Claims

1. A decoder, configured to generate an audio signal from a coded signal representing the audio signal, the decoder including:

a coded signal reader, configured to read the coded signal, thereby providing a plurality of indexes;

a scalar dequantization module, including:

a plurality of quantization index converters, each quantization index converter being configured to convert an index of the plurality of indexes onto a corresponding latent scalar value, so that a plurality of latent scalar values form a first latent audio signal representation of the audio signal; and

a first learnable section to provide a second latent representation from the first latent audio signal representation; and

a second learnable section including at least one learnable layer and configured to generate the audio signal from the second latent audio signal representation.

2. The decoder of claim 1, wherein each quantization index converter is configured to provide one single latent scalar value using at least one codebook which is different from the codebooks used by any other quantization index converter.

3. The decoder of claim 1, wherein all, or at least a multiplicity which is a subset of, the quantization index converters are configured to provide a respective plurality of the scalar values using at least one codebook which is a common codebook.

4. The decoder of any of claims 2-3, wherein at least one quantization index converter is a residual or multi-stage quantization index converter.

5. The decoder of any of claims 2-4, wherein at least one codebook is learnable.

6. The decoder of any of the preceding claims, wherein the second learnable section includes a styling or normalizing learnable element conditioning of the second latent representation or a processed version thereof.

7. The decoder of any of claims 2- 6, wherein at least one codebook has a variable-length representation in the bitstream.

8. The decoder of claim 7, configured so that, at least for two of the latent scalar values, or for all the latent scalar values, more frequent latent scalar values are converted from indexes with a representation that is more compact in the coded signal than indexes mapped onto less frequent scalar values.

9. The decoder of any of claims 2-8, wherein at least one codebook or quantization is non-uniform, where the value range to quantize is divided into unequal intervals, in such a way that more frequent intervals are smaller than less frequent intervals.

10. The decoder of any of the preceding claims, configured to select between at least one first decoding mode and one second decoding mode, wherein the first decoding mode is a first, low-quantization-index-converter-number, decoding mode and the second decoding mode is a second, high-quantization-index-converter-number, decoding mode, wherein the decoder is configured, in the first decoding mode, to provide, to the first learnable section, less latent scalar values in the first decoding mode than in the second decoding mode, the decoder thereby using less quantization index converters in the first decoding mode than in the second decoding mode.

11. The decoder of claim 10, configured to select between at least one first decoding mode and one second decoding mode, wherein the first decoding mode is a first, low-index number, decoding mode and the second decoding mode is a second, high-index number, decoding mode, and configured, in the second, high-index number, decoding mode, to use at least one codebook with a higher number of indexes, with higher resolution, and/or with higher bitlength than in the first, low-index number, decoding mode.

12. The decoder of any of the preceding claims, configured to select between at least one first decoding mode and one second decoding mode, so that:

in the second decoding mode, there are used the plurality of quantization index converters to provide the plurality of scalar values, each quantization index converter

being configured to provide one single scalar value, or a component thereof from one respective index of the plurality of indexes; and

in the first decoding mode, there is used one vectorial quantization index converter to provide multiple scalar values from one single index of the plurality of indexes.

13. The decoder of any of the preceding claims, configured to select between at least one first decoding mode and one second decoding mode, wherein the second decoding mode is multi-stage, with a second number of stages, and the first decoding mode is either single-stage or multistage with a first number of stages smaller than the second number of stages, so that:

in the first decoding mode, there are used the plurality of quantization index converters to provide the plurality of indexes, each quantization index converter being configured to convert one single index onto one single scalar value, or a plurality of indexes in the first number onto one scalar value; and

in second decoding mode, there is used at least one quantization index converter to convert indexes, in the second number, to provide at least one scalar value.

14. The decoder of any of the preceding claims, configured to select between at least one first classification decoding mode and one second classification decoding mode based on on a classification of the input audio signal, wherein the first classification decoding mode is trained for a first class of the classification and the second decoding mode is trained for a second class of the classification.

15. The decoder of any of claims 10-14, configured to select between the at least one first and second decoding mode based on a signalization written in the coded signal.

16. The decoder of any of the preceding claims, wherein the second learnable section is configured to change the dimension of the latent representation from the first latent representation to the second latent representation.

17. The decoder of any of the preceding claims comprising:

a first data provisioner configured to provide first data derived from an input signal;

a first processing block, configured to receive the first data and to output first output data in the given frame,

the decoder further comprising:

at least one conditioning learnable layer configured to process target data, from the second latent representation, to output conditioning feature parameters; and

a styling element, configured to apply the conditioning feature parameters to the first data or normalized first data.

18. The decoder of claim 17, configured to obtain the input signal from noise.

19. The decoder of any of the claims 16-18, further comprising at least one preconditioning learnable layer configured to receive the second latent representation and output target data representing the audio signal.

20. The decoder of any of the claims 16-19, wherein a first convolution layer is configured to convolute the target data or up-sampled target data to obtain first convoluted data using a first activation function.

21. The decoder of any of the claims 17-20, further comprising a normalizing element, which is configured to normalize the first data.

22. The decoder of any of the preceding claims, wherein the second learnable section is pre-trained with respect to the first learnable section.

23. An encoder for generating a coded signal in which an input audio signal is encoded, the encoder comprising:

a first learnable section including at least one learnable layer to provide a first latent representation of the input audio signal,

a scalar quantization module, to quantize the first latent representation, comprising:

a second learnable section to provide, from the first latent representation, a plurality of latent scalar values to be quantized; and

a plurality of quantizers, to provide a plurality of indexes, each quantizer being configured to quantize one single latent scalar value to be quantized and

to provide, from the one single latent scalar value, an index of the plurality of indexes; and

a coded signal writer configured to write the plurality of indexes in the coded signal.

24. The encoder of claim 23, wherein each quantizer, or at least one quantizer, is configured to quantize the respective latent scalar value using at least one codebook which is a quantizer-specific codebook.

25. The encoder of claim 23 or 24, wherein all, or at least a multiplicity which is a subset of the plurality of quantizers, are configured to quantize the respective latent scalar values using at least one codebook which is a common codebook.

26. The encoder of any of claim 23-25, wherein at least one quantizer is a residual or multi-stage quantizer.

27. The encoder of any of claims 23-26, wherein at least one codebook is learnable.

28. The encoder of any of claims 23-27, wherein at least one codebook has a variable-length bitstream representation.

29. The encoder of claim 28, configured so that, at least for two of the latent scalar values, or for a plurality of the latent scalar values, or for all the latent scalar values, more frequent latent scalar values are mapped onto indexes which are are more compact in the coded signal representation than the indexes mapped by less frequent scalar values.

30. The encoder of any of claims 23-29, wherein at least one codebook or quantization is non-uniform, where the value range to quantize is divided into unequal intervals, in such a way that more frequent intervals are smaller than less frequent intervals.

31. The encoder of any of claims 23-30, configured to select between at least a first encoding mode and a second encoding mode, wherein the first encoding mode is a first,

low-quantizers-number, encoding mode and the second encoding mode is a second, high-quantizers-number, encoding mode, the encoder being configured in such a way that, in the first, low-quantizers-number, encoding mode, the plurality of latent scalar values has less latent scalar values than in the second, high-quantizers-number, encoding mode, the encoder thereby using less quantizers in the first, low-quantizers-number, encoding mode than in the second, high-quantizers-number, encoding mode.

32. The encoder of any of claims 23-31, configured to select between at least one first encoding mode and one second encoding mode, wherein the second encoding mode is a second, high-index-number, encoding mode and the first encoding mode is a first, low-index-number, encoding mode, wherein the encoder is configured, in the second, high-index-number, encoding mode, to use at least one codebook with a higher number of indexes, with higher resolution, and/or with higher code-length, and/or with more quantization levels, and/or with higher index bit-lengthl than in the first, low-index-number, encoding mode.

33. The encoder of any of claims 23-32, configured to select between at least one first encoding mode and one second encoding mode, wherein the second encoding mode is a second, expanded-latent, encoding mode, and the first encoding mode is a a first, reduced-latent, encoding mode, wherein the second learnable section is configured, in the second encoding mode, to provide more latent scalar values than in the first encoding mode.

34. The encoder of any of claims 23-33, configured to select between at least a first encoding mode and a second encoding mode, so that:

in the second encoding mode, there are used the plurality of quantizers to provide the plurality of indexes, each quantizer being configured to quantize one single latent scalar value to provide the one index of the plurality of indexes; and

in the first encoding mode, there is used at least one quantizer to quantize multiple latent scalar values onto one single index.

35. The encoder of any of claims 23-34, configured to perform in parallel both a first encoding mode to provide a first coded signal version and a second encoding mode to

provide a second coded signal version, and to select between the first encoding mode and the second encoding mode by choosing to write, in the coded signal, the coded signal version, out of the first and the second coded signal versions, which minimizes the distortion with respect to the input audio signal.

36. The encoder of any of claims 23-35, configured to select between the first encoding mode and the second encoding mode based on a status of a communication link through which the coded signal is transmitted, so as to select, among the first encoding mode and the second encoding mode:

the encoding mode which provides a higher resolution, but higher bitlength, in case the communication link is comparatively highly performing; and

the encoding mode which provides a lower resolution, but lower bitlength, in case the communication link is comparatively poorly performing.

37. The encoder of any of claims 23-36, configured to select between at least one first classification encoding mode and one second classification encoding mode based on a classification of the input audio signal or a processed version thereof, wherein the first classification encoding mode is trained for a first class of the classification and the second classification encoding mode is trained for a second class of the classification.

38. The encoder of claim 37, wherein the first class is an unvoiced class, and the second class is a voiced class, wherein the first classification encoding mode is an unvoiced-oriented mode, and the second classification encoding mode is a voiced-oriented mode.

39. The encoder of any of claims 23-38, wherein the first learnable section is configured to reduce the dimension from the first latent representation to the second latent representation.

40. The encoder of any of claims 23-39, wherein the first learnable section comprises a format definer configured to define a multi-dimensional audio signal representation of the input audio signal, the multi-dimensional audio signal representation of the input audio signal including at least:

a first dimension, so that a plurality of mutually subsequent frames are ordered according to the first dimension; and

a second dimension, so that a plurality of samples of at least one frame are ordered according to the second dimension, to define a plurality of channels,

wherein the multi-dimensional audio signal representation is inputted to the at least one learnable layer of the first learnable section.

41. The encoder of any of claims 23-40, wherein the first learnable section is pre-trained with respect to the second learnable section.