Immersive audio properties for NCL media elements

Rafael Diniz Computer Science Department University of Brasília Brasilia DF Brazil rafael@riseup.net Marcelo F. Moreno Computer Science Department Federal University of Juiz de Fora Juiz de Fora MG Brazil moreno@ice.ufjf.br

ABSTRACT

With the introduction of immersive audio formats like MPEG-H and Dolby AC-4, TV terminals may allow end users to configure their experience by changing values of audio parameters directly or indirectly (audio profiles) via some kind of configuration user interface. In order to support more dynamic user interfaces for this kind of interactive audio setup, DTV apps should be able to manipulate such parameters as well. This contribution proposes the addition of immersive audio properties to media elements in NCL applications.

KEYWORDS

Immersive audio, interactivity, media properties, NCL

1 Introduction

ABNT NBR 15606-1 standard [1] specifies the data coding formats and protocols to be supported by middleware implementations for TV receivers under the Brazilian Terrestrial DTV system. Its most recent version (2018) includes new media formats, multimedia containers, streaming protocols and digital right management mechanisms, assembled into a new profile of TV receivers, called D-Profile TV receivers. Among these new media formats, immersive audio codecs MPEG-H and Dolby AC-4 are now included as optional support. Furthermore, these immersive audio codecs are also approved by the SBTVD Council to be supported as elementary streams in the broadcast, and new versions of the volumes of the ABNT NBR 15602 [2] containing support for these codecs should be published until the end of 2019.

In traditional channel-based audio coding, various sound sources are mixed to create a final channel-based mix for a specific target loudspeaker layout. Each audio channel in the final product has 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 their needs, which may be a specific playback device or their personal preferences [3].

Different from channel-based codecs, immersive audio codecs support object-based coding and, therefore, the possibility to

In: Future of Interactive Television Workshop (V WTVDI), Rio de Janeiro, Brasil. Anais Estendidos do Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia). Porto Alegre: Sociedade Brasileira de Computação, 2019. ©2019 SBC – Sociedade Brasileira de Computação. ISSN: 2596-1683 personalize the immersive experience by handling object's parameters.

The term 'object-based media' has become commonly used to describe the representation of media content by a set of individual assets, together with metadata describing their relationships and associations. At the point of consumption these objects can be assembled to create an overall user experience. The precise combination of objects can be flexible and 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. 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.[3]

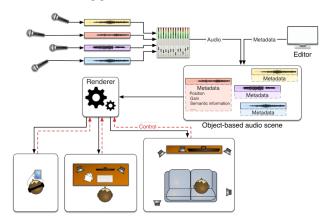


Figure 1. Conceptual overview of object-based audio production and consumption [2]

2 Proposal

In order to support immersive audio with interactivity support, DTV middleware specifications shall be modified in order to include an API that provides access to audio parameters. In the Brazilian Terrestrial DTV System and IPTV services compliant with ITU-T H.761, the Nested Context Language provides a harmonized API for handling media properties. The <media>

element allows for the reference to media objects of any type and, for some media types, specific properties may apply, via the property> element.

This contribution proposes the addition of immersive audio properties to media elements in NCL applications.

In immersive audio coding, each audio component (which can be a channel-based audio mix, an audio object or a HOA-based mix) has associated metadata properties like:

• the spatial change of its position, given in spherical coordinates (azimuth, elevation, radius)

 a linear gain that is to be applied by the object renderer with optional Loudness compensation

For instance, MPEG-H [4] defines the above metadata properties as well as interactivity control metadata, which specifies allowances and ranges for decoders to change some of those object metadata properties.

The following new properties should be added for immersive audio media objects:

| Property name | Description | Type | Example | Observation |
|---------------------------|---------------------------------------|-------------------|----------------------|--------------------------|
| (read-only) | | | | |
| soundPresetList | List of preset labels defined in | string | "3DStadium, | Labels are chosen by |
| | the audio coding, separated by | | Galvao, Adnet, | content producers and/or |
| | comma | | VisionImpairemen | broadcasters at encoding |
| | | | t," | time |
| soundObjectList | List of audio component labels | string | "actor1, actor2, | Labels are chosen by |
| | defined in the audio coding with | | carEngine, | content producers and/or |
| | at least one interactivity setting | | weather, bgMusic, | broadcasters at encoding |
| | allowed, separated by comma | | streetNoise, | time |
| | | | explosion" | |
| soundAzimuthOffsetRange | Allowed minimum and | negative integer, | -180, 180 | This property replaces |
| (label) | maximum offset for azimuth | positive integer | 0, 0 (no control | "balanceLevel" in |
| | control for <i>label</i> , in degrees | | allowed) | immersive audio media. |
| | (minAzOffset°, maxAzOffset°). | | | label must be an audio |
| | | | | component label |
| soundElevationOffsetRange | Allowed minimum and | negative integer, | -90, 90 | label must be an audio |
| (label) | maximum offset for elevation | positive integer | 0, 0 (no control | components label |
| | control for <i>label</i> , in degrees | | allowed) | |
| | (minElOffset°, maxElOffset°). | | | |
| soundDistFactorRange | Allowed minimum and | positive real, | 0.00025, 8.0 | label must be an audio |
| (label) | maximum distance change factor | positive real | 0.0, 0.0 (no control | component label |
| | for distance control of label | | allowed) | |
| | (minDistFactor, maxDistFactor) | | | |
| soundGainRange(label) | Allowed range for linear gain | negative integer, | -63, 31 | label must be an audio |
| | control for label, in decibels | positive integer | 0, 0 (no control | component label |
| | (minGain db, maxGain db) | | allowed) | |

Table 1. Read-only properties (producer- or broadcaster-defined components, ranges and allowances):

| Property name | Description | Type | Example | Observation |
|------------------------------|---------------------------------|---------|--------------------|-------------------------|
| (read-only) | | | | |
| soundAzimuthOffset(label) | Current azimuth offset for | integer | -45 | label must be an |
| | the object/channel | | 0 (may denote non- | audio component |
| | identified by <i>label</i> , in | | existent label) | label |
| | degrees | | | |
| soundElevationOffset (label) | Current elevation offset for | integer | 30 | <i>label</i> must be an |
| | the object/channel | | 0 (may denote non- | audio component |
| | identified by label, in | | existent label) | label |
| | degrees | | | |

| soundDistFactor(label) | Current distance factor for | positive real | 2.5 | label must be an |
|---------------------------|-------------------------------|---------------|--------------------|-------------------------|
| | the object/channel | | 0 (may denote non- | audio component |
| | identified by label | | existent label) | label |
| soundGain(label) | Current linear gain for the | integer | 15 | <i>label</i> must be an |
| | object/channel identified | | 0 (may denote non- | audio component |
| | by <i>label</i> , in decibels | | existent label) | label |
| soundLoudnessCompensation | Enable or disable loudness | boolean | true (enabled) | <i>label</i> must be an |
| (label) | compensation after a | | false (disabled) | audio component |
| | change in gain setting | | | label |
| soundPreset(label) | Current state of the preset | boolean | true (on) | <i>label</i> must be a |
| | identified by label. | | false (off or non- | preset label. |
| | | | existent label) | |

Table 2. Read-write properties (immersive audio interactivity):

2 Use case(s)

The support in NCL of immersive audio features will allow seamless integrated between hybrid DTV apps and the audio interactivity features allowed by immersive audio codecs.

REFERENCES

- [1] ABNT. ABNT NBR 15606-1: Data coding and transmission specification for digital broadcasting. Part 1: Data coding specification. 2018.
- [2] ABNT. ABNT NBR 15602 series Digital terrestrial television Video coding, audio coding and multiplexing.
- [3] WEITNAUER, Michael; MEIER, Michael. Object-based audio: The future of audio production, delivery and consumption. IRT website. 2018. Available at https://lab.irt.de/demos/object-based-audio/
- [4] ISO/IEC. ISO/IEC 23008-3:2019: Information technology High efficiency coding and media delivery in heterogeneous environments — Part 3: 3D audio. February 2019.
- [5] MALHAM, Dave. High order Ambisonic systems. Abstracted from "Space in Music – Music in Space" Mphil thesis. University of York. April 2003. Available at https://www.york.ac.uk/inst/mustech/3d audio/higher order ambisonics.pdf
- ABNT. ABNT NBR 15606-1: Data coding and transmission specification for digital broadcasting. Part 1: Data coding specification. 2018.