CLAIMS

1) A transform-based speech encoder configured to encode a speech signal into a bitstream; the encoder comprising

- a framing unit configured to receive a set of blocks; wherein the set of blocks comprises a plurality of sequential blocks of transform coefficients; wherein the plurality of blocks is indicative of samples of the speech signal; wherein a block of transform coefficients comprises a plurality of transform coefficients for a corresponding plurality of frequency bins;
- an envelope estimation unit configured to determine a current envelope based on the plurality of sequential blocks of transform coefficients; wherein the current envelope is indicative of a plurality of spectral energy values for the corresponding plurality of frequency bins;
- an envelope quantization unit configured to determine a quantized current envelope by quantizing the current envelope;
- an envelope interpolation unit configured to determine a plurality of interpolated envelopes for the plurality of blocks of transform coefficients, respectively, based on the quantized current envelope and based on a quantized previous envelope; and
- a flattening unit configured to determine a plurality of blocks of flattened transform coefficients by flattening the corresponding plurality of blocks of transform coefficients using the corresponding plurality of interpolated envelopes, respectively; wherein the bitstream is determined based on the plurality of blocks of flattened transform coefficients.

2) The transform-based speech encoder of claim 1, wherein

- the transform-based speech encoder further comprises an envelope gain determination unit configured to determine a plurality of envelope gains for the plurality of blocks of transform coefficients, respectively;
- the transform-based speech encoder further comprises an envelope refinement unit configured to determine a plurality of adjusted envelopes by offsetting spectral

energy values of the plurality of interpolated envelopes in accordance to the plurality of envelope gains, respectively;

- the flattening unit is configured to determine the plurality of blocks of flattened transform coefficients by flattening the corresponding plurality of blocks of transform coefficients using the corresponding plurality of adjusted envelopes, respectively.

3) The transform-based speech encoder of claim 2, wherein the envelope gain determination unit is configured to determine a first envelope gain for a first block of transform coefficients, such that a variance of the flattened transform coefficients of a corresponding first block of flattened transform coefficients derived using a first adjusted envelope is adjusted compared to a variance of the flattened transform coefficients of a corresponding first block of flattened transform coefficients derived using a first interpolated envelope.

4) The transform-based speech encoder of claim 3, wherein the envelope gain determination unit is configured to determine the first envelope gain for the first block of transform coefficients, such that the variance of the flattened transform coefficients of the corresponding first block of flattened transform coefficients derived using the first adjusted envelope is one.

5) The transform-based speech encoder of any one of claims 2 to 4, wherein the envelope gain determination unit is configured to insert gain data indicative of the plurality of envelope gains into the bitstream.

6) The transform-based speech encoder of any one of the preceding claims, wherein

- the current envelope is indicative of a plurality of spectral energy values for a corresponding plurality of frequency bands;
- a frequency band comprises one or more frequency bins;
- the envelope estimation unit is configured to determine the spectral energy value for a particular frequency band based on the transform coefficients of the plurality

of sequential blocks for the particular frequency band.

7) The transform-based speech encoder of claim 6, wherein a number of frequency bins per frequency band increases with increasing frequency.

8) The transform-based speech encoder of any one of claims 6 to 7, wherein the envelope estimation unit is configured to determine the spectral energy value for the particular frequency band based on a root mean squared value of the transform coefficients of the plurality of sequential blocks for the particular frequency band.

9) The transform-based speech encoder of any one of the preceding claims, wherein the envelope quantization unit is configured to insert envelope data into the bitstream indicative of the quantized current envelope.

10) The transform-based speech encoder of any one of the preceding claims, wherein

- a block of transform coefficients comprises MDCT coefficients; and/or
- a block of transform coefficients comprises 256 transform coefficients in 256 frequency bins; and/or
- a set of blocks comprises four or more blocks of transform coefficients.

11) The transform-based speech encoder of any one of the preceding claims, wherein

- transform-based speech encoder is configured to operate in a plurality of different modes comprising a short stride mode and a long stride mode;
- the framing unit, the envelope estimation unit and the envelope interpolation unit are configured to process the set of blocks comprising the plurality of sequential blocks of transform coefficients, when the transform-based speech encoder is operated in the short stride mode; and
- the framing unit, the envelope estimation unit and the envelope interpolation unit are configured to process a set of blocks comprising a single block of transform coefficients, when the transform-based speech encoder is operated in the long stride mode.

12) The transform-based speech encoder of claim 11, wherein, in the long stride mode,

- the envelope estimation unit is configured to determine a current envelope of the single block of transform coefficients comprised within the set of blocks; and
- the envelope interpolation unit is configured to determine an interpolated envelope for the single block of transform coefficients as the current envelope of the single block of transform coefficients.

13) A transform-based speech decoder configured to decode a bitstream to provide a reconstructed speech signal; the decoder comprising

- an envelope decoding unit configured to determine a quantized current envelope from envelope data comprised within the bitstream; wherein the quantized current envelope is indicative of a plurality of spectral energy values for a corresponding plurality of frequency bins; wherein the bitstream comprises data indicative of a plurality of sequential blocks of reconstructed flattened transform coefficients; wherein a block of reconstructed flattened transform coefficients comprises a plurality of reconstructed flattened transform coefficients for the corresponding plurality of frequency bins;
- an envelope interpolation unit configured to determine a plurality of interpolated envelopes for the plurality of blocks of reconstructed flattened transform coefficients, respectively, based on the quantized current envelope and based on a quantized previous envelope; and
- an inverse flattening unit configured to determine a plurality of blocks of reconstructed transform coefficients by providing the corresponding plurality of blocks of reconstructed flattened transform coefficients with a spectral shape, using the corresponding plurality of interpolated envelopes, respectively; wherein the reconstructed speech signal is determined based on the plurality of blocks of reconstructed transform coefficients.

14) The transform-based speech decoder of claim 13, wherein the quantized previous envelope is associated with a plurality of previous blocks of reconstructed transform coefficients, directly preceding the plurality of blocks of reconstructed transform

coefficients.

15) The transform-based speech decoder of any one of claims 13 to 14, wherein

- the plurality of sequential blocks of reconstructed flattened transform coefficients comprises a first block of reconstructed flattened transform coefficients at a first intermediate time instant;
- the envelope interpolation unit is configured to determine a spectral energy value for a particular frequency bin of a first interpolated envelope by interpolating the spectral energy values for the particular frequency bin of the quantized current envelope and of the quantized previous envelope at the first intermediate time instant;
- the first interpolated envelope is associated with the first block of reconstructed flattened transform coefficients.

16) The transform-based speech decoder of claim 15, wherein the envelope interpolation unit is configured to perform one or more of: a linear interpolation, a geometric interpolation, and a harmonic interpolation.

17) The transform-based speech decoder of any one of claims 15 to 16, wherein the envelope interpolation unit is configured to perform the interpolation in a logarithm domain.

18) The transform-based speech decoder of any one of claims 15 to 17, wherein the envelope interpolation unit is configured to determine the spectral energy value for the particular frequency bin of the first interpolated envelope by quantizing the interpolation between the spectral energy values for the particular frequency bin of the quantized current envelope and of the quantized previous envelope.

19) The transform-based speech decoder of any one of claims 15 to 18, wherein

- the plurality of sequential blocks of reconstructed flattened transform coefficients comprises a second block of reconstructed flattened transform coefficients at a second intermediate time instant;

- the envelope interpolation unit is configured to determine a spectral energy value for the particular frequency bin of a second interpolated envelope by interpolating the spectral energy values for the particular frequency bin of the quantized current envelope and of the quantized previous envelope at the second intermediate time instant;
- the second interpolated envelope is associated with the second block of reconstructed flattened transform coefficients;
- the second block of reconstructed flattened transform coefficients is subsequent to the first block of reconstructed flattened transform coefficients; and
- the second intermediate time instant is subsequent to the first intermediate time instant.

20) The transform-based speech decoder of claim 19, wherein a difference between the second intermediate time instant and the first intermediate time instant corresponds to a time interval between the second block of reconstructed flattened transform coefficients and the first block of reconstructed flattened transform coefficients.

21) The transform-based speech decoder of any one of claims 13 to 20, wherein
- the bitstream is indicative of a plurality of envelope gains for the plurality of blocks of reconstructed flattened transform coefficients, respectively;
- the transform-based speech decoder further comprises an envelope refinement unit configured to determine a plurality of adjusted envelopes by applying the plurality of envelope gains to the plurality of interpolated envelopes, respectively;
- the inverse flattening unit is configured to determine the plurality of blocks of reconstructed transform coefficients by providing the corresponding plurality of blocks of reconstructed flattened transform coefficients with a spectral shape, using the corresponding plurality of adjusted envelopes, respectively.

22) A transform-based audio encoder configured to encode an audio signal comprising a first segment into a bitstream; wherein the audio encoder comprises

- a signal classifier configured to identify the first segment from the audio signal as being a speech segment; wherein the first segment is to be encoded by a transform-based speech encoder;
- a transform unit configured to determine a plurality of sequential blocks of transform coefficients based on the first segment; wherein a block of transform coefficients comprises a plurality of transform coefficients for a corresponding plurality of frequency bins; wherein the transform unit is configured to determine long blocks comprising a first number of transform coefficients and short blocks comprising a second number of transform coefficients; wherein the first number is greater than the second number; wherein the blocks of the plurality of sequential blocks are short blocks; and
- a transform-based speech encoder according to any one of claims 1 to 12 configured to encode the plurality of sequential blocks into the bitstream.

23) The transform-based audio encoder of claim 22, further comprising a generic transform-based audio encoder configured to encode a segment of the audio signal other than the first segment.

24) The transform-based audio encoder of claim 23, wherein the generic transform-based audio encoder is an AAC or an HE-AAC encoder.

25) The transform-based audio encoder of any of claims 22 to 24, wherein
- the transform unit is configured to perform an MDCT; and/or
- the first number of samples is 1024; and/or
- the second number of samples is 256.

26) A transform-based audio decoder configured to decode a bitstream indicative of an audio signal comprising a first segment; the audio decoder comprising
- a transform-based speech decoder according to any one of claims 13 to 21 configured to determine a plurality of sequential blocks of reconstructed transform coefficients based on data comprised within the bitstream;

- an inverse transform unit configured to determine a reconstructed first segment based on the plurality of sequential blocks of reconstructed transform coefficients; wherein a block of reconstructed transform coefficients comprises a plurality of reconstructed transform coefficients for a corresponding plurality of frequency bins; wherein the inverse transform unit is configured to process long blocks comprising a first number of reconstructed transform coefficients and short blocks comprising a second number of reconstructed transform coefficients; wherein the first number is greater than the second number; wherein the blocks of the plurality of sequential blocks are short blocks.

27) A method for encoding a speech signal into a bitstream; the method comprising

- receiving a set of blocks; wherein the set of blocks comprises a plurality of sequential blocks of transform coefficients; wherein the plurality of sequential blocks is indicative of samples of the speech signal; wherein a block of transform coefficients comprises a plurality of transform coefficients for a corresponding plurality of frequency bins; determining a current envelope based on the plurality of sequential blocks of transform coefficients; wherein the current envelope is indicative of a plurality of spectral energy values for the corresponding plurality of frequency bins;
- determining a quantized current envelope by quantizing the current envelope;
- determining a plurality of interpolated envelopes for the plurality of blocks of transform coefficients, respectively, based on the quantized current envelope and based on a quantized previous envelope;
- determining a plurality of blocks of flattened transform coefficients by flattening the corresponding plurality of blocks of transform coefficients using the corresponding plurality of interpolated envelopes, respectively; and
- determining the bitstream based on the plurality of blocks of flattened transform coefficients.

28) A method for decoding a bitstream to provide a reconstructed speech signal; the method comprising

- determining a quantized current envelope from envelope data comprised within the bitstream; wherein the quantized current envelope is indicative of a plurality of spectral energy values for a corresponding plurality of frequency bins; wherein the bitstream comprises data indicative of a plurality of sequential blocks of reconstructed flattened transform coefficients; wherein a block of reconstructed flattened transform coefficients comprises a plurality of reconstructed flattened transform coefficients for the corresponding plurality of frequency bins;
- determining a plurality of interpolated envelopes for the plurality of blocks of reconstructed flattened transform coefficients, respectively, based on the quantized current envelope and based on a quantized previous envelope;
- determining a plurality of blocks of reconstructed transform coefficients by providing the corresponding plurality of blocks of reconstructed flattened transform coefficients with a spectral shape, using the corresponding plurality of interpolated envelopes, respectively; and
- determining the reconstructed speech signal based on the plurality of blocks of reconstructed transform coefficients.

29) A method for encoding an audio signal comprising a speech segment into a bitstream; wherein the method comprises

- identifying the speech segment from the audio signal;
- determining a plurality of sequential blocks of transform coefficients based on the speech segment, using a transform unit; wherein a block of transform coefficients comprises a plurality of transform coefficients for a corresponding plurality of frequency bins; wherein the transform unit is configured to determine long blocks comprising a first number of transform coefficients and short blocks comprising a second number of transform coefficients; wherein the first number is greater than the second number; wherein the blocks of the plurality of sequential blocks are short blocks; and
- encoding the plurality of sequential blocks into the bitstream according to claim 27.

30) A method for decoding a bitstream indicative of an audio signal comprising a speech segment; the method comprising

- determining a plurality of sequential blocks of reconstructed transform coefficients based on data comprised within the bitstream according to claim 29; and
- determining a reconstructed speech segment based on the plurality of sequential blocks of reconstructed transform coefficients, using an inverse transform unit; wherein a block of reconstructed transform coefficients comprises a plurality of reconstructed transform coefficients for a corresponding plurality of frequency bins; wherein the inverse transform unit is configured to process long blocks comprising a first number of reconstructed transform coefficients and short blocks comprising a second number of reconstructed transform coefficients; wherein the first number is greater than the second number; wherein the blocks of the plurality of sequential blocks are short blocks.