Claims

1. A method of processing a compressed representation of a spatial audio signal for generating a reconstructed representation of the spatial audio signal, wherein the compressed representation comprises a channel-based audio signal with a predefined number of channels and metadata, the metadata comprising direction information and energy information, with the direction information comprising indications of directions of arrival of one or more audio elements in an audio scene and the energy information comprising, for at least one frequency subband, respective indications of signal power associated with the directions of arrival, the method comprising:

generating audio signals of the one or more audio elements based on the channel-based audio signal, the direction information, and the energy information; and

generating a residual audio signal from which the one or more audio elements are substantially absent, based on the channel-based audio signal, the direction information, and the energy information,

wherein generating audio signals of the one or more audio elements comprises:

determining coefficients of an inverse mixing matrix M for mapping the channel-based audio signal to an intermediate representation comprising the residual audio signal and the audio signals of the one or more audio elements, based on the direction information and the energy information,

wherein determining the coefficients of the inverse mixing matrix M comprises:

determining, for each of the one or more audio elements, a panning vector $Pan_{down}(dir)$ for panning the audio element to the channels of the channel-based audio signal, based on the direction of arrival dir of the audio element;

determining a mixing matrix E that would be used for mapping the residual audio signal and the audio signals of the one or more audio elements to the channels of the channel-based audio signal, based on the determined panning vectors;

determining a covariance matrix S for the intermediate representation based on the energy information; and

determining the coefficients of the inverse mixing matrix M based on the mixing matrix E and the covariance matrix S.

2. The method according to claim 1, wherein an indication of signal power relates, for a given direction of arrival and a given frequency subband, to a fraction of signal power in the given

frequency subband for the given direction of arrival in relation to the total signal power in the given frequency subband.

3. The method according to claim 1 or 2, wherein the energy information includes indications of signal power for each of a plurality of frequency subbands.

4. The method according to any one of claims 1 to 3, further comprising:

panning the audio signals of the one or more audio elements to a set of channels of an output audio format; and

generating a reconstructed multichannel audio signal in the output audio format based on the panned one or more audio elements and the residual signal.

5. The method according to any one of claims 1 to 4, wherein the mixing matrix E is determined according to

$$E = (I_N|Pan_{down}(dir_1)| \dots |Pan_{down}(dir_P))$$

where I_N is an $N \times N$ identitiy matrix, with N indicating the number of channels of the channel-based signal, $Pan_{down}(dir_p)$ is the panning vector for the p-th audio element with associated direction of arrival dir_p that would pan the p-th audio element to the N channels of the channel-based signal, with $p = 1, \dots, P$ indicating a respective one among the one or more audio elements and P indicating the total number of the one or more audio elements.

6. The method according to claim 5, wherein the covariance matrix S is determined as a diagonal matrix according to

$$\{S\}_{n,n} = \mathrm{rms}(Pan_{down})_n \left(1 - \sum_{p=1}^{P} e_p\right)$$

for $1 \leq n \leq N$, and

$$\{S\}_{N+p,N+p} = e_p$$

for $1 \leq p \leq P$,

where e_p is the signal power associated with the direction of arrival of the p-th audio element.

7. The method according to any one of claims 1 to 6, wherein determining the coefficients of the inverse mixing matrix based on the mixing matrix and the covariance matrix involves determining a pseudo inverse based on the mixing matrix and the covariance matrix.

8. The method according to any one of claims 1 to 7, wherein the inverse mixing matrix M is determined according to

$$M = S \times E^* \times (E \times S \times E^*)^{-1}$$

where "×" indicates the matrix product and "$*$" indicates the conjugate transpose of a matrix.

9. The method according to any one of the preceding claims, wherein the channel-based audio signal is a first-order Ambisonics signal.

10. A program comprising instructions that, when executed by a processor, cause the processor to carry out all steps of the method according to any one of the preceding claims.

11. A computer-readable storage medium storing the program according to claim 10.

12. An apparatus comprising a processor and a memory coupled to the processor, wherein the processor is adapted to carry out all steps of the method according to any one of claims 1 to 9.