CLAIMS

What is claimed is:

1. A method of encoding an immersive voice and audio services (IVAS) bitstream, the method comprising:

receiving, using one or more processors, an input audio signal;

downmixing, using the one or more processors, the input audio signal into one or more downmix channels and spatial metadata associated with one or more channels of the input audio signal;

obtaining, using the one or more processors, a set of one or more bitrates for the downmix channels and a set of quantization levels for the spatial metadata from a bitrate distribution control table;

determining, using the one or more processors, a combination of the one or more bitrates for the downmix channels;

determining, using the one or more processors, a metadata quantization level from the set of metadata quantization levels;

quantizing and coding, using the one or more processors, the spatial metadata using the metadata quantization level;

generating, using the one or more processors and the combination of one or more bitrates, a downmix bitstream for the one or more downmix channels;

combining, using the one or more processors, the downmix bitstream, the quantized and coded spatial metadata and the set of quantization levels into the IVAS bitstream.

2. The method of claim 1, wherein the input audio signal is a four-channel first order Ambisonic (FoA) audio signal, three-channel planar FoA signal or a two-channel stereo audio signal.

3. The method of claim 1 or 2, wherein the one or more bitrates are bitrates of one or more instances of a mono audio coder/decoder (codec).

4. The method of claim 1 or 2, wherein the mono audio codec is an enhanced voice services (EVS) codec and the downmix bitstream is an EVS bitstream.

5. The method of claim 1 or 2, wherein obtaining, using the one or more processors, one or more bitrates for the downmix channels and the spatial metadata using a bitrate distribution control table, further comprises:

identifying a row in the bitrate distribution control table using a table index that includes one or more of a format of the input audio signal, a bandwidth of the input audio signal, an allowed spatial coding tool, a transition mode and a mono downmix backward compatible mode; and

extracting, from the identified row of the bitrate distribution control table, one or more of a target bitrate, a bitrate ratio, a minimum bitrate and bitrate deviation steps, wherein the bitrate ratio indicates a ratio in which a total bitrate is to be distributed between the downmix audio signal channels, the minimum bitrate is a value below which the total bitrate is not allowed to go and the bitrate deviation steps are target bitrate reduction steps when a first priority for the downmix signals is higher than or equal to, or lower, than a second priority of the spatial metadata; and

wherein determining the combination of the one or more bitrates for the downmix channels and the spatial metadata are based on one or more of the target bitrate, the bitrate ratio, the minimum bitrate and the bitrate deviation steps.

6. The method of claim 1 or 2, wherein quantizing and coding the spatial metadata for the one or more channels of the input audio signal using a set of metadata quantization levels is performed in a quantization loop that applies increasingly coarse quantization strategies based on a difference between a target metadata bit rate and an actual metadata bitrate.

7. The method of claim 1 or 2, wherein the quantization is determined in accordance with a mono codec priority and a spatial metadata priority based on properties extracted from the input audio signal and channel banded co-variance values.

8. The method of claim 1 or 2, wherein the input audio signal is a stereo signal and the downmix signals include a representation of a mid-signal, residuals from the stereo signal and the spatial metadata.

9. The method of claim 1 or 2, wherein the spatial metadata includes prediction coefficients (PR), cross-prediction coefficients (C) and decorrelation coefficients (P) for a spatial

reconstructor (SPAR) format, and prediction coefficients (P) and decorrelation coefficients (PR) for a complex advanced coupling (CACPL) format.

10. The method of claim 1 or 2, further comprising storing, streaming, or outputting the IVAS bitstream.

11. A system comprising:

one or more processors; and

a non-transitory computer-readable medium storing instructions that, upon execution by the one or more processors, cause the one or more processors to perform operations of any one of the method claims 1-10.

12. A non-transitory, computer-readable medium storing instructions that, upon execution by one or more processors, cause the one or more processors to perform operations of any one of the method claims 1-10.