

Claims

1. A method of processing a spatial audio signal for generating a compressed representation of the spatial audio signal, the method comprising:
- 5 analyzing the spatial audio signal to determine directions of arrival for one or more audio elements in an audio scene represented by the spatial audio signal;
- for at least one frequency subband of the spatial audio signal, determining respective indications of signal power associated with the determined directions of arrival;
- generating metadata comprising direction information and energy information, with the
- 10 direction information comprising indications of the determined directions of arrival of the one or more audio elements and the energy information comprising respective indications of signal power associated with the determined directions of arrival;
- generating a channel-based audio signal with a predefined number of channels based on the spatial audio signal; and
- 15 outputting, as the compressed representation of the spatial audio signal, the channel-based audio signal and the metadata.
2. The method according to claim 1, wherein analyzing the spatial audio signal is based on a plurality of frequency subbands of the spatial audio signal.
- 20 3. The method according to claim 1 or 2, wherein analyzing the spatial audio signal involves applying scene analysis to the spatial audio signal.
4. The method according to claim 3, wherein the spatial audio signal is a multichannel
- 25 audio signal; or
- wherein the spatial audio signal is an object-based audio signal and the method further comprises converting the object-based audio signal to a multichannel audio signal prior to applying the scene analysis.
- 30 5. The method according to any one of the preceding claims, wherein an indication of signal power associated with a given direction of arrival relates to a fraction of signal power in the frequency subband for the given direction of arrival in relation to the total signal power in the frequency subband.
- 35 6. The method according to any one of the preceding claims, wherein the indications of signal power are determined for each of a plurality of frequency subbands and relate, 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.

7. The method according to any one of the preceding claims, wherein analyzing the spatial audio signal, determining respective indications of signal power, and generating the channel-based audio signal are performed on a per-time-segment basis.

8. The method according to any one of the preceding claims, wherein analyzing the spatial audio signal, determining respective indications of signal power, and generating the channel-based audio signal are performed based on a time-frequency representation of the spatial audio signal.

9. The method according to any one of claims 1 to 3 or 5 to 8, wherein the spatial audio signal is an object-based audio signal that comprises a plurality of audio objects and associated direction vectors;

wherein the method further comprises generating the multichannel audio signal by panning the audio objects to a predefined set of audio channels, wherein each audio object is panned to the predefined set of audio channels in accordance with its direction vector; and

wherein the channel-based audio signal is a downmix signal generated by applying a downmix operation to the multichannel audio signal.

10. The method according to any one of claims 1 to 3 or 5 to 8, wherein the spatial audio signal is a multichannel audio signal; and

wherein the channel-based audio signal is a downmix signal generated by applying a downmix operation to the multichannel audio signal.

11. 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.

5 12. The method according to claim 11, wherein an indication of signal power associated with a given direction of arrival relates to a fraction of signal power in the frequency subband for the given direction of arrival in relation to the total signal power in the frequency subband.

10 13. The method according to claim 11 or 12, wherein the energy information includes indications of signal power for each of a plurality of frequency subbands and 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.

15 14. The method according to any one of claims 11 to 13, 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.

20 15. The method according to any one of claims 11 to 14, 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
25 audio signals of the one or more audio elements, based on the direction information and the energy information.

 16. The method according to claim 15, wherein determining the coefficients of the inverse mixing matrix M comprises:
30 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
35 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 .

17. The method according to claim 16, wherein the mixing matrix E is determined
5 according to

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

- where I_N is an $N \times N$ identity 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
10 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.

18. The method according to claim 17, wherein the covariance matrix S is determined as a diagonal matrix according to

$$15 \quad \{S\}_{n,n} = \text{rms}(Pan_{down})_n \left(1 - \sum_{p=1}^P e_p \right)$$

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
20 element.

19. The method according to any one of claims 16 to 18, 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.

25

20. The method according to any one of claims 16 to 19, wherein the inverse mixing matrix M is determined according to

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

- where " $\times$ " indicates the matrix product and " $*$ " indicates the conjugate transpose of a
30 matrix.

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

22. 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.

23. A computer-readable storage medium storing the program according to claim 22.

5

24. 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 21.