

Encoding and Decoding of Slot Positions of Events in an Audio Signal Frame

5

Specification

The present invention relates to the field of audio processing and audio coding, in particular to encoding and decoding slot positions of events in an audio signal frame.

10

Audio processing and/or coding has advanced in many ways. In particular, spatial audio applications have become more and more important. Audio signal processing is often used to decorrelate or render signals. Moreover, decorrelation and rendering of signals is employed in the process of mono-to-stereo-upmix, mono/stereo to multi-channel upmix, artificial reverberation, stereo widening or user interactive mixing/rendering.

15

Several audio signal processing systems employ decorrelators. An important example is the application of decorrelating signals in parametric spatial audio decoders to restore specific decorrelation properties between two or more signals that are reconstructed from one or several downmix signals. The application of decorrelators significantly improves the perceptual quality of the output signal, e.g. when compared to intensity stereo. Specifically, the use of decorrelators enables the proper synthesis of spatial sound with a wide sound image, several concurrent sound objects and/or ambience. However, decorrelators are also known to introduce artifacts like changes in temporal signal structure, timbre, etc.

20

25

Other application examples of decorrelators in audio processing are e.g. the generation of artificial reverberation to change the spatial impression or the use of decorrelators in multi-channel acoustic echo cancellation systems to improve the convergence behavior.

30

One important spatial audio coding scheme is Parametric Stereo (PS). Fig. 1 illustrates the structure of a mono-to-stereo decoder. A single decorrelator generates a decorrelated signal D (a “wet” signal) from a mono input signal M (a “dry” signal). The decorrelated signal D is then fed into a mixer along with the signal M. Then, the mixer applies a mixing matrix H to the input signals M and D to generate the output signals L and R. The coefficients in the mixing matrix H can be fixed, signal dependent or controlled by a user.

35

Alternatively, the mixing matrix is controlled by side information that is transmitted along with a downmix and contains the parametric description on how to upmix the signals of the downmix to form the desired multi-channel output. The spatial side information is usually generated during the mono downmix process in an accordant signal encoder.

Spatial audio coding as described above is widely applied, e.g., in Parametric Stereo. A typical structure of a parametric stereo decoder is shown in Fig. 2. In Fig. 2, decorrelation is performed in a transform domain. The spatial parameters can be modified by a user or additional tools, e.g. post-processing for binaural rendering/presentation. In this case, the upmix parameters are combined with the parameters from the binaural filters to compute the input parameters for the mixing matrix.

The output L/R of the mixing matrix H is computed from the mono input signal M and the decorrelated signal D.

$$\begin{bmatrix} L \\ R \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} M \\ D \end{bmatrix}$$

In the mixing matrix, the amount of decorrelated sound fed to the output is controlled on the basis of transmitted parameters, e.g. Inter-Channel Level Differences (ILD), Inter-Channel Correlation/Coherence (ICC) and/or fixed or user-defined settings.

Conceptually, the output signal of the decorrelator output D replaces a residual signal that would ideally allow for a perfect decoding of the original L/R signals. Utilizing the decorrelator output D instead of a residual signal in the upmixer results in a saving of bitrate that would otherwise have been required to transmit the residual signal. The aim of the decorrelator is thus to generate a signal D from the mono signal M, which exhibits similar properties as the residual signal that is replaced by D. Reference is made to the document:

[1] J. Breebaart, S. van de Par, A. Kohlrausch, E. Schuijers, "High-Quality Parametric Spatial Audio Coding at Low Bitrates" in Proceedings of the AES 116th Convention, Berlin, Preprint 6072, May 2004.

Considering MPEG Surround (MPS), structures similar to PS termed One-To-Two boxes (OTT boxes) are employed in spatial audio decoding trees. This can be seen as a generalization of the concept of mono-to-stereo upmix to multichannel spatial audio coding/decoding schemes. In MPS, there also exist Two-To-Three upmix systems (TTT

boxes) that may apply decorrelators depending on the TTT mode of operation. Details are described in the document:

- 5 [2] J. Herre, K. Kjörling, J. Breebaart, et al., "MPEG surround - the ISO/MPEG standard for efficient and compatible multi-channel audio coding," in Proceedings of the 122th AES Convention, Vienna, Austria, May 2007.

10 With respect to Directional Audio Coding (DirAC), DirAC relates to a parametric sound field coding scheme that is not bound to a fixed number of audio output channels with fixed loudspeaker positions. DirAC applies decorrelators in the DirAC renderer, i.e., in the spatial audio decoder to synthesize non-coherent components of sound fields. Directional audio coding is further described in:

- 15 [3] Pulkki, Ville: "Spatial Sound Reproduction with Directional Audio Coding", in J. Audio Eng. Soc., Vol. 55, No. 6, 2007

Regarding state-of-the-art decorrelators, reference is made to documents:

- 20 [4] ISO/IEC International Standard "Information Technology – MPEG audio technologies – Part1: MPEG Surround", ISO/IEC 23003-1:2007.

- [5] J. Engdegard, H. Purnhagen, J. Röden, L. Liljeryd, "Synthetic Ambience in Parametric Stereo Coding" in Proceedings of the AES 116th Convention, Preprint, May 2004.

25 IIR lattice allpass structures are used as decorrelators in spatial audio decoders like MPS [2,4]. Other state-of-the-art decorrelators apply (potentially frequency dependent) delays to decorrelate signals or convolve the input signals e.g. with exponentially decaying noise bursts. For an overview of state-of-the-art decorrelators for spatial audio upmix systems, reference is made to document [5]: "Synthetic Ambience in Parametric Stereo Coding".

30

In general, stereo or multichannel applause-like signals coded/decoded in parametric spatial audio coders are known to result in reduced signal quality. Applause-like signals are characterized by containing rather dense mixtures of transients from different directions. Examples for such signals are applause, the sound of rain, galloping horses, etc. Applause-like signals often also contain sound components from distant sound sources that are perceptually fused into a noise-like, smooth background sound field.

35

Lattice allpass structures employed in spatial audio decoders like MPEG Surround act as artificial reverb generators and are consequently well-suited for generating homogenous, smooth, noise-like, inversive sounds (like room reverberation tails). However, they are examples of sound fields with a non-homogeneous spatio-temporal structure that are still
5 immersing the listener: one prominent example are applause-like sound fields that create listener-envelopment not by only homogeneous noise-like fields, but also by rather dense sequences of single claps from different directions. Hence, the non-homogeneous component of applause sound fields may be characterized by a spatially distributed mixture of transients. These distinct claps are not homogeneous, smooth and noise-like at all.

10 Due to their reverb-like behavior, lattice allpass decorrelators are incapable of generating immersive sound fields with the characteristics, e.g. of applause. Instead, when applied to applause-like signals, they tend to temporally smear the transients in the signal. The undesired result is a noise-like immersive sound field without the distinctive spatio-
15 temporal structure of applause-like sound fields. Further, transient events like a single handclap might evoke ringing artifacts of the decorrelator filters.

USAC (Unified speech and audio coding) is an audio coding standard for coding of speech and audio and a mixture thereof at different bitrates.

20 The perceptual quality of USAC can be further improved in stereo coding of applause and applause-like sounds at bitrates in the range of 32 kbps when parametric stereo coding techniques are applicable. USAC coded applause items tend to exhibit a narrow sound stage and a lack of envelopment if no dedicated applause handling is applied within the
25 codec. To a large extent, stereo coding techniques of USAC and their limitations were inherited from MPEG Surround (MPS). However, USAC does offer a dedicated adaption for the requirement of proper applause handling. Said adaption is named Transient Steering Decorrelator (TSD) and is an embodiment of this invention.

30 Applause signals can be envisioned composed of single, distinct nearby claps temporally separated by a few milliseconds and superimposed noise-like ambience originating from very dense far-off claps. In parametric stereo coding at sensible side-information rate, the granularity of the spatial parameter sets (inter channel level difference, inter channel correlation, etc.) is much too low to ensure a sufficient spatial re-distribution of the single
35 claps, leading to a lack of envelopment. Additionally, the claps are subject to processing by a lattice allpass decorrelator. This inevitably induces a temporal dispersion of the transients and further reduces the subjective quality.

Employing a Transient Steering Decorrelator (TSD) within the USAC decoder results in a modification of MPS processing. The underlying idea of such an approach is to address the applause decorrelation problem as follows:

- 5 - Separate the transients in the QMF domain before the lattice allpass decorrelator, i.e.: split the decorrelator input signal into a transient stream s_2 and a non-transient stream s_1 .
- 10 - Feed the transient stream to a different parameter-controlled decorrelator, which is well-suited for transient mixtures.
- Feed the non-transient stream to the MPS allpass decorrelator.
- 15 - Add the outputs of both decorrelators, D_1 and D_2 to obtain the decorrelated signal D .

Fig. 3 illustrates a One-To-Two (OTT) configuration within the USAC decoder. The U-shaped transient handling box of Fig. 3 comprises a parallel signal path as proposed for the transient handling.

20

Two parameters that guide the TSD process are transmitted as frequency independent parameters from the encoder to the decoder (see Fig. 3):

- 25 - A binary transient/non-transient decision of a transient detector running in the encoder is used to control the transient separation with QMF time slot granularity in the decoder. An efficient lossless coding scheme is utilized for transmitting the transient QMF slot position data.
- 30 - Actual transient decorrelator parameters, which are needed for the transient decorrelator to steer a spatial distribution of transients. The transient decorrelator parameters denote an angle between the downmix and its residual. These parameters are only transmitted for time slots which have been detected at the encoder to contain transients.
- 35 In order to assess the quality of the above-described technology, two MUSHRA listening tests were conducted in a controlled listening test environment using high quality electrostatic STAX headphones. The testing was performed at 32 kbps and 16 kbps stereo configuration. Sixteen expert listeners participated in each of the tests.

Since the USAC test set does not contain applause items, additional applause items have been chosen to demonstrate the benefit of the proposed technology. The items listed in Table 1 have been included in the test:

Table 1: Items of the listening test:

Item	Properties
ARL_applause	applause with low to medium density (MPS testset item)
applause4s	very dense applause containing few distinct claps
Applse_2ch	dense multi-channel applause – front channels (MPS testset item)
Applse_st	dense multi-channel applause – stereo downmix (MPS testset item)
Klatschen	sparse applause signal

Regarding the regular twelve MPEG USAC listening test items, TSD is never active. However, these items do not remain exactly bit-identical since the TSD enable bit (indicating that TSD is off) is additionally included in the bitstream and thus slightly affects the bit-budget for the core-coder. Since these differences are very small, these items were not included in the listening test. Data is provided on the size of these differences to show that these changes are negligible and imperceptible.

A codec tool named inter-TES is part of USAC reference model 8 (RM8). Since this technique has been reported to improve the perceptual quality of transients including applause-like signals, inter-TES was always switched on in every test condition. In such a setting, the best possible quality is insured and the orthogonality of inter-TES and TSD is demonstrated.

The system tests have the following configurations:

- RM8: USAC RM8 system
- CE: USAC RM8 system enhanced by the Transient Steering Decorrelator (TSD)

Fig. 4 and 5 depict the MUSHRA scores along with their 95% confidence intervals for the 32 kbps test scenario. For the test data, Student's t-distribution was assumed. The absolute scores in Fig. 4 show a higher mean score for all items, for four out of five items there is a significant improvement in the 95% confidence sense. No item was degraded versus RM8.

The difference scores for USAC+TSD, as evaluated in a TSD core experiment (CE) with respect to USAC RM8 are plotted in Fig. 5. Here, a significant improvement for all items can be seen.

- 5 For the 16 kbps test setup, Fig. 6 and 7 depict the MUSHRA scores along with their 95% confidence intervals. Student's t-distribution of the data was assumed. The absolute scores in Fig. 6 show higher mean score for every item. For one item, significance in the 95% confidence sense can be seen. No item scored worse than RM8. The difference scores are plotted in Fig. 7. Again, a significant improvement for all items with respect to different
10 data was demonstrated.

The TSD tool is enabled by a bsTsdEnable flag transmitted in the bitstream. If TSD is enabled, the actual separation of transients is controlled by transient detection flags TsdSepData that are also transmitted in the bitstream and which are encoded in
15 bsTsdCodedPos in case TSD is enabled.

In the encoder, the TSD enable flag bsTsdEnable is generated by a segmental classifier. The transient detection flags TsdSepData are set by a transient detector.

- 20 As already pointed out, TSD is not activated for the twelve MPEG USAC test items. For the five additional applause items TSD activation is depicted in Fig. 8, displaying a bsTsdEnable logic state versus time.

If TSD is activated, transients are detected in certain QMF time slots and these are
25 subsequently fed to the dedicated transient decorrelator. For each additional test item, Table 2 lists percentages of slots within TSD activated frames which comprise transients.

Table 2: Transient slot percentage (transient slot density in % of all time slots of TSD frames)

30

Item	Transient slot density (%)
ARL_applause	23.4
Applause4s	20.1
applse_2ch	24.7
applse_st	23.8
Klatschen	21.3

Transmitting transient separation decisions and decorrelator parameters from the encoder to the decoder does require a certain amount of side information. However, this amount is
 5 overcompensated by the bitrate savings originating from the transmission of broadband spatial cues within MPS.

In consequence, the mean MPS+TSD side information bitrate is even lower than the plain MPS side information bitrate in plain USAC as listed in Table 3, first column. In the
 10 proposed configuration, as utilized for assessment of subjective quality, the mean bitrates listed in Table 3, second column, have been measured for TSD:

Table 3: MPS(+TSD) Bitrates in bits/second within a 32 kbps stereo codec scenario:

15

Item	MPS(+TSD) side information mean bitrate (bits/sec.)	
	plain USAC RM8	USAC with TSD
ARL_applause	2966	2345
Applause4s	2754	2278
applse_2ch	3000	2544
applse_st	2735	2253
Klatschen	2950	2495

The computational complexity of TSD arises from

- the transient slot position decoding
- 20 - the transient decorrelator complexity.

Assuming an MPEG Surround spatial frame length of 32 time slots, the slot position decoding requires (64 divisions + 80 multiplications) per spatial frame in the worst case, i.e., $64 \cdot 25 + 80 = 1680$ operations per spatial frame.

25

Ignoring copy operations and conditional statements, the transient decorrelator complexity is given by one complex multiplication per slot and hybrid QMF band.

This leads to the following overall complexity numbers of TSD, shown in comparison to the plain USAC complexity numbers in Table 4:

Table 4:

5 TSD decoder complexity in MOPS and relative to plain USAC decoder complexity:

	plain USAC complexity in MOPS	TSD: transient decorrelator complexity in MOPS	TSD: slot position decoder complexity in MOPS	Σ (TSD complexity) in MOPS	Σ (TSD complexity) relative to plain USAC
16kbps stereo ($f_s=28.8\text{kHz}$)	8.7	0.117	0.024	0.141	1.62 %
32kbps stereo ($f_s=40\text{kHz}$)	13.2	0.163	0.033	0.196	1.48 %

10 In summary, the listening test data clearly shows a significant improvement of subjective quality of applause signals in the difference scores of all items in both operation points. In terms of absolute scores, all items in the TSD condition exhibit a higher mean score. For 32 kbps, a significant improvement exists for four out of five items. For 16 kbps, one item shows significant improvement. None of the items scored worse than RM8. An improvement is achieved at, as can be seen from the data on complexity, negligible
15 computational costs. This further emphasizes the benefit of the TSD tool for USAC.

The above-described Transient Steering Decorrelator significantly improves audio processing in USAC. However, as has also been seen above, a Transient Steering Decorrelator requires information about the existence or non-existence of transients in a
20 particular slot. In USAC, information about time slots may be transmitted on a frame-by-frame basis. A frame comprises several, e.g., 32 time slots. It is therefore appreciated that an encoder also transmits information about which slots comprise transients on a frame-by-frame basis. Reducing the number of bits to be transmitted is critical in audio signal processing. As even a single audio recording comprises a vast number of frames this
25 means that even if the number of bits to be transmitted for each frame is reduced by just a few bits, the overall bit transfer rate can be significantly reduced.

The problem of decoding slot positions of events in an audio signal frame is however not limited to the problem of decoding transients. It would moreover be useful to decode slot

positions of other events as well, such as, whether a slot of an audio signal frame is tonal (or not), whether it comprises noise (or whether it doesn't) and the like. In fact, an apparatus for efficiently encoding and decoding slot positions of events in an audio signal frame would be very useful for a large number of different sorts of events.

5

When this document refers to slots or slot positions of an audio signal frame, slots in this sense may be time slots, frequency slots, time-frequency slots or any other kind of slots. It is furthermore understood that the present invention is not limited to audio processing and audio signal frames in USAC, but instead refers to any kind of audio signal frames and any
10 kind of audio formats, such as MPEG1/2, Layer 3 ("MP3"), Advanced Audio Coding (AAC), and the like. Efficiently encoding and decoding slot positions of events in an audio signal frame would be very useful for any kind of audio signal frame.

It is therefore an object of the present invention to provide an apparatus for encoding slot
15 positions of events in an audio signal frame with a few number of bits. Moreover, it is an object of the present invention to provide an apparatus for decoding the slot positions of events in an audio signal frame, encoded by an apparatus for encoding according to the present invention. The objects of the present invention are achieved by an apparatus for decoding according to claim 1, an apparatus for encoding according to claim 11, a method
20 for decoding according to claim 14 a method for encoding according to claim 15, a computer program for decoding according to claim 16, a computer program for encoding according to claim 17 and an encoded signal according to claim 18.

The present invention assumes that a frame slots number indicating the total number of
25 slots of an audio signal frame and an event slots number indicating the number of slots comprising events of the audio signal frame may be available in a decoding apparatus of the present invention. For example, an encoder may transmit the frame slots number and/or the event slots number to the apparatus for decoding. According to an embodiment, the encoder may indicate the total number of slots of an audio signal frame by transmitting a
30 number which is the total number of slots of an audio signal frame minus 1. The encoder may further indicate the number of slots comprising events of the audio signal frame by transmitting a number which is the number of slots comprising events of the audio signal frame minus 1. Alternatively, the decoder may itself determine the total number of slots of an audio signal frame and the number of slots comprising events of the audio signal frame
35 without information from an encoder.

Based on these assumptions, according to the present invention, the number of slot positions comprising events in an audio signal frame can be encoded and decoded using the following findings:

- 5 Let N be the total number of slots of an audio signal frame, and
let P be the number of slots comprising events of the audio signal frame.

It is assumed that both the apparatus for encoding as well as the apparatus for decoding are aware of the values of N and P.

10

Knowing N and P, it can be derived that there are only $\binom{N}{P}$ different combinations of positions of slots comprising events in an audio signal frame.

- 15 For example, if the slot positions in a frame are numbered from 0 to N-1 and if P=8, then a first possible combination of slot positions with events would be (0, 1, 2, 3, 4, 5, 6, 7), a second one would be (0, 1, 2, 3, 4, 5, 6, 8), and so on, up to the combination (N-8, N-7, N-6, N-5, N-4, N-3, N-2, N-1), so that in total there are $\binom{N}{P}$ different combinations.

- 20 Moreover, the present invention employs the further finding, that an event state number may be encoded by an apparatus for encoding and that the event state number is transmitted to the decoder. If each of the possible $\binom{N}{P}$ combinations is represented by a

- unique event state number and if the apparatus for decoding is aware which event state number represents which combination of slot positions comprising events in an audio signal frame (e.g. by applying an appropriate decoding method), then the apparatus for
25 decoding can decode the slot positions comprising events using N, P and the event state number. For a lot of typical values for N and P, such a coding technique employs fewer bits for encoding slot positions of events compared to other methods (e.g. employing a bit array with one bit for each slot of the frame, wherein each bit indicates whether an event occurred in this slot or not).

30

Stated differently, the problem of encoding the slot positions of events in an audio signal frame can be solved by encoding a discrete number P of positions p_k on a range of $[0 \dots N-1]$, such that the positions are not overlapping $p_k \neq p_h$ for $k \neq h$, with as few bits as possible. Since the ordering of positions does not matter, it follows that the number of unique

combinations of positions is the binominal coefficient $\binom{N}{P}$. The number of required bits is thus

$$bits = \text{ceil} \left(\log_2 \left(\binom{N}{P} \right) \right)$$

5

In an embodiment, an apparatus for decoding is provided, wherein the apparatus for decoding is adapted to conduct a test comparing an event state number or an updated event state number with a threshold value. Such a test may be employed to derive the positions of slots comprising events from an event state number. The test of comparing an event state number with a threshold value may be conducted by comparing, whether the event state number or an updated event state number is greater than, greater than or equal to, smaller than, or smaller than or equal to the threshold value. Furthermore, it is preferred that the apparatus for decoding is adapted to update the event state number or an updated event state number depending on the result of the test.

15

According to an embodiment, an apparatus for decoding is provided which is adapted to conduct the test comparing an event state number or an updated event state number with respect to a particular considered slot, wherein the threshold value depends on the frame slots number, the event slots number and on the position of the considered slot within the frame. By this, the positions of slots comprising events may be determined on a slot-by-slot basis, deciding for each slot of a frame, one after the other, whether the slot comprises an event.

According to a further embodiment, an apparatus for decoding is provided which is adapted to split the frame into a first frame partition comprising a first set of slots of the frame and into a second frame partition comprising a second set of slots of the frame, and wherein the apparatus for decoding is further adapted to determine the positions comprising events for each of the frame partitions separately. By this, the positions of slots comprising events may be determined by repeatedly splitting a frame or frame partitions in even smaller frame partitions.

In the following, embodiments of the present invention are described in more detail with respect to the figures, wherein:

35 Fig. 1 is a typical application of a decorrelator in a mono-to-stereo upmixer;

- Fig. 2 is a further typical application of a decorrelator in a mono-to-stereo upmixer;
- 5 Fig. 3 is a One-To-Two (OTT) system overview including a Transient Steering Decorrelator (TSD);
- Fig. 4 is a diagram illustrating absolute scores for 32 kbps stereo comparing RM8 USAC and USAC RM8+TSD in a TSD core experiment (CE);
- 10 Fig. 5 is a diagram displaying differential scores for 32 kbps stereo comparing USAC employing a Transient Steering Decorrelator versus a plain USAC system;
- 15 Fig. 6 is a diagram displaying absolute scores for 16 kbps stereo comparing RM8 USAC and USAC RM8+TSD in a TSD core experiment (CE);
- Fig. 7 is a diagram displaying differential scores for 16 kbps stereo comparing USAC employing a transient steering decorrelator versus a plain USAC system;
- 20 Fig. 8 displays TSD activity for five additional items depicted as logic status of the bsTsdEnable flag;
- 25 Fig. 9a illustrates an apparatus for decoding positions of slots comprising events in an audio signal frame according to an embodiment of the present invention;
- Fig. 9b illustrates an apparatus for decoding positions of slots comprising events in an audio signal frame according to a further embodiment of the present invention;
- 30 Fig. 9c illustrates an apparatus for decoding positions of slots comprising events in an audio signal frame according to another embodiment of the present invention;
- 35 Fig. 10 is a flowchart illustrating a decoding process conducted by an apparatus for decoding according to an embodiment of the present invention;

- Fig. 11 illustrates a pseudo code implementing the decoding of positions of slots comprising events according to an embodiment of the present invention;
- 5 Fig. 12 is a flow chart illustrating an encoding process conducted by an apparatus for encoding according to an embodiment of the present invention;
- Fig. 13 is a pseudo code depicting a process of encoding positions of slots comprising events in an audio signal frame according to a further embodiment of the invention;
- 10 Fig. 14 illustrates an apparatus for decoding positions of slots comprising events in an audio signal frame according to a further embodiment of the present invention;
- 15 Fig. 15 illustrates an apparatus for encoding positions of slots comprising events in an audio signal frame according to a an embodiment of the present invention;
- Fig. 16 depicts the syntax of MPS 212 Data of USAC according to an embodiment;
- 20 Fig. 17 illustrates the syntax of TsdData of USAC according to an embodiment;
- Fig. 18 illustrates an nBitsTrSlots table depending on MPS frame length;
- 25 Fig. 19 shows a table relating to bsTempShapeConfig of USAC according to an embodiment;
- Fig. 20 depicts the syntax of TempShapeData of USAC according to an embodiment;
- 30 Fig. 21 illustrates a decorrelator block D in an OTT decoding block according to an embodiment;
- Fig. 22 depicts the syntax of EcData of USAC according to an embodiment;
- 35 Fig. 23 illustrates a signal flow chart for the generation of TSD data;

Fig. 9a illustrates an apparatus 10 for decoding positions of slots comprising events in an audio signal frame according to an embodiment of the present invention. The apparatus for decoding 10 comprises an analysing unit 20 and a generating unit 30. A frame slots number FSN, indicating the total number of slots of an audio signal frame, an event slots number ESON indicating the number of slots comprising events of the audio signal frame, and an event state number ESTN are fed into the apparatus for decoding 10. The apparatus for decoding 10 then decodes the positions of slots comprising events by using the frame slots number FSN, the event slots number ESON and the event state number ESTN. Decoding is conducted by the analysing unit 20 and the generating unit 30 which cooperate in the process of decoding. While the analysing unit 20 is responsible for executing tests, e.g. comparing the event state number ESTN with a threshold value, the generating unit 30 generates and updates intermediate results of the decoding process, e.g. an updated event state number.

Furthermore the generating unit 30 generates an indication of a plurality of positions of slots comprising events in the audio signal frame. The particular indication of a plurality of positions of slots comprising events of the audio signal frame may be referred to as an "indication state".

According to an embodiment, the indication of a plurality of positions of slots comprising the events in the audio signal frame may be generated such that at a first point in time, the generating unit 30 indicates for a first slot, whether the slot comprises an event or not, at a second point in time, the generating unit 30 indicates for a second slot, whether the slot comprises an event or not and so on.

According to a further embodiment, the indication of a plurality of positions of slots comprising events may for example be a bit array indicating for each slot of the frame whether it comprises an event.

The analysing unit 20 and the generating unit 30 may cooperate such that both units call each other one or more times in the process of decoding to produce intermediate results.

Fig. 9b illustrates an apparatus for decoding 40 according to an embodiment of the present invention. The apparatus for decoding 40 inter alia differs from the apparatus 10 of Fig. 9a in that it further comprises an audio signal processor 50. The audio signal processor 50 receives an audio input signal and the indication of a plurality of positions of slots comprising the events in the audio signal frame which was generated by a generating unit 45. Depending on the indication, the audio signal processor 50 generates an audio output

signal. The audio signal processor 50 may generate the audio output signal, e.g., by decorrelating the audio input signal. Furthermore the audio signal processor 50 may comprise a lattice IIR decorrelator 54, a transient decorrelator 56 and a transient separator 52 for generating the audio output signal as illustrated in Fig. 3. If the indication of a plurality of positions of slots comprising the events in the audio signal frame indicates that a slot comprises a transient, then the audio signal processor 50 will decorrelate the audio input signal relating to that slot by the transient decorrelator 56. If, however, the indication of a plurality of positions of slots comprising the events in the audio signal frame indicates that a slot does not comprise a transient, then the audio signal processor will decorrelate the audio input signal S relating to that slot by employing the lattice IIR decorrelator 54. The audio signal processor employs the transient separator 52 which decides based on the indication whether a portion of the audio input signal relating to a slot is fed into the transient decorrelator 56 or into the lattice IIR decorrelator 54, depending on whether the indication indicates that the particular slot comprises a transient (decorrelation by the transient decorrelator 56) or whether the slot does not comprise a transient (decorrelation by the lattice IIR decorrelator 54).

Fig. 9c illustrates an apparatus for decoding 60 according to an embodiment of the present invention. The apparatus for decoding 60 differs from the apparatus 10 of Fig. 9a in that it further comprises a slot selector 90. Decoding is done on a slot-by-slot basis deciding for each slot of a frame, one after the other, whether the slot comprises an event. The slot selector 90 decides, which slot of a frame to consider. A preferred approach would be that the slot selector 90 chooses the slots of a frame one after the other.

The slot-by-slot decoding of the apparatus for decoding 60 of this embodiment is based on the following findings, which may be applied for embodiments of an apparatus for decoding, an apparatus for encoding, a method for decoding and a method for encoding positions of slots which comprise events in an audio signal frame. The following findings are also applicable for respective computer programs and encoded signals:

Assume that N is the (total) number of slots of an audio signal frame and P is the number of slots comprising events of the frame (this means that N may be the frame slots number FSN and P may be the event slots number ESON). The first slot of a frame is considered. Two cases may be distinguished:

If the first slot is a slot which does not comprise an event, then, with respect to the remaining N-1 slots of the frame, there are only $\binom{N-1}{P}$ different possible combinations of

the P slot positions comprising an event with respect to the remaining N-1 slots of the frame.

However, if the first slot is a slot comprising an event, then, with respect to the remaining
 5 N-1 slots of the frame, there are only $\binom{N-1}{P-1} = \binom{N}{P} - \binom{N-1}{P}$ different possible combinations of the remaining P-1 slots comprising an event with respect to the remaining N-1 slots of the frame.

Based on this finding, embodiments are further based on the finding that all combinations
 10 with a first slot where an event has not occurred, should be encoded by event state numbers that are smaller than or equal to a threshold value. Furthermore, all combinations with a first slot where an event has occurred, should be encoded by event state numbers that are greater than a threshold value. In an embodiment, all event state numbers may be positive integers or 0 and a suitable threshold value regarding the first slot may be $\binom{N-1}{P}$.

15 In an embodiment, an apparatus for decoding is adapted to determine, whether the first slot of a frame comprises an event by testing, whether the event state number is greater than a threshold value. (Alternatively, the encoding/decoding process of embodiments may also be realized, such that an apparatus for decoding tests, whether the event state number is
 20 greater than or equal to, smaller than or equal to, or smaller than a threshold value.) After analysing the first slot, decoding is continued for the second slot of the frame using adjusted values: Besides adjusting the number of considered slots (which is reduced by one), the number of slots comprising events is also eventually reduced by one (if the first slot did comprise an event) and the event state number is adjusted, in case the event state
 25 number was greater than the threshold value, to delete the portion relating to the first slot from the event state number. The decoding process may be continued for further slots of the frame in a similar manner.

30 In an embodiment, a discrete number P of positions p_k on a range of $[0...N-1]$ is encoded, such that the positions are not overlapping $p_k \neq p_h$ for $k \neq h$. Here, each unique combination of positions on the given range is called a state and each possible position in that range is called a slot. According to an embodiment of an apparatus for decoding, the first slot in the range is considered. If the slot does not have a position assigned to it, then the range can be reduced to N-1, and the number of possible states reduces to $\binom{N-1}{P}$. Conversely, if the

state is larger than $\binom{N-1}{P}$, then it can be concluded that the first slot has a position assigned to it. The following decoding algorithm may result from this:

```

    For each slot h
5      If state >  $\binom{N-h-1}{P}$  then
          Assign a position to slot h
          Update remaining state state := state -  $\binom{N-h-1}{P}$ 
          Reduce number of positions left P := P-1
    End
10  End

```

Calculation of the binomial coefficient on each iteration would be costly. Therefore, according to embodiments, the following rules may be used to update the binomial coefficient using the value from the previous iteration:

$$15 \quad \binom{N}{P} = \binom{N-1}{P} \cdot \frac{N}{N-P} \quad \text{and} \quad \binom{N}{P} = \binom{N}{P-1} \cdot \frac{N-P+1}{P}$$

Using these formulas, each update of the binomial coefficient costs only one multiplication and one division, whereas explicit evaluation would cost P multiplications and divisions on
20 each iteration.

In this embodiment, the total complexity of the decoder is P multiplications and divisions for initialization of the binomial coefficient, for each iteration 1 multiplication, division and if-statement, and for each coded position 1 multiplication, addition and division. Note
25 that in theory, it would be possible to reduce the number of divisions needed for initialization to one. In practice, however, this approach would result in very large integers, which are difficult to handle. The worst case complexity of the decoder is then N+2P divisions and N+2P multiplications, P additions (can be ignored if MAC-operations are used), and N if-statements.

30

In an embodiment, the encoding algorithm employed by an apparatus for encoding does not have to iterate through all slots, but only those that have a position assigned to them. Therefore,

35 For each position p_h , $h=1 \dots P$

$$\text{Update state state} := \text{state} + \binom{p_h - 1}{h}$$

The encoder worst case complexity is $P \cdot (P-1)$ multiplications and $P \cdot (P-1)$ divisions, as well as $P-1$ additions.

5

Fig. 10 illustrates a decoding process conducted by an apparatus for decoding according to an embodiment of the present invention. In this embodiment, decoding is performed on a slot-by-slot basis.

10 In step 110, values are initialized. The apparatus for decoding stores the event state number, which it received as an input value, in variable s . Furthermore, the number of slots comprising events of the frame as indicated by an event slots number is stored in variable p . Moreover the total number of slots contained in the frame as indicated by a frame slots number is stored in variable N .

15

In step 120, the value of $\text{TsdSepData}[t]$ is initialized with 0 for all slots of the frame. The bit array TsdSepData is the output data to be generated. It indicates for each slot position t , whether the slot with the corresponding slot position comprises an event ($\text{TsdSepData}[t] = 1$) or whether it does not ($\text{TsdSepData}[t] = 0$). In step 120 the corresponding values of all slots of the frame are initialized with 0.

20

In step 130 variable k is initialized with the value $N-1$. In this embodiment, the slots of a frame comprising N elements are numbered 0, 1, 2, ..., $N-1$. Setting $k = N-1$ means that the slot with the highest slot number is regarded first.

25

In step 140, it is considered whether $k \geq 0$. If $k < 0$, the decoding of the slot positions has been finished and the process terminates, otherwise the process continues with step 150.

30

In step 150, it is tested whether $p > k$. If p is greater than k , this means that all remaining slots comprise an event. The process continues at step 230 wherein all TsdSepData field values of the remaining slots 0, 1, ..., k are set to 1 indicating that each of the remaining slots comprise an event. In this case, the process terminates afterwards. However, if step 150 finds that p is not greater than k , the decoding process continues in step 160.

35 In step 160, the value $c = \binom{k}{p}$ is calculated. c is used as threshold value.

In step 170, it is tested, whether the (eventually updated) event state number s is greater than or equal to c , wherein c is the threshold value just calculated in step 160.

5 If s is smaller than c , this means that the considered slot (with slot position k) does not comprise an event. In this case, no further action has to be taken, as $TsdSepData[k]$ has already been set to 0 for this slot in step 140. The process then continues with step 220. In step 220, k is set to be $k:=k-1$ and the next slot is regarded.

10 However, if the test in step 170 shows that s is greater than or equal to c , this means that the considered slot k comprises an event. In this case, the event state number s is updated and is set to the value $s := s-c$ in step 180. Furthermore, $TsdSepData[k]$ is set to 1 in step 190 to indicate that slot k comprises an event. Moreover, in step 200, p is set to $p-1$, indicating that the remaining slots to be examined now only comprise $p-1$ slots with events.

15 In step 210, it is tested whether p is equal to 0. If p is equal to 0, the remaining slots do not comprise events and the decoding process finishes. Otherwise, at least one of the remaining slots comprises an event and the process continues in step 220 where the decoding process continues with the next slot ($k-1$).

20 The decoding process of the embodiment illustrated in Fig. 10 generates the array $TsdSepData$ as output value indicating for each slot k of the frame, whether the slot comprises an event ($TsdSepData[k]=1$) or whether it doesn't ($TsdSepData[k]=0$).

25 Returning to Fig. 9c, an apparatus for decoding 60 of an embodiment, wherein the apparatus implements the decoding process illustrated in Fig. 10 comprises a slot selector 90, which decides, which slots to consider. With respect to Fig. 10, such a slot selector would be adapted to execute process steps 130 and 220 of Fig. 10. A suitable analysing unit 70 of this embodiment would be adapted to execute processing steps 140, 150, 170, 30 and 210 of Fig. 10. The generating unit 80 of such an embodiment would be adapted to conduct all other processing steps of Fig. 10.

Fig. 11 illustrates a pseudo code implementing the decoding of the positions of slots comprising events according to an embodiment of the present invention.

35 Fig. 12 illustrates an encoding process conducted by an apparatus for encoding according to an embodiment of the present invention. In this embodiment, encoding is performed on

a slot-by-slot basis. The purpose of the encoding process according to the embodiment illustrated in Fig. 12 is to generate an event state number.

5 In step 310, values are initialized. p_s is initialized with 0. The event state number is generated by successively updating variable p_s. When the encoding process is finished, p_s will carry the event state number. Step 310 also initializes variable k by setting k to k:= number of slots comprising events in a frame - 1.

10 In step 320, variable "slots" is set to slots:=tsdPos[k], wherein tsdPos is an array holding the positions of slots comprising events. The slot positions in the array are stored in ascending order.

In step 330, a test is conducted, testing whether $k \geq \text{slots}$. If this is the case, the process terminates. Otherwise, the process is continued in step 340.

15 In step 340, the value $c = \binom{\text{slots}}{k+1}$ is calculated.

In step 350, variable p_s is updated and set to $p_s := p_s + c$.

20 In step 360, k is set to $k := k - 1$.

Then, in step 370, a test is conducted, testing whether $k \geq 0$. In this case, the next slot k-1 is regarded. Otherwise, the process terminates.

25 Fig. 13 depicts pseudo code, implementing the encoding of positions of slots comprising events according to an embodiment of the present invention.

Fig. 14 illustrates an apparatus for decoding 410 positions of slots comprising events in an audio signal frame according to a further embodiment of the present invention. Again, as in
30 Fig. 9a, a frame slots number FSN, indicating the total number of slots of an audio signal frame, an event slots number ESON indicating the number of slots comprising events of the audio signal frame, and an event state number ESTN are fed into the apparatus for decoding 410. The apparatus for decoding 410 differs from the apparatus of Fig. 9a in that it further comprises a frame partitioner 440. The frame partitioner 440 is adapted to split
35 the frame into a first frame partition comprising a first set of slots of the frame and into a second frame partition comprising a second set of slots of the frame, and wherein the slot positions comprising events are determined separately for each of the frame partitions. By

this, the positions of slots comprising events may be determined by repeatedly splitting a frame or frame partitions in even smaller frame partitions.

The “partition based” decoding of the apparatus for decoding 410 of this embodiment is based on the following concepts, which may be applied for embodiments of an apparatus for decoding, an apparatus for encoding, a method for decoding and a method for encoding positions of slots which comprise events in an audio signal frame. The following concepts are also applicable for respective computer programs and encoded signals:

- 5 Partition based decoding is based on the idea that a frame is split into two frame partitions A and B, each frame partition comprising a set of slots, wherein frame partition A comprises N_a slots and wherein frame partition B comprises N_b slots and such that $N_a + N_b = N$. The frame can be arbitrarily split into two partitions, preferably such that partition A and B have nearly the same total number of slots (e.g., such that $N_a = N_b$ or $N_a = N_b - 1$). By
- 10 splitting the frame into two partitions, the task of determining the slot positions where events have occurred is also split into two subtasks, namely determining the slot positions where events have occurred in frame partition A and determining the slot positions where events have occurred in frame partition B.
- 15 In this embodiment, it is again assumed that the apparatus for decoding is aware of the number of slots of the frame, the number of slots comprising events of the frame and an event state number. To solve both subtasks, the apparatus for decoding should also be aware of the number of slots of each frame partition, the number of slots where events occurred regarding each frame partition and the event state number of each frame partition
- 20 (such an event state number of a frame partition is now referred to as “event substate number”).
- 25 As the apparatus for decoding itself splits the frame into two frame partitions, it per se knows that frame partition A comprises N_a slots and frame partition B comprises N_b slots.

- 30 Determining the number of slots comprising events for each one of both frame partitions is based on the following findings:

- As the frame has been split into two partitions, each of the slots comprising events is now located either in partition A or in partition B. Furthermore, assuming that P is the number of slots comprising events of a frame partition, and N is the total number of slots of the
- 35 frame partition and that $f(P, N)$ is a function that returns the number of different combinations of slot positions of events of a frame partition, then the number of different

combinations of slot positions of events of the whole frame (which has been split into partition A and partition B) is:

Number of slots comprising events in partition A	Number of slots comprising events in partition B	Number of different combinations in the whole audio signal frame with this configuration
0	P	$f(0, N_a) \cdot f(P, N_b)$
1	P-1	$f(1, N_a) \cdot f(P-1, N_b)$
2	P-2	$f(2, N_a) \cdot f(P-2, N_b)$
...	...	...
P	0	$f(P, N_a) \cdot f(0, N_b)$

- 5 Based on the above considerations, according to an embodiment all combinations with the first configuration, where partition A has 0 slots comprising events and where partition B has P slots comprising events, should be encoded with an event state number smaller than a first threshold value. The event state number may be encoded as an integer value being positive or 0. As there are only $f(0, N_a) \cdot f(P, N_b)$ combinations with the first configuration, a
 10 suitable first threshold value may be $f(0, N_a) \cdot f(P, N_b)$.

All combinations with the second configuration, where partition A has 1 slot comprising events and where partition B has P-1 slots comprising events, should be encoded with an event state number greater than or equal to the first threshold value, but smaller than or
 15 equal to a second value. As there are only $f(1, N_a) \cdot f(P-1, N_b)$ combinations with the second configuration, a suitable second value may be $f(0, N_a) \cdot f(P, N_b) + f(1, N_a) \cdot f(P-1, N_b)$. The event state number for combinations with other configurations is determined similarly.

According to an embodiment, decoding is performed by separating a frame into two frame
 20 partitions A and B. Then, it is tested whether an event state number is smaller than a first threshold value. In a preferred embodiment, the first threshold value may be $f(0, N_a) \cdot f(P, N_b)$.

If the event state number is smaller than the first threshold value, it can then be concluded
 25 that partition A comprises 0 slots comprising events and partition B comprises all P slots of the frame where events occurred. Decoding is then conducted for both partitions with the respectively determined number representing the number of slots comprising events of the corresponding partition. Furthermore a first event state number is determined for partition A and a second event state number is determined for partition B which are respectively

used as new event state number. Within this document, an event state number of a frame partition is referred to as an “event substate number”.

However, if the event state number is greater than or equal to the first threshold value, the event state number may be updated. In a preferred embodiment, the event state number may be updated by subtracting a value from the event state number, preferably by subtracting the first threshold value, e.g. $f(0, N_a) \cdot f(P, N_b)$. In a next step, it is tested, whether the updated event state number is smaller than a second threshold value. In a preferred embodiment, the second threshold value may be $f(1, N_a) \cdot f(P-1, N_b)$. If event state number is smaller than the second threshold value, it can be derived that partition A has 1 slot comprising events and partition B has $P-1$ slots comprising events. Decoding is then conducted for both partitions with the respectively determined numbers of slots comprising events of each partition. A first event substate value is employed for the decoding of partition A and a second event substate value is employed for the decoding of partition B. However, if the event state number is greater than or equal to the second threshold value, the event state number may be updated. In a preferred embodiment, the event state number may be updated by subtracting a value from the event state number, preferably $f(1, N_a) \cdot f(P-1, N_b)$. The decoding process is similarly applied for the remaining distribution possibilities of the slots comprising events regarding the two frame partitions.

In an embodiment, an event substate value for partition A and an event substate value for partition B may be employed for decoding of partition A and partition B, wherein both event substate values are determined by conducting the division:

event state value / $f(\text{number of slots comprising events of partition B}, N_b)$

Preferably, the event substate number of partition A is the integer part of the above division and the event substate number of partition B is the remainder of that division. The event state number employed in this division may be the original event state number of the frame or an updated event state number, e.g. updated by subtracting one or more threshold values, as described above.

To illustrate the above described concept of partition based decoding, a situation is considered where a frame has two slots comprising events. Furthermore, if $f(p, N)$ is again the function that returns the number of different combinations of slot positions of events of a frame partition, wherein p is the number of slots comprising events of a frame partition and N is the total number of slots of that frame partition. Then, for each of the possible distributions of the positions, the following number of possible combinations results:

Positions in partition A	Position in partition B	Number of combinations in this configuration
0	2	$f(0, N_a) \cdot f(2, N_b)$
1	1	$f(1, N_a) \cdot f(1, N_b)$
2	0	$f(2, N_a) \cdot f(0, N_b)$

5

It can thus be concluded that if the encoded event state number of the frame is smaller than $f(0, N_a) \cdot f(2, N_b)$, then the slots comprising events must be distributed as 0 and 2. Otherwise, $f(0, N_a) \cdot f(2, N_b)$ is subtracted from the event state number and the result is compared with $f(1, N_a) \cdot f(1, N_b)$. If it is smaller, then positions are distributed as 1 and 1. Otherwise, we have only the distribution 2 and 0 left, and the positions are distributed as 2 and 0.

10

In the following, a pseudo code is provided according to an embodiment for decoding positions of slots comprising certain events (here: “pulses”) in an audio signal frame. In this pseudo code, “pulses_a” is the (assumed) number of slots comprising events in partition A and “pulses_b” is the (assumed) number of slots comprising events in partition B. In this pseudo code, the (eventually updated) event state number is referred to as “state”. The event substate numbers of partitions A and B are still jointly encoded in the “state” variable. According to a joint coding scheme of an embodiment, the event substate number of A (herein referred to as “state_a”) is the integer part of the division $\text{state}/f(\text{pulses_b}, N_b)$ and the event substate number of B (herein referred to as “state_b”) is the remainder of that division. By this, the length (total number of slots of the partition) and the number of encoded positions (number of slots comprising events in the partition) of both partitions can be decoded by the same approach:

15

20

25

Function $x = \text{decodestate}(\text{state}, \text{pulses}, N)$

1. Split vector into two partitions of length N_a and N_b .

2. For pulses_a from 0 to pulses

a. pulses_b = pulses - pulses_a

30

b. if state < $f(\text{pulses_a}, N_a) \cdot f(\text{pulses_b}, N_b)$ then
break for-loop.

c. state := state - $f(\text{pulses_a}, N_a) \cdot f(\text{pulses_b}, N_b)$

3. Number of possible states for partition B is
 $\text{no_states_b} = f(\text{pulses_b}, \text{Nb})$
4. The states, state_a and state_b , of partitions A and B, respectively, are the integer part and the remainder of the division $\text{state}/\text{no_states_b}$.
5. If $\text{Na} > 1$ then the decoded vector of partition A is obtained recursively by
 $\text{xa} = \text{decodestate}(\text{state_a}, \text{pulses_a}, \text{Na})$
 Otherwise ($\text{Na}==1$), and the vector xa is a scalar and we can set $\text{xa}=\text{state_a}$.
6. If $\text{Nb} > 1$ then the decoded vector of partition B is obtained recursively by
 $\text{xb} = \text{decodestate}(\text{state_b}, \text{pulses_b}, \text{Nb})$
 Otherwise ($\text{Nb}==1$), and the vector xb is a scalar and we can set $\text{xb}=\text{state_b}$.
7. The final output x is obtained by merging xa and xb by $\text{x} = [\text{xa} \ \text{xb}]$.

The output of this algorithm is a vector that has a one (1) at every encoded position (i.e. a slot position of a slot comprising an event) and zero (0) elsewhere (i.e. at positions of slots which do not comprise events).

In the following, a pseudo code is provided according to an embodiment for encoding positions of slots comprising events in an audio signal frame which uses similar variable names with a similar meaning as above:

Function $\text{state} \leftarrow \text{encodestate}(\text{x}, \text{N})$

1. Split vector into two partitions xa and xb of length Na and Nb .
2. Count pulses in partitions A and B in pulses_a and pulses_b , and set $\text{pulses}=\text{pulses_a}+\text{pulses_b}$.
3. Set state to 0
4. For k from 0 to $\text{pulses_a}-1$
 $\text{a. state} := \text{state} + f(k, \text{Na}) * f(\text{pulses}-k, \text{Nb})$
5. If $\text{Na} > 1$, encode partition A by
 $\text{state_a} = \text{encodestate}(\text{xa}, \text{Na});$
 Otherwise ($\text{Na}==1$), set $\text{state_a} = \text{xa}$.

```

6. If Nb > 1, encode partition B by
    state_b = encodestate(xb,Nb);
    Otherwise (Nb==1), set state_b = xb.
7. Encode states jointly
5    state := state + state_a*f(pulses_b,Nb) + state_b.

```

Here, it is assumed that, similarly to the decoder algorithm, every encoded position (i.e., a slot position of a slot comprising an event) is identified by a one (1) in vector x and all other elements are zero (0) (i.e., at positions of slots which do not comprise events).

10

The above recursive methods formulated in pseudo code can readily be implemented in a non-recursive way using standard methods.

According to an embodiment of the present invention, function f(p,N) may be realized as a look-up table. When the positions are non-overlapping, such as in the current context, then the number-of-states function f(p,N) is simply the binomial function which can be calculated on-line. There is

15

$$f(p,N) = \frac{N(N-1)(N-2)\dots(N-k)}{k(k-1)(k-2)\dots 1}.$$

20

According to an embodiment of the present invention, both the encoder and the decoder have a for-loop where the product f(p-k,Na)*f(k,Nb) is calculated for consecutive values of k. For efficient computation, this can be written as

$$\begin{aligned}
 f(p-k, N_a) f(k, N_b) &= \frac{N_a(N_a-1)(N_a-2)\dots(N_a-p+k)}{(p-k)(p-k-1)(p-k-2)\dots 1} \cdot \frac{N_b(N_b-1)(N_b-2)\dots(N_b-k)}{k(k-1)(k-2)\dots 1} \\
 25 \quad &= \frac{N_a(N_a-1)(N_a-2)\dots(N_a-p-k+1)}{(p-k+1)(p-k)(p-k-1)\dots 1} \cdot \frac{N_b(N_b-1)(N_b-2)\dots(N_b-k+1)}{(k-1)(k-2)\dots 1} \cdot \frac{p-k+1}{N_a-p-k+1} \cdot \frac{N_a-k}{k} \\
 &= f(p-k+1, N_a) f(k-1, N_b) \cdot \frac{p-k+1}{N_a-p-k+1} \cdot \frac{N_a-k}{k}.
 \end{aligned}$$

In other words, successive terms for subtraction/addition (in step 2b and 2c in the decoder, and in step 4a in the encoder) can be calculated by three multiplications and one division per iteration.

30

Similarly as in the method described before, the state of a long vector (a frame with many slots) may be a very big integer number, easily extending the length of representation in

standard processors. Therefore it will be necessary to use arithmetic functions capable of handling very long integers.

Regarding complexity, the method regarded here is, in difference to the slot-by-slot processes above, a split and conquer-type algorithm. Assuming the input vector length is a power of two, then the recursion has a depth of $\log_2(N)$.

Since the number of pulses remains constant on each depth of the recursion, then the number of iterations of the for-loop is the same at each recursion. It follows that the number of loops is pulses $\cdot \log_2(N)$.

As explained above, each update of the $f(p-k, N_a) \cdot f(k, N_b)$ can be done with three multiplications and one division.

It should be noted that subtractions and comparisons in the decoder can be assumed to be one operation.

It can be readily seen that partitions are merged $\log_2(N)-1$ times. In the joint encoding of states in the encoder, it is thus necessary to multiply and add $\log_2(N)-1$ times. Similarly, at the joint decoding of states in the decoder, it is necessary to divide $\log_2(N)-1$ times.

It should be noted that of the divisions, only the joint encoding of states in the decoder needs divisions where the denominator is a long integer. The other divisions always have relatively short integers in the denominator. Since divisions with long denominators are the most complex operations, those should be avoided when possible.

In summary, the number of long integer arithmetic operations is in the decoder

	Multiplications	$(3 \cdot \text{pulses} + 1) \cdot \log_2(N) - 1$
30	Divisions	$(\text{pulses}+1) \cdot \log_2(N)-1$
	Of which long denominator divisions	$\log_2(N)-1$
	Additions and subtractions	$\text{pulses} \cdot \log_2(N)$

Similarly, in the encoder there are

35	Multiplications	$(3 \cdot \text{pulses} + 1) \cdot \log_2(N) - 1$
	Divisions	$(\text{pulses}+1) \cdot \log_2(N)-1$
	Of which long denominator divisions	0
	Additions and subtractions	$(\text{pulses}+2) \cdot \log_2(N)$

Only $\log_2(N)-1$ divisions with a long denominator are required.

5 In further embodiments, above-described embodiments which comprise or which are adapted to employ recursive processing steps are modified such that some or all of the recursive processing steps are implemented in a non-recursive way using standard methods

10 Fig. 15 illustrates an apparatus for encoding (510) positions of slots comprising events in an audio signal frame according to an embodiment. The apparatus for encoding (510) comprises an event state number generator (530) which is adapted to encode the positions of slots by encoding an event state number. Furthermore the apparatus comprises a slot information unit (520) adapted to provide a frame slots number and an event slots number to the event state number generator (530). The event state number generator may implement one of the above-described methods for encoding.

15 In a further embodiment, an encoded audio signal is provided. The encoded audio signal comprises an event state number. In another embodiment, the encoded audio signal furthermore comprises an event slots number. Moreover, the encoded audio signal frame may also comprise a frame slots number. In the audio signal frame, the positions of slots comprising events in an audio signal frame can be decoded according to one of the above-described methods for decoding. In an embodiment, the event state number, the event slots number and the frame slots number are transmitted such that the positions of slots comprising events in an audio signal frame can be decoded by employing one of the above-described methods.

25 The inventive encoded audio signal can be stored on a digital storage medium or a non-transitory storage medium or can be transmitted on a transmission medium such as a wireless transmission medium or a wired transmission medium such as the Internet.

30 The following explains USAC syntax definitions adapted to support a Transient Steering Decorrelator (TSD) according to an embodiment:

35 Fig. 16 illustrates MPS (MPEG Surround) 212 data. MPS 212 data is a block of data comprising payload for the MPS 212 stereo module. The MPS 212 data comprises TSD data.

Fig. 17 depicts the syntax of TSD data. It comprises the number of transient slots (bsTsdNumTrSlots) and TSD Transient Phase Data (bsTsdTrPhaseData) for the slots in an

MPS 212 data frame. If a slot comprises transient data (TsdSepData[ts] is set to 1) bsTsdTrPhaseData comprises phase data, otherwise bsTsdTrPhaseData[ts] is set to 0.

5 nBitsTrSlots defines the number of bits employed for carrying the number of transient slots (bsTsdNumTrSlots). nBitsTrSlots depends on the number of slots in a MPS 212 data frame (numSlots). Fig. 18 illustrates the relationship of the number of slots in a MPS 212 data frame and the number of bits employed for carrying the number of transient slots.

10 Fig. 19 defines the meaning of tempShapeConfig. tempShapeConfig indicates the operation mode of temporal shaping (STP or GES) or the activation of transient steering decorrelation in the decoder. If tempShapeConfig is set to 0, temporal shaping is not applied at all; if tempShapeConfig is set to 1, Subband Domain Temporal Processing (STP) is applied; if tempShapeConfig is set to 2, Guided Envelope Shaping (GES) is applied; and if tempShapeConfig is set to 3 Transient Steering Decorrelation (TSD) is applied.

Fig. 20 illustrates the syntax of TempShapeData. If bsTempShapeConfig is set to 3, TempShapeData comprises bsTsdEnable indicating that TSD is enabled in a frame.

20 Fig. 21 illustrates a decorrelator block D according to an embodiment. The decorrelator block D in the OTT decoding block comprises a signal separator, two decorrelator structures, and a signal combiner.

D_{AP} means: all-pass decorrelator as defined in subsection 7.11.2.5 (All-Pass Decorrelator).

D_{TR} means: Transient decorrelator.

25 If the TSD tool is active in the current frame, i.e. if (bsTsdEnable==1), the input signal is separated into a transient stream $v_{X,Tr}^{n,k}$ and a non-transient stream $v_{X,nonTr}^{n,k}$ according to:

$$v_{X,Tr}^{n,k} = \begin{cases} v_X^{n,k} & , \text{ if TsdSepData}(n) = 1, 7 \leq k \\ 0 & , \text{ otherwise} \end{cases}$$

$$v_{X,nonTr}^{n,k} = \begin{cases} 0 & , \text{ if TsdSepData}(n) = 1, 7 \leq k \\ v_X^{n,k} & , \text{ otherwise} \end{cases}$$

30 The per-slot transient separation flag TsdSepData(n) is decoded from the variable length code word bsTsdCodedPos by TsdTrPos_dec() as described below. The code word length of bsTsdCodedPos, i.e. nBitsTsdCW, is calculated according to:

$$nBitsTsdCW = \text{ceil} \left(\log_2 \left(\frac{bsFrameLength}{bsTsdNumTrSlots + 1} \right) \right)$$

Returning to Fig. 11, Fig. 11 illustrates the decoding of the TSD transient slot separation data *bsTsdCodedPos* into *TsdSepData[n]* according to an embodiment. An array of length
 5 *numSlots* consisting of '1's for coded transient positions and '0's else, is defined as illustrated in Fig. 11.

If the TSD tool is disabled in the current frame, i.e. if (*bsTsdEnable*==0), the input signal is processed as if *TsdSepData(n)*=0 for all *n*.

10

Transient signal components are processed in a transient decorrelator structure D_{TR} as follows:

$$d_{X,Tr}^{n,k} = \begin{cases} e^{j\varphi_{TSD}^n} \cdot v_{X,Tr}^{n,k} & , \text{ if } bsTsdEnable = 1 \\ 0 & , \text{ otherwise} \end{cases} ,$$

15 where

$$\varphi_{TSD}^n = \pi \cdot 0.25 \cdot bsTsdTrPhaseData(n) .$$

The non-transient signal components are processed in all-pass decorrelator D_{AP} as defined
 20 in the next subsection, yielding the decorrelator output for non-transient signal components,

$$d_{X,nonTr}^{n,k} = \mathbf{D}_{AP} \left\{ v_{X,nonTr}^{n,k} \right\} .$$

The decorrelator outputs are added to form the decorrelated signal containing both
 25 transient and non-transient components,

$$d_X^{n,k} = d_{X,Tr}^{n,k} + d_{X,nonTr}^{n,k} .$$

Fig. 22 illustrates the syntax of *EcData* comprising *bsFrequencyResStrideXXX*. The
 30 syntax element *bsFreqResStride* allows for utilization of broadband cues in MPS. XXX is to be replaced by the value of the data type (CLD, ICC, IPD).

The Transient Steering Decorrelator in the OTT decoder structure provides the possibility to apply a specialized decorrelator to transient components of applause-like signals. The activation of this TSD feature is controlled by the encoder generated bsTsdEnable flag that is transmitted once per frame.

5

TSD data in the two channels to one channel module (R-OTT) of the encoder is generated as follows:

- 10 - Run a semantic signal classifier that detects applause-like signals. The classification result is transmitted once per frame: The bsTsdEnable flag is set to 1 for applause-like signals, otherwise it is set to 0.
- if bsTsdEnable is set to 0 for the current frame, no further TSD data is generated/transmitted for this frame.
- 15 - if bsTsdEnable is set to 1 for current frame, perform the following:
 - Switch on the broadband calculation of the OTT spatial parameters.
 - Detect transients in the current frame (binary decision per MPS time slot).
 - Encode the tsdPosLen transient slot positions in a vector tsdPos according to the following pseudocode, where the slot positions in tsdPos are expected
 - 20 in ascending order. Fig. 13 illustrates a pseudocode for encoding transient slot positions in tsdPosLen.
 - Transmit the number of transient slots ($bsTsdNumTrSlots = (\text{number of detected transient slots}) - 1$).
 - Transmit the encoded transient positions (bsTsdCodedPos).
 - 25 ○ For each transient slot calculate a phase measure that represents the broadband phase difference between the downmix signal and the residual signal.
 - For each transient slot encode and transmit the broadband phase difference measure (bsTsdTrPhaseData).

30

Finally, Fig. 23 illustrates a signal flow chart for the generation of TSD data in the two channels to one channel module (R-OTT).

35 Although some aspects have been described in the context of an apparatus, it is clear that these aspects also represent a description of the corresponding method, where a block or

device corresponds to a method step or a feature of a method step. Analogously, aspects described in the context of a method step also represent a description of a corresponding block or item or feature of a corresponding apparatus.

- 5 Depending on certain implementation requirements, embodiments of the invention can be implemented in hardware or in software. The implementation can be performed using a digital storage medium, for example a floppy disk, a DVD, a CD, a ROM, a PROM, an EPROM, an EEPROM or a FLASH memory, having electronically readable control signals stored thereon, which cooperate (or are capable of cooperating) with a
10 programmable computer system such that the respective method is performed.

- Some embodiments according to the invention comprise a data carrier having electronically readable control signals, which are capable of cooperating with a programmable computer system, such that one of the methods described herein is
15 performed.

- Generally, embodiments of the present invention can be implemented as a computer program product with a program code, the program code being operative for performing one of the methods when the computer program product runs on a computer. The program
20 code may for example be stored on a machine readable carrier.

Other embodiments comprise the computer program for performing one of the methods described herein, stored on a machine readable carrier or a non-transitory storage medium.

- 25 In other words, an embodiment of the inventive method is, therefore, a computer program having a program code for performing one of the methods described herein, when the computer program runs on a computer.

- A further embodiment of the inventive methods is, therefore, a data carrier (or a digital
30 storage medium, or a computer-readable medium) comprising, recorded thereon, the computer program for performing one of the methods described herein.

- A further embodiment of the inventive method is, therefore, a data stream or a sequence of signals representing the computer program for performing one of the methods described
35 herein. The data stream or the sequence of signals may for example be configured to be transferred via a data communication connection, for example via the Internet.

A further embodiment comprises a processing means, for example a computer, or a programmable logic device, configured to or adapted to perform one of the methods described herein.

- 5 A further embodiment comprises a computer having installed thereon the computer program for performing one of the methods described herein.

- 10 In some embodiments, a programmable logic device (for example a field programmable gate array) may be used to perform some or all of the functionalities of the methods described herein. In some embodiments, a field programmable gate array may cooperate with a microprocessor in order to perform one of the methods described herein. Generally, the methods are preferably performed by any hardware apparatus.

- 15 The above described embodiments are merely illustrative for the principles of the present invention. It is understood that modifications and variations of the arrangements and the details described herein will be apparent to others skilled in the art. It is the intent, therefore, to be limited only by the scope of the impending patent claims and not by the specific details presented by way of description and explanation of the embodiments herein.

Literature:

- 5 [1] J. Breebaart, S. van de Par, A. Kohlrausch, E. Schuijers, “High-Quality Parametric Spatial Audio Coding at Low Bitrates” in Proceedings of the AES 116th Convention, Berlin, Preprint 6072, May 2004
- 10 [2] J. Herre, K. Kjörling, J. Breebaart et al., “MPEG surround – the ISO/MPEG standard for efficient and compatible multi-channel audio coding,” in Proceedings of the 122th AES Convention, Vienna, Austria, May 2007
- [3] Pulkki, Ville; “Spatial Sound Reproduction with Directional Audio Coding” in J.Audio Eng. Soc., Vol. 55, No. 6, 2007
- 15 [4] ISO/IEC International Standard “Information Technology – MPEG audio technologies – Part1: MPEG Surround”, ISO/IEC 23003-1:2007.
- [5] J. Engdegard, H. Purnhagen, J. Röden, L.Liljeryd, “Synthetic Ambience in Parametric Stereo Coding” in Proceedings of the AES 116th Convention, Berlin, Preprint, May 2004

Claims

1. An apparatus for decoding (10; 40; 60; 410) an encoded audio signal having an
5 audio signal frame comprising slots and events associated with the slots,
comprising:
- an analysing unit (20; 42; 70; 420) for analysing a frame slots number indicating
10 the total number of slots of the audio signal frame, an event slots number indicating
the number of slots comprising the events of the audio signal frame, and an event
state number; and
- a generating unit (30; 45; 80; 430) for generating an indication of a plurality of
15 positions of slots comprising the events in the audio signal frame using the frame
slots number, the event slots number and the event state number.
2. An apparatus for decoding (10; 40; 60; 410) according to claim 1,
wherein the apparatus for decoding (10; 40; 60; 410) is adapted to decode the slot
20 positions of transients in an audio signal frame.
3. An apparatus for decoding (10; 40; 60; 410) according to claim 1 or 2,
wherein the analysing unit (20; 42; 70; 420) is adapted to conduct a test comparing
the event state number or an updated event state number with a threshold value.
- 25 4. An apparatus for decoding (10; 40; 60; 410) according to claims 3,
wherein the analysing unit (20; 42; 70; 420) is adapted to conduct the test by
comparing, whether the event state number or an updated event state number is
greater than, greater than or equal to, smaller than, or smaller than or equal to the
30 threshold value, and
- wherein the generating unit (30; 45; 80; 430) is furthermore adapted to update the
event state number or an updated event state number depending on the result of the
test.
- 35 5. An apparatus for decoding (10; 40; 60) according to claim 3 or 4,
wherein the apparatus for decoding (10; 40; 60) furthermore comprises a slot
selector (90),

wherein the slot selector (90) is adapted to select a slot as a considered slot,

wherein the analysing unit (20; 42; 70) is adapted to conduct the test with respect to a considered slot,

5

and wherein the threshold value depends on the frame slots number, the event slots number and on the position of the considered slot within the frame.

6. An apparatus for decoding (10; 40) according to claim 5,

10

wherein the analysing unit (20; 42; 70) is adapted to conduct the test comparing the event state number or an updated event state number with the threshold value, wherein the threshold value is $\binom{N-h-1}{P}$,

15

wherein N is the total number of slots of the audio signal frame, wherein P is the number of slots comprising the events of the audio signal frame or of a considered portion of the audio signal frame and wherein h is the position of the considered slot within the frame.

7. An apparatus for decoding (10; 40; 410) according to one of claims 1 to 4,

20

wherein the apparatus for decoding (10; 40; 410) further comprises a frame partitioner (440),

25

wherein the frame partitioner (440) is adapted to split the frame into a first frame partition comprising a first set of slots of the frame and into a second frame partition comprising a second set of slots of the frame, and wherein the apparatus for decoding (10; 40; 410) is further adapted to determine the slot positions comprising the events for each of the frame partitions separately.

8. An apparatus for decoding (10; 40; 60; 410) according to one of the preceding claims, further comprising:

30

an audio signal processor (50) for generating an audio output signal using the indication of a plurality of positions of slots comprising the events in the audio signal frame using frame slots number, the event slots number and the event state number.

35

9. An apparatus for decoding (10; 60; 410) according to claim 8,

wherein the audio signal processor (50) is adapted to generate the audio output signal according to a first method, if the indication of a plurality of positions of slots comprising the events is in a first indication state, and wherein the audio signal processor (50) is adapted to generate the audio output signal according to a different second method, if the indication of a plurality of positions of slots comprising the events is in a second indication state which is different from the first indication state.

10. An apparatus for decoding (10; 40; 60; 410) according to claim 9, wherein the audio signal processor (50) is adapted, such that the first method comprises employing a transient decorrelator (56) for decoding a slot, if the first indication state indicates that the slot comprises a transient and wherein the second method comprises employing a second decorrelator (54) for decoding a slot, if the second indication state indicates that the slot does not comprise a transient.

11. An apparatus for encoding (510) positions of slots comprising events in an audio signal frame, comprising:

an event state number generator (530) for encoding the positions of slots by encoding an event state number; and

a slot information unit (520), being adapted to provide a frame slots number indicating the total number of slots of the audio signal frame and an event slots number indicating the number of slots comprising the events of the audio signal frame to the event state number generator (530),

wherein the event state number, the frame slots number and the event slots number together indicate a plurality of positions of slots comprising the events in the audio signal frame.

12. An apparatus for encoding (510) according to claim 11, wherein the event state number generator (530) is adapted to generate an event state number by adding a positive integer value for each slot comprising an event.

13. An apparatus for encoding (510) according to claim 11, wherein the event state number generator (530) is adapted to generate the event state number by determining a first event substate number for a first frame partition,

by determining a second event substate number for a second frame partition, and by combining the first and the second event state number to generate the event state number.

- 5 14. A method for decoding positions of slots comprising events in an audio signal frame comprising:

10 analysing a frame slots number indicating the total number of slots of the audio signal frame, an event slots number indicating the number of slots comprising the events of the audio signal frame, and an event state number; and

15 generating an indication of a plurality of positions of slots comprising the events in the audio signal frame using frame slots number, the event slots number and the event state number.

15. A method for encoding positions of slots comprising events in an audio signal frame comprising:

20 receiving or determining a frame slots number indicating the total number of slots of the audio signal frame,

receiving or determining an event slots number indicating the number of slots comprising the events of the audio signal frame,

25 encoding an event state number based on the event state number, the frame slots number and the event slots number, such that an indication of a plurality of positions of slots comprising the events in the audio signal frame can be decoded by using frame slots number, the event slots number and the event state number

- 30 16. A computer program for decoding positions of slots comprising events in an audio signal frame implementing a method for decoding slot positions of the events in an audio signal frame according to claim 14.

35 17. A computer program for encoding positions of slots comprising events in an audio signal frame implementing a method for encoding slot positions of the events in an audio signal frame according to claim 15.

18. An encoded audio signal comprising an event state number, wherein the positions of slots comprising events can be decoded according to the method of claim 14.