Object-based Broadcasting – for Next Generation Audio Experiences

# THE EU PROJECT ORPHEUS

**Text** Andreas Silzle


Foto: BBC R&D

## INTRODUCTION

The media landscape has been subject to significant and on-going changes over the past years. The advent of mobile devices capable of playing media files with increasing sound quality, broadband mobile internet access and on-demand services has led to new and various media consumption patterns and consumer expectations. On the other hand, media production in general has not evolved to the same extent, which leads to an increasingly high effort to support an ever-growing number of target platforms and formats which cannot be fully leveraged today.

In particular, the way we are producing, transmitting and consuming audio has not significantly evolved over the past decades. Object-based audio is a promising concept that has the potential of transforming the whole way media is envisioned, created and consumed. Object-based audio delivers carefully curated individual elements of a programme all the way to the audience.

The EU research project ORPHEUS aims to create a complete object-based audio chain, implemented for a radio broadcasting scenario.

## 1. CONCEPT AND APPROACH

### 1.1. Project Background

Although digital and file-based workflows have recently found their way into media production, those workflows still try to some extent to map the old analogue and tape-based method of production and delivery to the digital world. For instance, in an audio production various sound sources are mixed in a digital audio workstation (DAW) to create a final channel-based mix for a specific target loudspeaker layout, see Fig. 1. Each audio channel in the final product is intended to be reproduced by a loudspeaker at a well-defined position. This fixed audio mix is transmitted to the end-user with basically no means to adapt it to a specific playback device or the user's personal preferences.

An object-based production approach, however, is able to overcome the above-mentioned obstacles. The term object-based media has become commonly used to describe the representation of media content by a set of individual audio assets, together with metadata describing their relationships and associations.

At the point of consumption these so-called objects are assembled to create an overall user experience. The precise

combination of objects is flexible and can be responsive to user, environmental and platform specific factors.
Essentially, the goal is to capture the creative intent of the producer and carry as much information as possible, required or desired, from the production side to the end-user, to ensure the best recreation possible on the consumer side (see Fig. 2). To achieve this, the final product of a production process will be an audio scene that is in turn composed of several objects. The metadata associated with each object includes, but is not limited to, the target position of the audio signal, its target loudness and a description of its actual content.
For playback, the object-based content needs to be ‚rendered' to the reproduction layout, such as a multichannel loudspeaker set-up. The term ‚rendering' describes the process of generating actual loudspeaker signals from the object-based audio scene. This processing takes into account the target positions of the audio objects, as well as the positions of the speakers in the reproduction room. It may further take into account user interaction such as a change of position or level.
An object-based approach, as mentioned above, can serve end-users more effectively by optimising the experience to best suit their access requirements, the characteristics of their playback platform and the playback environment or personal preferences of the listener. Moreover, it is highly beneficial for content producers, as workflows can be streamlined and only one single production needs to be created, archived and transmitted in order to support and serve a multitude of potential target devices and environments. This is enabled by the simple fact that the metadata of individual objects can be modified and adjusted, either by the end-user or along the production and transmission chain, without the need to change the audio material itself. This way, the key features of object-based media:

- Interactivity and personalisation,
- Accessibility,
- Immersive experiences and

can be achieved in a non-destructive, controlled and scalable way.

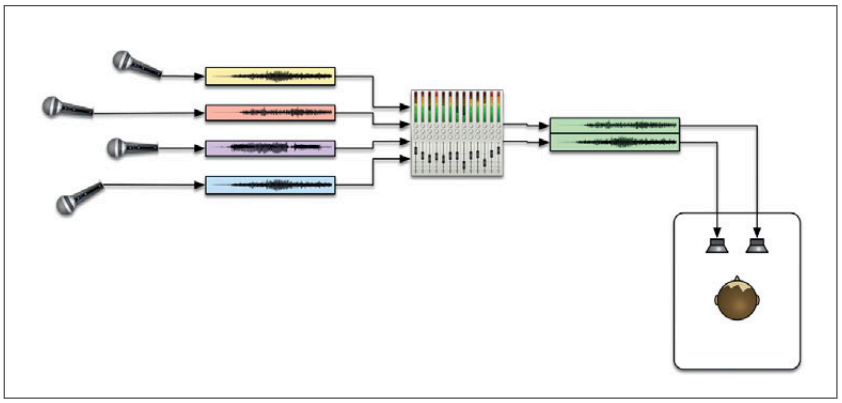
### 1.2. Setup-agnostic Content Representation

With object-based audio, sound is represented by a number of separate objects and associated side information as metadata, which defines, among other characteristics, the level, position or movements of the objects. As this metadata information is independent of the reproduction system, predefined locations or movements can be preserved in the best way possible, regardless of the loudspeaker reproduction layout.

### 1.3. Interactivity and Personalisation

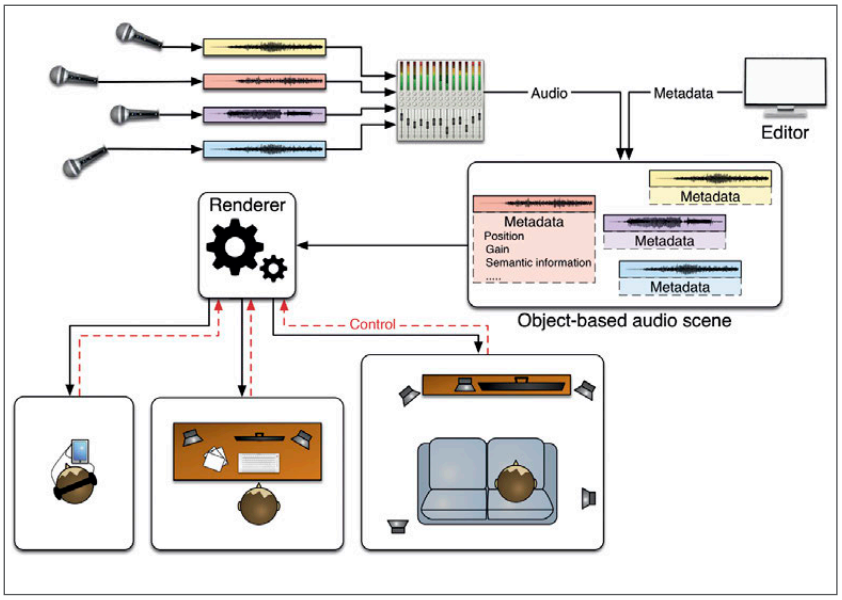
Because audio objects can be processed separately, object-based audio allows for interactivity and may be influenced by the user or application prior to the rendering process. With object-based audio, a listener can create their own mix and/or change their virtual listening position and adjust the spatial distribution or level balance of sound sources. Multiple language tracks can be defined as audio objects of which the user can select one for playback. Instead of transmitting a complete mix for each language version, original language and several dub versions, as well as the language-agnostic content, such as ambience, can be transmitted as separate audio objects.

**DR.-ING. ANDREAS SILZLE**
Senior Scientist, AudioLabs-IIS, Fraunhofer-Institut für Integrierte Schaltungen IIS, technical coordinator of ORPHEUS



Grafik: IRT

Figure 1: Conceptual overview of channel-based audio production and consumption

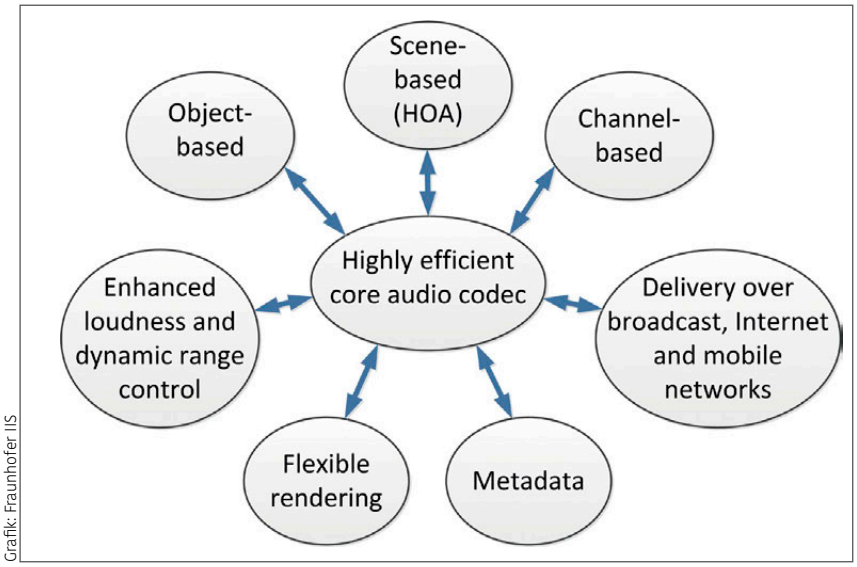



Grafik: IRT

Figure. 2: Conceptual overview of object-based audio production and consumption

Applied consistently, the object-based approach is far more than a unified format with additional metadata. Instead, it provides a description of the actual content of an object and its relation to other objects which enables a completely new and potentially non-linear media experience. In this context, the term ‚non-linear' refers to the possibility to adjust the narrative path of a story, thus enabling the option to select sections of special interest. A simple example might include news items, which can be reduced or expanded in length based on timing constrains or personal interests from short headlines to in-depth coverage, including all quotes and full length sound bites. This is not only beneficial for end-users from a personalisation point of view, but also enables producers to provide a single production that can automatically serve a large range of target platforms like mobile consumption, classical radio or social media that all have different granularity constraints.

Figure 3: Main properties of MPEG-H Audio




Grafik: Fraunhofer IIS

#### 1.4. Accessibility

Additional dialogue tracks such as spoken subtitles or audio description tracks can be provided with object-based audio. They can be selected in addition to or as a replacement of the main dialogue audio track to gain high accessibility. With object-based audio the level of single sounds and the balance between different sounds are adjustable in a convenient way. For example, in the presence of spoken subtitles, voice-over translation, audio description or simultaneous translation, the level of the main speech content/main audio track can be reduced automatically or by the user when an additional track is active (ducking, audio description receiver mix). For better speech intelligibility the balance between speech content and ambient sound may be changed depending on personal user-preference, the listening environment or the hearing abilities of the consumer. If the overall level is low, the low-level content, such as the dialogue, can be increased for easier comprehension, while at the same time the intensity of higher-level audio content is reduced, as seen in late night mode options.

## 2. USED FORMATS

In an IP-based broadcast production all the different specific video and audio connectors and cable types are replaced by Ethernet connectors and cables. Nevertheless, different formats operating on these unified connections are still necessary. The selected formats used in ORPHEUS are presented in the next sections. A requirement for the selection of these formats was that they are open formats and international standards.

#### 2.1. Audio Definition Model (ADM) and BW64

The most recently defined and most complete metadata model that supports the description of object-based audio (as well as channel-based and scene-based audio) is the Audio Definition Model (ADM). For more details see [1].

The BW64 is defined in Recommendation ITU R BS.2088. To describe the workflow with ADM metadata in BW64 files, an accompanying report ITU-R BS.2388 was published, describing typical use cases and recommended practices.

#### 2.2. MPEG-H: Delivery Format for Object-based Audio

MPEG-H 3D Audio is an audio coding standard developed by the ISO/IEC Moving Picture Experts Group (MPEG). The main features that make MPEG-H 3D Audio applicable for delivery of next generation audio ranging from highest-quality cable and satellite TV down to streaming to mobile devices are its flexibility with regard to input formats and its flexibility with regard to reproduction formats. An overview of MPEG-H audio metadata is provided in [2, 3]. The main properties of MPEG-H Audio are visualized in Fig. 3.

#### 2.3. MPEG-DASH

Because MPEG-H is primarily a compression format with the addition of 3D-rendering capability, it can be used like any other compression format together with codec-agnostic streaming formats such as DASH.

## 3. INITIAL REFERENCE ARCHITECTURE

One of the major objectives of ORPHEUS is the specification of reference architecture for an end-to-end object-based audio production workflow. The current version of the architecture contains five macroblocks, see Fig. 4.

#### 3.1. Recording

The purpose of the recording macroblock is to provide the tools and infrastructure required to conduct object-based recordings. Single objects as well as entire audio scenes will be captured, using both legacy microphones and novel microphone arrays which will be developed within the scope of ORPHEUS.

### 3.2. Pre-production and Mixing

The purpose of the pre-production and mixing block is to deliver tools for editing existing object-based content or creating such content from legacy audio material or other sources. The core of this block is the object-based DAW, which is extended by several tools and workflows to import, edit, monitor and export object-based content. For previewing the object-based content and simulating the user experience it is important to allow monitoring with different speaker setups (including binaural monitoring).

### 3.3. Radio Studio

The studio is the block, in which all of the different source material comes together and is combined into a 'radio programme'. The technical requirements in the control room of a radio studio include: capturing of audio signals, routing of external audio sources, playback of pre-recorded material, monitoring of audio quality and adding 'content metadata'. Traditionally, the 'mixer' is where audio channels are mixed together to form a single audio stream of the programme. In an object-based production, this concept is radically different in that elements are not usually mixed, but kept separate. The 'mixer' in this context performs the same overall function of a mixing desk, which is bringing together multiple sources into a single experience, but doing so through the manipulation of metadata. This handling is part of the IP Studio software of BBC R&D.

### 3.4. Distribution

This macroblock contains the modules and tools needed for distribution. Two distribution paths are considered:
a) Distribution of the object-based content via a content delivery network (CDN).
b) Distribution of a channel-based downmix of the object-based content, i.e. a pre-rendered version, delivered via DAB+ and/or DVB-S

### 3.5. Reception

This macroblock provides hardware and software solutions for the reproduction and reception of object-based audio content for the end-users. All solutions are comprised of two major components, a decoding/rendering module and a user interface. The decoding and rendering module automatically adapts the play-out of the object-based content according to the end-user environment's setup, i.e. from conventional loudspeaker setups to advanced multi-channel immersive audio systems or binaural rendering over headphones. The user interface allows for various personalised and interactive reproduction of audio content.
Several hardware and software solutions have to be considered in order to meet the requirements of different segments of the consumer electronics market, as well as different end-user listening habits and audio content consumption contexts.

The paper reflects only the authors' views. The Commission is not responsible for any use that may be made of the information it contains.

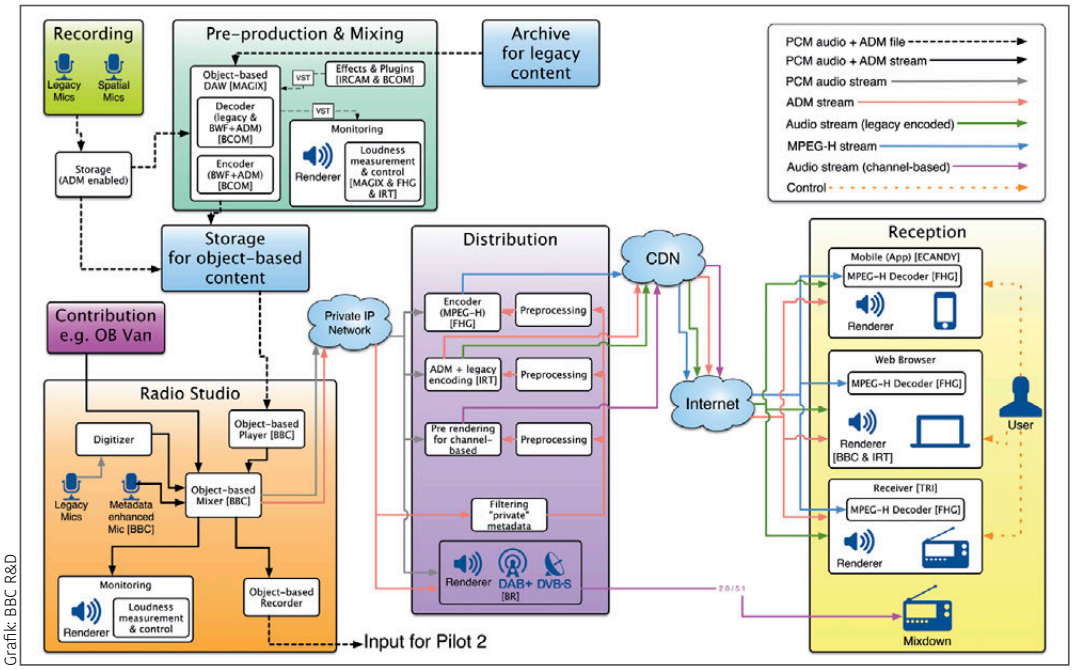



Figure 4: Pilot implementation architecture

## 4. SUMMARY

The goal of ORPHEUS is to bring the experience of object-based content to mass audiences at no additional cost. The project will demonstrate the new prodigious user experience through the realisation of close-to-market workflows and proof the economic viability of object-based audio as an emerging media and broadcast technology. ORPHEUS will publish reference architecture guidelines on how to implement object-based audio chains (http://orpheus-audio.eu/public-deliverables/). More details can be found in [4].

## 5. ACKNOWLEDGMENT

This work was funded in parts from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 687645 (ORPHEUS project).

References

[1] Füg, S., D. Marston, and S. Norcross. The Audio Definition Model – A Flexible Standardised Representation for Next Generation Audio Content in Broadcasting and Beyond. 141st AES Convention. 2016. Los Angeles, USA.

[2] Füg, S., et al., Design, Coding and Processing of Metadata for Object-Based Interactive Audio, in 137th AES Convention. 2014: Los Angeles, USA.

[3] Herre, J., et al., MPEG-H 3D Audio – The New Standard for Coding of Immersive Spatial Audio. IEEE Journal of Selected Topics in Signal Processing, 2015. 9(5).

[4] Silzle, A., et al. The EU Project ORPHEUS: Object-based Broadcasting – for Next Generation Audio Experiences. 29th Tonmeistertagung – VDT International Convention. 2016. Cologne, Germany.