Claims

1. An audio decoder (100; 300) for providing a decoded audio representation (122; 322; $\tilde{X}(k,n)$) on the basis of an encoded audio representation (110;310),

wherein the audio decoder comprises a filter (130; 360,370,380,338) for providing an enhanced audio representation (112; 312; $\hat{X}(k,n)$) of the decoded audio representation (122; 322; $\tilde{X}(k,n)$),

wherein the filter is configured to obtain a plurality of scaling values (136;336; M(k,n)), which are associated with different frequency bins or frequency ranges, on the basis of a plurality of spectral values (132; 326; $\tilde{X}(k,n)$) of the decoded audio representation which are associated with different frequency bins or frequency ranges, and

wherein the filter is configured to scale the plurality of spectral values of the decoded audio signal representation ($\tilde{X}(k,n)$), or a pre-processed version thereof, using the scaling values (136;336; M(k,n)), to obtain the enhanced audio representation (122;312;$\hat{X}(k,n)$);

wherein the filter (130; 360,370,380,338) comprises a neural network (380; 600) or a machine learning structure configured to provide the scaling values (136;336; M(k,n)) on the basis of the plurality of spectral values (132; 326; $\tilde{X}(k,n)$) describing the decoded audio representation (122; 322; $\tilde{X}(k,n)$), the plurality of spectral values being associated with different frequency bins or frequency ranges;

wherein the neural network (380;600) or the machine learning structure is trained such that a scaling for one or more of the plurality of spectral values (132; 326; $\tilde{X}(k,n)$) of the spectral decomposition of the decoded audio signal representation (122; 322; $\tilde{X}(k,n)$), or for one or more preprocessed

spectral values which are based on the plurality of spectral values of the spectral decomposition of the decoded audio signal representation, lies within a range between 0 and a predetermined maximum value,

wherein the maximum value is greater than 1.

2. The audio decoder (100; 300) according to claim 1,
wherein the filter (130; 360,370,380,338) is adapted to use a configurable processing structure, a configuration of which is based on a machine learning algorithm, in order to provide the scaling values (136;336; M(k,n)).

3. The audio decoder (100;300) according to claim 1 or claim 2,

wherein the filter (130; 360,370,380,338) is configured to determine the scaling values (136;336; M(k,n)) solely on the basis of the plurality of spectral values (132; 326; $\tilde{X}(k,n)$) of the decoded audio representation (122; 322; $\tilde{X}(k,n)$) in a plurality of frequency bins or frequency ranges.

4. The audio decoder (100;300) according to one of claims 1 to 3,
wherein the filter (130; 360,370,380,338) is configured to obtain magnitude values $\left|\hat{X}(k,n)\right|$ of the enhanced audio representation according to

$$\left|\hat{X}(k,n)\right| = M(k,n) * \left|\tilde{X}(k,n)\right|,$$

wherein M(k,n) is a scaling value,

wherein k is a frequency index,

wherein n is a time index,

wherein $\left|\tilde{X}(k,n)\right|$] is a magnitude value of a spectral value of the plurality of spectral values of the decoded audio representation; or

wherein the filter is configured to obtain values $\hat{X}(k,n)$ of the enhanced audio representation according to

$$\hat{X}(k,n) = M(k,n) * \tilde{X}(k,n),$$

wherein M(k,n) is a scaling value,

wherein k is a frequency index,

wherein n is a time index,

wherein $\tilde{X}(k,n)$ is a spectral value of the plurality of spectral values of the decoded audio representation.

5. The audio decoder (100;300) according to one of claims 1 to 4, wherein the filter (130; 360,370,380,338) is configured to obtain the scaling values (136;336; M(k,n)) such that the scaling values cause a scaling or an amplification for one or more of the plurality of spectral values (132; 326; $\tilde{X}(k,n)$) of the decoded audio signal representation (122; 322; $\tilde{X}(k,n)$), or for one or more preprocessed spectral values which are based on the plurality of spectral values (132; 326; $\tilde{X}(k,n)$) of decoded audio signal representation (122; 322; $\tilde{X}(k,n)$).

6. The audio decoder (100;300) according to one of claims 1 to 5, wherein the filter (130; 360,370,380,338) comprises a neural network (380; 600) or a machine learning structure configured to provide the scaling

values (136;336; M(k,n)) on the basis of the plurality of spectral values (132; 326; $\tilde{X}(k,n)$) describing the decoded audio representation (122; 322; $\tilde{X}(k,n)$), the plurality of spectral values being associated with different frequency bins or frequency ranges.

7. The audio decoder (100;300) according to claim 6, wherein input signals (372) of the neural network (380;600) or of the machine learning structure represent the logarithmic magnitudes, amplitude or norm of the plurality of spectral values of the decoded audio representation, the plurality of spectral values being associated with different frequency bins or frequency ranges.

8. The audio decoder (100;300) according to one of claims 6 to 7,

wherein output signals (336) of the neural network (380;600) or of the machine learning structure represent the scaling values (136;336; M(k,n)).

9. The audio decoder (100;300) according to one of claims 6 to 8, wherein the neural network (380;600) or the machine learning structure is trained to limit, to reduce or to minimize a deviation (MSE_{MA}) between a plurality of target scaling values (494, IRM(k,n)) and a plurality of scaling values (484, M(k,n)) obtained using the neural network (380; 580; 600) or using the machine learning structure.

10. The audio decoder (100;300) according to one of claims 6 to 9, wherein the neural network (380;600) or the machine learning structure is trained to limit, to reduce or to minimize a deviation (MSE_{SA}) between a target magnitude spectrum (510), a target amplitude spectrum, a target

absolute spectrum or a target norm spectrum ($|X(k,n)|$) and a magnitude spectrum (592), an amplitude spectrum, an absolute spectrum or a norm spectrum obtained using a scaling of a processed spectrum which uses scaling values (584) that are provided by the neural network (380;580;600) or by the machine learning structure.

11. The audio decoder (100;300) according to one of claims 6 to 10,

wherein the neural network (380;600) or the machine learning structure is trained such that the scaling for one or more of the plurality of spectral values of the spectral decomposition of the decoded audio signal representation, or for one or more preprocessed spectral values which are based on the plurality of spectral values of the spectral decomposition of the decoded audio signal representation, is limited to 2, or is limited to 5, or is limited to 10, or is limited to a predetermined value greater than 1.

12. The audio decoder (100;300) according to one of claims 6 to 11,

wherein the neural network (380;600) or the machine learning structure is trained such that the scaling values are limited to 2, or are limited to 5, or are limited to 10, or are limited to a predetermined value greater than 1.

13. The audio decoder (100;300) according to one of claims 6 to 12,

wherein a number of input features of the neural network (380;600) or of the machine learning structure is larger, at least by a factor of 2, than a number of output values of the neural network or of the machine learning structure.

14. The audio decoder (100;300) according to one of claims 6 to 13, wherein the filter (130; 360,370,380,338) is configured to normalize input features of the neural network or of the machine learning structure to a predetermined mean value and/or to a predetermined variance or standard deviation.

15. The audio decoder (100;300) according to one of claims 1 to 14,

wherein the neural network (380;600) comprises an input layer (610) , one or more hidden layers (612a-612d) and an output layer (614).

16. The audio decoder (100;300) according to claim 15,

wherein the one or more hidden layers (612a-612d) use rectified linear units as activation functions.

17. The audio decoder (100;300) according to one of claims 15 to 16, wherein the output layer (614) uses rectified linear units or bounded rectified linear units or sigmoid functions as activation functions.

18. The audio decoder (100;300) according to one of claims 1 to 17,

wherein the filter (130; 360,370,380,338) is configured to obtain short term Fourier transform coefficients $(132; 326; \tilde{X}(k,n))$ which represent the spectral values of the decoded audio representation, which are associated with different frequency bins or frequency ranges.

19. The audio decoder (100;300) according to one of claims 1 to 18,

wherein the filter (130; 360,370,380,338) is configured to derive logarithmic magnitude, amplitude, absolute or norm values (372) and to determine the scaling values (136;336; M(k,n)) on the basis of the logarithmic magnitude, amplitude, absolute or norm values, wherein the logarithmic magnitude values comprise logarithmic magnitude values of the absolute values of spectral domain representations of the decoded audio representation.

20. The audio decoder (100;300) according to one of claims 1 to 18,
wherein the filter (130; 360,370,380,338) is configured to determine a plurality of scaling values (136;336; M(k,n)) associated with a current frame on the basis of the plurality of spectral values (132; 326; $\tilde{X}(k,n)$) of the decoded audio representation (122; 322; $\tilde{X}(k,n)$), which are associated with different frequency bins or frequency ranges, of the current frame, and on the basis of the plurality of spectral values (132; 326; $\tilde{X}(k,n-1)$) of the decoded audio representation (122; 322; $\tilde{X}(k,n-1)$), which are associated with different frequency bins or frequency ranges, of one or more frames preceding the current frame.

21. The audio decoder (100;300) according to one of claims 1 to 20,
wherein the filter (130; 360,370,380,338) is configured to determine a plurality of scaling values associated with a current frame on the basis of the plurality of spectral values (132; 326; $\tilde{X}(k,n+1)$) of the decoded audio representation (122; 322; $\tilde{X}(k,n+1)$), which are associated with different frequency bins or frequency ranges, of one or more frames following the current frame.

22. An apparatus (200;400;500) for determining a set of values defining characteristics of a filter (130; 360,370,380,338) for providing an enhanced audio representation (112;312; $\hat{X}(k,n)$) on the basis of a decoded audio representation (122;322),

wherein the apparatus is configured to obtain spectral values (132;326;$|\tilde{X}(k,n)|$) of the decoded audio representation (122;322), which are associated with different frequency bins or frequency ranges, and

wherein the apparatus is configured to determine the set of values (382;412;512) defining the characteristics of the filter (130; 360,370,380,338), such that scaling values (136;336;484;584) provided by the filter on the basis of the spectral values of the decoded audio representation, which are associated with different frequency bins or frequency ranges, approximate target scaling values (494), or

wherein the apparatus is configured to determine the set of values (382;412;512) defining the characteristics of the filter (130; 360,370,380,338) such that a spectrum obtained by the filter on the basis of the spectral values (132;326;$|\tilde{X}(k,n)|$) of the decoded audio representation (122;322), which are associated with different frequency bins or frequency ranges and using scaling values (136;336;484;584) obtained on the basis of the decoded audio representation (122;322) approximates a target spectrum (510);

wherein the apparatus is configured to train a machine learning structure (380;480;580) such that the magnitude scaling for spectral values of the decoded audio signal representation, or for one or more preprocessed spectral values which are based on the spectral values of decoded audio signal representation, is limited to lie within a range between 0 and a predetermined maximum value,

wherein the maximum value is greater than 1.

23. The apparatus (200;400) according to claim 22,
wherein the apparatus is configured to train a machine learning structure (380; 480;580), which is a part of the filter (130; 360,370,380,338) and which provides scaling values (136;336; 484; 584; M(k,n)) for scaling magnitude values (362,372;462,472;562) of the decoded audio signal (122;322;432; 532) or spectral values (326;446;546) of the decoded audio signal, to reduce or minimize a deviation (MSE_{MA}) between a plurality of target scaling values (494; IRM(k,n)) and a plurality of scaling values (136;336;484;584; M(k,n)) obtained using the neural network on the basis of spectral values (326;446;546) of a decoded audio representation, which are associated with different frequency bins or frequency ranges.

24. The apparatus (200;500) according to claim 22,
wherein the apparatus is configured to train a machine learning structure (380;480;580) to reduce or minimize a deviation (MSE_{SA}) between a target spectrum (510;$|X(k,n)|$) and a spectrum (592; $|\hat{X}(k,n)|$) obtained using a scaling of a processed spectrum (532;546) which uses scaling values (584) that are provided by the machine learning structure.

25. The apparatus (200;400;500) according to one of claims 22 to 24,

wherein the apparatus is configured to train the machine learning structure (380;480;580) such that a scaling for spectral values of the decoded audio signal representation, or for one or more preprocessed spectral values which are based on the spectral values of decoded audio signal representation, lies within a range between 0 and 2 or lies within a range between 0 and 5 or lies within a range between 0 and 10.

26. A method (900) for providing an enhanced audio representation on the basis of an encoded audio representation,

wherein the method comprises providing (910) a decoded audio representation ($\tilde{X}(k,n)$] of the encoded audio representation,

wherein the method comprises obtaining (920) a plurality of scaling values (M(k,n)), which are associated with different frequency bins or frequency ranges, on the basis of spectral values of the decoded audio representation which are associated with different frequency bins or frequency ranges, and wherein the method comprises scaling (930) spectral values of the decoded audio signal representation ($\tilde{X}(k,n)$), or a pre-processed version thereof, using the scaling values (M(k,n)), to obtain the enhanced audio representation ($\hat{X}(k,n)$);

wherein the method comprises providing, using a neural network (380; 600) or a machine learning structure, the scaling values (136;336; M(k,n)) on the basis of a plurality of spectral values (132; 326; $\tilde{X}(k,n)$) describing the decoded audio representation (122; 322; $\tilde{X}(k,n)$), spectral values which are associated with different frequency bins or frequency ranges;

wherein the neural network (380;600) or the machine learning structure is trained such that a scaling for one or more spectral values (132; 326; $\tilde{X}(k,n)$) of the spectral decomposition of the decoded audio signal representation (122; 322; $\tilde{X}(k,n)$), or for one or more preprocessed spectral values which are based on the spectral values of the spectral decomposition of the decoded audio signal representation, lies within a range between 0 and a predetermined maximum value,

wherein the maximum value is greater than 1.

27. A method (1000) for determining a set of values defining characteristics of a filter for providing an enhanced audio representation ($\hat{X}(k,n)$) on the basis of a decoded audio representation,

wherein the method comprises obtaining (1010) spectral values ($\left|\tilde{X}(k,n)\right|$) of the decoded audio representation, which are associated with different frequency bins or frequency ranges, and

wherein the method comprises determining (1020) the set of values defining the characteristics of the filter, such that scaling values provided by the filter on the basis of the spectral values of the decoded audio representation, which are associated with different frequency bins or frequency ranges, approximate target scaling values, or

wherein the method comprises determining (1030) the set of values defining the characteristics of the filter such that a spectrum obtained by the filter on the basis of the spectral values of the decoded audio representation, which are associated with different frequency bins or frequency ranges and using scaling values obtained on the basis of the decoded audio representation approximates a target spectrum;

wherein the method comprises training a machine learning structure (380;480;580) such that the magnitude scaling for spectral values of the decoded audio signal representation, or for one or more preprocessed spectral values which are based on the spectral values of decoded audio signal representation, is limited to lie within a range between 0 and a predetermined maximum value,

wherein the maximum value is greater than 1.

28. A computer program for performing the method of claim 26 or 27 when the computer program runs on a computer.

29. An audio decoder (100; 300) for providing a decoded audio representation (122; 322; $\tilde{X}(k,n)$) on the basis of an encoded audio representation (110;310),

wherein the audio decoder comprises a filter (130; 360,370,380,338) for providing an enhanced audio representation (112; 312; $\hat{X}(k,n)$) of the decoded audio representation (122; 322; $\tilde{X}(k,n)$),

wherein the filter is configured to obtain a plurality of scaling values (136;336; M(k,n)), which are associated with different frequency bins or frequency ranges, on the basis of spectral values (132; 326; $\tilde{X}(k,n)$) of the decoded audio representation which are associated with different frequency bins or frequency ranges, and

wherein the filter is configured to scale spectral values of the decoded audio signal representation ($\tilde{X}(k,n)$), or a pre-processed version thereof, using the scaling values (136;336; M(k,n)), to obtain the enhanced audio representation (122;312;$\hat{X}(k,n)$);

wherein the filter (130; 360,370,380,338) comprises a Neural network (380; 600) or a machine learning structure configured to provide the scaling values (136;336; M(k,n)) on the basis of a plurality of spectral values (132; 326; $\tilde{X}(k,n)$) describing the decoded audio representation (122; 322; $\tilde{X}(k,n)$), spectral values which are associated with different frequency bins or frequency ranges;

wherein the neural network (380;600) or the machine learning structure is trained such that the scaling for one or more spectral values of the spectral decomposition of the decoded audio signal representation, or for one or more preprocessed spectral values which are based on the spectral values of the spectral decomposition of the decoded audio signal representation, is limited to 2, or is limited to 5, or is limited to 10, or is limited to a predetermined value greater than 1.

30. An audio decoder (100; 300) for providing a decoded audio representation (122; 322; $\tilde{X}(k,n)$) on the basis of an encoded audio representation (110;310),

wherein the audio decoder comprises a filter (130; 360,370,380,338) for providing an enhanced audio representation (112; 312; $\hat{X}(k,n)$) of the decoded audio representation (122; 322; $\tilde{X}(k,n)$),

wherein the filter is configured to obtain a plurality of scaling values (136;336; M(k,n)), which are associated with different frequency bins or frequency ranges, on the basis of spectral values (132; 326; $\tilde{X}(k,n)$) of the decoded audio representation which are associated with different frequency bins or frequency ranges, and

wherein the filter is configured to scale spectral values of the decoded audio signal representation ($\tilde{X}(k,n)$), or a pre-processed version thereof, using the scaling values (136;336; M(k,n)), to obtain the enhanced audio representation (122;312;$\hat{X}(k,n)$);

wherein the filter (130; 360,370,380,338) comprises a Neural network (380; 600) or a machine learning structure configured to provide the scaling values (136;336; M(k,n)) on the basis of a plurality of spectral values (132; 326; $\tilde{X}(k,n)$) describing the decoded audio representation (122;

322; $\tilde{X}(k,n)$), spectral values which are associated with different frequency bins or frequency ranges;

wherein the neural network (380;600) or the machine learning structure is trained such that the scaling values are limited to 2, or are limited to 5, or are limited to 10, or are limited to a predetermined value greater than 1.