



(51) International Patent Classification:
G10L 19/008 (2013.01)

(21) International Application Number:
PCT/IB2012/055964

(22) International Filing Date:
29 October 2012 (29.10.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/554,007 1 November 2011 (01.11.2011) US

(71) Applicant: **KONINKLIJKE PHILIPS ELECTRONICS N.V.** [NL/NL]; High Tech Campus 5, NL-5656 AE Eindhoven (NL).

(72) Inventors: **KOPPENS, Jeroen Gerardus Henricus**; c/o High Tech Campus, Building 44, NL-5656 AE Eindhoven (NL). **OOMEN, Arnoldus Werner Johannes**; c/o High Tech Campus, Building 44, NL-5656 AE Eindhoven (NL). **VAN DE KERKHOF, Leon Maria**; c/o High Tech Campus, Building 44, NL-5656 AE Eindhoven (NL).

(74) Agents: **COOPS, Peter** et al.; High Tech Campus, Building 44, NL-5656 AE Eindhoven (NL).

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM,

AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))

(54) Title: AUDIO OBJECT ENCODING AND DECODING

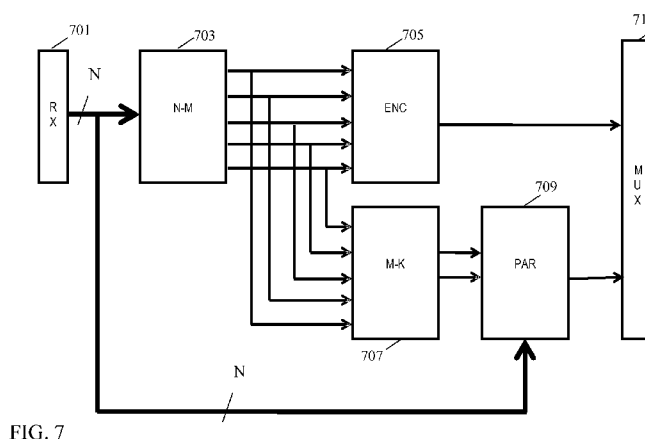


FIG. 7

(57) **Abstract:** An audio object encoder comprises a receiver (701) which receives N audio objects. A downmixer (703) downmixes the N audio objects to M audio channels, and a channel circuit (707) derives K audio channels from the M audio channels, $K = 1, 2$ and $K < M$. A parameter circuit (709) generates audio object upmix parameters for at least part of each of the N audio objects relative to the K audio channels and an output circuit (705, 711) generates an output data stream comprising the audio object upmix parameters and the M audio channels. An audio object decoder receives the data stream and includes a channel circuit (805) deriving K audio channels from the M channel downmix; and an object decoder (807) for generating at least part of each of the N audio objects by upmixing the K audio channels based on the audio object upmix parameters. The invention may allow improved object encoding while maintaining backwards compatibility.

AUDIO OBJECT ENCODING AND DECODING

FIELD OF THE INVENTION

The invention relates to audio object encoding and decoding and in particular, but not exclusively, to audio object encoding and/or decoding compatible with the MPEG SAOC (Spatial Audio Object Coding) standard.

5

BACKGROUND OF THE INVENTION

Multichannel audio is widespread and has become popular for many different applications including home cinema and multi-channel music systems. Audio encoding is often used to generate data streams that provide an efficient data representation of the audio signals. Such audio encoding allows an efficient storage and distribution of audio signals. Many different audio encoding standards have been developed for encoding and decoding of both traditional mono and stereo audio signals, as well as for encoding and decoding of multichannel audio signals. The term multichannel is henceforth used to refer to more than two channels. The use of dedicated audio standards allows for interworking and compatibility between many different systems, devices and applications, and it is therefore critical that efficient standards are adhered to. However, a significant problem arises when new standards are developed or existing standards are modified. In particular, modifications to standards may not only be time consuming and cumbersome to carry out but may also result in existing equipment not being suitable for the new or indeed for the existing standards. In order to facilitate introduction of new standards or standard modifications, it is desirable that these require as little modification to existing standards as possible. In some cases it is even possible to make modifications that are fully compatible with the existing standards, i.e. the modifications can be applied without any change to the existing standard specification. An example of this is bitstream watermarking. In bitstream watermarking specific bitstream elements are modified in a compatible fashion such that the bitstream can still be decoded according to the standard specification. Although the output has changed, the difference in quality is generally not audible.

MPEG Surround is one of the major advances in multi-channel audio coding and was recently standardized by Motion Picture Experts Group in ISO/IEC 23003-1. MPEG

Surround is a multi-channel audio coding tool that allows existing mono- or stereo-based services to be extended to multi-channel applications. Fig. 1 shows a block diagram of a stereo core coder extended with MPEG Surround. First the MPEG Surround encoder creates a stereo downmix from the multi-channel input signal. Next, spatial parameters are estimated from the multi-channel input signal. These parameters are encoded into the MPEG Surround bit-stream. The stereo downmix is coded into a bit-stream using a core encoder, e.g. HE-AAC. The resulting core coder bit-stream and the spatial bit-stream are merged to create the overall bit-stream. Typically the spatial bit-stream is contained in the ancillary data or user data portion of the core coder bit-stream. At the decoder side the core and spatial bit-stream are separated. The stereo core bit-stream is decoded in order to reproduce the stereo downmix. This downmix together with the spatial bit-stream is input to the MPEG Surround decoder. The spatial bit-stream is decoded to provide the spatial parameters. The spatial parameters are then used to upmix the stereo downmix in order to obtain the multi-channel output signal.

Since the spatial image of the multi-channel input signal is parameterized, MPEG Surround allows for decoding of the same multi-channel bit-stream onto rendering devices other than a multichannel speaker setup. An example is virtual surround reproduction on headphones, which is referred to as the MPEG Surround binaural decoding process. In this mode, a realistic surround experience can be provided using regular headphones. Fig. 2 shows a block diagram of the stereo core codec extended with MPEG Surround where the output is decoded to binaural. The encoder process is identical to that of Fig. 1. In the system, the spatial parameters are combined with the Head Related Transfer Function (HRTF) and the result is used to produce the so-called binaural output.

Building upon the concept of MPEG Surround, MPEG has standardized a system for encoding of individual audio objects. This standard is known as 'Spatial Audio Object Coding' (MPEG-D SAOC) ISO/IEC 23003-2. From a high level perspective, SAOC efficiently encodes sound objects instead of audio channels where each sound object may typically correspond to a single sound source in the sound image. In MPEG Surround, each speaker channel can be considered to originate from a different mix of sound objects whereas in SAOC data is provided for the individual sound objects. Similarly to MPEG Surround, a mono or stereo downmix is also created in SAOC. Specifically, SAOC also generates a mono or stereo downmix which is coded using a standard downmix coder such as HE-AAC. In this way, legacy playback devices will disregard the parametric data and play the mono or stereo downmix whereas SAOC decoders can upmix the signal to retrieve the original sound objects

or to allow them to be rendered in a desired output configuration. Object and downmix parameters are embedded in the ancillary data portion of the downmix coded bitstream to provide relative level and gain information for the individual SAOC objects, typically reflecting the downmix of these into the stereo/mono downmix. At the decoder side, the user can control various features of the individual objects (such as spatial position, amplification, and equalization) by manipulating these parameters, or the user can apply effects, such as reverb, to individual objects.

Fig. 3 shows a block-diagram for regular SAOC encoding. The SAOC encoder can be considered to be a preprocessing module situated before a conventional mono- or stereo encoder. The preprocessing consists of generating a stereo (or mono) downmix from a number N of object signals. Additionally object parameters are extracted and stored in an SAOC bitstream together with information on the downmix matrix M . The SAOC downmix information is encoded in two types of parameters. First the DMG (downmix gain) parameter indicates the gain applied to the object. The DCLD (downmix channel level difference) parameter signals the distribution of the object over the two channels in a stereo downmix. These parameters are both defined per object.

A SAOC decoder may perform the opposite operation. The received mono- or stereo downmix may be decoded and upmixed to a desired output configuration. The upmix operation includes the combined operation of an upmixing of the mono- or stereo downmix to generate the audio objects followed by a mapping of these to the desired output configuration based on a rendering matrix as illustrated in Fig. 4, where the mono or stereo input downmix is first upmixed to N audio objects based on the SAOC parameters. The resulting N audio objects are then downmixed to P output channels using a rendering matrix defining where the individual objects are positioned. Fig. 4 illustrates the conceptual SAOC decoding. However, typically the upmix matrix and the rendering matrix are combined into a single matrix and the generation of the output channels from the mono- or stereo downmix is performed as a single operation. An example thereof is shown in Fig. 5 which shows a specific example wherein P equals one or two, and where specifically for $P=2$ the output may be a binaural spatial output channel. Thus, the two output channels are generated using HRTF parameters applied to the individual objects to generate the desired binaural spatial image. Fig. 9 illustrates an example where $P>2$ and an MPEG Surround (MPS) decoding/processing is used to generate the P output channels.

However, an issue associated with SAOC is that the specification only supports mono- and stereo downmixes whereas there are a number of applications and use-

cases in which multi-channel mixes are used or even sometimes required, for instance DVD and Blu-Ray. It would therefore be desirable for SAOC to support such multi-channel applications, i.e. a multichannel downmix, but this would require substantial amendments to the SAOC standard specification which would be cumbersome, impractical, increase complexity and result in reduced backwards compatibility.

In particular, it would be advantageous if existing algorithms, functional units, dedicated hardware etc. developed for SAOC encoding and decoding could be reused while allowing improved support for multichannel audio.

Hence, an improved approach for object encoding and/or decoding (such as e.g. SAOC encoding/decoding) would be advantageous and in particular approaches allowing increased flexibility, reduced impact on standardised approaches, increased or facilitated backwards compatibility, allowing increased reuse of encoding and/or decoding functionality, facilitated implementation, multichannel support in object encoding, and/or improved performance would be advantageous.

SUMMARY OF THE INVENTION

Accordingly, the Invention seeks to preferably mitigate, alleviate or eliminate one or more of the above mentioned disadvantages singly or in any combination.

According to an aspect of the invention there is provided an audio object encoder comprising: a receiver for receiving N audio objects; a mixer for mixing the N audio objects to M audio channels; a channel circuit for deriving K audio channels from the M audio channels where $K = 1$ or 2 and $K < M$; a parameter circuit generating audio object upmix parameters for at least part of each of the N audio objects relative to the K audio channels; an output circuit for generating an output data stream comprising the audio object upmix parameters and the M audio channels.

The invention may allow audio encoding that can provide improved performance for multichannel rendering systems while supporting audio object encoding. The system may in some scenarios allow improved multichannel rendering and may in some scenarios allow improved audio object functionality. A low data rate can be achieved by combining M audio channels with audio object upmix parameters relating to K audio channels such that it is not necessary to include encoded data for the K audio channels in the output data stream.

The invention may allow multichannel support (with more than two channels) in audio object encoding systems providing audio object encoding (and/or decoding) based

on only mono and stereo signals. The encoding may generate an output data stream wherein a multichannel signal is provided together with associated audio object data, which however is not defined relative to the multichannel signal but rather relative to a mono or stereo signal that can be derived from the multichannel signal.

5 The invention may in many applications allow improved reuse and/or backwards compatibility with existing audio object encoding and/or decoding functionality.

 An audio object may be an audio signal component corresponding to a single sound source in the audio environment. Specifically, the audio object may include audio from only one position in the audio environment. An audio object may have an associated position
10 but not be associated with any specific rendering sound source configuration, and may specifically not be associated with any specific loudspeaker configuration.

 The output data stream may not include any encoding data of the K audio channels. In some embodiments, all of one, more or all of the N audio objects is generated from the K audio channels.

15 The derivation of the K channels may be performed in each segment, and the specific derivation may change dynamically, e.g. between segments. In many embodiments and/or scenarios M may be smaller than N.

 In accordance with an optional feature of the invention, the channel circuit is arranged to derive the K channels by downmixing the M audio channels.

20 This may provide a particularly advantageous system in many scenarios and applications. Particularly, it may allow reuse of functionality and may allow efficient audio object encoding and decoding. Specifically, the approach may allow the generated downmix to provide suitable components in the K audio channels for all audio objects also represented in the M audio channels.

25 In some embodiments, the downmixing may be such that each of the M audio channels is represented in at least one of the K channels, and in some embodiments in all of the K channels.

 In accordance with an optional feature of the invention, the channel circuit is arranged to derive the K channels by selecting a K channel subset of the M audio channels.

30 This may provide a particularly advantageous system in many scenarios and applications. Particularly, it may allow reuse of functionality and may allow efficient audio object encoding and decoding. In many embodiments it may reduce complexity and/or increase flexibility. The selection of K channels may be dynamically varied allowing different K channels to be selected in different time segments.

In accordance with an optional feature of the invention, the output data stream comprises a multichannel encoded data stream for the M audio channels, and the audio object upmix parameters are comprised in a part of the multichannel encoded data stream.

This may provide a particularly advantageous output data stream in many embodiments. In particular, it may allow a combined data stream which supports both multichannel audio directly and audio object encoding based on mono and/or stereo signals thereby allowing backwards compatibility. Thus a multichannel encoded data stream may be provided which contains the multichannel signal and audio object upmix parameters which are not provided relative to the encoded multichannel signal yet which still allows the object decoding based on the encoded multichannel signal.

In accordance with an optional feature of the invention, the output circuit is arranged to include mixing data representative of the mixing of the N audio objects to the M audio channels in the output data stream.

This may allow improved performance in many embodiments, and may in particular in many embodiments allow improved audio object decoding and functionality to be provided at the decoder. The mix data may e.g. be defined in the time frequency domain.

In accordance with an aspect of the invention, there is an audio object decoder comprising: a receiver for receiving a data stream comprising audio data for an M channel mix of N audio objects and audio object mix parameters for the N audio objects relative to K audio channels where $K = 1$ or 2 and $K < M$; a channel circuit deriving K audio channels from the M channel mix; and an object decoder for generating P audio signals from N audio objects at least partially generated by upmixing the K audio channels based on the audio object upmix parameters.

The invention may allow for audio object decoding and may in particular allow efficient audio object decoding based on a signal that directly supports multichannel rendering systems. The audio object decoder may generate the P audio signals without any audio encoding data being received for the K audio channels.

The invention may in many applications allow improved reuse and/or backwards compatibility with existing audio object encoding and/or decoding functionality.

The object decoder may be arranged to generate the P audio signals by upmixing the K channels to N audio objects and then mapping the N audio objects to the P audio channels. The mapping may be represented by a rendering matrix. The upmixing of the K channels to the N audio objects and the mapping of the N audio objects to the P output channels may be performed as a single integrated operation. Specifically, a KtoN upmix

matrix may be combined with an NtoP matrix to generate a KtoP matrix which is directly applied to the K channels to generate the P output signals. Thus, the object decoder may be arranged to generate P output channels based on the audio object upmix parameters for the N audio objects and a rendering matrix for the P output channels. In some embodiments, the N audio objects may be explicitly generated, and especially each of the P audio signals may correspond to a single audio object of the N audio objects. In some scenarios N may be equal to P.

In accordance with an optional feature of the invention, the channel circuit is arranged to derive the K channels by downmixing the M audio channels.

This may provide a particularly advantageous system in many scenarios and applications. Particularly, it may allow efficient audio object encoding and decoding. Specifically, the approach may allow the generated downmix to provide suitable components in the K audio channels for all audio objects also represented in the M audio channels. In some embodiments, the object decoder may be arranged to generate each of N audio objects by upmixing the K audio channels based on the audio object upmix parameters.

In some embodiments, the downmixing may be such that each of the M audio channels is represented in at least one of the K channels, and in some embodiments in all of the K channels.

In accordance with an optional feature of the invention, the data stream further comprises downmix data indicative of an encoder downmixing from M to K channels, and wherein the channel circuit is arranged to adapt the downmixing in response to the downmix data.

This may allow increased flexibility and/or improved performance in many embodiments. For example, it may allow adaptation of the downmix to the specific signal characteristics and may e.g. allow the downmix to be adapted to the N audio objects to provide suitable signal components of all N audio objects to allow the generation in the decoder of the objects.

In some embodiments, a fixed or predetermined downmix from M channels to K channels may be used in the encoder and the decoder. This may reduce complexity and may specifically obviate the need to include data indicative of the downmix in the data stream, thereby potentially allowing a reduced data rate.

In accordance with an optional feature of the invention, the channel circuit is arranged to derive the K channels by selecting a K channel subset of the M audio channels.

This may allow improved and/or facilitated audio object encoding in many embodiments. It may in many embodiments allow reduced complexity.

In accordance with an optional feature of the invention, the data stream further comprises additional audio object upmix parameters for the N audio objects relative to L audio channels where $L = 1$ or 2 and $L < M$, and the L audio channels and the K audio channels are different subsets of the M audio channels, and wherein the object decoder is further arranged to generate the P channels from N audio objects at least partially generated by upmixing the L audio channels based on the additional audio object upmix parameters.

This may allow improved audio object decoding in many embodiments. In particular it may allow the signal components of each audio object in more than K (and in particular all M) audio channels to be used in generating the audio object.

The subsets may be disjoint. In some embodiments, further upmixing may be based on one or more additional subsets of audio channels with associated audio object upmix parameters. In some embodiments, the combination of subsets may include all M audio channels.

In accordance with an optional feature of the invention, at least one of the P channels is generated by combining contributions from both the upmixing of the K audio channels based on the audio object upmix parameters and the upmixing of the L audio channels based on the additional audio object upmix parameters.

This may allow improved audio object decoding in many embodiments. In particular it may allow the signal components of each audio object in more than K (and in particular all M) audio channels to be used in generating the audio object.

In accordance with an optional feature of the invention, the data stream comprises mix data representative of the mixing of the N audio objects to the M audio channels, and wherein the object decoder is arranged to generate residual data for at least a subset of the N audio objects in response to the mix data and the audio object upmix parameters, and to generate the P audio signals in response to the residual data.

This may provide improved quality of one, some or all of the decoded audio objects in many embodiments. In many embodiments it may allow compatibility with standardized audio object decoding algorithms capable of receiving residual data, such as for example the SAOC standard. The residual data may specifically be indicative of a difference between an audio object generated from the K channels and the audio object upmix parameters, and the corresponding audio object generated on the basis of the M audio channels and the downmix data.

In accordance with an aspect of the invention, there is provided a method of audio object encoding comprising: receiving N audio objects; mixing the N audio objects to M audio channels; deriving K audio channels from the M audio channels where $K = 1$ or 2 and $K < M$; generating audio object upmix parameters for at least part of each of the N audio objects relative to the K audio channels; and generating an output data stream comprising the audio object upmix parameters and the M audio channels.

In accordance with an optional feature of the invention, there is provided a method of audio object decoding comprising: receiving a data stream comprising audio data for an M channel mix of N audio objects and audio object upmix parameters for the N audio objects relative to K audio channels where $K = 1$ or 2 and $K < M$; deriving K audio channels from the M channel mix; and generating P audio signals from N audio objects at least partially generated by upmixing the K audio channels based on the audio object upmix parameters.

These and other aspects, features and advantages of the invention will be apparent from and elucidated with reference to the embodiment(s) described hereinafter.

Brief Description of the Drawings

Embodiments of the invention will be described, by way of example only, with reference to the drawings, in which

Fig. 1 is an illustration of an MPEG Surround system in accordance with prior art;

Fig. 2 is an illustration of an MPEG Binaural Surround system in accordance with prior art;

Fig. 3 is an illustration of an MPEG SAOC encoder in accordance with prior art;

Fig. 4-6 illustrate examples of MPEG SAOC decoders in accordance with prior art;

Fig. 7 illustrates an example of elements of an audio object encoder in accordance with some embodiments of the invention;

Fig. 8 illustrates an example of elements of an audio object decoder in accordance with some embodiments of the invention;

Fig. 9 illustrates an example of elements of an audio object encoder in accordance with some embodiments of the invention;

Fig. 10 illustrates an example of an encoder output data stream in accordance with some embodiments of the invention;

Fig. 11 illustrates an example of elements of an audio object decoder in accordance with some embodiments of the invention; and

5 Fig. 12 illustrates an example of elements of an audio object decoder in accordance with some embodiments of the invention.

DETAILED DESCRIPTION OF SOME EMBODIMENTS OF THE INVENTION

The following description focuses on an object encoder and decoder system
10 wherein N audio objects are downmixed to M audio channels, i.e. wherein $M < N$. However, it will be appreciated that other mixes may be used and that M may in some embodiments and scenarios be equal to or larger than N.

Fig. 7 illustrates elements of an audio object encoder in accordance with some embodiments of the invention.

15 The encoder comprises a receiver 701 which receives N audio objects. Each audio object typically corresponds to a single sound source. Thus, in contrast to audio channels, and in particular audio channels of a conventional spatial multichannel signal, the audio objects do not comprise components from a plurality of sound sources that may have substantially different positions. Similarly, each audio object provides a full representation of
20 the sound source and. Each audio object is thus associated with spatial position data for only a single sound source. Specifically, each audio object may be considered a single and complete representation of a sound source and may be associated with a single spatial position.

Furthermore, the audio objects are not associated with any specific rendering
25 configuration and are specifically not associated with any specific spatial configuration of sound transducers. Thus, in contrast to traditional spatial sound channels which are typically associated with a specific spatial speaker setup, such as in particular a surround sound setup, audio objects are not defined with respect to any specific spatial rendering configuration.

The N audio objects are fed to an N to M downmixer 703 which downmixes
30 the N audio objects to M audio channels. In the example, $M < N$ but it will be appreciated that in some scenarios N may be equal to or even smaller than M. In the specific example of Fig. 7, M equals 5 but it will be appreciated that in other embodiments other numbers of channels may be used, including for example $M=7$ or $M=9$.

Thus, the N to M downmixer 703 generates an M channel multichannel signal in which the audio objects are spread over the channels. In contrast to the N audio objects, the M audio channels are traditional audio channels which typically comprise data from a plurality of audio objects and thus a plurality of sound sources with different positions.

5 Furthermore, the individual audio objects are generally spread over the M audio channels and often each of the M audio channels comprises a component from a given audio object, although in some scenarios some audio objects may only be represented in a subset of the M audio channels.

10 The N to M downmixer 703 generates a multichannel signal (henceforth used to denote the signal provided by the M audio channels) which may directly be rendered as a multichannel signal. Specifically, the multichannel signal formed by the M audio channels may be a spatial surround signal, and in the specific example the M audio channels may be respectively the front left, front right, centre, surround left and surround right channels of a five channel system (and accordingly $M=5$). Thus, the multichannel signal formed by the M
15 audio channels is associated with a specific rendering configuration and specifically each audio channel is an audio channel associated with a rendering position.

The N to M downmixer 703 can perform the downmix such that the individual audio objects are positioned as desired in the surround image provided by the M audio channels. For example, one audio object can be positioned directly to the front, another
20 object can be positioned to the left of the nominal listening position etc. The N to M downmix may specifically be manually controlled such that the resulting surround sound signal of the M audio channels provide the desired spatial distribution when the multichannel signal is rendered directly. The N to M downmix can specifically be based on an N to M downmix matrix that is manually generated by a person to provide the desired surround
25 signal from the M audio channels.

The M audio channels are fed to an M channel encoder 705 which proceeds to encode the M audio channels in accordance with any suitable encoding algorithm. The M channel encoder 705 typically employs a conventional multichannel encoding scheme to provide an efficient representation of the corresponding surround signal.

30 It will be appreciated that the encoding of the M audio channels is typically preferred but is not necessary in all embodiments. For example, the N to M downmixer 703 may directly generate a frequency domain or time domain representation of the signals which can be used directly. For example, it is possible to send the M audio channels to an object

decoder using un-encoded PCM data. However, an efficient encoding may substantially reduce the data rate and is therefore typically used.

The encoded multichannel signal may specifically correspond to a conventional multichannel signal and a conventional audio device receiving the multichannel signal can accordingly render the multichannel signal directly.

The encoder of Fig. 7 furthermore comprises functionality for providing audio object upmix parameters that allows the original N audio objects to be regenerated at a suitably equipped object decoding device. However, the audio object upmix parameters are not provided relative to the M audio channels but are instead provided relative to K audio channels where K is one or two. Thus, the encoder generates audio object upmix parameters relative to a mono or stereo signal. This allows compatibility with standards allowing only object encoding and decoding based on mono or stereo downmix signals from the original audio objects. This may in many scenarios allow standard audio object encoder or decoder functionality for mono or stereo signals to be reused with multichannel support. For example, the approach may be used to allow improved compatibility with SAOC.

The encoder comprises an M to K channel reducer 707 which receives the M audio channels from the N to M downmixer 703 and which then proceeds to derive K audio channels from the M audio channels with K being 1 or 2.

The M to K channel reducer 707 is coupled to a parameter circuit 709 which also receives the original N audio objects from the receiver. The M to K channel reducer 707 is arranged to generate audio object upmix parameters for at least part of each of the N audio objects relative to the K audio channels. Thus, audio object upmix parameters are generated which describe how (part or all of) the N audio objects can be generated from the mono or stereo signal received from the M to K channel reducer 707.

The M channel encoder 705 and the parameter circuit 709 are coupled to an output circuit 711 which generates an output data stream comprising the audio object upmix parameters received from the parameter circuit 709 and the encoded M audio channels received from the M channel encoder 705. However, the output data stream does not include any data of the K audio channels (whether encoded or not). Thus, an output data stream is generated which comprises an encoded multichannel signal that can be rendered directly by legacy multichannel devices even if no capable audio object decoding or processing. In addition, audio object upmix parameters are provided which can allow the original N audio objects to be regenerated at the decoder side. However, the audio object upmix parameters are not provided relative to the signal included in the data stream but instead relative to a

stereo or mono signal which is not included in the output data stream. This allows the operation to be compatible with audio object encoding and decoding approaches that are limited to mono and stereo signals. For example, existing SAOC encoding or decoding units may be reused while allowing multichannel support.

5 Furthermore, although the K audio channels are not included in the output data stream, they can be derived from the multichannel signal by the decoder. Accordingly, a suitably equipped decoder may derive the K audio channels and then generate the N audio objects based on the audio object upmix parameters. This can specifically be done using existing upmix functionality based on an underlying stereo or mono signal. Thus the
10 approach may allow a single output data stream to provide a multichannel signal which can be rendered directly by multichannel devices and audio object data related to a mono or stereo signal not included in the output data stream yet still allowing the original audio objects to be generated.

The output data stream may specifically comprise a multichannel encoded data
15 stream for the M audio channels where the multichannel encoded data stream also includes the audio object upmix parameters. Thus, a multichannel encoded data stream may be provided which comprises the multichannel signal itself plus data for generating the individual audio objects comprised in the multichannel signal but where this data is not related to the multichannel signal itself but rather to a mono or stereo signal which is not
20 included in the multichannel encoded data stream. The audio object upmix parameters may specifically be included in an ancillary, auxiliary or optional data field of the multichannel encoded data stream.

Fig. 8 illustrates an example of a decoder in accordance with some embodiments of the invention.

25 The decoder comprises a receiver 801 for receiving the output data stream from the encoder of Fig. 7. Thus, the receiver receives a data stream comprising audio data for an M channel downmix of N audio objects together with audio object upmix parameters for the N audio objects relative to K audio channels where $K = 1$ or 2 and $K < M$. In the example the audio data for the M channel downmix is encoded audio data.

30 The encoded audio data for the M channel downmix is fed to a multichannel decoder 803 which generates the M audio channels from the encoded audio data. The M audio channels are fed to an M to K channel processor 805 which derives the K audio channels from the M audio channels. The M to K channel processor 805 specifically performs the same operation as the M to K channel reducer 707 of the encoder of Fig. 7. The

resulting K audio channels are fed to an object decoder 807 which generates the N audio objects by upmixing the K audio channels based on the audio object upmix parameters. The object decoder 807 specifically performs the inverse operation of the parameter circuit 709 of Fig. 7.

It will be appreciated that in the example of Fig. 8, the object decoder 807 regenerates the N audio objects which can then be individually processed and/or mapped to a specific speaker configuration. Thus, in the example, P output signals are generated where $P=N$ and each output signal corresponds to one of the N audio objects.

In some embodiments, the mapping to a given speaker configuration may be combined with the upmixing of the object decoder 807, e.g. by applying a single matrix multiplication where the matrix coefficients reflect the combined matrix multiplication of the mapping of the K audio channels to the N audio objects and the matrix multiplication of the mapping of the N audio objects to the channels of the speaker configuration.

Specifically, P audio signals may be generated where each of the P audio signals may correspond to a spatial output channel of a given P-channel rendering configuration. This may be achieved by the object decoder 807 applying a rendering matrix which maps the N audio objects to the P audio signals. Typically, the object upmix matrix generating the N audio objects from the K audio channels is combined with the rendering matrix mapping the N audio objects to the P audio signals. Thus, a single combined object upmix and rendering matrix is applied to the K audio channels to generate the P audio signals. The combined object upmix and rendering matrix can specifically be generated by multiplying the object upmix matrix and the rendering matrix.

In some embodiments, the M to K channel processor 805 and the M to K channel reducer 707 may be arranged to generate the K channels by downmixing the M audio channels. In particular, the downmix may be generated such that all the audio objects have significant signal components in the downmix thereby allowing the upmixing based on the K channels to be efficient for all N audio objects.

An example of this approach is illustrated in Fig. 9. In the specific example, the object encoding is compatible with the SAOC standard, and thus an SAOC encoder is specifically used. In the specific example $M=5$ and $K=2$.

Furthermore, it is noted that in the example of Fig. 9 the generation of the K audio channels is performed by combining the operation that generates the M audio channels from the N audio objects and the operation that generates the K audio channels from the M audio channels into a single operation.

Specifically, the M audio channels may be generated by applying an encoder rendering matrix M_{Nto5} to the N audio objects to provide the M audio channels (a matrix multiplication may be performed for each frequency time tile as will be known to the person skilled in the art). Similarly, the K audio channels may be generated by applying a rendering matrix M_{5to2} to the M audio channels to provide the K audio channels (a matrix multiplication may be performed for each frequency time tile as will be known to the person skilled in the art). The sequential operation of these two matrix operations may be replaced by a single matrix operation performing the combined operation. Specifically, a single matrix multiplication by a matrix

$$M_{Nto2} = M_{5to2} \cdot M_{Nto5},$$

may be applied directly to the N audio objects as this is identical to applying the matrix M_{5to2} to the M (in the specific example 5) audio channels generated by the N to M downmixer 703 by the application of the matrix M_{Nto5} . Thus, in the decoder, the K channels are simply generated by multiplying the M (i.e. in the specific example 5) audio channels and the downmix matrix M_{5to2} .

It will be appreciated that any suitable approach or method for selecting or determining the rendering matrix M_{Nto5} may be used. Typically, a matrix is (semi)manually generated to provide the desired sound image.

Similarly, it will be appreciated that any suitable approach or method for selecting or determining the downmix matrix M_{5to2} may be used. In some embodiments a fixed or predetermined downmix matrix M_{5to2} may be used. This predetermined matrix may be known at the decoder which can accordingly apply it to the M audio channels to generate the stereo signal required for the audio object generation.

In other embodiments, the downmix matrix M_{5to2} may be a variable matrix which is adapted or optimized in the encoder dependent on the specific characteristics. For example, the downmix matrix M_{5to2} may be determined such that it is ensured that all audio objects are represented in a desirable way in the resulting stereo signal. In such embodiments, information on the downmix matrix M_{5to2} used at the encoder may be included in the output data stream. The decoder may then extract the downmix matrix M_{5to2} and apply this to the decoded M audio channels thereby generating the K audio channels to which the SAOC parameters can be applied.

When allowing an adaptive multichannel to stereo downmix, the data can be transmitted by employing the ancillary data structure in the syntax of the multichannel bitstream, e.g. similarly to the transmission of the SAOC data. This is illustrated in Fig. 10 which shows two different two options:

- 5 - the downmix parameters being transmitted in a separate container prior (or after) the SAOC container; and
- the downmix parameters being transmitted inside the SAOC container as a new entry in the SAOCExtensionConfig() field.

10 In some embodiments, the derivation of the K channels from the M audio channels is performed by selecting a subset of M audio channels.

For example, the SAOC encoding may be performed in response to only two audio channels, such as the front left and front right channels of a five channel surround signal formed by the M audio channels.

15 However, in many scenarios such an approach may lead to suboptimally decoded objects due to the selected subset channels potentially not including any signal components from a given audio object (in contrast to downmixed channels wherein the M audio channels can be downmixed to the K audio channels such that contributions from all M audio channels, and thus from all N audio objects, are included in the downmixed K channels).

20 Such problems may possibly be addressed by the decoder generating part or all of some of the N audio objects using other parallel approaches. For example, using the SAOC send effects interface functionality defining send effects to introduce a contribution generated as a send effect. The send effect may be defined such that it can provide a contribution to audio objects which cannot be generated with sufficient quality from the
25 selected K audio channels.

In some embodiments, contributions from the audio objects may be generated from a plurality of subsets of the M audio channels, where each subset is provided with suitable audio object upmix parameters. In some embodiments, each audio object may be generated from a single subset of the M audio channels with different audio objects being
30 generated from different subsets depending on how the objects have been downmixed to the M audio channels. However, typically the N objects will be distributed over more than K channels of the M audio channels and therefore the audio objects may be generated by combining contributions from upmixing of the different subsets of the M audio channels.

The encoder may thus have parallel parameter estimators which are fed different subsets of the N audio objects. Alternatively, all N objects are fed to each of the parallel parameter estimators. Rendering matrix M_{Nto5} is split such, and used as a downmix matrix in each parameter estimator, that the signal outputs of the parameter estimators

- 5 constitute the M channel mix. For example, one parameter estimator may produce K audio channels of the M audio channels and another parameter estimator may produce L audio channels of the M audio channels. E.g. one parameter estimator generates the front left and right channels and another estimator is generates the center channel. The parameter estimators additionally generate audio object upmix parameters for the respective channels.
- 10 The audio object upmix parameters for each individual parameter estimator is included in the output data stream as a separate set of audio object upmix parameters, e.g. specifically as a separate SAOC parameter data stream.

Thus, the encoder may generate a plurality of parallel SAOC compatible data streams each of which is associated with a stereo or mono subset of the M audio channels.

- 15 The corresponding decoder may then decode each of these SAOC compatible data streams individually using a standard SAOC decoder setup. The resulting decoded audio object components are then combined into the complete audio objects (or directly into output channels corresponding to the desired output speaker configuration). The approach may thus allow that all the signal components in the M audio channels can be exploited when
- 20 generating the individual audio object. Specifically, the subsets may be selected such that they together contain all of the M audio channels with each audio channel only being included in a single subset. Thus, the subsets may be disjoint and include all the M audio channels.

As a specific example, multiple SAOC streams can be included/ transmitted

25 with the M audio channel downmix, such that each stream operates on a mono or stereo subset of the multichannel downmix. With the objects possibly present in either specific, or in multiple streams, the rendering matrix used at the decoder side to distribute the audio objects to the desired output (speaker) configuration can be adapted to combine the individual contributions to the individual audio objects. The approach can provide a

30 particularly high reconstruction quality.

In comparison to the embodiment of Fig. 9, the N-to-5 matrix is in such a specific example not combined with a 5-to-2 downmix matrix to provide a K channel downmix of the five audio channels. Rather, the N-to-5 matrix is dissected and sent to three parallel SAOC encoders of which the bitstreams are all multiplexed into the bitstream.

For example

$$M_{dmx} = \begin{pmatrix} m_{11} & m_{12} & \cdots & m_{1N} \\ m_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ m_{51} & \cdots & \cdots & m_{5N} \end{pmatrix},$$

5 can be divided into

$$M_{dmx,1} = \begin{pmatrix} m_{11} & m_{12} & \cdots & m_{1N} \\ m_{21} & m_{22} & \cdots & m_{2N} \end{pmatrix},$$

$$M_{dmx,2} = (m_{31} \quad m_{32} \quad \cdots \quad m_{3N}),$$

10

$$M_{dmx,3} = \begin{pmatrix} m_{41} & m_{42} & \cdots & m_{4N} \\ m_{51} & m_{52} & \cdots & m_{5N} \end{pmatrix},$$

to provide three parallel SAOC streams that would typically work well for a typical five channel ordering of $\{L_f, R_f, C, L_s, R_s\}$ where L denotes left, R denotes right, C denotes centre, subscript f denotes front, and subscript s denotes surround.

15

Fig. 11 shows an example of a decoder for such an approach.

In some embodiments, the encoder may further be arranged to include downmix data representative of the downmixing of the N audio objects to the M audio channels into the output data stream. For example, the encoder rendering matrix describing the downmix of the N audio objects to the M audio channels may be included in the output data stream (i.e. in the specific example of Fig. 9, the matrix M_{Nto5} may be included).

20

The additional information may be used in different ways in different embodiments.

Specifically, in some embodiments the downmix data may be used to generate a subset of the audio objects based on the M audio channels. As there is more information available in the M audio channels than in the K audio channels, this may allow improved quality audio objects to be generated. However, the processing may not be compatible with a corresponding audio object encoding/decoding standard and may thus require additional

25

functionality. Furthermore, the computational requirements will typically be higher than for a standard (and typically heavily optimized) object decoding based on K signals. Therefore, the audio decoding based on the M audio channels and the downmix data may be limited to only a subset of the audio objects, and typically only to a very small number of the most dominant audio objects. The remaining audio objects may be generated using a standardised decoder based on the K channels. This decoding may often be substantially more efficient, e.g. by using dedicated and standardised hardware.

Furthermore, some encoding standards, such as SAOC, is capable of receiving residual data from the encoder where the encoded data reflects the difference between the original audio object and that which will be generated by a decoder based on the downmix and the audio object upmix parameters. Specifically, SAOC supports a feature known as Enhanced Audio Objects (EAO) which allows residual data to be provided for up to four audio objects.

In some embodiments the downmix data representative of the downmixing of the N audio objects to the M audio channels can be used to generate residual data at the decoder. Specifically, the decoder can calculate a specific audio object based on the downmix data, the M audio channels and the audio object upmix parameters. In addition, the same object can be decoded based on the K audio channels and the audio object upmix parameters. Residual data can be generated as an indication of the a difference between these. This residual data can then be used in the decoding of the N audio objects. This decoding may use a standardised approach for an object decoding standard which is based on K channels and which allows for residual data to be provided from the encoder.

In such an approach the additional information provided by the downmix data and the M audio channels is thus used to generate residual data information at the decoder rather than at the encoder. Thus, no residual data needs to be communicated. It will be appreciated that the object generated from the downmix data and the M audio channels may not be identical to the corresponding audio object before encoding but the additional information will typically still provide an improvement over the corresponding audio object generated from the K audio channels.

As a specific example, a standard SAOC decoder may be provided with a pre-processor which generates residual data that is fed to the SAOC decoder as if it were residual data generated at the encoder. Thus, the SAOC decoder may operate fully in accordance with the SAOC standard regarding EAO. In example of such a decoder is illustrated in Fig. 12.

The pre-processor may specifically calculate an audio object using the M_{Nto5} matrix. For example, an audio object may be generated from the 5 channel downmix using the following equation:

$$5 \quad \sqrt{\frac{M_{Nto5,1k}^2 \cdot OLD_k}{\sum_{i=1}^N M_{Nto5,1i}^2 \cdot OLD_i}} \cdot X_1$$

which reconstructs object k from downmix channel X_1 , where OLD is the linear representation of the OLD (Object Level Difference) parameter in the SAOC bitstream. This equation may be applied to each time-frequency tile of X_1 , using the corresponding SAOC parameters.

The above reconstruction assumes uncorrelated objects. By including the SAOC IOC parameters, it is possible to take inter-object correlations into account, e.g. by using the equation:

$$15 \quad \sqrt{\frac{M_{Nto5,1k}^2 \cdot OLD_k}{\sum_{i=1}^N \sum_{j=1}^N M_{Nto5,1i} \cdot M_{Nto5,1j} \cdot \sqrt{OLD_i \cdot OLD_j} \cdot IOC_{ij}}} \cdot X_1$$

This reconstruction is weighed with the gain of object k in downmix channel 1 ($M_{Nto5,1k}$).

Combining similar reconstructions from all 5 channels gives an object reconstruction that is weighed according to the gains to object k, i.e. the channel in which object k has the largest gain provides the largest contribution to the combined reconstruction \tilde{S}_k of object k:

$$\tilde{S}_k = \frac{\sum_{c=1}^5 \sqrt{\frac{M_{Nto5,ck}^2 \cdot OLD_k}{\sum_{i=1}^N \sum_{j=1}^N M_{Nto5,ci} \cdot M_{Nto5,cj} \cdot \sqrt{OLD_i \cdot OLD_j} \cdot IOC_{ij}}} \cdot X_c}{\sum_{c=1}^5 M_{Nto5,ck}}$$

25 where $\sum_{c=1}^5 M_{Nto5,ck}$ normalizes the reconstruction to the correct level.

As another example, an alternative weighed reconstruction could aim at ‘isolatedness’ of an object in a downmix channel.

Define:

$$B_{ck} = \sqrt{\frac{M_{Nto5,ck}^2 \cdot OLD_k}{\sum_{i=1}^N \sum_{j=1}^N M_{Nto5,ci} \cdot M_{Nto5,cj} \cdot \sqrt{OLD_i \cdot OLD_j} \cdot IOC_{ij}}},$$

then the alternative reconstruction

$$\tilde{S}_k = \frac{\sum_{c=1}^5 B_{ck}^2 \cdot \frac{B_{ck} \cdot X_c}{M_{Nto5,ck}}}{\sum_{c=1}^5 B_{ck}^2},$$

weighs each normalized sub-reconstruction $(B_{ck} \cdot X_c)$ of object k with its relative contribution to the corresponding downmix channel.

It will be appreciated that other approaches for generating the audio object from the M audio channels and the N to M downmix can be used in other embodiments.

In an SAOC encoder where Enhanced Audio Objects (EAO) are encoded, the corresponding residual signals are calculated as a difference between the original object signal and a reconstruction based on the mono or stereo SAOC downmix. These enhanced objects (X_{eao}) are therefore processed separately from the regular objects (X_{reg}).

The regular objects are downmixed according to a submatrix (D_{reg}) of the $K \times N$ downmix matrix (D), where $D = \begin{pmatrix} D_{reg} & D_{eao} \end{pmatrix}$ when $X = \begin{pmatrix} X_{reg} \\ X_{eao} \end{pmatrix}$. The result is a K-channel downmix:

$$Y_{reg} = D_{reg} \cdot X_{reg}$$

The EAOs are also downmixed using the corresponding submatrix D_{eao} , and the resulting downmix is combined with the downmix of the regular objects (Y_{reg}) into the SAOC downmix.

$$Y = Y_{reg} + D_{eao} \cdot X_{eao}$$

This downmix is expected at the input of the SAOC decoder.

- 5 Using downmix Y_{reg} and the EAOs as input signals, intermediate auxiliary signals are calculated using the $N_{eao} \times (K + N_{eao})$ matrix D_{aux} , where $N_{eao} = N - N_{reg}$ the number of EAOs.

$$Y_{aux} = D_{aux} \cdot \begin{pmatrix} Y_{reg} \\ X_{eao} \end{pmatrix}$$

10

The generation of the downmix Y and auxiliary signals Y_{aux} can be combined in a single matrix equation:

$$Y_{ext} = \begin{pmatrix} Y \\ Y_{aux} \end{pmatrix} = D_{ext} \cdot \begin{pmatrix} Y_{reg} \\ X_{eao} \end{pmatrix}$$

15

where

$$D_{ext} = \left(\begin{array}{cc|c} 1 & 0 & D_{eao} \\ 0 & 1 & \\ \hline & & D_{aux} \end{array} \right).$$

20

Matrix D_{aux} is chosen such that matrix D_{ext} is invertible and the EAO separation from the downmix is optimized. The elements of D_{aux} are defined in the SAOC standard and thus available in the decoder. In the SAOC decoder, using the inverse of D_{ext} , the EAOs (X_{eao}) can be separated from the regular objects (Y_{reg}) using the downmix (Y) and auxiliary signals (Y_{aux}) as an input.

25

In order to improve coding efficiency, the auxiliary signals are predicted from the downmix signals with prediction coefficients that are derived from data already available in the decoder.

$$\hat{Y}_{aux} = C \cdot Y$$

The prediction error $R = Y_{aux} - \hat{Y}_{aux}$ can be efficiently coded using the residual coding mechanism of the SAOC standard.

The residuals of this embodiment can be generated in the same way as described above using the M-channel object reconstruction \tilde{S} as the EAOs ($= X_{eao}$). Since
5 the individual objects are already mixed, these steps can be omitted. Thus giving

$$Y_{ext} = D'_{ext} \cdot \begin{pmatrix} Y \\ \tilde{S} \end{pmatrix},$$

with

$$10 \quad D'_{ext} = \left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \hline & D_{aux} & \end{array} \right),$$

and

$$Y'_{ext} = \begin{pmatrix} Y \\ Y'_{aux} \end{pmatrix}.$$

In case of four EAOs:

$$15 \quad D'_{ext} = \left(\begin{array}{cccccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \hline d_{ext,31} & d_{ext,32} & d_{ext,33} & \cdots & \cdots & d_{ext,61} \\ \vdots & & & \ddots & & \vdots \\ \vdots & & & & \ddots & \vdots \\ d_{ext,61} & \cdots & \cdots & \cdots & \cdots & d_{ext,66} \end{array} \right).$$

The residuals are then calculated as

$$R' = Y'_{aux} - \hat{Y}'_{aux},$$

20 with

$$\begin{pmatrix} Y \\ \hat{Y}'_{aux} \end{pmatrix} = C \cdot Y.$$

The resulting residuals (R') can then be inserted in the SAOC bitstream, in which the objects for which the residuals are calculated are identified as EAOs. The standard SAOC decoder can then proceed to perform a standard SAOC EAO decoding to generate the N audio channels.

5 This may provide improved quality of the decoded audio objects in many embodiments. In many embodiments it may allow compatibility with standardized audio object decoding algorithms capable of receiving residual data, such as for example the SAOC standard. The residual data may specifically be indicative of a difference between an audio object generated from the K channels and the audio object upmix parameters and the
10 corresponding audio object generated on the basis of the M audio channels and the downmix data.

It will be appreciated that the above description for clarity has described embodiments of the invention with reference to different functional circuits, units and processors. However, it will be apparent that any suitable distribution of functionality
15 between different functional circuits, units or processors may be used without detracting from the invention. For example, functionality illustrated to be performed by separate processors or controllers may be performed by the same processor or controllers. Hence, references to specific functional units or circuits are only to be seen as references to suitable means for providing the described functionality rather than indicative of a strict logical or physical
20 structure or organization.

The invention can be implemented in any suitable form including hardware, software, firmware or any combination of these. The invention may optionally be implemented at least partly as computer software running on one or more data processors and/or digital signal processors. The elements and components of an embodiment of the
25 invention may be physically, functionally and logically implemented in any suitable way. Indeed the functionality may be implemented in a single unit, in a plurality of units or as part of other functional units. As such, the invention may be implemented in a single unit or may be physically and functionally distributed between different units, circuits and processors.

Although the present invention has been described in connection with some
30 embodiments, it is not intended to be limited to the specific form set forth herein. Rather, the scope of the present invention is limited only by the accompanying claims. Additionally, although a feature may appear to be described in connection with particular embodiments, one skilled in the art would recognize that various features of the described embodiments

may be combined in accordance with the invention. In the claims, the term comprising does not exclude the presence of other elements or steps.

Furthermore, although individually listed, a plurality of means, elements, circuits or method steps may be implemented by e.g. a single circuit, unit or processor.

- 5 Additionally, although individual features may be included in different claims, these may possibly be advantageously combined, and the inclusion in different claims does not imply that a combination of features is not feasible and/or advantageous. Also the inclusion of a feature in one category of claims does not imply a limitation to this category but rather indicates that the feature is equally applicable to other claim categories as appropriate.
- 10 Furthermore, the order of features in the claims do not imply any specific order in which the features must be worked and in particular the order of individual steps in a method claim does not imply that the steps must be performed in this order. Rather, the steps may be performed in any suitable order. In addition, singular references do not exclude a plurality. Thus references to "a", "an", "first", "second" etc do not preclude a plurality. Reference signs
- 15 in the claims are provided merely as a clarifying example shall not be construed as limiting the scope of the claims in any way.

CLAIMS:

1. An audio object encoder comprising:
a receiver (701) for receiving N audio objects;
a mixer (703) for mixing the N audio objects to M audio channels;
a channel circuit (707) for deriving K audio channels from the M audio
5 channels where $K = 1$ or 2 and $K < M$;
a parameter circuit (709) generating audio object upmix parameters for at least
part of each of the N audio objects relative to the K audio channels;
an output circuit (705, 711) for generating an output data stream comprising
the audio object upmix parameters and the M audio channels.
10
2. The audio object encoder of claim 1 wherein the channel circuit (707) is
arranged to derive the K channels by downmixing the M audio channels.
3. The audio object encoder of claim 1 wherein the channel circuit (707) is
15 arranged to derive the K channels by selecting a K channel subset of the M audio channels.
4. The audio object encoder of claim 1 wherein the output data stream comprises
a multichannel encoded data stream for the M audio channels, and the audio object upmix
parameters are comprised in a part of the multichannel encoded data stream.
20
5. The audio object encoder of claim 1 wherein the output circuit (705, 711) is
arranged to include mixing data representative of the mixing of the N audio objects to the M
audio channels in the output data stream.
- 25 6. An audio object decoder comprising:
a receiver (801, 803) for receiving a data stream comprising audio data for an
M channel mix of N audio objects and audio object mix parameters for the N audio objects
relative to K audio channels where $K = 1$ or 2 and $K < M$;
a channel circuit (805) deriving K audio channels from the M channel mix;

and

an object decoder (807) for generating P audio signals from N audio objects at least partially generated by upmixing the K audio channels based on the audio object upmix parameters.

5

7. The audio object decoder of claim 6 wherein the channel circuit (805) is arranged to derive the K channels by downmixing the M audio channels.

8. The audio object decoder of claim 7 wherein the data stream further comprises
10 downmix data indicative of an encoder downmixing from M to K channels, and wherein the channel circuit (805) is arranged to adapt the downmixing in response to the downmix data.

9. The audio object decoder of claim 7 wherein the channel circuit (805) is arranged to derive the K channels by selecting a K channel subset of the M audio channels.

15

10. The audio object decoder of claim 9 wherein the data stream further comprises additional audio object upmix parameters for the N audio objects relative to L audio channels where $L = 1$ or 2 and $L < M$, and the L audio channels and the K audio channels are different subsets of the M audio channels, and wherein the object decoder (807) is further arranged to
20 generate the P channels from N audio objects at least partially generated by upmixing the L audio channels based on the additional audio object upmix parameters.

11. The audio object decoder of claim 10 wherein at least one of the P channels is generated by combining contributions from both the upmixing of the K audio channels based
25 on the audio object upmix parameters and the upmixing of the L audio channels based on the additional audio object upmix parameters.

12. The audio object decoder of claim 6 wherein the data stream comprises mix data representative of the mixing of the N audio objects to the M audio channels, and wherein
30 the object decoder (807) is arranged to generate residual data for at least a subset of the N audio objects in response to the mix data and the audio object upmix parameters, and to generate the P audio signals in response to the residual data.

13. A method of audio object encoding comprising:
receiving N audio objects;
mixing the N audio objects to M audio channels;
deriving K audio channels from the M audio channels where $K = 1$ or 2 and

5 $K < M$;

generating audio object upmix parameters for at least part of each of the N
audio objects relative to the K audio channels; and

generating an output data stream comprising the audio object upmix
parameters and the M audio channels.

10

14. A method of audio object decoding comprising:
receiving a data stream comprising audio data for an M channel mix of N
audio objects and audio object upmix parameters for the N audio objects relative to K audio
channels where $K = 1$ or 2 and $K < M$;

15

deriving K audio channels from the M channel mix; and

generating P audio signals from N audio objects at least partially generated by
upmixing the K audio channels based on the audio object upmix parameters.

15. A computer program product comprising computer program code means

20 adapted to perform all the steps of claims 13 or 14 when said program is run on a computer.

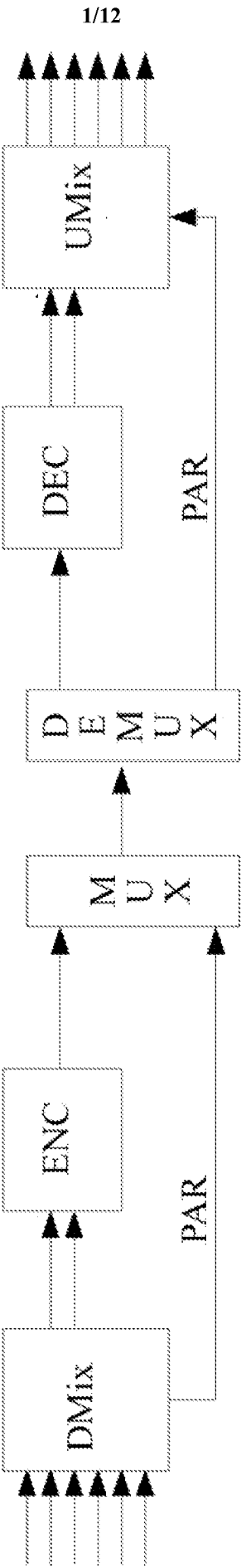
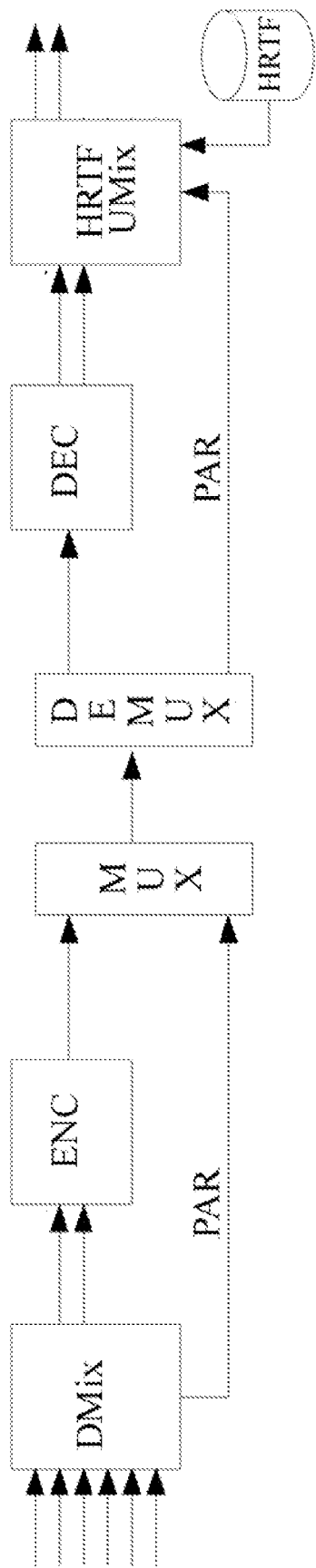


FIG. 1

Prior Art



Prior Art

FIG. 2

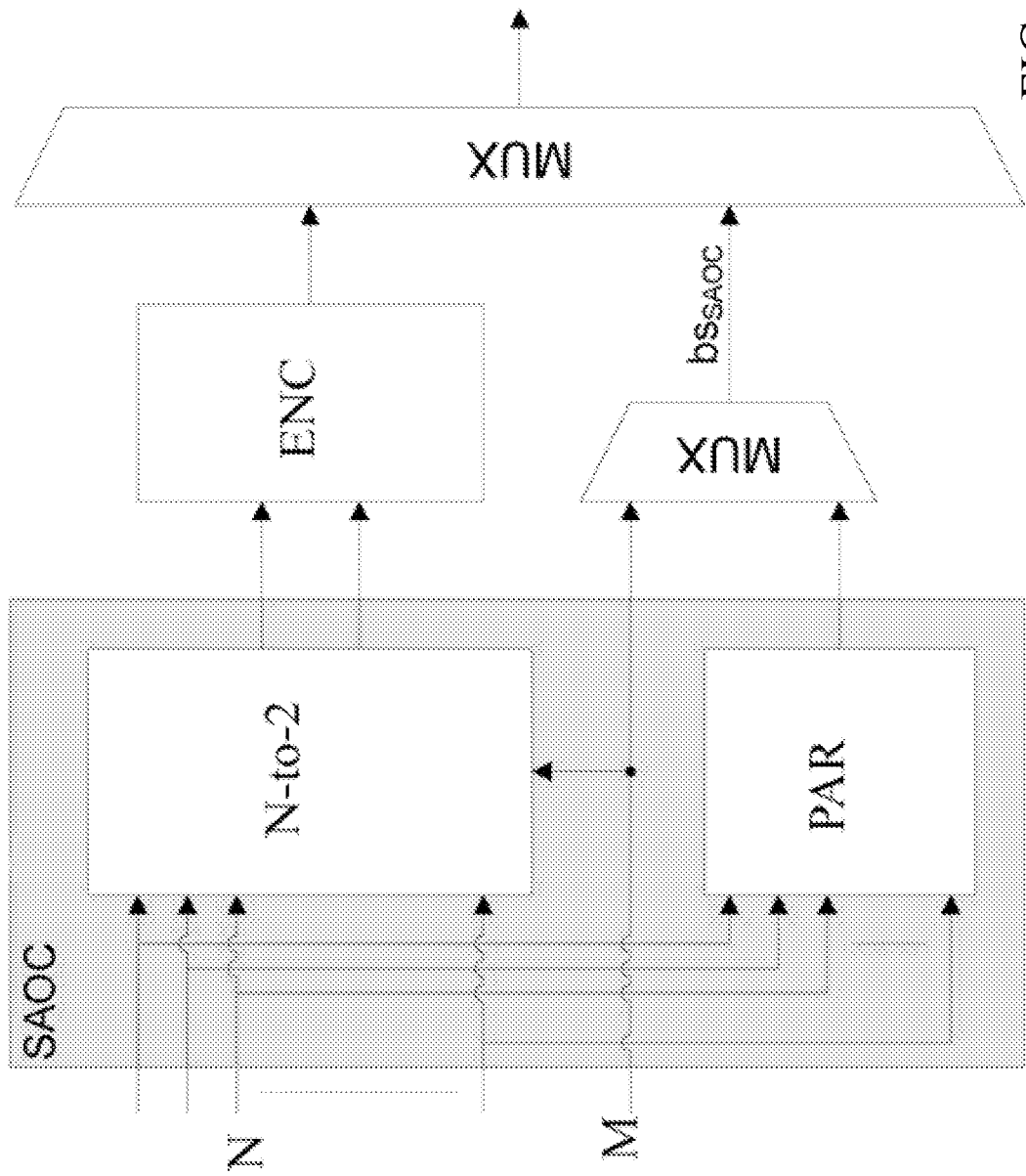


FIG. 3

Prior Art

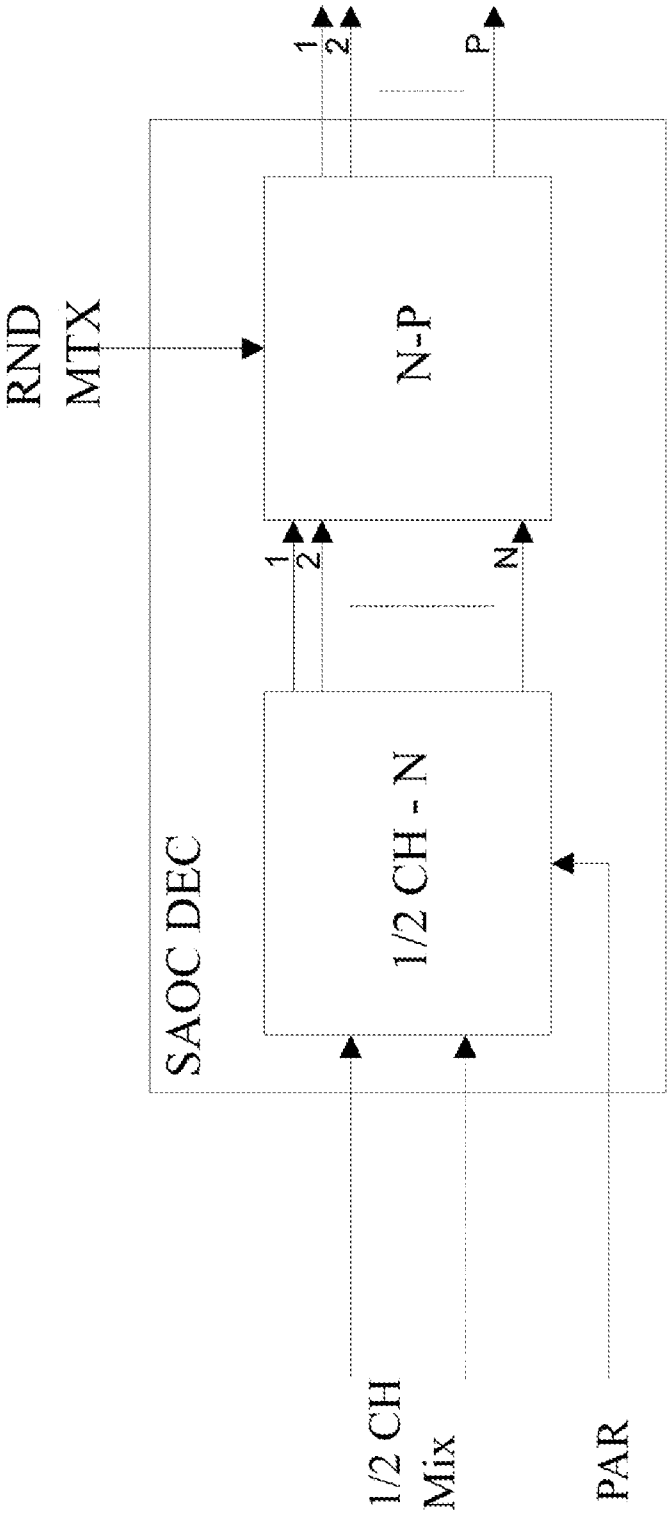


FIG. 4

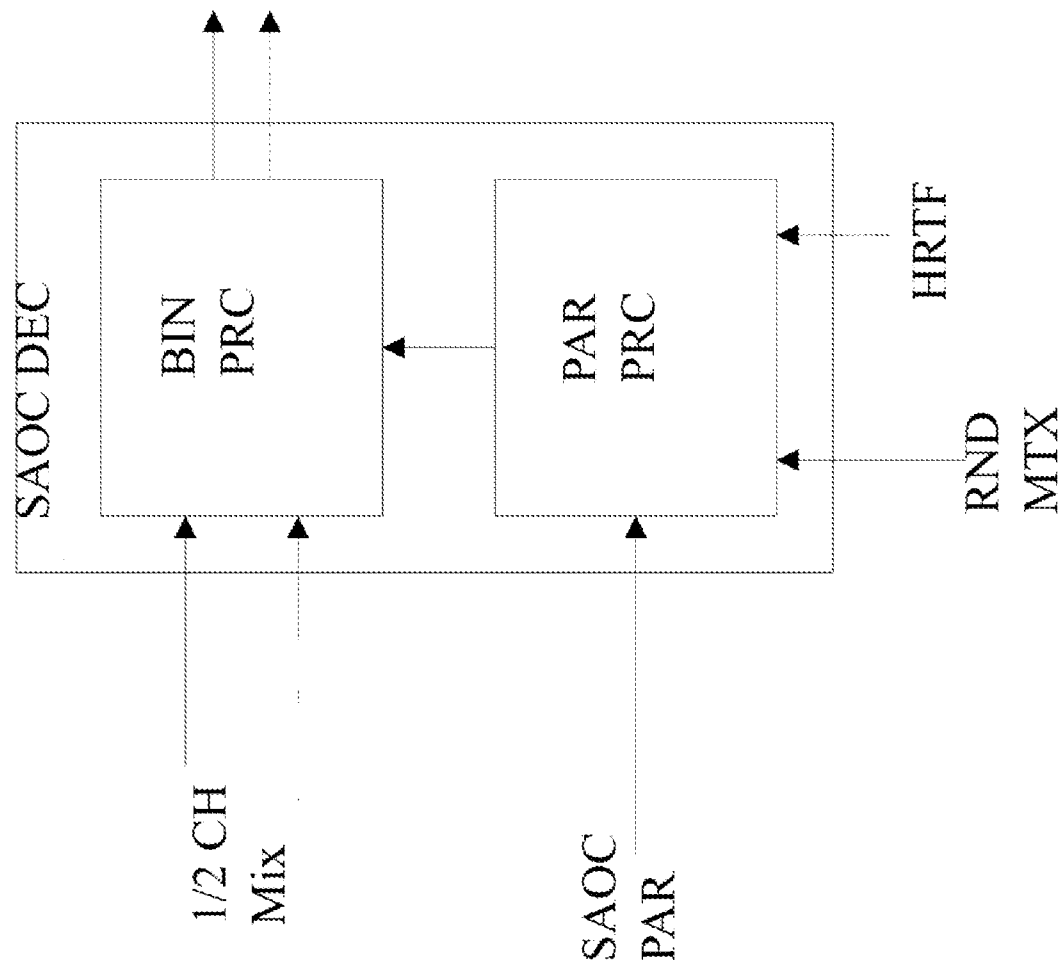


FIG. 5

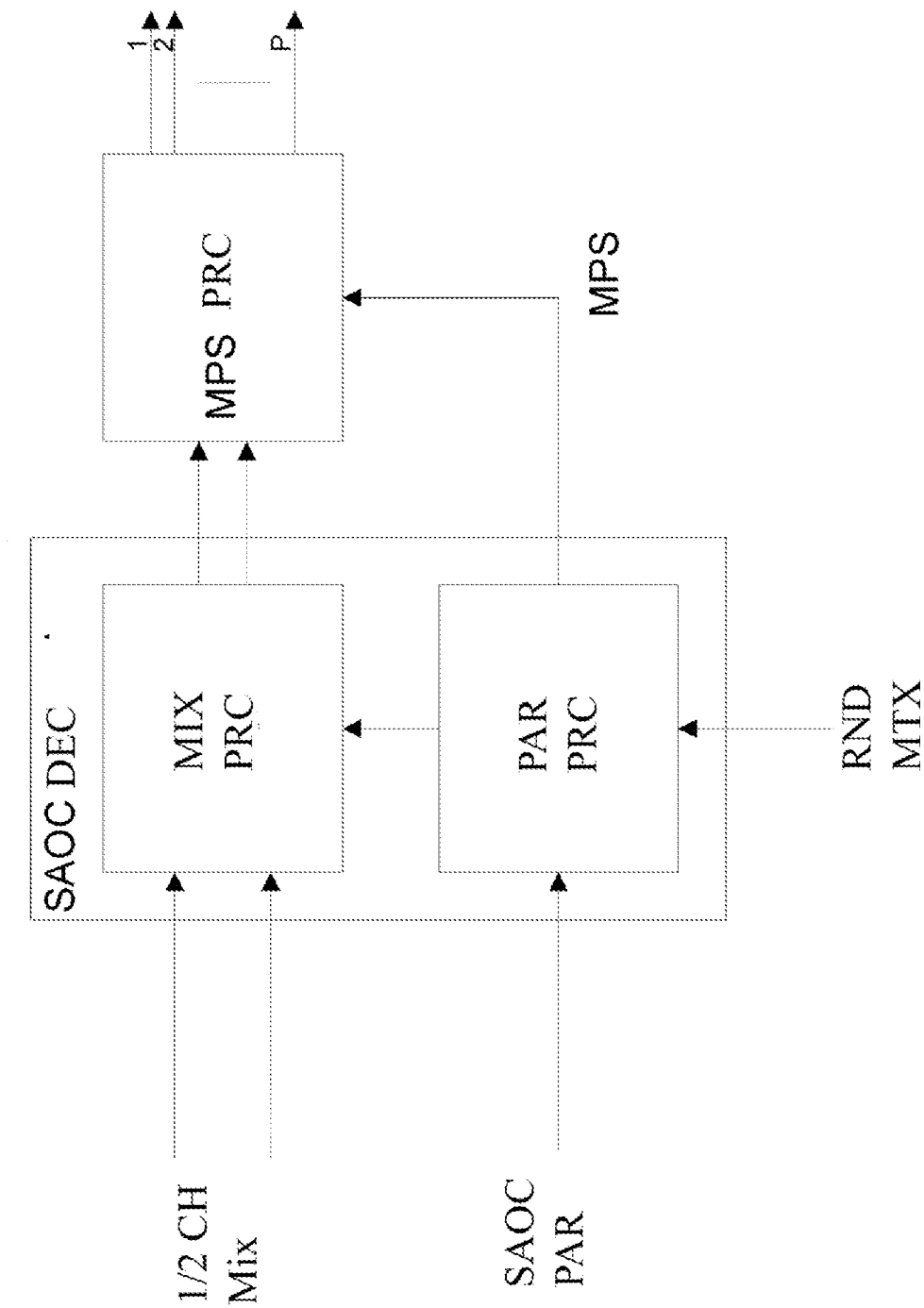


FIG. 6

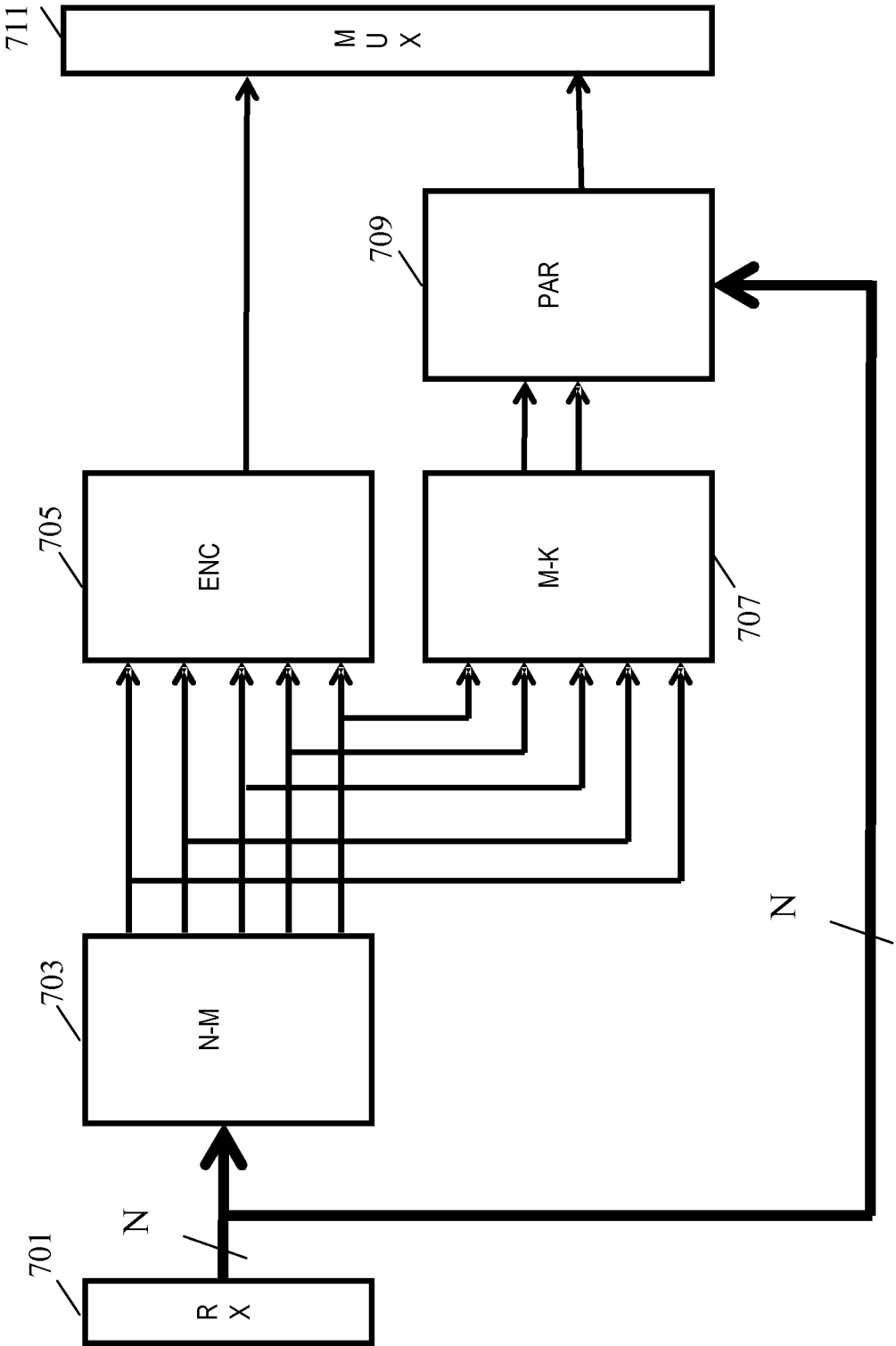


FIG. 7

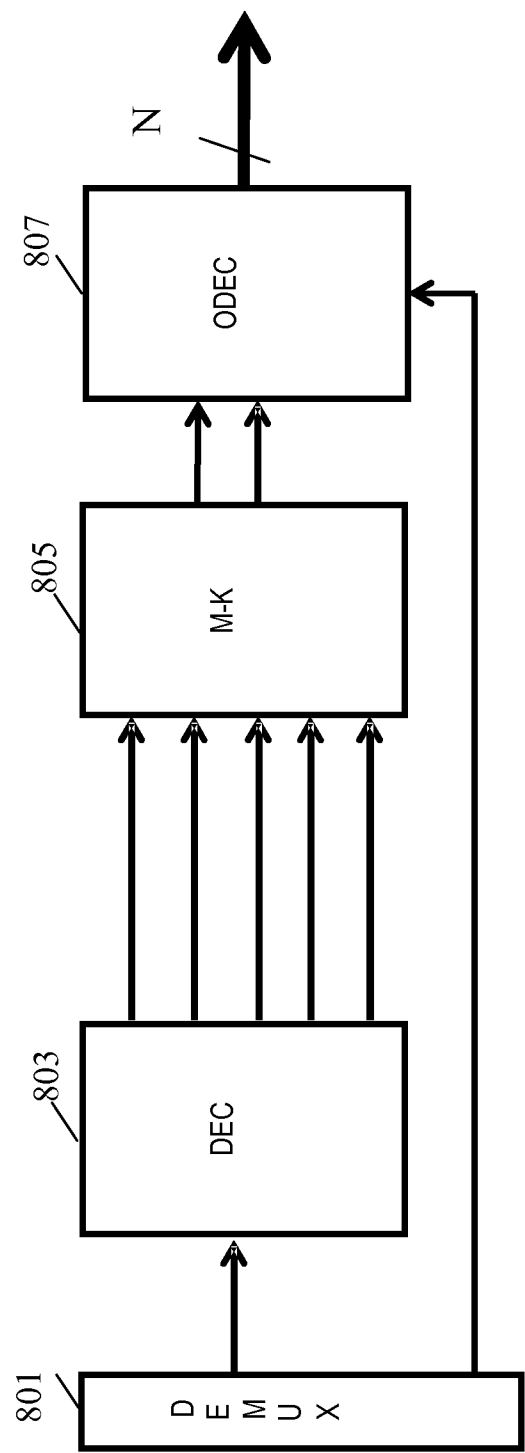


FIG. 8

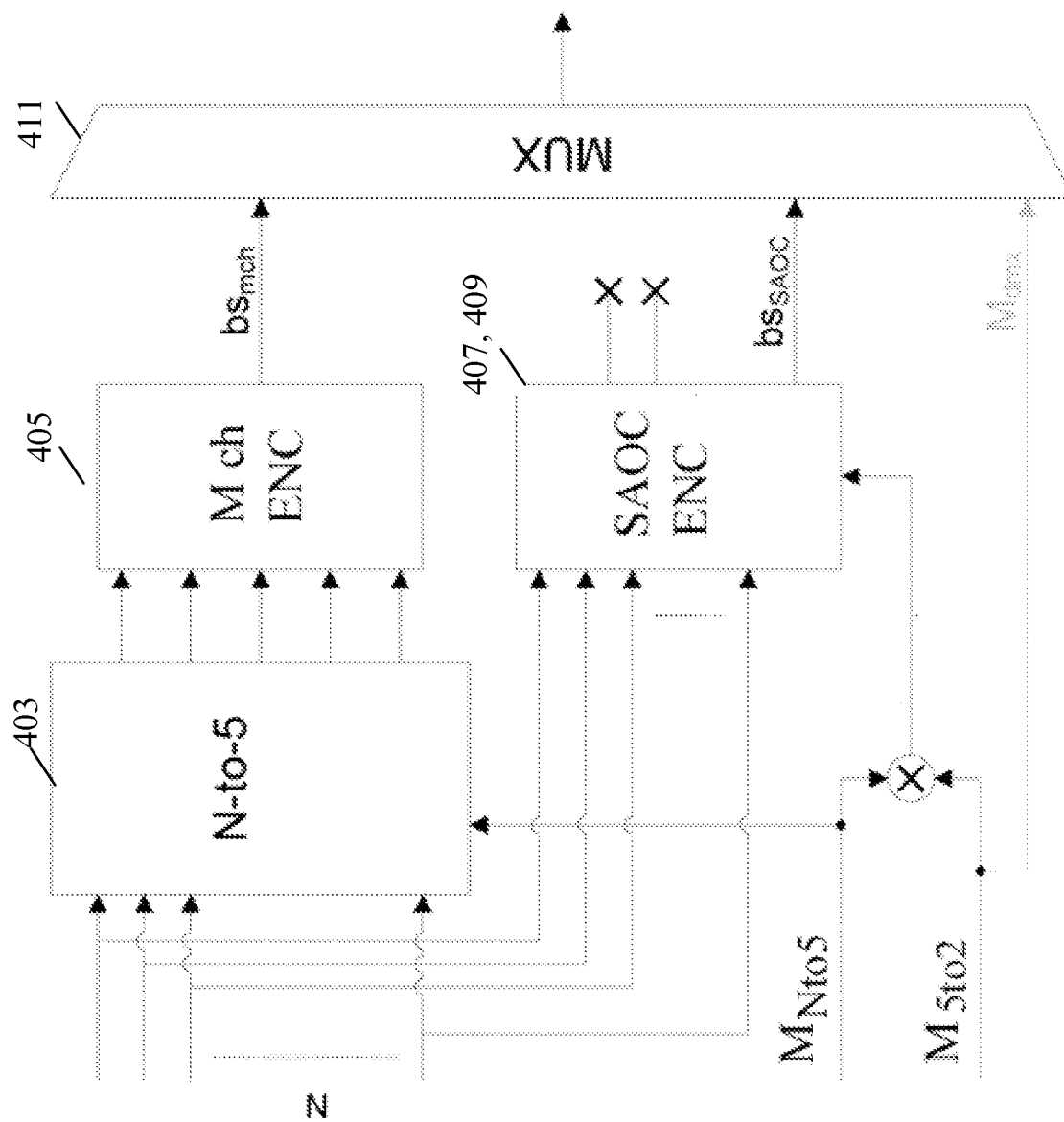


FIG. 9

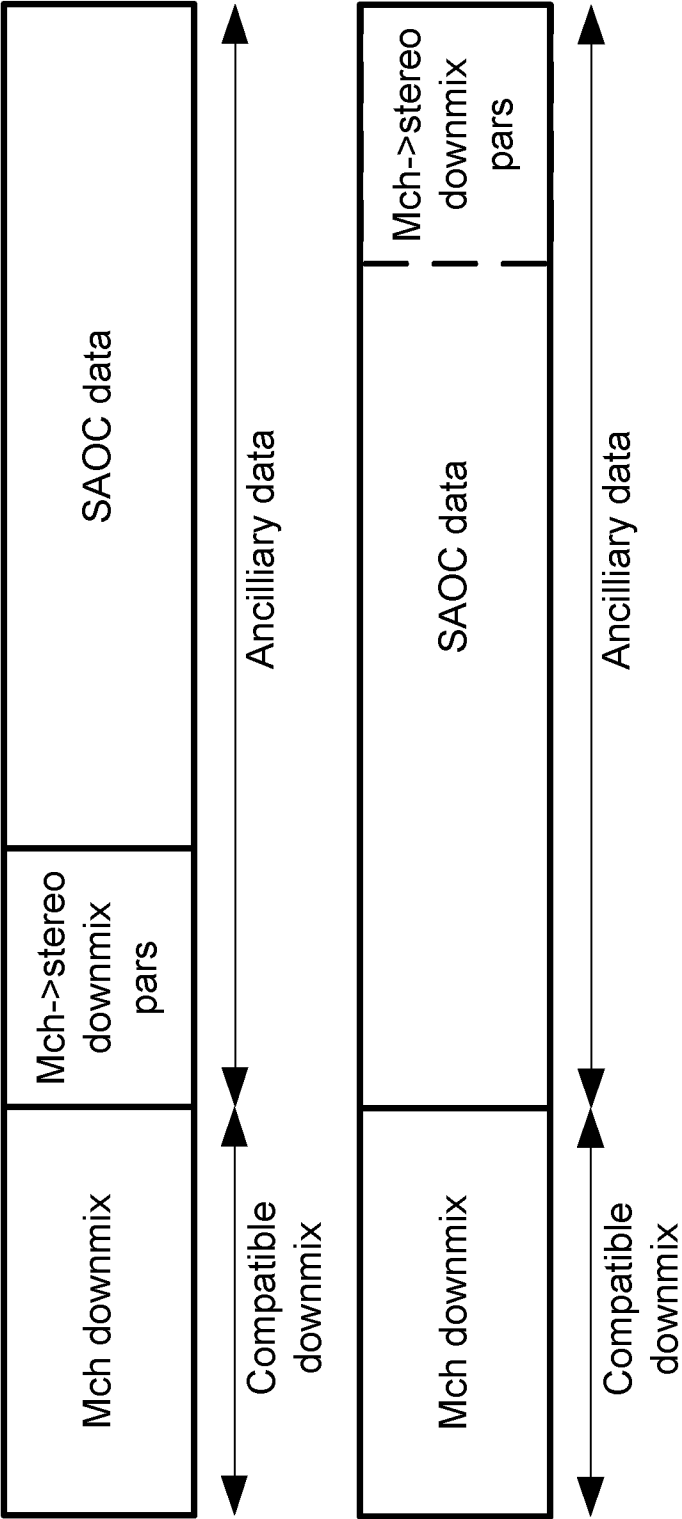


FIG. 10

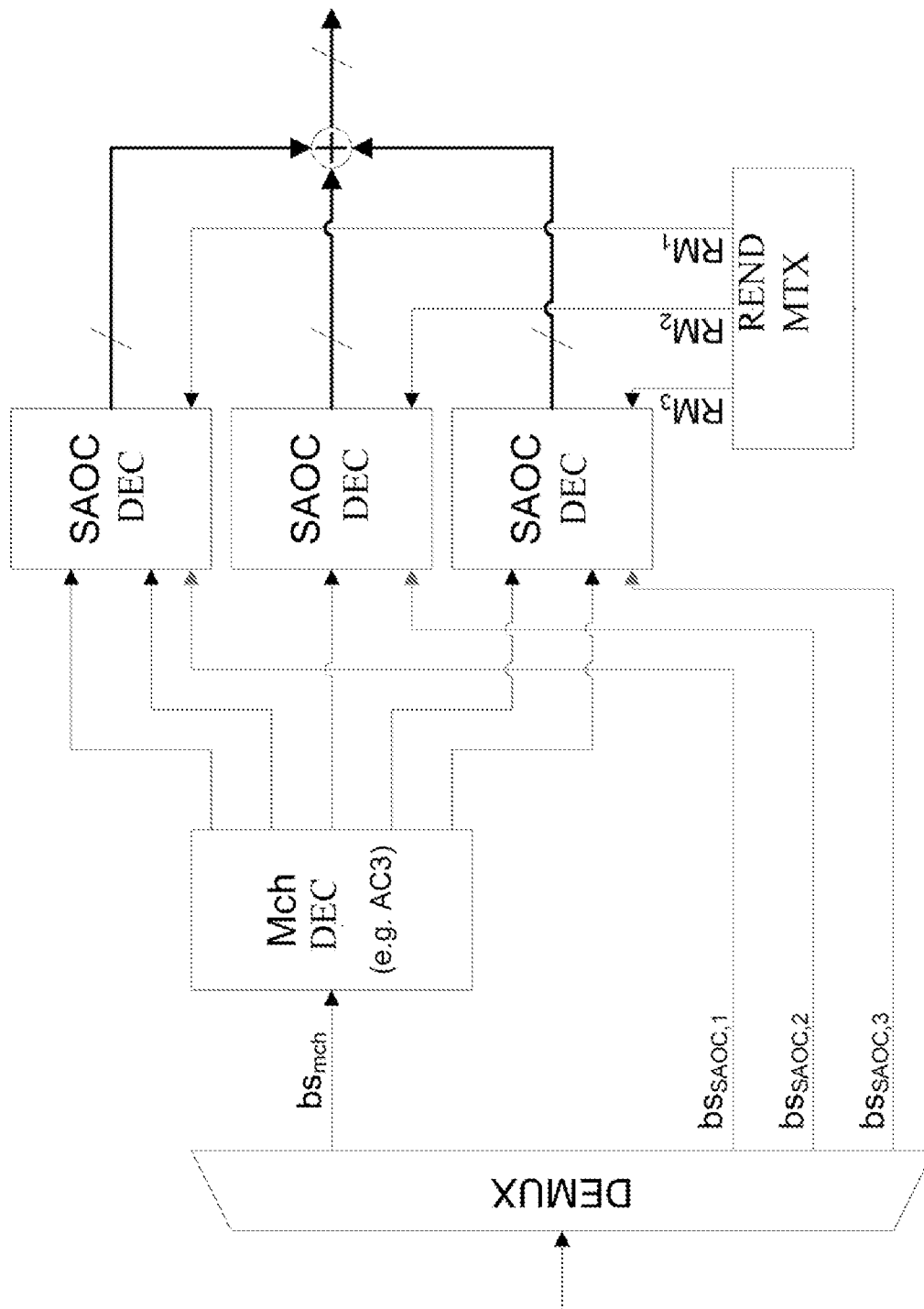


FIG. 11

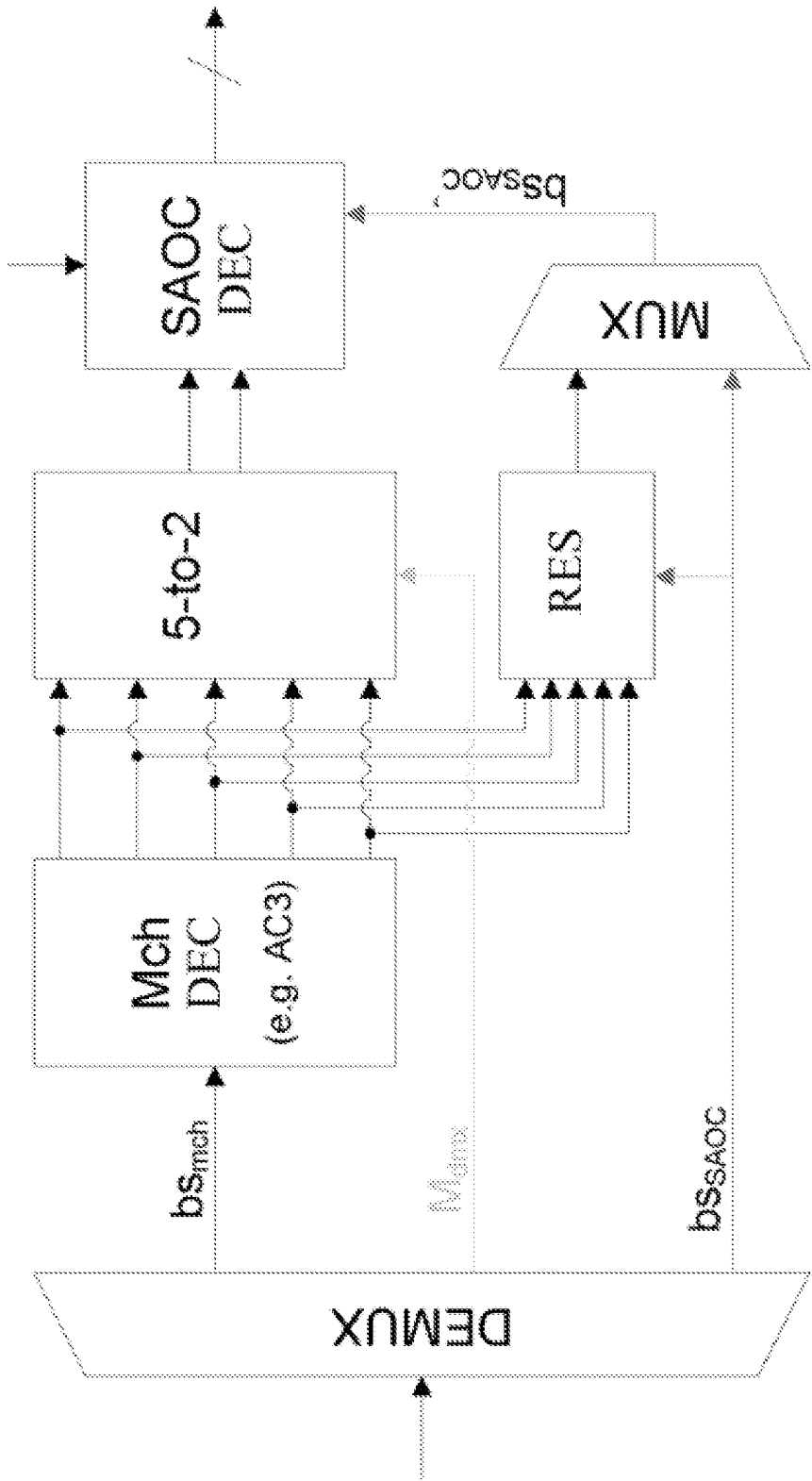


FIG. 12

INTERNATIONAL SEARCH REPORT

International application No

PCT/IB2012/055964

A. CLASSIFICATION OF SUBJECT MATTER

INV. G10L19/008

ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, WPI Data, INSPEC, COMPENDEX

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	ANONYMOUS: "ISO/IEC FDIS 23003-2: 2010, Spatial Audio Object Coding", 91. MPEG MEETING;18-1-2010 - 22-1-2010; KYOTO; (MOTION PICTURE EXPERTGROUP OR ISO/IEC JTC1/SC29/WG11),, no. N11207, 10 May 2010 (2010-05-10), XP030017704, ISSN: 0000-0030 cited in the application section 5.1; section D.1	1-15
A	----- WO 2008/046531 A1 (CODING TECH AB [SE]; ENGDEGARD JONAS [SE]; VILLEMOS LARS [SE]; PURNHA) 24 April 2008 (2008-04-24) page 19, line 21 - page 20, line 22; figure 9 page 21, line 1 - page 24, line 21; figure 14 -----	1-15



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

16 April 2013

Date of mailing of the international search report

24/04/2013

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

De Meuleneire, M

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/IB2012/055964

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2008046531	A1	24-04-2008	
		AT 503245 T	15-04-2011
		AT 536612 T	15-12-2011
		AU 2007312598 A1	24-04-2008
		AU 2011201106 A1	07-04-2011
		CA 2666640 A1	24-04-2008
		CN 101529501 A	09-09-2009
		CN 102892070 A	23-01-2013
		EP 2054875 A1	06-05-2009
		EP 2068307 A1	10-06-2009
		EP 2372701 A1	05-10-2011
		ES 2378734 T3	17-04-2012
		HK 1126888 A1	02-12-2011
		HK 1133116 A1	27-07-2012
		JP 2010507115 A	04-03-2010
		JP 2012141633 A	26-07-2012
		KR 20090057131 A	03-06-2009
		KR 20110002504 A	07-01-2011
		MY 145497 A	29-02-2012
		PL 2068307 T3	31-07-2012
		RU 2009113055 A	27-11-2010
		RU 2011102416 A	27-07-2012
		SG 175632 A1	28-11-2011
		TW 200828269 A	01-07-2008
		US 2011022402 A1	27-01-2011
		WO 2008046531 A1	24-04-2008