precision** p is the fraction of positive identifications that were actually correct(Goutte and Gaussier, 2005), such that

$$p = \frac{TP}{TP + FP} \tag{6.1}$$

- **Recall** r is is the fraction of actual positives that were identified correctly (Goutte and Gaussier, 2005), such that

$$r = \frac{TP}{TP + FN} \tag{6.2}$$

- $F1$ is a weighted harmonic average of p and r(Goutte and Gaussier, 2005), such that

$$F1 = 2\frac{p \times r}{p + r} \tag{6.3}$$

- **Accuracy** A is a weighted average of a test sensitivity and specificity (Alberg et al., 2004), such that

$$A = \frac{TP + TN}{TP + FP + TN + FN} \tag{6.4}$$

- **Cohen's Kappa** κ takes into account chance of agreement occurring randomly by looking at observed agreement P_o, and expected agreement P_e (McHugh, 2012), such that

$$\kappa = \frac{P_o - P_e}{1 - P_e} \tag{6.5}$$

These IAA measures can be used to infer agreement between annotators, but also their reliability (Martín-Morató and Mesaros, 2021). For the case of segmentation, Ren et al. (2018) uses time-based IAA with window frames constructed on increments of beats for pattern segmentation in music.

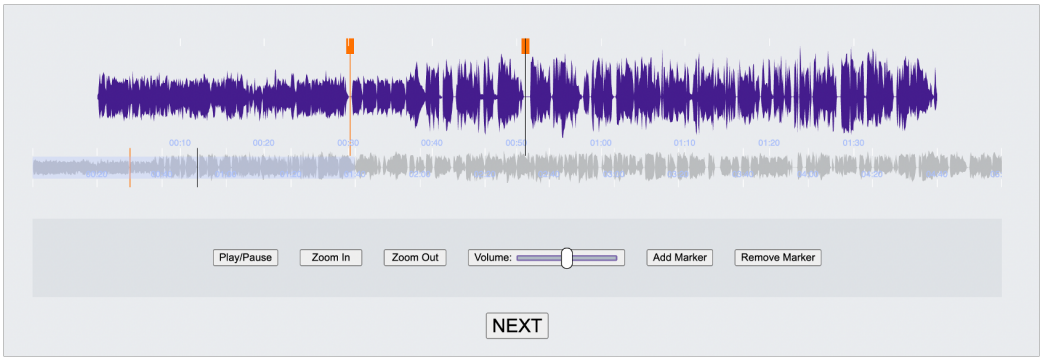


Figure 6.7: Experiment online annotation platform user-interface.

6.4.2 Method

Participants

Participants were recruited via internal BBC communication channels and prior interest registered at the questionnaire phase of the development of podCLIPR (see Section 7.3.2). Thirty-six producers who had expressed interest in innovative audio production tools were contacted via email. They were offered compensation for their time via internal systems. Ten participants out of the 36 contacted took part in an annotation task.

Material

The corpus of 50 5-minute podcast excerpts consisted of the top 40 shows on Apple Podcasts UK on the week of 25/07/23 [5]. Each show was cropped using FFmpeg into three separate excerpts:

- Excerpt 1: First five minutes
- Excerpt 2: Last five minutes

[5] https://web.archive.org/web/20230725104942/https://chartable.com/charts/itunes/gb-all-podcasts-episodes

- Excerpt 3: Random 5 minutes section from the middle of the podcast, excluding the first and last 5 minutes [6]

This segmentation of stimuli was chosen because it prevented the corpus from containing no chapter changes without requiring prior curation. The choice of five-minute duration was informed by experience: longer excerpts might trigger listener fatigue and restrict the end size of the annotated dataset by setting longer tasks, and shorter excerpts might not adequately represent chapter transition but rather smaller changes, and require additional preparatory filtering to prevent the dataset from containing no chapters at all.

These excerpts were filtered by me and another researcher from the BBC to exclude any potentially harmful, problematic or explicit content, and any technical mishap that could lead to risks for the participants (notably, audible pops and mismatched levels). This resulted in a set of 50, clean, excerpts.

Procedure

The expert participants were directed toward a bespoke web platform that enabled them to log their annotations. To make sure each excerpt was annotated twice by different participants, a file was only offered for a second annotator once all files had been annotated once. This meant each participant was not necessarily compared with every other participant.

I needed to build a special website because there wasn't an existing, available platform that would enable this study to run capturing the detail necessary to our research goals. The UI was designed to be easy to use,

[6] The pseudo-randomness was issued by a bash script that dealt with segmenting the files.

but minimal. Aside from accent colours in the wave-form display, grey tones were used as to not distract the participants from their task. An information panel was made available to participants at any point of the experiment via an "information" pop-up overlay. Before running the experiment, the platform was sent out to other members of the Music Computing and Psychology lab at the University of York and to the BBC R&D's audio team, to ensure usability and a bug-free experience that could otherwise greatly affect the results of the study.

The platform (see Figure 6.7) was built using the Peaks.js [7] library for the audio representation. To mitigate reported issues with this library, such as visual and audio synchronisation issues when using WAV, the files were converted to M4A. The platform registered not only the location of chapter markers, but also the time it took participants to annotate each file. The experiment was preceded by a training screen with a fiction podcast that contained audible transitions between scenes [8]. The user data was collected via Smartsheet (non-anonymous) and Glitch.com (completely anonymous) to comply with ethical requirements from all institutions involved.

10 expert audio producers were asked to annotate 10 excerpts each, with each excerpt rated by 2 experts utilising a randomised block design, to form a corpus of 50 podcast excerpts and place chapter markers where they believe a chapter ends and another begins. The phrasing of the task was purposefully open to interpretation, to minimise the potential bias that could be introduced by a more specific set of instructions:

"We ask you to listen carefully to each excerpt and annotate it

[7] https://github.com/bbc/peaks.js
[8] https://www.bbc.co.uk/programmes/m00146p6

> *with chapter markers. These markers should be positioned at the point you believe one segment is ending, and another is beginning. Some excerpts presented to you might not contain any chapters. Chapters might be equivalent to segments or scenes"*

Each file was annotated by two annotators for IAA metrics to be calculated, following a randomised block design. Following the application of IAA metrics, if there was significant agreement between participants, a ground truth corpus consisting of weighted averaged annotations could be created.

Analysis

To evaluate IAA, a window-based approach was chosen (Ren et al., 2018) – that was because, in the context of chapter annotations, temporal matches under a defined threshold are the focus, rather than a "degree of agreement" as often studied in IAA tasks. The influence of varying window sizes on agreement can therefore also be examined. The windows w were reported in seconds and were factors of the total stimulus length 300s so all the audio could be used systematically: [1,5,10,12,15,30,50,100,300].

One of the annotators was set as ground truth (GT), and the other as test (T), then switches. True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) were set as follows (Goutte and Gaussier, 2005):

- TP: in the given window, there is an annotation in the GT and an annotation in T
- TN: in the given window, there is no annotation in the GT and no annotation in T

- FP: in the given window, there is no annotation in the GT and an annotation in T
- FN: in the given window, there is an annotation in the GT and no annotation in T

An incremental counter tracked all TP, TN, FN, and FP. This counter did not allow for double counting (e.g. one annotation in GT and two in T would be +2 in the TP counter), as it would have over-represented cases when two markers were set at either end of a transition when actually only representing a single boundary. The counter ran per file, and not for a participant's complete set of ratings because each annotator was not being rated against every other annotator systematically.

In the case of TN, two values were captured: TN_detail, and TN, which assimilated two adjoining TN_detail into a single window, until one of the annotators denoted a boundary (e.g. if there was a chapter at the windows $\tau = 1$ and $\tau = 5$, TN_detail = 3, but TN = 1). This was to tackle the over-representation of TN one would get from large chapter-free sections.

From TP, TN, FN, and FP, different metrics were looked at, like Precision, Recall, F1, Accuracy, and Cohen's κ.

Although p, r and F1 are commonly used as metrics to determine agreement, in this case, using A in combination with κ gave a better overview of the actual agreement. Indeed, it is a well-reported design flaw that neither p, r or F1 take into account TN (Ren et al., 2018), which was problematic in this case, since a lack of annotation on both GT and T should be considered a success; moreover, in the case of the evaluation of pod-CLIPR, relying on these metrics would prioritise an algorithm that over-annotates, which is not

```
{
  "kmt": {
    "idPod": "Kermode & Mayo's Take",
    "pp1": {
      "id": "A",
      "bnd": [11.87, 92.51, 216.16]
    },
    "pp2": {
      "id": "B",
      "bnd": [92.46, 218.99]
    },
  "nagnt": {
    "idPod":  "The News Agents",
    "pp1": {
      "id": "C",
      "bnd": []
    },
    "pp2": {
      "id": "D",
      "bnd": []
    }
  }
}
```


Figure 6.8: Example annotations of two podcasts by two participants each.

ideal if the goal is human-like chapter suggestions.

Results

The corpus overrepresented sports programmes, because Wimbledon was happening at the time of the creation of the dataset, with several participants stating in the comments that Sports was not their area of expertise. One participant accidentally missed a question and was offered a chance to annotate the missing file again. Two participants who completed the task at the same time, resulted in a glitch and one file being annotated only once.

ID	A	B	C	D	E	Average
F	0.691	0.636	0.750	0.764	0.857	0.740
G	0.771	0.612	0.500	0.615		0.624
H	0.764	0.739	0.778	0.644	0.750	0.735
I		0.565	0.613	0.646	0.561	0.596
J		0.607	0.578	1.000	0.813	0.749
Average	0.742	0.632	0.644	0.734	0.745	0.694

Table 6.2: Heat map of average A per participant at a window frame of analysis $w = 10$ sec. Colour ranging from "red: lowest A across the set" to "green: highest A across the set". White is for "no data".

This file was removed from the final corpus, having a total of 49 excerpts (hence POD 49).

Figure 6.8 shows examples of JSON output from the comparison of producers' annotations on two audio files. Using a window of 10 seconds for analysis, participants A and B agreed there were boundaries at 92s and 216s, but disagreed on another boundary. Setting participant A as ground truth, and participant B as test, the counters would result in: TP = 2, TN = 4, FN = 1, FP = 0. The lack of annotations by participants C and D in Figure 6.8 further highlights the issues with using metrics like Precision, Recall and F1 to evaluate agreement in the context of this study: although participants completely agreed on the lack of boundaries in this excerpt, p, r, and F1 are all 0, because they do not take into account TN. This extreme example demonstrates the error of relying on these metrics for analysis. These values are less representative than Accuracy (A) or Cohen's κ for the purpose of this study. p, r, and F1 are available as supplementary material, but the following analysis focuses on A and κ.

To get a global look at agreement between annotators, the values were

ID	A	B	C	D	E	Average
F	0.596	0.551	0.636	0.692	0.774	0.650
G	0.681	0.505	0.401	0.564		0.537
H	0.627	0.639	0.679	0.587	0.709	0.648
I		0.515	0.443	0.539	0.469	0.492
J		0.504	0.487	1.000	0.729	0.680
Average	0.635	0.543	0.529	0.677	0.671	0.606

Table 6.3: Heat map of average κ per participant at a window frame of analysis $w = 10$ sec. Colour following κ's interpretation framework in McHugh (2012). "Light blue: Weak Agreement", "Blue: Moderate Agreement", "Dark Blue: Perfect Agreement". White is for "no data".

averaged if there were several comparisons between two users, and a non-symmetric agreement matrix was created. Table 6.2 and Table 6.3 represented said agreement matrix at a window frame $w = 10$ seconds.

This comparison was extended by looking at the average Cohen's κ and Accuracy per annotator per window size ranging from [1,50] and [1,300] respectively. Figure 6.9 and Figure 6.10 plot a trend line from a discrete floating average at each window for both κ and A in this range of windows. The larger the window size, the lower the κ, and the higher the accuracy – that is because the chances of agreement occurring at random get higher.

It took participants 6 minutes and 3 seconds on average to complete annotations on each file. This average was calculated by removing 2 outliers where one annotator took a long break, and one accidentally skipped a file.

6.4.3 Analysis

The scales of interpretation of F1 and A both depend on context, which should determine a "success" threshold. Bayerl and Paul (2011) mentions a threshold of 80% for IAA, unless the metric is chance corrected. For κ, the

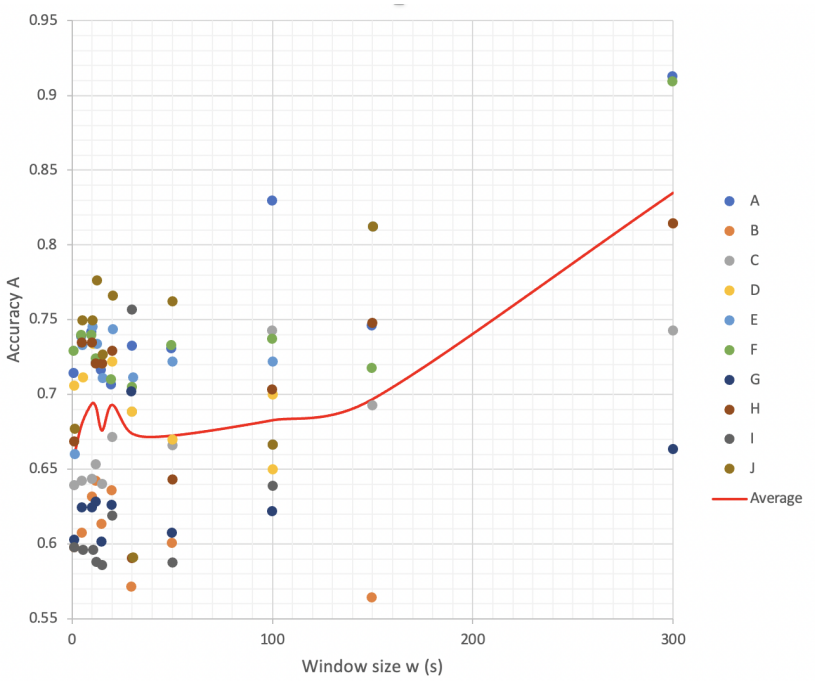


Figure 6.9: Plot of average A per annotator depending on window size. The trendline is a floating average at each discrete window size studied [1,5,10,12,15 etc.].

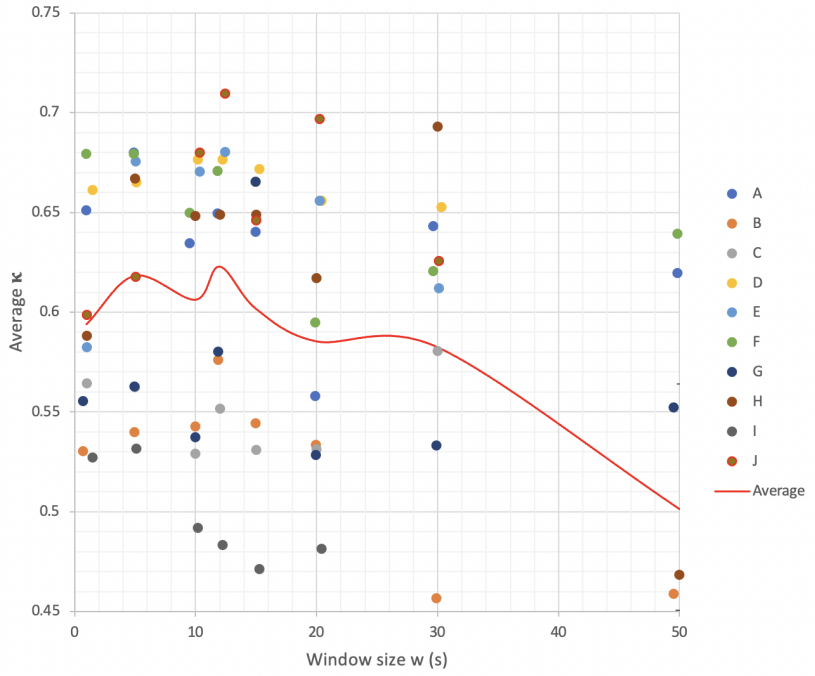


Figure 6.10: Plot of average κ per annotator depending on window size. The range of window sizes for stops at w=50 because for any larger w, the expected agreement nears 1, which means κ's denominator approaches 0 and therefore cannot be calculated.

"agreement scores" as defined by McHugh (2012) are followed.

Following these guidelines, the average κ across the windows examined has its minima at w=50s, which denotes "weak agreement", and its maxima at w=12s, denoting "moderate agreement". Choosing a window size maximising agreement in terms of κ and A sets a range between 5-15s, "moderate agreement". For subsequent tests, the window w=10s is chosen as a representative of that range of acceptable window sizes.

From this initial investigation, it is found that experts agree on what chapters are, although not perfectly. The exact definition of a chapter depends on the expert, potentially influenced by their background and preferred genre, and the context of their segmentation. Consequently, we can only expect an automatic chapterisation tool to come up with plausible suggestions, rather than a perfect performance. It also highlights the need for any suggestion to be easily editable by the producer.

Combining Annotations Into the POD 49 Corpus

For every TP match in the corpus at window=10s, the location of an average boundary marker was calculated, and use the average κ score of each participant across all files to weigh the combined annotation. This annotated corpus is accessible in the supplementary material, and online[9]. It consists of 49 annotated files, including 3 with 3 chapter markers, 18 with 2 markers, 15 with 1 marker, and 13 with no marker. By nature, POD 49 spans different genres and types, including Comedy extracts (16), News (6), Politics (6), Sport (6), Society and Culture (6), Science (3), History (3), TV & Film (3), Health (2), Business (2).

[9]https://annotated-podcast-corpus.glitch.me/

6.4.4 Discussion

By asking 10 audio producers to annotate a corpus of podcasts with chapter markers and looking at agreement metrics, it was found that there is a "moderate" agreement on Cohen's κ interpretation scale (McHugh, 2012). This points us to the empirical conclusion that experts share a universal understanding of what a chapter is, even though the precise definition does vary from one individual to the next, possibly a result of different expectations or backgrounds.

This has two effects within existing literature and projects looking at audio chapters:

- There is no perfect chapterisation, only plausible and implausible chapter marker positions. Over and under-annotating is part of the bell curve formed by the average accuracy scores of experts that can be deduced from averages reported in Table 6.2.

- There exists a ground truth that segmentation methods can be compared to. POD 49 is such a dataset, although this study justifies the creation of other datasets for more specific applications. The broad nature of POD 49 – taking segments from Apple's top 40 podcasts regardless of genre, production type etc. – indicates it might give an estimation of a segmentation system according to these broad guidelines. Were a system to be more specialised, or otherwise require a more specific evaluation (in terms of genre, format, production type, content etc.), a more specific dataset could be formed.

The creation of the POD 49 dataset also showcased a method of window-

based IAA analysis, expanding Ren et al. (2018)'s work in MIR, and Argaw et al. (2022) work in film and media. This method could be useful to create other datasets of time-based parameters, as it rewards correctly annotating as well as not annotating, which takes into account the importance of having an algorithm that does not over-annotate.

Overall, POD 49 can be used by the wider community to evaluate other segmentation algorithms, like NLP-based audio segmentation, as presented by Feldstein Jacobs (2022), or other methods for chapterisation (Barthet et al., 2011), and contributes to our understanding of a consensus surrounding the definition of chapters in audio and media.

6.5 Summary

The interviews conducted with creators at this stage to get practical feedback on the base concept for Podulr not only set further requirements for the project, but also provided more information on the target user of Podulr - beyond a tool made for tech-savvy producers to make NGP, Podulr has the potential to be used as an editing assistant or post-production tool, as a companion to the editors working in chunks, or the shows that work off long live recordings. Some missing features and components were highlighted, and some important questions arose from the discussions (e.g. How well would this perform on simpler podcasts or different types of speech content? How could the tool be scaled to a large organisation like the BBC? How will the content be distributed across platforms?).

The following study showed that there is an unspoken, implicit agreement amongst podcast producers as to what a chapter is; however, this lacks

a formal definition. An investigation into a more explicit definition of audio chapters could be performed to give a grounding for interactive media research.

The curation of POD 49 answered the question: Do expert podcast producers agree on what chapters are? Indeed, there is shown to be "moderate agreement" between the 10 expert producers prompted to annotate the files of POD 49. This analysis suggests that arriving at a correct chapterisation is not a fixed itinerary, but rather a collection of multiple pathways that all arrive at a plausible segmentation. This study showcased the use of IAA calculated per fixed window of time.

This phase of research adds more components to the answers to the research questions set:

RQ 2: How can we feasibly create and distribute immersive and personalised podcasts?

The need for simple tools well integrated within existing systems was further highlighted. Potential issues regarding distribution were discussed with producers, and lacking a way to unify the distribution process of interactive online audio, the idea to circumvent the issue altogether by hosting different versions of shows that could cater to different listeners was envisioned. This does carry its own set of problems, notably that organisations like the BBC only have a certain amount of content they can distribute, and that it requires more interactions from the user to get to their desired content.

RQ 3: Why is modular podcasting attractive to creators, and how can it be facilitated without making the production process more complex?

To understand the appeal of modular podcasting, the potential use cases

revealed in this Chapter can be referred to: *as a creative tool, as a catalogue manager, as a production assistant.* These three use cases were expanded upon by producers, citing for instance being able to help local radio productions cater their on-demand content to their audiences, or easily creating ad/ad free versions of their shows for patrons or fans. In addition to these specific use cases, producers also saw value in increasing accessibility and customisation of their programmes.

To tackle the issue of systematic chapterisation being time-consuming and poorly supported by current tools and platforms, Podulr was presented as an assistant to tag and segment chapters. This web app can be integrated at any cross-over point of the workflow, specifically intended to be used in conjunction with the post-production phase, but that could also easily be used on the go alongside the production (recording and editing) steps. To facilitate its use, making the app compatible across devices and the outputs as compatible as possible with existing structures is key.

7 Podcast Chapter Localisation through Intelligent Pattern Recognition (pod-CLIPR)

7.1 Introduction

The purpose of an automatic chapterisation tool in the context of podcasting was described in Chapter 6. The task of chapterising an existing file (adding chapter markers and associated metadata) is an important step in the production process – not only for accessibility reasons on the listener's side (e.g. by enabling easy navigation, or otherwise customisable content), but also to facilitate editing future episodes (e.g. by being able to easily access and re-use segments from prior shows) and supplementary material (e.g. highlights for social media). Although it is often a necessary step, podcast creators complain that "*the current system for chapter tagging and navigating is impractical and imperfect*" (Chapter 5).

The potential implications of automatic chapterisation go beyond simplifying navigation for users and assisting production. Interactivity has been implemented across various media, and in its many incarnations, the need for interactive media tools that do not further complexify already-arduous workflows has been highlighted (Chapter 5). Moreover, interactive media tools often face several barriers to being widely adopted: the lack of format or

standards, the lack of distribution solutions, and possibly most importantly, the requirement for additional, highly technical production work (Chapter 3).

The issues with chapterisation are widespread and recognised across the industry. They could therefore benefit from an AI-driven solution, in order to assist production, improve the listener experience, and simplify interactive audio production.

To complete Podulr, I put together a sound-based system for automatic podcast chapterisation, pod-CLIPR (Podcast Chapterisation through Intelligent Pattern Recognition), which uses audio event tags from a sound recognition convolutional neural network (CNN) (Kong et al., 2020) to identify and categorise changes in a soundscape, followed by a novel rule-based approach that infers potential chapter boundaries. This Chapter describes this system and its evaluation against the POD 49 dataset. This study demonstrates not only an efficient, sound-driven solution to audio chapterisation, and the use of inter-annotator agreement with a time-window approach, but also provides partial answers to *RQ 3: Why is modular podcasting attractive to creators, and how can it be facilitated without making the production process more complex?* and *RQ 4: What are the perceived benefits, risks, and costs of exploiting AI technologies for podcast production?*.

After introducing some of the previous work done in the field of chapter segmentation, some aspects of the methodology used are further detailed. Finally, the results of the study are analysed and discussed. This reflection showcases the successes and limitations of pod-CLIPR, which is shown to be able to suggest plausible chapter segmentations, where plausible means comparable to those of a human expert, as recorded in the POD 49 dataset.

All mathematical terms used in this Chapter are listed and defined in Appendix A.

7.2 AI for Automatic Media Segmentation

The idea of using AI for media segmentation is not novel – for instance, there already exist various applications of machine learning for video editing. Hidden Markov Models (HMMs) have been applied successfully to image and audio features for video segmentation (Boreczky and Wilcox, 1998). Soe (2021) gives an overview of previous systems and the technologies they employ – primarily, image analysis, transcript analysis, or motion and audio analysis. Soe (2021) also lists the various applications of AI for video editing: *Segmentation of videos, composition of video segments, visualisation of the timeline and video clips, smart manipulation of clips, creating transitions, and logging videos* (p. 3-4). This field is so proliferative that it warranted a systematic mapping analysis, covering the evolution of the research (Bieda and Panchenko, 2022).

Within Soe (2021)'s list of AI applications for editing, *Segmentation of videos* can also be applied for uses beyond the medium of film. AI-generation of highlights (also known as thumbnailing), a task that seeks the most salient shots in a full-length video (Ping and Chen, 2017), can be used to create trailers, social media promotions and more (Jiao et al., 2018). There are different approaches to automatically finding highlights in videos: NLP (Anne Hendricks et al., 2017), unsupervised learning from web-crawled training data (Yang et al., 2015), supervised learning from semantic embedded comments (Lv et al., 2016), or predicting time-sync comments (comments

attached to specific moments) (Ping and Chen, 2017).

In the audio domain, there is evidence of audio-based segmentation for video editing in Truong et al. (2016); Takiguchi et al. (2008). But, has AI been used to edit purely audio content? There are two facets to audio segmentation: dividing audio information into streams (Theodorou et al., 2014), and chapterisation, which can alternatively be described as a horizontal or temporal process. Both are different from personalised or AI-driven mixing (Oldfield et al., 2022; Sai Vanka et al., 2023) or mastering (Birtchnell, 2018), which make use of AI to finalise an audio file for distribution. This PhD focuses on chapterisation, as opposed to un-mixing audio into tracks.

Establishing clear chapter metadata could facilitate production and potentially offer some new creative applications. Podcast chapterisation has been attempted by Barthet et al. (2011) using a "Music or Speech" recognition system to separate musical segments and spoken ones. Feldstein Jacobs (2022) investigates the possibility of using LLMs for thematic analysis and segmentation, but the results presented are not replicable: not enough information on the datasets and resources used is made available to repeat either the methodology or findings.

7.3 pod-CLIPR

pod-CLIPR is a system that segments podcasts into chapters by analysing the output of an audio pattern recognition (APR) model. This sound-driven approach enables the segmentation to follow changes in the soundscape of a podcast. Even though pod-CLIPR is incapable of picking up on thematic changes in speech, the nature of podcast production – which often includes

clear transitions, edits, or even in the case of fiction, scenes with background sound effects and music – provides sufficient contrast for an audio-sensitive segmentation.

7.3.1 Audio Pattern Recognition

Artificial intelligence has been used extensively for audio analysis, for instance in musical beat tracking (Davies and Plumbley, 2007; Ellis, 2007), and processing (e.g. sound source separation (Défossez et al., 2019)), or synthesis (e.g. audio and music generation (Kreuk et al., 2023; Mehri et al., 2017; Barahona-Ríos and Collins, 2024)). The process of automatically tagging sound events, as described for instance by Kong et al. (2019a) and Kumar and Raj (2016a) has a plethora of applications. Here sound event recognition is used to map the evolution of a soundscape over time.

The particular iteration of sound recognition used in pod-CLIPR is a large-scale pre-trained audio neural network for audio pattern recognition (PANN-APR) Kong et al. (2020). It has been trained on the Audioset Ontology,[1] a dataset containing 5000 hours of audio with 527 sound classes. The model can be run on a whole audio file in WAV format, giving an overview of the total composition of a sound file, or frame by frame at a specifiable frame rate and window size. By averaging across a larger set time window, the granularity of the analysis can be modified. The detection window frame W_t is set as

$$W_t = 0.5 \text{ sec} \tag{7.1}$$

In comparison to other music or audio information retrieval time-based anal-

[1] https://research.google.com/audioset/ontology/index.html

ysis tasks, 0.5 sec might seem long, but this decision is informed by the end purpose of this algorithm: to output plausible chapters that producers can fine-tune if necessary on an online editing platform.

There is a difference between this window detection W_t and the best window frame $w = 10s$ found in the previous chapter (6). Where 10s was found to be the best fit to analyse overlapping annotations, 0.5 reflects the human effort of listening to key changes in audio to infer these boundaries. Any larger window sizes could result in glossing over whole sections of audio.

For an input audio file of length L sec, with N the number of frames in an audio file, and $N \in \mathbb{N}$, the output of PANN-APR is a state vector $\mathbf{s}_\tau$, where

$$\tau \in T = \{nW_t : n = 0, 1, \ldots, N = \lfloor L/W_t \rfloor\} \tag{7.2}$$

The state vector is notated

$$\mathbf{s}_\tau = (v_{1\tau}, v_{2\tau}, \ldots, v_{K\tau}) \tag{7.3}$$

with $v_{K\tau}$ indicating the probability ascribed by the model that the audio in the time window τ constitutes a particular category K.

These features can be filtered so that those with a probability below a threshold are not included in the output.

A cutoff threshold is set initially at 0.01, after testing different precision levels on test files. 0.01 was chosen because it returned incredibly comprehensive sets of labels and their probability quickly. In contrast, 0.001 returned many labels that seemed random or incorrect, and 0.1 missed some of the quieter components of the soundscape.

pod-CLIPR was developed using ISD techniques (Basil and Turner, 1975). A set of short podcast inputs were selected to test the algorithm as the development process went along. Parameters were set to get the best possible output with these files, following my estimation of where chapters began and ended.

This method informed the decision to include a jingle detection adding boundaries around musical moments (without this option, musical-led shows and magazine shows couldn't be properly segmented in the test dataset), to set the minimum chapter length and maximum transition time as $threshold\ duration = 8\,sec$ (longer transition times led to segments being ignored, while shorter transition times led to fades and transition segments would sometimes miss the longer fades between chapters in the test dataset) and to set an overall detection threshold (several thresholds were tested until the most reliable outputs were detected on the test dataset).

7.3.2 Reducing the Number of Categories

The model returns 527 feature probabilities. These features are mapped onto a set of 10 labels:

$$\begin{aligned}&\text{Labels} = \{\\&\text{“Music”, “Speech”, “Conversation”,}\\&\text{“Female speech; woman speaking”,}\\&\text{“Male speech; man speaking”,}\\&\text{“Narration; monologue”, “Outside; rural or natural”,}\\&\text{“Inside; small room”, “Singing”, “Sound effects”}\\&\}\end{aligned} \tag{7.4}$$

The values in Labels were determined by asking 20 podcast creators to answer a one-question questionnaire online: 100 top UK podcasts were analysed and all the tags returned by the sound recognition model used in CLIPR were examined.

This preliminary study was conducted to support the rule-based decisions made in pod-CLIPR. It was conducted with 20 BBC creators who had registered interest in previous stages of this doctoral project and internal lists of producers interested in helping audio R&D projects. The survey was hosted on Qualtrics.

All Sound Effects are grouped as one label, idem for individual Musical instruments (as opposed to the label “Music”), as these are sporadic sounds less likely to define a chapter or section. The 20 most common labels were presented to the participants so they ticked which labels out of this set they

considered pertinent to tag the contents of a podcast. The top 10 labels ticked by participants populate the set Labels, which helped simplify the set of rules defined and applied below.

Each state $\mathbf{S}_\tau$ is then defined as $\mathbf{S}_\tau = (v_{1\tau}, v_{2\tau}, \ldots, v_{10\tau})$, with $v_{1\tau}$ the value associated to the feature "Music" at a frame τ, $v_{2\tau}$, to "Speech", $v_{3\tau}$ to "Conversation", etc. There is no native "Sound Effects" label returned by the SER model. In order to compute this particular feature of incidental noises and effects, $v_{10\tau}$ is set as the sum of all probabilities of incidental noises and effects occurring at state τ.

7.3.3 Identification of Candidate Boundaries

The purpose of the steps described below is to determine whether there are high fluctuations between adjoining states $\mathbf{S}_\tau$ and $\mathbf{S}_{\tau+1}$. The flux between two states at windows τ and $\tau + 1$ is defined as

$$\Phi(\tau, \tau + 1) = \frac{1}{R} \sum_{k=1}^{R} |v_{k\tau+1} - v_{k\tau}| \tag{7.5}$$

With R the total number of features considered in *Labels*, $R = 10$. This calculation is depicted in Figure 7.1.

The use of flux as a feature is typical in systems attempting to detect events (e.g. onsets or beats) in audio (Collins et al., 2014).Lerch (2012) defines spectral flux, which measures the amount of change of the spectral shape between consecutive frames, where Φ measures the amount of change between state vectors $v_{k\tau}$ as computed from the SED model.

For the sequence of spectral flux values $(\Phi(\tau, \tau + 1))_{\tau \in T}$, let P_i denote the i^{th} percentile of the values therein. This configuration of pod-CLIPR uses

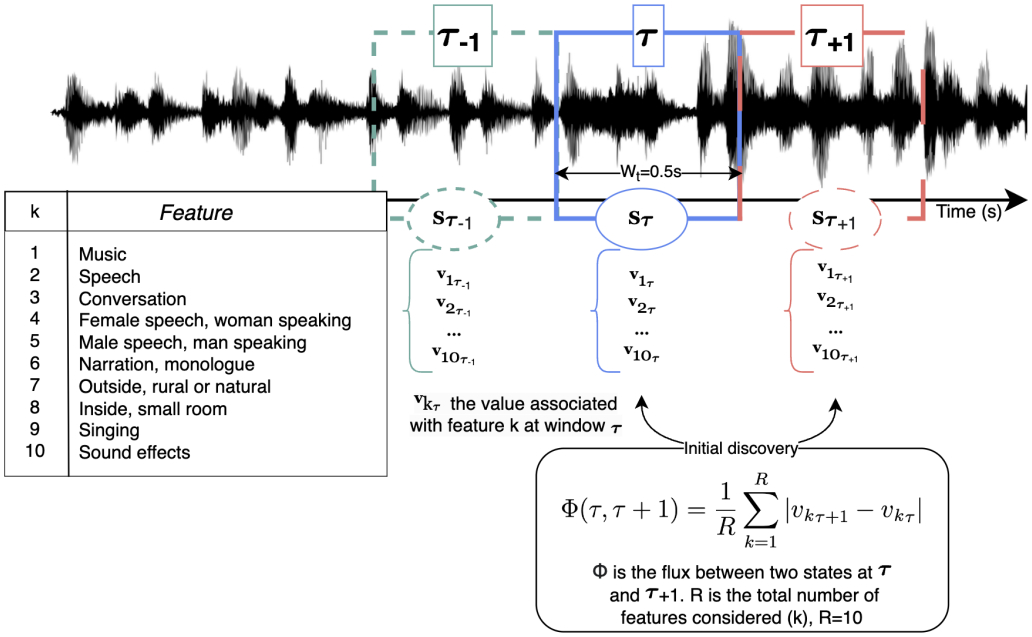


Figure 7.1: Diagram representing the key components of the initial discovery phase

P_{80} as a threshold to determine whether two adjoining states are significantly different. P_{80} was chosen during the iterative development for returning the most coherent answers for the test files used.

If $\Phi(\tau, \tau + 1) > P_n$, there is said to be high flux between states $\mathbf{S}_\tau$ and $\mathbf{S}_{\tau+1}$.

To determine if this is more than a momentary change, the comparison is extended to $\mathbf{S}_{\tau-1}$ and $\mathbf{S}_{\tau+1}$, focusing on the maximum feature value M_τ for some frame τ, defined by

$$M_\tau = \max_{k=1,\ldots,10} \{v_{k_\tau}\} \tag{7.6}$$

Two other quantities relevant to this consideration are:

- the absolute difference between M_j and $M_{j'}$, notated $d(j, j') = |M_{j'} - M_j|$;
- the number of standard errors from the mean of the values $(\mathbf{S}_\tau)_{\tau \in T}$, defined by $\epsilon_z = z\sigma_\tau/\sqrt{n}$, where σ_τ is the standard deviation of $(\mathbf{S}_\tau)_{\tau \in T}$, with $z = 0.95$, and n the sample size.

ϵ_z represents the inner-state variation between all the feature values $v_{k\tau}$ and the algorithm relies on its value as a threshold to compare major fluctuations between two states.

If the following three criteria hold:

$$\Phi(\tau, \tau + 1) > P_n \tag{7.7}$$

$$d(\tau, \tau - 1) > \epsilon_z \tag{7.8}$$

$$d(\tau + 1, \tau - 1) > \epsilon_z \tag{7.9}$$

then there is said to be a candidate boundary at $\tau \in T$.

7.3.4 Rules for Reducing False-Positives in Detected Boundaries

Inspecting these candidate boundaries for several recordings, and working with podcasts over a number of years, it is clear that podcasts contain certain idiosyncrasies that lead to five categories of false positives among candidate boundaries: 1) suspicious neighbours, meaning that there are too many candidate boundaries in some particular region; 2) transitions, short periods that fade one part into the next, resulting in two boundaries at either end of the fade, instead of a single boundary; 3) conversation interruptions,

where the speaker changes within a conversation involving multiple speakers; 4) dramatic pause, where the speaker takes a short pause in speech not equivalent to a topic change.

Finally, the value of music used as a cue is recognised, so another layer of detection is introduced for 5) jingles, where music is used as transition or introduction.

Each candidate boundary location is noted β_l, with l ranging from the first to last candidate boundary index. A rule is formalised to address each of the categories outlined above, updating the set of candidates after each rule is applied. This entails that the order of the rules can significantly alter the output of the algorithm. A flow diagram of pod-CLIPR can be seen in the supplementary material provided.

Suspicious neighbours examination. If $(\beta_{l+1} - \beta_l)/W_t < D$, the code proceeds to check if the flux is sufficiently large by comparing the composition of the state before β_l (or $\beta_l - 1$) with that of the state after β_{l+1} (or $\beta_{l+1} + 1$)

.

Our initial configuration of pod-CLIPR uses $D = 8$ sec, as we understand that shorter segments could scarcely be considered chapters.

If $\Phi(\beta_l - 1, \beta_{l+1} + 1) > \epsilon_z$, then β_l and β_{l+1} are not considered to be boundaries, and both are removed from the set of candidates. Else, β_l is removed and β_{l+1} is retained, to avoid doubling chapter markers at transitional periods.

Transitions check. To complement the "*Suspicious neighbours examination*", a final check on neighbouring candidate boundaries is performed.

If $\frac{|\beta_{l+1} - \beta_l|}{W_t} <$ *threshold duration*, it is assumed the segment detected is

transitional in nature, whether that is because it contains a cross-fade or sound effects denoting such transition. The code therefore concludes that β_{l+1} is not a boundary.

Conversation interruption inspection. To avoid false positives due to speaker changes within a conversation involving multiple speakers, the feature $k = 3$, "Conversation" is relied upon. If $v_{3\beta_l - a} > 0$ or $v_{3\beta_l + a} > 0$, with $a = 1$ or $a = 2$, then β_l is not a boundary but a change of speakers in a conversation, so β_l is removed from the expected boundaries set.

Dramatic pause scrutiny. ν_τ is the average value of all the non-speech related features at the window τ, that is: Music ($k = 1$), Outside; rural or natural ($k = 7$), Inside; small room ($k = 8$), Singing ($k = 9$), Sound effects ($k = 10$).

$$\nu_\tau = \frac{v_{1\tau} + v_{7\tau} + v_{8\tau} + v_{9\tau} + v_{10\tau}}{5} \tag{7.10}$$

If $|\nu_{\beta_l - 2} - \nu_{\beta_l + 2}| < \epsilon_{\beta_l}$, it is concluded that the environmental components of the soundscape two frames before and after the expected boundary are similar, and therefore that the high fluctuation at this expected boundary is due to a pause in speech, and therefore β_l is not an actual boundary.

Jingle detection. Two expected boundaries β_l are added at the beginning and end of a "nearly-consecutive" series of states where the maximum feature is $k = 1$,"Music", if they have not already been detected in the initial discovery phase. Two states are "nearly-consecutive" if $\frac{|\beta_{l+1} - \beta_l|}{W_t} < 5\,sec$, a duration which should account for the possible variations within a musical piece.

7.4 Method

To evaluate an algorithm, it is customary to compare its performance to a benchmark. In the case of media segmentation, Argaw et al. (2022) demonstrates the importance of having an expertly annotated corpus of extracts as said benchmark. The AVE dataset used in Argaw et al. (2022) includes several dimensions of information about each video file given to a group of 15 experts – for this specific goal, only one dimension of annotation was needed: chapter markers on a time-axis.

To evaluate pod-CLIPR, the corpus created in Section 6.4.1 and the output of pod-CLIPR across different parameters was compared to it. IAA metrics (see Section 6.4.1 for definitions) for the output of pod-CLIPR were calculated, treating the expert-made corpus as ground truth.

Although p, r and F1 are commonly used as metrics to determine agreement, in this case, using A in combination with κ gave a better overview of the actual agreement, for similar reasons as reported in Chapter6.

7.5 Evaluating pod-CLIPR

7.5.1 Set-up

The corpus created in Section 6.4.1 enabled us to evaluate pod-CLIPR. The effect of modifying the following parameters was observed:

P1 Jingle detection (original position/none/ending position)

P2 Transition length (4 sec, 8 sec, 10 sec, 12 sec, 14 sec),

P3 Flux detection threshold (70%, 80%,90%)

Window (s)	κ	Accuracy
10	0.606	0.694

Table 7.1: Average κ and Accuracy at a window frame w=10s for the experts who annotated the POD 49 dataset

An exploratory investigation was carried out, where the performance of the original configuration was compared to single parameter changes for P1, P2, and P3. The different values for each of these parameters were set by extending the original values to plausible alternatives surrounding it – with what values are deemed "plausible" determined in the iterative development process: a detection threshold over 90% returned almost no boundaries, while a threshold under 70% returned boundaries at every minor change in soundscape; a transition length over 14 sec ignored a very short chapter, while a transition length under 4 sec systematically tagged the beginning and end of all transitions as chapters. By looking at the highest of these scores, a best-performing configuration was hypothesised, and evaluated against the expert corpus. Cross-validation was used to draw conclusions on the performance of this optimal configuration across different shows.

7.5.2 Results

The POD 49 corpus was used as Ground Truth, noting note the average κ and A of human experts at a window frame at 10s (Table 7.1).

$w = 10s$ was used as a reference, as it was the window frame used to make the corpus, but it should be acknowledged that there was no "perfect" window frame, rather that the optimal window was contained in a range of 5-15s.

Configuration	w=1 sec	w=5 sec	w=10 sec	w=12 sec	w=15 sec	w=20 sec	Average
Original	0.560	0.559	0.556	0.546	0.544	0.531	0.549
No Jingle	0.588	0.580	0.584	0.568	0.570	0.564	**0.576**
End Jingle	0.561	0.515	0.512	0.518	0.490	0.481	0.513
Transition 4	0.527	0.498	0.493	0.472	0.452	0.469	0.485
Transition 6	0.534	0.535	0.534	0.521	0.512	0.502	0.523
Transition 10	0.575	0.577	0.576	0.564	0.567	0.563	0.570
Transition 12	0.590	0.586	0.592	0.569	0.581	0.570	0.581
Transition 14	0.592	0.588	0.600	0.573	0.593	0.586	**0.589**
Detection 70	0.573	0.577	0.576	0.578	0.573	0.566	**0.574**
Detection 90	0.580	0.564	0.550	0.568	0.558	0.530	0.558

Table 7.2: Different configurations' average κ scores across analysis window frames ranging from 1-20. "Light blue: Weak Agreement", "Blue: Moderate Agreement"

The algorithm pod-CLIPR was evaluated by varying 3 different parameters:

P1 Jingle detection (original position/none/ending position)

P2 Transition length (4 sec, 8 sec, 10 sec, 12 sec, 14 sec),

P3 Flux detection threshold (70%, 80%,90%)

This exploratory investigation looking at an original configuration compared to single parameter changes for P1, P2 and P3 was ran to compose a configuration of the best-performing set of parameters.

The same analysis script and method as put together to analyse the corpus was run. The corpus was set as GT and the algorithm's output as T. TP, TN, FP, and FN were counted in the same way. From this, A and Cohen's κ were calculated for each file and averaged for different window frames w. Results are seen in Table 7.2 and Table 7.3.

Using cross-validation, an "optimal configuration" is found. The highest κ for the entire set of POD 49 excluding one show was computed, and this was

Configuration	w=1 sec	w=5 sec	w=10 sec	w=12 sec	w=15 sec	w=20 sec	Average
Original	0.662	0.636	0.634	0.626	0.628	0.621	0.634
No Jingle	0.687	0.655	0.657	0.645	0.586	0.647	**0.646**
End Jingle	0.651	0.596	0.596	0.604	0.586	0.586	0.603
Transition 4	0.631	0.579	0.578	0.563	0.552	0.572	0.579
Transition 6	0.649	0.613	0.613	0.603	0.600	0.602	0.613
Transition 10	0.665	0.653	0.653	0.642	0.646	0.646	0.651
Transition 12	0.673	0.661	0.665	0.648	0.656	0.648	0.659
Transition 14	0.676	0.664	0.673	0.654	0.666	0.663	**0.666**
Detection 70	0.679	0.653	0.650	0.655	0.652	0.647	**0.656**
Detection 90	0.668	0.640	0.628	0.643	0.636	0.615	0.638

Table 7.3: Different configurations' average A across analysis window frames ranging from 1-20.

repeated until all shows have been excluded, following a leave-one-out cross-validation methodology. All sets examined return the best configuration such that: P1 = No jingle f, P2 = Transition length 14 sec, P3 = Flux detection threshold 70%. I call this optimised configuration of parameters *omega* to easily refer to it. This is indicative that no show particularly impacted the "optimal" configuration that can be inferred from Table 7.2 and Table 7.3. Table 7.4 showcases the results of cross-validation looking at Cohen's κ as an IAA metric at w=10s.

Using *omega* on the whole corpus, the average κ was 0.598 at w=10s, compared to 0.606 for the participants. Both just fall within the range of "moderate agreement"]0.59;0.79[. The average Accuracy was 0.674 at w=10s, compared to 0.694 for the human experts. At w=10s, the algorithm outperformed four humans in Accuracy and κ that were in lowest agreement with other experts.

Omega at different sound recognition thresholds: 0.001, 0.01, 0.1, and 0.5 was investigated. *Omega* performed better as the threshold diminished, that is to say, the more tags are returned by the model, the more accurate the

Show tested	κ
The rest is politics	0.706
The news agents	0.683
Leading	0.652
Sh**ged, married, annoyed	0.523
Test match special	0.714
The infinite monkey cage	0.697
DOAC	0.558
Off-menu with Ed Gamble and James Acaster	0.545
The rest is history	0.554
No such thing as a fish	0.552
My therapist ghosted me	0.545
Kermode and Mayo's take	0.554
F1 checkered flag	0.571
Elis James and John Robins	0.559
Nearlyweds	0.573
The Frank Skinner show	0.857
Desert island discs	0.458
Today in focus	0.544
ZOE science and nutrition	0.772

Table 7.4: Results of the cross-validation test performed on the POD 49 dataset per show

Threshold	κ	A
0.001	**0.655**	0.715
0.01	0.598	0.674
0.1	**0.617**	0.686
0.5	0.598	0.660

Table 7.5: Average Cohen's κ and Accuracy at w=10s for the *omega* configuration at different sound recognition thresholds. In bold, are the cells that outperform the average scores from human experts at w=10s. In blue are the cells that fall within "moderate agreement"

separation the rules could perform. Results can be seen in Table 7.5.

7.5.3 Analysis

Investigating the effect of changing P1 allowed us to observe that the Jingle detection function seems to hinder overall accuracy. This can be explained because the corpus is largely not musical. Perhaps in cases of shows like Desert Island Discs, where the music impacts the segmentation, having the ability to toggle the value of that parameter could still improve results overall. Looking at maximum transition times and chapter length (P2), the larger transition times seem to be linked to better results. This might be because increasing the minimum size of a chapter prevents the algorithm from over-annotating.

In terms of detection threshold (P3), using the 70% quantile as a threshold for what the algorithm considers a "significant difference" between frames returns more accurate boundaries. This is hypothesised to be linked to the sequential nature of these rules. If more boundaries pass through this initial discovery phase, more go through the following set of rules. By filtering out false positives returned by the discovery phase, the rules act as an effective

"accuracy" barrier: lowering the detection threshold narrows the set of false negatives, and widens that of false positives.

The *omega* configuration (no jingle detection, 70% quartile threshold for significant differences between frames, and 14 sec transition length) performs best at a 0.001 threshold for tagging in the sound recognition model. In this configuration, pod-CLIPR performs on a par with, if not better than, the average expert in terms of IAA.

Omega performs particularly well on conversational programs (e.g. The Frank Skinner show ($\kappa = 0.857$), ZOE Science and nutrition ($\kappa = 0.772$), Test match special ($\kappa = 0.714$), The rest is politics ($\kappa = 0.706$), The infinite monkey cage ($\kappa = 0.697$)), and particularly poorly on the show Desert island discs ($\kappa = 0.458$). This is thought to be because of a lack of music recognition when using *omega*.

From this evaluation, pod-CLIPR is shown to be an algorithm that assists, but cannot fully replace a creator's hand. Indeed, it could not replace it even if it had perfect accuracy with respect to the POD 49 corpus: as shown in Section 4.B., there is no perfect agreement amongst expert producers, only plausible and implausible chapter suggestions, with the segmentation of podcast audio depending not only on the content but also on the professional in charge of the segmentation.

7.5.4 Discussion

How automatic chapterisation solutions perform in comparison to expert podcasters is of great importance to understanding how AI can be used in an assistive fashion for podcast production. Investigating such systems requires

two steps - the creation of a baseline dataset, and an evaluation of a system against it.

The system presented in this study demonstrates that an audio-domain solution for automatic chapterisation is not only conceivable, but can yield results on a par with expert human annotators. pod-CLIPR offers a believable segmentation. The results of the cross-validation tests showcase how robust the results presented are. Beyond offering a reasonable approach to segmentation, pod-CLIPR sets a precedent; a "perfect" chapterisation does not exist, but a system that prioritises accuracy over precision enables plausible chapter markers to be proposed. Example outputs can be accessed online [2]. This analysis has highlighted the necessity for an automatic chapterisation solution to be integrated within a flexible user interface, that enables users to toggle certain rules (e.g. jingle detection) and fine-tune the proposed chapter boundaries.

At the moment, pod-CLIPR is based on a non-specific, sound recognition model (Kong et al., 2020), and subsequently maps its results on a set of 10 tags chosen by podcast producers to best describe audio in their work. It flows from this observation that a dedicated podcast-focused sound recognition model could be trained to focus on these features (or others) to forgo the mapping stage of pod-CLIPR.

About the set of rules and parameters examined, the customisable nature of the algorithm should be highlighted. The rules and conditions tested here were set iteratively, and performed well when evaluated; however other rules could be introduced to cater to more specific end conditions (e.g. recognise a particular sound effect in a comedy show that notes a transition, or focus

[2] https://annotated-podcast-corpus.glitch.me/

more on environmental noise and effects tags for a nature documentary). More than validating a specific algorithm, this validates the methodology of using sound event recognition in combination with a rule-based algorithm to produce plausible segmentation.

In this study, pod-CLIPR performed particularly well on shows such as The Frank Skinner Show, ZOE Science and Nutrition, The Rest is politics, The Newsagents, Leading, Test match special, and The Infinite Monkey Cage ($\kappa > 0.65$, high agreement with POD 49) – this spans across various genres, from politics, to news, sports, science and tech, and comedy. The lack of fiction is indicative of the test dataset used rather than of poor performance in this genre.

Through this study, the benefits of using the audio domain for segmentation became apparent: a lot of segment queues were audible (jingle, sound effects, long pauses, changes from monologue to conversation, different recording environments), thus chapters can be inferred without using a Large Language Model, or any other form of speech based thematic analysis. In terms of efficiency, the PANN-APR model used was trained for 3 days on a single card Tesla-V100-PCIE-32GB (Kong et al., 2020).

As this is a pre-trained model, no further training was necessary, but this initial cost should be taken into account when looking at efficiency overall (this model might not be particularly greedy, but others that could be used in place of this in similar methodologies and algorithmic structures might be, and therefore influence the perceived performance of the system).

On average, this PANN-APR model took 35s to return labels with no threshold for tag recognition over 527 classes at a window of 1024 samples

for a 5-minute file. The labelling threshold only influences the returned values and time taken to print out the results. This threshold was shown to influence minimally the output of pod-CLIPR. Analysing the the quality of output, energy costs and processing time of different chapterisation solutions could enable creators, businesses and researchers to choose adequate automatic segmentation solutions.

Douwes et al. (2021) proposes a framework to evaluate the environmental impact and efficiency of AI audio generation models. Using the same platform[3], the model is estimated to require 1.6 kWh to train, which is relatively low, especially compared to other tasks in the audio domain, like generation. However, Douwes et al. (2021) concludes that comparing models based on imprecise estimations is flawed, and argues that real energy costs should be recorded systematically for new models. Instead, it uses these estimations combined with MOS for each model investigated to create a Pareto space [4] to compare models. A similar Pareto space could be created with other chapterisation solutions (e.g. including an LLM model for thematic analysis), to maximise the quality of output, and minimise energy costs, time, etc. This would enable more comprehensive comparisons for creators, businesses and researchers to choose adequate automatic segmentation solutions.

Limitations

The impact of changing the architecture of pod-CLIPR was discussed briefly and rules on the results presented, but perhaps, equally important is the nature of the test dataset used. POD 49 lacks representation in some genres

[3]https://mlco2.github.io/impact/
[4]https://www.sciencedirect.com/topics/engineering/pareto-optimality

like fiction, and of narrative podcasts (Berry, 2020) in general.

The use of Accuracy and Cohen's κ to draw conclusions from the data collected, although justified, makes comparisons with the performance of other algorithms relying upon traditional IAA metrics such as F1, p and r more complicated. However, taking into account TN in this analysis as well as expected agreement, tackles issues that would have otherwise limited the work (e.g. scoring poorly when agreeing on lack of boundaries for extended periods of time, scoring better when the test subject is over-annotating etc.) POD 49 was devised using block design; this means that each annotation was only compared to one other. This method, in comparison to a repeated measures design, allows the generated dataset to be larger, but the results are not as representative of how the group of participants would have annotated each segment.

Future work

Once segments are created, it is important to consider how they would fit in larger distribution systems. At the moment podcasts support chapters inconsistently Chelsey (2021). Investigating format and standardisation of innovation in the podcasting industry should prove invaluable for producers and listeners alike.

Additionally, the system evaluated follows a hybrid model comprised of a neural network and a rule-based algorithm, but the rule-based portion could be replaced by another layer of machine-learning, circumventing the need for rules. How would this purely AI-driven solution compare to the hybrid approach presented in this chapter?

7.6 Summary

The evaluation of pod-CLIPR was performed by comparing the output of the algorithm against a benchmark, however alternative methods should be considered, like using a MOS (Douwes et al., 2021), which would allow us to determine the value of an assistive chapterisation tool as perceived by its users.

I look forward to other chapterisation options being evaluated against POD 49 and pod-CLIPR. De facto, pod-CLIPR is outperforming other segmentation solutions, because there is no other repeatable evaluation of an audio chapterisation tool available.

Through the programming and subsequent evaluation of pod-CLIPR, additional answers are brought to the following RQs:

RQ 3: Why is modular podcasting attractive to creators, and how can it be facilitated without making the production process more complex?

The performance of Podulr's underlying segmentation algorithm, pod-CLIPR was estimated to perform on a par with human experts, although the evaluation highlighted 1/ the need for some adaptability in the rules to best cater to different shows sonic lexicon, and 2/ that chapters are somewhat contentious by nature, so the results of pod-CLIPR should be easily amendable to fit each producer's specific expectations of segmentation.

RQ 4: What are the perceived benefits, risks, and costs of exploiting AI technologies for podcast production?

With a functioning automatic segmentation tool, the process of chapterisation could be eased, saving time and cutting costs for production companies. However, because a tool like Podulr still relies completely on human

input, it would require specific training. Also, on a computational level, going down the path of automatic segmentation naturally begs the question of optimisation of performances. This means that new models would be trained and tested, incurring unavoidable costs in research and development.

8 Podulr in Practice

8.1 Introduction

This chapter details the last phases of development and testing of Podulr, the modular podcasting app described in Chapter 3 and 6. The app's architecture is presented, including class diagrams, sketches, data structure, and libraries used. An initial version of Podulr was presented to 3 producers to gather feedback, and from these conversations, a beta version was put together and handed to producers for a final evaluation and "real world" use. This will bring to light the possible applications and limitations of pod-CLIPR through the Podulr GUI. Two use cases are described, each focusing on a different ways to use the tool as described in Chapter 6: as a catalogue manager (with *Inside Science*); as an editing assistant and creative tool(with *The Modular Book Club Podcast*).

This gives us insights into the next steps for Podulr, and how the app could be evaluated at different levels of the framework presented in Figure 4.1 in the future. With this exploration into these final steps of ISD through PD, their efficiency and relevance for a new media tool can be commented upon, providing the final components to answer the research questions set out in Chapter 1.

8.2 App Architecture

8.2.1 Set-up

Podulr is hosted on Glitch.com[1], a host and web code editor that enables easy and fast deployment of Node.js apps. This is well suited to rapid prototyping and agile web development as a whole, but specifically, to build a progressive web app compatible across devices that uses JS libraries and a mostly JS-based back-end (pod-CLIPR is equal parts Python and JS). At this stage, the wireframe and comments received at the prior round of interviews provide a solid footing to investigate possible libraries and individual elements that will be included within the final design.

One of the key aspects of the site is the editing window, featuring a waveform representation, that includes markers and segments. There are JS libraries that allow for such representations to be easily computed and interacted with. This includes packages such as Wavesurfer.js [2], Timeline.js [3], or P5.js [4]. In the end, the representation provided by Peaks.js is co-opted[5]. It is an open-source BBC dependency that includes metadata annotations for segments and markers, and is compatible with a wide array of file formats. It relies on an HTML canvas element to display an audio file. The resulting waveform can be accessed at different zoom levels and comes with pre-defined markers for cues and segments, all stored as objects (meaning that additional metadata can easily be added as Key/Value pairs) within an array. The

[1]https://glitch.com
[2]https://wavesurfer.xyz/
[3]https://timeline.knightlab.com/
[4]https://p5js.org/reference//p5.FFT/waveform- waveform method
[5]https://github.com/bbc/peaks.js

representation also features a movable playhead, and easy integration within a web Audio Context – meaning that it can be compatible with other web audio API processes (playback functionalities, other web Audio libraries etc.).

For audio-playing, playback speed, crossfades, and other nodes within the app's routing diagram, Tone.js[6] is used, for its more precise handling of audio elements than the standalone Web-audio API. This encourages a reflection on working formats within the app: the files uploaded cannot be compressed or loose information throughout the process. This entails a lossless approach, via buffers and media Blobs. Blobs are file-like objects of immutable raw data. They can be read and processed either as text or binary data, and allow web pages to include data not necessarily compatible with JS-native format [7]. Downloading audio that has been modified through the Web Audio API is a non-trivial issue. To tackle it, the data carried by media Blobs on the page need to be encoded as WAV files and made downloadable via a series of JS conversions. A simplified logic flow diagram is shown in Figure 8.1, which showcases the necessary steps to go from a file URL to a WAV media Blob.

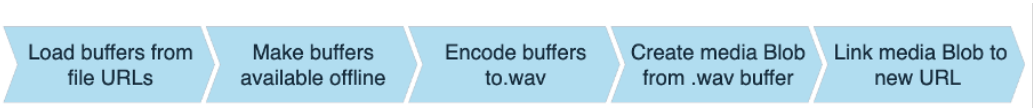


Figure 8.1: Process diagram for processing and downloading audio data through the Web Audio API

The metadata carried by the segments input by the user will be rendered as a text file, with chapter information in a format compatible with some of the most prominent podcast hosts (Acast, Buzzsprout, Spotify for Podcasters...), replicating the output format for cue markers of Audacity, for lack of

[6] https://tonejs.github.io/
[7] https://developer.mozilla.org/en-US/docs/Web/API/Blob

a more generalised file type.

Besides downloading considerations, the issue of file upload also emerged. The R&D nature of this app affords some ley-way regarding the upload solutions that would be acceptable for this project. Using Google Drive as an intermediary for hosting the files rather than directly accepting user's uploads, user files are collected through the Wget [8].

8.2.2 Initial Planning

Class diagrams are a useful tool within UML (Unified Modelling Language) to visualise the structure of a software system – mapping out its classes, attributes, operations and the intrinsic relationships between such objects. Figure 8.2 represents the initial class diagram drawn up for Podulr. There are 4 main classes (User Interactions, File Management, Audio Editor, Version Maker) relying on three key dependencies (pod-CLIPR, Peaks.js, Tone.js). This is a simplified diagram, which had to grow and adapt with the further iterations of the tool.

To bring more detail into the contents of each class:

- **UserInteractions** deals with all the graphical interactions with the page, including tracking mouse movements, drag-and-drop functionalities, keyboard links, and generally, event-listeners for interactions.

- **UserManagement** handles the upload of files and user data to the server, as well as retrieving and accessing the server output if a user logs in through a magic link.

[8] https://www.gnu.org/software/wget/

- **AudioEditor** deals with the representation and interactions with the sound files in the waveform editor portion of the app. This includes setting a web audio context loading buffers, playback functionalities, and dividing the buffers into chapters

- **VersionMaker** handles the creation and curation of different versions using the chapters set by the user, as well as the export functionalities for the audio and metadata output.

The initial design sketch was made with Canvas, using colours picked with a Coolors.co palette[9] to visualise and maximise potential contrast between background and element colours. The app's colour scheme lives between pink and blue, relying heavily on blurple (a mixed hue of purple and blue) – for its simple but playful esthetics, as well as acknowledging this hue has been used in tech-forward and innovative apps in the past years. The design was kept minimal, to fit the requirements gathered from prior interviews. Moreover, other producer considerations were integrated within this mock-up, such as the possibility to add or remove segments, and change playback speed or crossfade length.

[9]https://coolors.co/

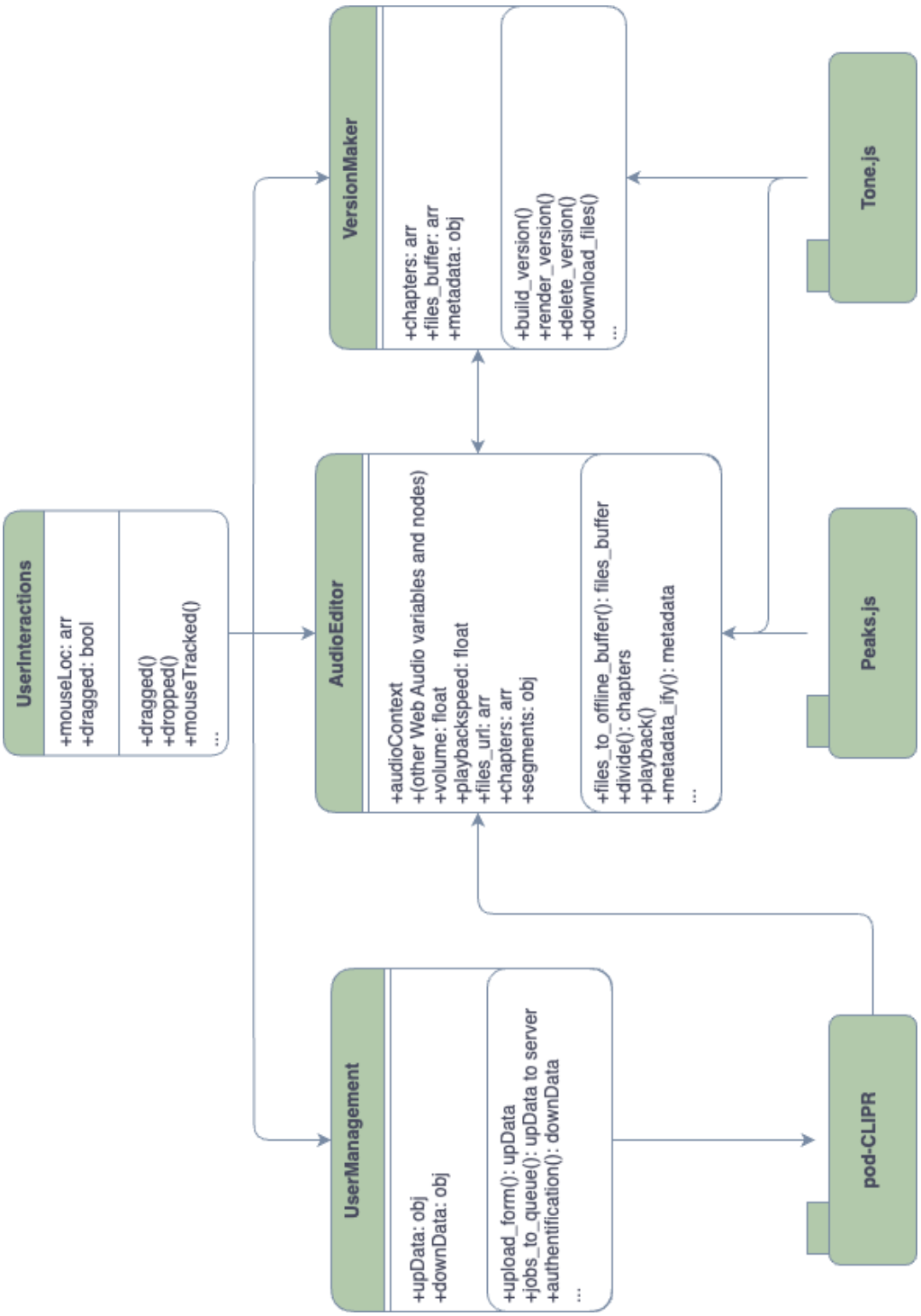


Figure 8.2: Initial class diagram of Podulr

The page is divided into two: at the top, a simple audio editor that enables

the user to interact with their audio file, placing, changing and deleting segment markers, and at the bottom, a version maker, that allows users to create and export various versions of their programmes. These two areas are divided through colour, with a soft gradient emerging from a central beam of "light". All the visual elements are made with .css, except the icons, which are royalty-free .svg files. The first design sketch is shown in Figure 8.3.

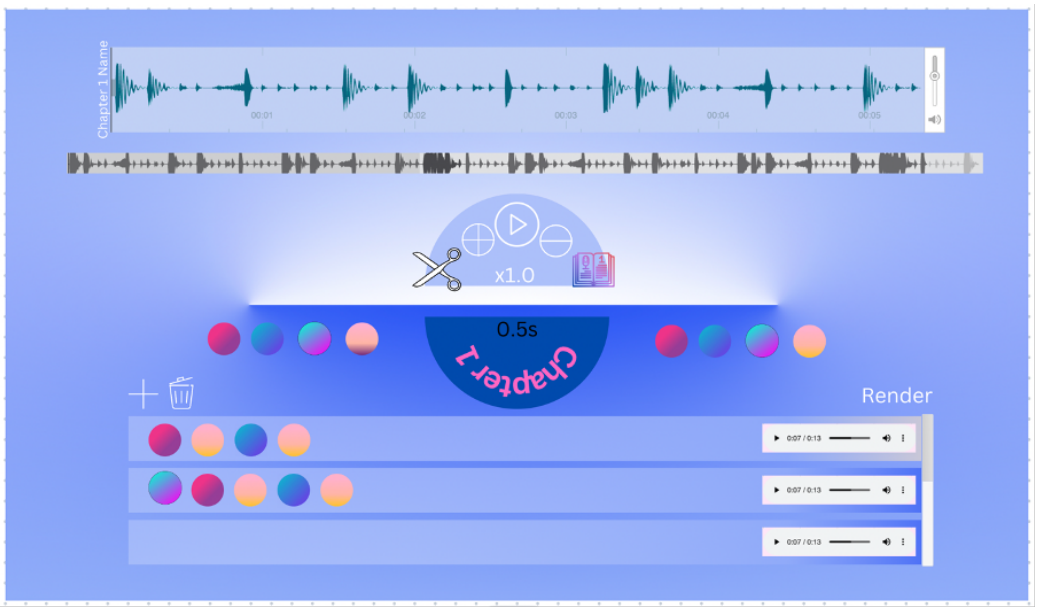


Figure 8.3: Initial design sketch for the main page of Podulr

In terms of structure, Podulr has a front end, client-facing side, and a back end, server-side divided amongst two distinct servers. Figure 8.4 represents how these interact. Server A acts as the interaction agent, passing and managing processing jobs to Server B, hosted on a university machine. Server B is where pod-CLIPR is deployed. This particular architecture means the project has access to far more disk space, with a completely customisable set-up. Where a solution like Amazon Web Services [10] could have enabled

[10]https://aws.amazon.com/

a single-server architecture, it would have incurred greater costs. This dual server architecture is used on other Music/Sound/Audio research software, such as Cocreate [11], Unmix [12], or Upmix [13].

Figure 8.4: Client-server architecture of Podulr

On the client side, the project is separated over two HTML pages, one for uploading files, and one with the audio editor seen in Figure 8.3. This editor can only be accessed if the user logs in via a magic link. JS and CSS scripts are broken into several files for legibility.

On the server side, Server A communicates user data that needs to be processed by pod-CLIPR with a second server, Server B. Server B sends requests for updates to Server A at repeated 2-minute intervals. The back end is comprised of a set of JS and Python scripts that deal with the download, processing via pod-CLIPR, upload, and sharing of processed data back to the client side.

8.2.3 Data Flow

The hidden data layer of Podulr is comprised of three different types of objects. *User objects* are written to user_session.json, and include information regarding each email address and their associated projects. *Project information* are saved to Jobs objects, written to jobs.json, and include the informa-

[11] https://cocreate.glitch.me/
[12] https://unmix.glitch.me/
[13] https://upmixai.glitch.me/

```
upData = {
    id: <project id>,
    emailAddress: <user email>,
    stampCreate: <date and time of upload>,
    origUrl: <URLs to user files>,
    isAnalysed: <boolean for user choice>,
    jingleDetection: <boolean for user choice>,
    idUser: <user id> (one per email),
    projectName: <project name>,
  };
```

Figure 8.5: Example project information going to university machine (client side to server side). <> denote placeholder values. Entries are strings unless specified otherwise.

tion related to each project that is passed down to the server for processing. Figure 8.5 is the template for Jobs objects. Finally, a *Podulr-Project Object (PPO)* is created in the back-end collating all the information necessary for a user to be authenticated to Podulr and access their processed data. Figure 8.6 is the template for a PPO object.

Server B involved pings Podulr to access the hidden data layer, gathering any new information written to Jobs.json or User_sessions.json. After processing any new jobs, it creates an associated PPO (Figure 8.6), communicates it back to the app, and sends an email to the user with a unique magic link to access the information contained in this PPO, triggering the authentication process.

```
PPO = {
    id: <project id>,
    name:  <project name>,
    idUser: <user id>
    metadata:{
        aiAnalysis: <boolean for user choice>
        jingleDetection: <boolean for user choice>
      }
    ,
    layers:
      {
        fileURL: <URLs to user files>,
        nFile: <number of files>,
        dateUploaded: <date and time of upload>,
        tagged: <boolean for whether the files have been tagged already>,
        name: <file name>,
        clipr_results: <boundaries resulting from analysis>,
      },
    ,
  };
```

Figure 8.6: Example PPO going from the university machine to the client (server side to client side). <> denote placeholder values. Entries are strings unless specified otherwise.

8.3 First Impressions

8.3.1 Interview Planning

Three interviews were planned with producers (two independents and one BBC creator) when the first usable version was available to check in with the app's development. The method used was similar to the previous set of interviews. The conversation took the form of a 30-minute Zoom semi-structured interview following a presentation of the tool and live demonstration. The interviews were not recorded, but thorough notes were taken throughout. These notes were then ordered within the table framework shown in Figure 6.6.

The first usable version of Podulr was rudimentary, but showcased the potential of the tool adequately. It included the following features:

- Uploading multiple user files
- Authenticating via magic link
- Opting in for automatic segmentation
- Responsive laptop version (tested on tablets but not phones)
- The editor has basic features (e.g. play/pause, zoom in/out, scroll, playback speed, volume, add/change/remove segments)
- The version maker has basic features (e.g. playable and draggable chapters in chapter bank, export to WAV, render as a media player to enable download, metadata text file download)

8.3.2 Procedure

Podulr was first presented to the participants, including possible use cases as they were discussed in the last set of interviews, and a live demonstration. The live demonstration was conducted using an <10 minutes extract of Siege, a BBC Radio 4 drama with very clear chapter delimitations [14]. This file, as well as a captured version of the live demonstration recorded for an industry presentation can be accessed in the supplementary material. This round of interviews was driven by a small number of questions, as the main purpose is to gather feedback and comments on an existing tool. The questions were as follows: 1/ Are there any other features you think would be useful for this tool? 2/ How do you feel about the way this tool exports metadata? 3/ What particular applications do you envision for this tool?

Additionally, participants were queried to know if they would like to fully beta-test the app and contribute to the use cases presented in Section 6.2.3.

8.3.3 Results

Notes were taken on digital post-its through each interview, subsequently ordered to fit the table in Figure 6.6. The anonymised data can be accessed in the supplementary material. A portion of the conversation was aimed at noticing current bugs or issues with Podulr. Notably, there were issues with Peaks.js' waveform representation linked to the user interactions with segment markers, the draggable playback speed rate and crossfade times did not print values in a consistently rounded format, and the chapter bank did not print out chapters in a coherent order.

[14] https://www.bbc.co.uk/programmes/m00146p6

In terms of changes required by the users, the following comments were made, followed by the modifications made to the code as a fix:

- The UI needs more/clearer icons, mainly: volume, zoom in and out, crossfade, add or remove segments and chapterise. So, the missing icons are added, and :*hover* or *:active* properties are attached to these objects to improve the UX.

- There needs to be a clearer distinction between the zoomed view and overview of the waveform, and the colour/font scheme of timestamps and chapter names need to be more legible. So, a border is added to the zoomed view and overview, as well as a different background colour. The colour and opacity of elements like text and markers are also changed to improve contrast and overall readability.

- The chapters' names need to be editable. So, an event listener is added, to respond to the user right-clicking on segments, spawning a dialogue box so the highlighted chapter can be renamed.

- There needs to be clear instructions for the user. So, dialogue screens are added to both the landing upload page and the editor page. The one on the landing page details upload requirements and the overall process, while the ones on the editor page are small tutorials summarising the functionalities behind each button. It can be accessed by clicking on an information button by the title of the page, or automatically when a user first loads a session.

- The site needs a load screen so the user does not just wait around for up to a minute while various elements appear on the site. So, an overlayed

load icon is added, that only fades once the audio buffers and graphical elements have been loaded.

- A total number of chapters should appear by the waveform overview for long files. So, a preview of chapter numbers is added

- More fine-tuning of the metadata should be accessible. So, a cogwheel icon is added to the version-making section, that enables the user to change metadata detail (author, project name, title of versions created)

- A mobile version should be implemented. So, the app's responsiveness is adapted for mobile. A full mobile version isn't developed for lack of time and resources.

- The question of whether the user could upload existing cue markers is raised.

8.4 Use Cases

8.4.1 Beta Version: Summary of Features

The final version of Podulr was a summation of the interviews and workshops held before. It took into account key feedback received in the last stage of development, and although the app would not sustain the combined traffic of many users, it was stable enough to be released in the real world, without the need to be constantly monitored by the developer. In addition to the features mentioned in Section 8.3.1, and the changes Section 8.3.3, some final elements were added to the design:

- A “download all” button, which downloads all the audio files and metadata at once.

- A key bind for the “a” key once chapters have been rendered, which creates a different version track per chapter.

- An information button to trigger the tutorial windows again

- An additional page disclosing information on Podulr’s use of AI, including the motivations and a link to the model and training dataset.

Because of recent changes in Google Drive security measures for automatically downloading files from the cloud, a file size limit of 100 MB was imposed on Podulr. Figure 8.7 and 8.8 show and detail screen captures of the app at this stage.

Figure 8.7: Screen capture of the landing page of Podulr. 1/ The file upload form for users to put in the details of their project; 2/ The tutorial pop-up which tells users of current file upload limits, and points to a contact email; 3/ Credits and affiliations, as well as a link to a statement regarding the use of AI in Podulr.

Figure 8.8: Screen capture of the main page of Podulr after a user is authenticated. 1/ Name and button to trigger tutorial pop-ups. 2/ Wave editing zone, with movable playhead and re-nameable chapters; 3: Control wheel for controlling playback options, adding segments, chapterising the project, and changing crossfade durations for exported versions 4/ Draggable chapter bank of segments created 5/ Version maker, with drag and drop capacities to put in and reorder chapters. Render button to export metadata and audio files. Cogwheel button to further customise the metadata. Download-all button available at the bottom right.

8.4.2 Podulr Within the Six Tensions Framework

In Chapter 2, the Six Tension Framework was proposed to help define podcasting, as an ever-growing concept. This framework sees podcasting at the convergence point of pairs of ideas in tension with one another, and podcast as a central idea constructed from the negative space formed by these elastic boundaries. Moreover, it was postulated that regardless of the future forms of podcasting, and the way it is influenced by other media, the Six Tensions Framework would still stand when defining what podcasting is and how it is experienced by its audiences.

To further illustrate this proposed framework, let us examine Podulr and the modular podcasts it can create under its lens. By using modular podcasting to create alternative versions of the same programme, all the pairs described are involved in different ways:

Personalisation and automation. Through offering additional personalisation for the users, modular podcasting must concurrently rely on additional automation to facilitate production and distribution. This automation prevents the degradation of the creative process - where a creator would have to deal with larger quantities of content - or of the consumption mode - whereby formats grow more complex and audiences are separated from content through mandatory interactivity.

Interactivity and immersion. To access customised content, listeners must interact with their podcast or podcast platforms. This requires a level of engagement that could disturb a more typical "passive mode" of listening. However, personalisation can reinforce feelings of immersion, by integrating the user and their environment in a product. The balance struck by modular

podcasting is to enable this personalisation, but highlight the need of minimally invasive interactions with the content (e.g. asking to choose versions before listening, relying on user preferences, or relying on recommendation algorithms).

Uniqueness and universality. Modular podcasts can cater to many different listener expectations, and by doing so can provide unique listening experiences. This uniqueness might hinder the universality of the media created. For instance, in the case of local news podcasts adapting to feature only relevant stories to a listener's geographical position, listeners might miss out on important stories from across the country. This can only be mitigated through careful editorial usage of these technologies. As reinforced by participants through their interviews and workshops, interactivity should be driven by a narrative goal.

Current audience and possible demographic. In the case of modular podcasting, "Current audience and possible demographic" is related to "Uniqueness and universality", as the uniqueness of the content will reach new demographics, either through a draw to the podcasts themselves or the underlying technology showcased, but also appeal to current audiences by simply catering shows to the specific requirements of existing audience members.

Mainstream and independent productions. There is a risk that modular podcasting used by large broadcasting corporations or networks could drown out smaller independent productions through the sheer volume of alternative content created. To prevent a shift amongst this pair of concepts, means of producing modular podcasting (both relevant tools and

formats) should be available to and usable by all podcasters regardless of their place of employment.

Art and technology. There is an intrinsic relationship between the intricacy of a creative process and the technological sophistication required to make, share, and preserve what is made. By enabling new creative projects, modular podcasting also calls for significant changes in the way personalised online audio is distributed and saved. The exploration of modular podcasting should come hand in hand with more technically rooted research to maintain podcasting's established compromise between embracing innovation and making, sharing, and preserving compelling pieces of art.

8.4.3 Evaluation With Creators: Onboarding Process

Podulr beta (c.f. Section 8.4.1) was shared with two producers to use and evaluate.

Jana is a producer for *Inside Science*, a BBC Radio 4 programme that presents various scientific topics in each episode. Jana was recruited for this evaluation through word-of-mouth via internal BBC communication channels, and had not participated in any other steps of this app's development. For this test, Jana was using episodes of *Inside Science* across a few weeks, in the context of using Podulr as a catalogue manager (to document and annotate the various segments present in their magazine-style show).

Jenn is a former bookseller and a current reviewer, podcaster, and editor with Riot New Media Group, also producing her own shows, focused on culture, specifically literature. Jenn had contributed to every step of the iterative development process, and therefore already has a working knowledge

of the research and Podulr. For this test, Jenn was using a book-related (Arts and entertainment) show, in the context of using Podulr as a creative tool, and editing assistant.

These two test users not only represent different goals for using Podulr, but also different connections to the research; by including Jana, who never heard of the tool until this evaluation stage, and Jenn, who was very familiar with and had contributed to Podulr, the ISD process finishes on a step that is both self-reflective (linking back to all prior phases), and outwards facing (bringing in an example of new users).

These evaluation interviews were conducted individually to cater to the different use cases represented by the profiles of the creators involved. Jana received a Zoom onboarding for 30 minutes, where the project was explained, and Podulr was demoed with an example file. Jenn received an email prompt, accompanied by a link and detailed instructions about the specificities of the beta version (file size limit, upload requirements etc.). After this, both were left to use Podulr for their project unsupervised, and feedback on any concerns or takeaways.

Through these interactions and the following usage of Podulr on two different projects, the following questions were answered:

1. How long did these creators use Podulr for?
2. How did Podulr perform for these specific projects?
3. How did Podulr integrate itself within these creators' normal workflow?
4. Is there anything to change or add to a next version of Podulr?
5. How do the creators feel about the way Podulr uses AI?
6. Would these creators use Podulr again?

8.4.4 Use Case 1: *Inside Science* (Podulr as a Catalogue Manager)

The primary goal of Jana was to annotate episodes of *Inside Science* with chapter information, to constitute a file base of segments accessible by the production team at later dates. This could be so that composite shows can be easily created (grouping various segments from different episodes related to a similar topic in one themed episode), so that segments can be shared across BBC productions, or simply, to maintain a detailed archive of the show.

1. How long did this creator use Podulr? Jana used Podulr for 30-minute sessions weekly for three weeks.

2. How did Podulr perform for this specific project? Jana opted for automatic segmentation using pod-CLIPR. Podulr missed a few chapters, but overall identified correct segments. Chapterisation is particularly relevant to *Inside Science*, as it relies on clear segments with oftentimes different recording environments.

3. How did Podulr integrate itself within this creators' normal workflow? In Jana's words, Podulr "*integrates itself in a workflow fairly naturally, as part of the post-production process*". There was a 5-10 minute learning curve, but shortly after, this creator began working on the uploaded files at 2x speed. After this period, Jana used the tool efficiently and estimated it would take her 10 minutes to annotate a full 30-minute show once she grew familiar with the controls.

4. Is there anything to change or add to the next version of Podulr? There were a few compatibility issues with Podulr. First, this

creator's default browser (Goodle Chrome) returned a rare error about loading offline audio buffers. The creator then switched to Edge, and the issue seems to be resolved. In terms of file formats, BBC producers tend to rely on Dropbox, which means Google Drive is impractical as a middle-man for file management. Moreover, if Podulr is to be adopted by more BBC producers, its reliance on Google Drive would make it hard, if not impossible, to be approved by the Information Security Department.

The lack of a progress bar under the load icon before the main page (Figure 8.8) is displayed was highlighted. For a 30-minute file, this load takes up to 45 seconds, which can be disconcerting for users if they are simply left looking at a rotating load icon.

An issue with Peaks.js appeared in one of the sessions, where segment markers when rapidly moved back and forth "disjoint" from the previous chapter (e.g. for two chapters, Chapter 1 with boundaries A;B, and Chapter 2 with boundaries C;D, B and C overlap as a cue marker. In this bug, B and C detach and can be moved independently, creating an additional chapter between B and C accidentally).

Finally, Jana raised the need for a "save button" on the page allowing one to preserve the data modified by the user for the project and access it at a later time.

5. How does this creator feel about the way Podulr uses AI? Jana had "*no qualms against this use of AI*". She acknowledged that a tool like Podulr could be useful for other shows across the BBC, especially longer shows that rely on segments and are broadcast live before being adapted as podcasts for the Sounds app. In her opinion, many parts of the production

workflow can be aided by AI, as long as human eyes are involved in the process, it could be really helpful to use these tools in more assistive, or administrative capacities.

This creator seemed satisfied with the amount of transparency regarding the underlying systems, even without having been a part of any ISD phases.

5. Would these creators use Podulr again? This producer was interested in integrating a more BBC compatible to their workflow more permanently.

8.4.5 Use Case 2: *Modular Book Club* (Podulr as a Creative Tool and Editing Assistant)

The primary goal of Jenn was to create different versions of a book club podcast episode, to produce "spoiler" and "no spoiler" versions of the same episode, in order to cater to different audiences.

1. How long did this creator use Podulr? Jenn used Podulr for around 3 hours across a few days, but working on a single project.

2. How did Podulr perform for this specific project? Jenn reported the following:

> *"Like all AI, it got close but often needed correction, and I actually ended up turning off the "auto-generate chapters" option the second time I uploaded because it was easier than correcting the auto-marked chapters."*

3. How did Podulr integrate itself within this creator's normal workflow? This creator used Garage Band for an initial edit, and uploaded the the finished episode in the post-production phase. They deemed that this initial upload step was "easy", although there was an initial bug with

authentication I had to fix so that they could access their file. This was due to a connectivity issue between Glitch.com and the backend systems.

4. Is there anything to change or add to the next version of Podulr? Beyond the issues flagged with authentication, which would require disassociating from Glitch.com to tackle, Jenn flagged some other problems encountered. First, she reported almost losing her project by accidentally pressing the "back" button, but being able to load the data back by clicking "forward". This highlights the need for a local save button.

Jenn also encountered an issue with the Peaks.js representation. Similar to Jana's experience, after a while handling segments, a phantom chapter in between two set segments appears when moving a cue marker repeatedly.

The chapter bank also struggled to open a buffer of 20 chapters. This was because of the architecture of the project, which relies original files' offline buffers to be segmented into different sections.

Finally, some cosmetic issues were highlighted, mainly that the text could be more legible, and that the chapters in the chapter bank would load in an ascending order systematically.

Jenn had very positive comments to make regarding the chapter naming feature, colour scheme, and tutorial windows.

5. How does this creator feel about the way Podulr uses AI? To this question, Jenn replies:

> *"I hope that the AI is observing privacy guidelines and that the content it learns from is protected according to best practices."*

This is in line with prior conversations with creators. However, the use of "hope" rather than "know" means that the statement on AI use, training

data, and research context of Podulr could be made even more apparent to the user.

6. Would these creators use Podulr again? Jenn was enthusiastic about future uses of Podulr, saying

> *"I really like the idea of being able to produce multiple episodes with simple chapter swapping/inclusion/exclusion, and hope to get to play with it again."*

8.5 Summary

Through these final conversations, the scope and limitations of Podulr became apparent. As expected from the evaluation carried out in Chapter 6, the performance of Podulr was acceptable to producers as is, particularly in the case of *Inside Science*, a magazine show with definite segments. For Jenn's Book Club episode, Podulr was helpful as an editing assistant, but the thematic nature of the changes within the content could explain the decision to turn off the automatic chapterisation. Through these two use cases, the research questions investigated in this thesis are partially answered:

RQ 2: How can we feasibly create and distribute immersive and personalised podcasts?

For both use cases, Podulr was integrated as a post-production tool. Although its intended purposes are different, its practical uses are similar (upload files>segment files> export segments and metadata)

These examples did not go all the way with distribution, so it remains to be examined how these projects would be distributed to audiences. For instance, the "spoiler"/"no spoilers" version of Jenn's Book Club podcast

could be simply hosted as two different files. In the case of Jana's Inside Science episodes, the goal was never to host different versions of a programme, but more so to assist producers in making composite works, or simply better archive their existing shows. This leaves the way that Podulr deals with formatting interactive media simply as a suggestion. The output metadata associated with the versions created on the UI uses simple syntax but isn't by default compatible or readable by podcast platforms.

RQ 3: Why is modular podcasting attractive to creators, and how can it be facilitated without making the production process more complex?

Podulr's solution to facilitating modular podcast production is deemed satisfactory in the two use cases, however, some more bugs were reported. Primarily, the reliance on third-party services and libraries seems to hinder the tool when used in the real world. For example: Glitch.com isn't always reliably making connections with other servers; Within this last phase of ISD, Google Drive changed its security measures, which made the wget method return errors for any file above 100 MB; Peaks.js has limitations when it comes to displaying and modifying segments which can greatly affect the overall user experience on Podulr.

A solution would be to detach the app from these components, requiring much software development work, but stabilising the product as a whole.

RQ 4: What are the perceived benefits, risks, and costs of exploiting AI technologies for podcast production?

Having the option of using AI for automatically segmenting chapters was appreciated by both participants, however, Jenn came to the conclusion that

it was faster to insert cue markers by herself, rather than to have to tweak the chapters that pod-CLIPR suggests. This speaks to the need for customisation and versatility of the AI tools offered to creators. In the case of Jana's *Inside Science* work, it seemed that AI could help speed up some already necessary, but repetitive and arduous tasks.

In both cases, the participants seemed comfortable with the use of AI in Podulr. Jenn's comment on the matter raises the issue of transparency: even if Podulr was transparent regarding its AI components, it seemed that the mention of its research context, training dataset, and underlying model are not prominent enough to be systematically understood by users. Jenn already knew of the ways in which Podulr uses sound recognition, through their prior involvement with the project, but it can be hypothesised that a new user might also miss the AI systems disclosure statement available through a link at the bottom of the page (c.f. Figure 8.7). This underlines the importance of not only being transparent, but encouraging users to reflect on AI processes actively.

9 Discussion & Conclusion

9.1 Discussion

The medium of podcasting has evolved over the last twenty years to stand alongside more traditional media such as music, radio, and television, in terms of the amount of content being produced, its reach, and its economic importance (RAJAR, 2023; Beniamini, 2023). The purpose of this PhD was to bring new insights into the inner workings of this flourishing medium – insights that may benefit researchers and industry professionals alike, and could provide grounding for further research and future innovations in the medium.

Beyond showcasing the missing links within the field of podcasting research through a thorough analysis of prior literature, this thesis proposes definitions and frameworks informed not only by context, but by a cohort of podcast producers whose opinions are collected repeatedly through interviews and workshops. This participatory approach is not only advantageous to understand the current practices of professionals, but also to justify the creation of tools for podcasting. Principles of PD for developing new multimedia tools (Markman, 2012; Meixner et al., 2017) structure the methodology of this thesis, and the pursuit of three research aims:

RA 1. Mapping the habits and expectations of podcasters

RA 2. Exploring the peculiarities of immersive and personalised podcasting

RA 3. Investigating and documenting an application of participatory design to the development of an AI-driven Next-Generation Podcasting tool, all the way from conception, to functional software

Each of these aims is tied to the target output of an end product. By having a functional tool as a goal, the consolidation of each of the ISD phases (Figure 4.2) results in a scientific justification for design requirements and decisions. And, through this project-specific justification, larger questions related to the fields of podcasting and personalised and immersive media are addressed (RQs). In this section, results presented in prior chapters are discussed, focusing on each RA independently, and on the creation and evaluation of an automatic chapterisation method for a modular podcasting system.

9.1.1 On the Topic of Podcasters' Habits and Expectations

"Guerilla Media" was an alternative term given to podcasting by the journalist who first reported on the phenomenon (Hammersley, 2004). Although the now commonplace "podcasting" ended up being the term adopted, "guerilla media" provides an accurate representation of what podcasting felt like at its inception: independent, irregular, and somewhat orthogonal to "mainstream" broadcasting. Podcasts past and present have been characterised as an example of independent media (Markman and Sawyer, 2014), but it is also

the case that many are now fully integrated into mainstream media, at the centre of a billion-dollar industry (Grand View Research, 2021). Podcasts are no longer only produced by amateurs or radio companies, but also by podcasting networks, and global corporations like Spotify and Apple.

The independence, or perhaps the impression of independence, showcased by the wide variety of podcasts produced every single day should not prevent from drawing conclusions from the medium as a whole. It is the hypothesis of this research that a formal generalisation of podcasting (whereby the medium is defined, its workflows are investigated, and the perspectives of actors are analysed) would allow the format to flourish, in turn providing justification and a framework for innovative research in the field. Particularly, it provides the necessary justification for the development of a modular podcasting app, Podulr.

After proposing a definition of podcasting and a framework for innovation, the first ISD phases conclude in seven relevant recommendations being made for designing podcasting tools [1]. Although these are not formulated according to software requirements specifications (SRS) document standards (IEEE Standards Association, 2018), they do answer some key questions that would enable the creation of such a document. The workflow generalisation, combined with the prior work done to create a framework (Six Tensions Framework) and the user goals gathered during the following workshop can be

[1]Podcasters 1) are interested in delivering better, more immersive and engaging experiences to their listeners, 2) have an already-complex workflow comprised of a wide range of tasks and skills, 3) are looking for ways to simplify this complex production process, 4) want their production tools to be efficient, compatible, useful, comfortable, good value for money, and no- code, 5) are looking for ways to adapt their podcasts to their listeners, 6) are concerned with accessibility and reaching as wide an audience as possible, and 7) are wary of unethical uses of AI in media

combined into a description of purpose, including definition, references, scope and overview of project (IEEE Standards Association, 2018, Figure 8). It also brings to light user characteristics, and some of the more general aims, objectives and constraints attached to a new podcasting tool.

It is the intention of this thesis to bring new insights into the inner workings of this flourishing medium – insights that may benefit researchers and industry professionals alike, and could provide grounding for further research and future innovations in the medium. However, the lack of academic literature covering the specific information necessary for similar research has not prevented the industry from bourgeoning with new tools and solutions for podcasters, as well as new distribution platforms for the listeners. These tools might rely on a similarly thorough investigation into the context they emerge from, but the private nature of the information these advances in podcasting rely on further settles the medium in proprietary entrenchments, which contradicts not only the foundational principle of independence of the medium, and also contribute to the lack of standards and formalisation in the industry.

As an example of the emergence of innovative podcasting tools in parallel timelines to this research, the past four years (2020-2024) have seen the creation of many "participatory podcasting" platforms. This concept is discussed in Chapter 2 and 3 (as a technology being explored within the research). Specific implementations of this idea are also discussed, in the form of Stereo or Clubhouse. More examples of this can be browsed in the app store: Swell, Cappucino.fm, and Riffr, Leher, Angle audio, or Tin Can, just to name a few.

Even though there was an initial enthrallment for Clubhouse, a self-described audio-based social networking app, its popularity eventually declined. Its initial appeal, linked to the impression of exclusivity and real-time conversation format, struggled to sustain user engagement over time, as concerns from both audiences and producers grew. On the user side, issues related to moderation and privacy particularly impacted listeners' experience. On the producer's side, it lacked robust monetisation strategies, and content that benefited from the platform's specific social features. The combination of these aspects led to a slow withdrawal of users and investors, marking Clubhouse as a cautionary tale of the volatile nature of new tools for podcasting. This example showcases that a layered investigation (see Figure 4.1) into a software's applications is essential to gaining a full understanding of how a new media tool will be integrated and accepted by all relevant actors.

Identifying what NGP means for one of the involved actors is only the first step in steadily developing and delivering NGP tools. As has been shown through various examples and conversations with creators, ensuring the new forms of podcasting have an adequate amount of editorial interest is paramount to the wide adoption of a new format. And, just as 3D animation requires specific modelling software, so will NGP require bespoke tools to produce new types of podcasts. By following this creator-first approach, this thesis is concerned primarily with finding out what innovative podcasts creators wish to make so that appropriate tools can be designed and interest can be better guaranteed.

9.1.2 On the Topic of Immersive and Personalised Podcasting, Format, and Standardisation

Historically, podcasts have been linked to new technology. In 2003, the idea to distribute audio files via RSS feed was novel. The MP3 file format itself was only a decade old (Witt, 2015), and being tech-savvy was a requirement for any upcoming podcaster. Since then, there have been many improvements to the systems behind podcasting, including new audio codecs, and the automatisation of RSS feeds offered by third-party providers. This evolution can be related to improvements to radio broadcasting systems. The switches from AM to FM, and FM to DAB are compared by Lax (2017).

> *"In both cases, then, we find instances where broadcast engineers argue that, so obvious are the improvements offered by the successor technology, take up will be rapid and yet, when this fails to materialize, an impression of some bemusement suggests itself in the engineers' writings."* - *(Lax, 2017, p.34)*

In contrast, and as argued in Chapter 2, podcasting is not envisioned as a linear replacement of radio, but rather as a parallel sibling media that shares some of its attributes and content. So, does this remark hold for the transition from RSS-based to streamed audio, and streamed audio to NGP formats?

This thesis makes some suggestions as to possible formats for NGP, that cater for more personalised or immersive audio, linking content and metadata, much like other Object-based standards would propose (Oldfield et al., 2015). This seems to be the prevalent theory for sharing interactive audio. But, as explored by Lax (2017), is this proposal for a standard simply *"a mismatch between broadcasters' early expectations and the subsequent responses*

by listeners"?

In 2022, BBC director general, Tim Davie, said in a speech to the Royal Television Society[2]:

> *"Imagine a world that is internet-only, where broadcast TV and radio are being switched off and choice is infinite [. . .] Over time this will mean fewer linear broadcast services and a more tailored joined-up online offer." - Tim Davie*

This mirrors the early enthusiasm from the BBC to switch to DAB (Lax, 2017) – so is the race to standardising personalised media bound to a similar slow adoption, turning the solution to what might seem a prevalent issue needing immediate resolution, into an agonisingly slow change in production and distribution pipelines? Evens (2020) comments on the slow integration and implementation of digital radio:

> *"All too often media objects are researched in isolation from the ecology in which they are produced, circulated and consumed, which is leading to one-sided analyses or oversimplified conclusions." - (Evens, 2020, p.516)*

And to a certain extent, this thesis *is* one-sided – although taking this creator-centric approach is thoroughly justified, it ignores the opinions of other actors, may that be distributors, audiences, or advertisers. But, as stated in the methodology, rather than considering this work as standalone, it is simply a step towards a final resolution. This is why rather than drawing full conclusions regarding formats, this research simply proposes hypotheses, or avenues of further investigation, which would fit the analysis carried out from a one-sided outlook.

[2]https://tinyurl.com/guardianTimDavie

The chameleonic nature of podcasting together with the growth of interest in immersive and personalised technologies is substantiated via the various studies carried out in this thesis. The subsection of technologies investigated as part of this survey of NGP is only a fragment of all the capabilities that could be integrated within such a concept. The restrictions imposed by the inherent resources of a PhD restricted the scope of the enquiry. However limited, this highlighted some technologies that were particularly interesting to podcast producers for NGP. Modular media might be seen as simply another facet of personalised media (Object-based, Enhanced, Adaptive, or Flexible - c.f. Chapter 3), but the term modular is especially useful as it carries the meaning of changing sections of a programme to fit different audiences' expectations, which is not referred to specifically by the other foregoing terms.

9.1.3 On the Topic of Developing AI-Driven Tools for Creators

AI was not a goal in of itself – rather, this work takes the approach of evaluating how multiple types of new technology that can improve or change the way podcasts are made or distributed could be integrated within NGP. AI only represents a fraction of the technologies observed. Throughout the process, the use of AI is driven by the interest of producers in the technological feats these models enable. The interest in different models really varies depending on their potential applications and use cases. There is a definite wariness triggered by ethically problematic AIs, mostly in the generative realm. Just as other traditional artists are concerned with the boom of AI synthesis (im-

age generation for visual artists, LLMs for writers, etc.)[3], so are podcasters, as they see AI as a potential threat to their livelihoods (c.f. Chapter 5). That podcasting has grown as an industry into a system being able to remunerate its actors solidifies the uniqueness and perceived value of the human work performed in this context. The generative AI debate goes beyond copying vs. inspiring (Yin et al., 2021), but strays onto more philosophical concerns that come hand in hand with conversations around deepfakes.

This PhD takes the approach to only work with technologies that producers are comfortable with – which situates the models of interest in the analytical sphere, rather than the generative. Sound recognition, the concept used for the tool created through this research, is remarkably unproblematic when used on the producer side and not the listeners. On the listener's side, there could always be issues related to privacy or concerns related to surveillance, but when it is used as a production assistant, ideally, only files that are completely owned by the rights-holding parties are processed. Of course, this cannot be systematically verified, but the users can be asked to confirm that they own the copyrights of the files they upload. Within the particular pipeline explored, the results of the sound event recognition occurring are not even communicated to the producer, only the output of the segmentation inferred from the tags. In theory, this prevents malicious uses of the app and technology.

This cautious approach and optimism are immediately derived from the conversations held with podcasters as part of the ISD process. This design method not only ensured the tool was adequately justified and could be adopted easily by users, but also targeted some of the inherent biases present

[3]https://www.humanartistrycampaign.com/

in new technology and AI research. Birhane et al. (2022) highlights the need for a change of methodologies when developing AI technologies :

> *"The field of artificial intelligence is faced with the need to evolve its development practices— characterized currently as technically-focused, representationally imbalanced, and non-participatory—if it is to meet the optimistic vision of AI intended to deeply support human agency and enhance prosperity." - (Birhane et al., 2022, p.1)*

Moreover, Birhane et al. (2022) highlights the possible use cases of PD within AI, to help with algorithmic improvement, methodological innovation, or collective exploration, along axes of empowerment and reflexive assessment. This is to distinguish performative PD from public and social-good-driven PD in AI. In the context of this research, the end user is involved quite transparently, with the aim of giving back to the community that has provided time and attention to the project, rather than to create currency out of the research and algorithmic outputs.

Returning to the concept of requirements gathering, implementing co-creation for new media tool development enables key ethical values to be integrated into a software requirements specification document (IEEE Standards Association, 2018), within the foundational descriptions of the project, but also its specific requirements. Such participatory methods help answer some of the challenges posed by integrating AI into creative tools can be complicated to tackle.

9.1.4 On the Topic of Automatic Chapterisation

Through the examination of the three research aims, the practical output of this PhD takes the form of an automatic chapterisation algorithm and web

app, Podulr. The PD process reveals the need for better chapterisation solutions, combined with a keenness to offer more customised programmes, and an interest in sound recognition-aided editing. This is accompanied by ideas for specific applications of such a tool, divided into three categories: Podulr as an editing assistant, a catalogue manager, or a creative tool. Through the evaluation of the underlying algorithm pod-CLIPR in Chapter 7 and case studies in Chapter 8, the best uses and limitations of the tool are outlined.

As a contrasting approach to chapterisation, tools like Fathom [4] use NLP for segmentation directly on the user side. Fathom takes a similar approach to personalisation as Ian Forrester's adaptive podcasting app (Dwornik, 2021), which sees all of the customisations happen directly on the user end. NLP is also used in tools like Podium [5] [6], an all-in-one AI-powered podcast editor, that combines transcription, semantic editing, thematic analysis, show-note generation and even a custom AI-driven chat-bot. In one of its blog posts, Podium states:

> *"If you're afraid or sceptical of using AI tools, I understand—there's a lot to digest and keep up with! The idea isn't that they create perfection and replace you, but rather give you a huge boost, something solid to build on—so that you're done in minutes and not hours. You're responsible for the perfection part." - Podium's statement on automatic chapterisation* [7]

Where Fathom removes the podcasters' agency in deciding where cue markers fall, Podium prides itself on assisting the producer in their task. This service is available at a cost, with monthly subscriptions going from $12

[4] https://hello.fathom.fm

[5] https://hello.podium.page

[6] The algorithms used are proprietary and the reliance on NLP is an educated guess from looking at this and other functionalities of the app

[7] https://hello.podium.page/blog/adding-chapters-to-your-podcast

to $284, which can be topped up with credits for additional hours of audio processing over the maximum afforded by a user's chosen plan.

Although the statement quoted of this app is positive and seemingly takes into account the producer's needs and expectations, there is an obscurity that comes with private development, meaning that those chapterisation solutions cannot be easily compared with other algorithms. This is an unfortunate by-product of the allure of all-encompassing creative AI tools in the tech world, where sometimes "about" sections perform some form of "ethical signalling" (in reference to "virtue signalling"), even though no claims can be substantiated. That is not to say by any means these apps are unethical in essence, but that they contribute to obfuscating uses of AI in podcasting. Particularly, this practice distances the public from a considerate self-reflection on AI uses and takes away from the user any possibility to engage with generative AI in an ethically, environmentally and morally conscious way (Born et al., 2021). When summarising the state of ethics in generative AI for music Hu highlights four main themes occurring in reports on the topic over the 2018-2024 period:

> ***Data transparency & auditability**: "Transparency in the data around AI systems, from records of training data to clear labelling of AI-generated works",*
>
> ***Human artist centrality**: "Advocacy for protecting, and not undermining, the integrity of human creativity and artistry",*
>
> ***Consent & control**: "The ability for artists and rights holders to maintain control over whether and how their work is used and interpreted by AI systems.",*
>
> ***Compensation & licensing**: "Fair compensation to music creators and rights holders for use of their works in AI systems, whether through one-off buyout fees, royalty payments, revenue-sharing, or other arrangements." - (Hu, p.7-8)*

This PhD exemplifies the importance of *Data transparency & auditability* when integrating new technologies into creative workflows. Users have different relationships, fears, and pre-conceptions, and those should be able to be systematically acknowledged if the purpose of a tool is to benefit creators in the long run. Without these considerations, and whether they are expressed through participatory design, co-creation, or simply an openness surrounding the models and data used, it is impossible to ascertain whether a new tool for next-generation media improves upon or worsens a globally biased and unfair system (Born, 2020).

9.2 Conclusions

9.2.1 Answering the Research Questions

In Section 1.3, I set out four research questions that spanned across all dimensions of the research carried out: from ontological concerns, to practical applications of new technologies in podcast production - these questions intended to strengthen our bases for podcasting research. In summary of each chapter, partial answers to these questions were provided. The following section brings together these conclusions into four distinct answers.

RQ 1: What is Next-Generation Podcasting?

In Chapter 2, I presented NGP as a term standing "for all the innovative podcasting techniques and ideas that are yet to take off". Through further investigation, the broadness of the term is refined. Asking producers what they would make if anything was possible, the importance of creating new,

engaging and personalised experiences for listeners is highlighted, as well as the willingness to adopt tools that would facilitate the production process. Within NGP lives the full scope of actors as described in the methodology (Chapter 4), and their expectations of the term might vary drastically. For producers, it appears that NGP encompasses new systems for producing (creator-centric) or delivering (listener-centric) content, with a particular focus on personalising listener experience and improving immersion. The idea of modularity is explored in this thesis, although other forms of customisation, might that be in terms of interface of content, could easily be subbed for this concept under different circumstances.

But, in the same way that the definition provided for podcasting is by nature subject to change, the assumptions made regarding NGP will evolve with time, and the moment a new tool or process is assimilated within our understanding of podcasting, it will no longer be a part of the concept of NGP. This is where the Six-Tensions Framework also helps assess advances in the industry – and enables us to think of podcasting as an object with malleable boundaries that adapt to technical innovation.

RQ 2: How can we feasibly create and distribute immersive and personalised podcasts?

The notions of "immersive" and "personalised" are built and explored collaboratively with podcasters throughout this thesis. Oftentimes, personalisation is perceived as a tool towards immersion, using customisation to further embed users in stories or allow them to engage at different levels. Upholding these concepts as goals for NGP matches the already existing motives for

podcast listening (Table 5.1), and encourages new forms of podcast experiences to be created. But to be able to create such experiences, there must be percipient records and analysis of current podcasting habits.

Through interviews and workshops, this thesis crystalises the current production behaviour of podcasters in English-speaking production networks (Chapter 5). Collecting this real-world information enables the curation of a list of recommendations (Section 5.2.5), that caters to designers, developers, and researchers, looking to innovate in the field of podcasting. The importance of efficient, compatible, useful, comfortable, good value for money, and no-code tools that simplify workflows are highlighted. This process not only exposes these key requirements from NGP software, but also solidifies the legitimacy of using PD and co-creation methods as a way to design and contribute to knowledge.

Conversations, a review of prior literature, and technological overview, confirm that although building immersive and personalised podcasting tools might be as simple as PD with target podcasters, the issue of distribution remains prevalent; if the ambition is host-agnostic output, then there must be a standardised format for playing, displaying, and interacting with such content. If the independent and individual nature of podcasting is preserved in this aspect, we can expect that the resulting landscape will be a multitude of apps, with host-specific formats and file requirements. It is possible that from this predicted wealth of distribution solutions, one will prevail and be adopted by other developers – but like radio format changes have taken decades to settle, it can be postulated that interactive online audio format standard changes would also be very slowly adopted.

RQ 3: Why is modular podcasting attractive to creators, and how can it be facilitated without making the production process more complex?

Modular podcasting is the practice of offering multiple variations of a single podcast to listeners. This can depend on listening contexts, habits, preferences, or choices, and be reflected, for instance, through variable length programmes or customisation of content. This process encourages podcasters to work in "chunks", a practice already followed by editors in larger organisations like the BBC, and for independents who work on long-running shows. The appeal of modular podcasts is three-fold: it enables listeners to receive personalised versions of their programmes, it generates new opportunities for storytelling, and it permits interactivity for audiences, without the need for producers to create new content.

The roadmap to developing a modular podcasting tool that fulfils the requirements listed by podcasters is made clear through our answer to RQ 2. But one important problem remains: this concept just adds another step to a convoluted production process. Chapterising podcast audio is a lengthy process, and without integrated export solutions in all DAWs and editing platforms, it can be a deterrent to engaging with modular podcasting. Beyond modular podcasting, offering accurate chapters can help listeners navigate content, and find more points of entry to a programme. For producers, keeping a detailed archive of chapters and their associated metadata can allow composite pieces to be created, which is especially useful for long programmes (both in duration and time running).

pod-CLIPR is a system that automatically chapterises podcast audio

based on a combination of sound event detection and rules. This system is tested on a dataset of expertly annotated English-language podcasts spanning across a wide array of genres. It performs on par with human producers – that is to say that the chapters it suggests are plausible, but that there are mild variations from one podcaster to the next as to what counts as a chapter.

This system is integrated into the back end of a user-friendly app. Its manageable nature is ensured by the constant feedback loop with producers involved in the design and development of the app. The reliance on a web app has the benefit of both being compatible with most workflows, but also requiring an additional tool on top of an existing myriad of software.

RQ 4: What are the perceived benefits, risks, and costs of exploiting AI technologies for podcast production?

Despite the fact that AI was not a pre-supposed direction of the tool developed, the final product relies on an SED model to operate. This approach to chapterisation is quite lightweight (and therefore, resource-friendly) when examined next to other segmentation systems that rely on custom-trained LLMs. The particular form of AI used by pod-CLIPR was seen as compatible with podcasters' opinions on machine learning. By shining a light on what an "acceptable" AI for podcasters is (a transparent, responsibly trained and deployed, ethically indisputable AI), the shadow of "unacceptable" AI practices is also revealed.

Indeed, unlike for example generative AI, this use of sound recognition circumvents issues related to ownership, copyright, authenticity, and quality.

By offering clear provenance of training datasets and annotation methods, it also goes around some more foundational ethical issues that come with these applications of machine learning.

This integrates itself within a larger moral minefield, where even "acceptable" AI in creative fields are sometimes marketed as "assistants", creating an impression of morally sourced models, even if these are not transparent enough regarding their training datasets and implementation for a potential user to apply due diligence regarding their use of AI. This can in turn entrench oblivious relationships with AI, furthering biases, misuses, and profit-driven innovation.

The logistical costs of such tools highlight the importance of providing a measured response to digital problems, as AI comes with environmental, time and human costs. Models should be compared with other systems available on the market and evaluated within the specific requirements of the workflows of involved parties and actors.

9.2.2 Review of Limitations

Despite the comprehensive approach taken in this research, it is important to address the limitations encountered, which may have impacted this thesis' scope, findings, and conclusion. The detail of the limitations encountered in individual studies is included in each result chapter, therefore this reflection will mainly focus on larger limitations pertaining to the work as a whole.

The overall methodological approach relies almost entirely on principles of participatory design and co-creation. Even though the benefits of such collaborative methods are well documented and utilised to explore the RA.

presented, they carry some unavoidable quandaries. The particular pool of participants involved cannot be perfectly representative of a target demographic, and therefore, results will skew depending on the composition of the cohort. This makes the output of the project, particularly Podulr, entirely subject to a relatively small group of podcasters' opinions, in turn affecting the conclusions drawn about immersive and personalised podcasting. The focus on the particular technologies Podulr relies upon is the result of a rigorous investigation into a set of the wider "English-speaking podcasters" group. The percentages shared communicate information not about the community as a whole, but simply about the group interviewed. Of course, hypotheses by extrapolations can be formulated, but mainly, this paragraph acknowledges the immense influence of the specificity of co-created products of this thesis as a whole.

Moreover, there are some issues related to the representation achieved in the group of participants involved throughout this project. In Chapter 5 when conducting the first interviews, I discuss the poor gender representation but note that it corresponds to the makeup of the industry. This cannot be said for other attributes of this particular set of podcasters. Their interest in new technologies is undeniable, making them more likely to be early adopters of new tools, but also have an optimistic view of innovative technologies, like AI. Also, the BBC might be exemplary of some of the traditional inner workings of large broadcasting corporations, but its public-funded quality sets it aside from other networks. This is reflected in the views of the employees who have shared their time and opinions on NGP.

Additionally, this application of PD relies on the primary investigator

for facilitating interviews, workshops, and analysing data. This intertwines issues of researcher bias within an already convoluted process. This issue is mitigated at each level where it could interfere with results: for interviews, the choice of loose questions and a semi-structured format enable interviewees to bring as much detail to their answers as they feel the need to; for workshops, other facilitators are brought in, and the prompts are derived from quotes from prior interviews; for data analysis, especially quantitative analysis, best practices to minimise annotator input are followed (although thematic analysis is by nature subjective, as topics cannot simply "emerge" but are brought to the surface by a researcher's a priori assumptions and point of view).

Furthermore, the choice of focusing on creators is justified at length, but it remains a decision, rather than an absolute truth that such products and research should stem from production sides rather than from any of the other actors involved within the podcasting industry. There is no way around this decision as a source of potential error in the work, simply the encouragement of subjecting this work to the scrutiny suggested in Chapter 3, evaluating its output with all concerned parties.

Finally, the issue of making global recommendations from datasets and participants who primarily are based out of the UK (or other English-speaking countries) must be emphasised. It was shown that the meaning of "podcast" changes depending on the time, but also of the place (e.g. French 2010s Podcasts as discussed in Chapter 2). This is a key factor when evaluating how the conclusions drawn in this work could apply to other parts of the world's podcast culture. Relationships to technologies differ widely, and thus expec-

tations from personalised media will vary. The choice of limiting this research to English-speaking podcasts is one of manageability and reasonable research expectations.

9.2.3 Summary of Contributions

This thesis makes the following contributions:

- A definition of podcasting alongside a framework for podcasting innovation. The following definition is theorised:

  A podcast is a piece of episodical, downloadable or streamable, primarily spoken audio content, distributed via the internet, playable anywhere, at any time, produced by anyone who so wishes.

  Alongside a framework for podcasting innovation (the Six Tensions Framework, Figure 7.1), which allows the ontology of podcasting to be as flexible and metamorphic as the medium itself.

- A contemporary workflow for podcasting. Through interviews with podcasters, a proposed generalisation of the current podcasting workflow is provided (Figure 5.2)

- A summary of expectations of producers for NGP, views on new technologies, and a reflection on the systems already in place and how they'll need to adapt to enable it. A definition of NGP is given, stemming from a review of literature, completed by the opinions of podcast producers, and culminating into a series of takeaways (Section 5.2.5), as well as an introduction to the concept of modular podcasting.

- A pipeline for automatic podcast audio chapterisation, pod-CLIPR (Podcast Chapter Localisation through Intelligent Pattern Recognition) comprising of a sound recognition model combined with a rule-based algorithm, and its evaluation. A reflection on the systems in place and possible distribution issues is also perceived throughout this examination.

- A reflection on participatory design for developing new media tools and a practical application in the form of the modular podcasting web app Podulr. A systematic application of an ISD methodology combined with a PD framework (Figure 4.2) to develop an innovative tool for new media. This results in concrete software outputs (pod-CLIPR and Podulr), and their respective evaluation, but also in an overview of the method in itself for this purpose.

9.2.4 Future Work

Although the contributions of this thesis are standalone, aspects of the research could be further explored in upcoming work. There are four axes of future research building upon the findings and methodologies developed in this dissertation: tool evaluation, new formats deployment, technical ameliorations, and wider communications and media concerns.

Evaluation

Evaluating the outputs of modular podcasting tools such as Podulr on other actors within the industry is invaluable to test and understand its possible longevity and integration within our concept of "podcast". By handing mod-

ular podcasts to all forms of audiences, from engaged to passive, the possible reach of these new programmes could be gauged. This data could be used in turn to understand how advertisers could engage with the content – and this would be a key factor in the adoption of new technologies with creators, as monetisation can be an incentive for production. In order to achieve this, modular podcasts would have to be created and made available to the public.

New formats

In parallel, the work presented surrounding formats to make and deliver NGP, but also, hypotheses made regarding immersive and personalised media formats more generally, warrant more research. In Figures 8.5 and 8.6, I share the data format used to communicate information between the client (the user), Podulr (the interface), and pod-CLIPR (the underlying algorithm). This format is JSON-based, and includes descriptions of a project as objects, where the author's information and directions are included, as well as the output from pod-CLIPR, and the links to the assets involved. These objects could be useful not only in a production context, but also to pass on content or experiences to consumers. Together with the data formats Podulr relies on for content edition and creation, the system used for Adaptive podcasting, where the bulk of the information is processed on a consumer device through the use of XML scripts, showcase how these resource description frameworks (RDFs) can be used in creating as well as distributing personalised media.

This thesis, although aiming to encapsulate both independent and network-produced podcasting, ended up focusing more heavily on the latter, because of the industry ties with the BBC. This meant that important questions

regarding the democratisation of new media tools and how they could be feasibly implemented in independent creators' work could not be examined as closely as within work that solely focused on independent producers' workflows.

Technical ameliorations

Another avenue for future work lies in improving Podulr. From the feedback provided in the case studies (Chapter 8), there are some obvious ameliorations to make. First, better compatibility of import features could be implemented, allowing users to upload files directly to the app's server without relying on a third party, and reading chapter locations from file headers or cue marker files if provided. Similarly, better compatibility of export features could be added, allowing users to download files in different formats, fitting whatever distribution pipeline envisioned. Then, offline usage could be investigated, culminating in reducing the work time on the page, rendering, accessing, and using buffers or graphical elements quicker. Finally, Podulr could be taken offline entirely, separating it from third-party systems, maybe integrating it as a plugin to existing DAWs.

pod-CLIPR could also benefit from further investigation and improvements. The impact of adding or changing rules could be observed in more detail. Furthermore, these rules could be customised so that they would cater to specific use cases; for instance, the BBC uses sound bites systematically to begin and end programmes on sounds. These sound-bites have specific fingerprints, where sound-event detection models would return similar outputs for these segments across shows. If a rule was written to specifically look for

this fingerprint, the segment could be tagged adequately.

pod-CLIPR could also be combined with music recognition systems, relying on a different kind of fingerprinting and hash tables, to match recognised musical moments to a database of existing songs. This could be used to easily clear or list music from podcasts or radio shows adapted as podcasts.

pod-CLIPR could be trained on more specific data. The SED model currently used is trained on a broad spectrum of classes, not specific to podcasts. To simplify the calculations, any returned tag is mapped onto a restricted set of features deemed important by twenty podcasters queried (Music, Speech, Conversation, Female speech; woman speaking, Male speech; man speaking, Narration; monologue, Outside; rural or natural, Inside; small room, Singing, Sound effects). A model directly trained to recognise these features could output more precise tagging. Further evaluation would be required, and this could be extended to a larger comparison dataset than POD 49, involving and representing more diverse genres and types of content.

More broadly, one can imagine a model that evades the need for a rule-dependent pipeline, by training a model to output boundary suggestions directly from audio files. A dataset like POD 49 would have to be provided as training data, but this approach could encapsulate some of the specificities of podcast chapterisation captured by the rules developed for pod-CLIPR.

pod-CLIPR, as its title suggests, is intended for podcast chapterisation. However, this system, combining an SED model with a rule-based system, could be extended to different media. For instance, in the context of film or video making, could this approach be used to rapidly infer chapters, regardless of the visual context? Maybe combining this system with some analysis

of the content of frames could lead to optimal results.

Communications and new media

Outside of technical ramifications, transferring the methodological approach taken through this research to other facets of podcasting could create interesting results. Indeed, participatory film-making (Manni et al., 2019), co-creation of musical pieces (Kelleher et al., 2019), or collective forms storytelling (Holloway-Attaway and Vipsjö, 2020), receive some academic scrutiny, but co-created podcasting (or podcasts created with and by a community) and its effects have not yet been investigated in depth. Yet PD could be used not just with podcasters, but also with listeners. This could not only contribute to the development of NGP, by creating new forms of audio content in partnership with audiences, but also enable the telling of important community-led stories. This approach could be invaluable in heritage and cultural spaces, but also in critical research benefiting minority groups, local clusters, or global issues centred around a people's unique voice.

Finally, within the context of new media in society, which this thesis has examined through its examination of the ontology of podcasting and scrutiny of the network of actors involved in the medium, more questions remain: How is the lack of gender balance in this industry influencing the tools and technologies being developed for podcasters? How do the fragmented representation and performative showcasing of minorities from larger media corporations speak to the commodification of diversity in our cultural landscape? What explains the rise and sustained interest in "fake podcasting" (the act of pretending to podcast, or perform the act of podcasting for

snippets without actually distributing content[8]) on social media, and how do these relate to the perceived sense of authority and authenticity from podcast hosts?

These questions can come with an impending sense of urgency, as with all media trends, there is only so much time for them to retain their cultural zeitgeist – before these questions leave the realm of "present investigation" to become a "look back" at the social currencies afforded by podcasting.

9.2.5 Conclusion

As this thesis draws to a close, it is evident that the ramifications of this project call for more research and creative applications. The uses of PD and ISD to design an NGP tool are observed through the lens of purpose-driven innovation. This process enables over 50 producers to share their opinion regarding their views of podcasting, and reveals some important aspects of the production process, like their workflow, or requirements for software. This bridges evident gaps in the academic literature, and as an essential by-product, gives grounding and justification for the development of Podulr, a modular podcasting tool enabling for AI-driven chapterisation of audio. The contributions listed in Section 9.2.3 have importance beyond the bounds of this particular research endeavour, offering key data to other designers, researchers, podcasters, and industry professionals regarding the future of podcasting (Chapter 5), the use of SED in a system for automatic segmentation (Chapter 7), and the impact of using co-creative processes in personalised media forays (Chapter 8).

[8]https://www.nbcnews.com/video/that-tiktok-podcast-may-not-be-real-182704197597

The research aims studied invited the parallel examination of the research questions presented. Through this, a characterisation of NGP is provided, including both a global definition informed by a review of the literature and cultural context, and a summary of the point of view of podcasters on what this term means to them. A method for integrating new tools for immersive and personalised podcasting in existing production and distribution systems is provided, highlighting the importance of participatory design, and a lack of finality when it comes to formats for creating and sharing Web Audio. The notion of modular podcasting is introduced, and this thesis presents its possible implementation of the concept into a web app, aided by an AI-driven chapterisation tool that suggests segmentations of audio files. This app is seen to have three possible applications (creative, archival, and practical), and the UI as well as the underlying algorithms are evaluated by human experts. Pod-CLIPR, the system for chapterisation, is shown to perform on par with experienced producers, and Podulr is seen as overall helpful and promising by participants. Finally, as the field of new technologies for NGP is refined into Podulr, the perceived benefits and downsides of relying on AI-driven technologies for podcasting are explored, emphasizing the importance of transparency, authenticity, quality, and safety, as well as the logistical and environmental costs of these solutions.

Overall, this research provides some necessary context to the fields of podcasting research and interactive audio production, drawing from HCI to develop and evaluate a collaborative product, and to conclude not just from the results of the work carried out as part of the iterative development process, but also from the application of this methodology in itself.

Even though podcasting is turning twenty years old in 2024, there is still a trend to approach it as a nascent format. But, podcasters are well established by now, and by considering their experiences, researchers can map the current cultural terrain, and plan for the future of this unique medium, which like this research, lies at the intersection of art and technology.

A

Appendix

A.1 Mathematical Symbols

- W_t: Detection window frame, set as 0.5 sec
- L: Length of an audio file in seconds
- N: Number of frames in an audio file
- τ: Time window index, $\tau \in T = \{nW_t : n = 0, 1, \ldots, N = \lfloor L/W_t \rfloor\}$
- $\mathbf{s}_\tau$: State vector at time window τ
- $v_{k\tau}$: Value associated with feature k at time window τ, $v_{k\tau} \in \mathbb{R}$, $0 \leq v_{k\tau} \leq 1$
- $\mathbf{s}_\tau = (v_{1\tau}, v_{2\tau}, \ldots, v_{K\tau})$: State vector in terms of feature probabilities
- R: Total number of features considered, $R = 10$
- $\Phi(\tau, \tau + 1)$: Flux between two states at windows τ and $\tau + 1$
- P_i: i^{th} percentile of the flux values
- M_τ: Maximum feature value for frame τ, $M_\tau = \max_{k=1,\ldots,10}\{v_{k_\tau}\}$
- $d(a, b)$: Absolute difference between a and b

- σ_τ: Standard deviation
- ϵ_z: Standard errors from the mean of the values, with z the confidence interval coefficient
- β_l: Candidate boundary location, with l ranging from the first to last candidate boundary index
- D: Threshold duration, set as $D = 8\,\text{sec}$

References

A blind legend, 2020. http://www.ablindlegend.com/. Accessed 09.2022.

MIDAS Spring 2020. Technical report, RAJAR, 2020.

Apple Podcasts Subscriptions and channels are now available worldwide, 2021. https://www.apple.com/newsroom/2021/06/apple-podcasts-subscriptions-and-channels-are-now-available-worldwide/. Accessed 06.21.

Entale, 2021. https://www.entale.co. Accessed 04.21.

The social audio app, 2022. https://www.clubhouse.com/.Accessed 04.22.

Flexible media, March 2023. https://www.bbc.co.uk/programmes/p0f8xhj4. Accessed 04.22.

Foley sound effects, 2024. https://www.adobe.com/uk/creativecloud/video/discover-sound-effects.html. Accessed 02.24.

Wil M P Aalas and Stefan Jablonski. Dealing with workflow change: identification of issues and solutions. *Comput Syst Sci & Eng*, 5:267–276, May 2000.

Pekka Abrahamsson, Outi Salo, Jussi Ronkainen, and Juhani Warsta. Agile software development methods: Review and analysis. *arXiv preprint arXiv:1709.08439*, 2017.

Chadia Abras, Diane Maloney-Krichmar, Jenny Preece, et al. User-centered design. *Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications*, 37(4):445–456, 2004.

Marie-Luce Bourguet Abriella Kazai, Mounia Lalmas and Alain Pearmain. Using metadata to provide scalable broadcast and internet content and services, 2018.

William C. Adams. *Conducting Semi-Structured Interviews*, chapter 19, pages 492–505. John Wiley & Sons, Ltd, 2015. ISBN 9781119171386.

Sharath Adavanne, Pasi Pertilä, and Tuomas Virtanen. Sound event detection using spatial features and convolutional recurrent neural network. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, page 771–775, March 2017. doi: 10.1109/ICASSP.2017.7952260.

Adori. Homepage, 2022. https://www.adorilabs.com/. Accessed July 2022.

Sarvesh Agrawal, Søren Bech, Katrien de Moor, and Søren Forchhammer. Influence of changes in audio spatialization on immersion in audiovisual experiences. *Journal of the Audio Engineering Society*, 70(10):810–823, October 2022. ISSN 1549-4950. doi: 10.17743/jaes.2022.0034. Publisher Copyright: © 2022 Audio Engineering Society. All rights reserved.

Alan B. Albarran, Tonya Anderson, Ligia Garcia Bejar, Anna L. Bussart, Elizabeth Daggett, Sarah Gibson, Matt Gorman, Danny Greer, Miao Guo, Jennifer L. Horst, Tania Khalaf, John Phillip Lay, Michael McCracken, Bill Mott, and Heather Way. "what happened to our audience?" radio and new technology uses and gratifications among young adult users. *Journal of Radio Studies*, 14(2):92–101, Nov 2007.

Anthony J Alberg, Ji Wan Park, Brant W Hager, Malcolm V Brock, and Marie Diener-West. The use of "overall accuracy" to evaluate the validity of screening or diagnostic tests. *Journal of General Internal Medicine*, 19 (5 Pt 1):460–465, May 2004.

Lisa Anne Hendricks, Oliver Wang, Eli Shechtman, Josef Sivic, Trevor Darrell, and Bryan Russell. Localizing moments in video with natural language. page 5803–5812, 2017.

Apple Newsroom. Apple Podcasts Subscriptions and channels are now available worldwide, June 2021. https://www.apple.com/newsroom/2021/06/apple-podcasts-subscriptions-and-channels-are-now-available-worldwide. Accessed 08.07.21.

Dawit Mureja Argaw, Fabian Caba Heilbron, Joon-Young Lee, Markus Woodson, and In So Kweon. The anatomy of video editing: A dataset and benchmark suite for ai-assisted video editing. In Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner, editors, *Computer Vision – ECCV 2022*, Lecture Notes in Computer Science,

page 201–218, Cham, 2022. Springer Nature Switzerland. ISBN 978-3-031-20074-8. doi: 10.1007/978-3-031-20074-8_12.

Mike Armstrong and Maxine Glancy. Whp 396, the role of the audience in media. june 2022. https://www.bbc.co.uk/rd/publications/role-of-audience-in-media-how-culture-framing-narration-shape-way-stories-are-understood. Accessed 06.23.

Mike Armstrong, Sally Bowman, Matthew Brooks, Andy Brown, Juliette Carter, Andy Jones, Max Leonard, and Thomas Preece. Taking object-based media from the research environment into mainstream production. *SMPTE Motion Imaging Journal*, 129(5):30–38, 2020.

Ron Artstein. *Inter-annotator Agreement*, page 297–313. Springer Netherlands, Dordrecht, 2017. ISBN 978-94-024-0879-9. doi: 10.1007/978-94-024-0881-2_11.

Association for Qualitative Research (AQR). Definition: Co-creation, 2022. https://www.aqr.org.uk/glossary/co-creation. Accessed 05.24.

Matthew P. Aylett, Leigh Clark, Benjamin R. Cowan, and Ilaria Torre. *Building and Designing Expressive Speech Synthesis*, page 173–212. Association for Computing Machinery, New York, NY, USA, 1 edition, 2021. ISBN 9781450387200.

W. Hugh Baddeley. The technique of documentary film production. revised edition. 1970.

Werner Bailer, Hannes Fassold, Jakub Rosner, and Georg Thallinger. Innovative tools for 3d cinema production, 2020.

Brian P Bailey and Joseph A Konstan. Authoring interactive media. *Encyclopedia of electrical and electronics engineering*, pages 1–13, 2000.

Adrián Barahona-Ríos and Tom Collins. Noisebandnet: controllable time-varying neural synthesis of sound effects using filterbanks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 32:1573–1585, 2024.

Julia Barnett. The ethical implications of generative audio models: A systematic literature review. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, page 146–161, Montreal QC Canada, August 2023. ACM. ISBN 9798400702310.

Mathieu Barthet, Steven Hargreaves, and Mark Sandler. Speech/music discrimination in audio podcast using structural segmentation and timbre recognition. In Sølvi Ystad, Mitsuko Aramaki, Richard Kronland-Martinet, and Kristoffer Jensen, editors, *Exploring Music Contents*, Lecture Notes in Computer Science, page 138–162, Berlin, Heidelberg, 2011. Springer. ISBN 978-3-642-23126-1. doi: 10.1007/978-3-642-23126-1_10.

Tom Bartindale, Alia Sheikh, Nick Taylor, Peter Wright, and Patrick Olivier. Storycrate: tabletop storyboarding for live film production. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, page 169–178, New York, NY, USA, May 2012. Association for Computing Machinery. ISBN 978-1-4503-1015-4.

Victor R. Basil and Albert J. Turner. Iterative enhancement: A practical technique for software development. *IEEE Transactions on Software Engineering*, SE-1(4):390–396, December 1975. ISSN 1939-3520. Conference Name: IEEE Transactions on Software Engineering.

Valentin Bauer, Anna Nagele, Chris Baume, Tim Cowlishaw, Henry Cooke, Chris Pike, and Patrick G. T. Healey. Designing an Interactive and Collaborative Experience in Audio Augmented Reality. In *Virtual Reality and Augmented Reality*, volume 11883, pages 305–311. Springer International Publishing, 2019.

Chris Baume. *Semantic Audio Tools for Radio Production.* PhD thesis, January 2018.

Chris Baume. "Even More or Less": A data-rich interactive podcast player. *Proceedings of ACM International Conference on Interactive Experiences for Television and Online Video (TVX2019)*, April 2019. doi: 10.5281/zenodo.2654885. ACM, New York, NY, USA, 10 pages.

Chris Baume, Mark D. Plumbley, Janko Ćalić, and David Frohlich. A contextual study of semantic speech editing in radio production. *International Journal of Human-Computer Studies*, 115:67–80, 2018.

Mohammad Bavarian, Heewoo Jun, Nikolas Tezak, John Schulman, Christine McLeavey, Jerry Tworek, and Mark Chen. Efficient Training of Language Models to Fill in the Middle, July 2022.

Petra Saskia Bayerl and Karsten Ingmar Paul. What determines inter-coder agreement in manual annotations? a meta-analytic investigation. *Computational Linguistics*, 37(4):699–725, 2011.

BBC. Radio 3 in immersive sound, 2020. https://www.bbc.co.uk/programmes/articles/29L27gMX0x5YZxkSbHchstD/radio-3-in-immersive-sound. Accessed July 2022.

BBC. BBC World Service International Podcast Competition: Meet the judges, 2021. https://www.bbc.co.uk/programmes/articles/5Zp6Rty3TLh0qnWr3Dl995K/bbc-world-service-international-podcast-competition-meet-the-judges. Accessed July 2022.

BBC Sounds. What are the codecs, bitrates and protocols used for BBC radio online?, 2021. https://www.bbc.co.uk/sounds/help/questions/about-bbc-sounds-and-our-policies/codecs-bitrates. Accessed 07.08.21.

BBC Taster. Monster, 2020. https://www.bbc.co.uk/taster/pilots/monster. Accessed November 2020.

BBC Taster. Instagramification, 2021. https://www.bbc.co.uk/taster/pilots/instagramification. Accessed November 2020.

BBC Taster. Spectrum Sounds, 2022. https://www.bbc.co.uk/taster/pilots/spectrum-sounds. Accessed July 2022.

Christoph Becker, Günther Kolar, Josef Küng, and Andreas Rauber. Preserving interactive multimedia art: A case study in preservation planning. In Dion Hoe-Lian Goh, Tru Hoang Cao, Ingeborg Torvik Sølvberg, and Edie Rasmussen, editors, *Asian Digital Libraries. Looking Back 10 Years and Forging New Frontiers*, Lecture Notes in Computer Science, page 257–266, Berlin, Heidelberg, 2007. Springer. ISBN 978-3-540-77094-7. doi: 10.1007/978-3-540-77094-7_35.

Valdecir Becker, Daniel Gambaro, and Thais Saraiva Ramos. Audiovisual design and the convergence between hci and audience studies. In Masaaki Kurosu, editor, *Human-Computer Interaction. User Interface Design, Development and Multimodality*, Lecture Notes in Computer Science, page 3–22, Cham, 2017. Springer International Publishing.

Oliver Bendel. The synthetization of human voices. *AI & SOCIETY*, 34(1): 83–89, March 2019. ISSN 1435-5655.

Nicole Beniamini. The Infinite Dial 2020. Technical report, Edison Research, 2020.

Nicole Beniamini. The infinite dial 2022. Technical report, Mar 2022. https://www.edisonresearch.com/the-infinite-dial-2022/. Accessed July 2022.

Nicole Beniamini. The Infinite Dial. Technical report, Edison Research, 2023.

Adan Benito, Thomas Vassallo, Joshua Reiss, and Parham Bahadoran. Fxive: A web platform for procedural sound synthesis. *Journal of the Audio Engineering Society. Audio Engineering Society*, 05 2018.

Tony Benson, Susanne Pedersen, George Tsalis, Rebecca Futtrup, Moira Dean, and Jessica Aschemann-Witzel. Virtual co-creation: a guide to conducting online co-creation workshops. *International Journal of Qualitative Methods*, 20, 2021.

Richard Berry. Will the iPod Kill the Radio Star?Profiling Podcasting as Radio. *Convergence: The International Journal of Research Into New Media Technologies*, 12:143–162, May 2006.

Richard Berry. A Golden Age of Podcasting? Evaluating Serial in the Context of Podcast Histories. *Journal of Radio & Audio Media*, 22(2):170–178, July 2015. ISSN 1937-6529. Publisher: Routledge.

Richard Berry. Part of the establishment: Reflecting on 10 years of podcasting as an audio medium. *Convergence*, 22(6):661–671, December 2016. ISSN 1354-8565. Publisher: SAGE Publications Ltd.

Richard Berry, 2020. https://richardberry.eu/there-are-just-3-types-of-podcast/. Accessed 05.24.

Jean-Samuel Beuscart and Kevin Mellet. La conversion de la notoriété en ligne. Une étude des trajectoires de vidéastes pro-am. *Terrains & travaux*, 26(1):83–104, 2015. ISSN 1627-9506.

Igor Bieda and Taras Panchenko. A systematic mapping study on artificial intelligence tools used in video editing. *International Journal of Computer Science and Network Security*, 22(3):312–318, March 2022. doi: 10.22937/IJCSNS.2022.22.3.40.

Frank Biocca and Mark R. Levy. *Communication in the Age of Virtual Reality*. Routledge, February 2013. ISBN 978-1-135-69357-2.

Abeba Birhane, William Isaac, Vinodkumar Prabhakaran, Mark Diaz, Madeleine Clare Elish, Iason Gabriel, and Shakir Mohamed. Power to

the people? opportunities and challenges for participatory ai. In *Equity and Access in Algorithms, Mechanisms, and Optimization*, page 1–8, Arlington VA USA, October 2022. ACM. ISBN 978-1-4503-9477-2. doi: 10.1145/3551624.3555290.

Thomas Birtchnell. Listening without ears: Artificial intelligence in audio mastering. *Big Data & Society*, 5(2), 2018.

Kurt Bittner and Ian Spence. *Managing iterative software development projects*. Addison-Wesley Professional, 2006.

Robert Bleidt, Arne Borsum, Harald Fuchs, and S. Merrill Weiss. Object-based audio: Opportunities for improved listening experience and increased listener involvement. *SMPTE Motion Imaging Journal*, 124(5): 1–13, July 2015.

Knut Blind. The impact of standardisation and standards on innovation. *Handbook of Innovation Policy Impact*, page 423–449, Nov 2013.

Jan o. Blom and Andrew F. Monk. Theory of personalization of appearance: Why users personalize their pcs and mobile phones. *Human–Computer Interaction*, 18(3):193–228, September 2003. ISSN 0737-0024. doi: 10.1207/S15327051HCI1803_1.

Michael Bommarito II and Daniel Martin Katz. GPT Takes the Bar Exam, December 2022.

Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, Erik Brynjolfsson, Shyamal Buch, Dallas Card, Rodrigo Castellon, Niladri Chatterji, Annie Chen, Kathleen Creel, Jared Quincy Davis, Dora Demszky, Chris Donahue, Moussa Doumbouya, Esin Durmus, Stefano Ermon, John Etchemendy, Kawin Ethayarajh, Li Fei-Fei, Chelsea Finn, Trevor Gale, Lauren Gillespie, Karan Goel, Noah Goodman, Shelby Grossman, Neel Guha, Tatsunori Hashimoto, Peter Henderson, John Hewitt, Daniel E. Ho, Jenny Hong, Kyle Hsu, Jing Huang, Thomas Icard, Saahil Jain, Dan Jurafsky, Pratyusha Kalluri, Siddharth Karamcheti, Geoff Keeling, Fereshte Khani, Omar Khattab, Pang Wei Koh, Mark Krass, Ranjay Krishna, Rohith Kuditipudi, Ananya Kumar, Faisal Ladhak, Mina Lee, Tony Lee, Jure Leskovec, Isabelle Levent, Xiang Lisa Li, Xuechen Li, Tengyu Ma, Ali Malik, Christopher D. Manning, Suvir Mirchandani, Eric Mitchell, Zanele Munyikwa, Suraj Nair,

Avanika Narayan, Deepak Narayanan, Ben Newman, Allen Nie, Juan Carlos Niebles, Hamed Nilforoshan, Julian Nyarko, Giray Ogut, Laurel Orr, Isabel Papadimitriou, Joon Sung Park, Chris Piech, Eva Portelance, Christopher Potts, Aditi Raghunathan, Rob Reich, Hongyu Ren, Frieda Rong, Yusuf Roohani, Camilo Ruiz, Jack Ryan, Christopher Ré, Dorsa Sadigh, Shiori Sagawa, Keshav Santhanam, Andy Shih, Krishnan Srinivasan, Alex Tamkin, Rohan Taori, Armin W. Thomas, Florian Tramèr, Rose E. Wang, William Wang, Bohan Wu, Jiajun Wu, Yuhuai Wu, Sang Michael Xie, Michihiro Yasunaga, Jiaxuan You, Matei Zaharia, Michael Zhang, Tianyi Zhang, Xikun Zhang, Yuhui Zhang, Lucia Zheng, Kaitlyn Zhou, and Percy Liang. On the Opportunities and Risks of Foundation Models, July 2022.

J. A. Bondy and U.S.R. Murty. *Graph theory with applications*. Wiley, 2002.

J.S. Boreczky and L.D. Wilcox. A hidden markov model framework for video segmentation using audio and image features. In *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '98 (Cat. No.98CH36181)*, volume 6, page 3741–3744 vol.6, May 1998. doi: 10.1109/ICASSP.1998.679697.

Georgina Born. Diversifying MIR: Knowledge and Real-World Challenges, and New Interdisciplinary Futures. *Transactions of the International Society for Music Information Retrieval*, 3(1), 2020.

Georgina Born, Jeremy Morris, Fernando Diaz, and Ashton Anderson. Artificial intelligence, music recommendation, and the curation of culture, 2021. White Paper.

Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, January 2006. ISSN 1478-0887.

Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, July 2008. Publisher: Taylor & Francis Group.

Nancy H. Brinson and Laura L. Lemon. Investigating the effects of host trust, credibility, and authenticity in podcast advertising. *Journal of Marketing Communications*, 29(6):558–576, August 2023. ISSN 1352-7266. doi: 10.1080/13527266.2022.2054017.

Mathias Broth. The studio interaction as a contextual resource for tv-production. *Journal of Pragmatics*, 40(5):904–926, May 2008.

Amy Bruckman. The combinatorics of storytelling: Mystery train interactive. 1990.

Richard James Burgess. *The art of music production: the theory and practice*. Oxford University Press, Oxford ; New York, fourth edition edition, 2013.

William Burns, Liming Chen, Chris Nugent, Mark Donnelly, Kerry Louise Skillen, and Ivar Solheim. Mining usage data for adaptive personalisation of smartphone based help-on-demand services. In *Proceedings of the 6th International Conference on PErvasive Technologies Related to Assistive Environments - PETRA '13*, pages 1–7, Rhodes, Greece, 2013. ACM Press.

Buzzsprout.com. How to start a podcast: Your lightning fast, no-sweat, guide for 2022, 2022. https://www.thepodcasthost.com/planning/how-to-start-a-podcast/. Accessed Oct. 2022.

Callin. Social podcasting, 2022. https://www.callin.com/. Accessed 04.22.

Cambridge Dictionary, 2024. URL https://dictionary.cambridge.org/dictionary/english/chapter.

Laura Carpenter. What is a chapter?, 2024. https://harpercollins.co.uk/blogs/glossary/what-is-chapter . Accessed November 2024.

Joe Casabona. Sounds Profitable's "The Creators" Report & the Future of Podcasting, July 2022. https://www.thepodcasthost.com/business-of-podcasting/sounds-profitables-the-creators-report/. Accessed 05.24.

Ashley Castleberry and Amanda Nolen. Thematic analysis of qualitative research data: Is it as easy as it sounds? *Currents in Pharmacy Teaching and Learning*, 10(6):807–815, 2018.

Pocket Casts. Listen to podcasts with the best free podcasting app - built by listeners, for listeners., 2023. Url:https://pocketcasts.com/ Accessed June 2024.

S. Cavanagh. Content analysis: concepts, methods and applications. *Nurse Researcher*, 4(3):5–16, May 1997.

Sachin Chachada and C.-C. Jay Kuo. Environmental sound recognition: a survey. *APSIPA Transactions on Signal and Information Processing*, 3: e14, January 2014. ISSN 2048-7703. doi: 10.1017/ATSIP.2014.12.

Sylvia Chan Olmsted and Rang Wang. Understanding podcast users: Consumption motives and behaviors, 2020.

Pinaki Shankar Chanda and Sungjin Park. Speech intelligibility enhancement using tunable equalization filter. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, volume 4, pages IV–613–IV–616, 2007. doi: 10.1109/ICASSP.2007.366987.

S. Chandrakala and S. L. Jayalakshmi. Environmental audio scene and sound event recognition for autonomous surveillance: A survey and comparative studies. *ACM Computing Surveys*, 52(3):63:1–63:34, June 2019. ISSN 0360-0300. doi: 10.1145/3322240.

Ioannis K. Chaniotis, Kyriakos-Ioannis D. Kyriakou, and Nikolaos D. Tselikas. Is Node.js a viable option for building modern web applications? A performance evaluation study. *Computing*, 97(10):1023–1044, October 2015. ISSN 1436-5057.

Gong Chao. Human-computer interaction: Process and principles of human-computer interface design. In *2009 International Conference on Computer and Automation Engineering*, pages 230–233, 2009. doi: 10.1109/ICCAE.2009.23.

Amelia Chelsey. Is there a transcript? mapping access in the multimodal designs of popular podcasts. In *Proceedings of the 39th ACM International Conference on Design of Communication*, pages 46–53, 2021.

Guoguo Chen, Shuzhou Chai, Guanbo Wang, Jiayu Du, Wei-Qiang Zhang, Chao Weng, Dan Su, Daniel Povey, Jan Trmal, Junbo Zhang, Mingjie Jin, Sanjeev Khudanpur, Shinji Watanabe, Shuaijiang Zhao, Wei Zou, Xiangang Li, Xuchen Yao, Yongqing Wang, Yujun Wang, Zhao You, and Zhiyong Yan. Gigaspeech: An evolving, multi-domain asr corpus with 10,000 hours of transcribed audio. (arXiv:2106.06909), June 2021. doi: 10.48550/arXiv.2106.06909. arXiv:2106.06909 [cs, eess].

Liming Chen, Kerry Skillen, William Burns, Susan Quinn, Joseph Rafferty, Chris Nugent, Mark Donnelly, and Ivar Solheim. Learning Behaviour for Service Personalisation and Adaptation. In Xizhao Wang, Witold Pedrycz, Patrick Chan, and Qiang He, editors, *Machine Learning and Cybernetics*, Communications in Computer and Information Science, pages 287–297, Berlin, Heidelberg, 2014. Springer.

Ronald Chenail. Interviewing the investigator: Strategies for addressing instrumentation and researcher bias concerns in qualitative research. *Qualitative Report*, 16:255–262, 01 2011. doi: 10.46743/2160-3715/2011.1051.

Michael Christel and Kyo Kang. Issues in requirements elicitation. Technical Report CMU/SEI-92-TR-012, Software Engineering Institute, Carnegie Mellon University, Pittsburgh, PA, 1992.

Anthony W. P. Churnside. *Object-based radio: effects on production and audience experience*. PhD thesis, University of Salford, 2015a.

Anthony WP Churnside. Breaking out - an audio experiment - bbc r&d, Nov 2015b. https://www.bbc.co.uk/rd/blog/2012-07-breaking-out-an-audio-experi. Accessed 19.07/21.

Victoria Clarke, Virginia Braun, and Nikki Hayfield. Thematic analysis. *Qualitative psychology: A practical guide to research methods*, 3:222–248, 2015.

CNET. The complete history of Apple's iPod, 2011. https://www.cnet.com/pictures/the-complete-history-of-apples-ipod/. Accessed July 2022.

Mark Coeckelbergh. The art, poetics, and grammar of technological innovation as practice, process, and performance. *AI SOCIETY*, 33(4):501–510, November 2018. ISSN 1435-5655. doi: 10.1007/s00146-017-0714-7.

Joseph Nathan Cohen. Podcast post-production, Oct 2021. https://queenspodcastlab.org. Accessed 05.24.

Tom Collins, Sebastian Böck, Florian Krebs, and Gerhard Widmer. Bridging the audio-symbolic gap: The discovery of repeated note content directly from polyphonic music audio. In *Audio Engineering Society Conference: 53rd International Conference: Semantic Audio*. Audio Engineering Society, 2014.

Joseph T. Colonel and Joshua Reiss. Reverse engineering of a recording mix with differentiable digital signal processing. *The Journal of the Acoustical Society of America*, 150(1):608–619, 2021.

The Business Research Company. Podcasting market trends, size, growth, demand analysis report 2024 to 2033, Jan 2024.

Christophe Couvreur, Vincent Fontaine, Paul Gaunard, and Corine Ginette Mubikangiey. Automatic classification of environmental noise events by hidden markov models. *Applied Acoustics*, 54(3):187–206, July 1998. ISSN 0003-682X. doi: 10.1016/S0003-682X(97)00105-9.

Rob Cover. Audience inter/active: Interactive media, narrative control and reconceiving audience history. *New Media Society*, 8(1):139–158, February 2006. ISSN 1461-4448. doi: 10.1177/1461444806059922.

C. Cox, W. Trojak, T. Dzanic, F. D. Witherden, and A. Jameson. Accuracy, stability, and performance comparison between the spectral difference and flux reconstruction schemes. *Computers Fluids*, 221:104922, May 2021. ISSN 0045-7930. doi: 10.1016/j.compfluid.2021.104922.

Kate Crawford. *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press, 2021.

James Cridland. 1 in 10 brits will launch their own podcast in 2022, say acast, Jan 2022. https://podnews.net/press-release/acast-new-year.

Karin Danielsson and Charlotte Wiberg. Participatory design of learning media: Designing educational computer games with and for teenagers. *Interactive Technology and Smart Education*, 3(4):275–291, January 2006. ISSN 1741-5659. doi: 10.1108/17415650680000068.

R. Dankert. Actor–network theory. *International Encyclopedia of Housing and Home*, page 46–50, 2012. doi: 10.1016/b978-0-08-047163-1.00606-8.

Matthew E. P. Davies and Mark D. Plumbley. Context-dependent beat tracking of musical audio. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(3):1009–1020, March 2007. ISSN 1558-7924. doi: 10.1109/TASL.2006.885257.

Matthew E P Davies, Norberto Degara, and Mark D Plumbley. Evaluation methods for musical audio beat tracking algorithms.

Robin De Croon, Joris Klerkx, and Erik Duval. Designing a useful and usable mobile emr application through a participatory design methodology: A case study. In *2014 IEEE International Conference on Healthcare Informatics*, page 176–185, Verona, September 2014. IEEE. ISBN 978-1-4799-5701-9. doi: 10.1109/ICHI.2014.31.

Brecht De Man, Joshua Reiss, and Ryan Stables. Ten Years of Automatic Mixing. September 2017.

Andy Dearden and H. Rizvi. Participatory design and participatory development: a comparative review. Indiana University, Bloomington, Indiana, USA, January 2008.

Alexandre Défossez, Nicolas Usunier, Léon Bottou, and Francis Bach. Music Source Separation in the Waveform Domain. https://hal.archives-ouvertes.fr/hal-02379796, April 2021.

Alan Dennis, Barbara Wixom, and David Tegarden. *Systems analysis and design: An object-oriented approach with UML*. John wiley & sons, 2015.

Descript. Homepage, 2022. https://www.descript.com/. Accessed 2022.

Patricia Devlin. It's Time to Tune In to Women Podcasters, September 2022. https://www.thepodcasthost.com/business-of-podcasting/women-podcasters/.

Torgeir Dingsøyr, Sridhar Nerur, VenuGopal Balijepally, and Nils Brede Moe. A decade of agile methodologies: Towards explaining agile software development. *Journal of Systems and Software*, 85(6):1213–1221, June 2012. ISSN 0164-1212.

Constance Douwes, Philippe Esling, and Jean-Pierre Briot. Energy consumption of deep generative audio models. (arXiv:2107.02621), October 2021. arXiv:2107.02621 [cs, eess].

Barbara Downe-Wamboldt. Content analysis: Method, applications, and issues. *Health Care for Women International*, 13(3):313–321, January 1992. ISSN 0739-9332.

Christopher Drew. Educational podcasts: A genre analysis. *E-Learning and Digital Media*, 14(4):201–211, 2017.

Mariam Durrani, Kevin Gotkin, and Corrina Laughlin. "Serial" , Seriality, and the Possibilities for the Podcast Format: Visual Anthropology. *American Anthropologist*, 117(3):1–4, September 2015. ISSN 00027294.

Weronika Dwornik. Adaptive Podcasting & Rabbit Holes Collective: Interview with Ian Forrester, March 2021. http://thewritingplatform.com/2021/03/interview-with-ian-forrester-rabbit-holes-collective/. Accessed 05.24.

Tore Dybå and Torgeir Dingsøyr. Empirical studies of agile software development: A systematic review. *Information and Software Technology*, 50 (9):833–859, August 2008. ISSN 0950-5849.

Alexandre Défossez, Nicolas Usunier, Léon Bottou, and Francis Bach. Music source separation in the waveform domain. November 2019.

Edison Research. SheListens: Insights on Women Podcast Listeners, 2019. https://www.edisonresearch.com/shelistens-insights-on-women-podcast-listeners/. Accessed 08.07.21.

Maura Edmond. All platforms considered: Contemporary radio and transmedia engagement. *New Media & Society*, 17(9):1566–1582, October 2015. ISSN 1461-4448. Publisher: SAGE Publications.

Benjamin Elizalde, Mirco Ravanelli, Karl Ni, Damian Borth, and Gerald Friedland. Audio-concept features and hidden markov models for multimedia event detection.

Daniel P. W. Ellis. Beat tracking by dynamic programming. *Journal of New Music Research*, 36(1):51–60, March 2007. ISSN 0929-8215. doi: 10.1080/09298210701653344.

Tyna Eloundou, Sam Manning, Pamela Mishkin, and Daniel Rock. GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models, March 2023.

Encyclopedia Britannica. radio | Definition, History, & Facts, 2010. https://www.britannica.com/topic/radio. Accessed 08.02.21.

Entale. Homepage, 2022. https://www.entale.co/. Accessed July 2022.

Tom Evens. Dab+ as a systemic innovation: Stakeholder interests and the introduction of digital radio. *European Journal of Communication*, 35(5): 502–517, October 2020. ISSN 0267-3231. doi: 10.1177/0267323120928218.

Adam Feldstein Jacobs. *Automatic Podcast Chapter Segmentation: A Framework for Implementing and Evaluating Chapter Boundary Models for Transcribed Audio Documents*. 2022.

Maria Angela Ferrario, Will Simm, Peter Newman, Stephen Forshaw, and Jon Whittle. Software engineering for 'social good': integrating action research, participatory design, and agile development. In *Companion Proceedings of the 36th International Conference on Software Engineering*, ICSE Companion 2014, pages 520–523, New York, NY, USA, May 2014. Association for Computing Machinery.

Adam Field, Pieter Hartel, and Wim Mooij. Personal dj, an architecture for personalised content delivery. In *Proceedings of the 10th international conference on World Wide Web*, pages 1–7, 2001.

Cameron Jones M. Floyd, Ingbert r. and Michael B. Twidale. Resolving incommensurable debates: A preliminary identification of persona kinds, attributes, and characteristics. 2:12–26, Apr 2008.

Melissa Ford. *Writing Interactive Fiction with Twine*. Que Publishing, Apr 2016.

Ian Forrester. Perceptive Radio, 2013. https://www.bbc.co.uk/rd/projects/perceptive-radio. Accessed 17.07.21.

Kim Fox and Yasmeen Ebada. Egyptian female podcasters: shaping feminist identities. *Learning, Media and Technology*, 47(1):53–64, January 2022. ISSN 1743-9884.

Matthew P Fox, Kareem Carr, Lucy D'Agostino McGowan, Eleanor J Murray, Bertha Hidalgo, and Hailey R Banack. Will Podcasting and Social Media Replace Journals and Traditional Science Communication? No, but... *American Journal of Epidemiology*, 06 2021a. ISSN 0002-9262.

Mim Fox, Siobhán McHugh, Denika Thomas, Felix Kiefel-Johnson, and Ben Joseph. Bringing together podcasting, social work field education and learning about practice with Aboriginal peoples and communities. *Social Work Education*, 0(0):1–17, September 2021b.

Jon Francombe, Tim Brookes, Russell Mason, James Woodcock, et al. Evaluation of spatial audio reproduction methods (part 2): analysis of listener preference. *Journal of the Audio Engineering Society*, 65(3):212–225, 2017.

Jon Francombe, James Woodcock, Richard Hughes, Russell Mason, Andreas Franck, Chris Pike, Tim Brookes, William Davies, Philip Jackson, Trevor Cox, Filippo Fazi, and Adrian Hilton. Qualitative Evaluation of Media Device Orchestration for Immersive Spatial Audio Reproduction. *Journal of the Audio Engineering Society*, 66(6):414–429, June 2018.

Matthias Frank, Franz Zotter, and Alois Sontacchi. Producing 3d audio in ambisonics. *Proceedings of the AES International Conference*, 2015, 03 2015.

Mark Frary. Power to the podcast: Podcasting is bringing a whole new audience to radio and giving investigative journalism a boost. plus, our handy guide to making your own podcasts, 2017.

Christopher Frauenberger, Judith Good, Geraldine Fitzpatrick, and Ole Sejer Iversen. In pursuit of rigour and accountability in participatory design. *International Journal of Human-Computer Studies*, 74:93–106, February 2015.

Karine Freire, Gustavo Borba, and Luisa Diebold. Participatory design as an approach to social innovation. *Design Philosophy Papers*, November 2011. doi: 10.2752/144871311X13968752924950.

Enrique Frias-Martinez, George Magoulas, Sherry Chen, and Robert Macredie. Automated user modeling for personalized digital libraries. *International Journal of Information Management*, 26:234–248, June 2006. doi: 10.1016/j.ijinfomgt.2006.02.006.

Ronald D. Fricker and Matthias Schonlau. Advantages and disadvantages of internet research surveys: Evidence from the literature. *Field Methods*, 14(4):347–367, November 2002. ISSN 1525-822X. doi: 10.1177/152582202237725.

Lindsay Harris Friel. What do podcasters *actually* care about? our survey says..., Mar 2022a. https://www.thepodcasthost.com/mindset/podcaster-cares/. Accesssed November 2024.

Lindsay Harris Friel. Podcast Statistics & Industry Trends: Avg. Listener Numbers, Gear, Formats & More, April 2022b.

Erin Friess. Personas and decision making in the design process: an ethnographic case study. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1209–1218, New York, NY, USA, May 2012. Association for Computing Machinery. ISBN 978-1-4503-1015-4.

Pennie Frow, Suvi Nenonen, Adrian Payne, and Kaj Storbacka. Managing co-creation design: A strategic approach to innovation. *British Journal of Management*, 26(3):463–483, 2015. ISSN 1467-8551. doi: 10.1111/1467-8551.12087.

Jonathan Furner. Definitions of "metadata": A brief survey of international standards. *Journal of the Association for Information Science and Technology*, 71(6):E33–E42, 2020.

Blasé Gambino. Reflections on accuracy. 22:393–404, December 2006. ISSN 1573-3602. doi: 10.1007/s10899-006-9025-5.

Min Gao, Kecheng Liu, and Zhongfu Wu. Personalisation in web computing and informatics: Theories, techniques, applications, and future research. *Information Systems Frontiers*, 12(5):607–629, November 2010.

Roberto García and Òscar Celma. Semantic integration and retrieval of multimedia metadata.

Jort F. Gemmeke, Daniel P. W. Ellis, Dylan Freedman, Aren Jansen, Wade Lawrence, R. Channing Moore, Manoj Plakal, and Marvin Ritter. Audio set: An ontology and human-labeled dataset for audio events. page 776–780, March 2017. doi: 10.1109/ICASSP.2017.7952261.

Enguerrand Gentet, Bertrand David, Sébastien Denjean, Gaël Richard, and Vincent Roussarie. Speech intelligibility enhancement by equalization for in-car applications. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6934–6938. IEEE, 2020.

Michael W. Geoghegan and Dan Klass. Podcasting How-To. In Michael W. Geoghegan and Dan Klass, editors, *Podcast Solutions: The Complete Guide to Podcasting*, pages 27–30. Apress, Berkeley, CA, 2005. ISBN 978-1-4302-0054-3. doi: 10.1007/978-1-4302-0054-3_3.

Bill Gillham. *The research interview [electronic resource] / Bill Gillham.* Real world research. Continuum, London, 2000. ISBN 978-0-8264-4797-5.

Lisa Glebatis Perks, Jacob S. Turner, and Andrew C. Tollison. Podcast Uses and Gratifications Scale Development: Journal of Broadcasting & Electronic Media: Vol 63, No 4. *Journal of Broadcasting & Electronic Media*, Vol 63(No 4):617–634, November 2019.

Cyril Goutte and Eric Gaussier. *A Probabilistic Interpretation of Precision, Recall and.* January 2005.

Grand View Research. Podcasting Market Size, Share, Industry Report, 2021-2028, 2021. https://www.grandviewresearch.com/industry-analysis/podcast-market. Accessed 07.22.

Dave Gray, Sunni Brown, and James Macanufo. *Gamestorming: A playbook for innovators, rulebreakers, and changemakers.* " O'Reilly Media, Inc.", 2010.

Jonathan Grudin and John Pruitt. Personas, Participatory Design and Product Development: An Infrastructure for Engagement. 2002.

Greg Guest, Kathleen M. MacQueen, and Emily E. Namey. *Applied Thematic Analysis*. SAGE, 2012.

Jiří Halák, Michal Krsek, Sven Ubik, Petr Žejdl, and Felix Nevřela. Real-time long-distance transfer of uncompressed 4k video for remote collaboration. *Future Generation Computer Systems*, 27(7):886–892, Jul 2011.

Ben Hammersley. Why online radio is booming, February 2004. http://www.theguardian.com/media/2004/feb/12/broadcasting.digitalmedia. Accessed 09.02.21.

Danielle Hancock and Leslie McMurtry. 'I Know What a Podcast Is': Post-Serial Fiction and Podcast Media Identity: New Aural Cultures and Digital Media. In *Podcasting: New Aural Cultures and Digital Media*, pages 81–105. Springer, July 2018.

Graham Harman. *Object-Oriented Ontology: A New Theory of Everything*. Penguin UK, March 2018. ISBN 978-0-241-26917-6.

Graham Harman. Object-oriented ontology (ooo). In *Oxford Research Encyclopedia of Literature*. 2019.

Rex Hartson and Pardha Pyla. *Front Matter*, page i–ii. Morgan Kaufmann, Boston, January 2019. ISBN 978-0-12-805342-3. doi: 10.1016/B978-0-12-805342-3.09989-6.

Rachel Heck, Michael Wallick, and Michael Gleicher. Virtual videography. *ACM Trans. Multimedia Comput. Commun. Appl.*, 3(1):4–es, feb 2007.

Einar Heiervang and Robert Goodman. Advantages and limitations of web-based surveys: evidence from a child mental health survey. *Social Psychiatry and Psychiatric Epidemiology*, 46(1):69–76, January 2011. ISSN 1433-9285. doi: 10.1007/s00127-009-0171-9.

Russ Hepworth-Sawyer and Craig Golding. *What is Music Production?: A Producer's Guide : the Role, the People, the Process*. Taylor & Francis, 2011. ISBN 978-0-240-81126-0. Google-Books-ID: UJoC_eibCzkC.

J. Highsmith and A. Cockburn. Agile software development: the business of innovation. *Computer*, 34(9):120–127, September 2001. ISSN 1558-0814. Conference Name: Computer.

María-José Higueras-Ruiz, Francisco-Javier Gómez-Pérez, and Jordi Alberich-Pascual. Historical review and contemporary characterization of

showrunner as professional profile in tv series production: Traits, skills, competences, and style. *Communication & Society*, 31(1):91–106, 2018.

Lissa Holloway-Attaway and Lars Vipsjö. *Using Augmented Reality, Gaming Technologies, and Transmedial Storytelling to Develop and Co-design Local Cultural Heritage Experiences*, page 177–204. Springer International Publishing, Cham, 2020. ISBN 978-3-030-37191-3. doi: 10.1007/978-3-030-37191-3_10.

Cherie Hu. Music AI Ethics Tracker. https://www.waterandmusic.com/data/ai-ethics-tracker. Accessed 05.24.

Chih-Pei HU and Yan-Yi CHANG. John w. creswell, research design: Qualitative, quantitative, and mixed methods approaches. *Journal of Social and Administrative Sciences*, 4(2):205–207, Jun. 2017. doi: 10.1453/jsas.v4i2.1313.

Mick Hurbis-Cherrier. *Voice and Vision. ; A Creative Approach to Narrative Film and DV Production.* Elsevier Science & Technology Books, 2011.

Sinh Huynh, Seungmin Kim, JeongGil Ko, Rajesh Krishna Balan, and Youngki Lee. Engagemon: Multi-modal engagement sensing for mobile games. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(1), mar 2018.

Hyperradio Radio France. Son 3D, 2021. https://hyperradio.radiofrance.com/son-3d/. Accessed July 2022.

IEEE Standards Association. *ISO/IEC/IEEE 29148:2018(E)*, page 1–104, November 2018. doi: 10.1109/IEEESTD.2018.8559686.

Lucian Ion and Neil Humphrey. White paper: 4k digital capture and post-production workflow, 2004.

Adam Feldstein Jacobs. Automatic podcast chapter segmentation: A framework for implementing and evaluating chapter boundary models for transcribed audio documents.

Stanisław Jedrzejewski. Radio in the new media environment. *Radio: The resilient medium*, pages 17–27, 2015.

Yifan Jiao, Zhetao Li, Shucheng Huang, Xiaoshan Yang, Bin Liu, and Tianzhu Zhang. Three-dimensional attention-based deep ranking model for video highlight detection. *IEEE Transactions on Multimedia*, 20(10): 2693–2705, 2018.

M. Cameron Jones, Ingbert R. Floyd, and Michael B. Twidale. Teaching Design with Personas. 2008.

Peter Jones. *Contexts of Co-creation: Designing with System Stakeholders*, page 3–52. Translational Systems Sciences. Springer Japan, Tokyo, 2018. ISBN 978-4-431-55639-8. doi: 10.1007/978-4-431-55639-8_1.

Ignas Kalpokas. *Chapter 2. The Malleable Self: Immersion, Self-Optimisation, and Gamification*. Emerald Publishing, 2021.

Karlheinz Kautz. Investigating the design process: participatory design in agile software development. *Information Technology & People*, 24(3):217–235, January 2011. Publisher: Emerald Group Publishing Limited.

Robin H. Kay. Exploring the use of video podcasts in education: A comprehensive review of the literature. *Computers in Human Behavior*, 28(3): 820–831, May 2012. ISSN 0747-5632.

Carol Kelleher, Hugh N. Wilson, Emma K. Macdonald, and Joe Peppard. The score is not the music: Integrating experience and practice perspectives on value co-creation in collective consumption contexts. *Journal of Service Research*, 22(2):120–138, May 2019. ISSN 1094-6705. doi: 10.1177/1094670519827384.

Bryce Kellogg, Vamsi Talla, and Shyamnath Gollakota. Bringing gesture recognition to all devices. In *11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14)*, pages 303–316, Seattle, WA, April 2014. USENIX Association. ISBN 978-1-931971-09-6.

T. Kemp, M. Schmidt, M. Westphal, and A. Waibel. Strategies for automatic segmentation of audio data. In *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.00CH37100)*, volume 3, page 1423–1426 vol.3, June 2000. doi: 10.1109/ICASSP.2000.861862.

NamHyeok Kim and Chanjun Park. Inter-annotator agreement in the wild: Uncovering its emerging roles and considerations in real-world scenarios. (arXiv:2306.14373), June 2023. arXiv:2306.14373 [cs].

Don Kimber and Lynn Wilcox. Acoustic segmentation for audio browsers. August 1999.

A. Baki Kocaballi. Conversational AI-Powered Design: ChatGPT as Designer, User, and Product, February 2023.

Q. Kong, Y. Xu, I. Sobieraj, W. Wang, and M. D. Plumbley. Sound event detection and time–frequency segmentation from weakly labelled data. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27 (4):777–787, April 2019a. ISSN 2329-9304. doi: 10.1109/TASLP.2019.2895254.

Q. Kong, Y. Xu, I. Sobieraj, W. Wang, and M. D. Plumbley. Sound Event Detection and Time–Frequency Segmentation from Weakly Labelled Data. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27 (4):777–787, April 2019b. ISSN 2329-9304. doi: 10.1109/TASLP.2019.2895254.

Qiuqiang Kong, Yin Cao, Turab Iqbal, Yuxuan Wang, Wenwu Wang, and Mark D. Plumbley. Panns: Large-scale pretrained audio neural networks for audio pattern recognition. (arXiv:1912.10211), August 2020. arXiv:1912.10211 [cs, eess].

Felix Kreuk, Gabriel Synnaeve, Adam Polyak, Uriel Singer, Alexandre Défossez, Jade Copet, Devi Parikh, Yaniv Taigman, and Yossi Adi. Audiogen: Textually guided audio generation. March 2023. doi: 10.48550/arXiv.2209.15352.

Klaus Kuhn. Problems and Benefits of Requirements Gathering With Focus Groups: A Case Study. *International Journal of Human–Computer Interaction*, 12(3-4):309–325, December 2000. ISSN 1044-7318. Publisher: Taylor & Francis.

Sari Kujala. User involvement: A review of the benefits and challenges. *Behaviour & Information Technology*, 22(1):1–16, January 2003. Publisher: Taylor & Francis.

Elvan Kula, Arie van Deursen, and Georgios Gousios. Modeling team dynamics for the characterization and prediction of delays in user stories. In *Proceedings of the 36th IEEE/ACM International Conference on Automated Software Engineering*, ASE '21, page 991–1002. IEEE Press, 2022. ISBN 9781665403375. doi: 10.1109/ASE51524.2021.9678939.

Anurag Kumar and Bhiksha Raj. Audio event detection using weakly labeled data. In *Proceedings of the 24th ACM international conference on Multimedia*, MM '16, page 1038–1047, New York, NY, USA, October 2016a. Association for Computing Machinery. ISBN 978-1-4503-3603-1.

Anurag Kumar and Bhiksha Raj. Audio Event Detection using Weakly Labeled Data. In *Proceedings of the 24th ACM international conference on*

Multimedia, MM '16, pages 1038–1047, New York, NY, USA, October 2016b. Association for Computing Machinery.

Kristal Kuykendall. Over Half of Students Surveyed See Coding Skills as Vital But Over a Third Lack Learning Access -, 2022.

C. Kyriakakis. Fundamental and technological limitations of immersive audio systems. *Proceedings of the IEEE*, 86(5):941–951, 1998.

Vincent Labatut and Hocine Cherifi. Accuracy measures for the comparison of classifiers. July 2012. doi: 10.48550/arXiv.1207.3790.

Stephen Lane, Paidi O'Raghallaigh, and David Sammon. Requirements gathering: the journey. *Communications of the ACM*, 38(5):31–32, May 1995. ISSN 0001-0782, 1557-7317.

Sally Bowman Lauren Ward, Maxine Glancy and Michael Armstrong. The impact of new forms of media on production tools and practices, 2020a. URL https://www.bbc.co.uk/rd/publications/whp-391-impact-new-forms-media-production-tools-practices.

Sally Bowman Lauren Ward, Maxine Glancy and Michael Armstrong. The Impact of New Forms of Media on Production Tools and Practices, September 2020b.

Stephen Lax. Different standards: Engineers' expectations and listener adoption of digital and fm radio broadcasting. *Journal of Radio Audio Media*, 24(1):28–44, January 2017. ISSN 1937-6529. doi: 10.1080/19376529.2017.1297147.

Sang-Heon Lee, Myoung-Kyu Sohn, Dong-Ju Kim, Byungmin Kim, and Hyunduk Kim. Smart TV interaction system using face and hand gesture recognition. In *2013 IEEE International Conference on Consumer Electronics (ICCE)*, pages 173–174, January 2013.

Samúel Lefever, Michael Dal, and Ásrún Matthíasdóttir. Online data collection in academic research: advantages and limitations. *British Journal of Educational Technology*, 38(4):574–582, 2007. ISSN 1467-8535. doi: 10.1111/j.1467-8535.2006.00638.x.

Alexander Lerch. *Instantaneous Features*, chapter 3, pages 31–69. John Wiley Sons, Ltd, 2012. ISBN 9781118393550. doi: https://doi.org/10.1002/9781118393550.ch3. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118393550.ch3.

Daniel J. Lewis. Apple podcasts statistics, 2023. https://web.archive.org/web/20230817105730/https://podcastindustryinsights.com/apple-podcasts-statistics/. Accessed August 2023.

Yu Liang, Aditya Ponnada, Paul Lamere, and Nediyana Daskalova. Enabling goal-focused exploration of podcasts in interactive recommender systems. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*, IUI '23, page 142–155, New York, NY, USA, 2023. Association for Computing Machinery.

C.P. Lim, S.C. Woo, A.S. Loh, and R. Osman. Speech recognition using artificial neural networks. In *Proceedings of the First International Conference on Web Information Systems Engineering*, volume 1, pages 419–423 vol.1, June 2000.

Mia Lindgren. Intimacy and Emotions in Podcast Journalism: A Study of Award-Winning Australian and British Podcasts. *Journalism Practice*, 0 (0):1–16, June 2021. ISSN 1751-2786.

Vivian Liu and Lydia B Chilton. Design Guidelines for Prompt Engineering Text-to-Image Generative Models. In *CHI Conference on Human Factors in Computing Systems*, pages 1–23, New Orleans LA USA, April 2022. ACM.

Lie Lu, Hong-Jiang Zhang, and Hao Jiang. Content analysis for audio classification and segmentation. *IEEE Transactions on Speech and Audio Processing*, 10(7):504–516, October 2002. ISSN 1558-2353. doi: 10.1109/TSA.2002.804546.

Brady D Lund and Ting Wang. Chatting about chatgpt: how may ai and gpt impact academia and libraries? *Library Hi Tech News*, 2023.

Guangyi Lv, Tong Xu, Enhong Chen, Qi Liu, and Yi Zheng. Reading the videos: Temporal labeling for crowdsourced time-sync videos based on semantic embedding. In *Proceedings of the AAAI conference on artificial intelligence*, volume 30, 2016.

Lewis E. MacKenzie. Science podcasts: analysis of global production and output from 2004 to 2018. *Royal Society Open Science*, 6(1):180932, 2019.

Virginia M. Madsen. Voices-cast: a report on the new audiosphere of podcasting with specific insights for public broadcasting. In Terry Flew, editor, *Communication, Creativity and Global Citizenship*, pages 1191–1210. ANZCA, 2009.

Danielle Magaldi and Matthew Berler. *Semi-structured Interviews*, page 4825–4830. Springer International Publishing, Cham, 2020. ISBN 978-3-319-24612-3. doi: 10.1007/978-3-319-24612-3_857.

Moira Maguire and Brid Delahunt. Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars. *All Ireland Journal of Higher Education*, 9(33), October 2017. ISSN 2009-3160.

Shoji Makino. *Audio source separation*, volume 433. Springer, 2018.

Sarah L. Malecki, Kieran L. Quinn, Nathan Zilbert, Fahad Razak, Shiphra Ginsburg, Amol A. Verma, and Lindsay Melvin. Understanding the Use and Perceived Impact of a Medical Podcast: Qualitative Study. *JMIR Medical Education*, 5(2):e12901, September 2019.

D. G. Malham. Approaches to spatialisation. *Organised Sound*, 3(2):167–177, August 1998. Cambridge University Press.

Kirsti Malterud. Systematic text condensation: A strategy for qualitative analysis. *Scandinavian Journal of Public Health*, 40(8):795–805, 2012. doi: 10.1177/1403494812465030.

Bruce Mamer. *Film Production Technique: Creating the Accomplished Image*. Cengage Learning, May 2013.

Simona Manni, Marian Ursu, and Jonathan Hook. Stepping through remixed: Exploring the limits of linear video in a participatory mental health film. In *Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video*, TVX '19, page 83–94, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450360173. doi: 10.1145/3317697.3323363.

Kris M. Markman. Doing radio, making friends, and having fun: Exploring the motivations of independent audio podcasters. *New Media & Society*, 14 (4):547–565, June 2012. ISSN 1461-4448. Publisher: SAGE Publications.

Kris M. Markman and Caroline E. Sawyer. Why Pod? Further Explorations of the Motivations for Independent Podcasting. *Journal of Radio & Audio Media*, 21(1):20–35, January 2014. ISSN 1937-6529. doi: 10.1080/19376529.2014.891211. Publisher:Routledge.

Keith Dana Martin. *Sound-source recognition: a theory and computational model*. Thesis, Massachusetts Institute of Technology, 1999.

Irene Martín-Morató and Annamaria Mesaros. What is the ground truth? reliability of multi-annotator data for audio tagging. In *2021 29th European Signal Processing Conference (EUSIPCO)*, page 76–80, August 2021. doi: 10.23919/EUSIPCO54536.2021.9616087.

Kerry Matthews. Research into podcasting technology including current and possible future uses, 01 2006.

Graham McAllister and Gareth White. Video Game Development and User Experience. pages 107–128. December 2010. ISBN 978-1-84882-962-6. doi: 10.1007/978-1-84882-963-3_7.

Mary L. McHugh. Interrater reliability: the kappa statistic. *Biochemia Medica*, 22(3):276–282, October 2012. ISSN 1330-0962.

Siobhan McHugh. How podcasting is changing the audio storytelling genre. *The Radio Journal – International Studies in Broadcast & Audio Media*, 14:65–82, January 2016.

Steven Mclung and Kristine Johnson. Examining the motives of podcast users. *Journal of Radio & Audio Media*, 17(1):82–95, May 2010a.

Steven Mclung and Kristine Johnson. Examining the Motives of Podcast Users. *Journal of Radio & Audio Media*, 17(1):82–95, May 2010b. ISSN 1937-6529. Publisher: Routledge.

Soroush Mehri, Kundan Kumar, Ishaan Gulrajani, Rithesh Kumar, Shubham Jain, Jose Sotelo, Aaron Courville, and Yoshua Bengio. Samplernn: An unconditional end-to-end neural audio generation model. February 2017. doi: 10.48550/arXiv.1612.07837.

Britta Meixner, Maxine Glancy, Matt Rogers, Caroline Ward, Thomas Röggla, and Pablo Cesar. Multi-screen director. *Adjunct Publication of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video*, 2017. doi: 10.1145/3084289.3089924.

Annamaria Mesaros, Toni Heittola, Tuomas Virtanen, and Mark D. Plumbley. Sound event detection: A tutorial. *IEEE Signal Processing Magazine*, 38(5):67–83, September 2021. ISSN 1558-0792. doi: 10.1109/MSP.2021.3090678.

Tomasz Miaskiewicz and Kenneth A. Kozar. Personas and user-centered design: How can personas benefit product design processes? *Design Studies*, 32(5):417–430, September 2011. ISSN 0142-694X.

Melanie Mitchell and David C. Krakauer. The debate over understanding in ai's large language models. *Proceedings of the National Academy of Sciences*, 120(13):e2215907120, 2023.

S. P. Mohanapriya, E. P. Sumesh, and R. Karthika. Environmental sound recognition using gaussian mixture model and neural network classifier. In *2014 International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE)*, page 1–5, March 2014. doi: 10.1109/ICGCCEE.2014.6922272.

J. Michael Moore and Frank M. Shipman III. A comparison of questionnaire-based and gui-based requirements gathering. In *Proceedings of the 15th IEEE International Conference on Automated Software Engineering*, ASE '00, page 35, USA, 2000. IEEE Computer Society. ISBN 0769507107.

Kasey Moore. Netflix is no longer making interactive shows and movies, 2024. https://www.whats-on-netflix.com/news/netflix-no-longer-making-interactive-shows-and-movies/; Accessed Apr 2024.

Veronica Morfi and Dan Stowell. Deep learning for audio event detection and tagging on low-resource datasets. *Applied Sciences*, 8(88):1397, August 2018.

Elan Moritz. Chatting with chat(gpt-4): Quid est understanding?, 2024. https://philarchive.org/rec/MORCWC-2. Accessed November 2024.

Anne E Mueller and Daniel L Segal. Structured versus semistructured versus unstructured interviews. *The encyclopedia of clinical psychology*, pages 1–7, 2014.

Nicolas M. Müller, Karla Pizzi, and Jennifer Williams. Human perception of audio deepfakes. In *Proceedings of the 1st International Workshop on Deepfake Detection for Audio Multimedia*, DDAM '22, page 85–91, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450394963. doi: 10.1145/3552466.3556531.

Steven Murdoch. Agent-oriented modelling in the production of 3D character animation. *Studies in Australasian Cinema*, 10(1):35–52, January 2016. ISSN 1750-3175. doi: 10.1080/17503175.2015.1133486. Publisher: Routledge _eprint: https://doi.org/10.1080/17503175.2015.1133486.

Declan Murphy. Quantization revisited: a mathematical and computational model. *Journal of Mathematics and Music*, 5(1):21–34, March 2011. ISSN 1745-9737. doi: 10.1080/17459737.2011.573674.

Simone Murray. Servicing 'self-scheduling consumers' public broadcasters and audio podcasting. *Global Media and Communication*, 5:197–219, 11 2009.

San Murugesan and Annamalai Ramanathan. Web personalisation - an overview. In Jiming Liu, Pong C. Yuen, Chun-hung Li, Joseph Ng, and Toru Ishida, editors, *Active Media Technology*, pages 65–76, Berlin, Heidelberg, 2001. Springer Berlin Heidelberg.

Music Radar. 6 ai powered intelligent plugins that could change the way you make music, 2022. https://www.musicradar.com/news/6-ai-powered-intelligent-plugins-change-music. Accessed July 2022.

Joschka Mütterlein. The three pillars of virtual reality? investigating the roles of immersion, presence, and interactivity. 2018.

Mudavath Nayak and Narayan K A. Strengths and weakness of online surveys. 24:31–38, May 2019. doi: 10.9790/0837-2405053138.

David R. Nelson and William V. Faux Ii. Evaluating Podcast Compositions: Assessing Credibility, Challenges, and Innovation. *The Journal of Social Media in Society*, 5(1):38–64, May 2016. ISSN 2325-503X.

Nic Newman, Richard Fletcher, Rasmus Kleis Nielsen, and Antonis Kalogeropoulos. Reuters institute digital news report 2019. *Reuters Institute for the Study of Journalism*, 2019.

Spotify Newsroom. Spotify Opens Doors for More Underrepresented Podcasters Through New Sound Up Programs, March 2021. https://newsroom.spotify.com/2021-03-31/spotify-opens-doors-for-more-underrepresented-podcasters-through-new-sound-up-programs/. Accessed July 2022.

Omar A. Niamut, Axel Kochale, Javier Ruiz Hidalgo, Rene Kaiser, Jens Spille, Jean-Francois Macq, Gert Kienast, Oliver Schreer, and Ben Shirley. Towards a format-agnostic approach for production, delivery and rendering of immersive media. In *Proceedings of the 4th ACM Multimedia Systems Conference*, page 249–260, Oslo Norway, February 2013. ACM.

Marta Ramona Novăceanu. From The Traditional Radio to Podcast or To the À La Carte. *World Journal of Research and Review*, 11(1), July 2020. ISSN 2455-3956. doi: 10.31871/WJRR.11.1.8.

NPR. Student Podcast Challenge, 2022a. https://www.npr.org/series/662609200/npr-student-podcast-challenge?t=1657918899275. Accessed July 2022.

NPR. A studio at your fingertips: 5 apps teachers are using to make student podcasts, 2022b. https://www.npr.org/2020/02/21/807372536/a-studio-at-your-fingertips-5-apps-teachers-are-using-to-make-student-podcasts.

Aditya Arie Nugraha, Antoine Liutkus, and Emmanuel Vincent. Multichannel audio source separation with deep neural networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 24(9):1652–1664, 2016.

Ulfa Octaviani and Chris Baume. A Remote Study on Enhanced Podcast Interaction, October 2020.

OFCOM. UK CMR 2007, 2007.

OFCOM. Online Nation – 2020 report, 2020.

Rob G. Oldfield, Max S. S. Walley, Ben G. Shirley, and Doug L. Williams. Cloud-based ai for automatic audio production for personalized immersive xr experiences. *SMPTE Motion Imaging Journal*, 131(7):6–16, August 2022. ISSN 2160-2492. doi: 10.5594/JMI.2022.3184849.

Robert Oldfield, Ben Shirley, and Jens Spille. Object-based audio for interactive football broadcast. *Multimedia Tools and Applications*, 74(8): 2717–2741, April 2015.

Tobias O.Nyumba, Kerrie Wilson, Christina J. Derrick, and Nibedita Mukherjee. The use of focus group discussion methodology: Insights from two decades of application in conservation. *Methods in Ecology and Evolution*, 9(1):20–32, 2018.

OpenAI. GPT-4 Technical Report, March 2023.

Marko Orel. Comparative analysis of software development methods between parallel, v-shaped and iterative. 169, August 2022a. ISSN 09758887. doi: 10.5120/ijca2017914605.

Marko Orel. *Collaboration Potential in Virtual Reality (VR) Office Space: Transforming the Workplace of Tomorrow*. Springer Nature, August 2022b. ISBN 978-3-031-08180-4. Google-Books-ID: NamAEAAAQBAJ.

Antti Oulasvirta and Jan Blom. Motivations in personalisation behaviour. *Interacting with Computers*, 20(1):1–16, January 2008. ISSN 0953-5438. doi: 10.1016/j.intcom.2007.06.002.

Yusuf Ozkan and Buket D. Barkana. Forensic audio analysis and event recognition for smart surveillance systems. In *2019 IEEE International Symposium on Technologies for Homeland Security (HST)*, page 1–6, November 2019. doi: 10.1109/HST47167.2019.9032996.

Michael O'Donoghue, Alan Hoskin, and Tim Bell. Guidelines for podcast production and use in tertiary education. page 2, 2008.

Yingwei Pan, Yue Chen, Qian Bao, Ning Zhang, Ting Yao, Jingen Liu, and Tao Mei. Smart director: An event-driven directing system for live broadcasting. *ACM Trans. Multimedia Comput. Commun. Appl.*, 17(4), nov 2021.

Bryan Pardo, Antoine Liutkus, Zhiyao Duan, and Gaël Richard. Applying source separation to music. *Audio Source Separation and Speech Enhancement*, pages 345–376, 2018.

Lawrence Pardoe, Lauren Ward, Hannah Clawson, Aimee Moulson, and Chris Pike. *Investigating user interface preferences for controlling background-foreground balance on connected TVs*. June 2020.

Bahadoran Parham, Adan L Benito, Thomas Vassalo, and Joshua D. Reiss. FXive: A Web Platform for Procedural Sound Synthesis. In *Audio Engineering Society*, Milan, Italy, May 2018. Journal of the Audio Engineering Society.

Helen Perks, Thorsten Gruber, and Bo Edvardsson. Co-creation in radical service innovation: a systematic analysis of microlevel processes. *Journal of product innovation management*, 29(6):935–951, 2012.

Elizabeth Perse and Jessica Butler. Call-In Talk Radio: Compensation or Enrichment. *Journal of Radio Studies*, 12:204–222, November 2005.

Virginie Petitjean. Sérialisation et logique de marque ou comment fidéliser les téléspectateurs : l'exemple de TF1. *Entrelacs. Cinéma et audiovisuel*, (HS), February 2008. ISSN 1266-7188.

Christopher William Pike. *Evaluating the Perceived Quality of Binaural Technology*. phd, University of York, January 2019.

Qing Ping and Chaomei Chen. Video highlights detection and summarization with lag-calibration based on concept-emotion mapping of crowd-sourced time-sync comments. *arXiv preprint arXiv:1708.02210*, 2017.

Podcast. *Oxford English Dictionary*. Oxford University Press, 3rd edition, 2008. https://www.oed.com/viewdictionaryentry/Entry/273003.

Podcast Insights. How To Start A Podcast: A Complete Step-By-Step Tutorial (2022 Guide), May 2018. https://www.podcastinsights.com/start-a-podcast/. Accessed July 2022.

Podcast Insights. 2021 Podcast Stats & Facts (New Research From Apr 2021), July 2021. https://www.podcastinsights.com/podcast-statistics/. Accessed 12.05.21.

Podnews.net. 1 in 10 Brits will launch their own podcast in 2022, say Acast, January 2022. https://podnews.net/press-release/acast-new-year. Accessed July 2022.

Karl Raimund Popper. *The logic of scientific discovery*. Routledge, 1992.

O J Postma and M. Brokke. Personalisation in practice: The proven effects of personalisation. *Journal of Database Marketing Customer Strategy Management*, 9(2):137–142, January 2002. ISSN 1741-2447. doi: 10.1057/palgrave.jdm.3240069.

Barbara F. Prince. Podcasts: The potential and possibilities. *Teaching Sociology*, 48(4):269–271, Oct 2020.

John Pruitt and Jonathan Grudin. Personas: practice and theory. In *Proceedings of the 2003 conference on Designing for user experiences*, DUX '03, pages 1–15, New York, NY, USA, June 2003. Association for Computing Machinery.

Radio Works. Acast introduces personalised podcast ads with A Million Ads, December 2018. https://radioworks.co.uk/acast-personalised-podcast-ads-a-million-ads/. Accessed 21.07.21.

Angela T. Ragusa, Anthony Chan, and Andrea Crampton. Ipods Aren't Just for Tunes. *Information, Communication & Society*, 12(5):678–690, August 2009.

RAJAR. Measurement of Internet Delivered Audio Services Winter 2019. Technical report, RAJAR, 2019.

RAJAR. Measurement of Internet Delivered Audio Services Spring 2020. Technical report, RAJAR, 2020.

RAJAR. Measurement of Internet Delivered Audio Services Autumn 2023. Technical report, RAJAR, 2023.

Paul Ramshaw. Is music production now a composition process? *Annual Conference on the Art of RecordProduction*, 2006.

Matthias Rauterberg, Oliver Strohm, and Christina Kirsch. Benefits of user-oriented software development based on an iterative cyclic process model for simultaneous engineering. *International Journal of Industrial Ergonomics*, 16(4-6):391–409, 1995.

Y. Ren, Oriol Nieto, Hendrik Vincent Koops, Anja Volk, and Wouter S. Swierstra. Investigating musical pattern ambiguity in a human annotated dataset. 2018.

Edison Research. Weekly Insights 1.11.2023 - Top Podcast Genres in the U.S. Q3 2022, January 2023. https://www.edisonresearch.com/weekly-insights-1-11-2023-top-podcast-genres-in-the-u-s-q3-2022/. Accesed 05.24.

Jemily Rime, Jon Francombe, and Tom Collins. How Do You Pod? A Study Revealing the Archetypal Podcast Production Workflow. In *ACM International Conference on Interactive Media Experiences*, IMX '22, pages 11–18, New York, NY, USA, June 2022a. Association for Computing Machinery.

Jemily Rime, Chris Pike, and Tom Collins. What is a podcast? considering innovations in podcasting through the six-tensions framework. page 13548565221104444, June 2022b. ISSN 1354-8565. doi: 10.1177/13548565221104444.

Jemily Rime, Alan Archer-Boyd, and Tom Collins. How will you pod? implications of creators' perspectives for designing innovative podcasting tools. 20(3), oct 2023. ISSN 1551-6857. doi: 10.1145/3625099.

Riverside.fm. How To Produce A Podcast in 2021- Riverside.fm, 2022. https://riverside.fm/blog/how-to-produce-a-podcast. Acessed July 2022.

Molly Robson. Intimacy in Isolation: Podcasting, Affect, and the Pandemic. *Perspectives in Biology and Medicine*, 64(3):388–407, 2021. ISSN 1529-8795. Publisher: Johns Hopkins University Press.

Frank Rose. The power of immersive media. *strategy+business*, page 54, February 2015.

Brent Rosso. Creativity and Constraints: Exploring the Role of Constraints in the Creative Processes of Research and Development Teams. *Organization Studies*, 35:551–585, March 2014.

B. Rousseau, W. Jouve, and L. Berti-Equille. Enriching multimedia content description for broadcast environments: from a unified metadata model to a new generation of authoring tool. In *Seventh IEEE International Symposium on Multimedia (ISM'05)*, December 2005.

Boris Rousseau, Parisch Browne, Paul Malone, and Mícheál Ó Foghlú. User profiling for content personalisation in information retrieval. Nicosia, Cyprus, 2004.

M. W. Rowe. The Definition of 'Game'. *Philosophy*, 67(262):467–479, 1992. ISSN 0031-8191. Publisher: Cambridge University Press.

Marie-Laure Ryan. Immersion vs. interactivity: Virtual reality and literary theory. *SubStance*, 28(2):110, 1999.

Soumya Sai Vanka, Maryam Safi, Jean-Baptiste Rolland, and György Fazekas. Adoption of ai technology in music mixing workflow: An investigation. Audio Engineering Society, May 2023.

Joni Salminen, Kathleen Wenyun Guan, Soon-Gyo Jung, and Bernard Jansen. Use Cases for Design Personas: A Systematic Review and New Frontiers. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, pages 1–21, New York, NY, USA, April 2022. Association for Computing Machinery.

Elizabeth Sanders and Pieter Jan Stappers. Co-creation and the new landscapes of design. *CoDesign*, 4:5–18, 03 2008. doi: 10.1080/15710880701875068.

Masanori Sano, Werner Bailer, Alberto Messina, Jean-Pierre Evain, and Mike Matton. The mpeg-7 audiovisual description profile (avdp) and its application to multi-view video. In *IVMSP 2013*, page 1–4, Seoul, Korea (South), June 2013. IEEE.

Daniela Schlütz and Imke Hedder. Aural parasocial relations: Host–listener relationships in podcasts. *Journal of Radio Audio Media*, 29(2):457–474, July 2022. ISSN 1937-6529. doi: 10.1080/19376529.2020.1870467.

Kristen M. Scott, Simone Ashby, David A. Braude, and Matthew P. Aylett. Who owns your voice? ethically sourced voices for non-commercial tts applications. In *Proceedings of the 1st International Conference on Conversational User Interfaces*, CUI '19, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450371872.

Kate Seers. What is a qualitative synthesis? *Evidence-Based Nursing*, 15(4): 101–101, October 2012. ISSN 1367-6539, 1468-9618.

Roneel V. Sharan and Tom J. Moir. An overview of applications and advancements in automatic sound recognition. *Neurocomputing*, 200:22–34, August 2016. ISSN 0925-2312. doi: 10.1016/j.neucom.2016.03.020.

Tzlil Sharon. Peeling the pod: towards a research agenda for podcast studies. *Annals of the International Communication Association*, 47(3):324–337, July 2023. ISSN 2380-8985. doi: 10.1080/23808985.2023.2201593.

Tzlil Sharon and Nicholas A. John. Imagining an ideal podcast listener. *Popular Communication*, 17(4):333–347, October 2019. ISSN 1540-5702, 1540-5710. doi: 10.1080/15405702.2019.1610175.

M. R. Sheldon, M. J. Fillyaw, and W. D. Thompson. The use and interpretation of the Friedman test in the analysis of ordinal-scale data in repeated measures designs. *Physiotherapy Research International: The Journal for Researchers and Clinicians in Physical Therapy*, 1(4):221–228, 1996.

Lindsey A. Sherrill. The "Serial Effect" and the True Crime Podcast Ecosystem. *Journalism Practice*, 16(7):1473–1494, August 2022. ISSN 1751-2786. Publisher: Routledge.

Ben Shirley and Rob Oldfield. Clean audio for tv broadcast: An object-based approach for hearing-impaired viewers. *Journal of the Audio Engineering Society*, 63(4):245–256, April 2015. ISSN 15494950. doi: 10.17743/jaes.2015.0017.

Ben Shirley, Melissa Meadows, Fadi Malak, James Woodcock, and Ash Tidball. Personalized object-based audio for hearing impaired tv viewers. *Journal of the Audio Engineering Society*, 65(4):293–303, April 2017. ISSN 15494950. doi: 10.17743/jaes.2017.0005.

Ben Shirley, Lauren Ward, and Emmanouil Theofanis Chourdakis. Personalization of object-based audio for accessibility using narrative importance. In *ACM International Conference on Interactive Experiences for Television and Online Video*, page 1–5. IEEE, June 2019.

Than Htut Soe. Ai video editing tools. what editors want and how far is ai from delivering? (arXiv:2109.07809), September 2021. doi: 10.48550/arXiv.2109.07809. arXiv:2109.07809 [cs].

Clay Spinuzzi. The methodology of participatory design. *Technical communication*, 52(2):163–174, 2005.

Spotify. Podcasts Homepage, 2022. https://open.spotify.com/genre/podcasts-web. Accessed July 2022.

Spotify for Artists. Audio file formats for Spotify, March 2021. https://artists.spotify.com/help/article/audio-file-formats. Accessed 07.08.21.

Spotify Newsroom. Get to Know Your Favorite Podcasts Even Better With New Polls Feature, September 2020a. https://newsroom.spotify.com/2020-09-23/get-to-know-your-favorite-podcasts-even-better-with-new-polls-feature/. Accessed 02.07.21.

Spotify Newsroom. The Trends That Shaped Streaming in 2020, December 2020b. https://newsroom.spotify.com/2020-12-01/the-trends-that-shaped-streaming-in-2020/. Accessed 02.07.21.

Jason E. Squire. *The movie business book*. Fireside, 1992.

Statista. U.S. podcast advertising share by genre 2019. https://www.statista.com/statistics/1124145/distribution-podcast-advertising-usa-genre/. Accessed 09.02.21.

Stereo. Homepage, 2022. https://stereo.com. Accessed July 2022.

Jonathan Steuer. Defining virtual reality: Dimensions determining telepresence. In Frank Biocca and Mark R. Levy, editors, *Communication in the Age of Virtual Reality*, chapter 3, pages 33–56. Routledge, 1995.

Bronson K. Strickland, Jarred M. Brooke, Mitchell T. Zischke, and Marcus A. Lashley. Podcasting as a tool to take conservation education online. *Ecology and Evolution*, 11(8):3597–3606, 2021. ISSN 2045-7758.

Peiqi Sui, Lin Wang, Sil Hamilton, Thorsten Ries, Kelvin Wong, and Stephen Wong. Mrs. dalloway said she would segment the chapters herself. In *Proceedings of the The 5th Workshop on Narrative Understanding*, page 92–105, Toronto, Canada, July 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.wnu-1.15.

John Sullivan, Patricia Aufderheide, Tiziano Bonini, Richard Berry, and Dario Llinares. Podcasting in transition: formalization and its discontent. *AoIR Selected Papers of Internet Research*, Oct 2020. ISSN 2162-3317.

John L. Sullivan. Podcast Movement: Aspirational Labour and the Formalisation of Podcasting as a Cultural Industry. In *Podcasting: New Aural Cultures and Digital Media*, pages 35–56. Springer International Publishing, Cham, 2018.

Charles S Swartz. *Understanding Digital Cinema: A Professional Handbook*. Routledge, New York, Oct 2004.

Tetsuya Takiguchi, Jun Adachi, and Yasuo Ariki. Audio-based video editing with two-channel microphone. In *2008 International Conference on Multimedia and Ubiquitous Engineering (mue 2008)*, page 282–287, April 2008. doi: 10.1109/MUE.2008.86.

Haining Tan, Tao Ye, Sadaqat Ur Rehman, Obaid Ur Rehman, Shanshan Tu, and Jawad Ahmad. A novel routing optimization strategy based on reinforcement learning in perception layer networks. *Computer Networks*, 237:110105, December 2023.

The Conversations Network, 2013. http://web.archive.org/web/20130729204535/http://www.conversationsnetwork.org/history.

The New York Times. 'Podcast Movies'? Feature-Length Fiction Stretches the Medium, 2022. https://www.nytimes.com/2021/12/24/arts/podcast-movies-fiction.html. Accessed July 2022.

The Orpheus Project. The Mermaid's Tears, 2017. http://mermaidstears.ch.bbc.co.uk/; Accessed July 2022.

The Podcast Host. How to Start a Podcast: Every Single Step for 2022, December 2021. https://www.thepodcasthost.com/planning/how-to-start-a-podcast/. Accessed July 2022.

The Simplecast Blog. What's the Difference Between Streams and Downloads?, September 2019. https://blog.simplecast.com/whats-the-difference-between-streams-and-downloads/. Accessed July 2022.

Limbik Theatre. Ambisonic stories, 2022. https://www.limbiktheatre.com/ambisonic-stories/ Accessed 09.2022.

Theodoros Theodorou, Iosif Mporas, and Nikos Fakotakis. An overview of automatic audio segmentation. *International Journal of Information Technology and Computer Science*, 6:1–9, October 2014. doi: 10.5815/ijitcs.2014.11.01.

Lennox Thomas, Julia MacMillan, Emilly McColl, C Hale, and S Bond. Comparison of focus group and individual interview methodology in examining patient satisfaction with nursing care. *Social Sciences in Health*, 1(4):206–220, 1995.

Amie L. Thomasson. Ontological Innovation in Art. *The Journal of Aesthetics and Art Criticism*, 68(2):119–130, May 2010. ISSN 0021-8529.

Chad C. Tossell, Philip Kortum, Clayton Shepard, Ahmad Rahmati, and Lin Zhong. An empirical analysis of smartphone personalisation: measurement and user variability. *Behaviour & Information Technology*, 31(10):995–1010, October 2012.

Catherine M. Traylor. Serialized killing : usability and user experience in the true crime genre. July 2019.

Ayushi Trivedi, Navya Pant, Pinal Shah, Simran Sonik, and Supriya Agrawal. Speech to text and text to speech recognition systems-Areview. page 9, 2018.

Anh Truong, Floraine Berthouzoz, Wilmot Li, and Maneesh Agrawala. Quickcut: An interactive tool for editing narrated video. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST '16, page 497–507, New York, NY, USA, October 2016. Association for Computing Machinery. ISBN 978-1-4503-4189-9. doi: 10.1145/2984511.2984569.

Matthew Turner, Robert Lowe, and Matthew Schaefer. Professional development and research engagement through podcasting. *ELT RESEARCH 35, 2020, ReSIG*, 02 2020.

G. Tzanetakis and F. Cook. A framework for audio analysis based on classification and temporal segmentation. volume 2, page 61–67 vol.2, September 1999. doi: 10.1109/EURMIC.1999.794763.

Stefan Uhlich, Marcello Porcu, Franck Giron, Michael Enenkl, Thomas Kemp, Naoya Takahashi, and Yuki Mitsufuji. Improving music source separation based on deep neural networks through data augmentation and

network blending. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 261–265. IEEE, 2017.

Marian Ursu, Davy Smith, Jonathan Hook, Shauna Concannon, and John Gray. Authoring interactive fictional stories in object-based media (obm). In *ACM International conference on interactive media experiences*, pages 127–137, 2020a.

Marian Ursu, Davy Smith, Jonathan Hook, Shauna Concannon, and John Gray. Authoring Interactive Fictional Stories in Object-Based Media (OBM). In *ACM International Conference on Interactive Media Experiences*, pages 127–137, Cornella, Barcelona Spain, June 2020b. ACM.

Marian F. Ursu, Maureen Thomas, Ian Kegel, Doug Williams, Mika Tuomola, Inger Lindstedt, Terence Wright, Andra Leurdijk, Vilmos Zsombori, Julia Sussner, Ulf Myrestam, and Nina Hall. Interactive tv narratives: Opportunities, progress, and challenges. *ACM Trans. Multimedia Comput. Commun. Appl.*, 4(4), nov 2008.

Yoni Van Den Eede. Imagining things: unfolding the "of" in philosophy of technology, through object-oriented ontology'. In Heather Wiltse, editor, *Relating to Things: Design, Technology and the Artificial*, chapter 10, pages 191–214. Bloomsbury Publishing, May 2020.

Connie K Varnhagen, Matthew Gushta, Jason Daniels, Tara C Peters, Neil Parmar, Danielle Law, Rachel Hirsch, Bonnie Sadler Takach, and Tom Johnson. How informed is online informed consent? *Ethics & Behavior*, 15(1):37–48, 2005.

Emmanuel Vincent, Tuomas Virtanen, and Sharon Gannot. *Audio source separation and speech enhancement*. John Wiley & Sons, 2018.

James Keith Vonderhaar. Production of a Radio Program Series. Master's thesis, Central Michigan University, United States – Michigan, 1983.

Jorge Vázquez-Herrero and Xosé López-García. When media allow the user to interact, play and share: recent perspectives on interactive documentary. *New Review of Hypermedia and Multimedia*, 25(4):245–267, October 2019. ISSN 1361-4568. doi: 10.1080/13614568.2019.1670270.

Scott Allan Wallick. Christopher lydon interviews : All the lydon interviews in one download, 2003. http://blogs.harvard.edu/lydondev/all-the-lydon-interviews-in-one-download/. Accessed 05.24.

Yuxuan Wang, R. J. Skerry-Ryan, Daisy Stanton, Yonghui Wu, Ron J. Weiss, Navdeep Jaitly, Zongheng Yang, Ying Xiao, Z. Chen, Samy Bengio, Quoc V. Le, Yannis Agiomyrgiannakis, Robert A. J. Clark, and Rif A. Saurous. Tacotron: Towards end-to-end speech synthesis. In *Interspeech*, 2017.

Zheng Wang, Jie Zhou, Jing Ma, Jingjing Li, Jiangbo Ai, and Yang Yang. Discovering attractive segments in the user-generated video streams. *Information Processing Management*, 57(1):102130, January 2020. ISSN 0306-4573. doi: 10.1016/j.ipm.2019.102130.

Lauren Ward. *Improving broadcast accessibility for hard of hearing individuals: using object-based audio personalisation and narrative importance.* PhD thesis, University of Salford, 2020.

Lauren Ward, Matthew Paradis, Ben Shirley, Laura Russon, Robin Moore, and Rhys Davies. Casualty accessible and enhanced (a&e) audio: Trialling object-based accessible tv audio. In *Audio Engineering Society Convention 147*. Audio Engineering Society, 2019.

Lauren Ward, Maxine Glancy, Sally Bowman, and Michael Armstrong. The impact of new forms of media on production tools and practices. *IBC2020, Sep*, 2020.

Lauren A. Ward. Accessible broadcast audio personalisation for hard of hearing listeners. *Adjunct Publication of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video*, 2017.

Jessie L Werner, Resa E Lewiss, Gita Pensa, and Alyson J McGregor. Women in podcasting: We should tune in. *The permanente journal*, 24, 2020.

Kris West and Stephen Cox. Finding an optimal segmentation for audio genre classification.

Heather Wiltse. *Relating to Things: Design, Technology and the Artificial.* Bloomsbury Publishing, May 2020. ISBN 978-1-350-12427-1.

Bob G. Witmer and Michael J. Singer. Measuring Presence in Virtual Environments: A Presence Questionnaire, Jun 1998.

Stephen Witt. *How music got free: the end of an industry, the turn of the century, and the patient zero of piracy.* Viking, New York, 2015.

Jasper Wyld. Collaborative storytelling and canon fluidity in the adventure zone podcast. *Convergence*, 27(2):343–356, 2021. doi: 10.1177/1354856520950555.

Bruno Xavier Leitão, Álan LV Guedes, and Sérgio Colcher. Toward web templates support in nested context language. In *Applications and Usability of Interactive TV: 8th Iberoamerican Conference, jAUTI 2019, Rio de Janeiro, Brazil, October 29–November 1, 2019, Revised Selected Papers 8*, pages 16–30. Springer, 2020.

Huan Yang, Baoyuan Wang, Stephen Lin, David Wipf, Minyi Guo, and Baining Guo. Unsupervised extraction of video highlights via robust recurrent auto-encoders. page 4633–4641, 2015.

Jing Yang, Amit Barde, and Mark Billinghurst. Audio augmented reality: A systematic review of technologies, applications, and future research directions. *Journal of the Audio Engineering Society*, 70(10):788–809, November 2022. ISSN 15494950. doi: 10.17743/jaes.2022.0048.

Zongyu Yin, Federico Reuben, Susan Stepney, and Tom Collins. "a good algorithm does not steal–it imitates": The originality report as a means of measuring when a music generation algorithm copies too much. In *Artificial Intelligence in Music, Sound, Art and Design: 10th International Conference, EvoMUSART 2021, Held as Part of EvoStar 2021, Virtual Event, April 7–9, 2021, Proceedings 10*, pages 360–375. Springer, 2021.

Zongyu Yin, Federico Reuben, Susan Stepney, and Tom Collins. Deep learning's shallow gains: a comparative evaluation of algorithms for automatic music generation. *Machine Learning*, pages 1–38, 2023.

Babangida Zachariah and Ogwueleka Francisca Nonyelum. A Comparative Analysis of Requirement Gathering Techniques, 2020.

Chenyan Zhang, Andrew Perkis, and Sebastian Arndt. Spatial immersion versus emotional immersion, which is more immersive? *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, 2017.

Emre Çakır, Giambattista Parascandolo, Toni Heittola, Heikki Huttunen, and Tuomas Virtanen. Convolutional recurrent neural networks for polyphonic sound event detection. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25(6):1291–1303, June 2017. ISSN 2329-9304. doi: 10.1109/TASLP.2017.2690575.