Immersive Audio Coding for Virtual Reality using a Metadata-Assisted Extension of the 3GPP EVS CODEC

D. McGrath[1], S. Bruhn[2], H. Purnhagen[2], M. Eckert[1], J. Torres[1], S. Brown[1], D. Darcy[3]

Dolby Laboratories: (1) Sydney, Australia, (2) Stockholm, Sweden, (3) San Francisco, USA

Motivation

- 3GPP set out to define the VRStream standard with these goals [1]:
  - Carriage of VR-Audio streams.
  - First-Order Ambisonics Bitrate: 128kbps + 10% metadata overhead
  - High Resolution Bitrate: 256-512kbps + 10% metadata overhead
  - Rendering to speaker array or binaural.

- Our additional goals:
  - Examine the validity of a "Compact Format" to represent arbitrary soundfields: composed of N^{th}-order Ambisonics + M dynamic objects
  - Examine the validity of a multi-channel codec built on one or more EVS monophonic codecs
  - Examine the use of a spatially-whitened first order Ambisonics (FOA) 4-channel signal for carriage by the core codec

SPAR Decoding Matrices

- Decoder upmixer makes use of matrices $\boldsymbol{C}$ and $\boldsymbol{P}$ [2]:

$$X'(k,\omega) = C(k,\omega)\times Y(k,\omega) + P(k,\omega)\times D(k,\omega)$$

where $\boldsymbol{Y}$ is the multi-channel signal from the EVS decoders, and $\boldsymbol{D}$ is a set of decorrelated signals derived from the W channel

- $\boldsymbol{C}$ = least-mean-squares upmix matrix (per time-frequency tile)
- $\boldsymbol{P}$ = diffuse energy compensation matrix

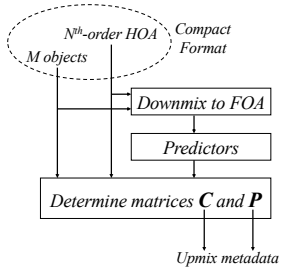
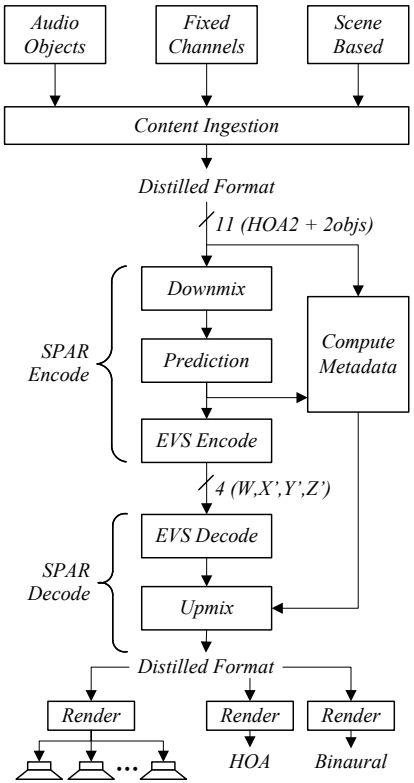


Fig 1. Metadata Calculation in SPAR Encoder

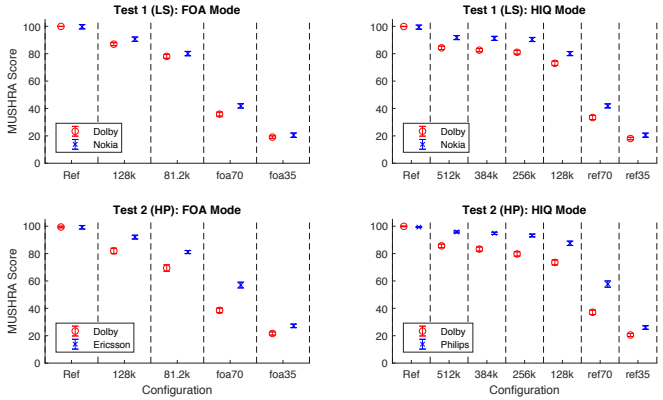
Codec Block Diagram



Subjective evaluation

- Listening test was carried out for the following playback conditions:
  - (LS) 7.1.4 Loudspeaker array
  - (HP) Headphones, with binaural rendering
- Test signals were grouped in 2 types:
  - (FOA) First Order Ambisonics
  - (HIQ) High resolution formats – HOA, Dynamic Objects, Multichannel
- MUSHRA testing methodology, subjects pre- and post-screened
- Tests carried out by Dolby and several independent cross-check labs

Results



* Number of listeners: (Dolby: 10-12, Nokia 9, Ericsson 10, Philips 12)

Thanks to Nokia, Ericsson and Philips for providing test subjects

[1] 3GPP. Virtual Reality (VR) streaming audio; Characterization test results. Technical Specification (TS) 26.818, 3rd Generation Partnership Project (3GPP), 09 2018. Version 15.0.0.
[2] H. Purnhagen, T. Hirvonen, L. Villemoes, J. Samuelsson, and J. Klejsa. Immersive audio delivery using joint object coding. In Audio Engineering Society Convention 140, May 2016.