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

ORPHEUS
Object-Based Audio Experience

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

D2.3: Interim Pilot Progress Report

Version: v1.0

Deliverable type	R (Document, report)
Dissemination level	PU (Public)
Due date	30/04/2017
Submission date	02/06/2017
Lead editor	Chris Baume (BBC)
Authors	Werner Bleisteiner (BR), Chris Baume (BBC), Niels Bogaards (ECANDY), Benjamin Duval (TRINNOV)
Reviewers	Nicolas Epain (BCOM), Andreas Silzle (FHG)
Work package, Task	WP2, T2.4
Keywords	

Abstract

This document describes the current status of Orpheus pilot phase 1, which has been divided into three stages. The first stage is a live production of an interactive object-based radio drama that can be experienced using a web browser. The second is a selection of material encoded using MPEG-H and made available through an iPhone and AV receiver. The third is an 'as-live' broadcast, live encoded using MPEG-H and made available over the Internet.

[End of abstract]

Document revision history

Version	Date	Description of change	List of contributor(s)
v0.1	10/04/2017	Added sections	Chris Baume (BBC)
v0.2	11/04/2017	sections BR added	Werner Bleisteiner (BR)
v0.3	3/5/17	Added live programme section	Chris Baume (BBC)
v0.4	5/5/17	Added stages B and C	Niels Bogaards (ECANDY), Benjamin Duval (TRINNOV)
v0.5	12/5/17	Added detail on stage B features. Moved background. Updated abstract/conclusion.	Nikolaus Färber (FHG), Chris Baume (BBC)
v1.0	2/6/17	Added suggested changes.	Uwe Herzog (EURESCOM), Werner Bleisteiner (BR), Andreas Silzle (FHG), Nicolas Epain (BCOM)

Disclaimer

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

Executive Summary

This document describes the current status of Orpheus pilot phase 1, which has been divided into three stages. The first stage is a live production of an interactive object-based radio drama that can be experienced using a web browser. The second is a selection of material encoded using MPEG-H and made available through an iPhone and AV receiver. The third is an ‘as-live’ broadcast, live encoded using MPEG-H and made available over the Internet.

Table of Contents

List of Figures

Abbreviations

AAC	Advanced Audio Coding
ADM	Audio Definition Model
BW64	Broadcast Wave format (64-bit)
MPEG-DASH	Dynamic Adaptive Streaming over HTTP

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.

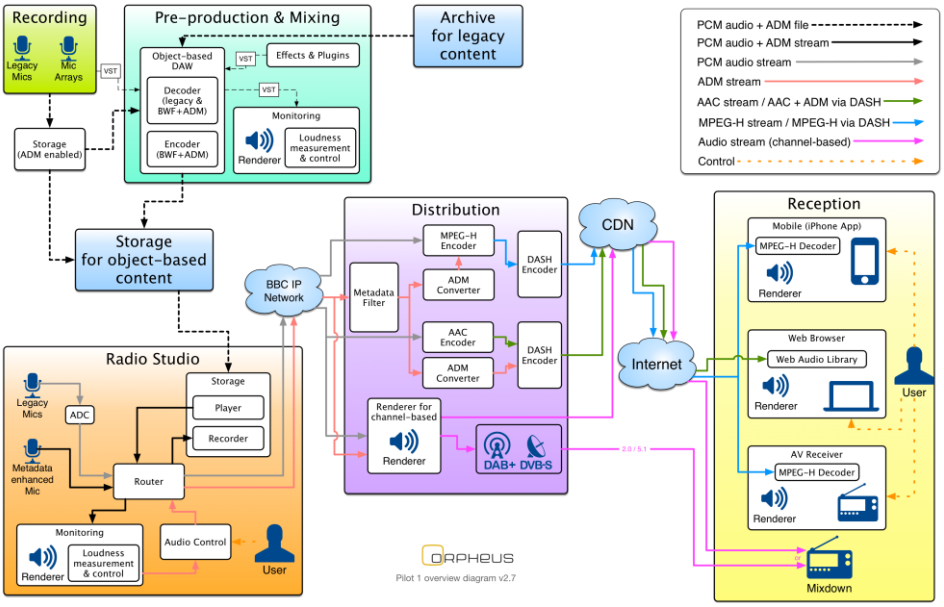

Figure 1: Pilot Phase 1 Infrastructure

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 has 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 is the IP-Studio[2] at the BBC (Radio Studio) - 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.

1.1 Pilot stages in 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.1.1.1 A – Live AAC+ADM

In this first phase, the MPEG-H encoder for delivery to the end-consumer is not yet fully integrated into the complete IP Studio and distribution infrastructure. Therefore, as operational workaround, the 'legacy' AAC encoding solution, as a proven reliable real-time possibility, alongside with the synchronous delivery of the object-based ADM metadata is chosen, in order to provide a first integration step. This distribution method also has the advantage that a broad audience can be addressed because the reception is based on commonly available HTML5 browsers.

A.1.1.2 B – As-live MPEG-H

Although MPEG-H live encoding facilities, along with MPEG-DASH streaming and MPEG-H reception and decoding in both app and AV receiver, are only likely to have been completed by the time of stage C, it has been decided to insert an interim "pseudo-live" Stage B, where the material used in stage A is offline encoded into a MPEG-H stream in advance and then broadcast "as live" immediately after completion. This will also allow to explore, test and evaluate the process of MPEG-H streaming distribution over MPEG-DASH extensively.

A.1.1.3 C – Live-encoded MPEG-H

In this final stage, the MPEG-H encoding capability is fully integrated and tested within each of the production, play-out, distribution and reception entities (IP Studio, Distribution, Reception macro blocks).

It is 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 should also be available to the public on the AppStore.

If feasible, an adapted and enhanced run-down scheme with additional live elements may be added in the IP studio at the BBC.

1.2 Selection of radio drama

Since its early days, from the 1920s onwards, radio in general has been a 'live' medium: announcers, musicians and actors all had to speak, play and act live in the radio studio and the output went

straight 'on air'. This was mainly due to the technical constraints in the days before high quality and durable recording facilities were available3.

Accordingly, the first full-length feature content for radio was transmitted live, then gradually adapted from existing performance arts genres, such as concerts, opera, dialogues and drama. Early broadcasters took up the challenge of trying to transmit these forms of content from renowned venues like concert halls, opera houses and theatres, but were hampered by the technical limitations of then available microphones, mixing and distribution technology.

As synthesis of both above mentioned pre-conditions, artists started to develop ideas of creating 'radio typical' art forms, to be realized exclusively in the radio studio, taking advantage of the 'acoustically controlled environment' and emerging possibilities to alter, manipulate and enhance words, sounds and music to create radio specific expressions and forms. The most common and advanced art-form of this kind has become the radio drama.

One of the most prestigious prodigies of this art form is Orson Welles's realization of the HG Wells novel "War of the World" staged and aired throughout live by Columbia Broadcasting System, New York on October 30, 1938 (luckily enough, by this time technology as well as a sense for radio-art being worthwhile to be preserved for posterity had been developed and it is still available for listening today). When aired, this radio play caused panic amongst listeners, who took the 'reports' on the attack for real. As Orson Welles embedded the plot of the novel, a fictional attack of the Martians, into the illusion of a regular radio programme going on, he played with – and uncovered - the suggestive power of radio itself. Ever since, radio drama has been the most advanced, challenging and yet still most 'open' art form in broadcasting.

[3] Additionally, legal restriction in many countries were imposed by the recording industry, fighting against recording facilities in broadcasting houses, fearing competition from commercial releases of broadcasting productions.

2 Pilot Stage A: live

2.1 Concept

The first Orpheus pilot programme will be a live object-based radio drama, produced and transmitted using a fully IP-based workflow. It has been commissioned especially for the project, and designed to show off some of the functionality that the object-based approach can offer.

The drama is being produced in partnership with BBC Radio Drama, who are running the editorial side of the production. The Orpheus partners are developing the technology for production, distribution and reception. We are also working with several freelancers such as a writer, graphic designer and web developer.

The working title of the drama is "**The Mermaid's Tears**". There are three main characters – two police officers (Dee and Bill) and a mother (Lesley) – and it is set in the mother's apartment. The police officers are investigating a poisoning, and interviewing the mother about what happened.

2.1.1 Features/functionality

We chose the following features that demonstrate the capabilities of the object-based approach, whilst integrating neatly with the narrative of the drama.

Multiple streams

The three characters in the story will move around the apartment into various rooms at different times. This introduces multiple parallel audio streams where action is happening simultaneously (see Figure 3). The users will have the option to choose which character they follow. Their experience of the story will therefore vary depending on who they follow, which gives a different perspective on what really happened.

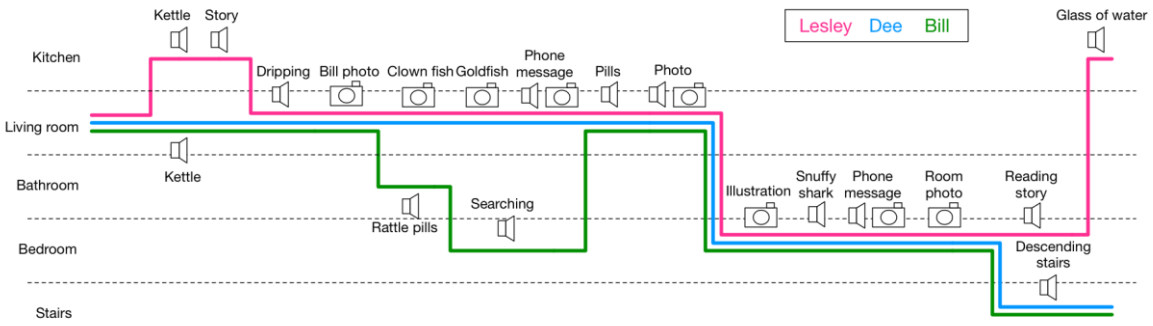

Figure 3: Draft diagram of character movements in the drama. Sound effects are marked with a speaker symbol, and images with a camera symbol.

Supplementary photos

Photos and images will appear at different points in the drama. These will be used to enhance the story by giving the audience a glimpse into the fictional world of the story, and to supplement the audio (e.g. displaying a text message). The images will be prepared in advance and triggered in the studio at the right moments.

2.2 Process and practicalities

This section gives an overview of the roles, tools, system and workflow that will be used to deliver the pilot.

2.2.1 Roles

The following roles are involved in producing the programme:

- **Producer**
  Leads all editorial aspects of the programme, including recruitment and leadership of the studio manager, writer, actors and graphic designer. Also performs essential administrative work such as booking spaces and paying contractors.
- **Studio manager**
  Also known as a sound engineer. Supervises all aspects of the sound, including sound design, equipment set-up and operation.
- **Executive producer**
  Oversees the work of the producer. Also coordinates the legal, financial and administrative to ensure the production can operate smoothly.
- **Writer**
  Works with the producer to develop the programme concept and theme. Writes the story and radio play script.
- **Actors**
  Perform the play in rehearsals and live broadcast, as directed by the producer.
- **Graphic designer**
  Creates graphical assets to give the programme a visual identity. Also responsible for creating the images that will be triggered during this object-based drama.

2.2.2 Tools

The tools below are being used to enable us to deliver this object-based production:

- **Pre-production interface**
  This will capture and store relevant metadata ahead of the programme's broadcast. Such information will include:
  - Title/descriptions for the programme
  - Labels/identities for the audio sources
  - Images to be triggered during broadcast
  - Pre-prepared audio clips to be played during broadcast
- **Pipeline configurator**
  This helps to set up routing for each IP Studio machine to direct the audio and metadata streams. For example, this allows us to arrange the inputs into a set of objects, and send those to a renderer, meter, recorder and the distribution chain.
- **Live production interface**
  This will control the metadata stream during the programme's broadcast, as controlled by the studio manager. The functionality will include:
  - Controlling the gain and position of each audio source
  - Assigning labels/identities to audio sources
  - Configuring the interactivity possibilities of the audio streams
  - Triggering playback of pre-prepared audio clips
  - Triggering images to be displayed
- **Distribution control interface**

This will set-up and control the distribution of the audio and metadata sources from the studio to the content distribution network.

2.2.3 System configuration

The draft system configuration for the programme is shown in Figure 4 (for the audio) and Figure 5 (for the metadata).

Audio

There are two rooms involved in the production – the recording studio (30C) where the actors perform, and the cubicle (30D) where the engineer monitors the mix. Three IP Studio machines will be used for the following operations:

- Studio
  - Routing
  - Rendering
  - Level metering
  - Object recording
- Audio playback
  - Audio clip storage
  - Audio clip playback
- Distribution
  - Audio encoding
  - DASH streaming

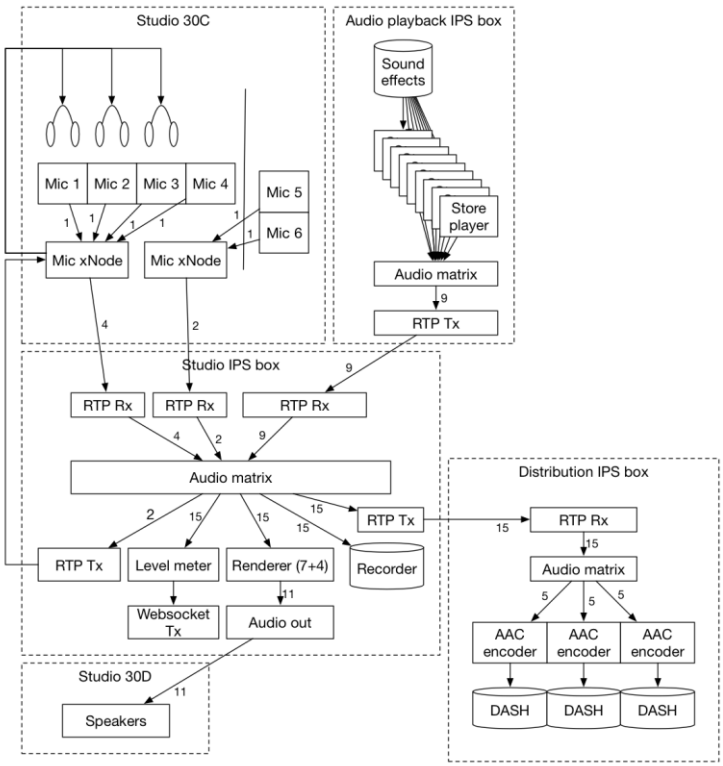

Figure 4: Draft system design for producing the audio

Metadata

Two systems are involved in generating and controlling metadata – the pre-production system and the live production system. Both have low overheads so will be hosted on virtual machines. Each system presents an interface which the production team can use to enter metadata before broadcast, then set up and control audio objects, and trigger audio playback and images at the correct times.

The output of these systems drives the audio renderer, which the studio manager uses to monitor the spatial audio output, and the distribution system which sends the metadata to the audience.

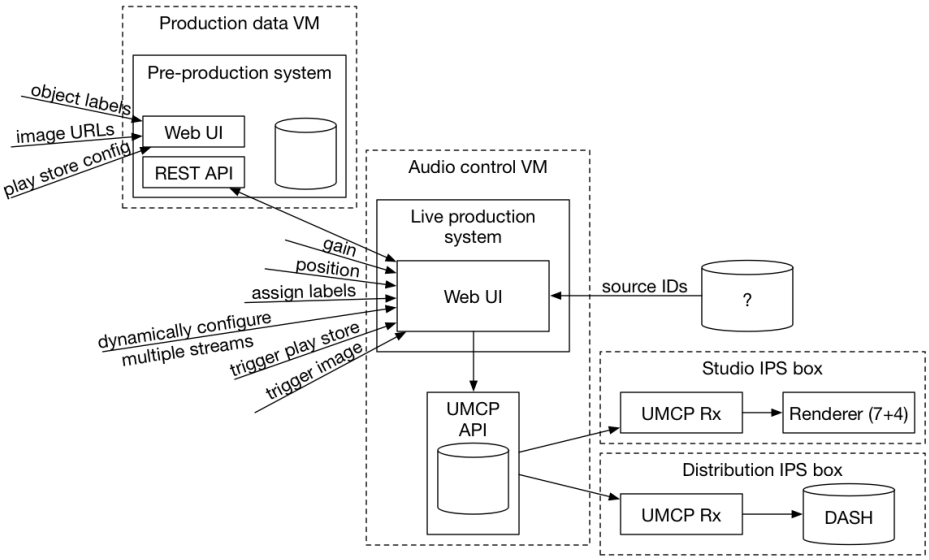

Figure 5: Draft system design for producing the metadata

2.2.4 Pilot production workflow

1. **Audio preparation**
   The programme contains many sound effects which need to be prepared in advance and triggered at the right time during broadcast. To achieve this, we write a list of sound effects that we need, and the sound engineer either records these or sources them from a sound effects library.
2. **Metadata preparation**
   The metadata which needs to be triggered during broadcast is prepared in advance. The images are created by the graphic designer and loaded into the pre-production interface by the producer. The title, description and labels/identities are also entered by the producer. The studio manager also adds the location of the audio clips that they prepared, so that they can later be triggered.
3. **Technical rehearsal**
   The system will be tested before the editorial team start work, to ensure that the tools and interfaces are working as expected. This will allow us to find problems and have the time to fix them without impacting on the production of the programme itself.
4. **Rehearsal**
   As with any live production, the actors, studio manager and producer will rehearse ahead of the live transmission to practise the performances and operation of the equipment. This will also act as a dry-run for the distribution system.

5. **Live broadcast**
   The drama will be performed live and transmitted over the web to a set of test listeners. The output will be recorded for later re-use.

2.3 End-user experience

2.3.1 Availability

Our live object-based radio drama in Stage A will only be available to a limited set of test users. After the live transmission, we will then make the programme available on BBC Taster for the wider public to experience the drama. BBC Taster is a platform for testing experimental formats and technologies, and getting feedback from the audience.

We will use Javascript and the Web Audio API to process the audio and metadata directly in the browser, so that users can experience the programme without having to install any special software.

2.3.2 Functionality

The following features will be made available to the audience:

- **Select output format**
  The audio for the programme is being rendered using their browser, which offers the possibility to provide output for various loudspeaker layouts and headphones. Our interface will give listeners the option of speaker output (based on their current configuration – e.g. stereo/5.1) or headphone output. We will also investigate including a head tracking function using data from available sensors. However, only a small fraction of listeners will have this capability, so this function will likely be added later.
- **Choose/switch character**
  As the programme can be experienced from multiple characters' viewpoints, users will be able to select which character they want to follow. They will also be able to swap character mid-way through the programme.
- **Image display**
  At points throughout the programme, images will be triggered and displayed to the audience.

3 Pilot Stage B: as-live

The aim of stage B is to evaluate the MPEG-H distribution on several platforms. This evaluation will be performed in a “pseudo-live” manner, with an offline encoding but an “as live” broadcasting. This will allow testing all of the stages of distribution of a MPEG-H streaming over MPEG-DASH. In addition, the end-user experience is exactly the same as for the live broadcast, such that user studies can be performed.

3.1 MPEG-H encoding

To simulate a live broadcast using MPEG-H as closely as possible before the entire processing chain is in place, the captured output from the Stage A live broadcast will be encoded offline by FHG shortly after it is broadcasted. This MPEG-H version can then be streamed to the MPEG-H clients; the iOS app by ECandy and the Trinnov AV-Receiver.

3.2 Additional material

To be able to test and present more aspects of the object-based delivery chain, Stage B of Pilot 1 will feature some additional content. Features that will likely be produced for this purpose include

- **Classical Music Orchestra**
  This item is a Higher Order Ambisonic (HOA) Sound Scene and is a recording of a classic music concert, such that the listener feels as being the conductor of the orchestra. Another example of an HOA Sound Scene could be a nature ambiance recoding from a rain forest with animals, wind and rain drops. The focus of this item is on an immersive 3D sound experience.
- **Soccer 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 his personal preference.
- **The Turning Forest**
  This is an existing radio drama available as ADM/BW64. 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 ADM/BW64 to MPEG-H.

3.3 End-user experience

3.3.1 Availability

Pilot 1 Stage B is not available to the public at large yet, but only to a selected focus group. The goal is to evaluate the technical completeness and correctness of the system, as well as the comprehensibility and usability of the iPhone app. The participants will have the opportunity to use the app with several reproduction systems, from headphones and conventional stereo loudspeaker systems to high-end multichannel loudspeaker systems.

3.3.2 Functionality

As Stage B includes a recording of the live broadcast from Stage A, we attempt to make all features listed in Section 2.1.1 also available in Stage B (Select Output Format, Choose/Switch Character, Image Display). Note, however, that this requires a conversion from ADM/BW64 into MPEG-H, which is not a trivial process. If necessary, the audio format and features have to be adapted to match the constraints given by the MPEG-H 3D Audio Low Complexity Profile at Level 3. In addition, the reproduction devices are not browser-based (as in Stage A) and much of the functionality has to be re-implemented on a different platform. Though this exercise is not without effort, it verifies the feasibility of converting the production format (ADM/BW64) into a modern distribution format (MPEG-H) and demonstrates 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 will be available on the iPhone app. The focus is on audio-related functionality that has a specific mapping to MPEG-H and object-based audio:

- **Browsing and Tune-In**
  While Stage A provides a single stream which lasts for a significant amount of time, Stage B will offer several short (<3 min) items as content. The user will be able to browse those and select one based on some summary information (picture, text). As we simulate a live broadcast, the item may not start from the beginning but start in the middle as typical for tuning into 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 he has the feeling of having missed something important. He can also skip forward again and catch up to the “live edge” of the stream. As we simulate a live broadcast it is not possible to skip forward beyond the current time.
- **Language Selection**
  Some items will have multiple languages which the user can select from. In particular “The Turning Forest” will have an English and German narrator. The user can also set his preferred language in which case the correct language is selected automatically.
- **Foreground/Background Volume Control**
  Some items will have the possibility to control the audio volume of foreground and background objects. In particular the “Soccer Match” will have the option to amplify the commentary while reducing the volume of the stadium atmosphere.
- **Persistency of User Settings**
  When “zapping” from item to item, it is often annoying to have to set the personal preferences again. For example, after setting the stadium/commentary to the desired prominence value and switching to another item/channel, it is desirable to re-apply the prominence value if the user switches back to the “Soccer Match”.
- **Loudness and Dynamic Range Control**
  Depending on the playback environment, it is 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).
- **Flexible Reproduction**
  Depending on the playback environment, different reproduction systems can be selected by the user. In particular, headphone/stereo, headphone/binaural, speakers/stereo, and speakers/surround. This will allow the user to experience different levels of immersion and requires the audio codec to provide different rendering and downmix formats.

3.4 Platforms/devices

Two devices will be used: an iPhone and an advanced A/V receiver.

On the iPhone, a privately distributed iOS app will be used to receive the MPEG-DASH / MPEG-H stream. The app is planned to support rendering to headphones (in stereo and binaural) and multichannel surround (5.1, 7.1) using AirPlay or a wired connection.

On the A/V receiver, a specific compilation of the GStreamer open source library will receive the MPEG-DASH stream and give the MPEG-H to the main signal processing operator of the unit. Then, a signal processing operator will use the FhG decoder to render the MPEG-H to a 7.1.4 loudspeaker layout.

3.5 Audience

As Stage B is a closed group stage of the Pilot, the audience will be selected by the various partners. The goal is to have a diverse group of listeners, to get feedback on as many aspects as possible; subjective sound quality evaluation, user interface analytics, clarity of the conveyed concepts, etc.

Specifically for the A/V Receiver, since the required setup is cumbersome and not easy to relocate, the tests will be performing at Trinnov's premises.

4 Pilot Stage C: live-encoded

This final stage is aimed at integrating and evaluating the complete pilot 1 broadcast chain. By this point, the MPEG-H encoding and decoding capability will be running in real-time, which will allow us to produce and distribute content to the iPhone and AV-receiver live.

In order to achieve this, the MPEG-H production library and MPEG-H encoder must be integrated with the BBC's IP Studio platform, and the metadata must be translated from ADM to the MPEG-H data structure. This provides a number of significant technical hurdles, causing us to separate this task out to the third stage of integration.

4.1 Additional material

Some additional material will likely become available in the period between Stage B and C, and this will be integrated in the Stage C public demo. Experiences from Stages A and B may also drive the production of additional content. It is unclear at the moment what this additional content may include, but are considering giving the opportunity for third-party broadcasters to try out the tools we have developed to produce and contribute some additional content.

4.2 End-user experience

Stage C's user experience can been seen as 'Stage A and B combined, but for a large audience'. As part of the audience is likely to be much less experienced with advanced audio technology, it will be important to adapt user interfaces and communication materials to be as clear as possible.

The program will have a more user friendly setup, compared to the more technical Stage B presentation. Stage C will use a second iteration of the iPhone and receiver interfaces, building on the lessons learned from the feedback received in Stage B.

The audience of this stage C will be a larger public for the iOS app: everyone who can download the iOS app to his personal devices will be able to experience the reception.

The A/V Receiver will also be presented to a larger audience, on a public event.

4.2.1 Availability

The same platforms and devices as in Stages A and B will be supported: an iOS app and an A/V Receiver.

The iOS app for Pilot 1 Stage C will be publicly available, probably on Apple's App Store. This cannot be guaranteed, as it will be subject to Apple's app approval process, but we can expect it to be publicly available. The content will be available for a limited amount of time, not longer than the end of the project.

The A/V Receiver embedding this functionality will not be available for sale yet, since the production of a hardware unit requires a large amount of testing. However, we aim to demonstrate the receiver at the International Broadcasting Conference (IBC) in September 2017.

4.2.2 Functionality

The functionality of Stage C will be identical to Stage B, however it will be encoded 'as-live'. In order to create a professional experience, we will join the material in Stage C together by recording 'links' from a radio presenter. This will turn the material into a object-based audio showreel, and we can use the link to explain a bit about the project and the technology.

5 Conclusions

The three stages of Orpheus pilot phase 1 will allow us to test our pilot architecture, and the tools we have developed to implement it. This includes live production and broadcast using ADM, translation of ADM to MPEG-H, and live encoding of MPEG-H. By the end of pilot phase 1, we should have a fuller understanding of which approaches work or not, and where there are weaknesses in the system where further research or development is needed.

Pilot phase 1 focusses on live broadcast, and it will be followed by a phase 2, which will investigate the possibilities made available when pre-producing audio experiences. We are currently at the stage of defining the scope of phase 2, and developing ideas around what the user experience could be.

[end of document]