Object-Based Radio: Effects On Production and Audience Experience

Anthony W P Churnside

School of Computing, Science & Engineering
The University of Salford, Salford, UK

Submitted in Partial Fulfilment of the Requirements of the
Degree of Doctor of Philosophy, August 2015

Contents

List of Figures

List of Tables

Declaration

This thesis is a presentation of the author's own work. The work described in chapters 3 and 5 was collaborative in nature and the roles of contributors are detailed below.

Collaborative Work

For the work in chapter 3 the author acted as supervisor for Dr Mark Mann during Dr Mann's traineeship at BBC R&D. The author conceived the experience and created the prototype audio player which was provided to Dr Mann. Under the supervision of the author, Dr Mann created the final audio player which allowed audio streams to be balanced and captured all user interaction with the player. The user interface design for the football study shown in figure 3.3 was designed by Ms Jasmine Cox. The audio streams which delivered the football experience were set up by Mr Matthew Paradis and Mr Andy Armstrong. In chapter 5 the concept of Perceptive Media was co-invented by the author and Mr Ian Forrester. The HTML5 audio player used in "Breaking Out" in chapter 5 was created by Happy Worm. The relationship with Happy Worm was managed by Mr Forrester, while the author defined the technical requirements and system architecture. The online survey questions for study in chapter 5 were written by the author in collaboration with Mr Forrester. The focus groups in chapter 5 were conducted by Dr Maxine Glancy based on the online survey questions, and the focus group responses quoted and analysed in this thesis were collected by Dr Glancy.

Previous Publications

Elements of this work have been published previously in journal and conference papers written or co-written by the author. Parts of the work in the literature review (chapter 2) formed the basis of "Emerging Technology Trends in Spatial Audio," [1] a summary of spatial audio research trends published in the SMPTE Motion Imaging Journal in 2012.

The work in chapter 3 was included in "Object-Based Audio Applied to Football Broadcasts: The 5 live Football Experiment," [2] a paper published in 2013 by Association for Computing Machinery. The work in chapter 5 was published in "The Creation of a Perceptive Audio Drama," [3] and presented at the New European Media Summit in 2012. Finally, concepts and conclusions discussed in chapter 6 and 7 were included by the author in his contribution to "Object-based Broadcasting - Curation, Responsiveness and User Experience," [4] a paper published in 2014 by the International Broadcasting Convention.

Data and Copyright

The data used for this thesis is the property of BBC R&D and where the author has created figures from data collected while employed by the BBC or where BBC R&D has previously published elements of this work the copyright is indicated in footnotes, as requested by the BBC.

Acknowledgements

I would like to thank supervisor Professor Trevor Cox and co-supervisor Dr Ben Shirley for their support, guidance and encouragement over the last four years.

I would also like to express gratitude to Dr Mark Mann, Andrew Bonney, Matthew Paradis and Andy Armstrong for their contribution to the football study in chapter 3.

I would like to thank Steve Brooke, Nadia Mollinari, Linda Marshall Griffiths and the rest of the cast for their roles in creating the object-based radio drama "Pinocchio" in chapter 4. I would also like to thank Tom Parnell and Stephen Rinker for their time creating audio mixes for the foreground vs background study in chapter 4.

I also would like to acknowledge Ian Forrester, Dr Maxine Glancy, Sharon Sephton, Henry Swindell and Sarah Glenister for their roles in the creation of object-based radio drama "Breaking Out" in chapter 5.

I would also like to thank, BBC R&D, BBC Radio Drama, BBC Radio 4, BBC Radio 5 Live for allowing me to experiment with their content and audiences.

Finally, I would like to thank Samantha Chadwick and Dr Frank Melchior for their professional support which ultimately allowed this research to take place.

Abstract

This thesis analyses the benefits of using object-based audio as a production and delivery format in order to enable new audience experiences. This is achieved though a series of case studies, each focusing on a different user experience enabled by the use of object-based audio. Each study considers the impact of using object-based audio on the creative process, production workflow and audience experience.

The first study analyses the audience's use of the ability to personalise the mix of a live football match. It demonstrates that there was not a single audio mix favoured by all, and the ability to change the mix was valued by the audience. While listeners did adjust the mix initially, they tended to leave it at that setting and did not interact much once they made their initial selection. While there were three favoured mixes, over 50% of listeners did not choose one of these three mixes, indicating that only offering three options would not satisfy everyone.

Modes of listening model the ways listeners deconstruct complex sound scenes into foreground and background categories ascribing different salience to foreground and background sounds. The second study uses this model to inform a series of card sorting exercises which result in similar foreground and background categories. However, rather than being unimportant, background sounds were present to convey ancillary information or to affect emotional responses and foreground sounds to expose plot or story events. This study demonstrated that this grouping was a meaningful categorisation for broadcast sound and evaluated how beneficial allowing different foreground and background audio mixes would be for audiences. It contains analysis of audio objects in the context of foreground and background sounds based on the opinions of the content creators. It also includes subjective testing of audience preferences for different mixes of foreground verses background audio levels across five different genres and four different loudspeaker layouts. It shows that there is no clustering of listeners based on their preference of foreground vs background balances. It also shows that there is significant variation of foreground and background balance preference between loudspeaker layouts.

The final study goes beyond tailoring audio levels, balances and loudspeaker layouts and analyses the benefit to audiences of being able to adapt the story of a drama in order to set it in a location that is familiar to the listener. It shows that being able to set a radio drama in the location where the listening is taking place improves audience's enjoyment of the programme. 75% of listeners who experienced the tailored version of the drama reported liking the story, compared with 65% of listeners who experienced a non-tailored version.

The three studies also analyse the impact of object-based content creation on production workflows by documenting the challenges faced and discussing possible solutions. For example, providing writers with constraints when they are designing dynamic content and allowing sound designers time to develop trust in the technology when mixing content for multiple loudspeaker layouts.

The original contribution to knowledge is to establish a new listening model applicable to constructed and designed sound experiences based on functional analysis of audio objects. This work also establishes, for the first time, a framework for the definition of an audio object based on the creator's intended range of audience experiences. In addition the thesis also provides insights into how audiences interact with object-based content experiences and insights about audience attitudes towards using personal data to personalise object-based content experiences. Each study addresses the potential advantages of delivering object-based audio, assess any impact on the quality of the audience's experience and analyses the challenges faced by production in the creation of these new experiences.

1

Introduction

The concept of audio objects has been applied in the computer games industry in order to enable interaction, however it is a relatively novel concept in a professional broadcasting environment. Audio objects are beginning to be adopted into film production workflows for delivery to cinemas with systems like Dolby Atmos [5], DTS MDA [6], and IOSONO [7].

The primary application of these systems, and perhaps the most thoroughly researched aspect of future audio formats is the potential to provide spatial audio and 3D sound. There is a large body of research concerned with 3D sound recording, production and reproduction. There is also a body of research that attempts to quantify benefits of delivering surround sound with height, although there is still much debate around how to quantify the benefits of delivering 3D surround sound. Consumer take-up of surround sound systems has been minimal, perhaps demonstrating that there is currently limited interest from audiences in hearing any form of surround sound beyond basic two channel stereo. This work analyses the benefits to using object-based audio beyond delivering spatial audio, and aims to understand the impact of the use of audio objects on the creative process, production workflow and audience experience.

1.1 Thesis Structure

This thesis is structured in three sections, the introduction and theory in chapters 1 and 2, a series of three studies in chapters 3 to 5, and the discussion and conclusion in chapters 6 and 7. For each of the studies content has been created by professional BBC content creators, the production processes of this content has been analysed, and finally the audience

experience of the content is assessed and analysed using a variety of methodologies. The content for each study is listed below.

- Study 1 in chapter 3 contains analysis of the production and reception of a live football match from Wembley stadium, where audiences were allowed to control the crowd vs commentary mix and the home vs away crowd levels.
- Study 2 in chapter 4 contains a analysis of the production of a radio drama "Pinocchio" and music recording of "Everything Everything", followed by a laboratory experiment to test listener's foreground vs background preference levels across a number of genres and playback systems.
- Study 3 in chapter 5 contains analysis of the production of an object-based radio drama "Breaking Out" which dynamically sets itself in the location where it is being experienced. Followed by an analysis of the audience reception of the content and attitude towards it.

1.2 Context

Audio broadcasting workflows can be split roughly into four stages; capture, production, distribution and reproduction (these are shown in figure 1.1). Workflows of specific genres can be broken down differently, for example radio drama has to be written, but all radio workflows can be modelled using these stages. Until now it is the final stage, the reproduction system, that has defined the capture, production and distribution stages. When producing 5.1 [9] surround audio content it is the ITU definition of the speaker positions that defines the whole production and mixing process; the programme is referred to as a "5.1 production". As audiences become more diverse, consumer electronics prices drop, and as consumer choice increases these listening conditions are diversifying. Consumer technology now allows CPU intensive processes like ambisonic decoding and binaural rendering to be performed in real-time. Although there is a trend towards increasing the number of channels, for example NHK's 22.2 surround sound [8] the channel paradigm carries with it a number of potential drawbacks, identified below.

- Inflexible loudspeaker positions. The loudspeaker locations define the whole broadcasting workflow and placing loudspeakers in locations other than those recommended will spatially distort the sound scene.
- Incompatibility between formats with different numbers of channels. Monophonic, stereophonic, and 5.1 surround are not compatible with each other, requiring compatibility checking and up-mix and down-mix rules to ensure the best compromise

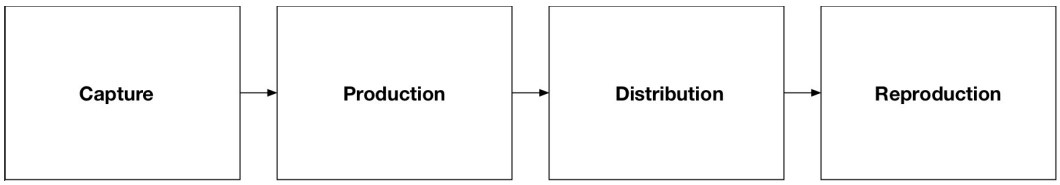


Figure 1.1: Broadcasting stages

when converting between formats.

- Inflexible data usage. While it might be sensible to represent a full orchestra using 22.2 channels, representing a single news reader in this way would be highly bandwidth inefficient.
- Limited ability to personalise content. For a listener with hearing difficulties who wishes the background music or sound effects to be quieter in the mix compared to the dialogue, a channel-based system lacks the flexibility to grant that wish.
- No clear way to deal with interactive content. Channel-based systems offer no obvious way for the user to interact with and explore audio scenes.

The use of audio objects for broadcasting content could address some of the points listed above. This work is concerned with developing the understanding of what an audio object is (from the point of view of both content producer and audience member), how this method of content production can allow new audience experiences and new forms of content, and whether these new experiences offer any improvement to current linear content experiences.

1.3 Aims

The aims of this thesis fall in to three general areas.

- Object-based content creation - Analyse the impact of using audio objects for content creation compared with existing traditional production workflows.
- Object-based content experience - Develop a series of test methods for assessing any benefits to listeners of using audio objects.
- Audio object definition - To understand how content producers view audio objects, what they consider to be an audio object and how they use audio objects in the context of both speech and non-speech content. In doing so, develop a framework for defining what an audio object is in the context of a piece of audio content.

2

Literature Review

2.1 State of the Art

High Definition Television began broadcasting 5.1 surround sound in 2006. Although 5.1 channels are the highest number of channels currently being broadcast in the UK, there have been some experimental 22.2 channel IP broadcasts. 22.2 channels is a standard proposed by Japanese broadcaster NHK [8].

Industry debates currently centre around three approaches to representing an audio scene.

- Channel-Based
- Scene-Based
- Object-Based

Each is a different approach and comes with advantages and disadvantages described below.

2.1.1 Channel-Based

Channel-based audio consists of discrete streams of audio data, each associated with a loudspeaker of a given position relative to the listener. Channel-based content requires a specific loudspeaker layout, and different loudspeaker layouts are often incompatible. Layouts in common use include Mono, Stereo and 5.1 Surround Sound. A channel-based approach has additional inflexibility, for example the inability to independently control any relative pa-

rameter of different sounds within a scene, such as temporal or spatial position, or loudness.

2.1.2 Scene-Based

Ambisonics and Higher Order Ambisonics are the key examples of a scene based approach to representing sound. Ambisonics was a technology developed in the 1970s. First order ambisonics uses four channels of audio data to represent a full 360 soundfield [10]. Original four channel (first order) Ambisonics has since been extended to include higher spherical harmonics which are demonstrated to provide more accurate localisation at a cost of computational complexity and bandwidth [11]. A scene-based approach does not require a specific loudspeaker layout, but due to the dependancy between signals still has the inability to independently control relative loudness of different sounds within a scene. However, certain Ambisonic transformations are possible, for example rotating the whole soundfield.

2.1.3 Object-Based

An object-based approach represents the scene as a set of independent sounds. Each of these sounds is accompanied by a set of metadata which contains descriptive data, such as level and position, to enable a renderer to reproduce the scene. This approach has the drawback of requiring some computational complexity on the part of the playback system in order to perform the rendering, but offers advantages of being free of a specified loudspeaker layout and being able to independently control different elements within a piece of content.

2.2 Object-Based Metadata Formats

A number of technical solutions for describing object-based audio scenes exist. These formats contain metadata that describe physical attributes of an accompanying audio signal. Different formats offer different capabilities due to the different physical attributes represented by the format. What follows is a brief critique of commonly used formats.

Virtual Reality Modelling Language (VRML) [12] allows the representation of virtual reality scenes. This language can reference audio files (PCM *.wav files) and define parameters such as start and stop times, pitch and location (using a coordinate system). VRML was developed for visual scenes and is not specifically geared to audio. The metadata associated

with VRML treats audio objects as sources. Perhaps the best known object-based format is MPEG-4. MPEG-4 includes a standard for object-based multimedia streaming and supports audio components. MPEG-4 uses Audio BIFS [13], (Audio Binary Format for Scene Description). Audio BIFS is similar to VRML, but extends the functionality. Although the MPEG-4 standard is widely used the BIFS is rarely implemented, the stamp "MPEG-4" mainly being associated with audio and video data compression standards and compatibility. Carouso [14], is a project that uses the MPEG-4 standard to deliver object-based media to allow real time interaction with 3D audio using Wave Field Synthesis. The project took a technical approach to audio object definitions basing the scene descriptions on physical signifiers rather than perceptual or experiential aspects of the sound. What is considered the biggest success of MPEG-4 is also considered its biggest failure; its complexity. More recently, MPEG-H Audio Open International Standard ISO/IEC23008-3 [15] published a standard which supports audio objects specifically to allow personalisation of audio content.

Fascinate [16] was an EU project that used a combination of Higher Order Ambisonics and audio objects to render audio scenes using ambisonics. This allowed real time spatial repositioning of audio scenes using transformations such as rotation and zoom. The project used a very high resolution panoramic video which was croppable using pan-scan techniques and the audio was rendered to match the visual cropping, panning and digital zooming. This research was driven by interaction around a visual display, therefore audio objects tended to be associated with visual objects within the scene, for example a football match was recorded and the noise from the football was treated as a separate audio object to that of the crowd. Another sports application discussed further in section 2.8 used SAOC (Spatial Audio Object Coding) [17]. SAOC is designed to carry parameterised audio objects in a backwards compatible stream. SAOC consists of a stereo stream and side-data to allow extraction of the audio objects. The approach to SAOC has been applied to MPEG Surround [18], a coding solution adopted by World DMB, to stream 5.1 surround sound over a backwards compatible stereo stream with a small amount of extra data to represent the additional signals. SAOC allows the user to adjust parameters such as volume and location of audio objects. This technology has been demonstrated for applications such as karaoke.

A number of other XML metadata based approaches such as SMIL (Synchronised Multimedia Integration Language) [19] exist which attempt to create a standard to allow users to create multimedia presentations. The Audio Scene Description Format [20] is a format for the storage and exchange of static, dynamic and interactive spatial audio content. This is an XML based format similar to SMIL but is dedicated to audio. Most recently, BWAV (Broadcast Wave Format) also offers the ability to contain audio objects and in January 2014 the European Broadcasting Union included support for audio objects in the BWAV audio definition model [21]. This format contains a metadata structure that is able to

represent audio objects with positions in space. While the work in this research paper contributed to the defining of this standard, the format is limited to linear content as the time variable within the audio definition model is linear, audio object start and lengths/end times are static and non-variable, unlike the positional metadata.

EDLs (Edit Decision Lists) and project files for DAWs (Digital Audio Workstations) such as Nuendo, Logic and Pro Tools carry similar metadata to the formats considered above. There are also a number of exchange formats for DAWs such as AES32 and OMF for transferring audio projects between digital audio workstations. These formats are channel-based rather than object-based and represent audio as a set of channel strips with associated production metadata. The production metadata would include routing and panning information, but these formats are intended as production rather than distribution or playback formats. There are also bespoke audio computer game formats, but there is little published about these as they are the property of computer games studios who tend to be protective of their intellectual property.

The formats described in this section support dynamic three dimensional audio positioning to various degrees, however, all these metadata formats for the representation of audio as objects or assets are born from technical discussions and specifications, there is little work conducted to form specification and understanding on the basis of desired audience experience. The parameters described by these formats tend to be founded on physical or signal level objects rather than perceptual characteristics (this is discussed further in section 2.10.1). The fact these formats focus on physical rather than perceptual attributes may be because the use of metadata to render the audio is primarily a physical and functional process, for example use for VBAP rules. While the use of a perceptual level objects in these formats might help producers and sound designers in constructing and manipulating sound experiences, it would also further abstract the data from rendering process which might explain why these formats seem to concern themselves with physical attributes rather than perceptual models. Also the systems tend to be designed by engineers rather than content producers or storytellers. This work demonstrates the need for an object definition based on the audience experience as desired by the producer of the content.

2.3 Object-Based Audio Sound Scene Creation

The largest body of research relating to object-based audio is around the rendering of objects using panning [22], binaural [23], ambisonic [10] and wave front reconstruction (WFR) [11] methods. The “Pinocchio” drama in chapter 4 uses a series of loudspeaker

arrays to assess the impact of playback array on foreground vs background mix preference. A combination of technologies were used to create the examples for the other two studies, what follows is an overview of the available approaches that was written to allow informed and justifiable decisions to be made during the production process.

2.3.1 Panning

Vector Base Amplitude Panning (VBAP) [22] is the most commonly used form of pairwise panning [24]. Sine and tangent laws are examples of basic pairwise panning laws [25]. These types of panning can exhibit the undesired characteristic of virtual sources appearing to come from the loudspeaker located nearest to the listener. [26]. Some of the current industry solutions use this approach, for example Dolby's Atmos and DTS's MDA player. Standard pairwise or vector base panning of audio objects is probably the simplest of production methods, however it allows only relative placement of audio objects, azimuth and elevation, but not distance. Distance can however be implied using relative level and mixing reverberation.

2.3.2 Binaural

Binaural sound is aimed at headphone listening. Binaural rendering takes binaural cues such as inter-aural level, time and frequency differences represented by head related transfer functions (HRTFs) to create the perception that sound is coming from a point in space. This can be performed in real time using HRTF data [23] and head tracking. A common approach to rendering binaural is to use virtual loudspeaker and room binaural impulse responses convolved with loudspeaker signals. Binaural sound would typically be listened to using headphones because the convolution processes described results in signals that need to be presented directly at the ear canal. Binaural is often used as an experimental rendering method because it is quick and simple to set up a pair of headphones and a laptop. However, unless content is produced specifically to be played back on headphones, binaural adds an extra layer of processing and margin of error to the rendering process which can negatively effect results and is more difficult to defend as an experiment method.

2.3.3 Wave Field Synthesis

Wave field synthesis (WFS) is based on Huygens' Principle [27], which states a waveform can be constructed from the superposition of elementary spherical waves. Therefore, using

a large number of closely packed loudspeakers it is theoretically possible to create any wavefront. Daniel [11] discusses WFS and its limitations due to loudspeaker numbers are mathematically demonstrated. The biggest drawback of WFS is that for a physically accurate reconstruction of a wavefront a very large number of closely packed loudspeakers is required.

2.3.4 Hybrid Methods and Commercially Available Production Systems

Recently, a number of hybrid methods of audio rendering have been made commercially available. IOSONO have a rendering system which primarily targets cinemas and live performance venues. This system is based on Wave Field Synthesis, however it has been optimised to work with fewer loudspeakers [7]. Dolby have recently announced Dolby ATMOS [5]. Details of the how the system works are limited, but it is driven by a combination of channel-based audio (5.1 and 7.1) for beds and atmospheres, and discrete audio objects which are rendered to the loudspeaker array using panning at the point of playback. While Dolby does not provide explicit guidelines for what sounds should be beds or atmospheres (and therefore treated as channel based) it is likely that it was Dolby's recognition of established surround sound recording and production tools and techniques which resulted in their continued use of a legacy channel based format as the foundation of Dolby ATMOS. SRS labs have recently announced what they term Multi Dimensional Audio (MDA) which takes an object-based approach [6], describing audio objects using metadata. The Soundscape Render [28] is an open source framework for the real-time rendering of audio spatially, using rendering methods such as Wave Field Synthesis, Vector Base Amplitude Panning, Ambisonic panning and Binaural Synthesis. The SSR uses Audio Scene Description Format [20] in order to represent the soundscenes.

2.4 Traditional Capture

Traditional radio production follows one of two approaches. The first is to treat the performance like it is a piece of theatre; the actors perform with each other around a stereo microphone and alongside them there is a foley artist who creates the sound effects live. These sound effects are often captured on a separate microphone and the recording is mixed live to a stereo file. Often recorded music is played in live which results in a recording that does not need much editing after the production.The second approach is more similar to Hollywood film production, initially dialogue is recorded clean, around a stereo microphone. After this sound effects and music is added in post production. The second approach requires a longer edit time but affords the content creators more control over the

sound [29]. The second approach is more similar to the production workflow likely to be needed for object audio based radio drama production.

2.5 Object Capture

In order to remain as independent as possible audio objects should be captured as cleanly as possible. Existing object-based production workflows are aimed at the US movie industry and are designed around Hollywood workflows. These typically include a lot of studio and vocal booth ADR (automated dialogue replacement) which results in clean audio objects. However, television and radio broadcast production budgets tend not to cover the additional expense of the ADR processes. There is limited published research into methods for capturing audio objects in the context of broadcasting workflows and more is required if object-based production is to be commercially successful in the broadcasting industry. There is a European project called S3A [30] which has been set up to further understanding of object based workflows in a broadcasting content, but it is yet to publishing findings relating to this work. One aim of this thesis is to increase knowledge of the impact of using audio objects for content creation of which clean object capture is an integral part.

2.6 Object Derivation

There has been research into deriving objects (clean signals) from both soundfield recordings and other traditional formats like mono and stereo [31]. While the need to separate signals comprised of a number of captured sounds into discrete objects is clear, an in-depth review of existing approaches to this is beyond the scope of this thesis. However, the ideas and applications explored by this thesis could be supported by source separation research, so references are included here for completeness. The most salient factor here is the classification or segmentation of sounds within a signal in order to extract them, rather than the functions and methods used to segment sounds from a mixed signal using independent component analysis. There have been studies [32] which attempt to segment audio content from within a stream attaching a semantic layer to them, for example basic types of sound such as speech, music, environmental sound, speech with music background, environmental sound with music background and silence. However, the majority of these approaches tend to be routed in the physical sound source model, rather then taking any perceptual route.

2.7 Reverberation

There is a notable lack of literature when it comes to reverberation for object-based audio. Industry tools are also very limited. Commercially available object-based systems achieve reverberation using a pragmatic approach. A common approach is to treat clean sounds with commercially available multichannel reverberation algorithms and treat those signals as a series of objects spatially distributed around the listener. Some of these algorithms treat early reflections and diffuse tails separately. Traditional approaches to reverberation, such as convolution, the use of tapped delay lines or filters tend to be implemented with specific speaker locations in mind, and therefore can suffer the same limitations of channel based production formats being tied to the reproduction system as described in section 2.1.1.

Physical modelling of an acoustic space would allow a more physically accurate reproduction of reverberation in 3D space compared with the signal processing methods described above. This can be performed using techniques such as the image/source method or ray tracing, but this is computationally expensive potentially requiring millions of operations every second [33]. In addition, for accurate modelling, acoustic characteristics of the space such as diffusion and diffraction should also be modelled. This can be achieved using techniques like vector based scattering [34], but this all adds to the computational expense. However, the need to create diffuse reverberation that is physically accurate, rather than perceptually pleasant for broadcasting applications is questionable. For example, BBC engineers mixing the Proms use artificial reverberation to enhance the audio mix and are less concerned with recreating an accurate sound of the Royal Albert Hall than they are with making a well balanced and pleasant sound. Some computer games use reasonably advanced reverberation models, for example, at a recent visit to Codemasters they demonstrated the production workflow for "Formula 1 2014." The production team used a reverberation model that included directional early reflections (from the audience seating area). The cars sounds were constructed from independent audio objects to allow the sound of the car to reflect its condition, for example the sound of a damaged exhaust would use a different audio object compared to that of a healthy exhaust.

IRCAM has developed the Spatialisateur project [35], its goal is to create a virtual acoustics processor which allows content creators to control the diffusion of sounds in a real or virtual space. This system is designed to overcome the inflexibility of channel-based systems, working with 2D and 3D systems.

2.7.1 Computer Games Production

Computer games creators treat certain audio as objects to allow interaction [36] with soundscapes. These computer games tend to take an empirical approach to object-based audio. A scientific approach to audio objects is not critical for most computer gamers. Computer game audio workflows are complicated and involve a combination of offline audio production and asset preparation and delivery to a computer programmer who is developing the game. Modern game development allows for interactive mixing, that is mixing audio while the game is being played, using specialist software tools [37]. In the past few years there have been software suites created that are designed to deal with the workflow complication faced, Wwise for example [38] aims to provide a solution to the whole audio pipeline.

2.7.2 HTML5 Audio

In the past few years the web audio API [39] has been developed through the w3c. This standardises audio playback and processing by enabling the browser to use javascript to buffer, schedule, trigger, and process online audio in the client browser. This browser based technology allows flexibility and the universal nature of supported web browsers allows the development of experimental audio experiences that can be distributed to a large audience for testing.

2.8 Object Audio Use in the Broadcast Industry

Despite the fact that audio objects are already widely used in the gaming industry there seems to be a lack of understanding of object-based audio in the broadcasting world. Certain broadcasters have taken small steps towards object-based audio, for example Absolute Radio have began to treat radio advertisements as separate objects [40], so people listening online hear their own targeted advertisements between songs while receiving the same radio programme as everyone else listening. Alongside this Capital Radio have recently introduced the ability for listeners to skip music tracks [41]. The BBC conducted the NetMix experiment [42] with Fraunhoffer which streamed a live tenis match with commentary and pitch/crowd sounds as two separate objects allowing the audience to create their own mix. There were a small number of participants in this study due the fact a bespoke piece of software needed to be downloaded and installed, but results suggested a bifurcation in audience preference for court vs commentary mix [43]. This experiment used an MPEG4 audio stream to transport the audio objects.

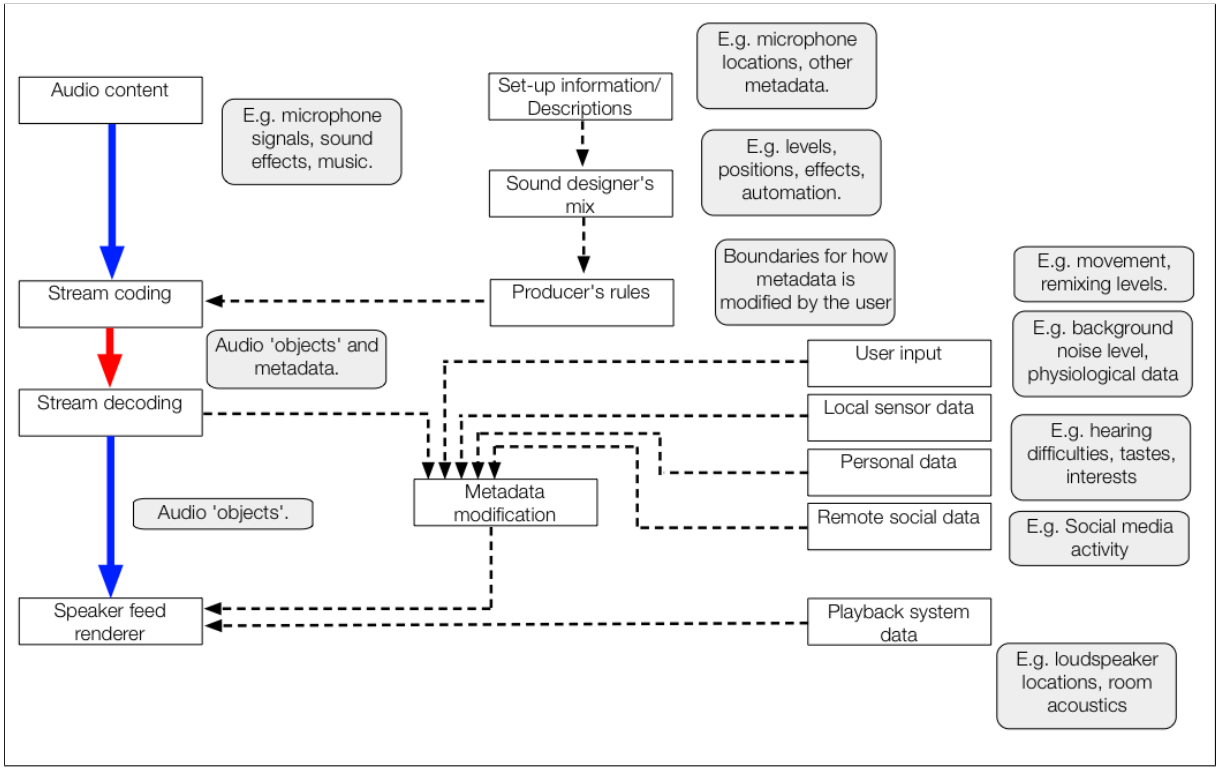


Figure 2.1: A block diagram for object-based audio distribution

2.9 Broadcasting Workflow Applications

There are standardisation discussions continuing at the EBU and ITU for common exchange formats to allow the sharing of content using flexible formats that support traditional channel-based content as well as object-based content. This work will contain design and analysis of some of the first experiments of broadcasting object-based audio. Figure 2.1 shows a possible broadcasting workflow for an object-based system.

2.10 Definitions

While terms like "audio object" appear in the literature, a common understanding of the definition of an audio object is not yet formalised. Soundscape literature [44] has considered definitions of an auditory object and broadly agree the term refers to the result of the

human's auditory system to extract and identify a perceptually independent unit from a complex acoustic environment based on the spectral, temporal and spatial characteristics of the sound. An audio object is not the same as an auditory object, auditory objects only exist with the human auditory system, however audio objects music exist in reality in order for constructed sound scenes to be created, distributed and re-created for the experience of an audience. A recent workshop run at the Audio Engineering Society Convention 133 discussed the practicalities of a move towards an object-based production and distribution system. Presentations from Dolby, DTS and France Telecom all acknowledged the differences between traditionally, physically/mathematically and perceptually motivated approaches to sound scene representation, each suggesting different approaches in order to provide a practical solution to object-based production workflows. From these differing opinions a technical definition of an audio object can be summarised as an audio signal with accompanying, time dependent metadata. This technical definition is fairly concise, however from an editorial perspective there are many questions unresolved. For a content creator, when producing an experience it is not clear which sounds should be independent audio objects. This problem of definition can be linked to the audience perception and the intended audience experience, and is discussed further in section 2.10.2 and 2.10.3.

2.10.1 Physical vs Perceptual

Soundscape and Neuroscience literature considers auditory objects as fully perceptual phenomena. The practical implementations of object based production and representation systems described in this chapter consider audio objects as having purely physical or acoustical characteristics. Whether audio objects are directly related to acoustical or perceptual objects in the audio scene is also unclear. A common assumption, and conceptually simple starting point, is to consider audio objects to be acoustic sources. This approach, routed in the physical, would decompose a sound scene into the acoustic sources that exist in that space. This is uncomplicated when dealing with a limited number of point sources in anechoic conditions. However, in reality sources rarely exist in anechoic conditions, nor are they truly point sources. To proceed compromises need to be made, these compromises are often justified as perceptually acceptable by engineers. Table 2.1 lists some physical parameters and gives a perceptual equivalent.

Figure 2.2 is a mind map of some potential audio object parameters, or ways in which audio objects might be conceived. In addition to the physical (objective) and perceptual (subjective) parameters some additional parameters identified in soundscape research [44], [48] have been included. There is also an "other" node. This is to take account of audio object parameters that fit into neither physical or perceptual categories, but are param-

Physical	Perceptual
Level	Loudness
Position (x, y, z)	Direction and distance
Source type (point, line, complex)	Size and shape

Table 2.1: Physical and perceptual descriptions of sound objects

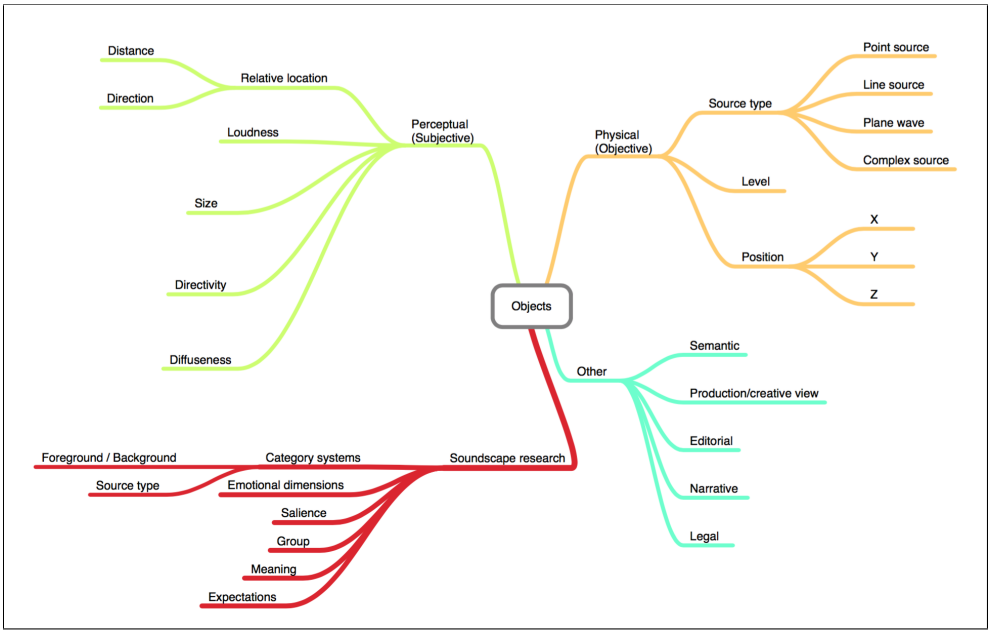


Figure 2.2: Different ways of considering objects

eters that are, or maybe will be in the future, important to the creative production process.

2.10.2 Intended Application

For an audio production, the decision about which sounds should be objects could be linked to the intended audience experience, in addition to level of independence required by the content (an interactive programme would probably require more audio objects than a linear programme). The decision could also take into account the consumption device, consumption context, the intended listener and condition of the listener. There is no agreed definition of an audio object within the audio research community.

2.10.3 Perception of Sound

The field of auditory attention is well researched [45] and contains many ideas that will be relevant to this work. Auditory attention considers the psychoacoustic interpretation of soundscapes in the physical, cognitive and affective domains. Such research will be important when considering audio objects as either 'foreground' or 'background' [49] sounds and when trying to assess the audience benefit of taking an object-based approach over a traditional channel-based production. Visual attention is considered "top-down" or "bottom-up", driven by the viewer or by some factor out of the viewers control, respectively[45]. The application of this theory to designed sound scenes is not trivial. Scenes are designed to direct the listener's attention towards what the creator considers to be important to the experience. In this context 'foreground' sounds would be those considered important, using a salience based approach rather than basing categorising on proximity or loudness.

There have been a number of important papers on auditory objects which consider auditory objects as subjective, having been identified by the listener from acoustic events [46]. However, these concepts have not been applied to designed and constructed soundscapes. There have be studies which specifically consider the reproduction of soundscapes [47], these are based on realistic reproduction of soundscapes, not the production of designed and constructed scenes of the kind created for radio.

The human auditory system has the capability to focus on particular sounds of interest in a sound scene (and therefore block out other audible sound) [50]. This phenomenon is know as The Cocktail Party Effect [51]. The ability to differentiate between different speech signals coming from the same location (for example one loudspeaker) still exists, although it is more challenging for the listener compared with spatial distributed signals [52]. Therefore it could be hypothesised that the correct background/foreground balance is more critical for audio systems with fewer loudspeakers. Literature states that increasing the spatial distribution of speech and noise sources only slightly improved the intelligibility of speech against noise [53]. However, the literature concerning speech and noise primarily considers intelligibility measures rather than quality of experience or listener preference.

To understand the perception of an audio object it is helpful look at the literature relating to visual perception. Gestalt Psychology [54] concerns the mental organisation of perceived stimuli, and is most closely linked to visual perception. Table 2.2 shows the auditory analogue of the visual interpretation.

With the intended application in mind it should be possible to use the theory shown in table 2.2 to help determine the most appropriate way to translate physical sounds into

Term	Visual	Auditory
Proximity	Visual stimuli that appear close together in space are grouped	Sounds that arrive from a similar location are grouped
Similarity	Similar shaped stimuli are grouped	Similar timbre and pitched events are grouped
Continuity	Visual stimuli that follow a regular spatial pattern	Sound which follow a regular pitch pattern are grouped
Closure	Boarders are filled in to make shapes where plausible	Interrupted sounds are perceived as continuous where plausible
Simplicity	Common shapes are perceived as objects	Sounds with simple harmonic relationships are grouped
Common fate	Visual stimuli that move together are grouped together	Auditory stimuli with similar rhythmic patterns are grouped

Table 2.2: Gestalt object perception

perceptual audio objects.

2.11 Sound Taxonomies

2.11.1 Computer Game Audio

van Tol proposes a framework for interpreting computer games audio [55]. Figure 2.3 presents a two dimensional framework for describing audio within games, grouping sound in one of four categories; Zone, Effect, Affect and Interface. Interface describes non-diegetic sounds that are linked to the player's actions. Effect is a category for describing sounds from the game world that result from actions and events in the world of the computer game. Both these categories could be considered foreground categories. Zone describes diegetic sounds associated with the place in which the action is occurring. Affect is a category for sounds linked to the place of action that are non-diegetic; music for example.

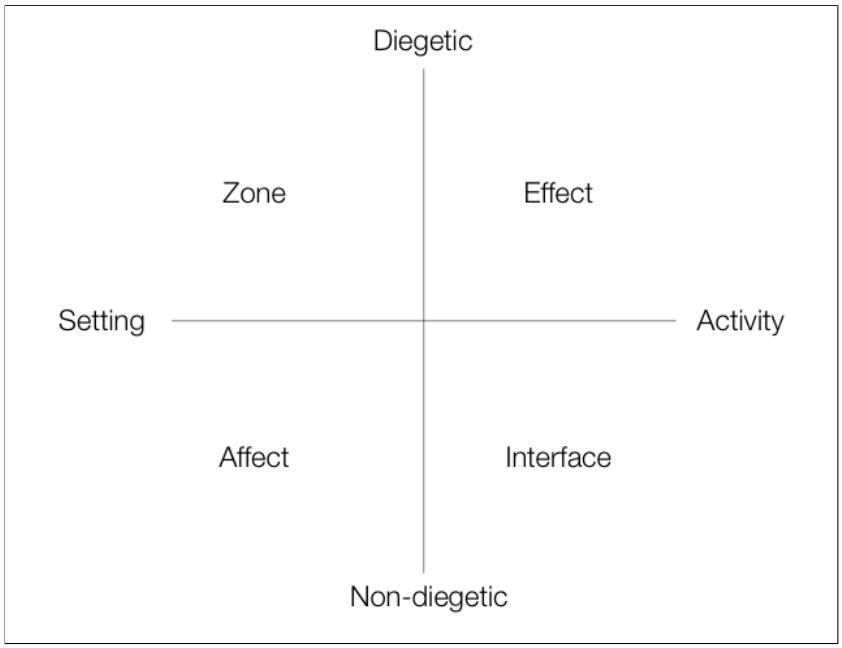


Figure 2.3: Framework for describing games audio [55]

2.11.2 Soundscapes

In recent years there has been research conducted into human perception of soundscapes. There have been a number of models proposed for people's perception of soundscapes. Common sound categories that have been found useful in the literature include 'sound' and 'noise', and similar to that 'foreground' and 'background' [56]. Schafer [59] outlines a classification of sound types below.

- Natural sounds (birds, wind)
- Human sounds (laughing, talking)
- Sounds and society (music,)
- Mechanical sounds (machines, cars)
- Quiet and silence (room tone)
- Sounds as indicators (clock, alarm)

However, this taxonomy is problematic because it mixes source with semantic characteristics, a 'sound as an indicator' could also be 'human sound', for example someone shouting a warning. Other common categories include 'event sequences', where individual sounds

can heard, and 'amorphous sequences' where individual sounds cannot be distinguished [57]. Another common categorisation of sounds within soundscapes was 'natural', 'human' and 'mechanical' [58].

Much of the soundscape research is aimed at understanding people's perception of real soundscapes and involve tasks such as asking listeners to rate how 'annoying' or 'pleasant' a sound is. The aim of much of the research is to build understanding of the impact of the built environment on the sonic experience, in the knowledge that the nature of a noise can be as significant as the level of that noise to a listener's annoyance.

Two key differentiating factors between the soundscape and games research are the concepts of diegesis and interactivity. The first highlights the difference between naturally occurring and constructed soundscapes. The second on the difference between interaction and non-interaction with a scene. Diegesis is a concept that does not apply to natural soundscapes, as all the sounds within a natural soundscape are by their nature diegetic. Constructed sounds scenes often augment an experience by adding music or narration which exist outside of the world of the experience. The computer games research also includes the assumption of interactivity, which is never the case for soundscape research and not often the case for radio content. The assumption of the need for interaction within the model described in figure 2.3 limits its application to non-interactive audio content.

2.11.3 Foreground and Background

The terms foreground and background are primarily associated with visual perception, referring to the proximity of visual stimulus, foreground being physically or perceptually closer to the subject than background. These terms are now commonly used as classifications by sound designers and audiences. In 2011 there was a major study into the cause of intelligibility problems on television [60]. It found intelligibility problems caused by background noise and background music accounted for a quarter of the complaints at 13% & 11% respectively. The other issues identified were problems with the speech itself, for example foreign accents, mumbled speech or speaking too quickly. Related to this there has been some research into the effect of background sound on the audibility of speech for hearing impaired listeners [61]. This work demonstrated the transmission of foreground and background elements of a sports event as separate audio streams which were dynamically rendered at the receiving end in real-time. Foreground and background categories rarely feature in music research with the exception of musical source separation [62] where the research has considered foreground as the singer and the background everything else within the music.

Rumsey [63] proposed a framework which grouped sound into three categories: 'individual sources', 'ensemble sources' and 'attributes relating to the environment'. Ensemble sources being a group of individual sources which are perceived by the listener as a single source, due partly to their spatial proximity to each other. Ensemble source width is identified as a key parameter and its existence disallows a single ensemble source to come from two separate locations. This is probably a fair assumption in non-constructed soundscapes, but sound designers often create spatially separate 'ensemble sources', for example double tracking [64] guitars of vocals and panning them hard left and right. Rumsey's categorisation of sound aligns with work of the QESTRAL project [65] where auditory streams that could be perceived as objects by the listener were considered foreground, and background streams were sounds that could not be identified as a single object. The QESTRAL foreground and background streams were created using noise signals from a single location for foreground or decorrelated noise from multiple simultaneous locations for background. These object categorisations are founded on spatial attributes, focusing on the importance of proximity rather than other Gestalt parameters identified in table tab:gestalt.

2.12 Listening Modes

Schaeffer [66] theorised there were four listening modes. This are shown in table 2.3. The écouter listening sense captures the attention of the listener for example attention being drawn to a noise connoting danger. Entendre is selective listening, the subject can hear a sound's characteristics without directing enough of their attention to infer any meaning. Ouïr refers to peripheral listening, sensing the world around without drawing a listener's direct attention. During comprendre listening the subject attaches an emotional meaning to the sound. While this is relatively early work (1966), it is of interest because Schaefer came to this theories from the perspective of sound creator, and how he thought listeners engaged with content he created.

	Abstract	**Concrete**
Objective	Comprendre	Écouter
Subjective	Entendre	Ouïr

Table 2.3: Four listening modes according to Schaeffer [66]

Schaeffer's listening theories have been built upon in the literature and consolidated by Chion [67] to three listening modes. Causal listening is the act of listening for information

about the source or cause of the sound. Semantic listening is listening for meaning, for example listening to speech. Reduced listening is the focused listening to a sound's physical characteristics without engaging in causal (thinking about the cause) or semantic (thinking about the meaning). Huron [68] has suggested six listening modes motivated by emotion which are shown in table 2.4. The usefulness of these listening modes to sound designers and audiences has not been tested in practice.

Name	**Response**
Reflexive	Sudden Physiological
Denotative	Identification of the source
Connotative	Physical and passively learned meanings
Associative	Arbitrary learned associations
Empathic	Perception of the source's state of mind
Critical	Reflected listening

Table 2.4: Huron's six activating systems

A paper by Turri [69] suggests a hierarchical model shown in table 2.5. It is possible to map these against some of the previously established models, in fact some of the eight listening modes proposed by Turri are the same as previous theories. A focus group was conducted as part of this research the focus groups were not conducted to explore these ideas and the theory was applied to comments made in the focus groups post-hoc.

There is listening model research that categorises background foreground listening modes [70] based on listener attention. This listening model can be linked to classification of background and foreground sounds. This link is made in the literature [71], as is the fact that listener attention, and therefore the foreground vs background classification of a sound is dependent on the listening context. There have been other listening modes suggested by the literature. Gaver [72] suggests two listening modes: everyday listening (listening for information) and musical listening (attention on sonic characteristics). Raimbault [73] suggests a different two categories of listening: holistic (listening to the whole soundscape) and descriptive (focusing on a single sound within a soundscape) which could be analogous to background and foreground categories, respectively. Many of these models can be considered to align with the 'top-down' and 'bottom-up' model of attention from the visual perception research.

Type	Mode	Response
Pre-conscious	Reflexive	Sudden physiological
	Connotative	Non-contextual meaning
Source-orientated	Causal	Cause of the sound
	Empathetic	Emotional associations
Context-orientated	Functional	Purpose of the sound
	Semantic	Contextual meaning
	Critical	Reflective meaning
Quality-orientated	Reduced	Physical qualities

Table 2.5: Turri's listening mode hierarchy

While some of these listening models have proven useful for sound designers and the analysis of sound scenes, they tend not to be founded on experimental activity or directly supported by empirical evidence.

2.13 Assessing Experience

A key part of this thesis will involve assessing new audience experiences to judge whether the new experience that has been enabled by using audio objects is improved compared with the non-object-based traditional/reference experience. There are a number of recognised methods for assessing audience perception that are discussed in the literature. This section gives an overview of them.

2.13.1 Subjective Testing

Traditional audio testing is well defined in recommendations such as MUSHRA ITU-R recommendation BS.1534-1 [74] and Recommendation ITU-R BS.1116-1 [75]. These clearly define approaches to assessing audio systems and processes where there is a clear reference, a hidden reference and a number of anchors. Recent literature has explored other methods of subjective testing which can be used to investigate audio experiences, some examples of this are cited below.

2.13.2 Quality of Experience

Quality of Experience (QoE) is more often becoming used to represent a measurement of the user's perception of an experience. The QoE is not a physical measurement, but a concept which takes a holistic view of multiple elements of the user's perception of an experience [76]. This concept takes a more holistic view compared to approaches described in section 2.13.1, aiming to understand the user experience, taking into account wider contexts [77].

2.13.3 Qualitative Testing Methods

Psychology offers a number of methods for assessing perceptions of multimedia content [78]. Commonly used approaches include focus groups, retrospective surveys, beta testing and usability testing. Many of these approaches have arisen from testing of computer games.

Method	Approach
Focus Groups	Typically 3-12 consumers who discuss experience together
Retrospective Surveys	Consumers are asked to self-report responses to questions
Beta Testing	Beta testers normally volunteer to look for technical problems with an experience
Usability Testing	Consumer behaviour is observed whilst using the content under test

Table 2.6: Common computer games perception testing

2.13.4 Analysing Engagement

The paper "Measuring Narrative Engagement" [79] aims to assess the engagement of an episode of linear television. The paper defines a set of dimensions to allow the measurement of narrative engagement:

- Understanding Narrative
- Attentional Focus

- Narrative Presence
- Emotional Engagement

While these could be useful metrics when designing experiments, it is still unclear whether these narrative dimensions can equate or map to quality of experience.

There have also been number of studies which attempt to measure engagement using physiological means, such as EEG scans [80], perspiration [81] and muscle movements (smiles or frowns) [82]. Many of these approaches are fairly new fields of research and more work is needed to determine how effectively this physiological data can be gathered and interpreted.

2.14 Conclusion

Given the state of the literature there are key areas where knowledge is lacking and where there is scope for research. There is limited literature on measuring the preference, enjoyment or engagement of personalised or interactive experience enabled by object-based production. There is also limited literature on the framework for defining and classifying audio objects in broadcast content. Finally, production tools and workflows for object-based are limited and the impact on the production and creative workflow of using audio objects is not clearly covered by existing literature. This thesis aims to fill some of these gaps and provide a framework for defining an audio object in the content of a broadcasting production workflow.

3

Study 1: Live Football

3.1 Introduction

Traditional broadcasting is a compromise. There is a tacit assumption that the sound mix that is broadcast meets with the approval of the majority of listeners and that audience preference of the mix is evenly distributed around the broadcast mix. A potential outcome of object-based audio is that not all of the audience need to experience the same audio mix. There have been previous investigations into object-based audio experiences, for example Netmix [42], which allowed the audience to control the background/foreground mix of a tennis match. To allow audiences this level of control broadcasters need to produce a set of common assets (text, pictures, sound and video) together with metadata to determine how these assets are rendered in response to the type of device asking to present them. The Netmix experiment used proprietary technology to deliver the commentary and background audio signals separately in order to give the listener control of the balance between the two [43]. The number of responses to the NetMix trial was fairly low. These numbers were limited, due in part to the requirement that audiences downloaded and installed a bespoke player for the experience.

This study extends the functionality of NetMix to allow listeners to control the balance of the crowd/commentary mix and choose the relative loudness of the home and away crowds. The experiment for this study was conducted on Monday 27th May 2013 during the English Football League Championship Play-off Final between Crystal Palace and Watford from Wembley Stadium, London. The experiment enabled the interaction by broadcasting three live streams over IP, with one pair of stereo microphones pointing at the Crystal Palace fans, one pair pointing at the Watford fans and a mono feed from the commentary box. The user interface employed the HTML5 web audio Javascript API [39] to control

the streams, enabling the listener to alter the relative balance between all three streams. The aim of this study was to determine the demand for such an object-based experience, to understand how listeners used the interface and how they responded to events during the broadcast.

3.2 Implementation

The live radio broadcast chain can be divided crudely into three parts; production, distribution and consumption. The following sections describe the methodologies followed for each part of the broadcasting workflow.

3.2.1 Production

Four microphones were used to pick up the crowd noise; each acted as a left or right channel for one end of the stadium. Practicalities of the football stadium layout and infrastructure at the location determined the physical placement of the microphones. A compromise between being too close (with a risk of bad language being broadcast) and too distant (a lack of presence leading to difficulty in differentiating between stadium ends) was achieved by positioning the microphones near the corner flags, just in front of the advertising hoardings, pointing at the crowd. These requirements had a direct influence on the audio objects used to create the scene which is discussed further in chapter 6. A radio transmitter attached to each of the four Sennheiser shotgun 416 broadcast microphones transmitted the microphone signals to radio receivers at the Radio 5 live commentary box as shown in Figure 3.1.

A computer and sound card converted the three audio signals into three 128 kbps AAC encoded streams which were sent back to the BBC control room over IP. The two streams for the crowd noise were stereo and a third stream for the commentary (which was captured from a lavalier microphone) was mono. AAC was chosen due to its superior sound quality to MP3 and is a compression format that is commonly used by the BBC for streaming audio content of this type.

3.2.2 Distribution

The audio streams from the stadium were received at a server room at the BBC, transcoded to MP3 and Ogg Vorbis formats. These compression methods were chosen in order to cater

Figure 3.1: Left, crowd noise microphone. Right, radio receivers.[1]

for all major web-browsers on all major operating systems. The streams were distributed by an Icecast server and the number of concurrent streams was initially limited to 3000 (or 1000 listeners) in order to manage bandwidth use and therefore cost. This was extended to 12000 (3 x 4000 listeners) during the course of the trial in order to cope with unexpected demand on the service. Figure 3.2 shows a high level systems diagram for the experiment.

3.2.3 Consumption

The user interface was designed to be as simple and intuitive as possible. There were design constraints imposed by Radio 5 Live which coloured the usage data. The result is shown in Figure 3.3. Listeners were able to mix between one end of the stadium and the other, with the left side 100% Crystal Palace and the right 100% Watford. The audience could control the mix by dragging a microphone icon over a plan view of the stadium. Only the position of the mouse on the left-right/x axis was used to determine the desired balance. Listeners were also able to change the balance between the commentary and the crowd. Audiences were allowed to control the balance using a simple drag bar at the bottom of the interface. At the extremes, either commentary or crowd would be 9dB above the other. Listeners were also able to stop, start and control the overall volume on the interface. All audience interaction with the interface was logged. The position of the microphone was logged every 500 ms. The position of the commentary balance bar was logged when the mouse was released after dragging.

[1] Images originally published in [2], ©BBC 2013.

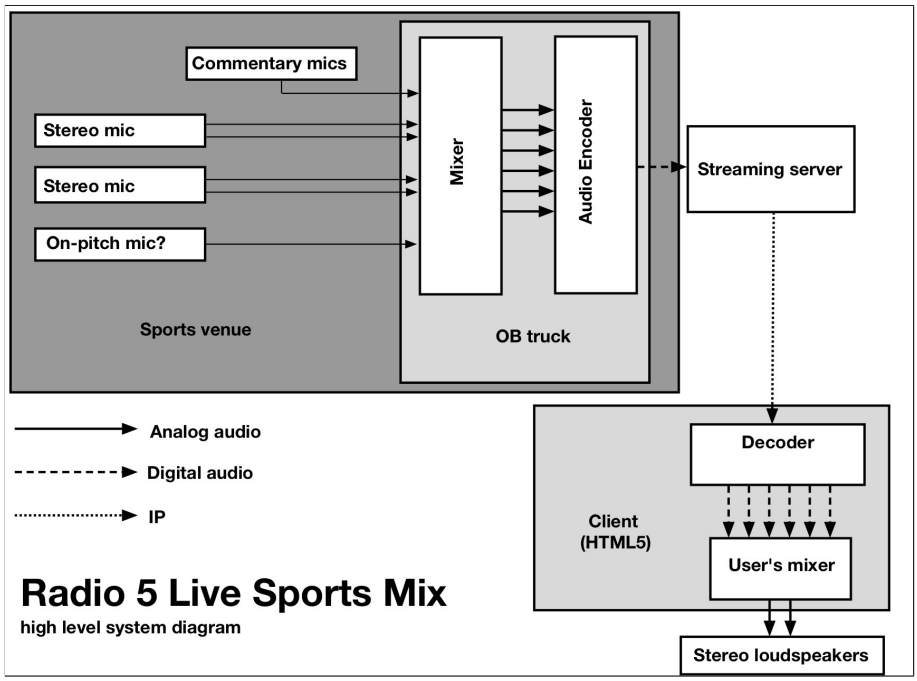


Figure 3.2: A high level system diagram for the Radio 5 Live experiment

3.2.4 Experimental Limitation Critique

The limitation of not being able to completely turn off the commentary, and the user interface design were determined and approved by Radio 5 Live. To maintain a level of quality for their audiences whatever the user setting, 5 Live wanted the commentary always to be audible, just at a lower level. Both these issues are likely to be a source of bias in the data. The user interface has clear visual anchor points in the middle and at either end of both the pitch end and commentary vs crowd control. The anchor point in the centre grounded the broadcast mix, the inference here that this is the 'correct' mix and this is likely to result in many people sitting the control here, rather than selecting a position based purely on the sound. Similarly, the user interface had hard limits at either end of each control which will have a similar bias effect. The fact that it was not possible to turn off either crowd or commentary is also going to introduce bias to the results, probably increasing the number of people who move the control to the extremities. The foreground vs background study (chapter 4) considers and addresses these experimental design limitations.

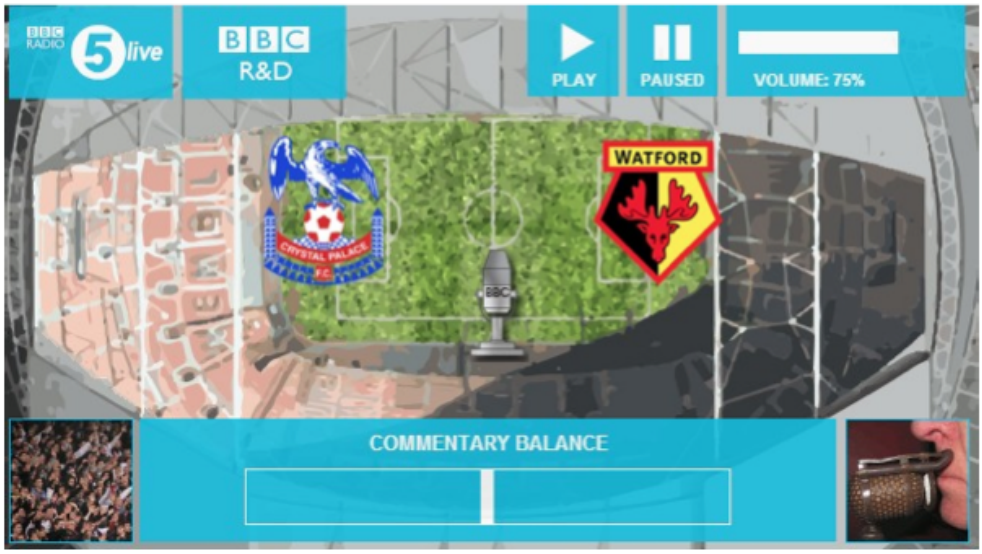


Figure 3.3: The football experiment user interface.[2]

3.3 Interaction Analysis

The broadcast had 5286 listeners who attempted to engage with the experience. A limitation in some browsers resulted in some people not being to receive concurrent streams, and some users' browsers were limited due to incompatibility. There were 2692 successful listeners, logs of their interactions were recorded during the match. The majority of listeners (65%) chose a microphone position within the first 30 seconds of joining the broadcast. Listener's interactions with the crowd vs commentary balance exhibited a similar behaviour. This is illustrated in figure 3.4. Time of arrival of each listener is adjusted relative to the listener's first interaction with the commentary. The 1075 listeners to the first half are represented in the figure, number of listeners are plotted against normalised time.

There are three clear peaks shown in figures 3.4 and figure 3.5. The highest peak was an even mix of commentary and crowd sound. This represents the balance as it would be on the traditional broadcast. A second peak appears where audiences set the slider to provide maximum crowd sounds and minimum commentary. The third, more subtle, peak at maximum commentary. A considerable number of listeners chose a level other than the three peaks. Roughly 22% rested in the centre, 6.5% set the control to maximum commentary and 5.5 % set the control to maximum crowd noise. Meaning that around 66% of people set the control outside of the three peaks.

[2]User Interface designed by Jasmine Cox. Originally published in [2]. ©BBC 2013.

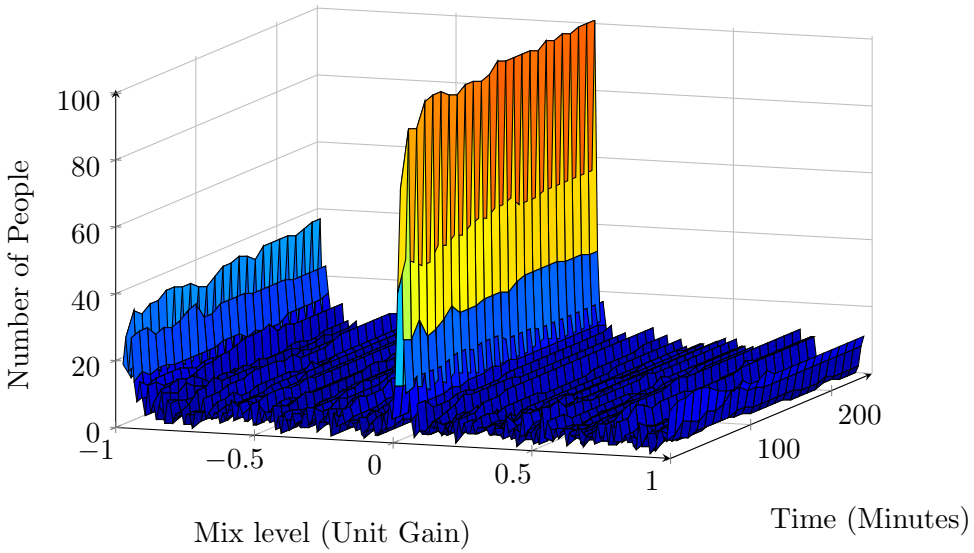


Figure 3.4: A histogram showing how listeners adjusted balance over time.[3]

The same behaviour was evident for listeners when selecting microphone position, however there was an equal balance of listener's choice of stadium ends. Figure 3.5 shows the final crowd vs commentary balance. Table 3.1 shows results from some statistical analysis of the data. The distribution is not Gaussian. The Interquartile range of the data set is 0.1422, with a mode of 0. This leads to the conclusion that half the audience were satisfied by the broadcast mix which in turn means half the audience chose to change the broadcast mix. This could be used as evidence to demonstrate the value of object-based audio for this type of experience.

Data	**Result**
-0.0407	Mean
0.3996	Standard Deviation
0	Median
0.1422	Interquartile Range

Table 3.1: Crowd/commentary mix with time (where 0 is equal commentary and crowd, +/-1 is maximum commentary/crowd)

Of particular interest was whether or not listeners reacted to events during the match. Figure 3.6 shows the total activity against time. A few key incidents in the match and on air/social media announcements for the trial are marked. At Wembley, the broadcast

[3]Based on data published in [2]. ©BBC 2013.

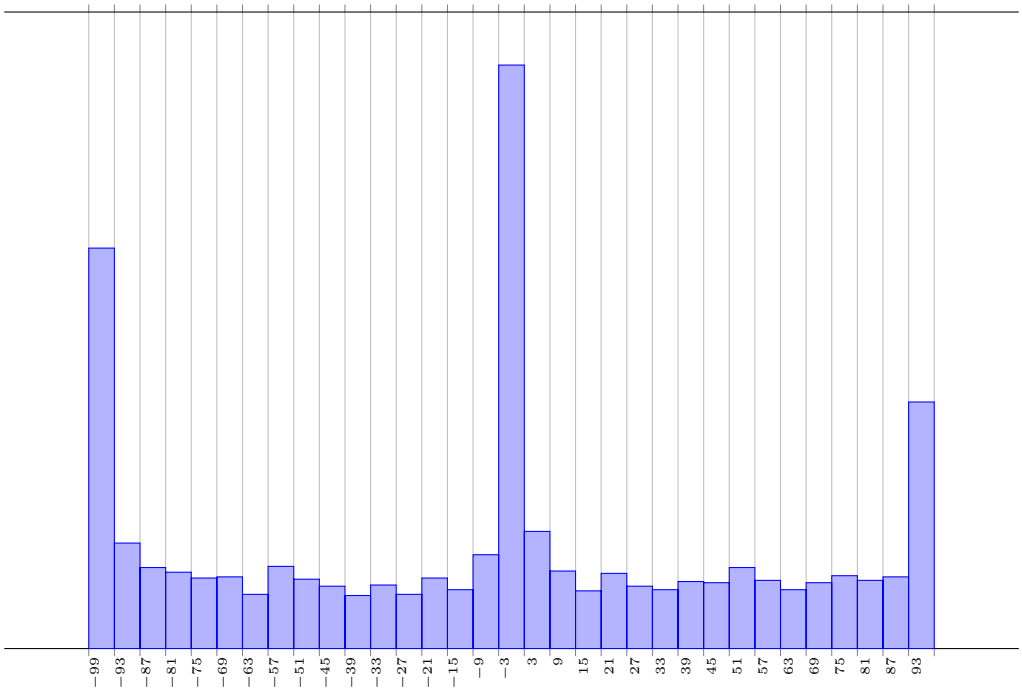


Figure 3.5: Histogram for final choices of the crowd/commentary mix (where 0 is equal commentary and crowd, +/-100 is maximum commentary/crowd)[4]

went smoothly, though a small adjustment in the relative crowd/commentary level needed to be made a few minutes into the broadcast. Unfortunately, the Icecast server needed to be restarted 20 minutes into the broadcast because new connections were being refused. This configuration error caused a spike in activity during the match (see Figure 3.6). The Ogg Vorbis commentary stream failed to restart at the beginning of the second half which also caused a spike in activity.

Figure 3.7 shows how users interacted with the controls (commentary and stadium end controls) over time. These figures are normalised to show the level of interaction relative to the time the user joined the broadcast. The graph suggests that there is an initial burst of activity, but after 10 minutes of interaction users stopped using the controls, preferring to set the controls and leave them. This suggests that events on the field did not have a major effect on the way that listeners responded to setting the microphone position and commentary balance. These peaks can be accounted for by the events identified in figure 3.6 which lead to new listeners joining the service rather than current listeners reacting to events on the pitch. Figure 3.8 shows that people joining the broadcast was the major fac-

[4]Based on data published in [2]. ©BBC 2013.

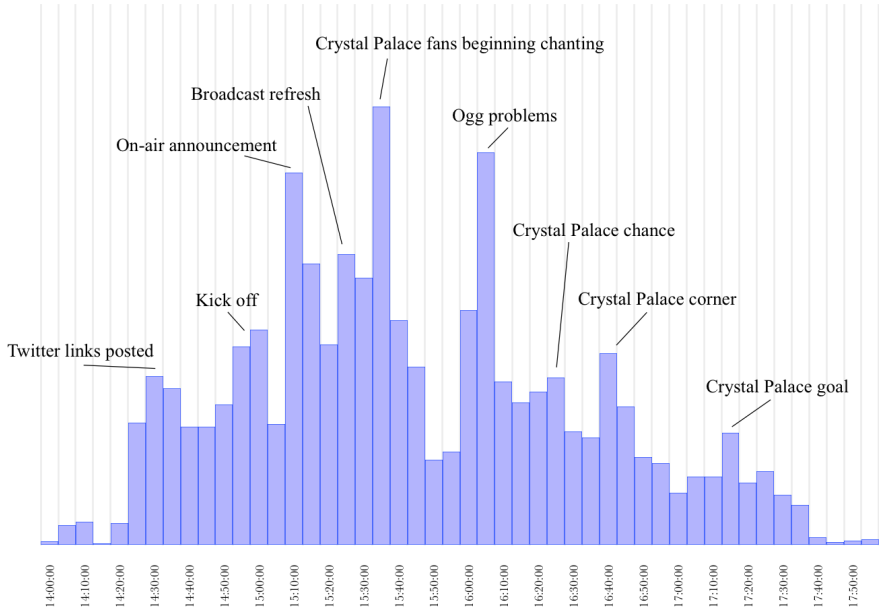


Figure 3.6: Graph showing total activity contextualised with broadcast and stadium events.[5]

tor causing spikes of activity which can be account for by the new listener's burst of activity.

The histogram shown in figure 3.8 shows the combined activity, with the initial burst of activity (5 minutes) exhibited by users when they joined the broadcast removed from the data. The absence of certain peaks correlating to the events identified in figure 3.8 along with remaining peaks being lower than those shown in figure 3.6 suggests that listeners did not alter their preferred crowd vs commentary balance in response to these events to the extent suggested by figure 3.6. After ten minutes figure 3.9 shows even lower peaks. The biggest exception being the ogg stream failure the time of which correlates with a significant peak in activity. This, taken in conjunction with figure 3.7 suggests that many listeners selected a position to set the sound balance during the first few minutes of listening and left the sound balance in that position for the duration of the match, unless there was a technical error.

[5]Based on data published in [2]. ©BBC 2013.

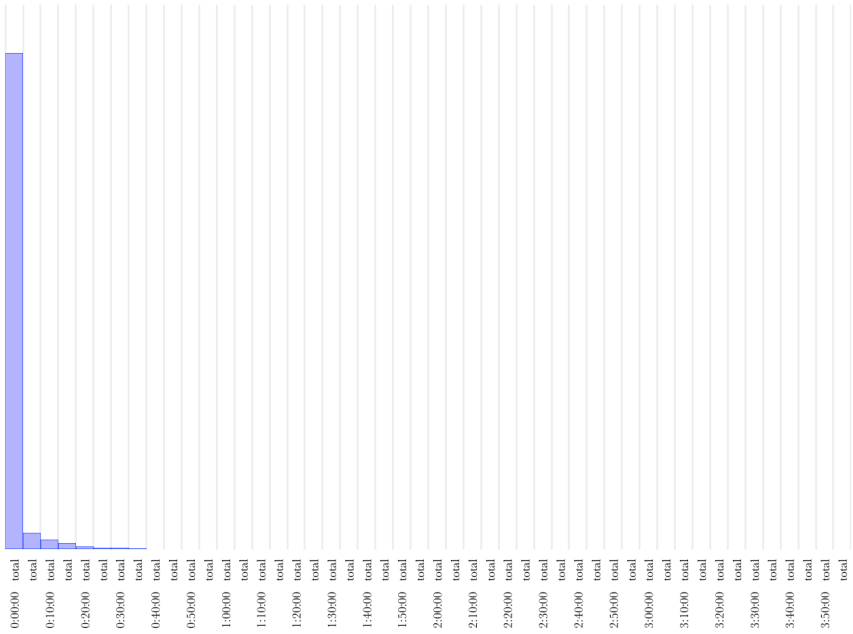


Figure 3.7: A graph showing user interactivity with time, normalised to their joining the broadcast.[6]

3.4 Listener Feedback

Listeners were invited to answer a short questionnaire after the broadcast. This was designed to assess the listener's enjoyment of the experience and was administered online. The questionnaire is included in appendix A.1. There were 701 responses. The key results from it are shown in figures 3.10, 3.11 and 3.12.

Approximately three in four participants felt being able to control the crowd vs commentary mix resulted in an improved experience compared to normal broadcasts. Just under three quarters of participants preferred being able to control the stadium end mix and the experience over traditional radio coverage. 60% of listeners considered the interface easy to use with a roughly even split between those listeners who wanted more control and those who did not want more. This suggests that listeners chose not to adjust their preferred balance during a broadcast rather than this being the result of a usability problem. However, while this design made the experience more understandable for the audience, it is

[6]Based on data published in [2]. ©BBC 2013.

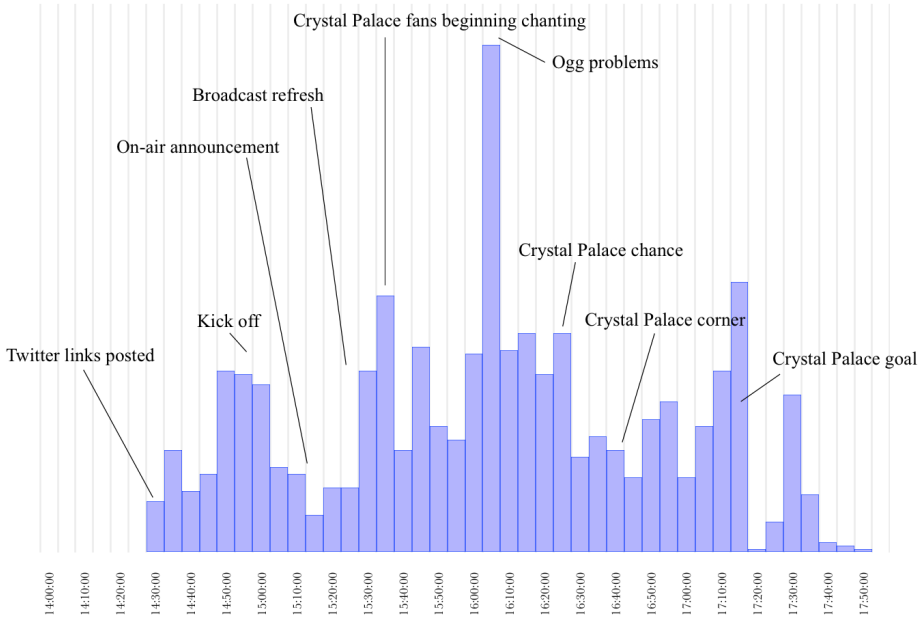


Figure 3.8: Commentary changes after first 5 minutes of user activity removed.[7]

[7]Based on data published in [2]. ©BBC 2013.

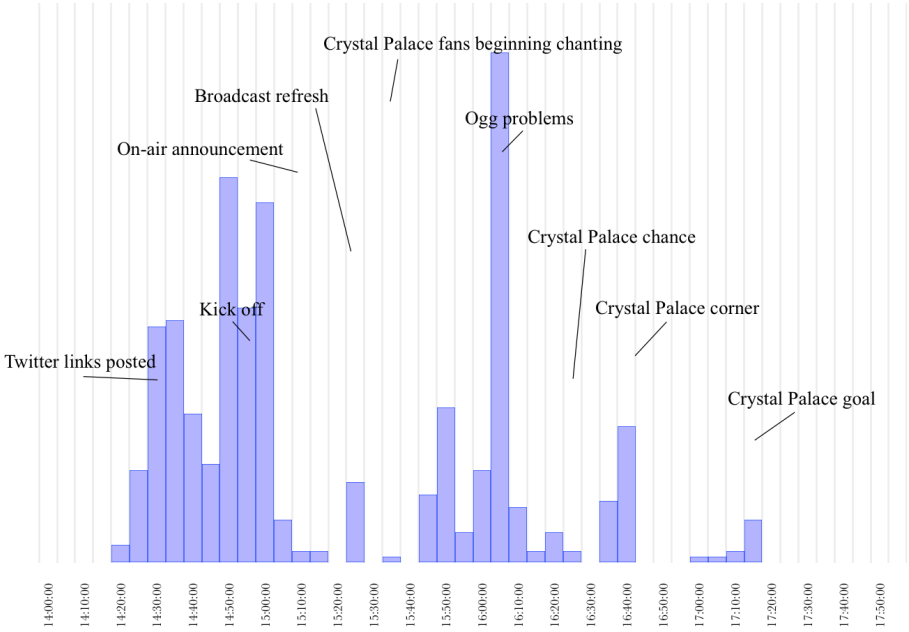


Figure 3.9: Commentary changes after first 10 minutes of user activity removed.[8]

highly likely that the design of the user interface will have impacted on people's choice of mix.

[8]Based on data published in [2]. ©BBC 2013.

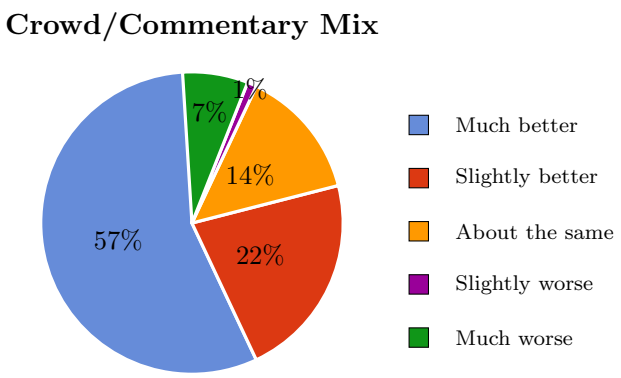


Figure 3.10: Impact of being able to control the commentary/crowd mix.[9]

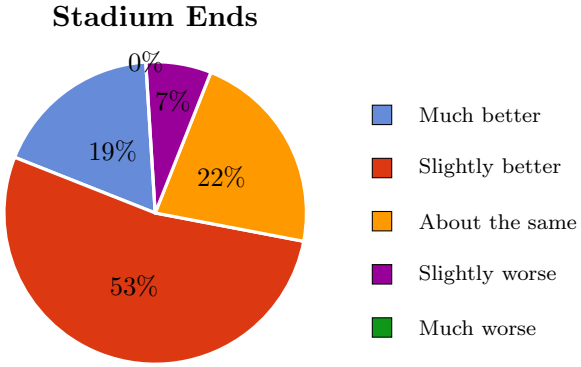


Figure 3.11: Impact of being able to choose ends.[10]

3.4.1 Social Listening

Listeners also provided qualitative feedback and comments on social media that were monitored using an account that was set up for the experiment. There were 55 mentions in total, examples of these messages are shown in figures 3.13 to 3.16. BBC Radio 5 live typically broadcasts online using a 56 kbps G.722 mono codec, therefore this broadcast provided an improvement in sound quality which was reflected in some of the responses to the experiment on social media (see figure 3.16). The responses from social media generally fell into two categories, those hard of hearing who appreciated being able to isolate the commentary, for example 'Brilliant idea. Now I can reduce crowd volume and actually hear

[9] Based on data published in [2]. ©BBC 2013.
[10] Based on data published in [2]. ©BBC 2013.

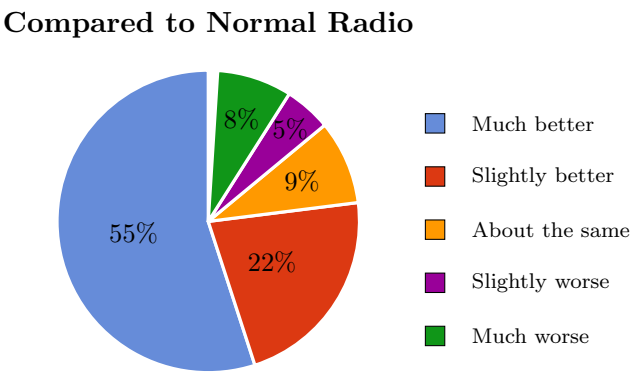


Figure 3.12: Compared to traditional radio.[11]

the sometimes mumbled comments of the expert summarisers' (see figure 3.13) and those who liked being able to turn the commentary down and listen to the crowd, for example 'love it, great to be able to hear the atmosphere more over the commentary' (see figure 3.14). This expression of preference for a particular experience rather than enjoying the capability to interact with the content is reflected by the quantitate user behaviour data.

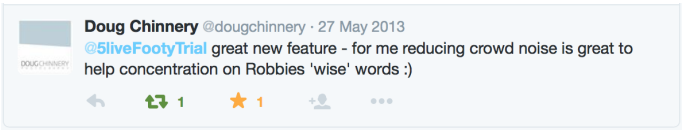


Figure 3.13: Feedback supporting commentary boost on social media

Figure 3.14: Feedback supporting crowd boost on social media

3.5 Discussion and Summary

The audience feedback suggests that the ability to personalise the audio mix is valued by three quarters of the participants of this experiment. One observation that can be made about the audience's behaviour was they would initially interact with and explore

[11] Based on data published in [2]. ©BBC 2013.

Figure 3.15: General feedback on social media

Figure 3.16: Non-object-based benefits of the football trial

the limits of the controls before settling on a final mix as illustrated in figure 3.7. Audiences did not react to events in the stadium by interacting with the user interface, so it appears the key motivation for listeners to use the controls was to improve the general listening experience rather than actively interact with the experience. This is backed up by the comments observed on social media. It is possible that audiences valued having the ability to alter the mix, even if they chose not to change in from a typical broadcast balance.

No firm conclusions can be made about what the ideal mix or mixes should be, nor the degree of choice a listener would like in choosing what the mix should be or if they want to choose a mix at all. The limitations placed on the user interface design by BBC Radio 5 Live appear to influence the audience behaviour, encouraging the mix to be set in the middle or at either end of the slider. It is impossible to accurately gauge the bias caused by the user interface requirement stipulated by 5 Live, but the shape of the histogram in figure 3.5 suggests that there is likely to have been a major bias. While the shape of the distribution is highly likely to have been influenced by the user interface design some results are still valuable. For example, roughly twice as many people chose to increase the contribution of the crowd and lower the commentary. This could have been due to the nature of the experience. It was a live football match and the ability to choose which side of the crowd to listen to may have encouraged listeners to turn up the crowd level

relative to the commentary. However, comments from social media back up the quantitive data, suggesting some listeners value the ability to turn up the commentary while others like to turn it down. Both of these types of listener get value from flexibility enabled by the object-based nature of the broadcast. However, figure 3.5 confirms the presence of three peaks: maximum commentary, maximum crowd and equal commentary and crowd which could be used to justify broadcasting three static audio mixes rather than a set of audio objects. However, 66% of the audience did not settle on one of these three peaks. This presents a strong argument against providing three audio mixes to cater for the three preference peaks, but delivering audio objects to allow more flexibility to cater for all the listener preferences.

Another influencing factor was the self selecting nature of the respondents to the experiment, this experiment was publicised using the BBC 5 Live and Research & Development social media accounts which attract a certain (male, middle-aged) demographic. In addition the need for a modern browser could have influenced the results in a similar fashion.

A further experiment to determine people's preference for different foreground and background mixes which addresses some of the experimental limitations occurring in this experiment appears in chapter 4.

4

Study 2: Audio Object Classification

4.1 Introduction

The second study has been designed to help understand the perception of what an audio object is by content creators, and to address some of the design limitations of the football study conducted in chapter 3 in order to understand if and how preferences for different foreground vs background mixes vary with person, genre and playback system. The decision to use foreground vs background as the chosen categories was influenced by the large number of complaints the BBC receives relating to background noises being too loud, and the need to understand audience perception of foreground vs background levels in constructed sound scenes.

4.2 Context

The aim of this chapter is to understand how sounds are grouped by content creators for speech and music based content, and to extend the previous chapter by addressing some of the experimental limitations. Chapter 2 highlights the gaps in literature for analysis of soundscapes that have been designed, and the resulting lack of listening models for constructed sound scenes where there is a clear intention behind the presence of all of the audio objects within a sound scene.

Chapter 2 also discusses the balance between listener attention when directed by the listener as apposed to when listener attention is directed by external factors. When a sound scene is constructed by a creative team there is intention behind the inclusion or exclusion of each audio object within that scene. Scenes are designed to tell and story and to do so the listener's attention is directed by the content creators, which could be considered an external factor. With this in mind it is important to highlight the significance of the content designer's intention when it comes to what the audience is experiencing and where their attention should be directed. This is why the main body of work in this section involves the study of content creator's approach to audio design using audio objects and the analysis of the creator's view of the content and audio objects with in it.

4.3 Approach

In order to best understand sound categorisation two types of production (speech based and non-speech based) were analysed. These two categories were chosen as the approach to design and creation for each is quite different, and different production teams are responsible for each category. The creators of each type of content were asked to perform a series of tasks to illustrate their understanding of how the different sounds within a scene should be categorised. Following these exercises a series of clips were created using two audio objects: foreground and background. An interface was then designed to allow subjects to set their preferred foreground vs background mix.

4.3.1 Choice of Content

The drama "Pinocchio" was chosen for a number of reasons. First, it is highly representative of radio drama produced by the BBC. Secondly, it was thought the drama was suited to an immersive with-height audio production, due to the dramatic scenes and locations in the story. While there were domestic scenes there were also fantastical scenes, underwater and etherial characters which would allow the creative exploration on the part of the Producer and Sound Designer. A musical track by "Everything Everything" was chosen for the non-speech example. This choice was mainly down to availability. The track is a fairly typical pop song, with a traditional instrument line up.

4.3.2 Production System Choice

IOSONO was chosen for the production system for "Pinocchio", "Everything Everything" and the other test clips used in the foreground vs background study. This was mainly due to system availability. The number of object-based production systems available are very limited. The IOSONO system allows object-based production, but bases its concept of audio objects as sound sources/channels in a DAW. There was a risk that the IOSONO approach (using sources as objects) could have influenced the Sound Designer's grouping of objects. However, the Producer considered the audio objects in the same way as the Sound Designer and had no contact with the IOSONO system therefore it is unlikely the IOSONO approach to treat sources as objects influenced the Sound Designer's results. IOSONO was also used for the production of all the 9.0 content. With a loudspeaker layout as sparse as 9.0 (compared to WFS systems) the IOSONO production system uses a VBAP rendering algorithm. The added advantage of the IOSONO system is its ability to equalise room responses. The listening room was calibrated using the room equalisation technique. It is unlikely that these choices will have influenced the results, due in part to the fact the production environment and test environment used were in the same place and using the same system.

4.4 Speech Based Content

4.4.1 Production

The radio drama was created by a highly experienced team; a radio programme Producer with over 12 years experience producing and a Sound Designer with over 25 years experience making radio drama.

Recording

The radio drama "Pinocchio" was recorded at Media City in the radio drama studio known as MPAS (Multi-Purpose Audio Studio). This facility features a number of different acoustic spaces, including a dead area shown in figure 4.4.1, a live area and a number of smaller rooms. These are designed to simulate different acoustic environments in which a radio drama might be set. For example, a scene-based outside would be recorded in the dead area, a scene-based indoors would be recorded in one of the smaller rooms.

Figure 4.1: MPAS dead area.

Microphone Techniques

"Pinocchio" was recorded using traditional stereo recording techniques. The main microphone used was a stereo microphone. Additional microphones were also used for reverberation, distance effects and foley sound effects. The recording was made in the same way as it would have been for any stereo production of this type.
Alternative methods of capturing audio exist and are used in television and film production, for example ADR (Automatic Dialogue Replacement) where actors re-perform their lines in studios when location recordings are not of an acceptable quality. These approaches result in much cleaner audio objects, however after some discussion with the producers and actors around the disadvantages, costs and risks associated with recording actors in isolated studios the decision was made to record the actors performances around a stereo microphone as is done with the majority of radio drama. Traditional stereo microphone techniques allow the recording of a natural performance, recording the actor's movements around the microphone and capture some of the acoustic environment compared with recording acoustically isolated actors.

Figure 4.2: The mixing setup for "Pinocchio" showing (left to right) the Production Assistant, a visitor, the Sound Designer, the Producer (at the back) and me.

Mixing Set-Up

The mixing set-up was designed in a way that allowed the Sound Designer to work using a system that had a large collection of plugins with which he was familiar (Pro Tools). The spatial audio rendering system used a Nuendo plugin for its panning. Therefore the Pro Tools computer was connected via MADI to a spatial audio system. This consisted of a MacPro running Nuendo 5, connected via MADI to an IOSONO spatial audio rendering system which was, in turn, connected via MADI to a set of 28 DACs. These DACs were connected to the loudspeakers. A set of 26 loudspeakers played back a 3D surround sound version of the play. Parallel sets of stereo, 5.0 and 9.0 loudspeakers allowed the team to monitor down-mixed versions to ensure quality of the broadcast versions. The intention was for all of the panning and panning automation to be handled by the system running Nuendo while all of the other mixing (levels and effects) be handled by Pro Tools. Figure 4.3 shows the block diagram of the set-up. Figure 4.2 shows a photograph of the mixing taking place.

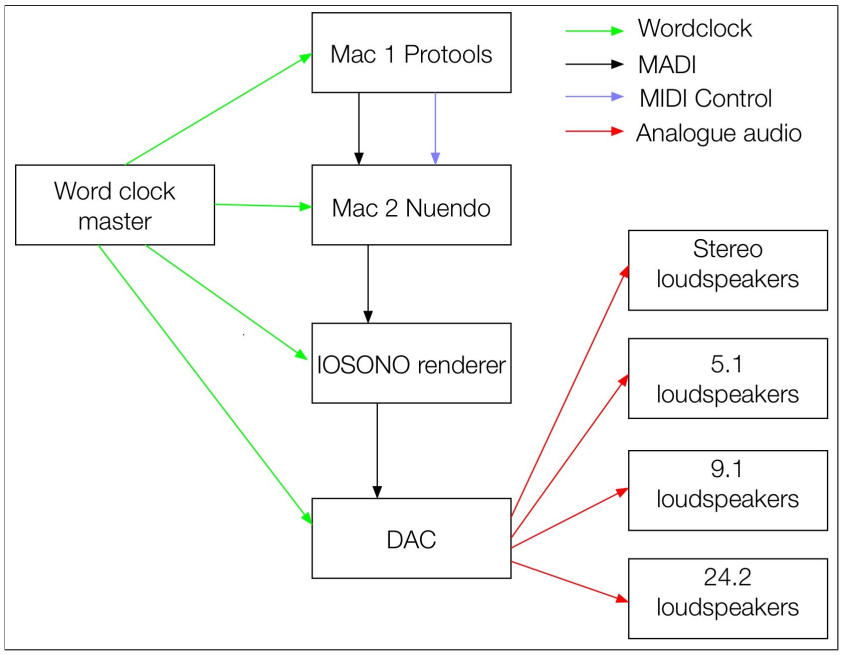


Figure 4.3: Block diagram for the mixing setup for "Pinocchio".

Mixing Process

The mixing process began at BBC audio studios where a rough dialogue mix was created using Pro Tools 9. This involved arrangement of the audio recordings in the correct order by the Sound Designer, based on the script and notes from the Producer. This mix also allowed the Producer to choose the best takes, and to understand the pace of the piece in order to get it to roughly the right length. The required length of "Pinocchio" was 56 minutes and 46 seconds. Once the dialogue mix was finished the editing computers were moved into a listening room containing the surround sound rig and IOSONO wave field synthesises spatial audio renderer. The were four parallel loudspeaker arrays used for the mixing, a stereo pair, a 5.0 set-up, a 9.0 set up and 26 loudspeaker (24.2) set up. The 9.0 array was a standard 5.0 set-up with an additional four high loudspeakers in the corners. The 24.2 array had a square of 16 loudspeakers at regular one metre intervals at ear height, a square of eight high loudspeakers, with two subwoofers at floor height.

The entire play was mixed as audio objects and positional metadata rather then directly to a set number of channels. This is the first time this has ever been attempted for a UK radio broadcast.

While the Sound Designer used all the parallel loudspeaker layouts described in section 4.4.1 to monitor the different renders at the beginning of the production, as his confidence in the system grew more time was spent mixing using the full loudspeaker array and less time was spent monitoring the down mixes.

The story required the creation of a number of contrasting soundscapes. Characters such as the Blue Fairy and the Cricket, and locations such a busy fairground and inside a shark meant "Pinocchio" had a lot of sonic variety. Typically these sound scenes were constructed using a combination of atmospheric sounds (such as waves, wind or babble/chatter), contextual sound effects (such as a groan of the shark) and the effects or treatment applied to the voices (normally reverberation). These effects were applied using Pro Tools plugins with which the Sound Designer was familiar. In addition many of the sound effects were from sound effects archives and existed as channel-based formats such as 2.0 stereo and 5.1 surround. This meant that while there was no panning in Pro Tools, there were groups of channels associated with different sounds, for example a five channel surround sound reverb. These channel groups were treated as separate audio objects and each channel was positioned using the Nuendo panning system and rendered by the IOSONO system to that position in space.

4.4.2 Audio Object Analysis

This is the first time a production workflow using audio objects has been used for linear UK radio drama production. In order to understand the way in which radio drama Producers consider audio objects, this section describes how two subjects, the Sound Designer and Producer of "Pinocchio", reported their understanding of the audio objects within the speech based content.

Method

The decision to use a reductive card sorting method to categorise audio objects was made in order to understand the content creators' classification of audio objects. Card sorting is an established methodology in interaction design for understanding and arranging data for information architecture [86]. Card sorting for this type of experiment offers a number of advantages such as providing a physical interaction and focal point for discussions with the subjects and allowing them to see the audio objects and keep them in their minds while undertaking the task. There are two approaches to card sorting; open sorting, where subjects are asked to create their own categories, and closed sorting where the categories

are predefined. This research used a combination of the two approaches, beginning with open card sorting and gradually working down to two pre-defined categories; foreground and background. This would allow the insight from asking the subjects to choose their own groups, but also allow a qualitative result that could be used to create the foreground and background sounds in the foreground vs background preference research.

One scene was selected with which the subjects were familiar. The subjects were able to listen to the scene prior to the interview. They were also provided with the final version of the script for that section in advance of the exercise. The script is included in appendix B.1. The section of the play chosen for discussion was a one minute clip which included a range of different voices, sound effects, spot mic recordings and non-diegetic music. A list of the sounds used to create the scene are shown in table 4.1.

	Sound	Source
1	Pinocchio - voice over	Recorded in a semi anechoic room with a single close microphone (Mono)
2	Pinocchio - dialogue	Recorded in a semi anechoic room with a stereo microphone approximately one metre from the performer. Captured at the same time as objects 3, 5, 6 and 7. (Stereo)
3	Coachman - dialogue	Recorded in a semi anechoic room with a stereo microphone approximately one metre from the performer. Captured at the same time as objects 2, 5, 6 and 7. (Stereo)
4	Splash sound of Pinocchio landing in the sea	From BBC sound effects library. (Stereo)
5	The sound of the coachman putting the rope around Pinocchio's feet	Recorded in a semi anechoic room with a stereo microphone approximately two metres from the performer. Captured at the same time as objects 2, 3, 6 and 7. (Stereo)
6	The sound of the coachman putting the stone around Pinocchio's neck	Recorded in a semi anechoic room with a stereo microphone approximately two metres from the performer. Captured at the same time as objects 2, 3, 5 and 7. (Stereo)
7	Pinocchio's reins whilst walking to the cliff edge	Recorded in a semi anechoic room with a stereo microphone approximately two metres from the performer. Captured at the same time as objects 2, 3, 5 and 6. (Stereo)
8	Sound of the wind	From BBC sound effects library. (Stereo)
9	Market chatter as they leave the market	Recorded in a semi anechoic room with a stereo microphone approximately one metre from the performer. (Stereo)
10	Underwater sound	From BBC sound effects library. (Stereo)
11	Sound of the birds	From BBC sound effects library. (Stereo)
12	Sound of the waves	From BBC sound effects library. (Stereo)
13	Non-diegetic music	Provided by the composer. (Stereo)

Table 4.1: Audible sounds

The sounds were identified by a group of 2 expert listeners (who were unfamiliar with this radio play, although probably familiar with the story) who made a list of audible sounds with in the scene. These sounds were not analogous to the audio sources. For example, the underwater sound was created from a large number of recordings, including a recording of a microphone in a bag being dragged along a riverbed. However, the resulting underwater soundscape was considered one sound. Similar can be said of the wind and of the waves. The resulting sound cards are perhaps more analogous to auditory objects than audio sources.

Each sound was given a different card and the subjects were asked to perform a number of card sorting exercises. This approach was chosen as the physical representation of the sounds allowed the subjects to move them around easily and provided a physical focal point for discussion. The card sorting exercises began open and became more and more closed and focused, as shown in the list tasks below.

1. Arrange the cards in whatever way makes most sense to you.
2. Arrange the cards in order of importance to the scene.
3. Group the cards into as few or as many categories as you like.
4. Gradually reduce these into a smaller number of categories step by step until you have only two categories remaining.
5. Group the objects into two categories - explicitly foreground and background.

During the process the subjects were probed for their reasons for placement and thought process during the exercise. The names of the categories were created by the subjects during the experiment.

Producer results

The Producer preferred three categories. The first category she described as 'text'.

> "These text based sounds, they're part of the story or in the script."
>
> Producer, "Pinocchio".

The Producer was quick to put all speech into this category, noting the narration was more important than the dialogue. After some debate the splash sound was also placed into this category. The three spot effects were given their own category, while the more diffuse effects and the music were placed into an 'Atmospheres' category. The final arrangement

is shown in table 4.5.

> "You could tell the whole story using only the narration if you wanted"
>
> Producer, "Pinocchio".

Pinocchio voice over	Pinocchio dialogue	Coachman dialogue	Splash of Pinocchio landing in the sea
The sound of coachman putting the rope around Pinocchio's feet	The sound of coachman putting the stone around Pinocchio's neck	Pinocchio's reins whilst walking to the cliff edge	Wind sound
Market chatter as they leave	Underwater sound	Birds sound	Waves sound
Non-diegetic music			

Table 4.2: Three categories: text (green), spot effects (yellow) and atmospheres (red)

When asked to group sounds down into only two categories, 'foreground and background' table 4.3 shows the spot effects were considered foreground sound by the Producer.

Pinocchio voice over	Pinocchio dialogue	Coachman dialogue	Splash of Pinocchio landing in the sea
The sound of coachman putting the rope around Pinocchio's feet	The sound of coachman putting the stone around Pinocchio's neck	Pinocchio's reins whilst walking to the cliff edge	Wind sound
Market chatter as they leave	Underwater sound	Birds sound	Waves sound
Non-diegetic music			

Table 4.3: Two categories: foreground (green) and background (red)

Sound Designer

The Sound Designer first grouped the cards into five categories (table 4.4). Speech included any speech that needed to be intelligible. Spot effects included the sound effects that were recorded as part of the production process. The Splash was given its own category because it was a spot effect, but not one that could be recorded as part of the production. Atmospheres were grouped together, and music was given its own category.

Pinocchio voice over	Pinocchio dialogue	Coachman dialogue	Splash of Pinocchio landing in the sea
The sound of coachman putting the rope around Pinocchio's feet	The sound of coachman putting the stone around Pinocchio's neck	Pinocchio's reins whilst walking to the cliff edge	Wind sound
Market chatter as they leave	Underwater sound	Birds sound	Waves sound
Non-diegetic music			

Table 4.4: Five categories: speech (red), other effects (blue), spot effects (yellow), atmospheres (pink) and music (red)

When asked to reduce the number of groups the Sound Designer grouped the objects in a similar way to the Producer with the exception of the splash, which was placed in the spot effects group (table 4.5).

Pinocchio voice over	Pinocchio dialogue	Coachman dialogue	Splash of Pinocchio landing in the sea
The sound of coachman putting the rope around Pinocchio's feet	The sound of coachman putting the stone around Pinocchio's neck	Pinocchio's reins whilst walking to the cliff edge	Wind sound
Market chatter as they leave	Underwater sound	Birds sound	Waves sound
Non-diegetic music			

Table 4.5: Three categories: speech (green), spot effects (yellow) and atmospheres (red)

When forced to reduce the cards into two groups the Sound Designer grouped the cards in exactly the same way as the Producer (table 4.6).

Pinocchio's voice over	Pinocchio dialogue	Coachman dialogue	Splash of Pinocchio landing in the sea
The sound of coachman putting the rope around Pinocchio's feet	The sound of coachman putting the stone around Pinocchio's neck	Pinocchio's reins whilst walking to the cliff edge	Wind sound
Market chatter as they leave	Underwater sound	Birds sound	Waves sound
Non-diegetic music			

Table 4.6: Two categories: foreground (green) and background (red)

4.4.3 Function and Importance

The two subjects were asked to arrange the cards in order of their importance. In undertaking this task they were asked to justify their choices. Both based their importance order on the intended function of each sound. These intended functions were captured and are included in the results.

At one point during the category reduction exercise the Sound Designer and the Producer agreed on grouping the sound cards into three categories, described by the Producer as Text, Effects and Atmospheres and the Sound Designer as Speech, Spot Effects and Beds. These groups were linked to the function of each sound.

Pinocchio voiceover	Pinocchio voiceover
Pinocchio dialogue	Pinocchio dialogue
Coachman dialogue	Coachman dialogue
Splash of Pinocchio landing in the sea	Splash of Pinocchio landing in the sea
The sound of coachman putting the rope around Pinocchio's feet	The sound of coachman putting the rope around Pinocchio's feet
The sound of coachman putting the stone around Pinocchio's neck	The sound of coachman putting the stone around Pinocchio's neck
Non-diegetic music	Pinocchio's reins whilst walking to the cliff edge
Pinocchio's reins whilst walking to the cliff edge	Wind sound
Wind sound	Underwater sound
Underwater sound	Birds sound
Birds sound	Waves sound
Waves sound	Market chatter as they leave
Market chatter as they leave	Non-diegetic music

Table 4.7: Importance left: Sound Designer, right: Producer. Arranged in order of importance of individual sound

Both the Sound Designer and the Producer arranged the cards based on which sounds they thought the audience should be focusing on. The Producer and Sound Designer arranged the sound cards based on importance, placing the most important sounds at the top, and the least at the bottom. The results were identical with one exception; the music. The Sound Designer placed much more importance on the music saying at times it was at the foreground. Both considered the Pinocchio voiceover the most important sound. The top three sounds for both were the three speech sounds. The next group of sounds were the spot effects. The least important sounds were beds and atmospheres.

> "At times the music is the foreground, it's all you should be listening to."
>
> Sound Designer, "Pinocchio".

The Producer gave the music the least importance, placing it at the bottom of the list. The Sound Designer considered the wind more important than the waves because the wind had an intended unsettling emotional effect, whereas the waves only told the audience the action was taking place near the sea. Results are shown in table 4.7.

> "The wind has an unsettling feeling, it puts you on edge a bit, but the waves tell you they're near the sea. They sound nice but aren't as important as the wind."
>
> Sound Designer, "Pinocchio".

> "All the sounds are included for one reason or another, some are included because they are described in the script by the writer, others are added by Steve [the Sound Designer] during the mix, but there's always a reason for them."
>
> Producer, "Pinocchio".

	Sound	Function
1	Pinocchio - voice over	To tell the story from Pinocchio's point of view
2	Pinocchio - dialogue	Acting out the story, exposition and emotion
3	Coachman - dialogue	Acting out the story, exposition and emotion
4	Splash sound of Pinocchio landing in the sea	Informing the listener that Pinnichio has been thrown into the sea
5	The sound of the coachman putting the rope around Pinocchio's feet	Informing the listener that Pinnichio was tied up, conveying the feeling of claustrophobia, being trapped
6	The sound of the coachman putting the stone around Pinocchio's neck	Informing the audience that Pinnichio was tied up, conveying the feeling of claustrophobia, being trapped
7	Pinocchio's reins whilst walking to the cliff edge	Informing the listener the characters have moved toward the cliff edge
8	Sound of the wind	To create an unsettling feeling in the mind of the audience
9	Market chatter as they leave the market	To inform the listener they are alone when this sound stops
10	Underwater sound	To inform the listener that Pinnichio is under the water
11	Sound of the birds	To inform the listener that they are outside, in the countryside
12	Sound of the waves	To inform the listener that they are by the sea.
13	Non-diegetic music	Accentuate the emotions already being conveyed by other audio objects

Table 4.8: Function of the sounds, according to the Producer and Sound Designer

In identifying the importance of the sounds, the Producer and Sound Designer justified the inclusion of all the sounds in the scene, explaining why they were present. These justifications are shown in table 4.8. The justifications for the sounds were easily grouped into three categories each which had a clear function. The three functions were:

- To further the story through conveying the occurrence of an action or event.

- To convey an emotion or feeling to the listener.
- To provide contextual information such as a location or time of day.

These reasons are grouped in table 4.9 when their foreground and background category is also shown. It is unsurprising that the Sound Designer and Producer agreed on these justifications given they originally made the decision to include them jointly. These functions are shown alongside the foreground or background classification.

	Sound	Function	Category
1	Pinocchio - voice over	Events/actions	Foreground
2	Pinocchio - dialogue	Events/actions and emotion	Foreground
3	Coachman - dialogue	Events/actions and emotion	Foreground
4	Splash sound of Pinocchio landing in the sea	Events/actions	Foreground
5	The sound of the coachman putting the rope around Pinocchio's feet	Events/actions and emotion	Foreground
6	The sound of the coachman putting the stone around Pinocchio's neck	Events/actions and emotion	Foreground
7	Pinocchio's reins whilst walking to the cliff edge	Events/actions	Foreground
8	Sound of the wind	Context and emotion	Background
9	Market chatter as they leave the market	Context	Background
10	Underwater sound	Context	Background
11	Sound of the birds	Context	Background
12	Sound of the waves	Context	Background
13	Non-diegetic music	Emotion	Background

Table 4.9: Function of the sounds, according to the Producer and Sound Designer linked to the foreground and background categorisation.

4.4.4 Conclusions and Discussion

Neither Sound Designer nor Producer placed only speech as foreground, both included sound effects which they considered important to the story. Both felt that different sounds

moved between foreground and background depending on what was going on in the story. Neither Sound Designer nor Producer thought two categories (foreground and background) were as useful/natural as three. Both approached the importance task by a combination of assessing where the audience's attention was directed and an analysis of the intended function of each sound. The difference in opinion of the importance of the music can probably be explained by the difference in responsibilities of the two roles. The Producer was responsible for turning the script, written words, into a recording, taking a holistic view of telling the story. The Sound Designer is focused on the soundscape and overall sonic experience of the production.

Foreground and background sound categorisation was linked to the importance, foreground sounds were generally considered more important than background sounds. However, the functional analysis performed by the Producer and Sound Designer to arrive at an importance list revealed that every sound in the scene was important and every sound had a function. Functions ascribed to sounds by the Sound Designer and Producer were one of three types, providing information about story events and actions, providing contextual information for example a location or time of day, or conveying a feeling or emotion. This leads to two taxonomies:

- Attentional - Foreground and background classification which is based on where the listener's attentions is (or should be) directed.
- Functional - Based on why the sound was included by the production team.

The foreground / background, attentional categorisation is perhaps more important to the audience than the functional categorisation. This is evidenced in the feedback from the audibility research conducted by the BBC [60] and the complaints the BBC receives relating to background sounds being too loud in comparison with foreground sounds.

These two taxonomies are related. Sounds that convey plot information, which include dialogue, narration and spot effects were considered foreground sounds by both Producer and Sound Designer. Sounds which conveyed contextual information or conveyed a feeling or emotion, but did not convey plot actions or events were considered background sounds.

These results seem to relate to listening model literature, whereby complex scenes be categorised into foreground and background sound based on listener attention being "top-down" or "bottom-up". However, many experiments using these categories use real of simulated soundscapes as stimulus. This is problematic when using entirely constructed soundscapes, where every sound is placed for a reason. The importance exercise conducted here identified all the sounds as having some importance, all the sounds were desirable and had a function. The background sounds had a different function and were considered less

important compared with the foreground sounds. Applying "top-down" or "bottom-up" listening modes to fully constructed sound scenes is not straight forward. While the listener can choose to focus their attention where they want, the soundscape is constructed in a way that directs the audience's attention. In addition, considering background sound as unwanted, as one might in the design of communication systems, is also incorrect when applied to sound design. Background sound is present to perform specific functions, namely to provide contextual information or convey emotion.

4.5 Non-Speech Based Content

4.5.1 Production

Recording

A performance of popular music group "Everything Everything" was recorded at Media City in BBC 6 Music's live music studio. The space is designed to be able to accommodate live bands with an acoustically 'live' and acoustically 'dead' end. "Everything Everything" are a five piece popular music band consisting of guitars, bass, drums, keyboards, backing tracks and vocals. The band were set up in a circle, facing each other. The performance was mixed live in stereo for the BBC 6 Music broadcast and a multitrack was taken for this research.

Microphone Techniques

Traditional close microphone techniques were used to record the band, totalling 24 microphone or line sources. Each of these signals were split in two, one was mixed for the live broadcast, the other was sent to a multitrack recorder for a post production mix for the purposes of this research.

Post Production Mixing Set Up

The mix was created using Nuendo running on a Mac Pro, connected via MADI to an IOSONO spatial audio rendering system which was in turn connected via MADI to a set of DACs. These DACs were connected to the loudspeakers. One set of 26 loudspeakers was used to play back the mix. Although the mix was object-based a set of channel-based

reverberation effects were used.

Mixing Process

An experienced BBC sound engineer mixed the music recordings. Unlike the "Pinocchio" production described in section 4.4 a single DAW was used, which simplified the process. The levels were mixed as they would with a stereo mix. Some time was spent experimenting with the positioning of the sources, initially audio objects were placed in positions reflective of the positions of the real sources in physical space. For example the drums were all positioned slightly to the left of centre and the bass to the right. Having spent some time experimenting with this approach the engineer changed to mix the sounds more like a stereo mix, positioning the kick drum and bass in the front centre and left and right overheads at +/- 60 degrees. Having spent time experimenting with a mix that represented the physical space and a mix that was entirely constructed, the engineer favoured a fully constructed mix. This differed to a similar exercise that was conducted with a classical engineer mixing an orchestral recording. The classical engineer preferring to locate the microphone sources in positions reflected in the physical positions of the microphones in the physical space.

4.5.2 Analysis

Method

One multitrack recording of a performance of pop band "Everything Everything" was chosen by a BBC 6 music sound engineer. The sounds audible in this performance are shown in table 4.10. The same card sorting exercise given to the "Pinocchio" production team members described in section 4.4.2 was used here with the BBC 6 Music sound engineer.

Groupings

Table 4.10 shows how the subject preferred five categories: main vocals (green), backing vocals (red), guitars (pink), drums (orange) and everything else (yellow).

Bass	Jon Guitar	Alex Guitar	Drums SPDS
Bass Keys	Floor Tom	Jon Vox	Alex Vox
Jer Vox	Snare	Rim	Keys
Track	Kick	Snare 2	Tom Rack
Tom Floor 1	Floor Tom 2	Hi Hat	Overheads

Table 4.10: Five groupings

When asked to reduce the number of groups the subject initially grouped down to two groups: vocals, key and guitars in one group and everything else in the other as shown in table 4.11.

Bass	Jon Guitar	Alex Guitar	Drums SPDS
Bass Keys	Floor Tom	Jon Vox	Alex Vox
Jer Vox	Snare	Rim	Keys
Track	Kick	Snare 2	Tom Rack
Tom Floor 1	Floor Tom 2	Hi Hat	Overheads

Table 4.11: Two groupings

However, on reflection the subject regrouped into three groups: vocals, instruments and drums as shown in table 4.12.

Bass	Jon Guitar	Alex Guitar	Drums SPDS
Bass Keys	Floor Tom	Jon Vox	Alex Vox
Jer Vox	Snare	Rim	Keys
Track	Kick	Snare 2	Tom Rack
Tom Floor 1	Floor Tom 2	Hi Hat	Overheads

Table 4.12: Three groupings

When asked to reduce it further the subject chose two groups which they referred to as 'human stuff' and 'mechanical stuff'. This result is shown in table 4.13.

Bass	Jon Guitar	Alex Guitar	Drums SPDS
Bass Keys	Floor Tom	Jon Vox	Alex Vox
Jer Vox	Snare	Rim	Keys
Track	Kick	Snare 2	Tom Rack
Tom Floor 1	Floor Tom 2	Hi Hat	Overheads

Table 4.13: Alternate two groupings

When asked to categorise the sounds into foreground and background the subject mooted the idea that lead vocals would be foreground and everything else would be background, however, he dismissed this idea because the music was a 'piece of art' and therefore was all foreground. The subject did however suggest a useful foreground and background categorisation would be direct sound as foreground and room sound/reverberation as background. There is some sense in splitting direct and diffuse sound, given the direct sound could be considered independent (a classical orchestra can perform in different acoustic environments). However, this is restrictive and to be useful there would need to be evidence of different listeners preferring a different diffuseness for the same direct signal.

4.5.3 Importance

When asked to arrange the cards into an order of importance the subject physically arranged the cards in a similar way to how the piece of content was spatially mixed, with the main vocals front centre and the others placed on five levels of diminishing importance. This result is shown in table 4.14.

		Jon Vox				
	Alex Vox		Jer Vox			
Bass keys	Keys	Track	Jon Guitar	Alex Guitar	Bass	
Snare	Hi Hat	Kick	Snare 1	Rim	Overheads	Drums SPDS
	Tom rack	Tom Floor 1	Floor Tom	Tom Floor 2		

Table 4.14: Importance

4.5.4 Conclusions and Discussion

The “human stuff” and “mechanical stuff” categories identified by the engineer can be related to soundscape categorisation literature [56] where sounds were considered ‘natural’, ‘human’ or ‘mechanical’. Modes of listening are more difficult to apply to music when it is considered a ‘piece of art’. If the music is perceived by the audience as a whole, single auditory event, breaking it down into audio objects is challenging. The various ‘top-down’, ‘bottom-up’ listening modes are valid for music, but listener’s can chose to critically engage their attention on certain specific aspects of the music, or sit back and listen in a more holistic way.

The engineer considered music as a 'piece of art' and a whole which should remain as such. There was far less clarity in the intended function of different sounds within music compared with speech based content. Functional analysis of different musical genres (orchestral for example) might yield more conclusive results, however it is likely that the argument that it a 'piece of art' could still be made. There is some suggestion that the direct sound might be considered foreground with the indirect or reverberant sound considered background. If the direct vs diffuse categorisation is valid there may be a link to listening descriptively and direct sound and listening holistically and diffuse sound. However, in general the results from the music exercises were far less conclusive than those of the speech based content.

4.6 Background vs Foreground Listening Test

Listener attention research identifies foreground and background as two distinct categories which humans use when listening to complex sound scenes. Existing ANL (acceptable noise level) research focuses on listening to speech against background noise, classifying all background sound as unwanted. Previous research [87] was identified no correlation between listeners who self-reported liking levels of day-to-day background noise and listeners who found higher levels of background noise acceptable in ANL testing. As the card sorting research earlier in this chapter identified, the broadcast content that has been created by Producers and Sound Designers ideally contains no unwanted sound. Therefore while the foreground and background sound categories identified by the literature have been used for the following experiment, semantically the categories have slightly different meanings. Background sounds are still important to the perception of the sound scene, conveying emotion or contextual information, while foreground sounds convey event and plot information. The card sorting activities show these two categories are meaningful, the results of the following experiment cannot be directly compared to results from experiments where background sounds are considered irrelevant or noise.

A listening test was conducted to discover how audience's preferences for foreground vs background mixes varies. The football study in chapter 3 suggested that preferences for foreground vs background mix varied from person to person, however due to the live and uncontrolled nature of the test, further experimentation was required to fully and reliably demonstrate this phenomenon. A listening test was designed to address the experimental limitations of the football study shown in chapter 3.

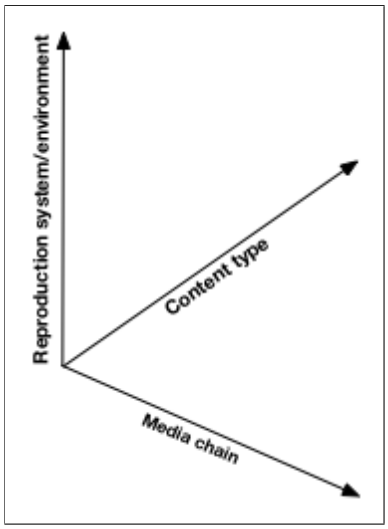


Figure 4.4: Perception of audio object problem space

4.6.1 Figure Ground Problem Space

Figure 4.4 shows a possible representation of the problem space this experiment investigated. The reproduction system and environment, the broadcast chain and the content type (genre) being shown as orthogonal and independent dimensions of the problem space. The impact of object-based content to allow dependent level control of foreground and background sounds on the genre and reproduction environment will be explored in this study. This leads to a null hypothesis: "loudspeaker layout does not affect preferred foreground verses background balance."

4.6.2 Method

The listening test was initially carried out in a controlled environment in BBC R&D's listening room in Salford. A 9.0 surround sound loudspeaker array was set up and listeners were seated in the sweet spot.

A user interface for the testing phase (shown in figure 4.5) was designed which aimed to

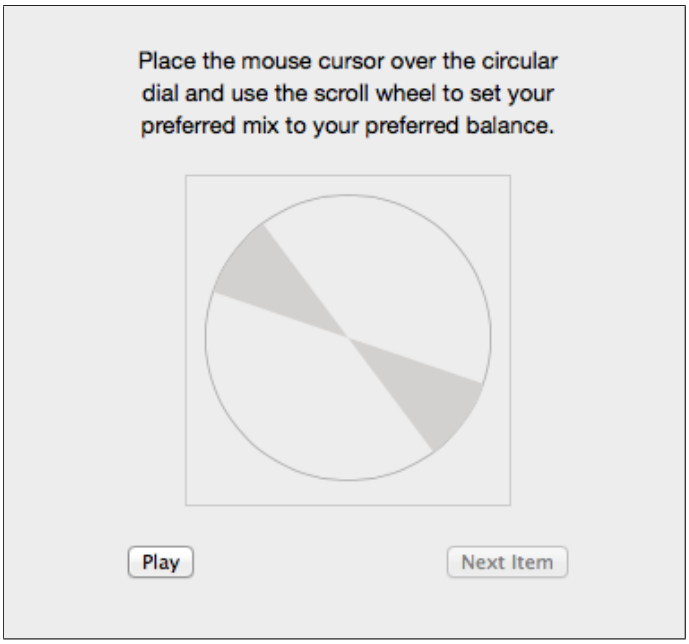


Figure 4.5: Foreground vs background user interface

remove the following influencing factors of the football study experiment.

- No visual cue of the current background/foreground mix is provided.
- There are no end-stops: the control is continuously variable crossfading the foreground and background audio objects in and out.
- It is possible to balance the mix with only foreground or only background sounds if desired.
- The initial background/foreground mix is randomised.

A hardware input was preferred over a software dial, therefore a scroll wheel interface was designed because the mouse scroll wheel is a commonly owned continuous rotation interface familiar to most. The scroll wheel balanced the foreground and background mix using constant power panning laws. Equations 4.1 to 4.4 are constant power panning laws. These panning laws ensure a constant perception loudness when balancing between equal loudness foreground and background signals.

given:

$$g^2_{foreground} + g^2_{background} = 1 \tag{4.1}$$

therefore:

$$g_{foreground} = cos(p') \tag{4.2}$$

and:

$$g_{background} = sin(p') \tag{4.3}$$

where:

$$p' = \pi(p + 1) \tag{4.4}$$

where:
$g_{foreground}$ is the gain applied to the foreground.
$g_{background}$ is the gain applied to the background.
p represents the balance position ranging from -1 to +1.

4.6.3 Additional 9.0 Production

Table 4.15 shows the different genres used for this listening experiment. 9.0 versions of the recordings were produced by professional BBC sound engineers. These 9.0 recordings were specifically created for this experiment using additional microphones for capture and were produced using the same methodology as "Pinocchio" and "Everything Everything". Production of the audio for the additional clips was not as complex as "Pinocchio", typically involving only level balancing, dynamic range control, some equalisation and spatial positioning of the objects. Engineers used the IOSONO system used for "Pinocchio" and "Everything Everything" to mix the audio objects while monitoring using the same 9.0 system that would be used for testing the subjects, in the same listening room. The same IOSONO room equalisation FIR filtering was used for both production and listening test stages of the research.

4.6.4 Clips

There is no recommendation for a 9.0 to 5.0 down mix yet, therefore the 5.0 down-mix equations were arrived at using the ITU recommended 5.0 to 2.0 down-mix equation as a basis and further listening and consultation with BBC sound engineers supported this decision. The 9.0 mixes were down-mixed to 5.0, 2.0 and mono using the down-mix equations shown in equations 4.5, 4.6 and 4.7. The clips were about 20 seconds long. This length was chosen in order to strike a balance between clips being long enough for the listeners to comprehend the narrative in terms of foreground and background, and short enough to maintain a constant loudness regardless of the background/foreground mix. Clips also needed to be chosen carefully to ensure there was no ambiguity with which sounds were classified as foreground and which sounds were considered background sounds (as noted in section 4.4 an audio object could move from foreground to background depending on what was happening in the story.) These foreground and background classifications were based on results from the card sorting exercises detailed earlier in section 4.4. Clip order was randomised. A hidden reference was included to confirm subject and experiment reliability by presenting the subject with the reference clip twice in order to test repeatability.

$$\begin{aligned} L_{5.0} &= L_{9.0} + \frac{1}{\sqrt{2}} TFL_{9.0} \\ R_{5.0} &= R_{9.0} + \frac{1}{\sqrt{2}} TFR_{9.0} \\ C_{5.0} &= C_{9.0} \\ L_{s,5.0} &= L_{s,9.0} + \frac{1}{\sqrt{2}} TRL_{9.0} \\ R_{s,5.0} &= R_{s,9.0} + \frac{1}{\sqrt{2}} TRR_{9.0} \end{aligned} \tag{4.5}$$

$$\begin{aligned} L_o &= L_{5.0} + \frac{1}{\sqrt{2}} C_{5.0} + \frac{1}{\sqrt{2}} L_{s,5.0} \\ R_o &= R_{5.0} + \frac{1}{\sqrt{2}} C_{5.0} + \frac{1}{\sqrt{2}} R_{s,5.0} \end{aligned} \tag{4.6}$$

$$M = L_o + R_o \tag{4.7}$$

Where:

$L_{9.0}$ is the left loudspeaker signal for the 9.0 surround array.
$R_{9.0}$ is the right loudspeaker signal for the 9.0 surround array.
$C_{9.0}$ is the centre loudspeaker signal for the 9.0 surround array.
$L_{s,9.0}$ is the rear left loudspeaker signal for the 9.0 surround array.
$R_{s,9.0}$ is the rear right loudspeaker signal for the 9.0 surround array.
$TFL_{9.0}$ is the top front left loudspeaker signal for the 9.0 surround array.
$TFR_{9.0}$ is the top front right loudspeaker signal for the 9.0 surround array.
$TRL_{9.0}$ is the top rear left loudspeaker signal for the 9.0 surround array.
$TRR_{9.0}$ is the top rear right loudspeaker signal for the 9.0 surround array.

$L_{5.0}$ is the left loudspeaker signal for the 5.0 surround array.
$R_{5.0}$ is the right loudspeaker signal for the 5.0 surround array.
$C_{5.0}$ is the centre loudspeaker signal for the 5.0 surround array.
$L_{s,5.0}$ is the rear left loudspeaker signal for the 5.0 surround array.
$R_{s,5.0}$ is the rear right loudspeaker signal for the 5.0 surround array.

L_o is the left loudspeaker signal for the 2.0 stereo array.
R_o is the right loudspeaker signal for the 2.0 stereo array.
M is the centre loudspeaker signal for the monophonic array.

Clip	Genre	Description	Foreground	Background	Panning
1	Drama	A 30 second section from radio drama, "Pinocchio"	Described in section 4.4	Described in section 4.4	IOSONO VBAP
2	Documentary	A 30 second section from radio documentary "A Cornish Gardener"	Dialogue and narration	Atmosphere and music	IOSONO VBAP
3	Pop Music	A 30 second section of "One Day Like This" by Elbow	Direct/close instrument microphones	Room microphones and Artificial Reverberation	IOSONO VBAP
4	Classical Music	A 30 second orchestra recording of a generic film score	Direct/close instrument microphones	Room microphones	IOSONO VBAP
5	Sports	A 30 second clip of Champion's League Final, Watford v Crystal Palace	Commentary	Crowd and pitch	IOSONO VBAP
5	Panel Show	A 30 second clip of I'm Sorry I Haven't a Clue	Contestants	Audience Laughter and applause	IOSONO VBAP

Table 4.15: Listening test clips

4.6.5 Test Environment

The test was conducted in a BBC R&D listening room, which is a soundproofed and acoustically treated room. There were four parallel loudspeaker layouts, mono, stereo, 5.0 and 9.0. IOSONO room equalisation processing was used to equalise the direct sound. For each loudspeaker the response was measured in four locations around the "sweet spot" (central listening position). Based on these measurements FIR filters were calculated each of the loudspeaker signals was processed in real time using these FIR filters. Consistent levels when switching between the layouts were measured and maintained throughout the experiment. Nine loudspeakers were set up and subsets of those nine were used for each of

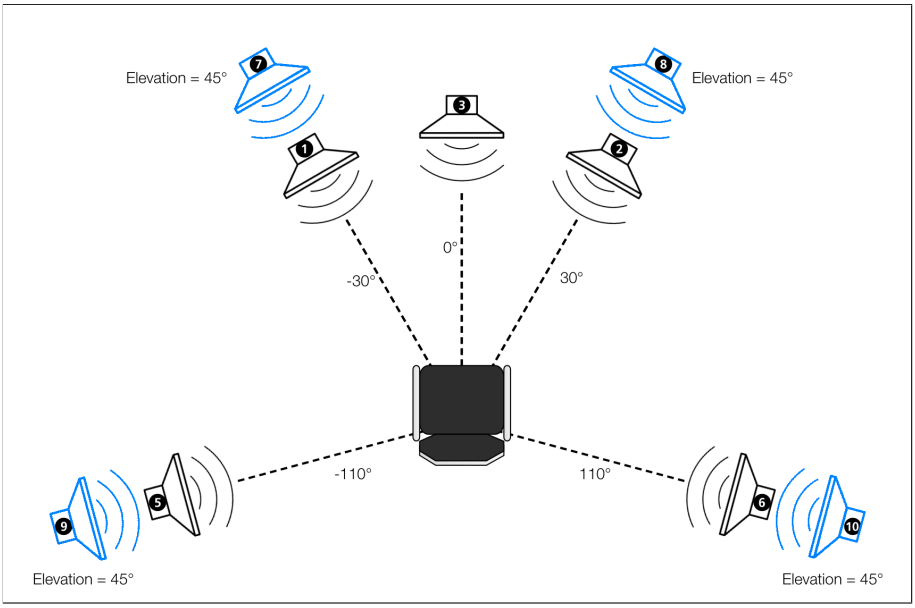


Figure 4.6: Loudspeaker layouts for the listening test.

the four layouts under test. Table 4.16 shows the loudspeaker subsets and figure 4.6 shows the layout positions. All loudspeakers were the same distance from the central listening position, the loudspeakers shown in blue are elevated at +45°. The subject's view of the loudspeaker positions was obscured during the test using acoustically transparent cloth to ensure the presence of the loudspeakers didn't influence the subject's results (shown in figure 4.7).

Figure 4.7: The listening test set up, with curtain to obscure the loudspeaker placement. One of the expert listeners standing aside of the rear right loudspeaker

Loudspeaker Name.	**No.**	**1.0**	**2.0**	**5.0**	**9.0**
Front left	1		X	X	X
Front right	2		X	X	X
Front centre	3	X		X	X
Rear left	5			X	X
Rear right	6			X	X
Top front left	7				X
Top front right	8				X
Top rear left	9				X
Top rear right	10				X

Table 4.16: Loudspeaker use table

4.6.6 Subjects

The panel was made up of 19 subjects. The age range was fairly evenly spread with subjects in 20-29, 30-29 and 40-49 age brackets. They were all experienced listeners having undertaken subjective listening tests before, but not all were expert listeners and none were required to pass a screening tests beyond the familiarisation stage of the experiment. Data was verified using results of the hidden reference, and analysis of the time spent by each subject on each example to ensure there were no anomalies.

4.6.7 Instructions

Subjects were asked to set their preferred balance using the mouse wheel to control the audio mix. No other information was given to the subjects. Before beginning the test, subjects were presented with some simple video instructions explaining how to control the test. Following the video their was a training stage which required the user to set a specific mix level to ensure they were dextrous enough to reliably control the experiment. Finally there was a familiarisation phase which allowed the subject to become comfortable with the controls with a sample clip. Subjects were not given any information about what the control was supposed to do. Subjects were presented with the user interface shown in figure 4.5 and asked to use the dial to set the balance to their preferred mix.

4.7 Results

In order to limit the range of results sensible clipping limits were defined. A maximum perceptible difference of 40dB was measured by expert listeners and differences beyond this were clipped as larger differences were considered imperceptible. A minimum noticeable difference between foreground and background of 1dB was measured by expert listeners using the test material and foreground vs background differences below this were considered equal as smaller differences were considered imperceptible.

The results for speech and music are presented separately. The responses given in the card sorting exercises detailed earlier in section 4.4 found the approach to identifying foreground and background categories for music more challenging, compared to speech based content for which there was clear consensus between the Sound Designer and Producer. As a result of these different results the method for classification of foreground and background for speech is different from the classification for music, therefore it makes more sense to present

the results separately.

Figures 4.8 to 4.11 are box plots (indicating interquartile range and outliers) of the results of the relative change in dB between foreground and background levels. 0dB represents a broadcast mix as set by a BBC engineer. A positive dB represents a boost to the foreground relative to the background, a negative dB represents a lowering of the foreground relative to the background. The relative level changes resulting from controlling the scroll wheel were based on panning laws to ensure altering the scroll wheel maintained a constant overall sound level.

Figure 4.8 shows the width of the preference distributions using a box plot graph. There is a preference of increasing background (non-direct sound) relative to clean sound. Notably, this preference for more non-direct sound is common across different reproduction systems as is shown in figure 4.9. 5.0 surround sound has a noticeably narrower distribution contrasted with other loudspeaker layouts. This result is noticeable but less pronounced with speech based content, as shown in figure 4.11. The narrow distributions found in 5.0 surround sound speech based content have an interquartile range of only 6dB.

Two two way ANOVAs for speech and music were performed. As with the card sorting exercises the results from the speech based content were more conclusive. The loudspeaker systems returned results of $F(3, 287) = 0.511, p < 0.05$, therefore the null hypothesis that loudspeaker layout does not affect preferred foreground verses background balance cannot be rejected. However, the results for genre give $F(3, 287) = 0.003, p < 0.05$ and the null hypotheses that genre does not affect preferred foreground verses background balance can be rejected. The music results allow no such rejection of null hypothesis. This statistical analysis is shown in figure 4.7 for speech and figure 4.7 for music.

Further analysis of the variance differences was conducted to explore the significance of 5.0 surround sound having a lower standard deviation and interquartile range than the other systems under test. Figure 4.7 shows the variance of results for each of the loudspeaker layouts (1.0 mono, 2.0 stereo, 5.0 surround and 9.0 with height). The variances of the foreground vs background results of each loudspeaker layout was compared. Differences between all layout variances is insignificant with the exception of 5.0 surround sound which as a variance of 41.76 contrasted with the other variances which are in the region of 111 to 118. Pairwise ANOVA testing shown in table 4.17 reveal that 5.0 surround sound has a significantly different variance and standard deviation than all the other loudspeaker layouts. This confirms the observations made using the box plots.

Comparison	P-Value
Mono to Stereo	0.87054
Mono to 5.0 Surround	0.00002
Mono to 9.0 Surround	0.9067
Stereo to 5.0 Surround	0.00004
Stereo to 9.0 Surround	1.46945
5.0 Surround to 9.0 Surround	0.00001

Table 4.17: Pairwise ANONA testing

The histogram plots shown in 4.12 and 4.13 allow the comparison of the speech and music results from this experiment, but it is also easy to compare results with the foreground vs background element of the football study in figure 3.5 in chapter 3. This comparison of results further highlights the influence of the user interface, the results from the football study are almost symmetrical (with a skew of -0.008) about the centre which reflects the UI design. In comparison the speech results in the foreground vs background study have an asymmetrical distribution with a long tail towards values representing louder background sounds (with a skew of -1.7).

ANOVA						
	SS	df	MS	F	p-Value	F crit
Factor 1 (Genre)	2,721.56481	3	907.18827	4.66702	0.00335	2.63606
Factor 2 (System)	449.35781	3	149.78594	0.77057	0.51128	2.63606
Factor 1 and 2	1,120.11303	9	124.457	0.64027	0.76225	1.91258
Within Groups	55,787.87222	287	194.38283			
Total	60,078.90787	302	198.93678			

Table 4.18: Statistical summary of foreground vs background mix preference data for speech

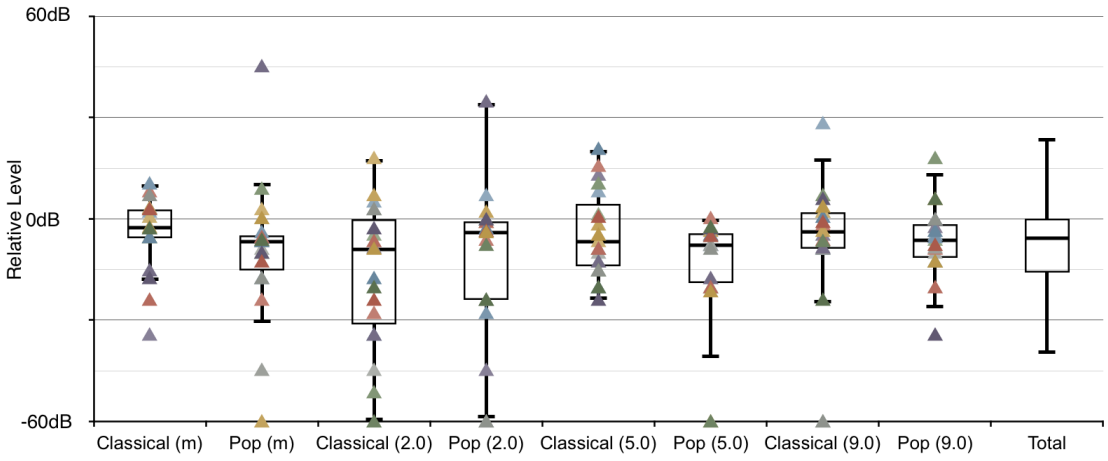


Figure 4.8: Boxplot showing foreground vs background mix set by participants for music content for different genres.

ANOVA						
	SS	df	MS	F	p-Value	F crit
Factor 1 (Genre)	285.29887	1	285.29887	0.87527	0.35108	3.90731
Factor 2 (System)	1,523.95617	3	507.98539	1.55846	0.20213	2.66789
Factor 1 and 2	1,032.67461	3	344.22487	1.05605	0.36988	2.66789
Within Groups	46,611.3807	143	325.95371			
Total	49,453.31036	150	329.68874			

Table 4.19: Statistical summary of foreground vs background mix preference data for music

Figures 4.14 shows how foreground vs background mix preferences changed with different age groups for speech based content. As a result of a panel of expert listeners identifying a just noticeable difference of 1dB, differences below this value are ignored. Younger subjects aged between 20-29 expressed a preference for more foreground than background contrasted to listener aged 40-49 who expressed a preference for louder background sounds. A chi-square test of goodness-of-fit was performed to determine whether the age influenced preference. Results for foreground vs background mix preference was not equally distributed across age groups ($X^2(4, N = 288), p < 0.00445$).

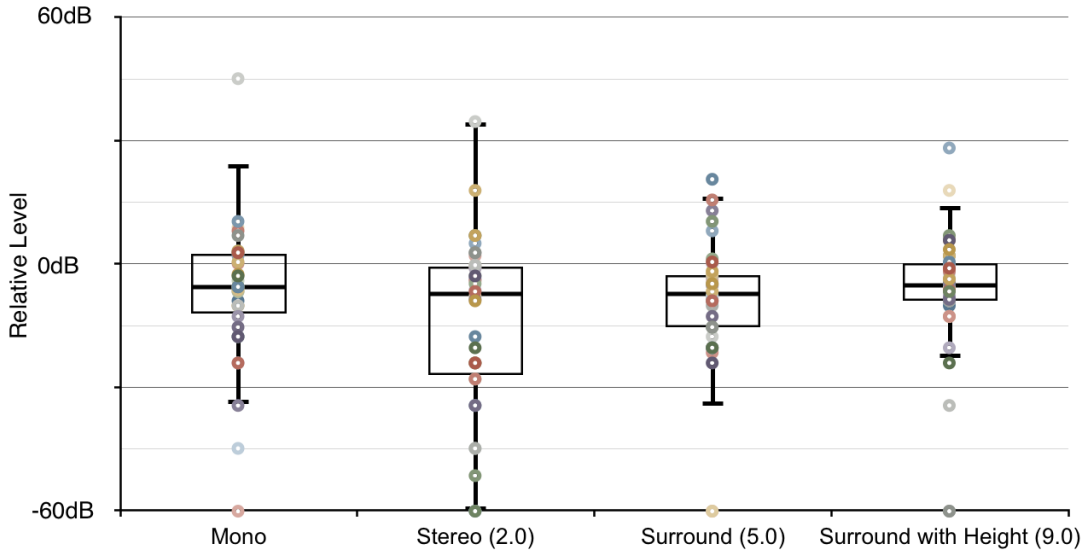


Figure 4.9: Boxplot showing foreground vs background mix set by participants for music content for different systems.

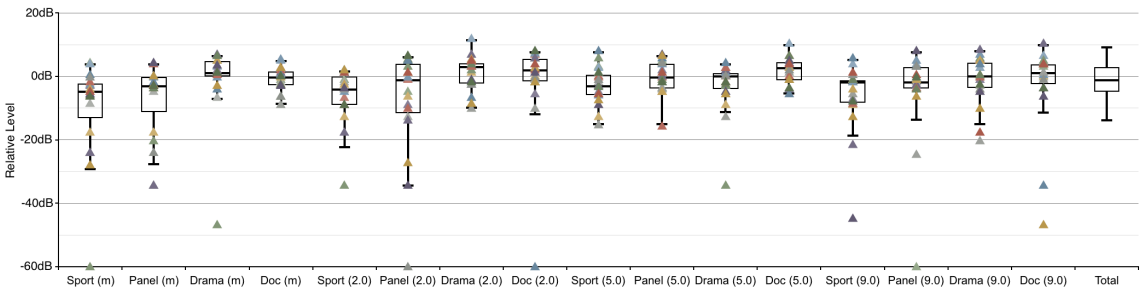


Figure 4.10: Boxplot showing foreground vs background mix set by participants for speech content for different genres.

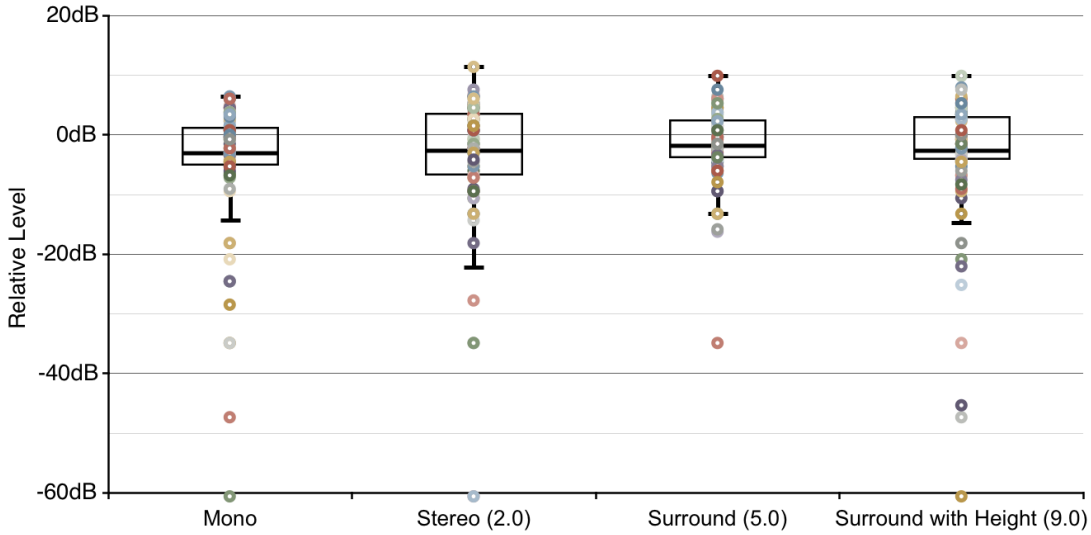


Figure 4.11: Boxplot showing foreground vs background mix set by participants for speech content for different systems.

4.8 Discussion and Summary

The results show there is no clustering of listener preferences of foreground vs background mixes, suggesting that the three clusters shown in the football study in chapter 3 are likely to be caused by the influence of the user interface design. The distributions of the preferences are generally clustered around the balance that would be broadcast however, the distributions suggest that many people would prefer a different mix to that which channel-based broadcasting current supplies and the tacit assumption that there is a normally distributed preference around the broadcast signal is not correct.

For speech content 89% of the time a noticeably different mix to that which would have been broadcast was chosen (greater than 1dB from equal to the broadcast mix). Roughly half of the time the background was boosted relative to the foreground and 38% of the time foreground was set louder than background. The results where foreground was set higher the average ratio of foreground to background sound level was 4.3dB with a standard deviation of 4.2dB. in cases where background was set higher the average ratio of foreground to background sound level was 8.9dB with a standard deviation of 10.1dB.

For music 95% of the time subjects set the balance to something different to a typical

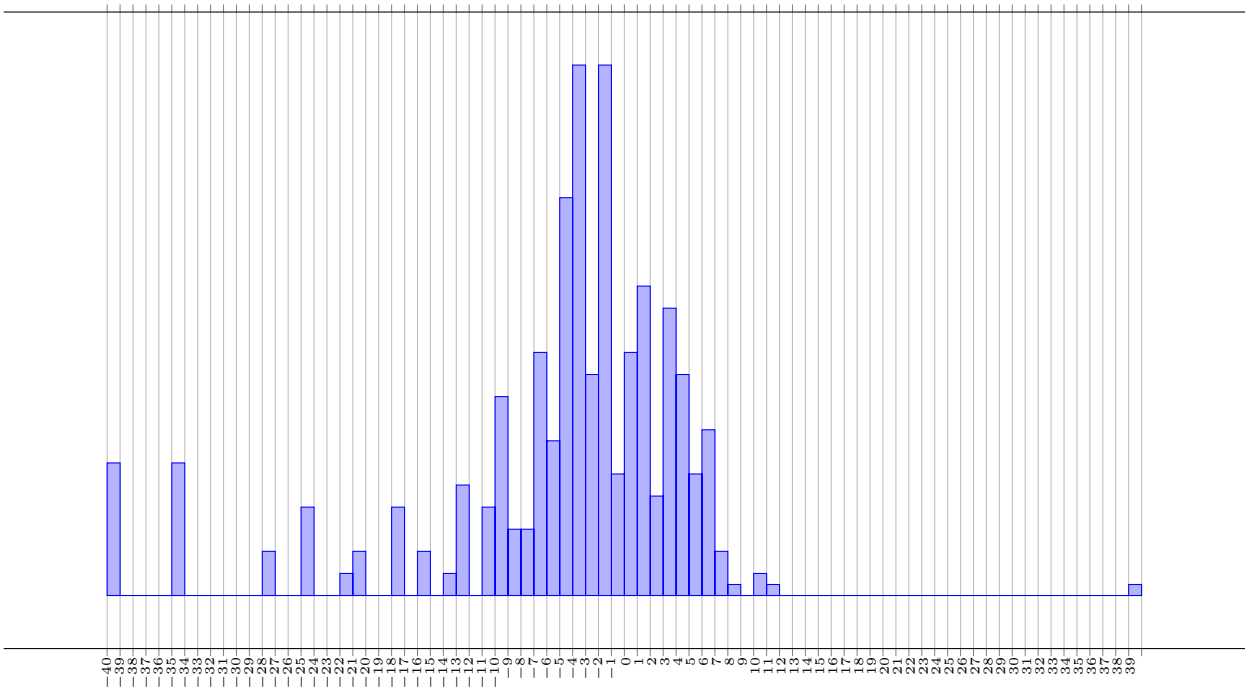


Figure 4.12: Histogram for foreground vs background preference for speech, values in dB represent relative foreground vs background mix.

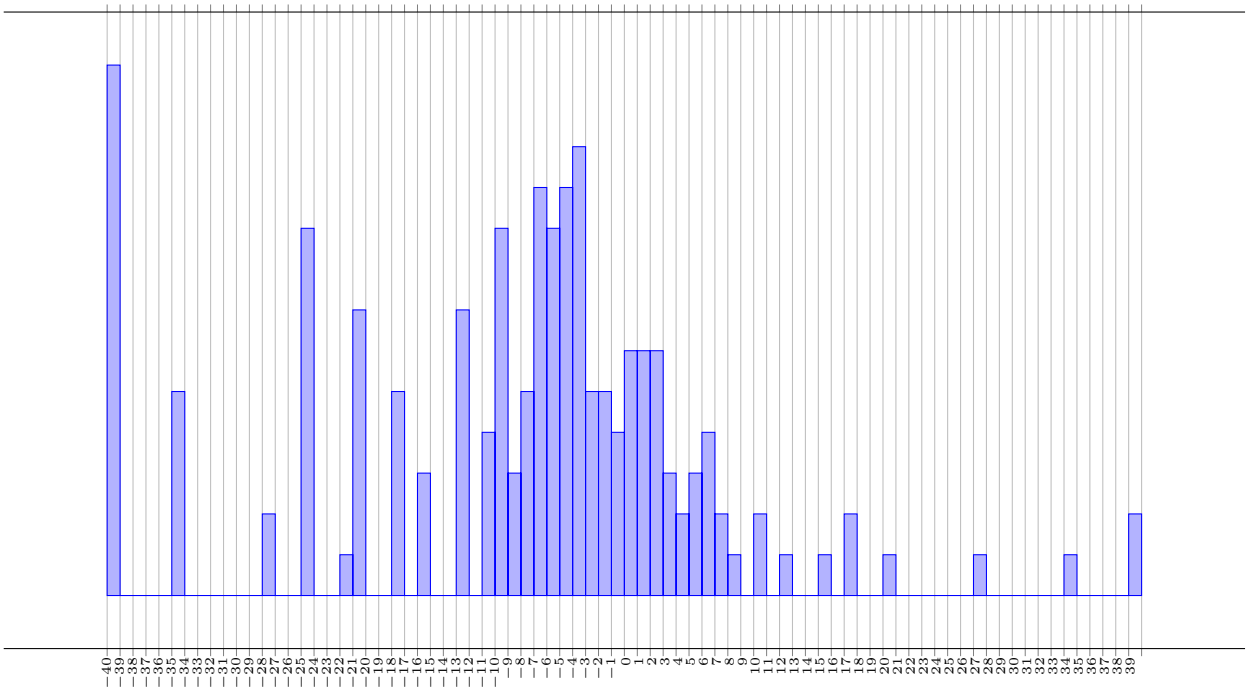


Figure 4.13: Histogram for foreground vs background preference for music, values in dB represent relative foreground vs background mix.

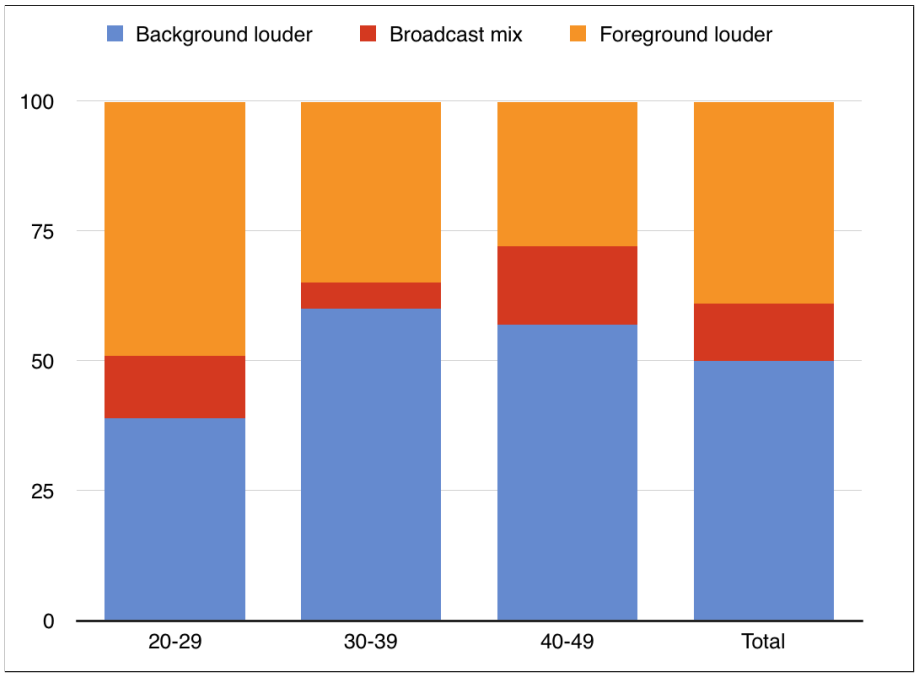


Figure 4.14: Foreground or background mix preferences for speech for different age ranges.

broadcast balance, 23% of the time the foreground was set louder than background, 71% of the time background was set louder than foreground levels. For the cases with higher foreground the average ratio of foreground to background sound level was 9.5dB with a standard deviation of 10.4dB. For the cases with higher background levels the average ratio of foreground to background sound level was -14.3dB with a standard deviation of 12.1dB.

These are considerable differences and suggest the audiences enjoyment of a piece of broadcast content could be improved by using audio objects to personalise the foreground vs background mix. This supports the results from the experiment in section 3 that much of the population would prefer a mix different to that which is typically broadcast.

The effect of varying the loudspeaker layout on foreground vs background mix was shown to be insignificant. This is an unexpected result. As described in sections 4.4 and 4.5 the content was created by experienced BBC engineers, following established mixing conventions. This resulted in the majority of foreground sounds panned to the front loudspeakers and the background sounds being fairly evenly distributed around the loudspeaker array. This could suggest that people can listen through the system to the programme material. There has been research[89] which has shown that for subjects perceiving soundscapes the

effect of reproduction system does not influence listening test results. However, for speech based content the genre (sport, drama, documentary or panel show) was shown to influence the preferred background verses foreground balance.

Although the difference in preferred foreground vs background mix preference between loudspeaker layout was not significant, the variance of the surround sound results is significantly different from the variance of the other three loudspeaker layouts. The variance and standard deviation for 5.0 surround sound was significantly lower than all the other loudspeaker layouts. This suggests there was more consensus in the preferred foreground vs background mix for 5.0 surround sound. This is a curious result which is difficult to explain.

Another unexpected but significant result is that younger listeners preferred louder foreground levels compared to older listeners. This result warrants further work to understand the significance. Specific age data was not captured, only age brackets of ten years. In addition, while the ages were reasonably well distributed between age brackets subjects were not recruited to fill specific age quotas. There were no subjects over the age of 50, which might not be old enough to get presbycusis seriously enough to affect their foreground verses background balance. The result is surprising because one might expect older audiences to prefer louder foreground compared to background sounds. The results in from the large scale audibility study conducted by "The Voice of the Viewer and Listener" [60] found the cause of most of the complaints about BBC sound is the background sound being too loud. The key differentiation with the audibility research report and this thesis is the audibility research assessed intelligibility. The study presented here focuses on the audiences preference. The findings here and the findings of the audibility research are not mutually exclusive. It is possible that older audiences suffer more from poor intelligibility due to background sound levels that are too loud, while at the same time preferring louder background sounds than younger audiences (as long as the levels do not adversely affect intelligibility). In addition, the subjects used for this research were experienced listeners whose analytical hearing might be better than listeners who complain about poor audibility of BBC services.

The results for music were less conclusive. While categorisation of diffuse sounds as background might suggest the reverberation is superfluous, one of the key findings of this work is that background sounds in broadcast content is not superfluous. In the context of speech based content the background performs an important function. Diffuse musical sounds also perform an important function which can be likened to the speech based background function of conveying contextual information. Reverberation can provide contextual information about the environment in which the performance is taking place, and how far away the performance is. However, comparing histograms for speech and music shown in

figures 4.12 and 4.13 respectively, the different shapes show the results from the music were more widely distributed and there was less consensus than speech, which had less than half the variance (110) than music (228). This evidence of greater subjectivity supports the conclusion that music is an art form and breaking it down into foreground vs background categories makes little sense. Genres of music vary so widely that further work exploring if this is true beyond popular music would be valuable.

Soundscape research uses scenes that are naturally occurring, or are designed to model naturally occurring sounds. This is not the case with broadcast sound, the scene is constructed based on the collective vision of a production team lead by the Producer. Because the presence of every sound is considered and there is intent behind its placement it maybe that the terms foreground and background are not semantically appropriate for use in a listening model for constructed broadcast sound. The foreground and background categories could be termed active and passive sounds, or informational and contextual sounds. These categories are analogous to some of the soundscape research, for example 'event sequences' can be considered analogous to the foreground category, while 'amorphous sequences' might be likened to the background. However the significant difference between this research and the soundscape research is soundscape research analysed a listener's perception and reaction to an uncontrolled, or perhaps managed sound scene. This research considered the intention behind a highly contrived and controlled sound scene.

Truax [70] listed three listening modes listening "in-search", "listening in-readiness" and "background listening". Listening in-search requires the listener to be actively engaged with the sounds, a listening mode that no doubt the Producer and Sound Designer are hopeful the audience has adopted. However, based on the results from this chapter while audiences might engage with foreground sounds by listening in-search, background sounds (those which have the function of conveying contextual information) could be engaged with using a less intensive listening mode from the Truax model.

5

Study 3: Location Based Drama

5.1 Introduction

This chapter describes the design and creation of an object-based audio drama entitled "Breaking Out". The audio drama was designed to take advantage of the adaptive nature of audio objects and provide a narrative that varies, depending on the listener's geographic location and the date/time the listener is experiencing the drama. This chapter considers the impact of object-based production and personalisation on the creative process (the story design and writing), and the recording and production workflows. This section also includes analysis of results from a series of three focus groups[11] and a retrospective survey of the listener's opinions for this drama. There was a control group who did not receive a personalised drama. This chapter aims to analyse the advantages of audio objects to audiences beyond providing them with the capability of additional control over foreground and background levels.

5.2 Context

Radio drama workflows have not changed significantly since techniques were developed 75 years ago [90]. Linear, static narratives are created and scripted. These scripts are rehearsed and performed by actors. These performances are recorded and mixed with music and sound effects in order to create a single static, normally stereo recording. This recording is then transmitted to the audience and each individual audience member experiences the same linear narrative, at the same time. The most significant change in radio drama

[11] The focus groups were conducted by Dr Maxine Glancy of BBC R&D

production was perhaps the move from analogue audio tapes to digital audio workstations [91]. While this change meant the roles in radio drama production altered slightly and multichannel post-production was made more easy, the general approach to the production process has remained largely unchanged. There has also been some work producing surround sound radio drama, although this is very niche, considering how little drama is created compared to other audio content.

5.3 Perceptive Media

Perceptive media is a term coined by the author and Forrester [3] to describe a piece of content that automatically undergoes adaptations in response to information about the individual viewers or listeners [92]. A similar concept to context aware computing [93], perceptive media adapts stories and narratives to suit individual audience members. The aims of this study are to explore two things:

- How a Producer might use audio objects to approach the creation of such a piece of perceptive content.
- How the audience react to a piece of content created from audio objects that adapts in response to data about them.

5.4 Narrative Adaptation

The first stage in the creation of the experimental audio drama "Breaking Out", was to understand what data the drama would collect about the audience and how the drama would react to that information. As this was to be a web based drama distributed online some examples of the type of data it is possible for a browser to detect are shown in table 5.2. For this investigation, only non-sensitive data that was easily collectable from the listener was used. The listener was not required to manually provide any information (such as their name, social network login details etc.). It was decided that the drama would be delivered by IP and would be a browser based experience. There is the potential for a hybrid (broadcast/IP) solution, but it was felt a simpler prototype using IP would be more appropriate at this stage in the investigation. There are limited data that can be determined by an internet browser. It was decided the drama adaptations would depend on a single piece of information; the user's present location. In order to allow realtime adaptation of parts of the script a browser based (JavaScript) text-to-speech engine [94] was employed to voice variables that changed in response to the listener's location creating

audio objects in real time.

Datum type	Method/Examples
Audio	Analysis of audio from attached microphones.
Video	Analysis of video from attached webcams.
Geo-location	Location can be determined and used to look up - time of day, language, local events/news, weather.
User-Agent	Can be used to identify the browser, operating system from which the device may be inferred.
Feature Detection	JavaScript can be used to detect features such as screen resolution.
Referrer string	The previously visited website can can determined form the referee string.
Downloaded files	Previously downloaded torrents can be determined for, the user's IP address.

Table 5.1: Data sources

5.5 Production Workflow

The processes followed to create and deliver this audio drama differed to that of a traditional radio drama. Figure 5.1 shows a workflow typical of a traditional radio drama compared to that used for this drama production.

Traditional radio workflows see the programme fully designed and produced with a single 'final mix' created before it is distributed to the audience. Using an object-based production workflow, a single 'final mix' never exists and the 'final mix' experienced by the listener is created at the point of consumption. The fact that programme makers don't have full control of the final mix is daunting to some producers who are familiar with traditional workflows. However, it can be argued that the belief of control that exists with channel-based content is misplaced, given the variety of devices and listening environments used by audiences to consume audio content. This study used a similar workflow as that described in the "Pinocchio" study in chapter 4 for the mixing, however the production

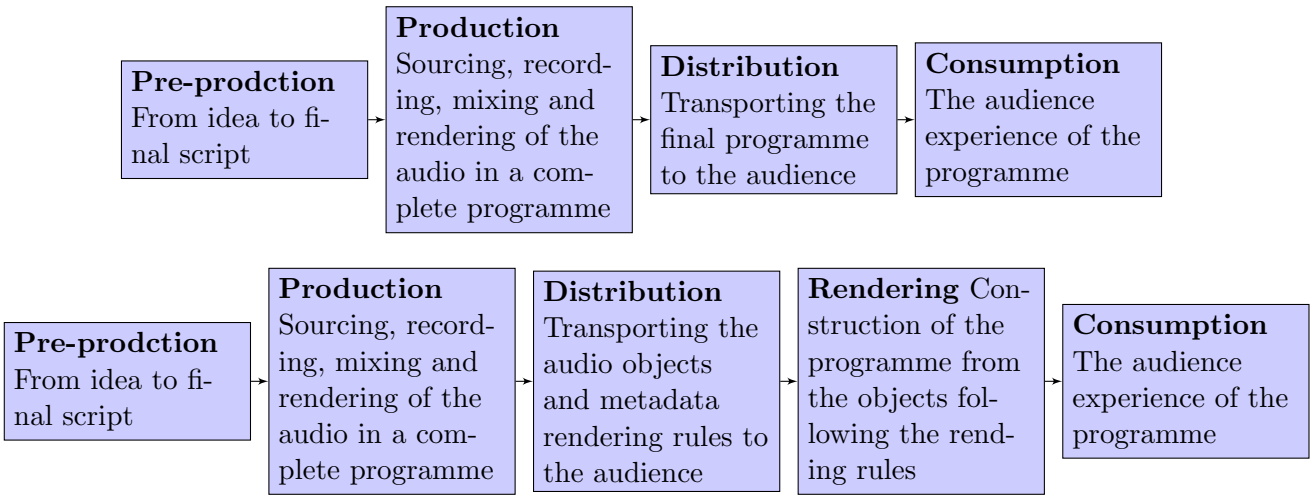


Figure 5.1: Workflow diagram, top: traditional workflow, bottom: object-based workflow.

process was more complex.

5.6 Design Process

In traditional radio drama workflows a scriptwriter is given a brief and delivers a first draft to the drama producer. The Producer then provides feedback to the writer. This is an iterative process which concludes with a final version of the radio drama script. This project used a similar process, but the writer (who would normally deliver the final script) had to work with the Producer to develop a creative idea that used the location detecting technology for the benefit of the story. Conversations with the writer during the process revealed that once they understood the technology they started to enjoy the creative process of working with constraints and variables.

5.7 Technology Constraining the Writer

In order to allow the localisation of the story constraints were placed on the writer. Although intelligible, the browser based JavaScript text-to-speech engine did not sound particularly natural, therefore it was stipulated that one of the characters should be robotic; a computer or artificial intelligence, to make the text-to-speech voice more acceptable to

the audience within the context of the drama.

5.8 The Writer Constraining Technology

Due to the nature of perceptive media a final version of the script never existed and each time the radio drama is listened to it is different. This leads to a very large number of final versions. The need for these different versions meant the Producer had to work closely with the writer to identify how the location data could enhance the story. While the overall arc of the story remained the same regardless of listener location, variables were woven into the narrative to allow the the story to be set in the location of the listener.

The personalised variables identified are listed below:

- Town where the story is set.
- Three well known places located nearby where the story is set.
- The weather at the time and place the listener is listening.
- Films being shown at a cinema near to the listener's location.
- The date the listener is listening.
- The news on the date the listener is listening.
- The social network used by the listener.

The variables which could be changed needed to be identified, and the writer needed to determine what the possible variables could be based on the location where the audience is listening. This meant the writer giving up some control over the script, while maintaining a control of the general story arc. The writer was open to exploring and experimenting with the capabilities of the technology, however it is likely that not all writers would be open to creating malleable stories, the route through which can be influenced by factors outside of their control.

5.9 Production Process

The drama featured three characters; two actors and a computer. The drama also called for the sound effects of a lift, such as 'dings', doors opening and closing, lift movements and buttons being pressed as well as music. The script was analysed and a list of required

sounds was made. The non-dynamic sounds were either sourced using the BBC sound effects library or recorded in the BBC R&D listening room.

5.9.1 Clean Capture

The adaptation of the drama occurred at the point of listening, not the point of production. This meant the individual sounds of the actors speaking and the sound effects needed to be recorded separately and treated as separate audio objects along the whole of the production chain. While the "Pinocchio" study in chapter 4 considers the content as two audio objects; foreground and background, this drama was more complex, requiring multiple audio objects. Performances of each of the actors, the lift sound effects and the music were captured discretely in dry acoustic conditions (see figure 5.2). These remained as separate audio objects and were played back with the correct timing, level and processing at the point of consumption.

5.9.2 Rendering

Once all the individual audio files were captured the drama script was translated into a JavaScript which contained the following metadata for each audio object.

- A unique ID.
- The location of the audio file (a URL).
- A textual description of the sound or in the case of dialogue the words delivered.
- Timing/synchronisation information.
- Processing information, such as the acoustic environment/room.

5.9.3 Robot Voice

The text-to-speech was also performed in the client browser, effectively creating audio objects in real time. Other more advanced and natural sounding text-to-speech algorithms are available [95]. However, the browser based requirement for this audio drama meant the speech resulting from the javascript text-to-speech engine was clearly synthetic. This limitation was thought acceptable for the purpose of this study. It is believed that client based text-to-speech technology will improve in future, and the constraints placed on the writer which required one of the characters to be robotic helped negate the issue of the

Figure 5.2: Recording environment. [12]

synthetic voice impacting negatively on audience engagement with the drama.

5.9.4 Variables

The variables identified in section 5.8 are populated using a number of online resources, such as the BBC's weather information [96], the BBC news podcast feed [97] and a cinema film release RSS feed. When text based variables are returned by these feeds the variables are voiced by the text-to-speech engine, in the case of audio files (for example the news podcast) the audio is played using the Web Audio API [39] as defined by w3c [98].

5.9.5 Sound Effects

The majority of the drama occurs in a small lift. In traditional radio drama production the sense of a small room would be created either by recording the actors inside a room with similar acoustic properties to the desired space, or by processing clean recordings in order to make it sound like they took place inside the desired space. The lift's speech generation is performed inside the browser so it is impossible to pre-process to make it sound like it

[12] Originally published in [3], ©BBC 2012.

took place inside a lift. Client-side processing had to be used to make the actor and the lift's voice sound like they were spoken in the same acoustic space. Using Web Audio API convolution all the reverberation was applied by the web-browser using impulse responses of a small room.

5.10 Architecture

Figure 5.3 shows a simple system diagram indicating the sources of the different data used in the audio drama. There were a number of different types of audio object that were used in order to create the audio drama, these are shown in table 5.2.

Audio	Source	Type
Harriet's voice	Recorded by the production	Stereo mp3 objects
Sound effects/ Music	Sourced from sound effects and music libraries	Stereo mp3 objects
Room Impulse Responses	Captured during the Production	Stereo PMC wav
Lift voice	Online RSS feeds	Dynamic text-to-speech
News report	Online podcast	Stereo mp3 pulled from RSS feed

Table 5.2: Audio object formats

5.11 Control Panel

The object-based approach to content production allowed for the location based personalisation. Some of the additional flexibility allowed by the use of audio objects was enabled by the addition of a control panel. This provided a GUI for controlling variables such as foreground/background mix, pace (the gap length between foreground files) and level and reverberation controls for each audio object. Figure 5.4 shows this control panel. This control panel was added for testing and demonstration purposes but was made unavailable for users during the web experiment to avoid any distraction and the possibility of influencing

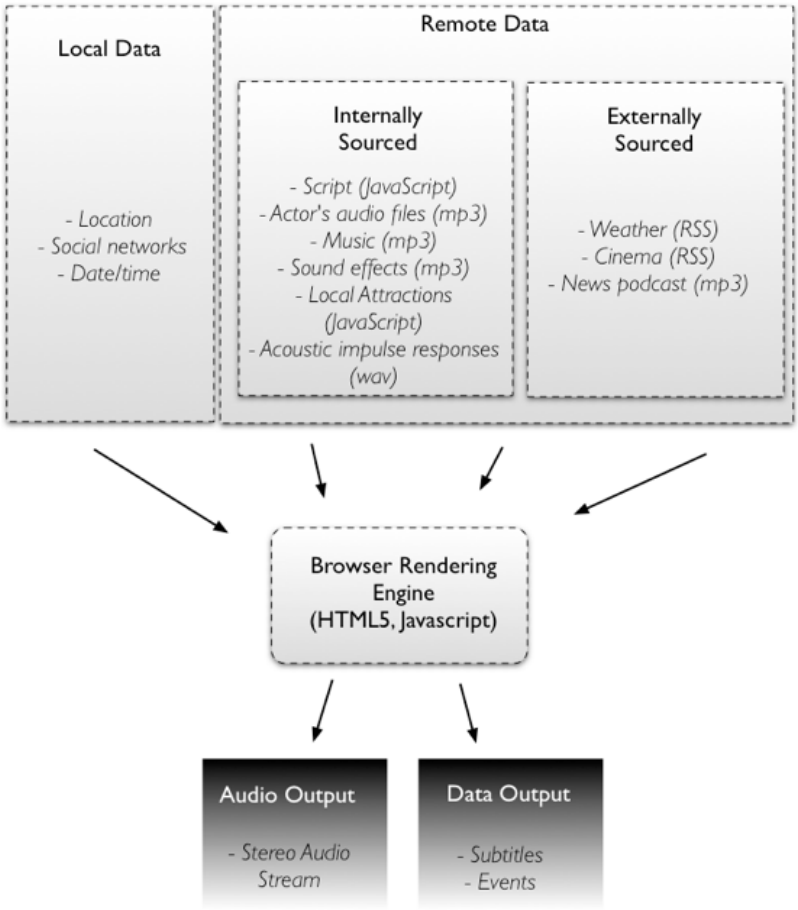


Figure 5.3: High level system diagram for "Breaking Out". [13]

the audiences' enjoyment of the experience.

5.12 Listening Test

The evaluation consisted of two separate research methods:

- Online listeners were asked to listen to "Breaking Out" then complete a retrospective

[13] Originally published in [3], ©BBC 2012.

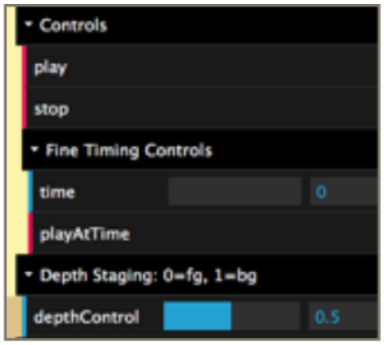


Figure 5.4: The "Breaking Out" control panel, made unavailable during the experiment.[15]

questionnaire.

- Three focus groups, each in a difference location, in which participants talked in detail about "Breaking Out" and the perceptive media experience in general, after listening to the drama.

5.12.1 Methodology

This study was designed to understand whether the overall experience of a piece of audio drama could be improved through the use of audio objects to enable a personalised experience. Local references were taken from the location of the server of the participant's internet provider or their GPS location depending on what data was available. A subset of the respondents were served a generic version of the play that was not located in their area. There were a few cases where the location look-up got the incorrect location or the listener denied their browser access to their location. Where location data was incorrect (did not match a self-reported postcode) or denied the results were excluded from this experiment. The majority of online participants found the drama virally from social networks or word of mouth. Once participants had listened to the drama they were directed to the online survey. The online survey was used as a data capture method in order to record information from online participants. In addition to the online questions a set of focus groups were conducted. Focus groups were conducted for a number of reasons, firstly to ensure a representative cross section of BBC audience. The online survey was self-selecting and attracted a skewed demographic who were not, on the whole, radio drama listeners. For the focus groups, participants were recruited using an agency to represent a cross-section of the audience, at three locations across the UK; Manchester, Glasgow and London. These people were all regular listeners of radio dramas. It was hoped the focus groups would verify the online data, and provide additional insight. The questions were written to assess

[15] Originally published in [3], ©BBC 2012.

the audiences' engagement and enjoyment of the experience, in order to allow the comparison of the listener's enjoyment of the personalised version with their enjoyment of the generic version. The respondents in the focus groups all received a localised version of the story. Copies of the questions used for the online survey (which also formed the basis of the focus group) are shown in appendix C.1.

5.12.2 Demographics

At the time of collating the results there were 755 responses to the online survey. 81% of the respondents were male. The age range of online respondents was skewed, with 50% of the online respondents aged between 18 and 30 years old and 53% between 31 and 45 years old. The focus groups were more evenly spread both in terms of age and gender. The bias towards younger male respondents becomes more evident when focus group responses are contrasted with online respondents, this is discussed later in this chapter. Another key difference between focus group respondents and online respondents is the online respondents did not tend to regularly listen to radio drama. The focus groups were based in three UK locations, whereas the online respondents locations were reasonably well distributed across the UK as shown in figure 5.5.

Each focus group had four participants with a total of 12 participants overall. Demographics (age, gender and background) was evenly spread. Participants were selected on the basis they regularly listened to radio drama, for example there were a number of Archers [99] fans.

> "The Archers was always on in the house when I was growing up."

Focus group subjects found radio listening a calming experience, and considered listening to radio drama to be a comforting experience. Over half of the focus groups reported that radio drama formed a backdrop to another activity (for example ironing), and was habitual, being part of or giving structure to a routine.

Figure 5.5: "Breaking Out" online listener locations.[16]

Regularity of online participants listening to radio dramas	
Daily	8%
Weekly	12%
Monthly	7%
Once in a while	37%
Never	36%

Table 5.3: Respondent listening habits.[17]

[16] Originally published in [3], ©BBC 2012.
[17] Based on data originally published in [3], ©BBC 2012.

> "I don't take radio dramas in all the time, good because it's not visual and not too distracting, you don't have to look-up from your work."
>
> Respondent 1, 31, Glasgow.

> "Sometimes I like to listen to random work, for something interesting. If I like it I carry on listening."
>
> Respondent 2, 38, Glasgow.

Focus group participants tended not to plan their radio drama listening, rarely arranging to listen to one-off radio dramas. Respondents decision to listen to radio drama was ad-hoc, primarily linked to the type of activity in which the listener was engaged.

5.12.3 Results

Results from the online survey and the focus groups are presented in this section.The online survey data is illustrated within the pie charts which do not include data from the focus groups. Certain aspects of this experiment were identified as notable by the listeners, such as the robotic voice and the graphical user interface and are therefore specifically analysed in more detail.

Impact of Robotic voice

The ability to synthsis certain lines in the play meant that one of the characters had their lines performed by a javascript text-to-speech engine. This voice was clearly synthetic, and despite being incorporated into the story by casting the text-to-speech engine as the voice of the lift listeners still commented on the synthesis. Listeners noted that modern lifts have a more naturalistic voices and this made the "Breaking Out" lift voice old fashioned or retro. Many people thought the flat emotion-less performance was amusing, warming to the style of dialogue.

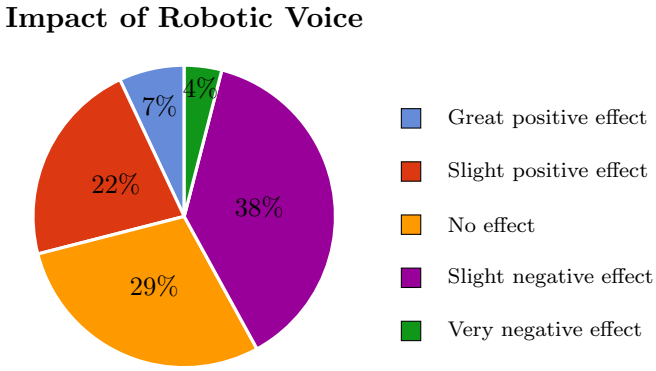


Figure 5.6: Online survey results reflecting how the robotic voice affected the enjoyment of the drama.[18]

> "The voice itself is a bit robotic, it reminded me of knight-rider. You mentioned new technology, but it seemed like old technology to me, it was a bit dry. But I actually enjoyed, I thought it was good. I liked having the Glasgow stuff in it."
>
> Respondent 2, 38, Glasgow.

Although approximately 30% of respondents reported the robot voice positively impacted their enjoyment of 'Breaking Out', the repeated use of this was questioned in the focus groups:

> "I don't know how many plays I could listen to when it's just a robot. You haven't got endless possibilities. If I had to tell you where I was anyway, I'd rather get a really good play than all the plays having to contain a robot."
>
> Respondent 3, 33, Manchester.

One respondent found the robotic voice slightly difficult to understand and alleviated this problem by using the subtitles feature. Using the subtitles resulted in the listener watching the experience. This led to some confusion as to the audio-visual nature of the content, whether is was a visual experience or a radio experience.

[18] Based on data originally published in [3], ©BBC 2012.

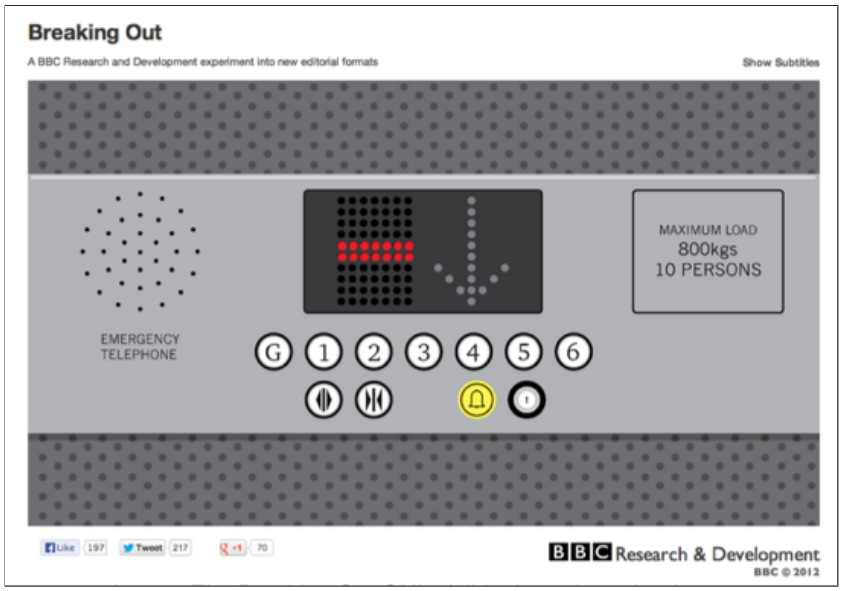


Figure 5.7: The "Breaking Out" visual design, visible throughout the drama, this contains no interactive elements.[19]

[19]Based on data originally published in [3], ©BBC 2012.

Graphical User Interface (GUI): Audio Experience or Audio-Visual Experience?

There was some confusion as to whether the drama was a radio programme or an audio visual experience. The GUI shown in figure 5.7 included some minimalistic animation to allow users to monitor at what stage there were when waiting for the audio to cache. This use of graphics was confusing for some who expected some more to happen, as illustrated by the following quotes:

> "It phased me that you had to watch it and listen to it. I was concentrating on the lift so much that I thought 'something will pop-up in a minute'."
>
> Respondent 5, 65, London.

> "I wanted to see her shoes walking out the door. It should be one or the other. You either have the visuals that go with it, and it's TV, or you don't."
>
> Respondent 6, 25, London.

> "I was waiting for something exciting to happen on the screen, like a visual drama and it didn't happen. I was thinking, as the arrow was going up, it was going so slowly, I thought 'please god, just get there', but I shouldn't have been watching it. I should have got up and wandered away."
>
> Respondent 4, 53, Glasgow.

The Effect of Localisation

A higher proportion of listeners reported liking the personalised version of "Breaking Out", with 79% liking it slightly or a lot, contrasted to the non-personalised version, with only 67% liking it slightly or a lot (a chi-square test revealed a significant difference $p < 0.001$). This suggests the audience experience of a piece of radio drama can be improved by this type of personalisation.

Recommending a Personalised Audio Drama to Others

Listeners to the personalised version of "Breaking Out" were more likely to recommend the drama to a friend than those who received the generic version. 56% of listeners to the personalised version would agreed they would recommend the drama to a friend, contrasted

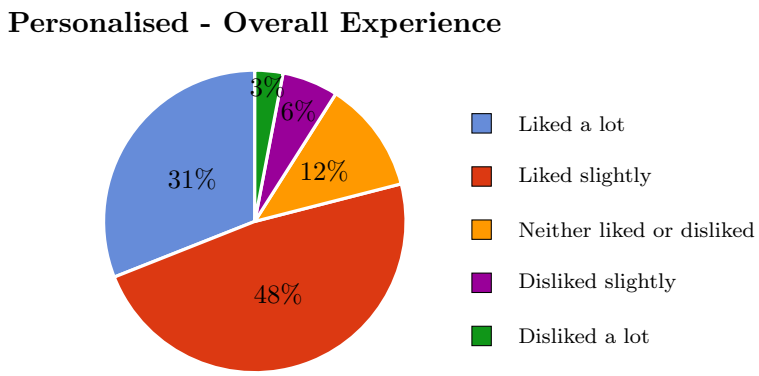


Figure 5.8: Online survey results reflecting how much participants liked the overall experience of listening to "Breaking Out" with personalised content.[20]

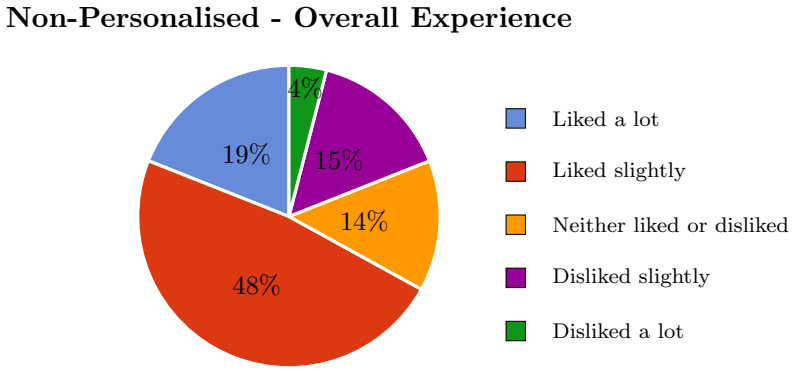


Figure 5.9: Online survey results reflecting how much participants liked the overall experience of listening to "Breaking Out" without personalised content.[21]

to 38% of listeners to the generic version. This suggests that listening to the localised version of the radio drama resulted in a better experience than listening to the generic version.

Visibility of Dynamic Personalised References

Figure 5.12.3 illustrates that those online listeners who received the personalised version found the local references to be either very or moderately local. Two thirds of the fo-

[20] Based on data originally published in [3], ©BBC 2012.
[21] Based on data originally published in [3], ©BBC 2012.

cus group listeners said the location based personalisations were more noticeable than the weather, date or film references, considering many of them to be 'local'.

The local knowledge of the listeners clearly influenced their responses, not all references were recognised by everyone. 6% of online listeners who received the personalised version did not recognise any locations. The majority of participants recognised as least some of the local references.

Story/Plot

Participants who received personalised references enjoyed the plot to a greater extent compared with participants who had listened to the audio drama without personalised content (see figures 5.12.3 and 5.12.3). 82% of respondents who listened to the personalised version reported liking the plot compared with 72% of listeners to the generic version (chi-squared test gives $p < 0.02$). These figures suggest that receiving personalised content within an audio drama can improve the enjoyment of the storyline.

It became apparent in the focus groups that the perception of the use of local references was more pronounced in those groups outside the London area. Participants in the Manchester and Glasgow focus groups where aware (from the first mention of a local reference) that "Breaking Out" was set in their locale. 63% of London based listeners recognised the location references, compared to 66% of non-London listeners. However there were comments from online listeners in London that back up the findings from the focus group.

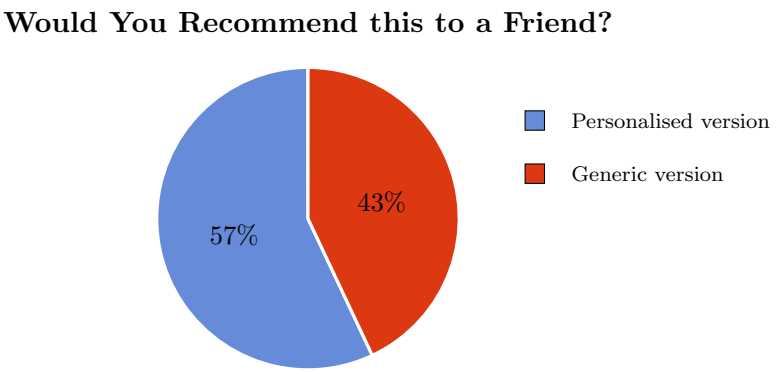


Figure 5.10: Online participants who noticed the localised content were 15% more likely to recommend a personalised audio drama to a friend (a chi-square test revealed $p < 0.02$).[22]

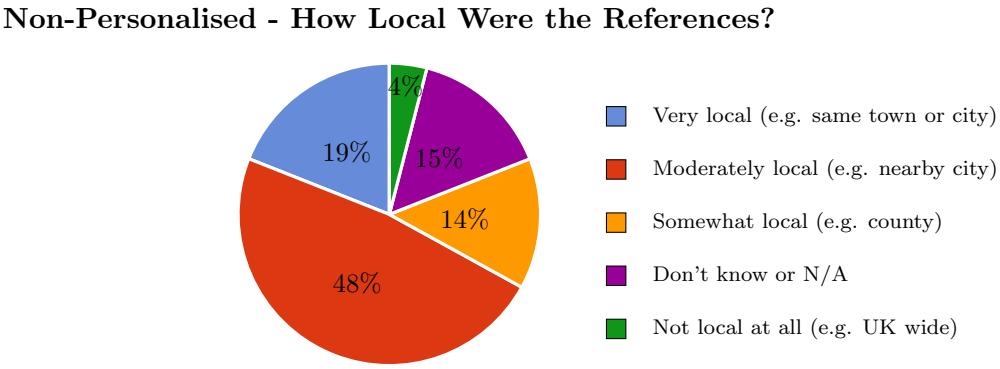


Figure 5.11: The majority of personalised references in "Breaking Out" were either categorised as 'very local' or 'moderately local'.[23]

[22] Based on data originally published in [3], ©BBC 2012.
[23] Based on data originally published in [3], ©BBC 2012.

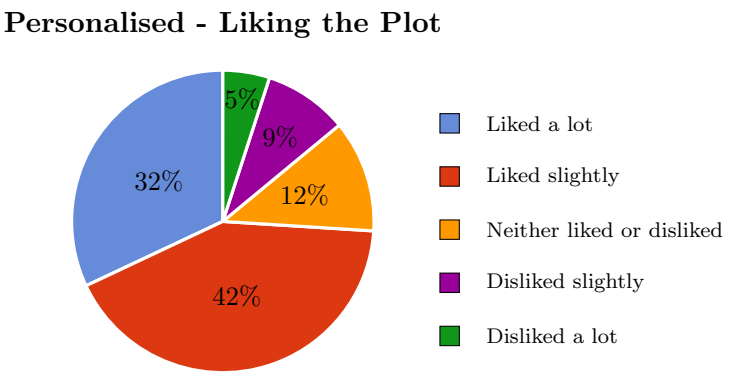


Figure 5.12: Nearly 75% of online participants who listened to "Breaking Out" with personalised references reported liking the plot.[24]

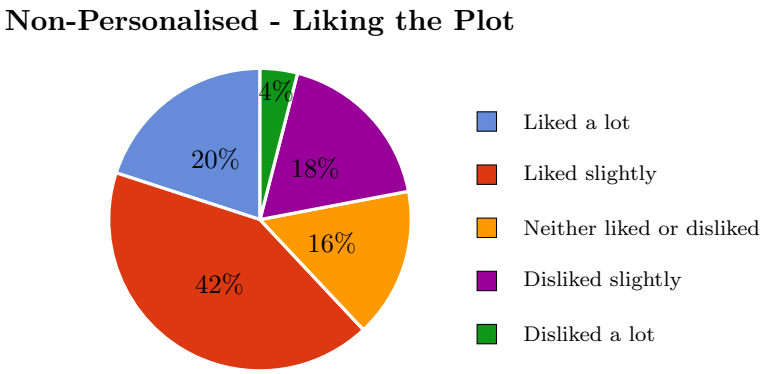


Figure 5.13: 65% of online participants who listened to "Breaking Out" without personalised references reporting liking the plot.[25]

[24] Based on data originally published in [3], ©BBC 2012.
[25] Based on data originally published in [3], ©BBC 2012.

> "I found that because all the locations were london based. I assumed it was a standard capital-centric viewpoint and made no association to the dynamic regional nature of it. But as a recent arrival in London, I could well imagine that in regional places this would have more resonance."
>
> Respondent 754, 31-45, London N19.

> "I really enjoyed the item but as I am based in London did not realise the location based personalisation until I partook in this questionnaire. Being from the North East though I would appreciate the Location based more as everything is very London centric in the media."
>
> Respondent 244, 18-30, London E1.

> "I think this is an interesting endeavour... must admit, that because I am a London person and used to London-centric things, I suppose it was less novel to hear St Paul's, Natural history museum, etc mentioned, and I couldn't quite make out exactly what the lift was saying when I though they might have been talking about restaurants... maybe Belgos? Anyway, the more recognised landmarks I guess reap slightly less reward than something a little more obscure which makes you think, oh, I am complicit in this understanding/imagining of the city. But of course, working with only general location information and little else about the listener makes it difficult to provide something more tailored. But I had tried this when living in a place other than London, the local references would have been even more rewarding."
>
> Respondent 845, 18-30, London W1.

While 77.5% of listeners outside London reported liking the experience contrasted to 70.1% inside London, the significance of these results is not enough to reject a null hypothesis ($p < 0.26$). This is likely due to the geographic distribution of the focus groups being limited to three cities contrasted to the online listeners who were widely distributed across the country as shown in figure 5.5.

> "Usually I sit down and it's about London, and it [was] good to hear it was about Glasgow."
>
> Respondent 2, 38, Glasgow.

> "It definitely sucked me in more, especially when it mentioned Chorlton, that's not a place everyone has heard of, so it definitely sucked me in a bit more."
>
> Respondent 3, 33, Manchester.

However, listeners in the London focus groups were less well aware of the local references, non of the London-based participants were aware of the local references until completing the questionnaire after the listening session.

> "It didn't twig until I did the survey afterwards, I thought it was part of the thing."
>
> Respondent 6, 25, London

> "I think if you lived in other parts of the country you would pick up on it more because you aren't used to hearing about your area."
>
> Respondent 7, 47, London

London focus group listeners wanted the local references to be more local, referring to places in their distract rather than the wider London area. Focus group listeners in Manchester and Glasgow were content with references from anywhere within the city. A number of respondents commented on the accent of the main character, although that was not intended to be a part of the localisation.

> "I'm West London, and that guy was more cockney."
>
> Respondent 6, 25, London.

> "I thought she was a southerner. She had a southern accent, which I noticed straightaway. A more Mancunian accent would have worked better. By using place names and what have you, if she had a Manchester accent it would have been more authentic."
>
> Respondent 11, 46, Manchester.

> "I also thought she wasn't from Manchester, which didn't necessarily put me off it, cause a lot of people who live in Manchester aren't from Manchester, but it was something I noticed."
>
> Respondent 3, 33, Manchester.

> "I wondered why she didn't have a Scottish voice. I thought is it a flaw, but it's a diverse culture. It popped into my head, you're doing something here, you have local reference points here, it's a new technology but you don't have a Scottish voice."
>
> Respondent 2, 38, Glasgow.

Focus group information suggests the level of personalisation could be increased through use of regional accents to help localised the story even more.

Online listeners who recognised more of the references reported enjoying the drama more than those who recognised fewer of the local references. 42% of listeners who heard the localised version said they liked the experience 'a lot' compared with 32% who did not get the local references (figure 5.12.3, $p > 0.01$). These results show that online listeners who heard the drama containing local references preferred the experience to listeners who did not hear local references. This result is reflected in the focus group responses.

> "I didn't know how they knew where I was, I thought it was just because the BBC is near Manchester. I enjoyed the affect of hearing about places I knew. I like things I can relate to. Something that I can experience."
>
> Respondent 8, 29, Manchester.

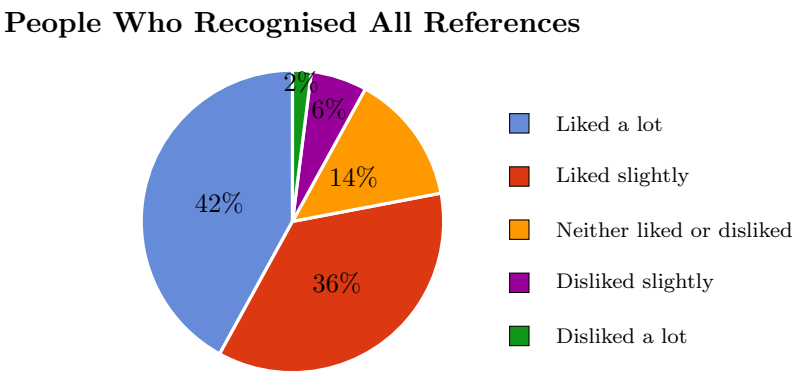


Figure 5.14: 42% of online participants, who recognised all of the local references, liked the references a lot.[26]

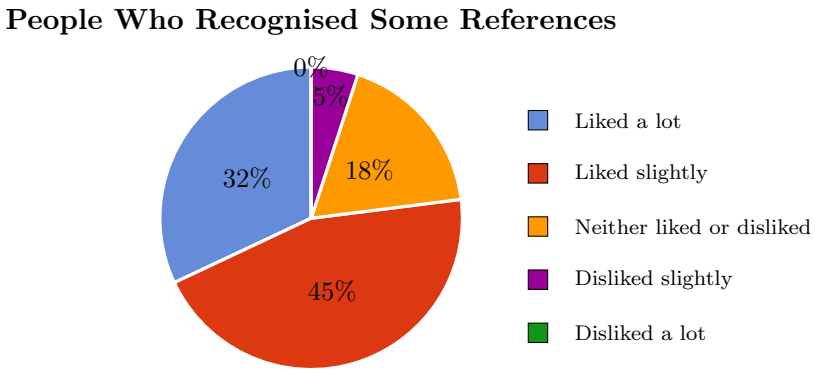


Figure 5.15: Online participants, who recognised some of the references, did not like the use of references as strongly as those that recognised all of the references.[27]

Escapism

Another influence on the listener's enjoyment of the experience which seemed to affect the London focus group more than Manchester and Glasgow groups was the mention of escapism. London listeners wanted to hear mentions of locations to which they had attachments that were not in the city: places they were familiar with, for example holiday locations and other places they have visited: places they felt at home, rather than their home.

[26] Based on data originally published in [3], ©BBC 2012.
[27] Based on data originally published in [3], ©BBC 2012.

> "It's like escapism. You listen to things that are not necessarily like your present mood. If I was listening to something that was too familiar I might not want it. I might not want the noise I hear outside every day."
>
> Respondent 6, 25, London.

> "It was too familiar, most of the plays I listen to, in my minds-eye it's a fiction I listen too, it was too familiar, this was just a bit grim, too real, not enough fiction or fantasy. The setting was too real."
>
> Respondent 2, 38, Glasgow.

5.12.4 Enhancing Engagement

Most (82%) of the online listeners who experienced the personalised version of the drama agreed that the experience was more engaging than traditional radio dramas.

Focus groups spoke of staying with the drama a 'little bit longer' as a result of the localisation. Non-London listeners said they were listening out for the local references. In some cases listeners suggested it became a challenge to recognise the local references, which arguably influenced the mode of listening and pushed the radio drama towards a more game-like experience.

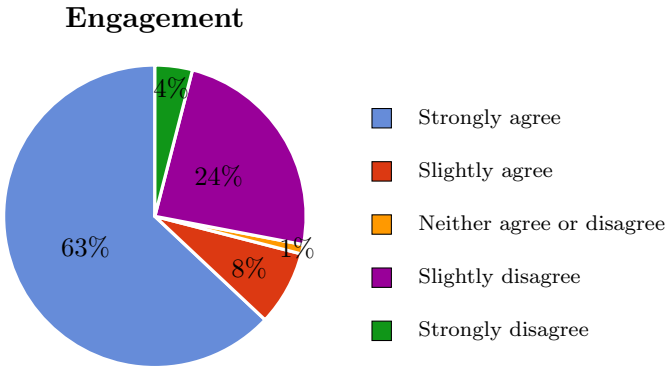


Figure 5.16: 82% of the online audience reported personalised content made them more engaged with "Breaking Out".[28]

> "It wouldn't make a difference as to whether I listened to the play or not, but I'd be more enthused by the local stuff. I could begin to relate to things that are in the story. I don't think it made a great deal of difference to my enjoyment of the story. I enthuse about it more because of that, but whether or not I would seek out programmes that did that I don't know."
>
> Respondent 9, 60, Manchester.

> "I knew all the references. It's nice that they were all the obvious ones as well. It made you imagine it more vividly, more than just skimming over it."
>
> Respondent 3, 33, Manchester.

> "I was listening out for things because I'd heard 'King Tuts' and 'the Science Centre', so I wondered what they were going to mention next. I couldn't believe this story, in which she's talking to this elevator. I couldn't visualise where they were because I couldn't see that if she's been in the house there's no point in telling her about these things because she wouldn't even know what they are."
>
> Respondent 1, 31, Glasgow.

Location/Setting of a Storyline

Both the online results and the focus group quotes suggested that the use of localisation made listeners feel closer to the story. Roughy 75% of online listeners agreed that they felt

[28] Based on data originally published in [3], ©BBC 2012.

closer to the setting of the storyline as a result of the localised content.

Characters

Participants in the three focus groups and the online survey were also asked if the use of perceptive media made them identify (feel closer to, sympathise/empathise) with the characters.

The personalisation used in "Breaking Out", while improving listeners' enjoyment of the drama, it did not make the listeners feel closer to the main character. There were comments from the focus groups that suggested that this could have been due to a lack of empathy with the character's condition.

> "I couldn't relate to her at all. I didn't notice the northern, southern or any accent. I just couldn't understand her problem."
>
> Respondent 9, 60, Manchester.

There were some suggestions from the focus groups about how their connection with the main character could have been improved.

> "A regional voice would be good. [Would that make you feel closer to the character?] Yes."
>
> Respondent 2, 38, Glasgow.

> "It didn't make me relate to her [the main character] more, but knowing the places that were mentioned it made me listen more intently than I normally would, just because I could recognise them."
>
> Respondent 10, 29, Manchester.

The mental heath problem being experienced by the main character was very specific, and a more likeable main character may have resulted in more positive responses from focus group members.

5.12.5 Perceptive Media and Product Placement

The ability to personalise content to this extent could be of interest to advertisers. During focus groups a few of participants had some concerns about the 'product placement' nature

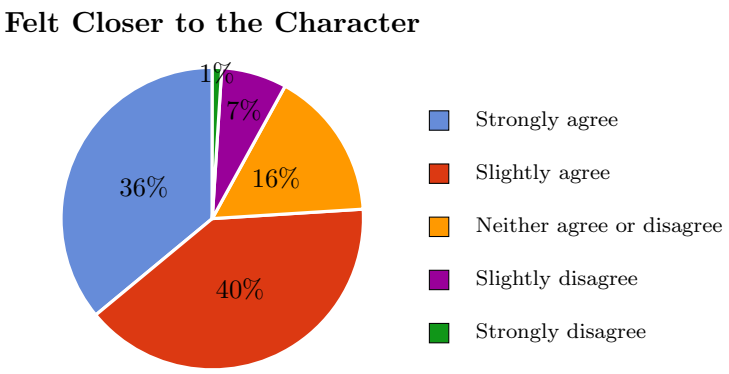


Figure 5.17: There was a mixed response to how participants felt about the main character. In this case the personalised content did not bring the audience closer to the character. [29]

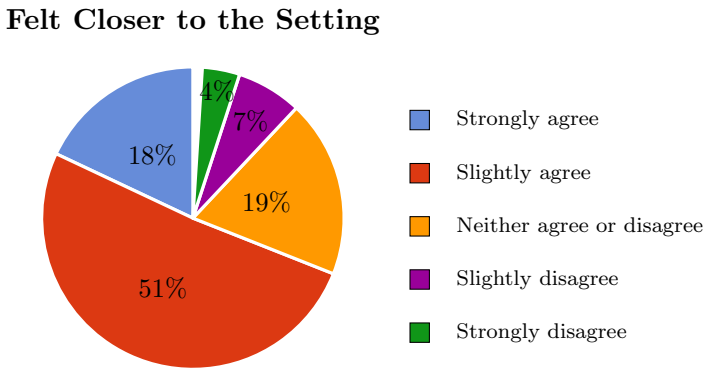


Figure 5.18: 75% of the online participants reported that the personalised content made them feel closer to the setting of "Breaking Out". [30]

of the experience, and there followed some discussion about the selection of the locations to be included and how this might be commercially influenced.

[29] Based on data originally published in [3], ©BBC 2012.
[30] Based on data originally published in [3], ©BBC 2012.

> "It could be used as a very sneaky localised advertising platform, not something the BBC would want to be accused of, but potentially more lucrative than banner adverts online."
>
> Respondent 21, 31-45, Belfast BT36.

> "It was really good. It was a good range [of references]. However, I listened with cynicism, I thought it was product placement, but this is just because you aren't used to hearing it."
>
> Respondent 1, 31, Glasgow.

> "I picked up on the local references but I wasn't bothered if it was advertising. I picked up on the local references, but I did not pick-up on the weather. I remember the date. Most of it was lost on me, it was only when I got to the end when I saw the questions and realised that this would have been so different if you'd lived in Manchester or Edinburgh. Then I appreciated it more."
>
> Respondent 4, 53, Glasgow.

5.12.6 Use of Personal Information

While focus group listeners were generally comfortable with the use of approximate location there were some concerns about how personal information was being used.

Respondents from the focus groups who were over the age of 55 had different views to those under 55. Younger respondents were more open to the use of personal information. This result is evident from the online survey too, which had a skew towards a younger demographic who were more comfortable with the use of personal information in order to improve an experience as shown in table 5.4.

Everyone in the focus groups said they would want to control these type of experience with a user account, as long as it did not take too much time or effect. They would not want to give their details over and over again or answer questions in order to get a personalised experience such as this. In addition to this focus group listeners did not want their data to result in them being contacted by Facebook or pestered by marketing email.

70% of online listeners enjoyed the passively-generated content used in "Breaking Out". This is reminiscent of the work in the football study in chapter 3, and the idea that the

Willingness to share data	18-30	45-60
Definitely yes	8%	2%
Probably yes	28%	26%
Unsure	18%	16%
Probably not	22%	25%
Definitely not	24%	30%

Table 5.4: Willingness of different age ranges to share data (chi-square test $p < 0.01$).[31]

Data	Yes	No
Exact location (within metres)	57%	28%
Your Mood (via your temperature, heart-beat, etc)	50%	31%
Personal details (your name, age, gender, race)	44%	36%
Social networking data	31%	50%
Internet browser history	18%	65%
Names of family/friends	15%	61%
Online purchase history	13%	73%

Table 5.5: Types of information people are willing to share for Perceptive Media.[32]

user does not want to have to interact with a piece of content, but they do enjoy it being personalised to suit them.

As discussed, personalised media experience and their reliance on the user explicitly submitting personal information generates a range of responses from audience members, reflecting their various needs and concerns. An illustration of these can be seen in the following selection of responses from focus groups.

[31] Based on data originally published in [3], ©BBC 2012.
[32] Based on data originally published in [3], ©BBC 2012.

> "My preference would be for me to have to give my consent. So I can choose which kinds of information would be sent or used. Politically I'm uncomfortable with that. It turns me into a 'consumer'. First of all I'd like to do it via an account, then as you get used to the technology the way you think about it may change. Then maybe do it through [tracking my] social media, but initially through an account."
>
> Respondent 6, 46, Manchester.

> "I'd rather give some quick information beforehand rather than it look at my cookies in the browser, etc."
>
> Respondent 1, 38, Glasgow.

> "I would open an account so that possibly in the future things could change, but I want absolute control and be able to cancel things. I'm rather negative about those sorts of things. I'm concerned it's a bit '1984', Orwellian stuff. My feeling is big-brother is getting too much information on us, and this step is one step too far for me. I would open an account because I'm interested, but I wouldn't give consent for any other information."
>
> Respondent 9, 60, Manchester.

> "I don't like the idea assumptions being made of what I'm interested in. It's nice to have something more random. Something that's not relevant to me might be something I'm actually interested in! ...as well as privacy issues and that kind of thing. I think I'd want to be able to control it, every time. Not necessarily entering the details every time, but having an account."
>
> Respondent 10, 29, Manchester.

> "I think including my name, or including a friend's name may be funny when it's added to the story. It would be interesting to see what they could do with it. But it's just knowing that I would get dumped with a ton of emails or personalised product placement. I understand and empathise with the need for it, but I don't necessarily want it on my doorstep."
>
> Respondent 4, 53, Glasgow.

To gain some understanding of how subjects might like to submit and control the use of personal information subject were asked:

> "How much effort would you be be willing to make in order to have a perceptive media audio drama?"

Most of the online respondents stated that they were willing to make their personal information available, either using a specific account or via their social media accounts, in order to enable a personalised experience such as "Breaking Out". 12% of online listeners would not be willing to share any data in order to enable a personalised service as shown in figure 5.12.6.

Comments from focus group members who were unwilling to share personal data in this way suggest they would be interested in personalised content that reacted to their mood rather than using their personal data. Focus groups talked about how mood based personalisation might effect their enjoyment of content of this type.

"If I'm post-work I'd be in a bad mood, and if I was really anxious and the actor is really anxious 'then I don't know if that would have a positive psychological effect on me."

Respondent 12, 36, London.

"You could set it to be really soothing, if you are agitated."

Respondent 7, 47, London.

"Mood is probably the least important, because it's not giving any information about you, so I wouldn't have any problems with that."

Respondent 2, 38, Glasgow.

"So it could ask 'do you want to have a personalised program this evening, or do you want to go with the standard one?"

Respondent 12, 36, London.

"How quickly would you get bored of it always being about you? If it's serialised it would be good though. You could be drawn into it that way. One off plays you wouldn't, a serial you would."

Respondent 1, 38, Glasgow.

"Then it could start suggesting things to you, and it could be really dynamic 'when you log into Facebook there'll be an app that will pop-up with a message, just like a serial ''have you seen what your friends are doing?' it could really push the boundaries on turning your life and your friends lives into a radio soap."

Respondent 6, 25, London.

"I don't mind name and location, but not address, family stuff. [Tracking my] online shopping would annoy me because if it was too 'made for me' I'm not interesting. It's too much. I'd like to open an account, I don't like my Facebook and twitter being linked. If it had 'tick' or 'no' for what you want to do, and then you can change the settings as you go along."

Respondent 3, 33, Manchester.

Many of the focus group respondents were comfortable with the use of physiological data which could be down to the anonymous nature of the data. The fact that physiological data could be gathered without requiring the user to enter any information creating a more passive experience, making the content seem more like traditional radio and less like a data

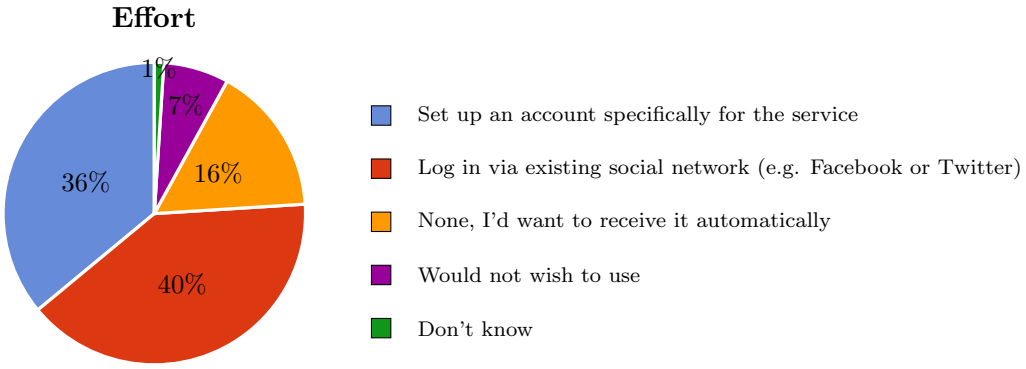


Figure 5.19: Most participants would be happy to make some level of effort to access a broadcast using perceptive media.[33]

driven experience.

While the concept of "Breaking Out" was understood by the users interviewed in the focus groups, it was clear they they did not have an understanding of exactly what was happening to their data. Respondents assumed that their data was being collected and stored somewhere, despite the fact all their data was processed locally. The browser required the user to grant the application access to their location, many users inferred this meant their data was being collected. More clarity as to how the data was being used and that it was not being transmitted or stored anywhere would be beneficial to educate and reassure users in future. However, the potential value of this data is likely to mean commercial implementations of this type of experience in the future would collect this data. In this scenario users would have to be made aware and agree to the user of their data in this way.

[33]Based on data originally published in [3], ©BBC 2012.

5.12.7 Wider Effects of Perceptive Media

There was some discussion and enquiry about the nature of experiencing perceptive media in a social context.

> "If you personalise it everyone has their own experience of it. You could talk about it...'what was your like?' ...'I had sun'...'I had rain'. It takes that familiarity out which uses the social aspect of listening to something, a radio."
>
> Respondent 6, 25, London.

There was further speculation about where perceptive media could take media experiences in the future.

> "Could you join the information of 3 or 4 people listening together at the same time?"
>
> Respondent 7, 47, London.

> "Yes, get 4 of my friends listening to personalised versions of The Archers, in 'City Life'. It's like a more advanced version of a podcast -a radio / social platform!"
>
> Respondent 6, 25, London.

These types of comments not only reflect that some of the participants accept perceptive media as a new format in broadcasting, they also provide indications of the types of perceptive media features they might like to experience in the future.

5.12.8 Social Listening

The launch of "Breaking Out" received considerable interest from mainstream technology press [100] [101] [102] and as a result was talked about across social media channels. There were 295 social media posts mentioning this perceptive media project, with 762 retweets or shares. The majority of the online conversation was not feedback, but were just posts sharing the link to the experiment. However, there were a small number of comments which did more than just share the experiment. These comments reflect those comments from the focus groups. For example, there were a number of comments that showed listeners thought the experiment was 'creepy', shown in figure 5.20. There was a comment which highlighted concerns about the use of personal data (shown in figure 5.21). There were other comments about possible advertising applications (see figure 5.22) and the impact that this would have on the creative process (see figure 5.23).

Figure 5.20: Audience fears expressed on social media

Figure 5.21: Comments on social media about personal data usage

Figure 5.22: Comments on social media about advertising applications

Figure 5.23: Comments on social media about effects on creative process

5.13 Discussion and Summary

This chapter covers the creation and audience experience of "Breaking Out". "Breaking Out" is an object-based radio drama which was created to help understand the impact of using audio objects to create personalised content experiences. The experience used the listener's location and the content was adapted in response to that data. After listening to the drama 744 self-selected listeners filled in an online questionnaire. In addition three geographically separate focus groups were conducted. Listeners who heard a personalised version of the drama rated their enjoyment of the experience more highly that those who did not hear a personalised version. They also felt closer to the location, and liked the overall experience more than listeners to the non-personalised version. The personalisation also made listeners more likely to recommend such an experience to a friend.

None of the negative impacts on the audience experience were directly related to the use of audio objects, but were linked to traditional production techniques and not caused by the use of audio objects.

Aspects of the experience that listeners thought had a negative impact were aspects relating to the production as a whole, rather than its object-based nature. Specifically, the use of visual stimuli detracting from the audio experience and the audience perception that the main character was unsympathetic. These could have negatively impacted on the audiences engagement levels but would have had the same influence on the personalised and non-personalised versions.

While there are a small number of studies in the literature that demonstrate systems and methods of location based storytelling [103] the research tends to demonstrate working systems from a technical standpoint rather than trying to assess the technology's impact on the audience experience. Other storytelling research that incorporates location into narrative [104] tends to place the story onto the location, rather than placing the location into the story. These location based experiments also tend to require the audiences to move around a specific location to navigate the narrative and the whole experience becomes more interactive and game-like, unlike the sit-back experience that "Breaking Out" provides.

This study has shown how audio objects could be employed to create personalised audio experiences in the future. It has demonstrated that the ability to personalise content using object-based audio and that using location information can improve the audience experience. The results suggest there is an audience benefit to providing personalised experiences such as this. The study also explored some of the creative and technical challenges involved

in producing content of this type.

6

Discussion

6.1 Production

Radio production workflows have not changed significantly since broadcasting began in the 19th century. The biggest change in workflow was the move from analogue to digital as a production format which allowed non-linear editing using digital audio workstations, rather than splicing analogue tape to create edits. The studies in this thesis assess the impact of using audio objects as an output format on conventional digital production workflows.

6.1.1 Spatial Mixing of Audio Objects

All the samples in the foreground vs background study were produced by BBC engineers using an object-based audio production system that allowed full 360°spatial audio production. There was a notable difference in the engineer's approach across different genres. The production team's aims for the sport, classical music and drama were different to the aim for popular music, as observed in chapter 4. The production of the sport example aimed to create an audio experience that was an idealised version of the experience the listener would receive if they were present at the event. Classical music takes the same approach. These genres do not recreate an experience as it would if the listener was present at an event, the aim of the production team is to create an idealised sound, that is quite often considered better than the sound an audience would experience had they been present. For the sport production the engineer built the sound scene around an imagined position at the event, positioning audio objects spatially around them as if they were actually present. The same is true of the drama production, although because the sound scene being created has never actually existed the scene being built is fictional. It is however, representative of

a physical space which could theoretically exist. However, popular music of the type used in this research is not based on any kind of physical reality. In the "Everything Everything" production in chapter 4 different audio objects used different artificial reverberation settings. Positions of the audio objects were based on an imaginary sonic space that would be impossible to recreate in a physical space. For example, each drum was a separate audio object which, in reality are physically quite close together. However, the way the BBC engineer placed them in the 9.0 mix was very widely spatially distributed around the listener, placing right and left overhead drum microphone sources at +90°and -90°respectively. It is worth noting in the card sorting exercises in section 4.5 the overheads were grouped as one object, despite being two sources coming from opposite directions. This spatial positioning of sources is unlike the classical music example used in the foreground vs background study, which used positions that were reflective a real space; the audio sources were positioned based on the physical layout of the microphones in the concert hall and therefore the orchestra. The insight from these production workflow observations should influence the producer's vision of the range of possible audience experiences (see discussion in section 6.1.4) and their decision of how to break down the sound scene into discrete audio objects, as discussed in section 6.3.2.

6.1.2 Listening Models Based on Function

Crafted broadcast content contains no incidental or accidental sound, everything audible is there as a result of a considered decision. Existing listening consider the perception of natural or real world soundscapes. For this reason existing foreground vs background attention models used for understanding complex sound scenes are not directly applicable to broadcast applications. Analysis of the Producer's and Sound Designer's views of audio objects enabled existing foreground vs background listening models to be adapted and applied to broadcast audio content. The film theory "mise-en-scène" means the placement of objects within a scene to add meaning to a story. Its use in film theory arises from the idea that every object within a piece of film or theatre is placed there for a reason. This is also true of broadcast audio, it is a construction and all the audio objects within a sound scene are audible for a reason. While existing models tend to categorise sound into foreground and background types based on either proximity or salience (which in evolutionary terms are probably highly correlated) these models are arrived at using natural scenes where sounds have a cause, but not all have a function.

The work in the "Pinocchio" study in chapter 4 breaks down a scene into its constituent objects and asks the production team to explain the reason for the inclusion of each of these objects. Different functions align with the foreground or background classifications. Foreground sounds being audio objects that convey plot events or emotion. Background

sounds are still important, but will convey context or emotion rather than story events. Sounds that signify a plot or story event are not limited to speech, but include many sound effects. It is interesting to note this discrimination of foreground and background is based on the intended function of the audio objects, rather than the type of sound, for example speech vs sound effects. This categorisation works at a perceptual rather than a physical level. Many of the listening models covered in Chapter 2 revolve around three modes of listening: focusing on the source, the meaning or physical characteristics of the sound. While most of these listening models are arrived at and justified using a theoretical approach the three categories identified by producer and sounds designer in chapter 4 were a result of empirical exercises.

One categorisation of the music suggested by the engineer was 'human stuff' and 'mechanical stuff'. This is reminiscent of results from soundscape research which reports listeners grouping sounds into the types 'natural', 'human' and 'mechanical' [58]. The soundscape research was motivated by assessing listener preference of natural (or modelled natural) soundscapes, whereas this research is about categorisation of objects within a constructed sound scene. Of course, unlike the drama production with used sound effects and foley recordings, the music productions used in this research contained no 'natural' sounds.

These representations of sound are very much driven by the producer's viewpoint. This differs to existing listening models. Existing models can be applied to the listening mode of the content producers: during the production they were highly engaged in focused listening "Top Down". However, audience modes of listening were not always such. The nature of the experiment in the foreground vs background mix resulted in subjects listening "in search" due to the task required of the subject and the unusual laboratory-like location. However, while the listeners to "Breaking Out" knew they were experiencing an experimental work, the focus groups which followed revealed that listeners often had the radio on in while their attention was elsewhere; Truax's "background listening", "holistic listening" or listening "bottom up". The ultimate aim of the Producer is to create a piece of engaging content, and therefore to shift listeners from "background listening" through "listening in-readiness" to listening "in-search".

Results from the football study and "Breaking Out" location based drama study demonstrate that the capabilities enabled by objects can make more engaging content. Results from from the foreground vs background study show the distribution of foreground vs background preferences are not closely clustered or symmetrically distributed around the correct broadcast mix. This also suggests audience engagement with audio content could be improved by using foreground and background audio objects to deliver a personalised balance to suit their preferred foreground vs background audio mix.

6.1.3 Working Towards a More Than a Single Vision

Perhaps the biggest impact on production using audio objects is the need for content creators to design content for a range of experiences rather than single fixed version. The "Pinocchio" study in chapter 4 illustrates how the content creators began to trust the technology to deal with down-mixing by monitoring down-mixes more as the production process went on. The "Breaking Out" study in chapter 5 shows the the writer, who would traditionally create a static story actually enjoyed the process of working with constraints and incorporating variables into the story. This is a change in mindset required from traditional content creators is highly analogous to the effect of responsive web design [106] on traditional web-designers.

> "The control which designers know in the print medium, and often desire in the web medium, is simply a function of the limitation of the printed page. We should embrace the fact that the web doesn't have the same constraints, and design for this flexibility. But first, we must 'accept the ebb and flow of things'."
>
> John Allsopp, "A Dao of Web Design" [107].

The processes of content creation for "Pinocchio" described in chapter 4 and "Breaking Out" in chapter 5 were clearly iterative and additive, starting with a blank DAW session and building the content by recording or sourcing and processing audio. Interviews with the Producer and Sound Designer show there is a clear intention and vision upon which decisions made during the process are based. The creators have a vision of how the audio should sound, and they follow the process to get as close to that vision as possible within the time available. One of the key benefits identified in the literature of object-based audio is the ability to abstract the production process from the playback environment. This allows content creators to design and create the content once and distribute it across a multitude of different channels and consumption devices. As a result of this executing a creator's vision now needs to result in a range of experiences.

A consistent experience cannot be achieved across different reproduction systems or in different contexts. Broadcast audio content is not heard in the same environment by the whole audience. Some of the audience are likely to be listening in far better quality listening conditions, using higher quality listening equipment. The quality of experience of the audience will vary from listener to listener. For example the quality of experience of a listener in a high quality listening room is not comparable to a listener driving a noisy car

for obvious physical and perceptual reasons. This is not new news for content producers, and does not represent a major shift in reality (even if it does require a shift in creator's mindset). Engineers already have had to deal with creating content for different systems. Additional loudspeakers added to enable 9.0 surround mixes to be played back in 4 adds another dimension to the variety of quality of experiences possible, but even this does not cause a paradigm shift.

A common approach taken by sound engineers when mixing for a range of listening environments and systems is to monitor the mix on a second, low budget set of loudspeakers to check the audio mix is satisfactory on those loudspeakers. During the "Pinocchio" production in chapter 3, the monitoring environment was capable of playing back mono (1.0), stereo (2.0), surround (5.0), surround with height (9.0) and wavefield synthesis (24.2). By the end of the mixing process the Sound Designer was only occasionally using the parallel rendering to check the stereo (2.0) reproduction in a similar way as mastering engineers use smaller loudspeakers to quality check their mastering. Some content creators might perceive the need to create content for a range of experiences a creatively threatening scenario.

However, the Sound Designer's behaviour observed in the "Pinocchio" study was to check the rendering to fewer loudspeakers less and less as the production process continued. Ultimately trusting the IOSONO system to render to mono (1.0), stereo (2.0) and surround (5.0) without regularly checking the those mixes, choosing instead to focus on the surround sound with height mix. This is even more significant because the production was due to be broadcast on Radio 4, who were only interested in the two channel stereo mix, therefore it was crucial that the stereo mix was correct. From this monitoring behaviour it is possible to predict that content creators will become more comfortable with defining a range of experiences as they become more confident in the technology's ability to maintain quality of the audience experience.

The process of defining a range of experiences is observed in different ways in the football study in chapter 3 and the "Breaking Out" location based study in chapter 5. The production team's vision for the football study was to allow the audience to determine where they sat in the football stadium. This could range from one end of the stadium to the other and could result in different levels of commentary compared to crowd noise. The content creator defined this range but they did not listen to every possible version of the mix, trusting the technology to constrain the experience within their defined range of experiences. In the production of "Breaking Out" the writer defined the variables which could change in response to the audience's location. The writer did not then listen to every possible version of the play. To do this would have been impractical due to the number of possible versions. Both of these examples demonstrate that it is practical for a content

creator to envisage and define a range of audience experiences.

6.1.4 Constraining the Range of Intended Experiences

Foreground and background objects are both important. Using foreground and background classification based on function, as established in chapter 4, to remove all the background audio objects would result in a story that conveyed plot events without any contextual information. This would probably have a detrimental effect on the audience experience. If a content producer wanted to allow the control of foreground vs background balance and they wanted to maintain the quality of experience, the content producer would need to set limits for the amount of personalisation allowable. Section 5.5 describes how these constraints were manifest in the perceptive media study in chapter 5. These constraints were dictated by the Producer's vision of the range of experiences which itself was constrained by the capabilities of the text to speech technology. These constraints define a range of experiences that are not as open ended as a typical computer game, but this is a major shift in creative process for traditional linear storytellers.

Section 4.5.4 concludes that it is harder to group audio objects of non-speech content than speech content because it is seen as a 'piece of art'. In music broadcasting, the artist normally works relatively independently of the sound engineer. The artist will perform the music and the engineer's job is to balance that music to a stereo mix. Musicians tend to vary in how much contact they have with the sound engineers, some are very hands on while others barely say a word to the engineer. Current music production is geared around the distribution of stereo content (compact disc and stream/download). While there is a layer of abstraction between broadcaster and musician (the broadcaster treating the music as a immutable 'piece of art'), and music distribution channels are primarily stereo, it is unlikely object-based music will be broadcast until an audience benefit and demand has been more clearly demonstrated.

6.2 Experience

6.2.1 Foreground vs Background Preferences

Foreground and background preferences were not influenced by loudspeaker layout. This is discussed is section 4.8 in the foreground and background study. Although these preferences were not influenced by loudspeaker layout, the range of preferences expressed by different

subjects suggest that audiences would benefit from treating foreground and background sound as independent audio objects. While peer reviewed literature available for foreground vs background mix preferences of constructed audio content is limited for normal hearing there is a study which specifically considered listeners with hearing impairments [61]. There is also a large body of journalistic material covering complaints from BBC audiences about background sound being too loud in programmes [108] [109] [110] [111]. This, coupled with the results in the foreground vs background study in chapter 4 provide a strong argument for using foreground and background audio objects for speech based audio content.

6.2.2 Personalised, Not Interactive

All three studies in chapters 3, 4 and 5 provide evidence to support an argument for the benefits of using object-based audio. The 5 Live Football study in chapter 3 and "Pinocchio" in chapter 4 have results based on audience preference. "Breaking Out" in chapter 5 assessed the impact of personalisation on audience experience. The football study and "Pinocchio", chapters 3 and 4 respectively, asked the audience to interact with the content in order to arrive at meaningful results. Application of the conclusions of the foreground vs background study in chapter 4, and to a lesser extent the football study in 3, suggest a benefit for audiences could be achieved delivering a personalised content experience, not an interactive one. The football study demonstrates that, in the context of live sport, audiences chose not to interact with the experience even though they were presented with a user interface that allowed interaction. Events of the football pitch did not result in more interaction. "Breaking Out" in chapter 5 did not provide the option to interact, there was a menu that allowed interaction but this was hidden from users. There is a debate within the broadcasting industry concerning audience demand for interactive experiences. One of the benefits of object-based content over channel-based content that is often cited during this debate is its ability to allow active interactivity instead of passive consumption. For example, the Spatial Audio Object Coding (SAOC) [17] specification identifies applications such as interactive re-mixing, rich media and gaming as benefits of using SAOC. These benefits apply to the concept of audio objects in general and are not limited to a specific format such as SAOC. The football study in chapter 3 provides some evidence relating to the audience's demand for interaction. 65% of the listeners experiencing the football match stopped interacting with the experience after 30s, and events occurring on the field did not result in additional audience interaction. This suggests a limited audience demand for the ability to interact with linear sports content, people preferring to set a preferences and then listen with no more interaction with the content.

6.2.3 Audience Awareness

It is harder to quantify the value of personalisation if the audience are unaware it is taking place. If the audience does not know it is happening (as may be the case with the personalised drama study covered in chapter 5) they are certainly not conscious of its value. However, evidence from all three studies shows personalisation can improve the audience experience, so while the listeners may not be aware of its value they certainly can benefit from the personalisation. During the experiment in chapter 5 focus groups revealed that listeners in London did not notice the personalisation as much as non-London based listeners. This result could be seen as evidence of the media's London-centric nature; it is likely that the London audience were desensitised to stories being set in the capital an therefore their experience may not have been improved as much as those not based in London. This was noticeable from the comments in the focus groups, but the difference in enjoyment of London and non-London based listeners from the online study was insignificant.

6.2.4 Filter Bubbles

There have been a number of studies which look at the effects of personalising internet searches and recommendation systems based in explicit and implicit information from the user. The term 'filter bubble' was used by Eli Pariser to describe cases when the type content audiences are exposed to shields them from viewpoints that contradict their own. Nguyen et al. [112] found that the personalisation of a recommendation system narrowed the diversity of the user's recommendations slightly over time. This literature is centred around recommendation systems rather than content personalisation, however it is easy to see how the conclusions made in the literature can be applied to content personalisation. "Breaking Out" in chapter 5 sets itself in the location of the listener, referring to local locations which 96% of listeners had heard of. This is the beginning of a filter bubble. Although a relatively new term, filter bubbles are not a new phenomenon; a reader might choose to read one particular newspaper which would confirm and reinforce their viewpoints, resulting in confirmation bias. Some of the results of "Breaking Out" can be related to the filter bubble theory. Focus group listeners in London were less aware of the location based personalisation than listeners outside London. This suggests that media listened to by the focus groups is London-centric, a criticism often levelled at the media. As well as enforcing filter bubbles, these bubbles could be burst by the type of location based personalisation used in "Breaking Out" by presenting them with unfamiliar locations. This effect of personalised media on filter bubbles is worthy of further investigation.

6.2.5 Personal Data

Results from the "Breaking Out" study reveal further questions concerning content personalisation. The ability to personalise content requires personal information. For personalisation of foreground vs background audio mixes content creators (or rendering systems) would need to be aware of the individual listener's foreground vs background level preference. For live sport events, such as the one studied in chapter 3, the team that the listener supported would also need to be known. The location based study from chapter 5 needed to register the listener's geolocation in order to set the drama correctly. The studies within this thesis demonstrate that personalisation can improve the audience experience. The more personal data that is known, the more can be done to personalise the content. Data and privacy is a big topic of contention [113], and while the studies within this thesis show audience experiences can be improved with insights from this data, consumers and audiences are becoming increasingly concerned about the privacy of their personal data. It is worth noting that the system design of "Breaking Out" in chapter 5 did not collect any location information, all the personalisation was preformed dynamically in the client browser on the listener's computer. The only location data gathered was from those listeners who agreed to complete the online survey and provided their location. This demonstrates it is possible to design personalised experiences that do not collect personal data.

6.2.6 Advertising

In 1993 the film "Demolition Man" [114] was released. In one of the scenes fast food chain "Taco Bell" paid for product placement for the US release. However, for the international release the references to "Taco Bell" (as US fast food chain) would have been missed by audiences so "Pizza Hut" was used instead. A second version of the scene was created, dubbing the reference to "Pizza Hut" and replacing the "Taco Bell" graphic [115]. This is a clear example of location influencing content for non-creative purposes. The second version of the scene in "Demolition Man" was created in post production by dubbing the dialogue, however object-based audio would have enabled the switching of this reference in real time to suit the individual listener. It is clear that this technology could have value to advertisers. Targeted advertising appears in many forms of digital media, by tracking user data and online behaviours. The use of personalised content and product placement is familiar to audiences. Comments from listeners collected as part of the "Breaking Out" audio drama study highlight that creators of object-based content need to be mindful of the blurred lines between personalised references and advertising and the impact this might have on the user.

6.3 A Missing Definition

In the course of creating the content for this study the need for an agreed definition of the term *Audio Object* has become apparent. This section aims to provide a definition. It reflects the evolution of how the term *Object* has been used in the audio research community, how it is now becoming understood in linear production of the type created for chapter 3 and chapter 4 and how this understanding can be applied to non-linear or personalised story telling of the type explored in chapter 5.

6.3.1 What is Not an Object?

Object-based audio is a concept that has been developed in response to a perceived need amongst content producers and audio technologists for a playback system agnostic content format. Problems with compatibility between surround sound, stereo and mono formats coupled with the diversification of listening systems have resulted in complicated production workflows to simultaneously deliver multiple formats, or a drop in quality of audience experience for people listening in non-ideal set ups. Object-based production abstracts the playback system from the production environments in order to maximise the quality of the audiences experience for any given playback system.

Early object-based production systems such as IOSONO's first production system [7] do not distinguish between an audio source and an audio object. In some scenarios this is might be a fair assumption, for example in the case of an interview where each person was using a separate microphone, this could be considered two audio objects: interviewer and interviewee. However, applying this to an orchestra that was captured using 100 microphones would result in 100 separate audio objects. It is hard to justify the need for all the data and bandwidth required to transport that amount data to the audience. Based on this, it seems incorrect to define an audio object as a source.

Current schemas and formats that support audio objects either define an audio object implicitly by the type of signal and metadata they carry or provide such flexibility that the need for a object definition is circumvented. MDA [6] and IOSONO's [7] streaming formats carry a single or set of audio signals accompanied by positional and level information. These data can be traced to the object as a source school of thought. The recent EBU BWAV standard [21] contains metadata fields with the flexibility to avoid the need for a semantic audio object definition. However, a framework that allows content producers to move from individual sources to discrete audio objects is required.

6.3.2 Towards a Audio Object Definition

To arrive at a sensible definition of an audio object it is important to consider the desired user experience. If the results of the football study in chapter 3 are taken at face value and the influence of the user interface is ignored there would be an argument for delivering the match as three mixes associated with the three peaks (a home mix, an away mix and a broadcast mix) rather than individual objects that can be mixed dynamically by the listener. If the desired user experience was to allow a listener to choose between three different mixes these three mixes would be all that need to be created. However, because the user experience was defined as one that the listener can dynamically mix home crowd, away crowd and commentary a more granular set of audio objects is required. In addition, the football case study had a user experience requirement of minimising the risk of audible bad language being broadcast but avoiding a lack of presence leading to difficulty in differentiating between stadium ends. This resulted in the design and creation of audio objects that were not made from single microphone signals so close to the crowd that individual voices were audible, neither were they a mix of voices that resulted in a more diffuse sound that would make it difficult to differentiate between home and away sides. A compromise defined by the desired user experience had to be reached. It can be understood from this example that the granularity required of the sound, and therefore the number and characteristic of audio objects, is defined by the user experience.

Taking the slightly more complex case of an orchestra which consists of 100 microphones feeds: It is clear that there is a process to go though to turn the 100 microphone feeds into a smaller set of audio objects. Like the football example, the exact nature of these audio objects should be defined by the desired user experience (or range of user experiences). For example, if the desired user experience of this orchestral recording was a mono audio experience played back though a single loudspeaker there would be no reason to have more than a single mono audio object to represent the whole orchestra. Anything more than this would be unnecessary. If the intended audience experience is a 22.2 surround sound audio mix it is likely that a larger number of audio objects would be needed. It is defining the range of experiences that will allow a production team to define and create a set of audio objects from a set of sources. A single audio object is a sound that is relatively immutable for the intended range of audience experiences. Levels and positions of sources within that audio object should not change relevant to each other. If the relative level and position of sources don't need to change in order to cater for the intended range of experiences then these sources can be considered the same audio object. Figure 6.1 shows this in a system diagram which emphasis how difficult it is to know an efficient number (and nature) of objects to create from a set of sources without first constraining the range of experiences as discussed in section 6.1.4.

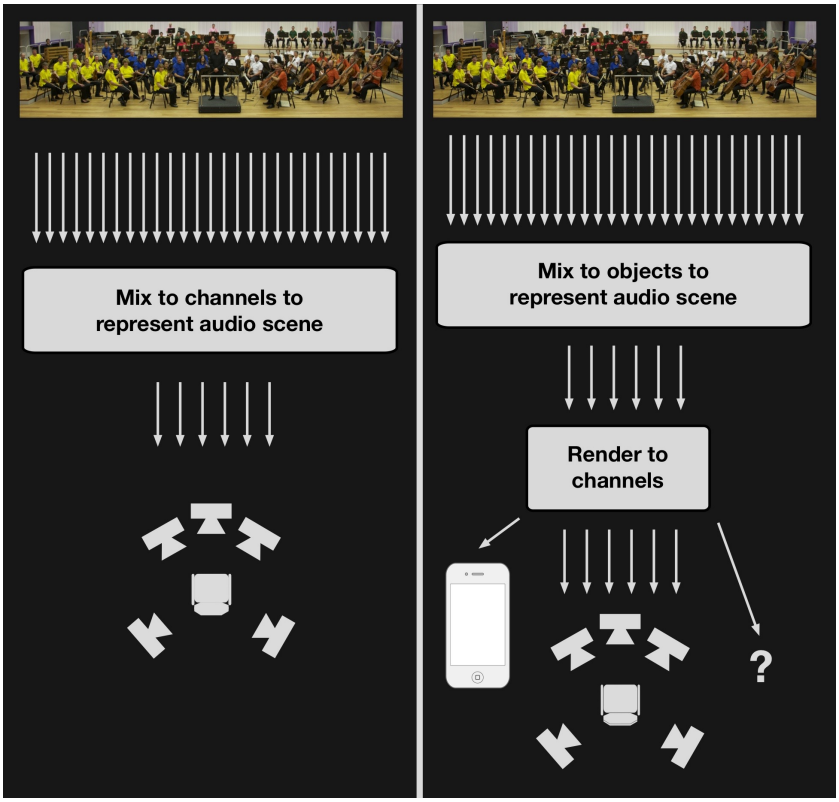


Figure 6.1: From sources to objects in the case of an orchestra.

6.3.3 Applying this to Non-linear Content

This definition is influenced by linear broadcasting scenarios explored by the football and "Pinocchio" in chapters 3 and 4 respectively, and the desire to provide high quality spatial audio that is not tied to a specific reproduction system. Things are different when it comes to non-linear content experiences such as the one investigated in 5, however the same logic can be successfully applied to non-linear audio. Rather than defining objects based on constant relative level and position in space, objects can have the additional requirement of having consistent relative position in time. Therefore if the relative temporal position (in addition to levels and spatial position) of sources does not need to change in order to cater for the intended range of user experiences then these sources can be considered the same audio object. Adding the variable of temporal position is novel because audio objects have previously been about spatial audio and dealing with different reproduction systems, not non-linear content mutable in the time domain. For reasons of efficiency, when breaking down an audio scene into audio objects, objects should be made as large as they can be, and as small as they need to allow the intended audience experience. Discussion of the work in this thesis results in the audio object definition below.

A single immutable chunk of data with associated metadata that allows it to be combined with other objects to create a range of experiences within the constraints defined by a content designer.

This definition fundamentally differs from that of an acoustic source or an auditory object. An acoustic source exists in the physical world and refers to the cause of a vibration in the audible spectrum. An audio object contains audio data that could have originated from a single acoustic source, a group of acoustic sources or the influence of an environment, depending on the intended listener experience as defined by the content designers. An auditory object exists only in the mind of a listener and may or may not relate to an acoustic source. Unlike an auditory object an audio object is defined by the content producer, not the listener. Audio objects are more akin to Lego®bricks which can be used to assemble an experience, but may not be perceived by the audience as separate objects.

7

Conclusions

This thesis tackles object-based audio in a professional broadcasting environment, a subject into which there is limited research. The work takes a user (audience and content creator) centric approach, but is differentiated from much of the literature in the field of object-based content because it focuses on applications beyond spacial audio.

The first study uses a live sports event as a platform to explore how the audience use, and whether or not they benefit from an interactive experience enabled by the use of audio objects as an alternative to a traditional channel-based linear broadcast experience. The study uses test material that allows listeners to a football match to control the mix between home and away fans and control the balance between the crowd and the commentary levels, dynamically during the live event. This study found that listeners chose not to interact with the experience beyond setting their preferred balance, treating it as personalised experience rather than an interactive experience. Even significant events on the pitch during the match did not result in peaks in interaction. This study also revealed that the choice of commentary vs crowd mix was distributed across the range the user interface allowed. Across the range of possible mixes there were three clear peaks where audiences set their preferred balance. These three peaks matched the user interface limiteds of maximum commentary, maximum crowd and equal commentary and crowd. Despite these three peaks being influenced by the user interface design there were still 66% of listeners who set a balance outside of these peaks. The football study also describes the production and distribution challenges encountered and solutions developed. The majority of the challenges being infrastructural, for example the stadium shape, design and infrastructure limiting the number of microphones and their positions. The editorial requirement to disguise any bad language balanced with the need to audibly differentiate between stadium ends resulted in a compromise between diffuseness and clarity of audio object.

The second study extends the first study to a number of different genres, analysing the production process and in doing so understanding the content Producer's view of audio objects. These audio objects are paired down into a diminishing number of categories using card sorting exercises and interviews with the Producers, Engineers and Sound Designers of the content. This allowed the construction of a new listening model appropriate for constructed broadcast content which groups sound into categories based on the Producer's intended function of the audio object. This new model addresses limitations in existing listening models which were developed to model listener responses to natural and managed soundscapes rather than constructed and designed soundscapes. The card sorting exercises, interviews and analysis of the production process demonstrated that broadcast sounds are highly constructed, and the presence of each audio object is carefully considered by the Producers based on their intended function of the audio object. Through the card sorting exercises and interviews which took place during the "Pinocchio" study a set of these functions were identified as representing actions or events, conveying emotion or providing contextual information. These functions were then linked to foreground and background categories. Content examples from six different genres were created with foreground and background as separate audio objects, based on the function of the sounds. This clips were used in an experiment to understand how the audiences' preference for different foreground and background mixes varies across the population, and between loudspeaker layouts. The study did not identify a significant effect of loudspeaker layout on listener's preferred foreground vs background mix, however it did discover listener's preference of foreground vs background preference is not normally distrusted around the broadcast mix, but is skewed towards preferring more background sounds. It also found that the majority of listeners preferred a mix different to that which would have been broadcast. These results are reflective of the results in the football study.

The final study builds on the first two. Based on results from the football study that showed listeners did not interact with the audio content, but benefited from experiencing something different to the broadcast mix, a content experience was created that used audio objects to dynamically change variables within a story. This work includes an analysis of the design and creation of an audio drama that dynamically locates the story at the listener's location. This content was used as test material to understand the impact of using audio objects on both content creators and their workflow, and listeners and their experience. This study found that personalising content based on location improved the audiences' enjoyment of the content. Listeners to the personalised radio drama reported a better overall experience and enjoyment of "Breaking Out", they felt closer to the location of the storyline, and liked the overall experience more than listeners to a generic version. The result from the "Breaking Out" study also highlighted some of the potential societal and commercial implications of experiences enabled by using object-based audio.

Reenforcing filter bubbles and the potential use of this technology by advertisers to more effectively promote products were identified as areas of concern if adopting an object-based audio approach to creating personalised audio content. In addition, this study revealed that personalised content which enhances the audience experience of audio content also raises concerns around data privacy because personalisation has to be driven by personal data.

The analysis of the production of object-based audio content from each of the studies in this thesis has made it clear that content designers should move from executing a single vision of their content experience towards envisioning a range of experiences. Their role as content producers should be to define and constrain a range of experiences. This was put in to practice with all three studies: the live football experiment where BBC 5 Live constrained the range of levels the audio could be set to (not allowing 100% commentary or 100% crowd mixes); the foreground vs background experiment where the creation of the foreground and background categorisation was defined by the content producers; and the location based drama where the dynamic variables within the play were defined by the writer.

Another key conclusion resulting from the analysis of the three studies is that audio objects are not equivalent to audio sources and should be defined by more than a spatial position. The discussion in chapter 6 provides a framework for defining an audio object. This framework for object definition is not founded on physical aspects of a sound scene (for example from where the sound objects are coming), but begins with the vision of the range of experiences intended by the content creator. There may be instances where this framework for object definition falls down, for example more esoteric art forms, but the framework can successfully be applied to audio content experiences which have been designed; where there is a clear intention behind the production and storytelling.

7.1 Further Work

There are a number of specific surprising results which warrant further investigation in the future. One of the key findings of this research is a new listening model to describe and classify audio objects within a piece of created audio content in the context of professional broadcasting. This differs from existing soundscape models of listening and sound classification. Further work to unify these different listening models would be beneficial. Perhaps using semi-constructed content as test material, for example highly observational documentary. The foreground vs background experiment in chapter 4 also identified younger subjects as preferring louder foreground sounds. Further investigation of the link between

age and foreground vs background mix preference would help to explain this result. This research found that there was more consensus amongst listeners about the foreground vs background mix preference when using 5.0 surround sound compared to other loudspeaker layouts, a further investigation would help explain this result.

This thesis has explored only a limited number of applications of object-based audio. There are almost limitless applications of audio objects, beyond personalising foreground and background balance and story elements relative to geographic location. An exploration of the benefits of using audio objects beyond using foreground vs background mix, and location to personalise content would be useful further work. For example a study to understand the impact of using local accents, in addition to local references on listener enjoyment and engagement.

Discussions in chapter 6 consider the impact on production workflow when creating object-based audio content that can result in a range of audience experiences. Further ethnographic research is needed to understand how Producers and storytellers can adapt their existing workflows to allow them to deliver personalised experiences and define and constrain a range of audience experiences. This thesis also makes way for a further investigation in the social science field to understand the implications of delivering highly personalised content experiences instead of static broadcast content. This would help explore such phenomena as filter bubbles and the audiences' attitude to personal data collection and use as discussed in section 6.2.

References

[1] A Churnside, S Spors, F Melchior, "Emerging Technology Trends in Spatial Audio", SMPTE Mot. Imag. J; September 2012; 121:(6) 95-100.

[2] A Churnside, M Mann, A Bonney, F Melchior "Object-Based Audio Applied to Football Broadcasts: The 5 live Football Experiment," ACM international workshop on Immersive media experiences, 2013.

[3] A Churnside, I Forrester, "The Creation of a Perceptive Audio Drama," Paper presented at the NEM Summit, October 2012.

[4] A Churnside, M Shotton, F Melchior, M Evans, M Brooks, M Armstrong, "Object-based Broadcasting - Curation, Responsiveness and User Experience," International Broadcasting Convention, page 12.2, 2014.

[5] Dolby Laboratories Inc., "Dolby ATMOS", Dolby white paper, 2012.

[6] SRS Labs, MDA http://www.srslabs.com/landing.aspx?id=2459, accessed 30 June 2015.

[7] F Melchoir, U Heusinger, J Leibetrau, "Perceptual evaluation of a spatial audio algorithm based on wave field synthesis using a reduced number of loudspeakers", Audio Engineering Society Convention Paper 8575, Presented at the 131st Convention 2011 October 2023 New York, USA.

[8] K Hamasaki, S Komiyama, K Hiyama, H Okubo, "5.1 and 22.2 Multichannel Sound Productions Using an Integrated Surround Sound Panning System-, NAB Broadcast Engineering Conference, April 2005, Las Vegas.

[9] ITU-R BS 775-3, "Multichannel stereophonic sound system with and without accompanying picture", International Telecommunications Union, Geneva, Switzerland, 1992-1994-2006-2012.

[10] D Malham, "Spherical Harmonic Coding of Sound Objects - the Ambisonic 'O' Format", 19th International Conference: Surround Sound - Techniques, Technology, and Perception, June 2001.

[11] J Daniel, R Nicol, S Moreau, "Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging", 114th AES Convention, 2003, Amsterdam, The Netherlands.

[12] "Virtual Reality Modeling Language (VRML)" ISO 14472-1:1997, 1997.

[13] E D Scheirer, R Vanaanen, J Huopaniemi, "AudioBIFS: Describing Audio Scenes with the MPEG-4 Multimedia Standard", IEEE Transactions on 1 (3), 237-250, Sept 1999.

[14] The Carrouso Project, http://emt.emt.iis.fhg.de/projects/carrouso/ accessed 30 June 2015.

[15] R Bleidt, "Introducing the MPEG-H Audio Alliance's New Interactive and Immersive TV Audio System", HPA Tech Retreat, Palm Springs, February, 2015.

[16] R G Oldfield, B G Shirley, "Format agnostic recording and spatial audio reproduction of football broadcasts for the television', Institute of Acoustics Reproduced Sound Conference, Brighton, UK, November 16th, 2011.

[17] J Breebaart, J Engdegård, C Falch, O Hellmuth, J Hilpert, A Hoelzer, J Koppens, W Oomen, B Resch, E Schuijers, L Terentiev, "Spatial Audio Object Coding (SAOC) - The Upcoming MPEG Standard on Parametric Object Based Audio Coding", 124th AES Convention, May 2008, Amsterdam, The Netherlands.

[18] J Breebaart, S Disch, C Faller, J Herre, G Hotho, K Kjörling, F Myburg, M Neusinger, W Oomen, H Purnhagen, J Rödén, "MPEG Spatial Audio Coding / MPEG Surround: Overview and Current Status", Philips Research Laboratories, New York, 2005.

[19] SMIL, latest specification, http://www.w3.org/TR/REC-smil/ accessed 30 June 2015.

[20] M Geier, J Ahrens, and S Spors. "ASDF: Ein XML Format zur Beschreibung von virtuellen 3D-Audioszenen." 34th German Annual Conference on Acoustics (DAGA), Dresden, Germany, March 2008.

[21] "Audio Definition Model" Tech 3364, EBU, 2014.

[22] V Pulkki, "Virtual Sound Source Positioning Using Vector Base Amplitude Panning", JAES Volume 45 Issue 6 pp. 456-466; June 1997.

[23] C Tsakostas, Floros, Y Deliyiannis, "Real-time Spatial Mixing Using Binaural Processing", 4th Sound and Music Computing Conference, July 2007, Greece.

[24] D Griesinger, "Stereo and Surround Panning in Practice", Audio Engineering Society Convention Paper 5564, Presented at the 112th Convention 2002 May 1013 Munich, Germany.

[25] A D Blumlein, U.K. patent 394,325, 1931.

[26] M A Gerzon, "Panpot Laws for Multispeaker Stereo", Audio Engineering Society, Vienna, March 1992.

[27] A J Berkhout, "A Holographic Approach to Acoustic Control", J. Audio Eng. Soc., pp. 977-995, 1988.

[28] J Ahrens, M Geier, S Spors, "Introduction to the SoundScape Renderer (SSR)", Quality and Usability Lab Deutsche Telekom Laboratories Technische Universitt Berlin, November 13, 2012.

[29] T Crook, "Radio Drama: Theory and Practice", Psychology Press, 1999.

[30] The S3A Project, http://www.s3a-spatialaudio.org/wordpress/ accessed 9 March 2016.

[31] P Comon, C Jutten, "Handbook of Blind Source Separation: Independent Component Analysis and Applications", Academic Press, 2010.

[32] T Zhang, J Kuo, "Audio content analysis for online audiovisual data segmentation and classification", Trans. Speech Audio Processing, vol. 9, pp. 441-457, 2001.

[33] J O Smith, Physical Audio Signal Processing, http://ccrma.stanford.edu/~jos/pasp/, online book, accessed 30 June 2015.

[34] C L Christensen, J H Rindel, "A new scattering method that combines roughness and diffraction effects", Forum Acusticum, Budapest, 2005.

[35] Spatialisateur User Manual, IRCAM, October 2012 http://forumnet.ircam.fr/wp-content/uploads/2012/10/Spat4-UserManual1.pdf accessed November 2013.

[36] J Friberg, D Gärdenfors, "Audio games: new perspectives on game audio", ACE '04 Proceedings of the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology, New York, USA.

[37] A Brandon, "Audio for Games: Planning, Process, and Production", New Riders, 2004

[38] Audiokinetic Wwise Overview, Audiokinetic, https://www.youtube.com/watch?v=EjZKqDF3F3k, accessed 30 June 2015.

[39] C Rogers, Web Audio API, http://www.w3.org/TR/webaudio/, 2013, accessed 30 June 2015.

[40] Marketing Week, http://www.marketingweek.co.uk/sectors/media/radio/absolute-radio-offers-location-based-personalised-ads/3027824.article, accessed 30 June 2015.

[41] Capital Radio, http://www.capitalxtra.com/news/about-my-capital-xtra/, accessed 30 Nov 2015.

[42] NetMix experiment, http://www.bbc.co.uk/blogs/5live/2011/06/netmix.shtml accessed 30 June 2015.

[43] H Fuchs, O Hellmuth, S Meltzer, F Ridderbusch, S Tuff, "Dialog enhancement: Enabling user interactivity with audio," Paper presented at the NAB, 2012.

[44] J K Bizley, Y. E. Cohen, "The what, where and how of auditory-object perception", Nature reviews Neuroscience, Oct, 2013.

[45] M Cooke, "Modelling Auditory Processing and Organisation", Cambridge University Press, September, 1993.

[46] Davies, W. J. (2015). "Cognition of soundscapes and other complex acoustic scenes", Internoise 2015. San Francisco.

[47] Davies, W. J., N. S. Bruce and J. E. Murphy (2014). "Soundscape reproduction and synthesis" Acta Acustica United with Acustica 100(2): 285-292.

[48] Griffiths, T. D. and J. D. Warren (2004). "What is an auditory object?" Nature Reviews Neuroscience 5(11): 887-892.

[49] H Scheich, F Baumgart, B Gaschler-Markefski, C Tegeler, C Tempelmann, H J Heinze, F Schindler, D Stiller, "Functional magnetic resonance imaging of a human auditory cortex area involved in foreground-background decomposition.", European Journal of Neuroscience, Volume 10, Issue 2, pages 803-809, February 1998.

[50] J Blauert, "Spatial Hearing: The Psychophysics of Human Sound Localization" .MIT Press, 1983.

[51] J P L Brokx, S G Noteboom, "Intonation and the perceptual separation of simultaneous voices", Journal of Phonetics, 10:23-36, 1982.

[52] A S Bregman, "Auditory Scene Analysis: The Perceptual Organization of Sound", MIT Press, 1990.

[53] W Noble, S Perrett, "Hearing speech against spatially separate competing speech versus competing noise", Perception and Psychophysics, Volume 64, Issue 8, pp 1325-1336, 2002.

[54] G Humphrey, "The psychology of the Gestalt." Journal of Educational Psychology, 15(7), 401-412. 1924.

[55] R van Tol, S Huiberts, "IEZA: A Framework For Game Audio", 2008, http://www.gamasutra.com/view/feature/3509/ieza_a_framework_for_game_audio.php accessed 30 June 2015.

[56] W J Davies, M D Adams, N S Bruce, R Cain, A Carlyle, P Cusack, D A Hall, K I Hume, A Irwin, P Jennings, M Marselle, C J Plack, J Poxon, "Perception of soundscapes: An interdisciplinary approach", Applied Acoustics, 2013.

[57] B Schulte-Fortkamp, A Fiebig, "Soundscape analysis in a residential area: An evaluation of noise and people's mind", Acta Acustica united with Acustica, 2006.

[58] S R Payne, W J Davies, M D Adams, "Research into the Practical and Policy Applications of Soundscape Concepts and Techniques in Urban Areas." Department for Environment, Food and Rural Affairs: London, 2009.

[59] R. Schafer. "The Soundscape: Our Sonic Environment and the Tuning of the World", Destiny Books, 1977.

[60] P Menneer, "VLV Audibility of Speech on Television Project", VLV, 2011.

[61] B G Shirley, "Improving Television Sound for People with Hearing Impairments", University of Salford, 2013.

[62] B Raj, P Smaragdis, M Shashanka, R.Singh, "Separating a foreground singer from background music", International Symposium on Frontiers of Research on Speech and Music, Mysore, India, 2007.

[63] F Rumsey, "Spatial quality evaluation for reproduced sound: Terminology, meaning , and a scene-based paradigm", J. Audio Eng. Soc. 50: 651-666, 2002.

[64] O Julien "The diverting of musical technology by rock musicians: the example of double-tracking", Popular Music, 18, pp 357-365, 1999.

[65] P Jackson, M Dewhirst, R Conetta, "QESTRAL (Part 3): System and metrics for spatial quality prediction", AES 125th Conv. San Francisco, 2008.

[66] P Schaeffer, "Traité des objets musicaux" Paris: Editions du Seuil, 1966.

[67] M Chion, "Audio-Vision: Sound on Screen", New York: Columbia University Press, 1990.

[68] D Huron, "A six-component theory of auditory-evoked emotion", Proceedings of ICMPC7, 673-676, 2002.

[69] K Tuuri, M Mustonen, A Pirhonen, "Same sound - different meanings: A novel scheme for modes of listening", Proceedings of Audio Mostly (pp. 13-18), Ilmenau, Germany: Fraunhofer Institute for Digital Media Technology IDMT, 2007.

[70] B Truax, "Acoustic communication (2nd ed.)", Ablex Publishing, 2001.

[71] M. Thorogood, J. Fan, P Pasquier, "BF-Classifier: Background/Foreground Classification and Segmentation of Soundscape Recordings", Proceedings of the Audio Mostly 2015 on Interaction With Sound, p.1-6, Thessaloniki, Greece, 2015.

[72] W.W. Gaver,"What in the world do we hear? An ecological approach to auditory scene perception", Ecological Psychology 5, 1-29, 1993.

[73] M. Raimbault, "Qualitative Judgements of Urban Soundscapes: Questionning Questionnaires and Semantic Scales", Acta Acustica united with Acustica 92, 929-937, 2006.

[74] ITU-R BS.1534, "Method for the subjective assessment of intermediate quality levels of coding systems," International Telecommunications Union, Geneva, Switzerland, 1997.

[75] ITU-R BS.1116-1, "Methods for the Subjective Assessment of Small Impairments in Audio Systems In- cluding Multichannel Sound Systems," International Telecommunications Union, Geneva, Switzerland, 1997.

[76] A Mellouk, H A Tran, S Hoceini, Quality-of-Experience for Multimedia, ISBN: 978-1-84821-563-4, Wiley-ISTE, October 2013.

[77] P Le Callet, S Maller, A Perkis, "Qualinet White Paper on Definitions of Quality of Experience", European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), Lausanne, Switzerland, Version 1.2, March 2013.

[78] J P Davis, K Steury, R Pagulayan, "A survey method for the assessing perceptions of a game: The consumer play test in game design", The International journal of computer games research, Vol 5, October 2005.

[79] R Busselle, H Bilandzic, "Measuring Narrative Engagement", Media Psychology, 2009.

[80] B. Abelson, "Zombies, brains, and tweets The neural and emotional correlates of social media", Harmony Institute, September 2013.

[81] R L Mandryk, K M Inkpen, T W Calvert, "Using psychophysiological techniques to measure user experience with entertainment technologies", Journal of Behaviour and Information Technology, 2005.

[82] J Peacock, S Purvis, R L Hazlett, "Which Broadcast Medium Better Drives Engagement? Measuring the Powers of Radio and Television with Electromyography and Skin-Conductance Measurements", Journal of Advertising Research, Vol. 51, No. 4, 2011.

[83] C Kyriakais, "Fundamental and Technological Limitations of Immersive Audio Systems", Proceedings of the IEEE, VOL. 86, NO. 5, 1998.

[84] K Pihkala, "Extending SMIL with 3D audio", HUT, Telecommunications Software and Multimedia Laboratory

[85] J D Reichbach, R A Kemmerer, "SoundWorks: an object-oriented distributed system for digital sound", Computer, March 1992.

[86] W Hudson, "The Encyclopedia of Human-Computer Interaction, 2nd Ed.", The Interaction Design Foundation, Denmark, 2014.

[87] M C Freyaldenhoven, D Fisher Smiley, R A Muenchen, T N Konrad, "Acceptable Noise Level: Reliability Measures and Comparison to Preference for Background Sounds", Journal of the American Academy of Audiology, Volume 17, Number 9, pp. 640-648(9), October 2006

[88] 5.1-Channel Production Guidelines, Dolby Laboratories Inc.

[89] C Guastavino, B F G Katz, J D Polack, D J Levitin, D Dubois: Ecological validity of soundscape reproduction. Acta Acustica united with Acustica 91 (2005) 333-341.

[90] M Swaguchi,"Practical Surround Sound Production Part-1: Radio Drama," AES, 2001.

[91] J McCarthy, J Stewart, "Producing Radio Drama Using Networked Digital Audio Workstations," AES, 1995.

[92] Perceptive Media, http://thenextweb.com/media/2012/02/08/the-bbc-is-experimenting-with-perceptive accessed 30 June 2015.

[93] Context Aware Stories, http://connectedsocialmedia.com/6359/future-lab-context-aware/ accessed 30 June 2015.

[94] eSpeak, http://espeak.sourceforge.net/ accessed 30 June 2015.

[95] S Lemmetty, "Review of Speech Synthesis Technology," Helsinki University, 1999.

[96] Weather Source, http://open.live.bbc.co.uk/weather/feeds/en/ accessed 30 June 2015.

[97] News podcast source, http://downloads.bbc.co.uk/podcasts/radio/newspod/rss.xml accessed 30 June 2015.

[98] W3C (HTML5 audio group), http://www.w3.org/ accessed 30 June 2015.

[99] "The Archers", [radio soap], created by the BBC, UK, 1950 - present.

[100] D Lee, "BBC shows off 'perceptive radio' that can alter scripts" http://alistapart.com/article/responsive-web-design accessed 30 June 2015.

[101] M Bryant, "The BBC unveils its first 'Perceptive Media' experiment, and you can try it now" http://thenextweb.com/media/2012/07/03/the-bbc-unveils-its-first-perceptive-media-experiment-and-you-can-try-it-now/ accessed 30 June 2015.

[102] S Gibbs, "The BBC Opens Up Its First "Perceptive Media" Experiment and You Can Try It Out Right Now" http://www.gizmodo.co.uk/2012/07/the-bbc-opens-up-its-first-perceptive-media-experiment-and-you-can-try-it-out-right-now/ accessed 30 June 2015.

[103] D Crow, P Pan, L Kam, G Davenport, "M-views: A system for location based storytelling", ACM UbiComp, Seattle, WA, 2003.

[104] J Paay, J Kjeldskov, A Christensen, A Ibsen, D Jensen, G Nielsen, "Location-based storytelling in the urban environment", Proceedings of the 20th Australasian conference on computer-human interaction: Designing for habitus and habitat, ACM New York, NY, USA, 2008.

[105] M Marszalek, C Schmid, "Semantic Hierarchies for Visual Object Recognition", CVPR 2007, IEEE Conference on Computer Vision & Pattern Recognition, Minneapolis, United States, IEEE Computer Society, pp.1-7, 2007.

[106] E Marcotte, 2010. "Responsive web design. A List Apart." http://alistapart.com/article/responsive-web-design accessed 30 June 2015.

[107] J Allsopp, "A Dao of Web Design", Published in CSS, Layout & Grids, Typography & Web Fonts, Accessibility, 2000 http://alistapart.com/article/dao accessed 30 June 2015.

[108] H Furness, "Broadchurch viewers switch on subtitles to catch plot amid complaints of mumbling", Daily Telegraph, 2015, http://www.telegraph.co.uk/news/celebritynews/11327873/Broadchurch-viewers-switch-on-subtitles-to-catch-plot-amid-complaints-of-mumbling.html accessed 30 June 2015.

[109] "Jamaica Inn: BBC receives more than 100 complaints over 'mumbling"', Guardian, 2015, http://www.theguardian.com/media/2014/apr/22/jamaica-inn-bbc-mumbling-sound-levels accessed 30 June 2015.

[110] D Walker, "Quirke 'mumbling': BBC "looking into" complaints made about sound quality of Gabriel Byrne crime drama", Mirror, 2014, http://alistapart.com/article/responsive-web-design accessed 30 June 2015.

[111] O Gillman, "BBC faces deluge of complaints from listeners over bizarre internet coverage of fictional flood disaster in the Archers", Mail Online, 2015, http://www.dailymail.co.uk/news/article-2982334/BBC-faces-deluge-complaints-listeners-bizarre-internet-coverage-fictional-flood-disaster-Archers.html accessed 30 June 2015.

[112] T T Nguyen, P-M Hui, F Maxwell Harper, L G Terveen, J A Konstan, "Exploring the Filter Bubble: The Effect of Using Recommender Systems on Content Diversity," University of Minnesota, 2014.

[113] T R Graeff, S Harmon, "Collecting and using personal data: consumers' awareness and concerns," Journal of Consumer Marketing, 2002.

[114] "Demolition Man," [film], Directed by Marco Brambilla, Warner Bros, USA, 1993.

[115] "Demolition Man," [clip], Directed by Marco Brambilla, Warner Bros, USA, 1993. https://www.youtube.com/watch?v=gpRzusd9Yi8 accessed 30 June 2015.

Appendices

Appendix A

Appendices Relating to the Football Study in Chapter 3

A.1 Questions Used for Online Study

Included in the following pages are questions written by the thesis author to user understand the audiences' perception and reception of the 5 Live Football Experiment.

5 Live Football Experiment

This is a short survey to find out what you thought of the 5 Live football experiment.

0%
100%

Football Questions

* **The application gave you the ability to choose from which end of the stadium you heard the match. Please rate how this ability impacted upon your overall experience of match.**
Choose one of the following answers

- Much better
- Slightly better
- About the same
- Slightly worse
- Much worse

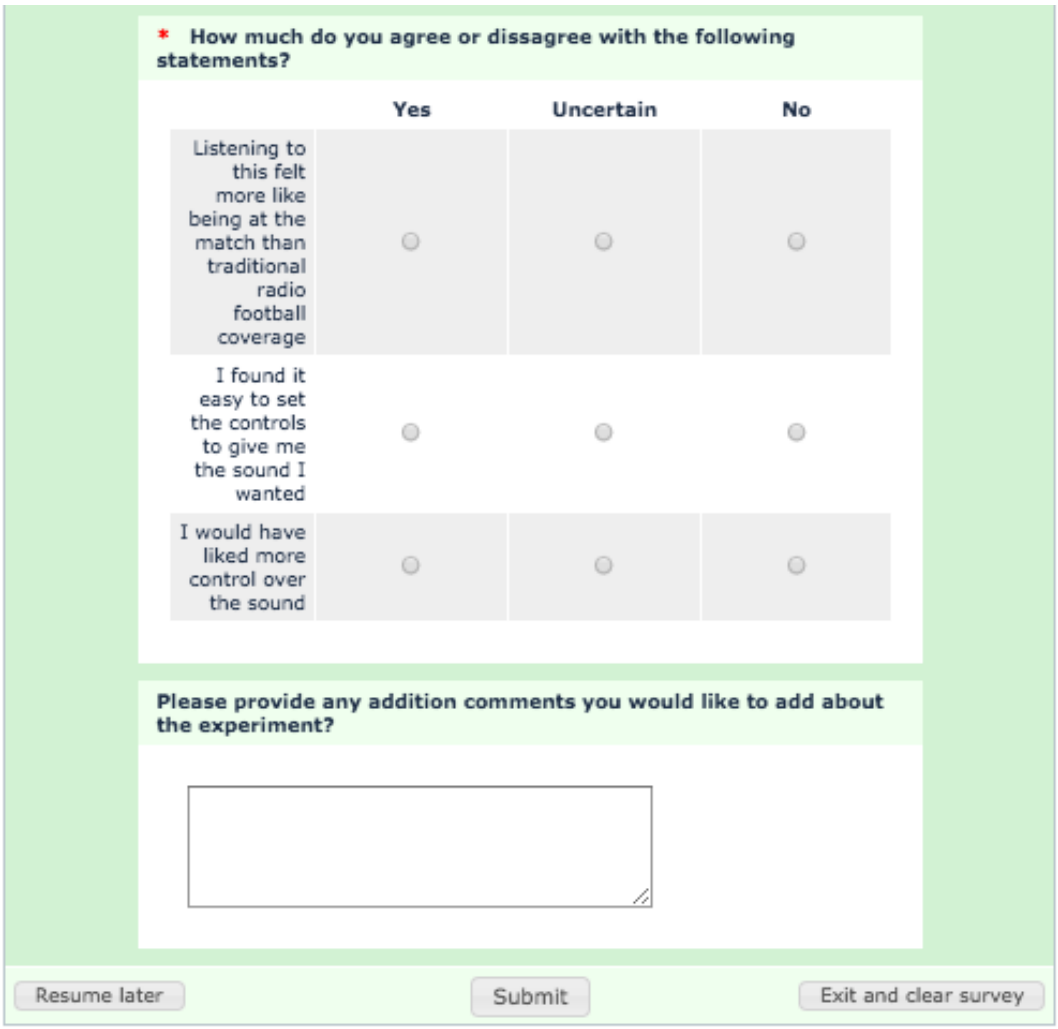
*
The application gave you the ability to control the balance between the pitch sound and the commentary.
Please rate how this functionality impacted upon your overall experience of match.

Choose one of the following answers

- Much better
- Slightly better
- About the same
- Slightly worse
- Much worse

* **How did this experience compare to listening to more traditional radio coverage of a football match?**
Choose one of the following answers

- Much better
- Slightly better
- About the same
- Slightly worse
- Much worse



* **How much do you agree or dissagree with the following statements?**

	Yes	Uncertain	No
Listening to this felt more like being at the match than traditional radio football coverage	○	○	○
I found it easy to set the controls to give me the sound I wanted	○	○	○
I would have liked more control over the sound	○	○	○

Please provide any addition comments you would like to add about the experiment?

Resume later | Submit | Exit and clear survey

Appendix B

Appendices Relating to the Foreground vs Background Study in Chapter 4

B.1 Production Script

Written by Carlo Collodi's dramatised by Linda Marshall Griffiths.

Included in the following pages is a section of the production script reproduced with kind permission of the drama's author Linda Marshall Griffiths. This is the final production script for the section of the recording used for the object identification and classification test with the Producer and Sound Designer and used in the foreground vs background preference experiment.

COACHMAN.	That's it then.	
	HE TAKES PINOCCHIO'S REIN.	
	Now we must take the long walk my friend.	
	THEY WALK.	
	EXT. DAY EDGE OF CLIFF BY SEA *THE SEA CAN BE HEARD.*	
PINOCCHIO	(V/O) And I smell the sea, and I think of my father so long ago... I'm glad, he's not alive to see this.	
COACHMAN.	Now my friend, let's make this quick. Stone round the neck, quick shove, splash - you won't know anything. The tide's high today.	
	Let me just tie this rope round your leg, to pull you back in by. Don't look at me, it's no good crying now. I've got to get my money's worth - your skin's so hard I intend to drown you then make a drum of you. I play, you see - in my village band so you'll have some use.	
	We're not bad, got gigs!	
	Goodbye.	
	HE SHOVES PINOCCHIO INTO THE SEA. PINOCCHIO GOES DOWN AND DOWN.	
	EXT. DAY UNDER WATER	
	PINOCCHIO GOES DOWN FAST.	
PINOCCHIO	(V/O) I'm going down fast.	

A light...
A blue.

(*OUT-LOUD, SWALLOWED BY WATER*) Blue Fairy!!

BLUE FAIRY (V/O, A WHISPER) Remember when we met?

PINOCCHIO (V/O) I remember?

STILL GOING DOWN.

BLUE FAIRY (V/O) What you said you wanted?

PINOCCHIO (V/O) No.

BLUE FAIRY (V/O) Remember...

PINOCCHIO (V/O) For my father to love me.

BLUE FAIRY. (V/O) No, he always loved you.

PINOCCHIO (V/O) I don't know...
Help me, I'm still going down.

HE SHOUTS OUT IN THE WATER.

PINOCCHIO (V/O) The fishes!

Ow, ow.
They eating the skin and hair...

(*SCREAMING IN THE WATER*) Ow, ow!!!

HE STRUGGLES IN THE WATER, STILL GOING DOWN – A GREAT NOTE BUILDING AS HE CONTINUES TO SINK.

(V/O) But they never reached the bone. For I am no donkey you see and a wooden boy is

still underneath, I let go the rope round my leg and still I keep going down.

HE GOES DOWN, THE WATER ALL ROUND HIM - DEEPER.

BLUE FAIRY (V/O, A REAL WHISPER) What is it you wanted?

PINOCCHIO (V/O) Don't leave me, I can't see you anymore.

BLUE FAIRY (V/O) What was it?

PINOCCHIO (V/O) I wanted to be a boy all right.
To have a heart!

SHE BREATHES INTO THE WATER AND LIFTS AWAY.

BLUE FAIRY (V/O) You have a heart...

PINOCCHIO (V/O) And the blue light is so far above me. Just a speck of light, spread on the surface and shrinking as I still go down.

As I get deeper and the water presses in.

Listen...

EVERYTHING IS VERY STILL.

Listen.
Can you hear that?

SILENCE.

No sound.
No heart beating.

The heart I have is no longer wood, it has turned to stone and it is taking me down down.

	A CRY OF A BOY DEEP DOWN IN THE WILDERNESS. THEN BENEATH THAT A GREAT ROAR.	
	Something's moving towards me faster, deadlier...	
	A GREAT WAVE DEEP...	
	A great opening gulf! He heard me... He smelt my fear, Fear he loves, the great shark bites.	
	PINOCCHIO IS SWALLOWED BY THE GREAT SHARK.	
	INT. DAY SHARK	
	A GREAT CAVE, WATER RUNS, ECHOES. DRONES, SPLASHES.	
PINOCCHIO.	Hello! Hello, is anyone here? So dark. So black.	
	All around me. An inkwell.	
	STARTS TO CRY.	
	Don't leave me here, please don't leave me here. I'm scared of the dark, I can't...	
	I'm scared.	
	SUDDENLY A LIGHT IS STUCK.	
	A light. The blue light of a candle. I got lost, I didn't mean...	

	Don't move away, wait...	
	HE SCRAMBLES AFTER THE LIGHT.	
	I'm coming after you - I'm not going to lose my way this time...	
PINOCCHIO (cont.)	Wait...	
	A MAN SINGS SOMEWHERE NEAR.	
	Hello? Help! I can see your lantern!! Hello!	
	HE RUNS FORWARD.	
	Will somebody help me?	
GEPPETTO.	Lucky I lit my candle - it is so low now I've been sitting in the dark a long time.	
PINOCCHIO.	I saw it.	
GEPPETTO.	Who are you?	
	PINOCCHIO COMES CLOSE, GEPPETTO GASPS.	
PINOCCHIO.	Papa?	
GEPPETTO.	Pinocchio?	
	PINOCCHIO RUNS INTO HIS ARMS.	

Appendix C

Appendices Relating to the Personalised Content Study in Chapter 5

C.1 Questions Used for Online Study and Focus Groups

Included on the following page are questions written by the thesis author to user understand the audiences' perception and reception of the location based audio drama "Breaking Out". These questions were used for the online study and as a basis for discussion for the focus groups.

0%
100%

Media habits and general impressions

We would like to know a little more about your media habits and what you thought of the play.

* **How often do you listen to radio/audio dramas?**

- Daily
- Weekly
- Monthly
- Once in a while
- Never

* **Please tell us which browser you used to listen to the audio drama?**
Choose one of the following answers

- Chrome
- Firefox
- Internet Explorer
- Safari
- Don't know

* **Breaking Out featured a robotic lift voice. This has been generated using a prototype technology that is still in development.**
How much did the qualtity of the robotic voice affect your enjoyment of the drama?

	Negatively affected a lot	Negatively affected slightly	Had no affect	Positively affected slightly	Positively affected a lot
Please select:	○	○	○	○	○

* **Thinking about Breaking Out, please rate the following in terms of how much you liked or disliked them.**

	Disliked a lot	Disliked slightly	Neither liked or disliked	Liked slightly	Liked a lot	Don't know
Overall experience	○	○	○	○	○	○
Story/plot	○	○	○	○	○	○

Next ▸ Exit and clear survey

Use of regional references

Breaking Out made use of your approximate location in order to use regional references in the story. Your computer determines your geographic location and plays the most appropriate regional references to you during the drama.

This process can provide a localised experience. For instance, local landmarks, regional weather, events in your area and local news can all be incorporated into the story.

* **Did you notice the regional references in Breaking Out?**

- ○ Yes
- ○ No

* **Thinking about the regional references, how local to you were they?**

- ○ Very local (e.g. same town or city)
- ○ Moderately local (e.g. nearby city)
- ○ Somewhat local (e.g. county)
- ○ Not local at all (e.g. UK wide)
- ○ Don't know or N/A

* **How much did you like the use of regional references in the play?**

	Disliked a lot	Disliked slightly	Neither liked or disliked	Liked slightly	Liked a lot	Don't know or N/A
Please select:	○	○	○	○	○	○

* **The use of regional references in Breaking Out made me...**

	Strongly disagree	Disagree	Neither agree or disagree	Agree	Strongly agree	Don't know
feel more engaged with the story	○	○	○	○	○	○
more familiar with the setting	○	○	○	○	○	○
feel closer to the main character	○	○	○	○	○	○

* **Roughly how many of the regional references in Breaking Out had you heard of before?**

- ○ All of them
- ○ Some of them
- ○ None of them
- ○ Don't know or N/A

We would like to know how we could develop this editorial format. We'd also like to know if, and how, you would recommend it to others.

*** In Breaking Out, your approximate location was used to personalise the play. Would you be willing to let the following other types of information be used in this way too?**

	Definitely not	Probably not	Unsure	Probably yes	Definitely yes
Exact location	○	○	○	○	○
Personal details (name, gender, age, race)	○	○	○	○	○
Names of friends and family	○	○	○	○	○
Online purchase history	○	○	○	○	○
Your mood (via heartbeat or temperature)	○	○	○	○	○
Internet browser history	○	○	○	○	○
Your social network activity	○	○	○	○	○

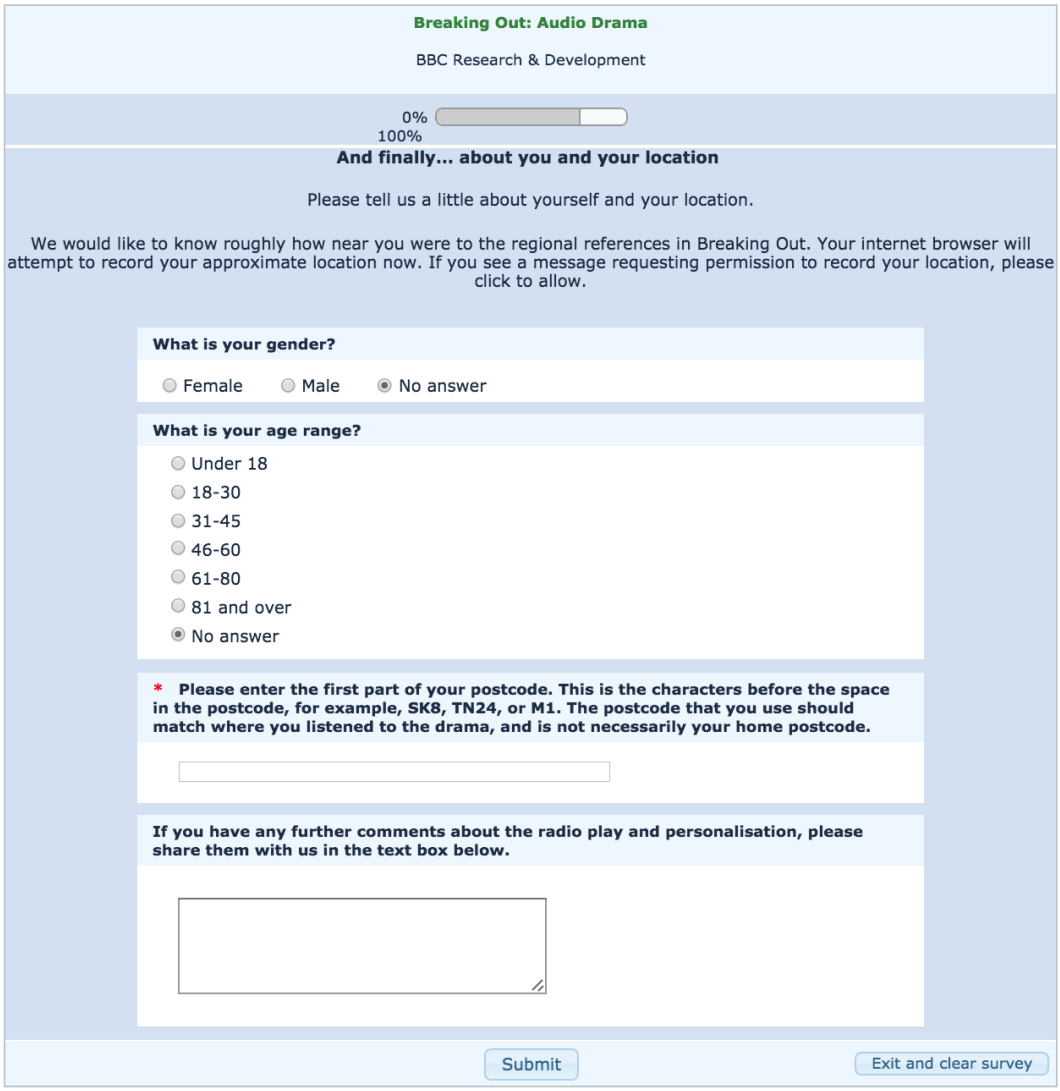
*** How much effort would you be willing to make in order to have a personalised audio drama?**

- ○ None, I'd want to receive it automatically
- ○ Log in via existing social network (e.g. facebook or twitter)
- ○ Set up an account specifically for the service
- ○ Don't know
- ○ Would not wish to use

*** Does the fact that the play is personalised mean that you are more or less likely to recommend it to someone else?**

- ○ Less likely
- ○ No change
- ○ More likely
- ○ Don't know

How would you describe this personalised audio drama experience to other people?


Breaking Out: Audio Drama
BBC Research & Development
0%
100%
And finally... about you and your location
Please tell us a little about yourself and your location.
We would like to know roughly how near you were to the regional references in Breaking Out. Your internet browser will attempt to record your approximate location now. If you see a message requesting permission to record your location, please click to allow.
What is your gender?
Female
Male
No answer
What is your age range?
Under 18
18-30
31-45
46-60
61-80
81 and over
No answer
* Please enter the first part of your postcode. This is the characters before the space in the postcode, for example, SK8, TN24, or M1. The postcode that you use should match where you listened to the drama, and is not necessarily your home postcode.
If you have any further comments about the radio play and personalisation, please share them with us in the text box below.
Submit
Exit and clear survey