Grant Agreement No.: 687645
Research and Innovation action
Call Topic: H2020 ICT-19-2015

Object-based broadcasting – for European leadership in next generation audio experiences

D2.5: Final Pilot Progress Report

Version: v1.0

Deliverable type	R (Document, report)
Dissemination level	PU (Public)
Due date	31/05/2018
Submission date	30/05/2018
Lead editor	Werner Bleisteiner (BR)
Authors	Werner Bleisteiner (BR), Niels Bogaards (EC), Michael Meier (IRT), Niko Färber (FHG IIS), Chris Baume (BBC)
Reviewers	Andrew Mason (BBC), Michael Weitnauer (IRT)
Work package, Task	WP2, T2.4, T2.5
Keywords	Object-based audio (OBA), pilot, MPEG-H, interactive functionalities, on-demand consumption

Abstract

This document is based upon the previous D2.3 Interim Pilot Progress Report. It describes the final status of both phases of the ORPHEUS pilot. The first phase was divided into three stages, with the first being a live production of an interactive object-based radio drama experienced using a web browser, the second a selection of material encoded using MPEG-H and made available through an iPhone and AV receiver, and the third being an 'as-live' broadcast, live encoded using MPEG-H and made available over the Internet.

In the second phase object-based audio productions from pilot phase 1 were enhanced with interactive functionalities for on-demand consumption.

Document revision history

Version	Date	Description of change	List of contributor(s)
v0.15	22/03/2018	Initial structure	W. Bleisteiner (BR)
v0.2	06/04/2018	Chapter 1 reworks and additions	W. Bleisteiner (BR)
v0.31	19/04/2018	Chapter 5 contributions	W. Bleisteiner (BR)
v0.33	23/04/2018	Input from elephantcandy	N. Bogaards (EC)
v0.34	24/04/18	Chapter 6 contributions	W. Bleisteiner (BR)
v0.4	25/04/18	Exec. Sum. and Chapter 7	W. Bleisteiner (BR)
v0.45	25/04/18	Chapter 5.5	N. Bogaards(EC)
v0.5	27/04/18	Appendix A	W. Bleisteiner (BR)
v0.6	04/05/18	Chapter 4& 5 contributions (consolidated)	Niko Färber (FHG IIS) Michael Meier (IRT)
v0.7	08/05/18	Corrections by reviewer	Andrew Mason (BBC)
v0.8	09/05/18	Added Section 2	Chris Baume (BBC)
v0.9	09/05/18	Corrections by reviewer	Andrew Mason (BBC)
v0.95	11/05/18	check previous contributions	W. Bleisteiner (BR)
V0.96	15/05/18	Changes to some captions. Correction of name in figure 14 2nd reviewer annotations	Andrew Mason (BBC) Michael Weitnauer (IRT)
v0.97	17/05/18	Preparation for approval	W. Bleisteiner (BR)
v1.0	23/05/18	Final for publication, editorial check	W. Bleisteiner (BR), A. Köhler, U. Herzog (EURES)

[1] http://creativecommons.org/licenses/by-nc-nd/3.0/deed.en_US

Executive Summary

This document describes both ORPHEUS pilots in creative and technical perspectives.

Pilot Phase 1: Live multiplatform object-based audio was executed between June and September 2017 and comprised three stages:

- Stage A was a live production of an interactive object-based radio drama at the BBC that could be experienced using a web browser only, using the AAC audio codec plus serial ADM to deliver object-based audio metadata.
- Stage B contained a selection of pre-produced content, offline encoded into MPEG-H and made available on the preliminary version of the ORPHEUS iOS app and the Trinnov AV receiver.
- Stage C of pilot phase 1 was a 'live encoded' object-based audio broadcast, streamed using MPEG-H and made available as an object-based audio stream on the ORPHEUS iOS app and the Trinnov AV receiver.

Pilot Phase 2: Enhanced object-based audio for on-demand consumption was accomplished in April 2018. In this final phase, content provided by the ORPHEUS partners BR, BBC and FHG IIS, that had been used as test material for pilot 1 stage B, was refined with additional interactive features and provisioned for delivery in a published version of the ORPHEUS app, demonstrating the full scale of for object-based audio functionalities.

Table of Contents

List of Figures

List of Tables

Abbreviations

AAC	Advanced Audio Coding
ADM	Audio Definition Model
BW64	Broadcast Wave format (64-bit)
CDN	Content delivery network
DAW	Digital audio workstation
IP	Internet Protocol
LOI	Level of importance – model applied for variable-length structuring
MPEG-DASH	Dynamic Adaptive Streaming over HTTP
MOS	Mean Opinion Score

1 Introduction

The aim of the project pilot is to demonstrate a full chain of object-based audio production, distribution and reproduction, based upon the pre-defined pilot implementation architecture.

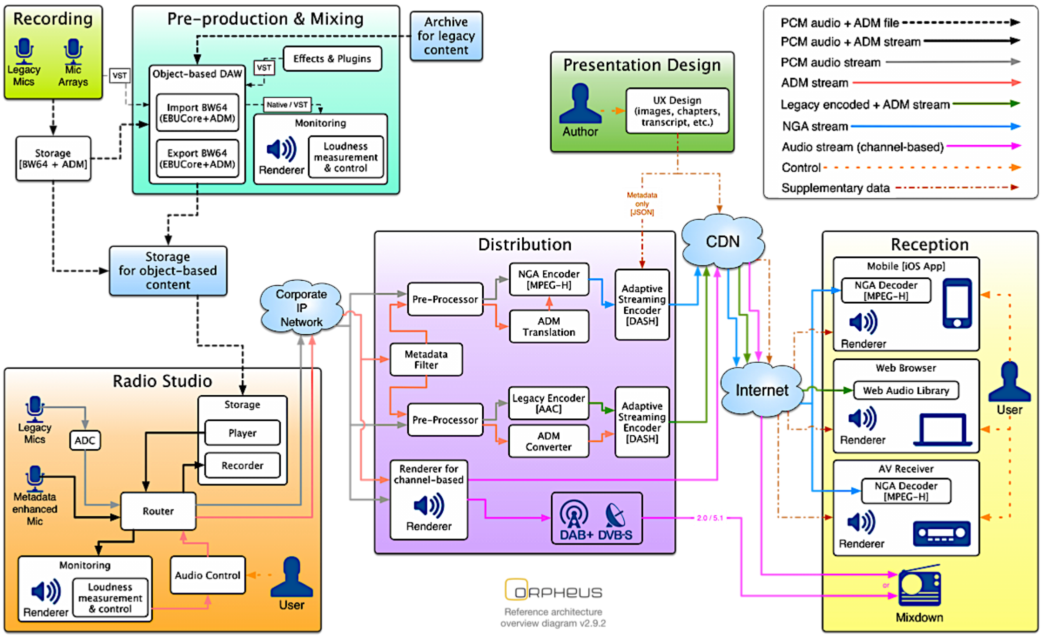

Figure 1: ORPHEUS Reference architecture Infrastructure

All entities (macro blocks) of a radio broadcast infrastructure are involved (end-to-end chain): Recording - Pre-production - Radio Studio – Distribution and Reception.
The pilot in ORPHEUS had two main phases:

- Pilot Phase 1: Live multiplatform object-based audio
- Pilot Phase 2: Enhanced object-based audio for on-demand consumption

The central hub for the production of pilot phase 1 (the "Radio Studio") was the IP Studio installed in Studio 30/D at the BBC[2] - for the production and play-out of a live programme (on air).

Figure 2: ORPHEUS Radio Studio at BBC

[2] c.f. *ORPHEUS-D3.4-v1.1_Implementation and documentation of a live object-based production environment* for detailed description of facilities. http://orpheus-audio.eu/wp-content/uploads/2016/12/orpheus-d3-4-v1-1_implementation-and-documentation-of-a-live-object-based-production-environment.pdf

1.1 Pilot phase 1

During the development of the technical solutions in the project it became obvious that a decoupling of various processes and parallel task management serves best in order to solve both the complexity of the implementation and integration process as well as the possibility to ensure backward compatibility to existing infrastructure. For this reason, pilot phase 1 has been divided in three sub-stages.

A – Live AAC+ADM

In this first phase, the MPEG-H encoder for delivery to the end-consumer was not yet fully integrated into the complete IP Studio and distribution infrastructure. Therefore, as an operational workaround, the 'legacy' AAC encoding solution, as a proven, reliable, real-time possibility, with synchronous delivery of the object-based ADM metadata was chosen, in order to provide a first integration step.

B – As-live MPEG-H

Next, an interim "as-live" Stage B was inserted, where the material used in stage A was encoded offline into an MPEG-H stream in advance and then broadcast "as live" immediately after completion. This allowed us to explore, test and evaluate the process of MPEG-H streaming distribution over MPEG-DASH extensively.

C – Live-encoded MPEG-H

In this final stage, the MPEG-H encoding capability was fully integrated within each of the production, play-out, distribution and reception entities (IP Studio, Distribution, Reception macro blocks). It was then possible to stream the audio plus ADM directly from the BBC IP Studio, encode it live and deliver it over various distribution channels in object-based formats to the ORPHEUS app and the ORPHEUS AV receiver, as well as hosted pre-rendered for 'legacy broadcast' dissemination. At this time, a solid version of the ORPHEUS app was also made available on the B2B AppStore.

This complete working chain of pilot phase 1 was successfully presented to the EU project coordinator and the reviewers during the 2nd review of ORPHEUS at the BBC Broadcasting House in London on December 13th, 2017.

Figure 3: ORPHEUS 2nd project review at BBC

1.2 Pilot phase 2

By definition in the ORPHEUS project, pilot phase 2 was to develop and demonstrate "enhanced object-based audio for on-demand consumption". We had also decided in the planning stage of the project, that demonstration content produced during pilot phase 1 should serve as input for pilot phase 2.

The original idea to set-up a common editorial and production team, consisting of experienced staff from the BBC and BR, commissioning and creating an imaginary "OPRHEUS Object-based Radio Show" that provides the content for the two pilot phases, unfortunately turned out to be too complicated and impractical. Therefore the BBC focussed on adapting existing productions for on-demand delivery – whereas BR pursued to collaborate with regular broadcasting productions and enhance them gradually with object-based features and experience.

Figure 4: 'Hybrid' working with OBA Test Rack in BR's studio #9 control room

The main restriction in this approach was the concessions, that all productions at BR needed to be made under 'real working conditions', with mainly the existing and, for regular broadcast productions, approved tools in BR's studios. This means that the standardized and strictly-administered production tools (i.e. ProTools) in a radio drama studio, had to be used as the initial recording and editing DAW. Fortunately, the professional studio architecture at BR is not self-contained, so that various interfaces offer possibilities to hook-up with external gear.

So we were able to attach a specially-designed 'ORPHEUS OBA Test Rack' to the existing equipment in order to simulate, create and experiment within a 'hybrid' standard/object-based production environment.

Four of the ORPHEUS pilot phases pieces from BR were pre-produced under these conditions:

- *Experience object-based audio* by David Globig
- *The Art of Foley* by Max Bauer, Andrea Kilian and Bernhard Jugel
- *Mozart Gigue KV 574 in 360º* by *Passo Avanti*
- *Heute im Stadion (Football Match)*

The first three are typical 'pre-produced' pieces, recorded, edited and mixed in a studio in advance, well before going on air. Whereas the *Football Match* is typical for live radio – but for this purpose it was also 're-enacted' and recorded in the studio.

All of these pieces are available in the ORPHEUS app. There they may be experienced with additional program info, chapter selection, transcript, and dial wheel navigation, in variable length.

Compared to today's representation of audio on-demand – still being merely recordings from on air transmissions – these functionalities demonstrate impressively these 'enhanced object-based audio features'.

2 Pilot Phase 1 - Stage A: live

In deliverable 2.3[3], we described the design of Pilot Phase 1 Stage A - our experimental live object-based broadcast, entitled "The Mermaid's Tears". In this section, we start by reflecting upon the delivery of our pilot and discuss the ways in which our plans changed from the original design. We go on to describe the interfaces that we built to be able to deliver the pilot. Our intention is to share our experience in producing a live object-based broadcast so that others who are interested in doing the same can learn from what we found. Finally, we present the results of our large-scale feedback through BBC Taster, in which over 1000 people listened to our pilot production.

2.1 Pilot delivery

The aim of Pilot Phase 1 Stage A was to broadcast a live object-based audio stream. We chose to create a radio drama, as this format made it easier to integrate interactive and immersive audio elements into the programme. However, the existing drama production workflow is entirely based on pre-production. In the early days of radio, drama was broadcast live, but this is no longer the case, and the people with the skills required to achieve this have since disappeared. For this reason, we decided to first pre-produce the drama as normal, then turn it into a live production afterwards.

2.1.1 Live broadcast

We started producing our live broadcast by following the existing drama production workflow to create the drama in stereo using a DAW. As each individual sound source remained separate in the DAW, we then unpicked the drama recording into a number of individual objects for live playout. We loaded the objects into a "sequence store" in the BBC's IP Studio platform, which could later be triggered for live playout using our production interface.

We encountered a technical issue with playback of the audio objects that could not be resolved by our transmission date. Unfortunately, this meant that we were unable to include the pre-produced elements in our live broadcast. Instead, we used two sets of stereo microphones to capture the audio for each location in the drama. Due to the limited audience experience of the live broadcast, we only used a small set of test listeners to test that the technology behind the pilot operated correctly. We then focussed our efforts on creating a fully-featured pre-produced experience for BBC Taster.

2.1.2 Metadata representation and emission

We used the Audio Definition Model (ADM) as a method of describing our audio objects, and the Universal Media Composition Protocol (UMCP) to represent and transmit the ADM metadata within the IP Studio production platform. ADM was designed for use in object-based audio file storage rather than live streaming. Several ORPHEUS partners are involved with developing Serial ADM (sADM), which provides support for live object-based audio streams. However, at the time of our live broadcast, our work on sADM was not sufficiently advanced that we were able to use it for our pilot. For this reason, we broadcast the UMCP directly to the audience.

By the time we were ready to put our pre-produced pilot onto BBC Taster, the sADM specification and our tools were sufficiently advanced that we were able to use them. We created a system that converted the UMCP 'production' metadata into the sADM 'emission' metadata. This involved flattening any hierarchical structures as the production metadata could contain nested groups whereas the emission metadata could not.

[3] c.f. ORPHEUS-D3.2_Interim Pilot Progress Report https://orpheus-audio.eu/wp-content/uploads/2017/06/orpheus-d2-3_interim-pilot-progress-report.pdf

2.2 Interfaces

Due to the lack of 'off-the-shelf' solutions available, we developed custom user interfaces for both the production and audience end. Each of these are described below.

2.2.1 Pre-production interface

When designing the system for delivering our pilot, we proposed creating a 'pre-production system' that would capture static metadata about the programme. This would then populate the live audio control interface for presentation to the audience, or be fed directly to the audience via the broadcast metadata.

We did not create a pre-production interface for several reasons. Firstly, we had to create a custom user interface for presenting the drama to the audience. This afforded us the possibility of including relevant metadata, such as triggered images, directly into the interface. A generic user interface requires static metadata, but a custom interface does not. Secondly, in designing and implementing such a pre-production system, we would have to create a data model that would be sufficiently generic to future object-based broadcasts. This is a significant challenge, and one which we did not have time to complete before the pilot broadcast.

Our solution was to capture any dynamic metadata directly in the audio control interface, and to place any static metadata directly in the presentation interface. However, it would be preferable in the long term not to have to develop custom presentation interfaces for each programme. A generic presentation interface would need to be populated with metadata, and a pre-production interface and suitable data model would then be required to capture this.

2.2.2 Live production interface

We could not find any 'off-the-shelf' products that would allow us to capture the metadata we required for our live object-based broadcast. To deliver our pilot, we designed and implemented an 'audio control interface' to capture and distribute this metadata (see Figure 5).

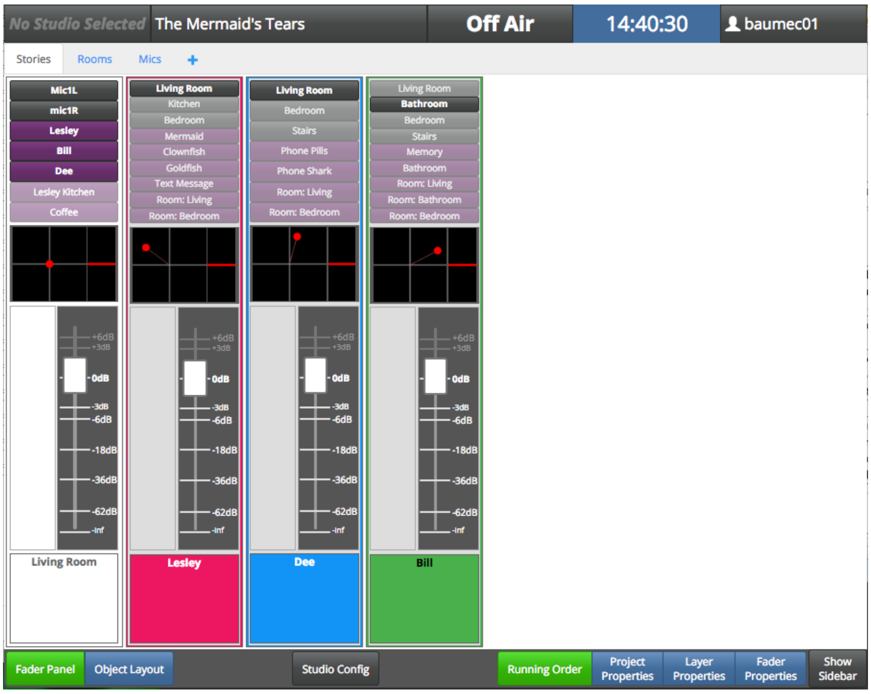

Figure 5: The Orpheus 'audio control interface', used to capture and distribute the required metadata.

The interface was developed as a web-based tool using the React UI framework and the UMCP API. It captured the hierarchical structure of the programme and objects, in addition to the gain, position and label of each audio object. The interface was designed to be similar to a traditional mixing desk to give it familiarity, but powerful enough to create a wide range of object-based content.

Our interface could be used to create interactive experiences by marking groups of objects as 'mutually exclusive', indicating that they should not be played back at the same time. We also implemented a method of triggering events that could be used for any purpose. For our pilot, we used these 'triggerables' for pop-up images.

2.2.3 Distribution control interface

Our pilot was distributed through 15 audio channels and a metadata stream. These were served to the user using MPEG-DASH. Due to the current limitations of web browsers, we grouped the 15 channels into three groups of 5. The DASH stream was generated in the BBC's IP Studio platform within the production system. We then used a script to automatically copy the packets of the DASH stream to a CDN that served the files to the audience.

Although in our original design, we envisaged the need for a 'distribution control interface', we ended up using a script controlled using a command line. This fulfilled our needs for the pilot, but such an interface required good technical knowledge and is not very 'user friendly'. When creating a large-scale broadcast system, the broadcast engineers will likely want to have a GUI for observing the status of, and controlling, the distribution of programmes.

2.2.4 Presentation interface

We created a custom interface (see Figure 6) for presenting and playing our pilot to the listeners. The interface was implemented as a website so that it could be accessed and experienced without having to install special software. We designed our interface to include the following features:

- Character selection (Lesley, Dee, Bill)
- Output format selection (stereo, surround, binaural)
- Triggered images and overlays

We worked with a graphic designer who created the image assets, and a front-end web designer who coded the website. We used the BBC's *bbcat-js* library[4] to receive the audio and metadata, and to render the audio to the selected output format.

[4] The BBC Audio Toolkit JavaScript library is, at the time of writing this report, in the process or being made ready for release as open source software

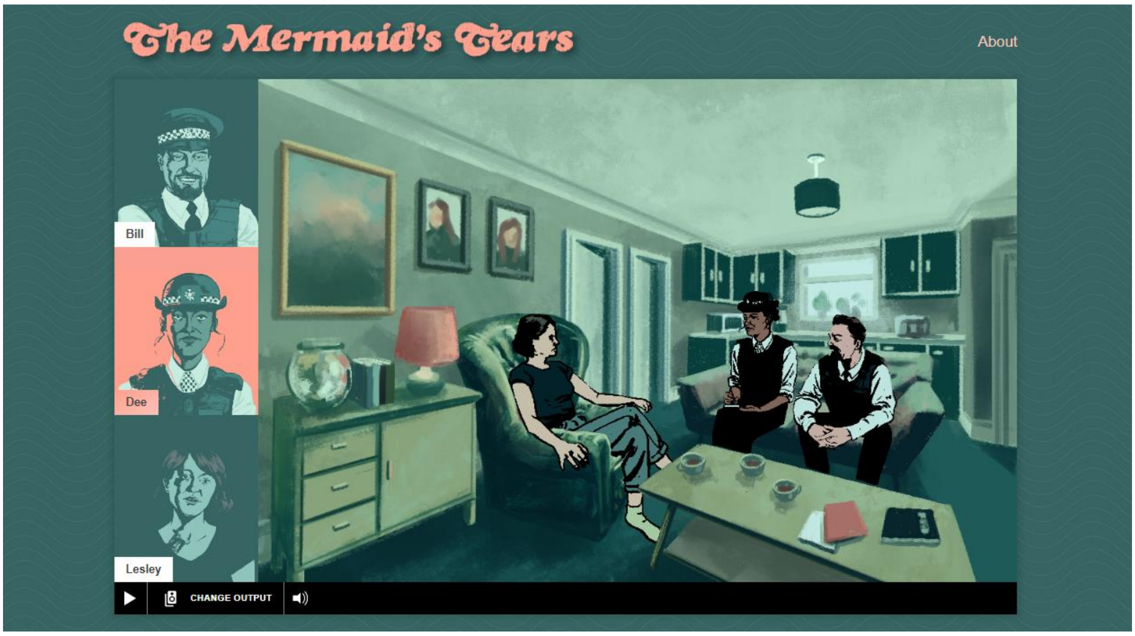

Figure 6: The custom presentation web interface developed for The Mermaid's Tears.

2.3 Feedback

We evaluated Pilot 1 Stage A through BBC Taster. Taster is an online platform that the BBC uses to share prototypes with the general public and receive feedback. We put The Mermaid's Tears on BBC Taster as a 3-month pilot. During that period, over 1000 people listened to the drama. 50% of listened on mobile phones, 39% on desktop and 11% on tablet.

The BBC Taster platform includes a feedback mechanism where the audience can give the pilot a star rating and complete a simple questionnaire. 115 people provided feedback on our pilot during its 3-month run.

2.3.1 Star rating

Participants were asked to give their experience an overall rating, shown in Figure 7. Nearly half of the participants rated our pilot with 5 stars.

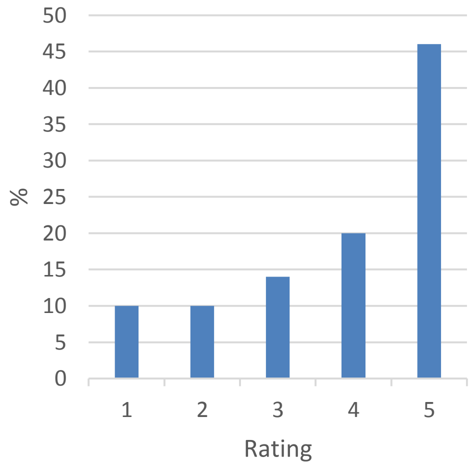

Figure 7: Results of overall star rating for Pilot phase 1, stage A.

2.3.2 Feature comparison

Object-based audio supports a variety of new features. We were interested in finding out which of these the participants preferred. We were unable to support a ranking input due to technical constraints, so we asked participants to select the feature they enjoyed the most.

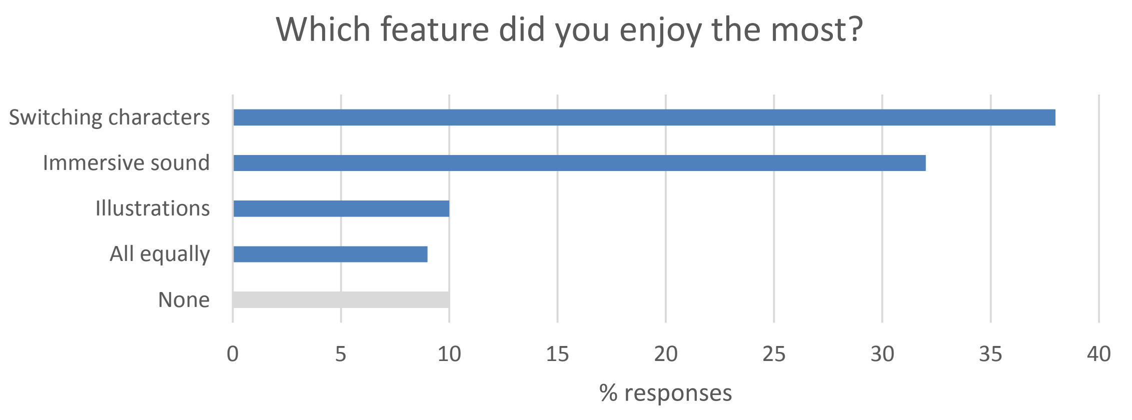

Figure 8: Comparison of the appreciation of the different features offered

2.3.3 Parallel story-lines

Our drama contained three parallel story-lines. We wanted to learn whether having multiple simultaneous stories made the experience richer, or whether listeners felt the experience was incomplete as they were missing out on the other stories.

We asked participants “Did having a choice of story-lines feel richer or incomplete?”. Nearly half the participants found it a richer experience, whilst less than 20% said it felt incomplete.

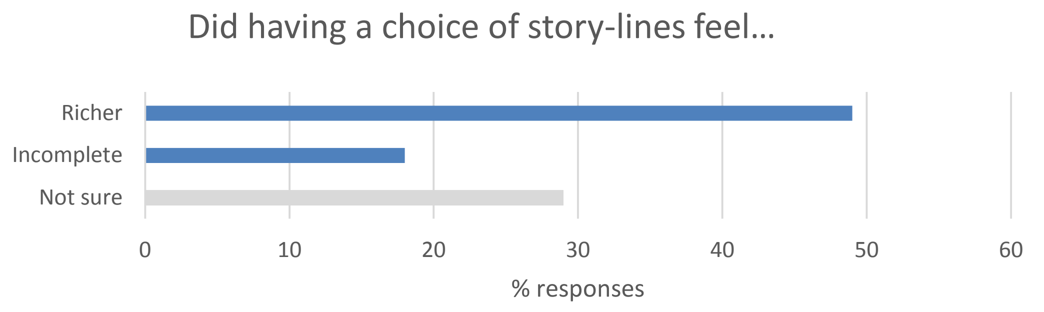

Figure 9: Rating of the experience of parallel story-lines

The three stories in the drama each provided a different experience. We wanted to find out how many times the participants listened to the drama. Nearly 60% only listened once, but over 35% listened twice. Only 2 participants out of 115 listened more than twice.

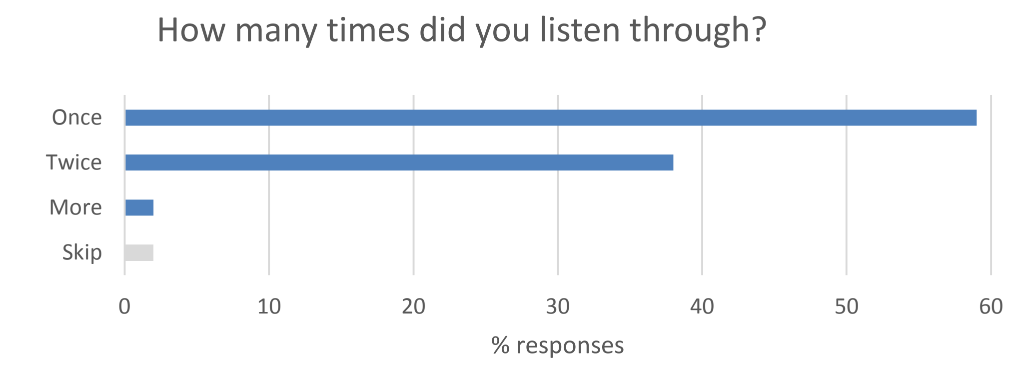

Figure 10: Proportion of respondents listening different numbers of times through

2.3.4 Relationship with broadcaster

Every prototype on the Taster platform includes some standard questions about how the experience affects the participant's relationship with the BBC.

Participants were asked whether the BBC should do more stuff like this. Nearly 90% of participants said yes.

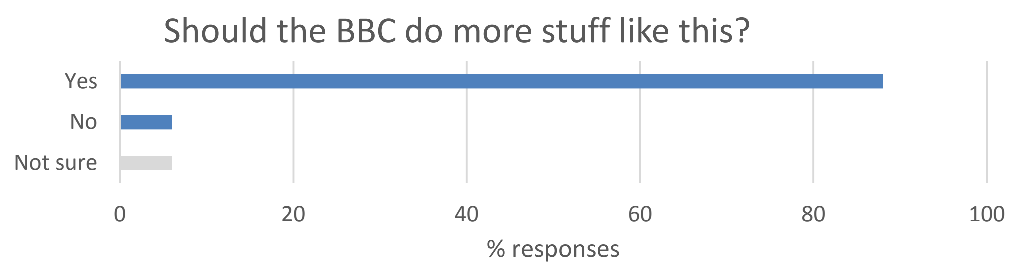

Figure 11: Proportion of respondents wanting more of this to be made

We also asked whether our pilot made the participants feel better, the same, or worse about the BBC. Over two-thirds of participants said it made them feel better about the BBC, with only one participant saying it made them feel worse.

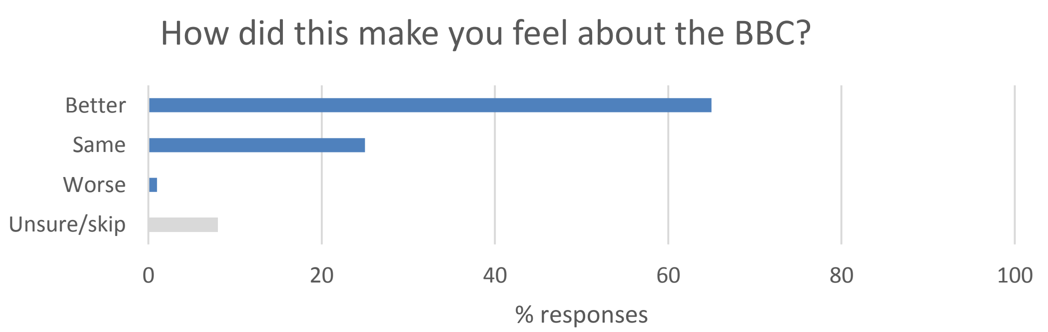

Figure 12: Effect on the audience's opinion of the BBC

3 Pilot Phase 1 - Stage B: as-live

In Stage B of Pilot 1, object-based radio programmes were distributed using MPEG-H over MPEG-DASH to the ORPHEUS Radio iOS client app and the Trinnov AV receiver. While the aim was to stay as close as possible to the live scenario of Pilot 1 Stage A, additional material was produced to experiment with a larger variation of encoding and interaction configurations. Pilot 1 Stage B was used extensively to demonstrate to external audiences the concepts and early results developed in the project. In the ORPHEUS Radio app, extra metadata was used to simulate some functions.

3.1 MPEG-H encoding

The integration of an MPEG-H encoder into the production chain was identified as a significant development effort, both at FHG and at the BBC. In order to be able to evaluate the ORPHEUS concepts in an end-user setting as quickly as possible, Pilot 1 Stage B used MPEG-H files that were encoded offline by FHG. The source material was produced in a number of different formats and therefore the MPEG-H conversion was mainly a manual process. Shortly after it was broadcast live in Pilot 1 Stage A, The Mermaid's Tears was encoded into MPEG-H. The MPEG-H files were then streamed to the iOS app and the Trinnov AV receiver using MPEG-DASH, in the same way as a live encoded stream would be, hence the term 'as live'.

3.2 Additional material

To be able to test and present more aspects of the object-based delivery chain, Stage B of Pilot 1 also featured early versions of material dedicated for Pilot Phase 2 with basic playback functions:

- ***Herbst***
  Herbst (Autumn) is a poem accompanied by 3D recorded nature sounds. The listener can choose between a German and an English version of the spoken text. Additionally, the sound of a humming bee can be interactively located anywhere in 3D space
- ***Football Match***
  An extract from a soccer match in which the stadium atmosphere is recorded as a 5.1 surround "background object" while the commentary is the "foreground object". The user can control the prominence level between the two objects according to their personal preference.
- ***The Art of Foley***
  A full radio documentary about the creation of sound effects for films and radio drama. This BR production was originally made in 5.1 surround sound. The narration can be rendered in English, German or French.
- ***The Turning Forest***
  This is an existing radio drama available as BW64+ADM. However, the distribution as MPEG-H requires converting a high number of objects into a lower number of channels. In addition, BR produced a German narrator for this drama, which allows testing the selection of one out of two languages. This item is a test case for the conversion process from BW64+ADM to MPEG-H.
- ***Mozart Gigue***
  As a pure music piece, Mozart's Gigue KV 574, performed by Passo Avanti, highlights some other aspects of object-based audio: each instrument is available as a separate audio object, that can be manipulated by the listener. In the app, a choice can be made between four spatial scenes, where the instruments are positioned in different locations with respect to the listener. One of the modes features a slow rotation where the ensemble circles virtually around the listener.

3.3 End-user experience

3.3.1 Availability

The goal of Pilot 1 Stage B is to evaluate the technical completeness and correctness of the system, as well as the comprehensibility and usability of the iOS app. While the prototypes of the app and the AV-Receiver were not available publicly yet, the results were demonstrated at workshops, fairs and other closed settings. The Pilot 1 Stage B app also served as the basis for the end-user tests by b<>com and FHG.

3.3.2 Functionality

As Stage B includes a recording of the live broadcast from Stage A, the goal was to make all features listed in Section 2.1.1 also available in Stage B (Select Output Format, Choose/Switch Character, Image Display). The conversion from BW64+ADM version of the Mermaid's Tears into MPEG-H was not a trivial process: the audio format and features needed to be adapted to match the constraints given by the MPEG-H 3D Audio Low Complexity Profile at Level 3. In addition, the specific interaction and graphics that were created for the browser-based Stage A were not readily available in the ORPHEUS app and were only partly implemented for the platform. However, this exercise verified the feasibility of converting the production format (BW64+ADM) into a distribution format (MPEG-H) and demonstrated the universality of the ORPHEUS architecture with respect to distribution formats (AAC and MPEG-H) and end-user devices (Browser, iPhone, AVR).

The following features were available on the iOS app. The focus is on audio-related functionality that has a specific mapping to MPEG-H and object-based audio:

- **Navigation and Tune-In**
  While Stage A provides a single stream which lasts for a significant amount of time, Stage B offers several items of content. The user can navigate between those and select one, based on some summary information (a picture, some text). As we simulate a live broadcast, the item might not start from the beginning but start in the middle, as is typical if tuning in to a live event.
- **Skip-Back and Catch-Up**
  The user can decide to skip back in time and e.g. listen to the start of the current chapter if they have the feeling of having missed something important. They can also skip forward again and catch up to the "live edge" of the stream.
- **Language Selection**
  Some items have multiple languages from which the user can select.
- **Foreground/Background Balance**
  In some items the possibility to control the balance between the audio volume of foreground and background objects was enabled. For instance the "Football Match" has the option to amplify the commentary while reducing the volume of the stadium atmosphere.
- **Immersive audio**
  Depending on the playback environment, different reproduction systems can be selected by the user. This could be headphone/stereo, headphone/binaural, speakers/stereo, and speakers/surround. This allows the user to experience different levels of immersion, and requires the audio codec to provide different rendering and downmix formats
- **Direct object manipulation**
  In the MPEG-H file, interaction features and ranges can be specified for individual objects. The can be as simple as a toggle to turn an object off (muting it), but the user can also position objects interactively in 3D space or select 3D rendering scenes where groups of audio objects may be adjusted in a relative or absolute way.

- **Loudness and Dynamic Range Control**
  Depending on the playback environment, it is sometimes desirable to compress the audio to different target loudness levels. For example, when being in a noisy environment such as a train, the user typically prefers to increase the loudness and reduce the dynamic range (without causing clipping or distortions).

3.4 Platforms/devices

Two platforms are targeted in Pilot 1 Stage B: the ORPHEUS Radio iOS app and the Trinnov Altitude 32 AV receiver.

The iOS app, which has a version for iPhone and for iPad, receives MPEG-H streams over MPEG-DASH. The audio can be rendered to headphones in stereo and binaural, to loudspeakers using AirPlay, and to multi-channel surround systems (so-called “5.1”, and “7.1”) using a wired HDMI connection.

On the AV receiver, a specific compilation of the GStreamer open source library was integrated to receive the MPEG-DASH stream and pass the MPEG-H to the main signal processing operator of the unit. Then, a signal processing operation uses the FhG decoder to render the MPEG-H to a 4+7+0 loudspeaker layout (that is, four above head height, seven at head height, none below head height).

3.5 Audience

Stage B was a closed group stage of the Pilot and its audience included both the employees of the various project partners as well as selected workshop participants. A broader public could experience the results of Stage B at fairs (IBC, and Medientage). Combined with the test subjects of the JOSEPHS and BCOM user studies, several hundreds of people have experienced the results of Pilot 1 Stage B.

3.6 Feedback

Two end-user studies were carried out using the results from Pilot 1 Stage B, one by BCOM and one by FHG at the JOSEPHS lab; extensive documentation for these studies will be made available in *D5.6 Report on Audio subjective tests and User tests*. Additional feedback was gathered in a non-formal way at the ORPHEUS Workshop in London, the IBC and Medientage tradeshows and from individual experts at the various project partners. Overall, the concepts and technology were well received: visitors at the booths indicated that they understood and appreciated the functionalities and were able to manipulate the app autonomously.

Figure 13: ORPHEUS demo at the EBU IBC stand

4 Pilot Phase 1 - Stage C: live-encoded MPEG-H

This final stage aimed at an object based live transmission based on MPEG-H. The main difference from Stage A (see Section 2) was that MPEG-H was used as an emission codec, instead of AAC with ADM metadata. Compared to Stage B (see Section 3), the main difference was that the production and encoding was operated in real time, i.e. "live". Hence, Stage C combined the features of the stages achieved previously and aimed to be the most complete integration scenario of Pilot Phase 1.

The overall implementation architecture of Pilot 1 Stage C is illustrated in Figure 14 and explained in more detail below. Before that, we first describe the live production which was used to demonstrate Stage C in the trial.

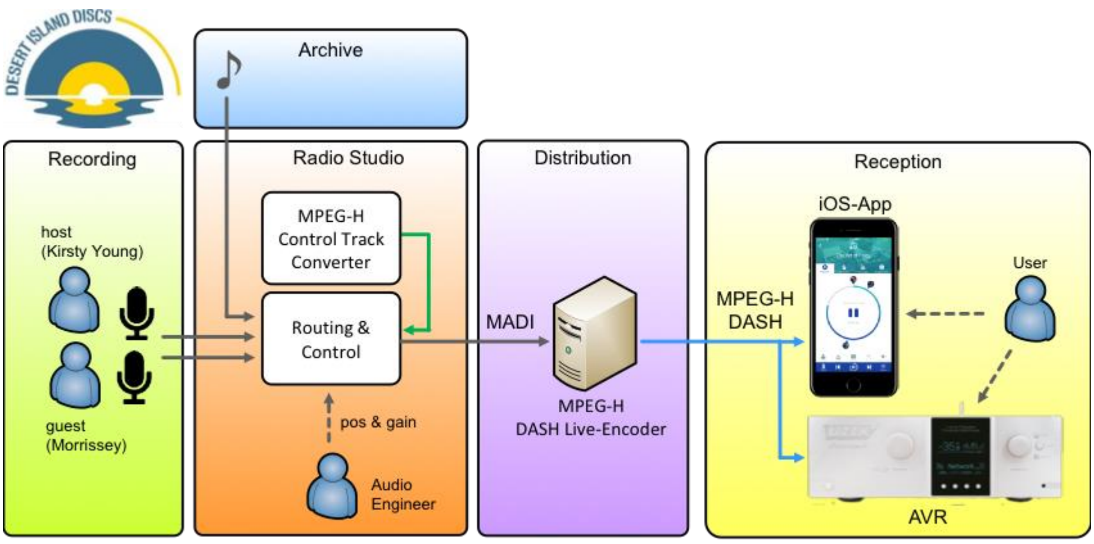

Figure 14: Overview of Pilot Phase 1C Implementation Architecture

4.1 Live Production: Desert Island Discs

In order to demonstrate the scope of Pilot 1 Stage C, the following live event was produced. The scenario was inspired by a radio show of BBC Radio 4 called **"Desert Island Discs"** in which the host (Kirsty Young) interviews invited guests (here, Steven Patrick Morrissey) to share the soundtrack of their life by asking, "Eight tracks, a book and a luxury: what would you take to a desert island?".

To reproduce this scenario, two actors read the transcript of the show, one taking the role of Kirsty and the other the one of Morrissey. Their voice was recorded live in the studio and ingested as two audio objects in the IP Studio system. The position and gain of the readers' voice could be controlled live by the sound engineer during the show. In addition, the concept of the show allowed alternation between the spoken word and the selected music items. This was used to replay archived material, and an immersive (4+7+0) recording of a classical concert was used for the pilot.

Though the demonstration scenario follows a relatively simple structure, it combined channel-based and object-based audio in a realistic live production scenario. Hence, it was sufficient as a proof of concept. The architecture implemented, however, could have supported more complex scenarios as well.

4.2 Platform and Devices

4.2.1 Integration in IP Studio

In Stage C, the signal pipeline in IP Studio was configured to support the live production scenario outlined above. The two microphone sources are carried as AES67 streams from Axia xNodes, while the pre-recorded sources (music) are carried as RTP streams entirely within IP Studio. The audio mix (pan and gain) is controlled with metadata, which is carried within IP Studio using UMCP and received from the live production interface over web sockets. It is converted into MPEG-H metadata and embedded into an "MPEG-H Control Track" for carriage over a MADI interface to the "MPEG-H DASH Live Encoder" (see below). Because the structure of the audio scene was relatively simple, it was not necessary to implement a fully dynamic meta-data handling system capable of dealing with any possible object hierarchy or level of interactivity. Instead, knowledge of the nature of production was used to initialise the renderer component when the pipeline was started. More specifically, the audio scene used consisted of an immersive 4+7+0 channel bed with two dynamic audio objects, which was maintained throughout the live transmission without any configuration changes.

Fraunhofer provided a dedicated PC with an external MADI card for live audio input (RME MADIface, 128-Channel MADI USB interface). The PC ran the *MPEG-H DASH Live-Encoder*, which implements multi-channel audio input, MPEG-H encoding and the generation of DASH content, consisting of MP4 fragments (fmp4) and the Media Presentation Description (mpd). The DASH content was then served from a local HTTP server, which connected to the Local Area Network (LAN) at BBC. The clients (elephantcandy iOS-App, Trinnov AVR) accessed the content via a Wireless LAN access point.

The input of the MPEG-H DASH Live Encoder was provided from the IP Studio and included 15 channels of uncompressed PCM audio, transmitted over MADI. One critical point for this architecture is the transmission of audio-related metadata. For that purpose, Fraunhofer has developed the *MPEG-H Control Track* (not within ORPHEUS), which allows the transport of MPEG-H metadata over a PCM channel using a data modem. This Control Track was carried over the 16th PCM channel. The important point of the Control Track is that object-based audio can be transported over legacy broadcast infrastructure, such as MADI or SDI. It therefore allows an easier integration of OBA into existing workflows and studio infrastructure. In the case of BBC, the same 64-channel MADI infrastructure which is used to connect the loudspeakers in the studio could also be used to interface to the MPEG-H Live Encoder.

4.2.2 Reception Devices

Two reception devices for the MPEG-H stream were used in Stage C: The iOS-App developed by elephantcandy and the AV Receiver developed by Trinnov. For the iOS-App, only minor changes had to be implemented compared to Stage B and we therefore refer to Section 3 for the description. The only difference in the user interface was that an additional content "Live from the BBC" was available for starting the live stream.

The integration of MPEG-H decoding and DASH reception into the AV Receiver by Trinnov was successfully completed for the Stage C pilot. The result was a fluent reception and decoding of the streamed content, and a correct rendering to the selected loudspeaker configuration. A GUI running on a tablet was available for interactivity. The GUI is constructed automatically when the streaming starts and allows adjustment of stream-specific parameters of MPEG-H in real time. The GUI is using web technology and can be accessed (after authorization) from any browser that connects to the WiFi Access Point of the AV receiver, see Figure 15.

Figure 15: Trinnov-AVR Altitude32 with Web-GUI running on a tablet

4.3 Audience

Unfortunately it was not possible to make Stage C available in a broad trial as initially planned. However, the production was streamed live using MPEG-H during the review meeting on Dec. 13th 2017 with a representative of the European Commission and independent assessors being present, see Figure 3. A short documentary video from this event is available on the ORPHEUS website[5].

Both reception devices were demonstrated to a broader audience during the 2nd ORPHEUS workshop (May 15th, 2018).

In addition, Trinnov publicly introduced and demonstrated their *Altitude32* with the integrated MPEG-H decoding options on the High End trade show in Munich between May 10th and 13th 2018.

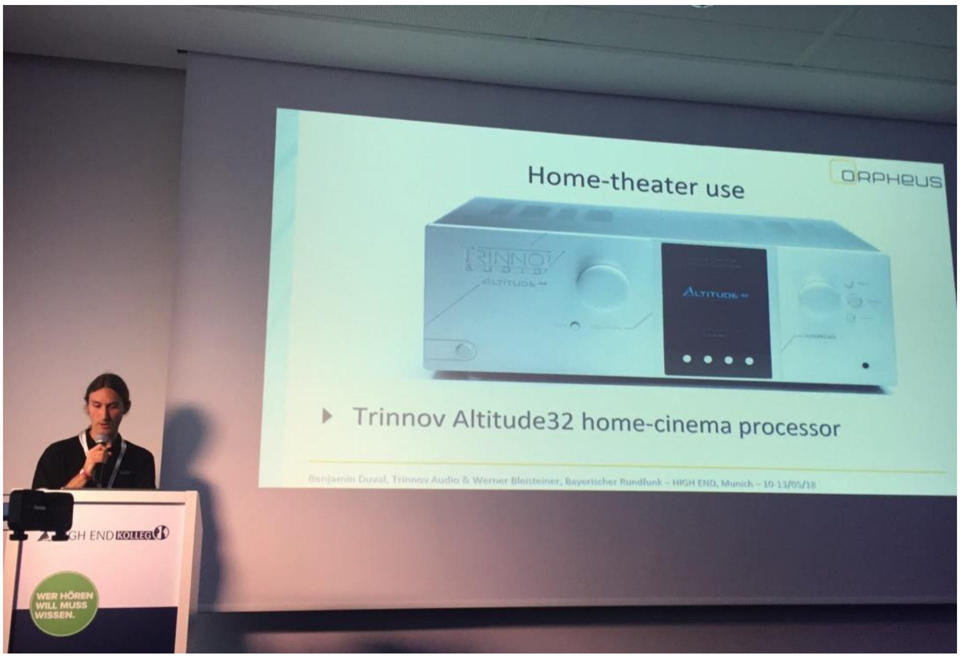

Figure 16: Trinnov presentation at High End Munich

[5] https://orpheus-audio.eu/videos/orpheus-demo-ec-review-2017-12-14-binauralsound/

5 Pilot Phase 2

For the ORPHEUS project, Fraunhofer IIS, BBC R&D and BR created and provided a total of 8 productions for demonstration of the various new possibilities and features of object-based audio[6]. The pieces are representative of typical examples of radio broadcasting genres.

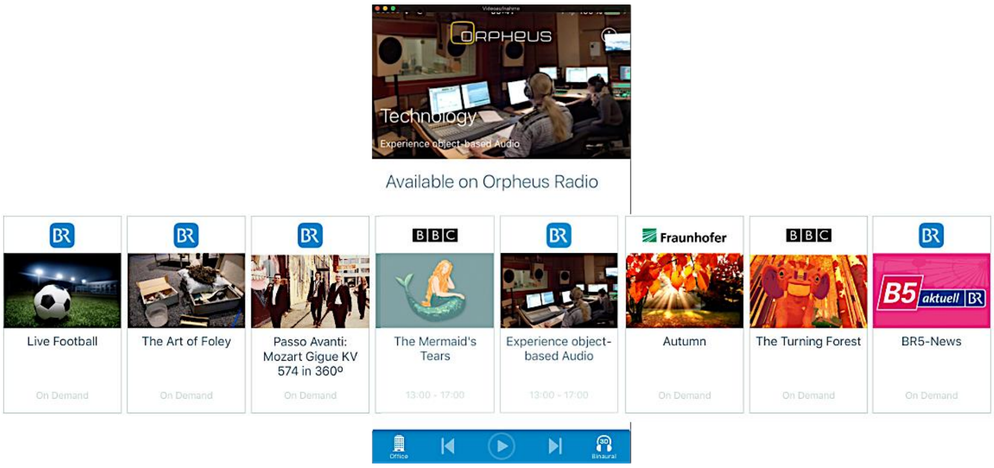

Figure 17: Overview of all ORPHEUS pilot productions in the app

The following table gives an overview of the main features:

Title	Genre	P.	Language	OBA features	Interaction	Transcript	Variable-length
Heute im Stadion (Live Football)	(live) report	BR	DE	FG/BG DRC 3D Audio	---	DE	LOI 1-2 3 segments
Die Kunst des Geräuschemachens / The Art of Foley / L'Art de Bruitage	radio documentary	BR	DE / EN /FR	FG/BG DRC 3D Audio	---	DE /EN /FR	
Passo Avanti: Mozart Gigue KV 574 in 360º	music	BR	./.	FG/BG DRC 3D Audio	4 positions for instruments	./.	
The Mermaid's Tears	radio drama	BBC	EN	FG/BG DRC 3D Audio	3 perspectives of the story	no	
Erlebe objekt-basiertes Audio / Experience object-based audio	journalistic contribution	BR	DE / EN	FG/BG DRC 3D Audio	---	DE /EN	LOI 1-5: 43 segments
Herbst / Autumn	soundscape	FHG	DE /EN	FG/BG DRC 3D Audio	audio-object prominence and positioning	DE / EN	LOI 1-2: 13 segments
The Turning Forest	radio drama	BBC	EN / DE (BR)	FG/BG DRC 3D Audio	---	---	
B5 aktuell Newshour	news program	BR	DE	DRC			LOI 1-5: 183 segments

Table 1: Overview of ORPHEUS pilot productions and features implemented

All of these productions were provisioned for on-demand consumption in the ORPHEUS app. All of the features described are available in it[7].

[6] The credits of all pilot pieces of the ORPHEUS project are shown in Appendix A

[7] Note: the ORPHEUS receiver from Trinnov, the Altitude 32 AVR, at present can only reproduce object-based audio features implemented within the MPEG-H audio codec, i.e. language selection, foreground/background volume

5.1 Concepts for advanced on-demand consumption

5.1.1 Features and functionalities

In the course of the developments of the ORPHEUS iOS app for Pilot Phase 1B and 1C the most essential usability features for users to access and handle object-based audio on-demand streams had already been implemented.

- **Browsing and Tune-In**
  The user can browse various content pieces, getting an overview with titles, pictures, and a short description of what is presently available, make a selection, and start playback.
- **Skip-Back and Catch-Up**
  The user can navigate within the selected program item using skip bw/fw buttons.
- **Language Selection**
  Four of the ORPHEUS object-based audio pilot productions offer the possibility of switching language between English and German ("The Art of Foley" offers French as third option). In general, the default playback language corresponds to the selected OS language, if available. Otherwise it is set to English.
- **Foreground/Background Volume Control**
  All of the ORPHEUS demonstration productions with spoken word offer individual adjustment of the balance of foreground speech and background noise/music.
- **Loudness and Dynamic Range Control**
  Depending on the playback environment, the audio may be compressed to different target loudness levels and ranges. For example, when in a noisy environment such as a train, the user typically prefers to increase the loudness and reduce the dynamic range (without causing clipping or distortions).
- **Persistence of User Settings**
  The variety of choices for preferences, parameters or connections can be stored in 'Profiles', stored on the device. These profiles are at present still activated manually but this can be automated. The set of parameters in a 'Profile' persists when changing the content – but may of course be readjusted, if necessary.

Although not being 'essential', some more of the features and additional functionalities we had defined as being for "advanced on-demand consumption" were also able to be implemented during the final stage of Pilot Phase 1B and 1C:

- **Dial wheel navigation**
  A 'dial wheel' displays segments of the program with assigned markers for 'points of interest' or chapters.
- **Additional Info on Programme Section**
  Clicking on a marker opens an overlay window providing additional information on this specific content segment, along with an illustrating picture. If desired, the user may jump directly to this segment by hitting "replay section".
- **Transcript with speaker recognition and depiction**
  Any of the pre-produced ORPHEUS pilot pieces offer transcripts of the spoken word. Apart from helping to make content more intelligible and comprehensible, this may also be used

---

adjustment, dynamic range control, 3D audio. Other functionalities created within the macro block 'Presentation Design' in the reference architecture – see Figure 1: ORPHEUS Reference architecture Infrastructure - are not implemented.

to navigate within the content to certain points of interest: the audio playback point follows the position of a click in the transcript.

- **Audio playback with variable length**
  Within the concept of object-based audio - with content being assembled according to rules provided by additional metadata - a functionality inconceivable with legacy technology seems to be feasible: to make audio adaptive in length.
  This offers the user a choice. They may decide how much time they want to spend on a topic, or how deep they want to go into a certain subject. From another perspective, this issue may be regarded as if various 'edits' or 'versions' of a program can be made available. We certainly found this one of the most challenging tasks – yet feasible, given a pragmatic approach, and basing it upon inherent concepts of media creation.

5.1.2 'The Inverted Journalistic Pyramid' as model for variable-length and depth

Anyone working in journalism or public relations sooner or later gets to know the 'inverted pyramid'. It is the most common communications model to describe how text information is structured effectively. Especially in news journalism the most important guideline is to 'get straight to the point'.

It supposedly emerged with the rise of news dissemination with technologies like the telegraph, and the telephone. Due to their early unreliability with regular breakdowns, the most important things had to be put first, but this also kept communications costs low, and saved time. So, this led eventually to a "telegraphic style" for newspaper writing, too.

The minimum requirement of this model is a 3-tier inverted pyramid, comprised of

- a lead paragraph on
  who, what, where, when, why, and how something happened
- a body paragraph, offering one or more arguments, discussing questions, telling a story, or more issues on the subject
- a tail, giving additional information or pointing towards further related stories or news items.

Figure 18: Basic 3-tier 'inverted pyramid'

Today, this principle is omnipresent in all kinds of media.

- Newspapers, online news and blog posts
  - Big printed headlines, with condensed, often fragmented and sharp wording
  - Bold printed lead paragraphs providing the most important information in full sentences, serving also to draw the reader into the story so that he continues reading the full article
- Radio and TV News
  - Show Opener with headline overview
  - News anchor moderating or introducing the item or a
    News Reader reading a news bulletin (written in the pyramid news style)
  - Reporter on location providing details and background or even commenting

Moving down, the information provided is gradually enriched, leading more and more to "the full picture" of a story – but in return it gets more and more time consuming, given the fact that reading or listening and watching are linear perceptual processes. Still, any of the tiers may offer 'breaking points', if the recipient feels the need to leave the narration by having 'had enough' or having 'no more time'.

So, time is inversely related to depth of information.
We have studied and discussed this model in detail and found it appealing to investigate whether we could apply it and combine it with ideas of, and proposed technologies for, variable-length productions in object-based broadcasting.
We are well aware that presently there are other promising approaches based upon technologies like automated data extraction, deep learning or artificial intelligence. However, these are not fields of our own expertise and to include them would have overwhelmed the ORPHEUS project. Also, being "close to real-world workflows and applicability" is one of our credos.

Last but not least, this principle is already well known to broadcast journalists, as they work with it - consciously or subconsciously – every day. So it's worthwhile to "first keep it simple", by taking established practices and workflows - but still gain an exciting new feature for interactivity and personalisation. Not to mention that this still leaves all the crucial decisions of creating content completely in the hands of the responsible journalists and editors.

5.2 Suitable content and implementation of variable-length

5.2.1 *Experience Object-Based Audio*

It is obvious that a realization of the variable-length functionality described above depends on suitable content. From a mere technical and production perspective it seems fairly easy to provide such a feature using edit decision lists (EDL) to create alternative rundowns. Yet, as the content is spoken word, the meaning and semantics transported is at risk of losing its comprehensibility if this technical solution is applied only according to technical rules. So, the semantics of the piece have to contain a good amount of hierarchical logic and the way that the information given has been structured. As was said above, the 'inverted pyramid' often becomes an internalized rule of journalistic writing, and good journalists obey it by instinct.

This was definitely the case when BR scientific reporter David Globig conceived his piece "Experience object-based audio" ("Erlebe objekt-basiertes Audio") that he has been originally commissioned to create exclusively for ORPHEUS. David Globig is a long-term and widely-experienced author of scientific reports at BR. He is renowned for being able to explain complex matters with ease – and he has a special fondness for audio subjects too.

Although not having been constructed initially for demonstrating the 'variable-length' feature, "Experience object-based audio" appeared, at a glance, to be an ideal object of study for it.

As a first step, the manuscript of the piece was analysed to identify the hierarchy of information in it, and the possibilities of assigning different "levels of importance" (LOI). Early on, it seemed appropriate to assign not just 3 tiers – *Lead*, *Body*, *Tail* – but to extent the basic model with two more tiers, making 5, to enable more distinction and granularity[8].

[8] As in typical uses of the *Mean Opinion Score* (MOS) in audio evaluation, a 5 level categorization should allow distinct decisions even in complex tasks. If not needed, it may easily be collapsed down to fewer categories.

Figure 19: 5-tier 'inverted pyramid' applied for the ORPHEUS concept of variable-length

The so-called metadata includes the position in a given space where the sound is supposed to be coming from – and how loud it's supposed to be. When the sound is played back, the system specifies and calculates which speaker is to be used for output of the signal in order to produce the desired audio impression.

In one cinema, a sound might be directed to speaker Number 23, installed obliquely in front on the left. In another cinema, it might be speakers 17 and 18 that cover the sound from that particular spot.

So what about sound reproduction on headphones? Object-based audio works there, too. The trick is simply to lead the ear, or rather the brain, to believe the sound is coming from several separate speakers.

But first you have to know how we actually hear, from which direction sound comes.

*STATEMENT Weitnauer – German Original, ducked Voiceover:
For one thing, there's the so-called 'interaural time difference'. In other words, if I imagine an object is slightly to the left of me, then the sound reaches my left ear a little before my right ear. Simply owing to the fact that it takes a little longer to get there. And then there is / the 'interaural level difference'. A slightly higher volume reaches my left ear than my right.

These differences can be artificially added to the reproduction of a sound object for earphones.

This is enough to render the difference between left and right aural perception.

It's more complicated, however, for up and down, in front and rear. This is where the shape of our pinna comes in.

Stephan Peus spent many years working on so-called "dummy head microphones" at Georg Neumann GmbH, a company based in Berlin.

*STATEMENT Peus – German Original, ducked Voiceover:
Our pinna 'bend' the signals quite heavily. The auricle is folded in a pretty complex way and fitted out with hollows or cavities. This gives a very specific quality to sounds coming from different directions. And these differences, these aural, or what we call 'spectral', differences are detected and processed in the brain. Which enables us to determine whether the signal is coming from the front, the side, above or from somewhere else.

This, too, can be digitally imitated.

*STATEMENT Peus – German Original, ducked Voiceover:
Present-day computer technology is so sophisticated and processing speed so fast that we can recreate, we can simulate, 'head-related transfer functions', as they're called, in other words the way the outer ear influences incoming sound signals. And we can impart that influence to the signal.

Ideally, when reproduced through headphones, the whole thing then sounds as though we were actually hearing the object-based soundtrack through dozens of loudspeakers in a cinema.

(So) how should we picture object-based Audio?

It's actually something we know from everyday life: A car driving by in the street is an object that makes noise. It moves amongst other objects that make noise too – and that may be moving too: pedestrians talking to one another, a cyclist, leaves blown by the wind across the asphalt.
They are all discrete elements, separate individual objects.
And a sound engineer or sound designer can put together a whole scene from such separate individual elements - in a virtual space.

Figure 20: Second 2 page of the manuscript of "Ex-OBA" with assigned LOI in colour

Unlike a typical 'news item', that is strictly laid out top-down, this piece proved to be more of a narration, starting with an introduction, then outlining of the 'issue', then 'getting to the point' and eventually, garnishing the findings with some nicely explained and decorated examples, before finishing with an outlook. Nevertheless, this sort of dramaturgy can be assigned with 'levels of importance', creating a total of 43 single segments.

Transferred (for easy recognition here colourised) onto the narration track of the piece, and unfolded in the 5 levels*,* Figure 21 depicts the centralised structure of the piece and makes a 'non-inverted' pyramid shape. The run times go from 1:05 up to full length 13.01. Note: the 'mute' parts in blue denote the statements from interviewees.

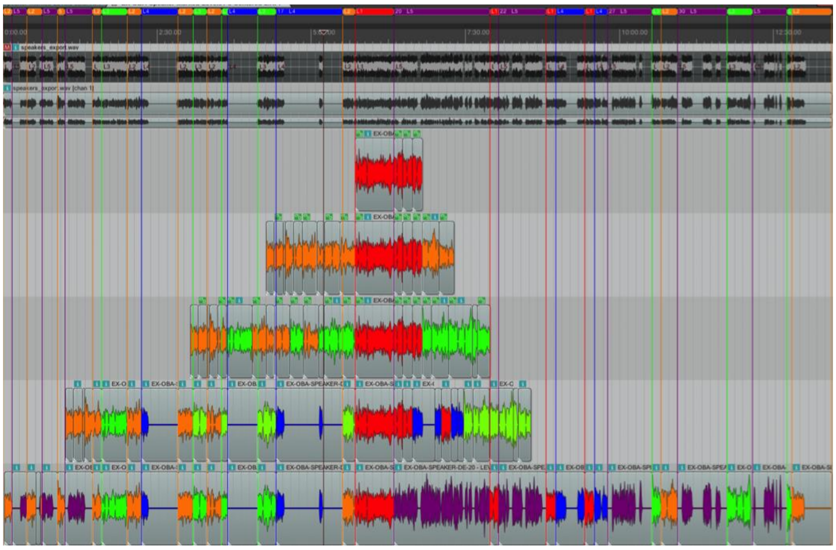

Figure 21: narrator track of Pilot piece 'EX-OBA" structured and unfolded in LOI 1-5

It is important to emphasize that the logic of the LOI principle applied here is an "unfolding" one, i.e. any level above 1 includes the lower-numbered levels: L1=L1, L2=L1+L2, L3=L1+L2+L3, etc.

5.2.2 News Hour from BR's "B5 aktuell" news radio channel

An additional experiment was conducted, trying out the applicability of the LOI concept with classic news radio content. For this purpose, a clean feed - also known as an 'air check'[9] – of one hour of the B5 aktuell was recorded, analysed, and processed[10].

B5 aktuell, established in 1991, is BR's renown 24 hours radio news channel. It operated throughout according to a well-defined 'hour clock' in a clear and accurately executed structure. One hour on *B5 aktuell* is comprised of:

- Moderated news – a news editor presents the top news plus with short bulletins from correspondents (mins. 0:00 to 8:00)
- Bavarian news – with Bavarian news editor and local correspondents' bulletins (mins. 8:00-13:00)
- Stock news (mins. 13:00-14:00)

Figure 22: B5 aktuell news hour clock

[9] Such "check" recordings are routinely made for use in subsequent investigation of problems or complaints and compliance with regulations

[10] The recording was made on December 2017 from 15:00 to 16.00 CET

- Traffic (mins. 14:00-15:00)
- Classic news by news reader (mins. 15:00-20:00)
- Background report from correspondent (mins. 20:00-25:00)
- Cultural News with news editor and contributions
  (mins. 25:00-28:00) etc.

Here, the 5-tier model of the LOI concept was also applied, with the different elements assigned as:

1. Headline
2. Bulletin – lead paragraph
3. Story explained - background
4. Statement – broader picture
5. Additional information – full extent

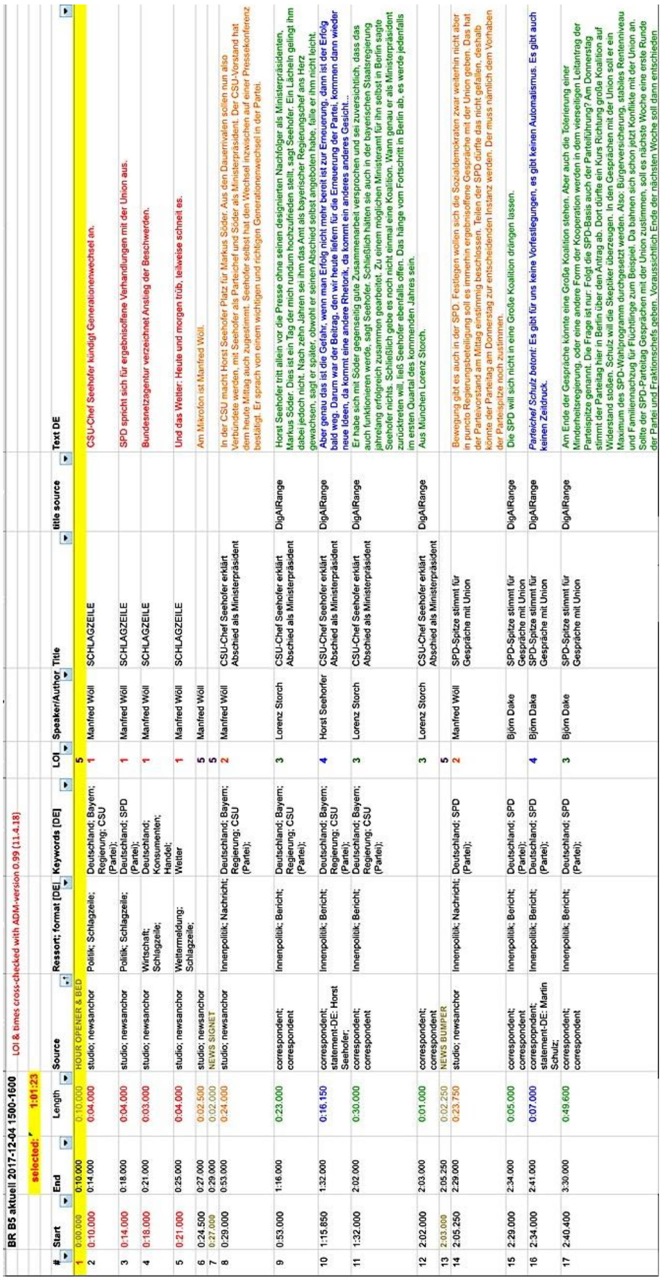

Figure 23: Analysis of B5 aktuell news hour

In this way, the 60-minute program was divided into 183 segments, generating variable-length options – without any further subject or keyword selections - from 2 minutes upwards.

5.3 Process and practicalities

5.3.1 Content conceptualisation

Many of the ideas for the content that BR contributed to the ORPHEUS pilot phase 2 were sketched during or in the course of a “sprint” in the internal creative workshop “NEXT AUDIO” with creators, editors producers and sound engineers held on June 1st and 2nd 2016. Below is some background on the conceptualisation and genesis of BR’s pieces: the idea is, “content is crucial”.

Figure 24: NEXT AUDIO workshop 2016 at BR Studio #9

Football Match

A group of colleagues, partly from BR’s Sports department, worked on a draft of a concept called “Liga Live 4.0” to enhance the weekly Saturday afternoon “Bundesliga-Konferenz” (German 1st League), using possibilities of object-based broadcasting. This included:

- Selectable audio reproduction format, i.e. binaural for headphone listening
  - ➢ at present all such radio transmissions are only in mono
- Individual pre-selection of favourite matches to follow
  - ➢ at present the content is ‘directed’ by the broadcaster
- ‘Goal alarm’ from parallel matches with the possibility of switching, and skipping back to recap on the scene
- Additional information on players, background stories, and related statistics

The realisation of such an ambitious concept would certainly have not been possible within the ORPHEUS project. But it proved that media professionals understood immediately the scope of the new opportunities of the object-based media approach. So within our abilities we decided to pick at least one of the users’ requests, the ‘immersive audio’ aspect, and work it out in order to create and evaluate an example for a ‘football on radio’ experience as it has not been done before. Capturing (re-enactment): January 26th 2017, object-based mastering: July 2017.

The Art of Foley

One of the participants in the NEXT AUDIO workshop, Bernhard Jugel, also works as a producer and director for BR’s *Radio Documentary and Features Department*. Early in 2017 he started to work on a project on the history of The Art of Foley along with foley artist Max Bauer and actress Andreas Kilian and, pertinently assuming this being a worthwhile, if not ideal, subject for an object-based production, offered ORPHEUS to collaborate.

Figure 25: The Art of Foley production

The main idea was to use the object-based production in order to create a 'binaural' version of the 1-hour documentary in addition to the regular stereo mix. As a good deal of the quotes in the story where originally from English and French sources, the option of creating an additional abridged ORPHEUS version in three languages became obvious, as well as applying and implementing further typical object-bases media features for demonstrating in the pilot phase 2[11].
Capturing and mix: April 3rd to 12th 2017 – Additional language recording and editing: June 15th-17th 2017, Object-based mixing: May 9th-10th, Object-based mastering: August 2017.

Passo Avanti: Mozart Gigue KV 574 in 360º

The Munich-based ensemble *Passo Avanti* likes to take on artistic experiments applying new technologies. Having heard about the ORPHEUS project, they wanted to know if object-based audio would enable 3D Audio for Virtual Reality and 360º video, especially for a new jazzy arrangement they had created of Mozart's piano piece, the Gigue KV 574. During rehearsals we found out that this particular version would ideal for exploring the various capturing, editing and mixing challenges for object-based music production. Being clearly laid out and straightforward in its musical scoring too, it is ideal for examining 3D audio perception questions.

Figure 26: Passo Avanti recording the Mozart Gigue KV 574

[11] binaural rendered version in German, English and French of "The Art of Foley" are available on ORPHEUS community at ZENODO: https://zenodo.org/record/1203365#.WtdKYcgaQxM

Capturing: January 28th 2017- Mixing: July 26th-27th, Object-based mastering: November 2017

Experience object-based audio (EX-OBA)

This piece was conceived to explain and demonstrate some basic principles and applications of object-based audio so to speak “in medias res”. BR science reporter and presenter David Globig, who also participated in the NEXT AUDIO workshop, took up the challenge of creating a concise and comprehensive, “explanation piece” – typical radio – with the usual elements, text, statements, background sounds and music and writing a story around it. The original German version went on air in BR’s Bayern2 Science Magazine in June 2017[12].

Figure 27: David Globig and Christian Schimmöller working on EX-OBA

Capturing: January 25th-27th, mixing: June ,
Object-based mastering 1st version : September 2017
Pilot Phase 2 conceptualisation: October 2017
Object-based mastering 2nd version: March 2018

B5 aktuell News Hour

The material for this demonstration was kindly provided from an internal “air-check” recording and reassembled and edited to full length.

Capturing: December 4th 2017
Editing: January 2018-04-19
ADM finalisation: March 2018

[12] binaural rendered versions in German and English of “Experience Object-based Audio ” are available on ORPHEUS community at https://zenodo.org/record/1203361#.WtdRj8gaQxM

5.3.2 Recording and pre-production

It is important to note that at the time BR started to collect, capture, and produce the material for the pilots – January to April 2017 – neither the Sequoia DAW nor any other of the ORPHEUS tools, were available in a condition that could be already used for a production or to create the object-based audio production meta-data format ADM.

As outlined before, most of BR's pilot productions were initiated for, or created in collaboration with, programmes that went on air. Also, BR production studios and facilities are strictly administered in order to maintain continuous and reliable operation. Installation or use of non-approved hardware and software in regular production workflows is not allowed. Therefore we had to find ways and means to work around this, in order to be able experiment with object-based audio tools and production environments.

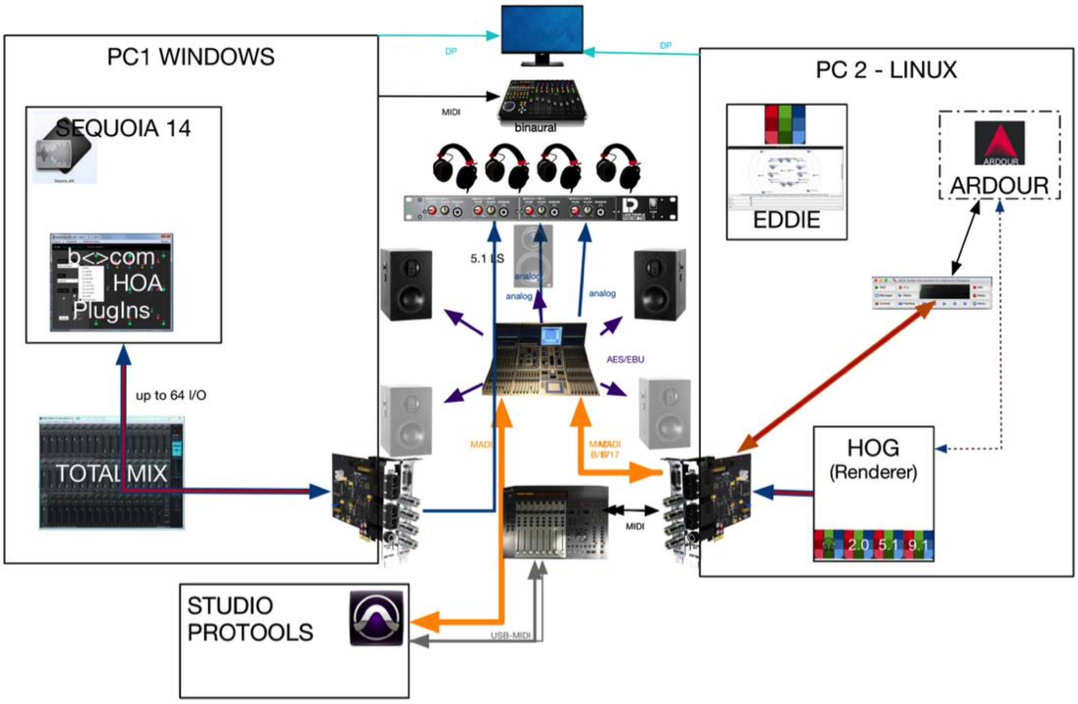

Figure 28: ORPHEUS OBA-TEST-RACK PCs' connections to standard BR studio

We took a "hybrid approach" by attaching a specially designed 'ORPHEUS OBA Test Rack' to the existing equipment. This consists of 2 standard (but low noise) PCs:

- PC1 operating Windows with MAGIX Sequoia including b<>com HOA Plug-ins
- PC2 operating LINUX with IRT's experimental OBA production system (ARDOUR optionally for standalone DAW playback operations)
- each PC has a RME HDSP MADI FX sound card inserted

Via MADI and MIDI we were able to hook up and synchronise the two machines to any of our typical studio equipment as peripheral or out-board gear.

So for all BR pre-productions in BR Studio #9 we used the *ProTools* DAW installed there for recording (BWF 48 kHz, 24bit) and basic editing tasks. The basic tracks arranged on the timeline were defined as 'audio objects' by assigning them to a maximum of 32 busses, routed via MADI to the outboard the IRT production system. In there, they were panned in 3D and spatially rendered by the HOG in binaural, stereo or 5.1 surround sound during the mixing process. Eventually the rendered signal was re-routed back to the studio *ProTools*, to be recorded as "pre-rendered" versions (binaural, stereo mixes or 5.1 versions)[13] for regular channel-based broadcasting play-out.

5.3.3 Finalisation as ADM in MAGIX Sequoia

The process to create the genuine object-based ADM production format could only be commenced when the tools developed by the ORPHEUS partners were available. For this purpose, the original 'hybrid' production sessions from ProTools and the IRT experimental production system were combined and transferred at IRT.

Figure 29 shows the imported BWF+ADM of the "Experience Object-based Audio" demo piece in MAGIX Sequoia. It consists of:

- 2 grouped tracks for the narrators (DE + EN) – green frame
- 7 grouped tracks for the interview statements (DE + duckings for voice-over in EN)
- 32 tracks for ambience and music objects used for the background

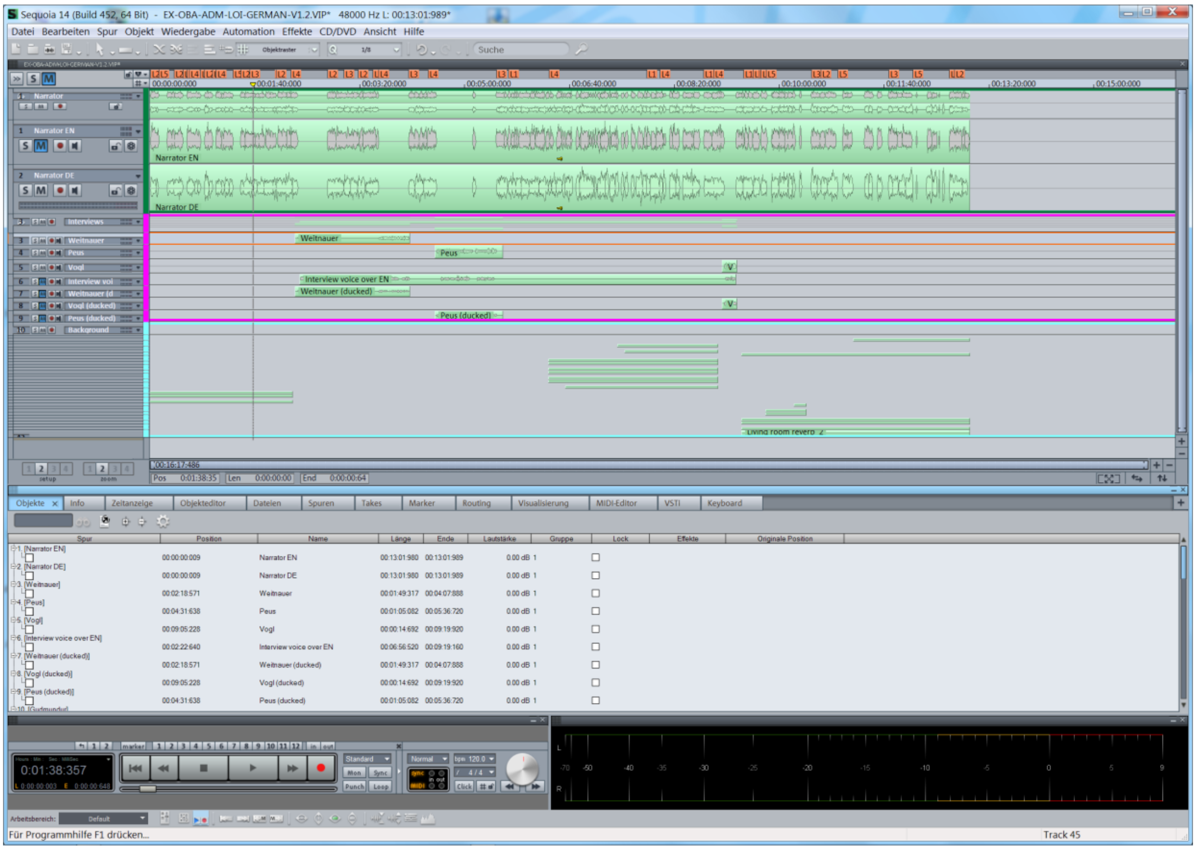

Figure 29: EX-OBA production imported as ADM in SEQUOIA with LOI markers

Above the timeline there are markers labelled L1 to L5 that were manually inserted to define the segments for the variable-length function. These time stamps are eventually exported into the EBUCore[14] metadata format field

<technicalAttributeByte typeLabel="ContentLevel">N</technicalAttributeByte> (N=1-5)

that is enclosed, along with the ADM metadata, within the BW64 audio file.

[13] screencasts of EX-OBA and FOLEY illustrating this workflow are available from IRT's LAB website: https://lab.irt.de/bringing-object-based-audio-to-the-production/

[14] *EBU Tech 3293 EBU Core Metadata Set* https://tech.ebu.ch/docs/tech/tech3293.pdf

In a same way, the one hour news program from B5 aktuell was finalised as BWF+ADM. Here, there is only one foreground audio object for speech, plus two audio objects for the occasional sound design elements in stereo.

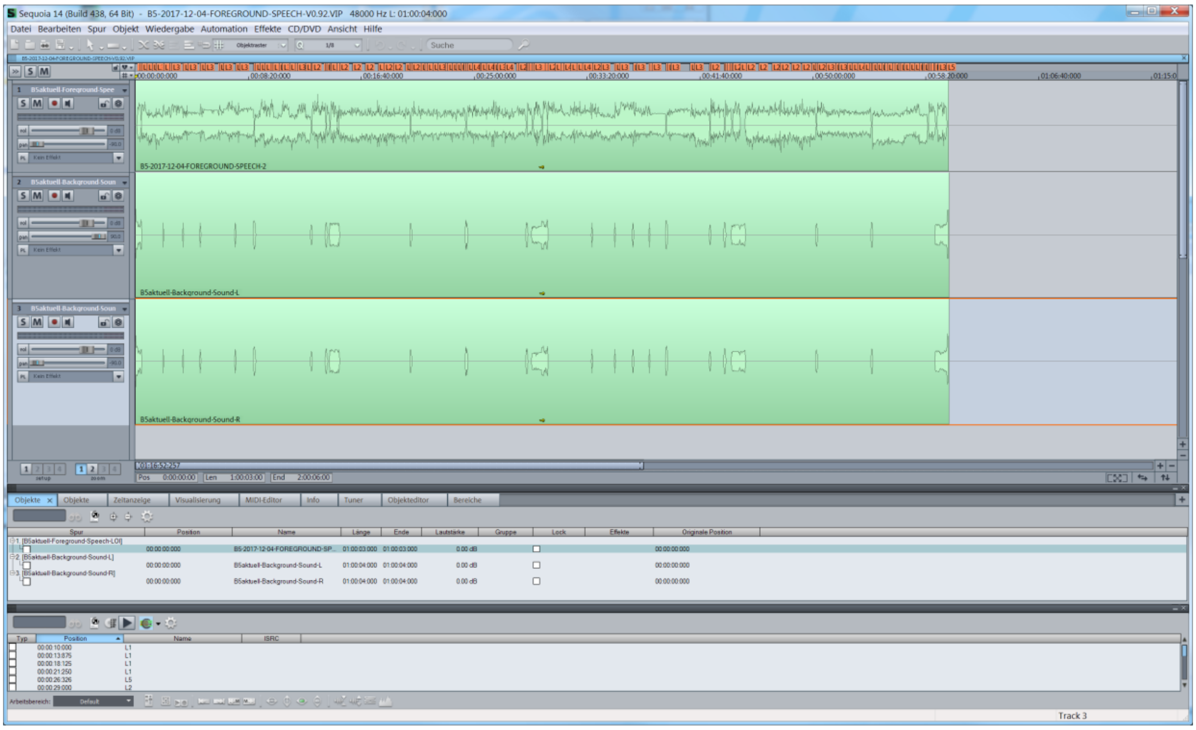

Figure 30: B5 aktuell news hour voice and background tracks with LOI markers in SEQUOIA

5.3.4 ADM Pre-processing (IRT)

The "Experience Object-based Audio" demonstration, as produced, features a total of 42 individual objects, where 37 are needed for the German version and 38 are needed for the English version. While this level of detail and additional information – including the grouping and annotation of the individual objects – provides valuable details e.g. for further (post-)production steps or archiving purposes, it exceed the limits imposed by the MPEG-H low complexity profile level 3.

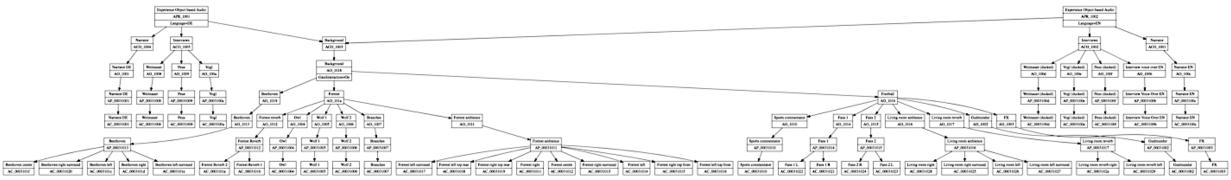

Figure 31: Schematic overview of all elements of the original "EX-OBA" production.

To be able to encode the demonstration piece with MPEG-H, the original ADM metadata and audio had to be transcoded into a low complexity version, again as ADM metadata, within a BW64 audio file.

This process is not straight-forward, as the "Experience Object-based Audio" piece uses and combines a wide variety of features of the ADM and object-based audio in general:

- Multiple languages, represented as distinct *AudioProgrammes*. Both languages must still be available in the transcoded version
- Content and objects shared between *AudioProgrammes*
- GainInteractivity to control foreground/background balance: The transcoding process must ensure that parts used for foreground/background control can still be adjusted separately
- Immersive content: Parts of the demonstration make extensive use of 3-dimensional

positioning of audio objects to show the immersive potential of next-generation audio.

The most common strategy to reduce the complexity of an audio scene is to combine (or pre-render) a set of elements into a, so-called, bed, i.e. to render parts of the scene to a virtual loudspeaker setup that replaces the multiplicity of individual elements.

The elements from the “Experience Object-based Audio” demo can be put into three disjoint sets:

- All “background” objects referenced by the *AudioObject* with the *GainInteractivity* flag set
- All elements that belong to the German language version
- All elements that belong to the English language version

Furthermore, given the fact that the immersive features need to be represented if pre-rendered to a bed, the *4+7+0* speaker setup has to be chosen as the format for the bed. This is due to the fact that it’s the only speaker setup supported by MPEG-H that features 4 height speakers in addition to the 7 speakers within the horizontal plane.

Given the above constraints, the principles of combinatorial logic can be used to search for an optimal solution.

The final result is presented in Figure 32.

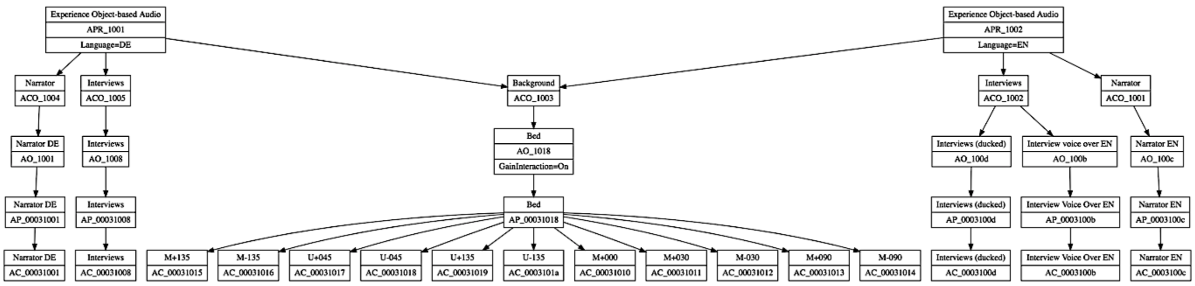

Figure 32: Schematic representation of the pre-processed “EX-OBA” ADM, including the pre-rendered interactive background element shared by both AudioProgrammes.

It can be seen that all the background elements have been pre-rendered into the 4+7+0 bed. Furthermore, the individual interview elements have been collapsed into a single “interviews” object for each language, which is easily possible as the interview parts are never active simultaneously.

Thus, the resulting BW64 audio file including ADM metadata could be encoded with MPEG-H while still containing all essential features of the original production.

5.3.5 Encoding into MPEG-H

Being the counter-part to Pilot Phase 1 - Stage B: as-live, the encoding of MPEG-H files also for Pilot 2 was executed offline by FHG. The source material was produced in various formats and therefore the MPEG-H conversion was mainly a manual process.

5.4 Platform/devices (EC)

The biggest challenge for the iOS app in Pilot phase 2 is the support for variable-length playback. In a mobile context, variable length makes a lot of sense as listeners are likely to be time-constrained, or may want to grab only headlines first and listen to a longer version or the entire programme later. On a smartphone, users are also expecting more control over their content, as opposed to conventional radio receivers where interaction is typically limited to just selecting a channel.

For the variable-length support, the app features a user-centric design that clearly states the amount of time the program will take at the current setting. The length can be reduced or expanded and a graphic shows which slices in time would be included for the selected playback length.

The variable-length LOI markers are supplied as additional metadata, streamed in parallel to the MPEG-H over MPEG-DASH. While the Trinnov AV receiver does not support variable-length at this moment, the transparency of the JSON metadata assures that the content produced in Pilot phase 2 can be played back on the device, just without the variable-length functionality

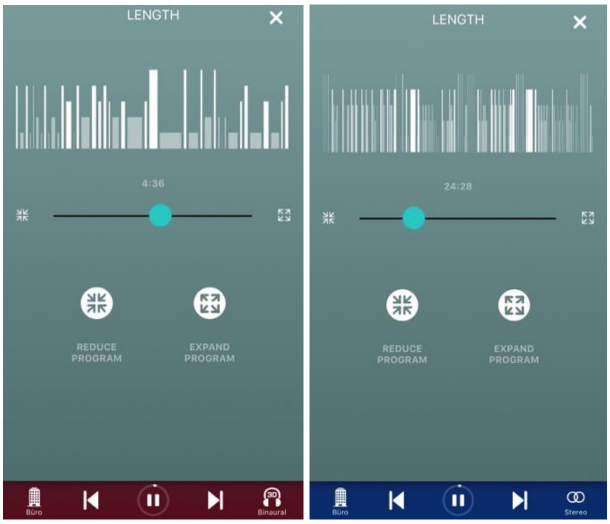

Figure 33: "EX-OBA" and B5 aktuell newshour variable-length screen

5.5 Audience

The version of the ORPHEUS Radio app made for Pilot phase 2 was distributed via Apple's iOS Business AppStore to employees and selected professional colleagues of the project partners upon invitation or upon registration to interested parties of the audio community via the ORPHEUS website.
While still a prototype, the goal is to expose a larger audience of mainly broadcasting professionals to the new concepts and content that was developed during the project.

5.6 Feedback

As explained before, Pilot Phase 1 Stage B is also to be considered as preliminary to Pilot Phase 2. Most of the key functionalities implemented in the ORPHEUS app – except the variable-length feature - were therefore already tested and evaluated during the public user tests at JOSEPHS in Nuremberg. Details are at presented still under preparation and to be published in the ORPHEUS *D5.6 Report on Audio subjective tests and User tests*.

6 For further study

The final representations of the pieces for ‘enhanced object-based audio for on-demand consumption’ in Pilot Phase 2 of ORPHEUS demonstrate impressively the advantages and possibilities of object-based audio.

Yet it has to mentioned once again, that most of them were created in a ‘hybrid’ way. In the beginning we were using tools, infrastructure and workflows in operating broadcast production studios and transferring them gradually during the parallel on-going development process for the various ORPHEUS object-based audio tools and implementations. So, there were, inevitably, some bespoke solutions to be designed and elaborated for certain functionalities in order, finally, to achieve these stunning results.

For all ORPHEUS partners, this was also a constant process of reflection on how to implement the means of metadata creation for the object-based audio format in such a way that in production, creators and engineers have a convenient software and GUI, in order to make use of typical object-based audio features. As we all are still at the starting point of object-based production, there is still a lot to explore and develop.

Here are some examples from our specific experience that need further study and improvement:

Use of EBUCore within BW64 for additional descriptions

Coming out of the production DAW, ADM parameters are just one part of the metadata within the BW64 container format. ADM itself is usually part of EBUCore, which is a much more extensive metadata set that has been designed “to describe audio, video and other resources for a wide range of broadcasting applications including archives, exchange and production in the context of a Service Oriented Architecture”[15].

Within ORPHEUS we have used the ADM audio parameters for the transformation process into the MPEG-H distribution codec[16]. EBUCore metadata, i.e. the technical attribute “ContentLevel”, was used for enabling the “Level of Importance” functionality in the variable-length feature, by extracting this information and conversion into JSON and passing it on to the MPEG-DASH stream.

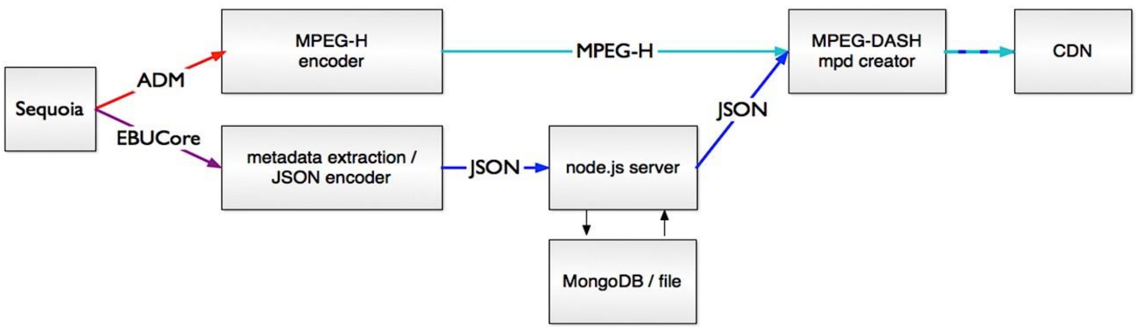

Figure 34: overview of audio and metadata flow for Pilot Phase 2

[15] *EBU Tech 3293 EBU Core Metadata Set* https://tech.ebu.ch/docs/tech/tech3293.pdf

[16] c.f. *ORPHEUS-D4.3 Final report on the work on representation, archiving and provision of Object-based Audio*, https://orpheus-audio.eu/wp-content/uploads/2018/02/orpheus-d4.3_final-report-on-representation-archiving-and-provision-of-oba.pdf

Regarding the extensive capabilities and functions of EBUCore, a lot of information might be regarded as valuable and appropriate in future use cases, e.g.

- identifying single audio objects embedded within an 'object-based audio composition'
- providing information on genealogy of the used material
- attaching links and references to additional material
- providing key-word for elements
- semantic web applications etc.

Presentation Design

To create appealing representations for object-based audio there's more to be delivered to the end user device than just audio for rendering. We have already emphasized the importance of a 'Presentation Design' by adding it as an integral part of our Reference Architecture described in D2.4[17]. Within the scope of ORPHEUS the 'Presentation Design' was used to enable accessibility and usability for the various new functionalities offered by object-based audio and additional metadata in an attractive and easy way. Some of the functionalities were unprecedented within an audio-only context, e.g. offering transcripts for radio, but also we didn't want to restrict ourselves from the start to given sets and legacy standards, so we opted for a "design-oriented approach".

Some of the features for 'Presentation Design' may be assisted by existing technologies and standards:

- Program Schedule -> EPG
- Item Playing -> DL+
- Item Visuals-> RadioDNS

The possibilities to incorporate or to interact with these standards are to be examined in future tests and development projects.

Some other functions we exemplarily introduced to audio streaming need fundamental development:

- Transcript Generation

  Ideally a transcript should be generated during the capturing or recording. It should always stay attached to the audio throughout the production process, regardless of any editing made on either the audio or the transcript level. If different persons are not identified as different audio objects at the time, automated speaker detection should be possible and implied within the transcript.
  For an appealing representation, names and icons of the speaker are used in the GUI.

  Solution used in ORPHEUS:
  Transcripts were partly generated by using TRINT[18], that offers parts of similar functionalities for generating transcripts as web services to create subtitles for video (".srt" format). These were transferred into the JSON files that contained this and other additional metadata for the demonstrators.

  Future approaches:
  Functionalities of transcript generation are to be considered for implementation into pre-production DAWs.

[17] c.f. *ORPHEUS-D2.4: Final Reference Architecture Specification and Integration Report* https://orpheus-audio.eu/wp-content/uploads/2018/03/orpheus-d2.4_final-reference-architecture-specification-and-integration-report_v1.0.pdf

[18] https://trint.com/

EBUCore[19] has defined an element <TextLine> that allows linking text to "parts/segments/scenes or a timestamp".
For distribution purposes, more investigation in respect of the EBU-TT metadata[20] elements seems appropriate.

- Chapterisation, Points of Interest

  In a pre-production, the Point of Interest or Chapters are either already given, or can be defined accurately by the producer or editor. Ideally these segments are enriched with additional short description and pictures.

  Solution used in ORPHEUS:
  Chapters and Points of Interest were manually identified and their related time stamps inserted in the JSON file, along with chapter title, descriptions and links to pictures and icons.

  Future approaches:
  The timestamps of chapters can be identified with markers within the DAW and exported within EBUCore in the BW64 file – however, any additional description of pictures could only be added if a DAW would offer this feature.

As an alternative or interim solution until these functionalities are available in DAWs and broadcast production systems, a web-based solution could serve. A mock-up is depicted here.

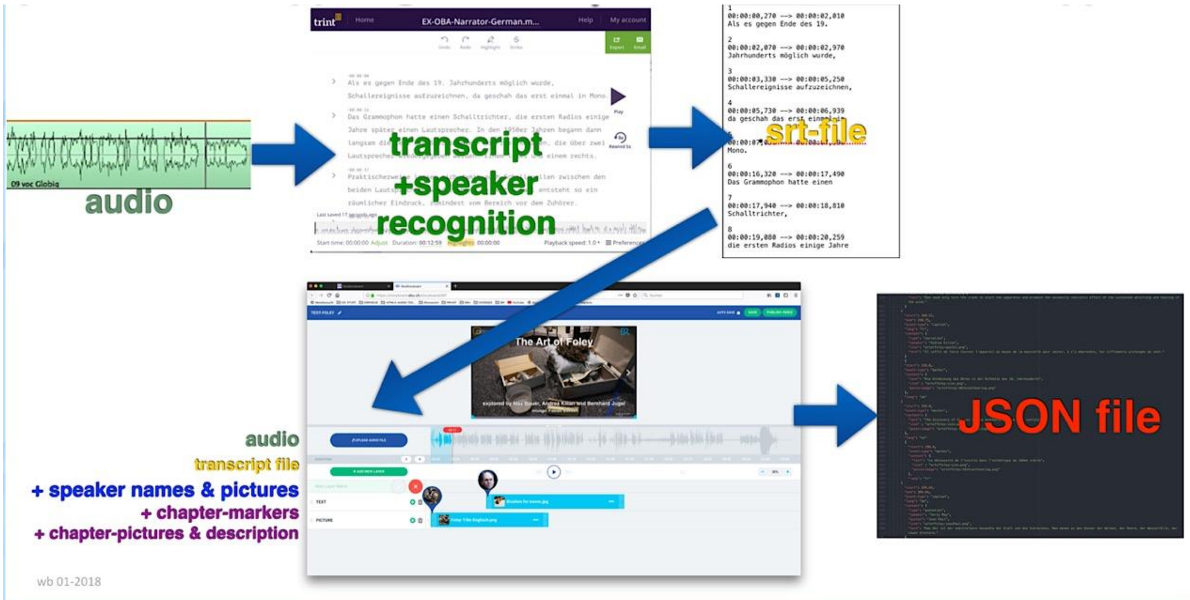

Figure 35: mock-up of a web-based workflow to generate additional JSON metadata

- An (object-based) audio file is loaded into a web service
- first a transcript including speaker detection is generated giving a basic timeline structure
- additional visual elements and pictures may be added in layers as instantaneous or continuing elements along the timeline
- a JSON file plus a container of attached assets is generated, ready for the provisioning.

[19] *EBU Tech 3293 EBU Core Metadata Set* https://tech.ebu.ch/docs/tech/tech3293.pdf

[20] *EBU Tech 3390 EBU-TT part M, Metadata Definitions* https://tech.ebu.ch/files/live/sites/tech/files/shared/tech/tech3390.pdf

Variable Length

The most challenging new feature was creating a variable-length functionality based upon editorial and producer-decided ‘levels of importance’. This leads to manifold considerations of how to implement this attractively and easy to handle into a production workflow. Although the results we present are convincing, other, possibly more complex material, may be more demanding.

- **UI in the DAW to create markers and identify segments**
  We have taken the simplest approach possible, using just marker named “L1, L2….L5”. Additional enhancement would include referencing colours for single segments.
- A “**flexible timeline**” (horizontal group function) is conceivable
- **Instant Pre-listening**
  The various edits created by such a segmentation process must be instantly available for pre-listening
- **Crossfading background sound**
  In case there are several audio objects audible at the time, e.g. as background sounds or music, and a move from point A to B would result in an unpleasant audible jump or cut generated by these elements, a crossfade function is desirable.
- **Alternative transition**
  Related to the crossfade mentioned above, a move from A to B could also require for an ‘alternative transition’, e.g. an abridged version or short insert in order to maintain the semantic integrity and comprehensibility of a message given.

7 Conclusions

The medium commonly called 'radio' offers a plethora of different formats and genres: music, news journalistic contributions, spoken word, documentaries, radio drama. And all of these can be created live as well as pre-recorded.

For almost a century, since it was introduced, radio implied solely being 'on air', transmitted on radio waves, a linear broadcast, tied to the passing of time. Even more: there was just one 'mix', a pre-formatted relationship of items of sound available for every reception device, any listening condition and audience.

Today, with Internet technology expanding more and more and replacing legacy broadcast distribution, we find that audience habits and requirements dramatically change. Media need to be more individual and targeted in order to be accepted and appreciated. The development of object-based media reflects these needs and offers unprecedented possibilities of accessibility, adaptability, personalisation and immersion.

It was our prime intention in the ORPHEUS project to demonstrate in the pilot phases the new possibilities that object-based audio can offer to established content formats. Therefore we attempted to apply the new features and functionalities to a full range of typical audio broadcasting genres and their related workflows.

We are confident that the examples and results presented here fulfil this convincingly. We hope, too, that these prototypes may serve not just as best-practice but also as inspiration for others to explore and create new content in the emerging object-based audio domain, to bring exciting new audio experiences to audiences.

Appendix A Production Credits for ORPHEUS’ Pilot Pieces

Heute im Stadion / Football Match (BR)

Footballmatch Bundesliga SC Freiburg - FC Bayern München,
January 20, 2017
Reporter: Hans-Peter Pull (BR) (re-enactment)
Producer: Werner Bleisteiner (BR)
Sounddesign & Mix: Christian Schimmöller, Helge Schwarz (BR)

Die Kunst des Geräuschemachens / The Art of Foley / L’Art de Bruitage (BR)

Written by: Max Bauer, Andrea Kilian, Bernhard Jugel
Translations: Eric Rosencrantz (EN), Étienne Gilllig (FR)
Produced by: Max Bauer, Andrea Kilian, Bernhard Jugel
Editor: Katja Huber (BR)
Voices: Narrator Susanne Schröder (DE), Anne Alexander Sieder (EN)
Katja Schild (FR)
Quotations: Andrea Kilian, Petra Mörk, Jerzy May, Heinz Peter (DE)
Tania Higgins, David Creedon (EN)
Andreas Kilian, Étienne Gillig (FR)
Recording, Sounddesign & Mix: Gerhard Wicho, Fabian Zweck

Mozart Gigue KV 574 in 360º by Passo Avanti

Arrangement: Alexander von Hagke
Musicians: Alexander von Hagke (Clarinet), Eugen Bazijan (Cello),
Julia vom Deich (Violine), Vladimir Grizelj (Guitar)
Recording & Mix: Christian Schimmöller (BR), Yannik Grewe (FHG IIS)

The Mermaid’s Tears (BBC)

Written by: Melissa Murray
Producer/Director: Jessica Mitic
Sound Designer: Caleb Knightley
Cast: Sarah Ridgeway (Lesley), Chetna Pandya (Dee), Simon Ludders (Bill),
Lara Laight (Sally)
Engineering: Chris Baume, Matthew Firth, Matthew Paradis, Andrew Mason,
Richard Taylor, Andrew Nicolaou

Erlebe objekt-basiertes Audio / Experience Object-based Audio (BR)

Author: David Globig (BR) - Translation: Eric Rosencrantz (EN)
Speaker: David Globig (DE), David Creedon (EN)
Sounddesign & Mix: Christian Schimmöller

Herbst / Autumn (FHG IIS)

Author: Axel Horndasch
Sound Design and Mix: Julian Klapp
Footage: Valentin Schilling

The Turing Forest (BBC R&D, S3A)

Written by: Shelley Silas - Translation: Werner Bleisteiner (DE - BR)
Voices: Graeme Hawley (EN), Carsten Fabian (DE - BR)
Sound Director: Chris Pike (BBC)
By kind permission of BBC R&D and the S3A project
[end of document]