CLAIMS

1) A method (600) for encoding a multi-channel input signal (101) which comprises N different channels; wherein the input signal (101) comprises a sequence of frames; wherein the method (600) comprises, for a current frame of the sequence of frames,

- determining (601) whether the current frame is an active frame or an inactive frame using a signal and/or a voice activity detector;
- determining (602) a downmix signal (103) based on the multi-channel input signal (101); wherein the downmix signal (103) comprises less than or equal to N channels;
- determining (603) upmixing metadata (105) comprising a set of parameters for generating, based on the downmix signal (103), a reconstructed multi-channel signal (111) comprising N channels; wherein the upmixing metadata (105) is determined in dependance of whether the current frame is an active frame or an inactive frame; and
- encoding (604) the upmixing metadata (105) into a bitstream,

wherein the method (600) further comprises

- determining a covariance, in particular a covariance matrix, for the current frame of the multi-channel input signal (101), by performing temporal smoothing over multiple frames from the sequence of frames, in dependance of whether the current frame is an active frame or an inactive frame; and
- determining one or more parameters of the upmixing metadata (105) based on the covariance for the current frame.

2) A system comprising:
one or more processors; and
a computer-readable medium storing instructions that, when executed by the one or more processors, cause the one or more processors to perform operations of claim 1.

3) The method (600) of claim 1, wherein the method (600) comprises

- determining that the current frame is an active frame following a set of one or more preceding inactive frames from the sequence of frames by more than a pre-determined burst duration threshold; and

- determining the covariance for the current frame based on samples, in particular only based on samples, from the current frame.