

(21) Application No: 1020087.1

(22) Date of Filing: 26.11.2010

(71) Applicant(s):
The University of Surrey
(Incorporated in the United Kingdom)
GUILDFORD, Surrey, GU2 7XH, United Kingdom

(72) Inventor(s):
Ikhwana Elfitri
Ahmet Kondo
Banu Gunel

(74) Agent and/or Address for Service:
Barker Brettell LLP
100 Hagley Road, Edgbaston, BIRMINGHAM,
B16 8QQ, United Kingdom

(51) INT CL:
G10L 19/00 (2006.01) **G10L 19/14** (2006.01)
H04S 7/00 (2006.01)

(56) Documents Cited:
WO 2008/078973 A1 **WO 2008/039039 A1**
US 20020090113 A1

(58) Field of Search:
INT CL **G10L**
Other: **WPI EPODOC TXTE**

(54) Title of the Invention: **Spatial audio coding**
Abstract Title: **Spatial audio coding**

(57) A system for coding multi-channel audio signals for transmission to a decoder comprises: an input arranged to receive an audio signal (106); a trial coding signal generator (101) arranged to output a trial coding signal; a model (102) of the decoder arranged to receive, as an input, the trial coding signal and to synthesize, as an output, a trial re-synthesized audio signal; and optimization means (103) arranged to compare the resynthesized audio signal with the received audio signal thereby to determine the form of an optimized coding signal, and to transmit the optimized coding signal.

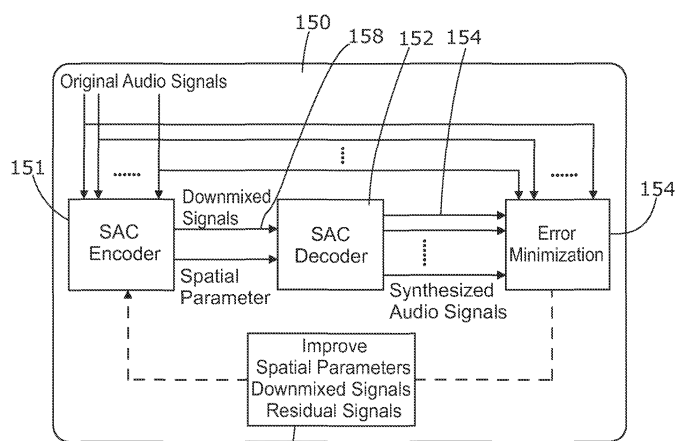


Fig. 15

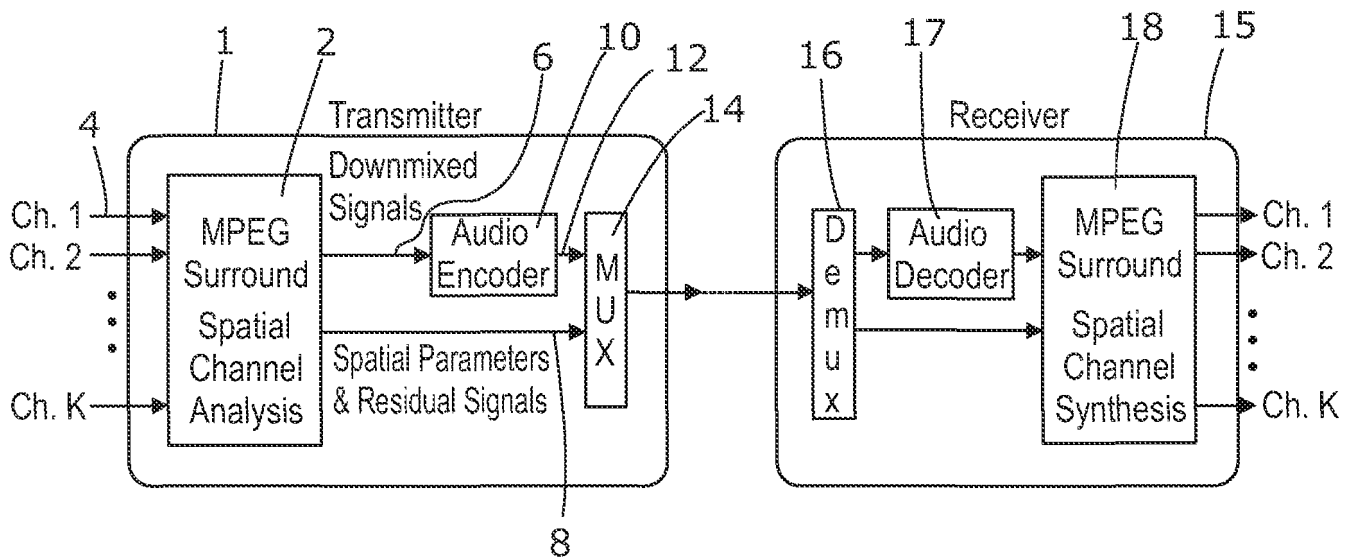


Fig. 1

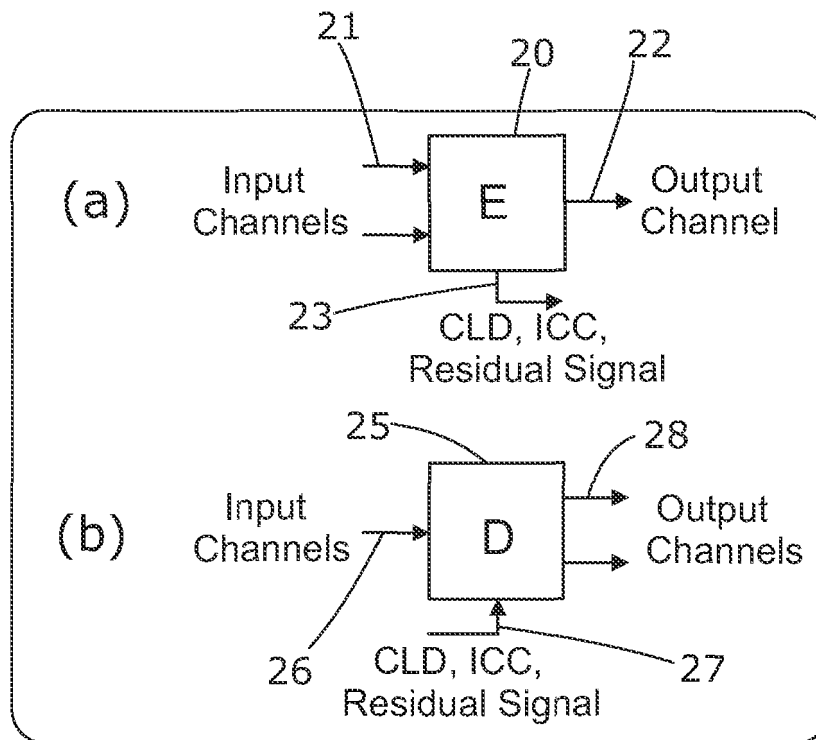


Fig. 2

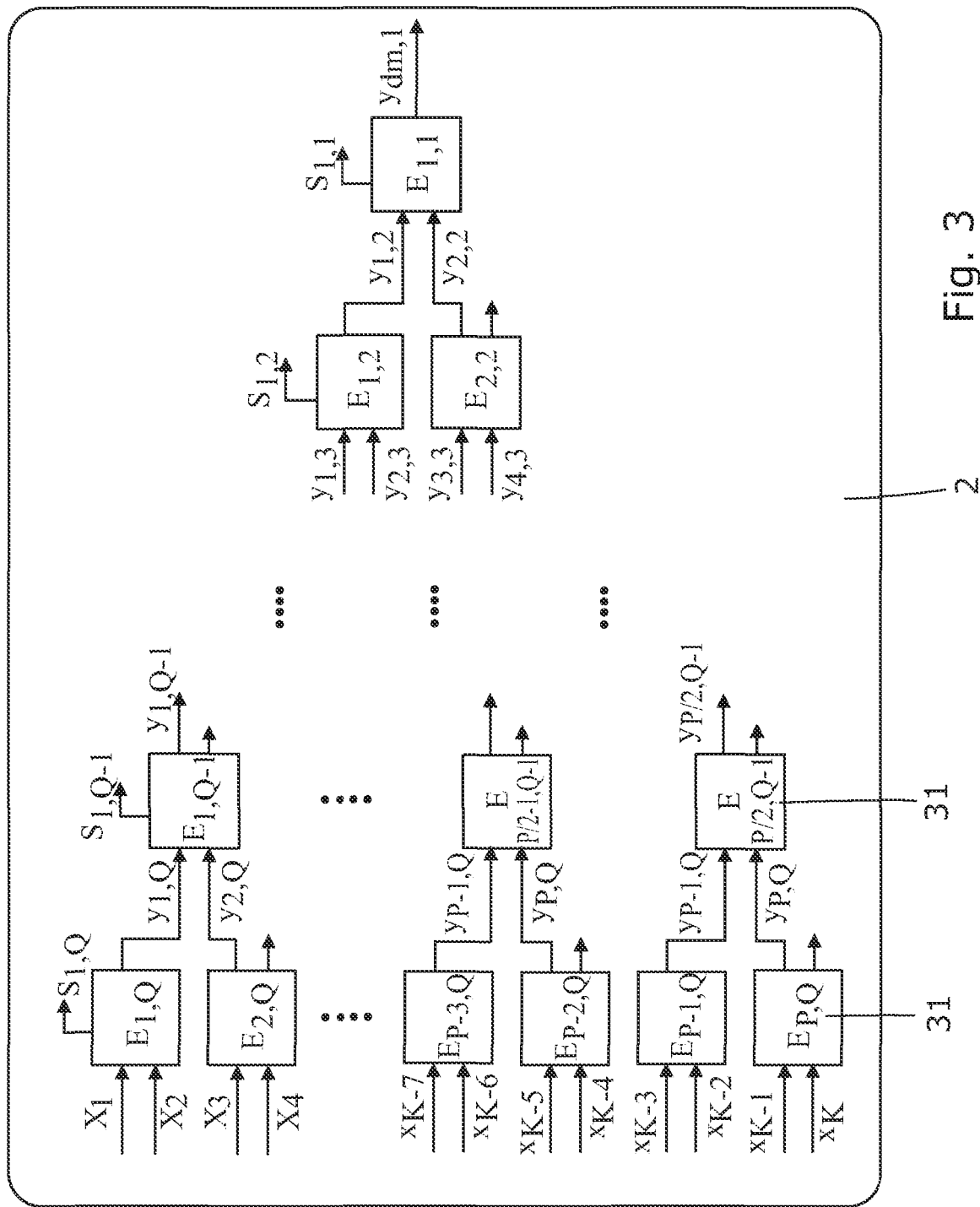


Fig. 3

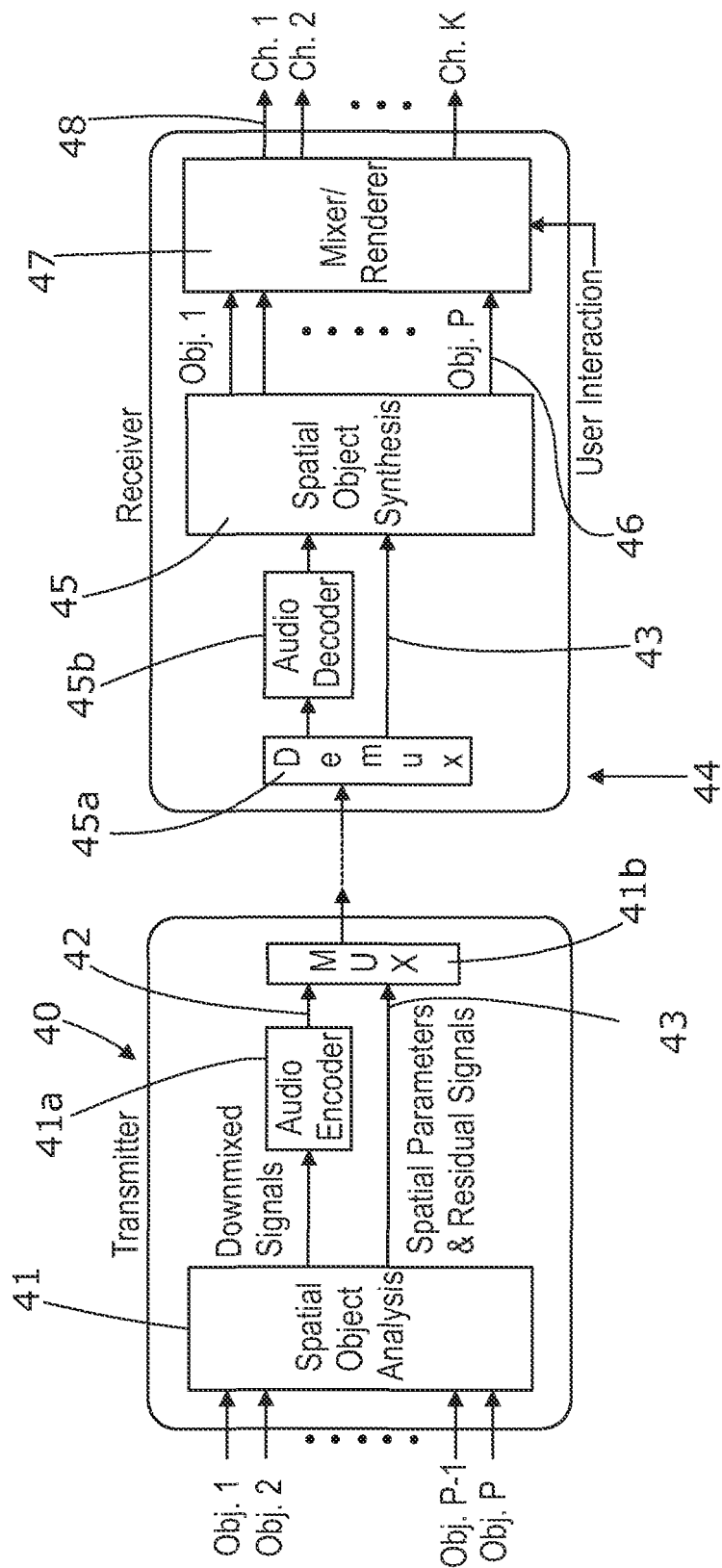


Fig. 4

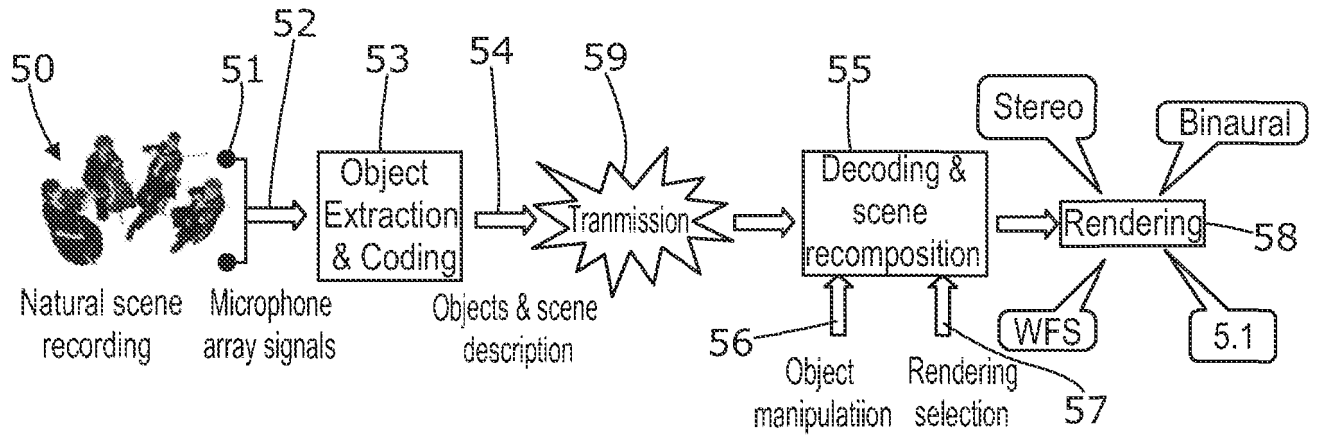


Fig. 5

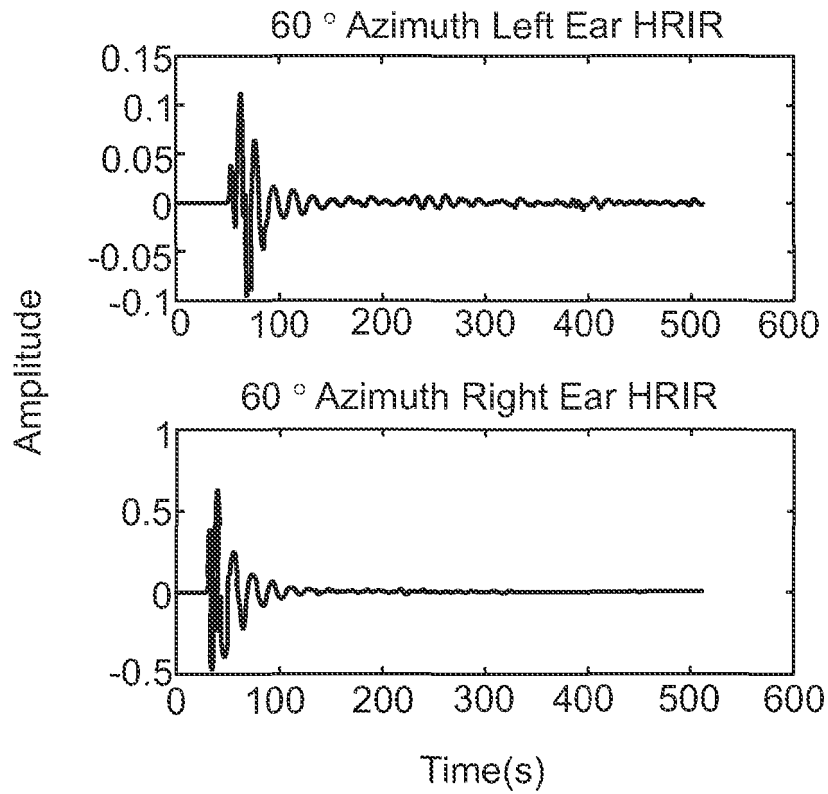
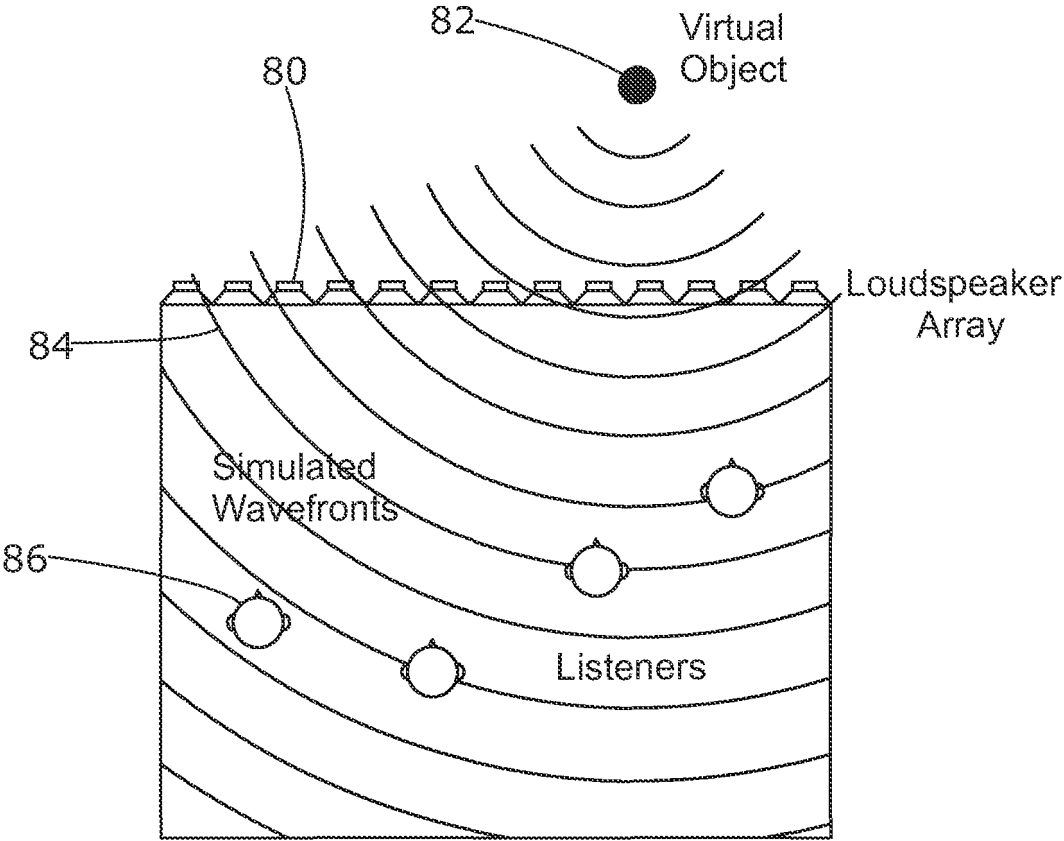
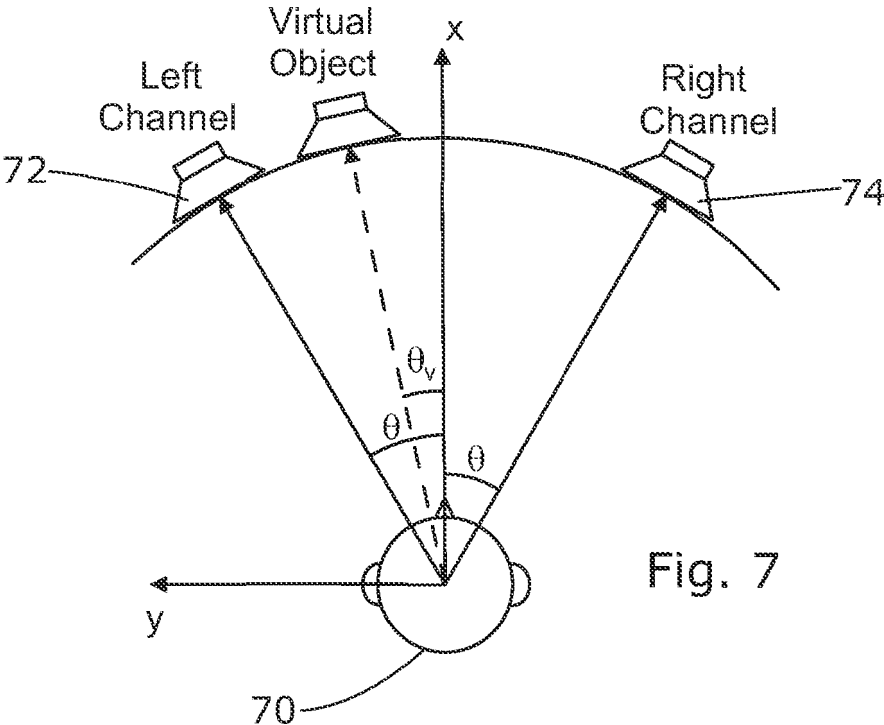


Fig. 6



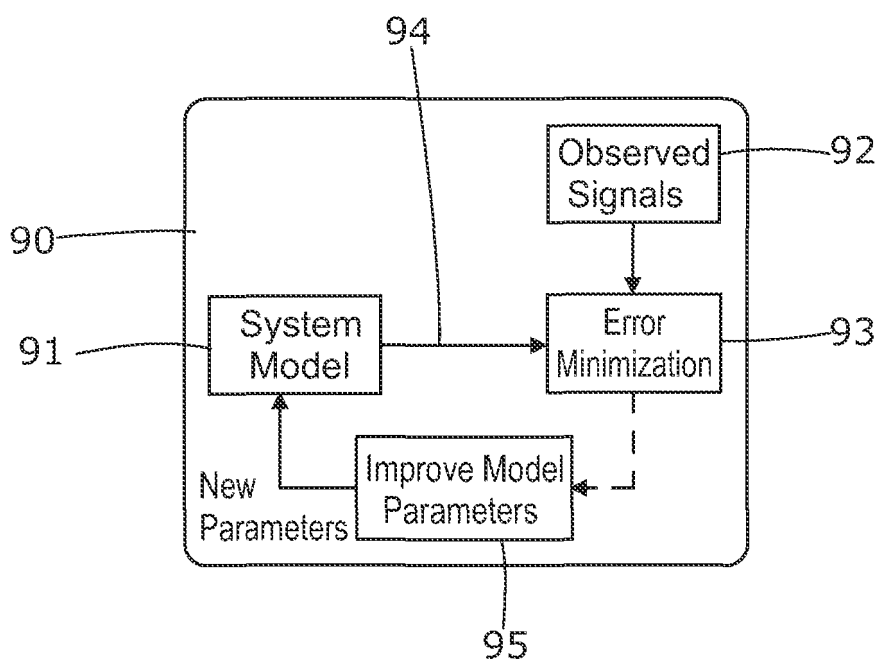


Fig. 9

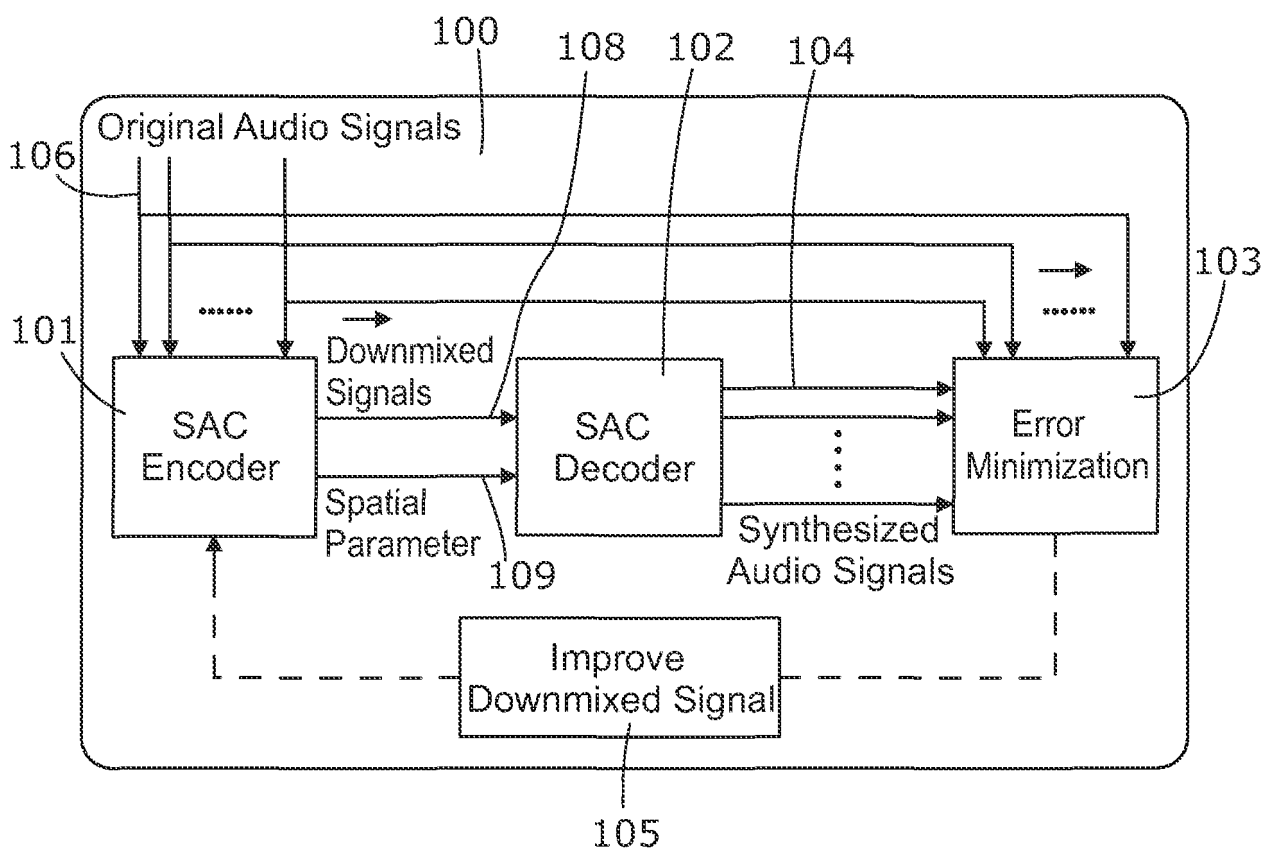


Fig. 10

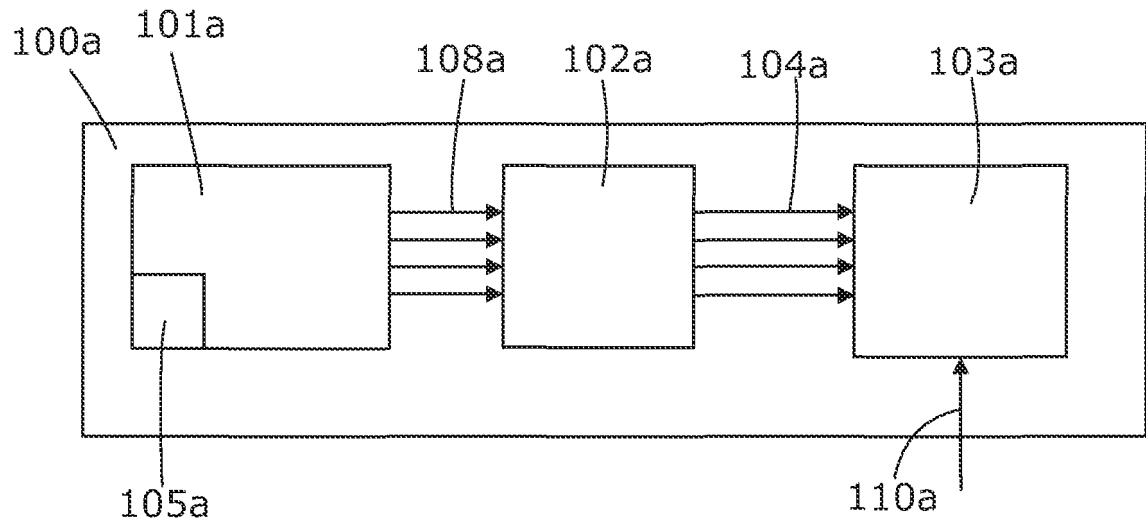


Fig. 10a

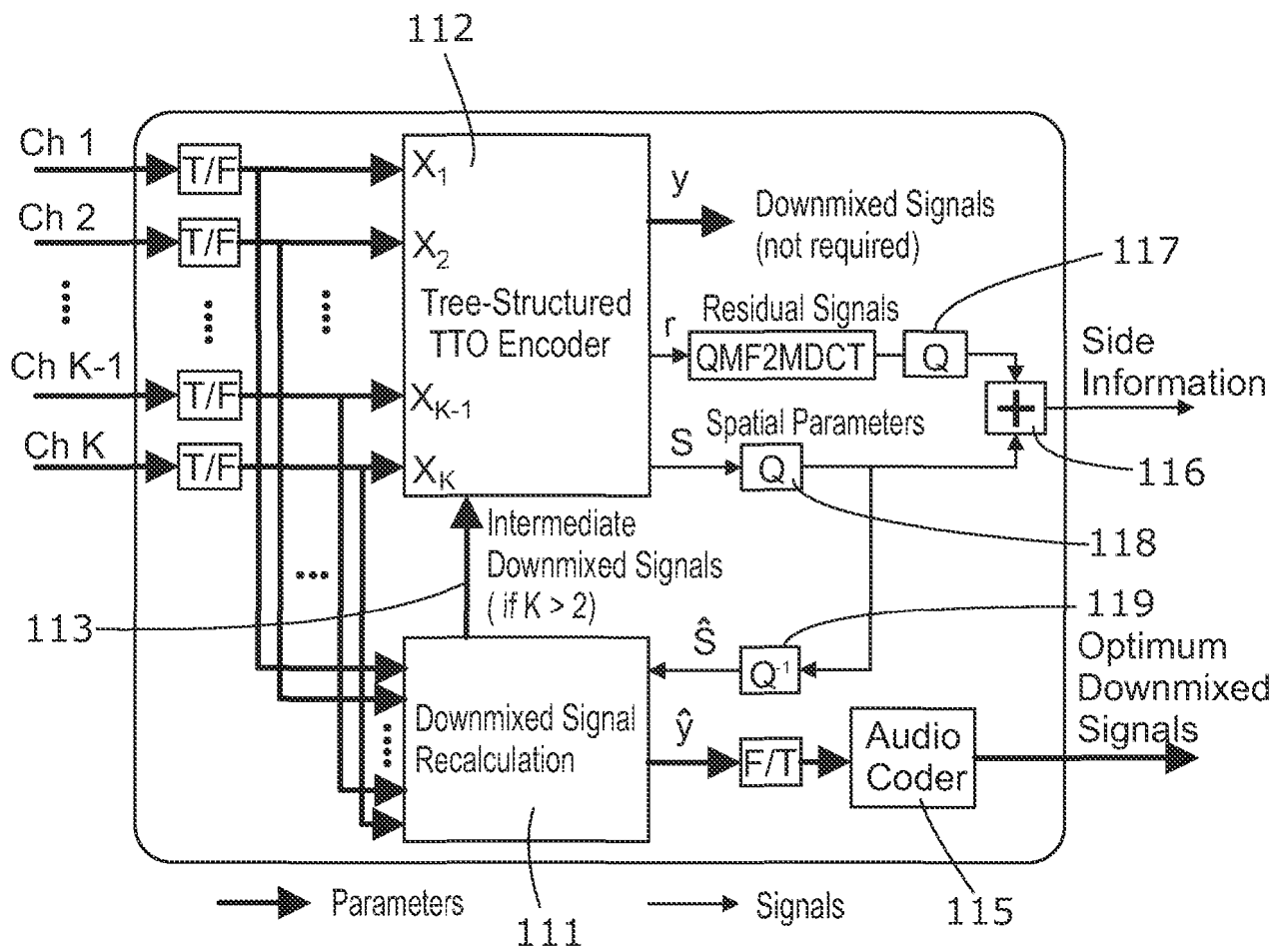
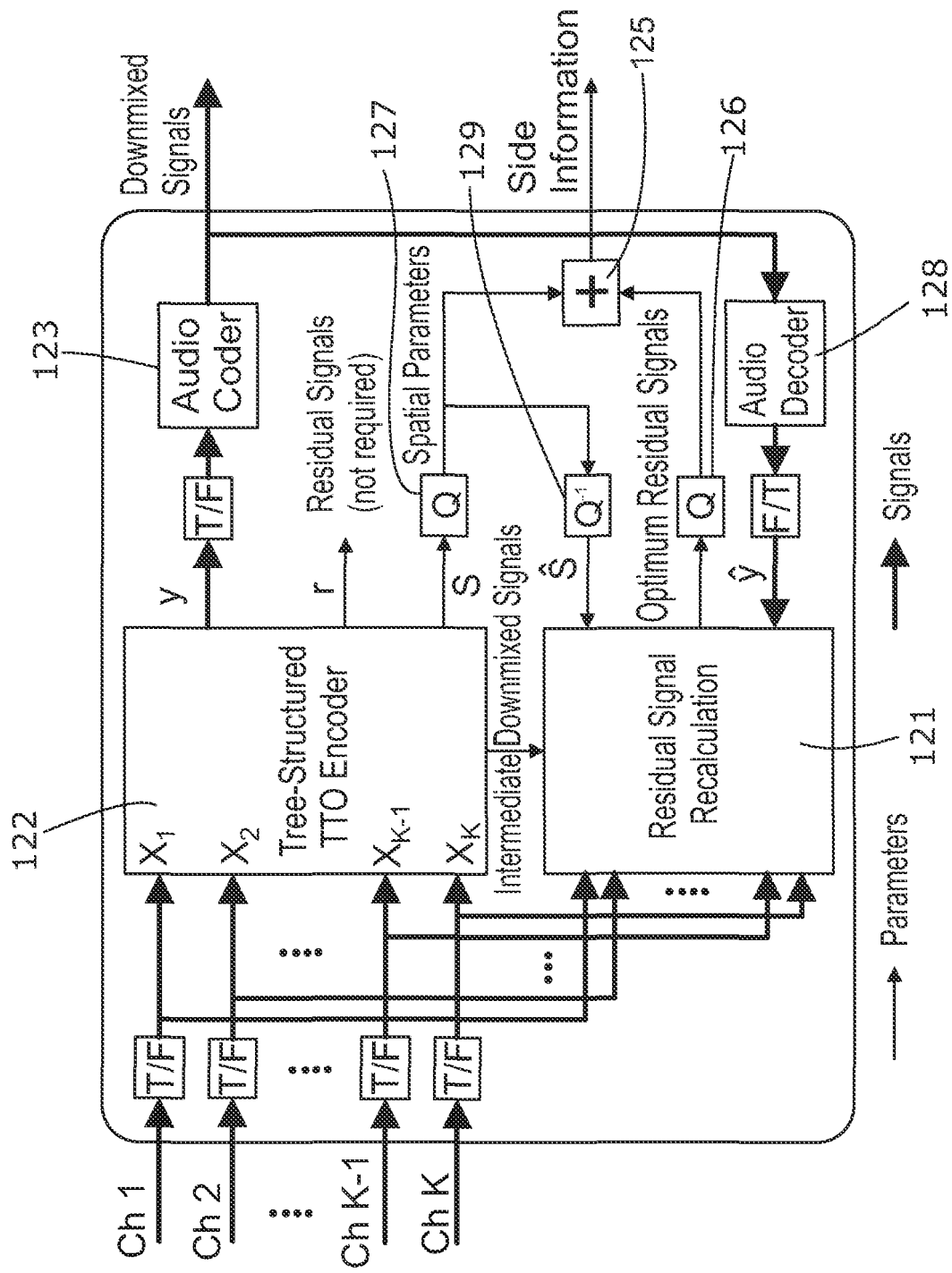


Fig. 11



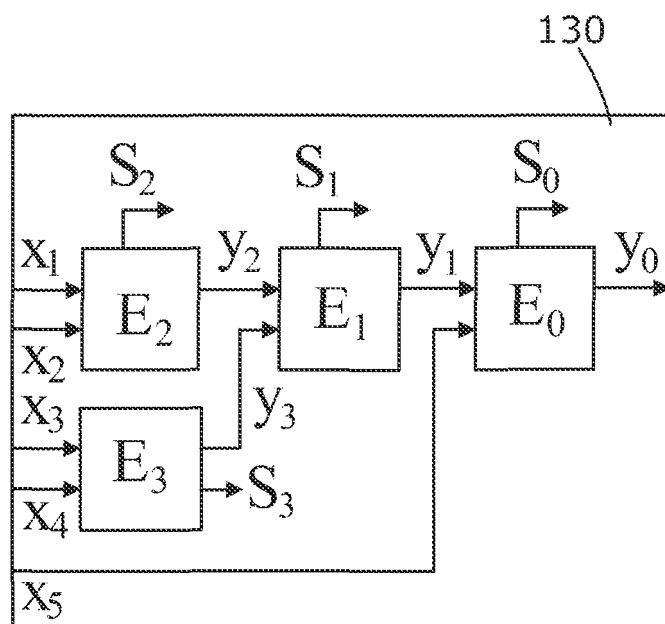


Fig. 13

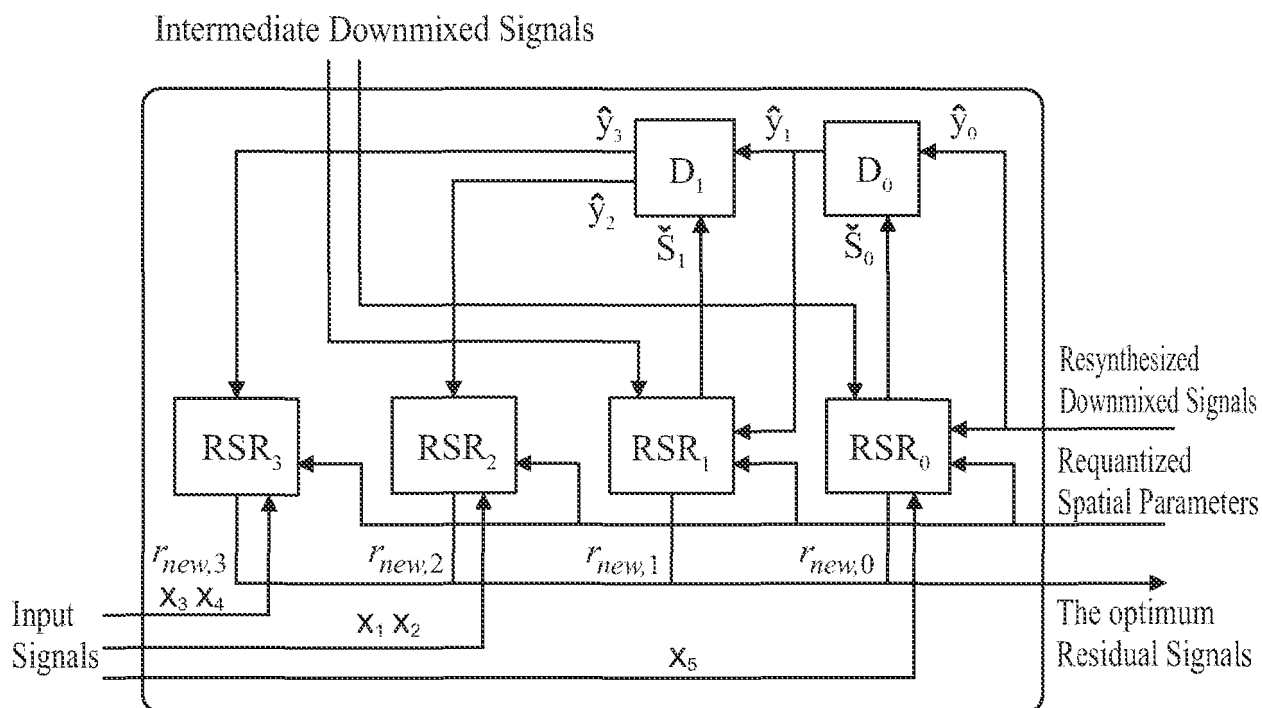


Fig. 14

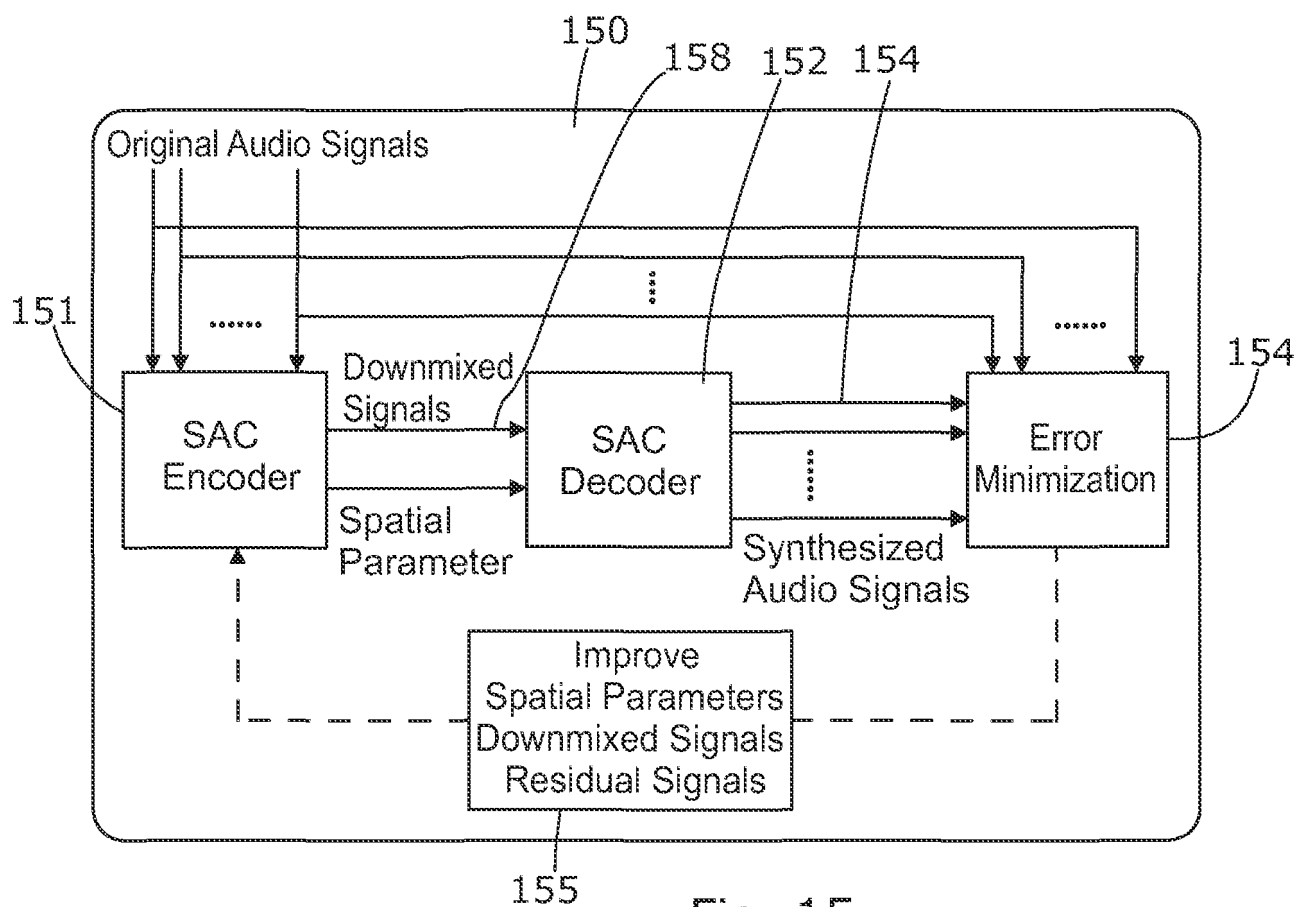


Fig. 15

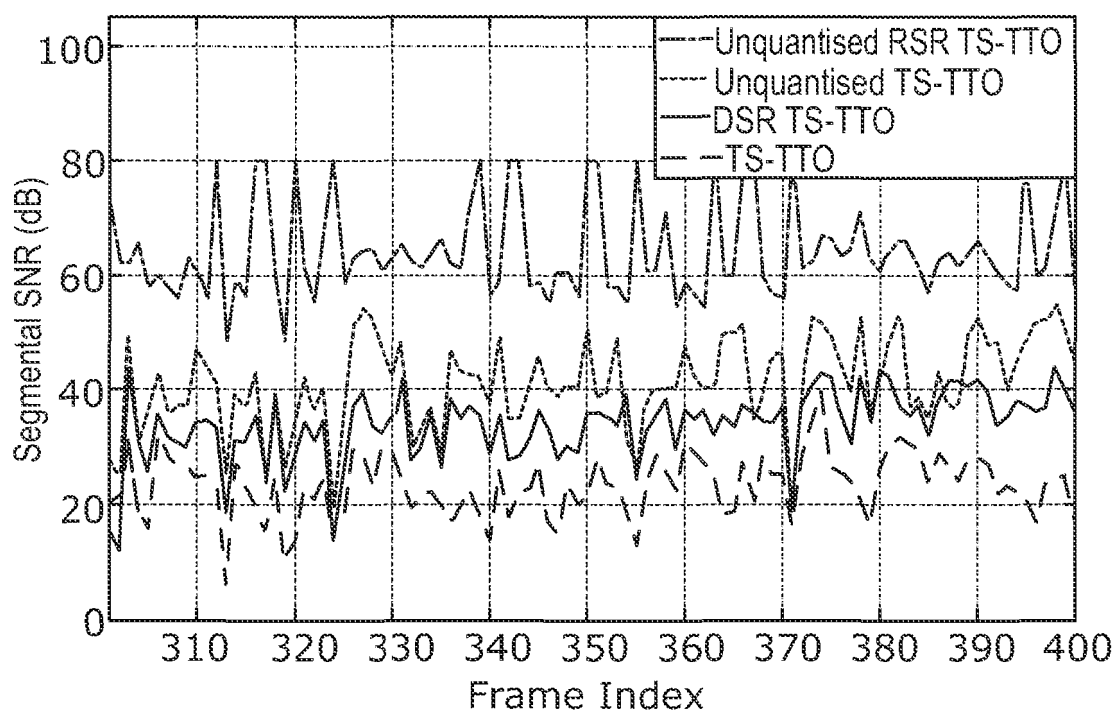


Fig. 16

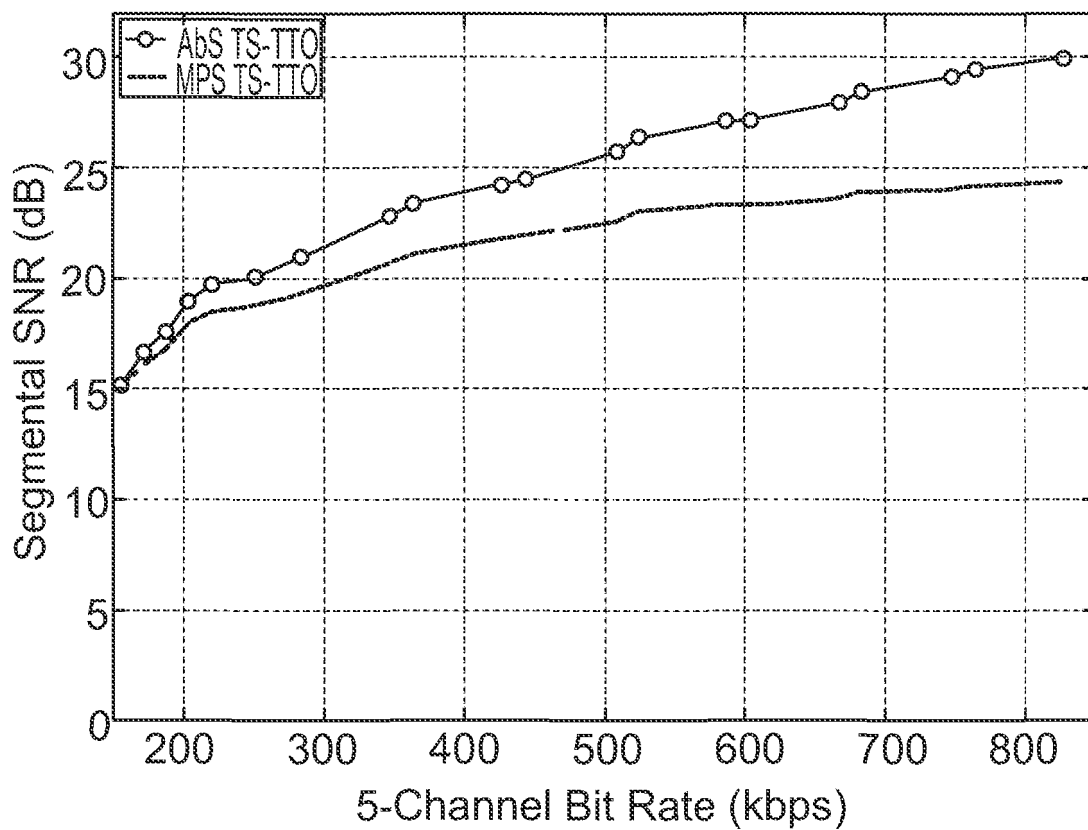


Fig. 17

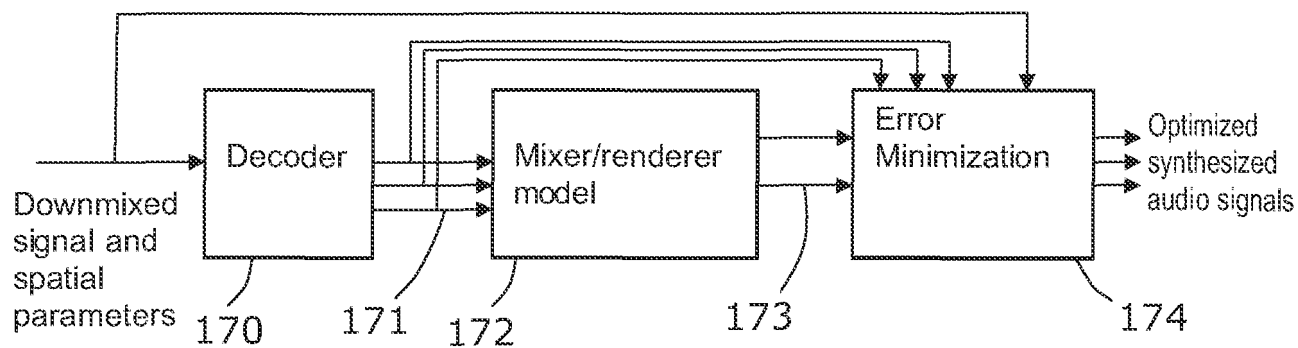


Fig. 18

Spatial Audio Coding

Field of the Invention

The present invention relates to spatial audio coding and has a wide
5 variety of applications such as cinema industry, home entertainment,
digital audio broadcasting (DAB), computer games, music downloading
and streaming services and other internet applications, such as
multichannel teleconferencing.

10 Background to the Invention

Spatial audio relates to the audio capturing and reproduction systems that
convey information about a 3D sound scene. Such a sound scene can
describe, for example, individual sounding objects, their positions, and
the acoustics of the environment in which they are situated. In order to
15 convey this spatial information and improve the realism of perceived 3D
scenes, multichannel audio systems have been increasingly implemented
in many applications. The importance of 3D audio accompanying 3D
video is projected to increase further in the near future with the spread of
3D displays.

20

There are various 3D audio reproduction systems available today, such as
binaural, stereo, 5.1, 7.1 and similar multichannel systems, Ambisonics
and wave field synthesis (WFS). These are based on either fooling the
ears by creating an auditory illusion at the ear location, or creating the
25 exact wavefront of a simulated sound source using secondary sources.
The fidelity of the auditory environments reproduced with these
techniques differs according to the listening position, the acoustics of the
listening environment and the loudspeaker setup. However, major
difficulties lie with the preparation of the content in a way that enables
30 user interaction with, or manipulation of, the sound scene, and the

preservation of the spatial fidelity during both coding for efficient transmission and rendering for different reproduction systems.

Traditionally, it was required that the audio capturing and reproduction techniques would match. For example, stereo material recorded with a stereo microphone technique would be played through a stereo loudspeaker system. Any deviations would cause problems with unmatched number of input and output channels and low quality spatial impression. While up/down mixing techniques alleviated some of these problems, the fidelity of the processed audio material remained unsatisfactory. It was then accepted that clean, unmixed versions of the individual sound sources, or sound objects, should be available for scalable and interactive rendering that is both independent of the reproduction system and flexible in allowing modification of the composition of the original 3D sound scene. While synthetic audio objects are easily available, isolating individual audio objects from a natural scene that contains several sound sources is also possible.

Traditionally, multichannel audio signals have been compressed by using a number of mono audio coders. Although this discrete multichannel audio coding technique was able to reproduce high quality audio signals, it was not preferable since the bandwidth requirements increased linearly with the number of channels. In order to represent multichannel audio signals in fewer bits, spatial audio coding techniques have become popular that are based on finding the appropriate spatial parameters to represent multiple audio signals in a compact form.

Several algorithms have been proposed to encode multiple audio signals, such as parametric stereo (PS), binaural cue coding (BCC), spatial audio scene coding (SASC), directional audio coding (DirAC), spatial squeeze surround audio coding (S^3 AC), MPEG Surround (MPS), and finally the

spatial audio object coding (SAOC), whose standardization is ongoing. All of these approaches have unique advantages and disadvantages.

Coding Multichannel Audio

5

For digital transmission and storage, multichannel audio signals should be represented in a more compact form. This can be done either by exploiting the statistical redundancy within each channel or the interchannel redundancy between the channels. Among the techniques in the first group, Dolby AC-3 and MPEG Surround advanced audio coding (AAC) are the most powerful ones. These apply transform coding as well as exploiting the human auditory system. Each channel of the multichannel audio content can be coded and transmitted by such an audio coder. However, this method is clearly not effective in terms of the amount of bits to be transmitted, since it increases linearly with the number of channels. The techniques in the second group represent multiple channels in fewer channels, usually mono or stereo, which is called downmixing. Matrix surround audio codecs, such as Dolby Surround/Pro Logic and Lexicon Logic 7, aim to increase or decrease the number of channels by up/down conversion to meet the bandwidth requirements and the reproduction system. However, this is usually done without ensuring the preservation of spatial information essential for 3D perception.

25 Spatial audio coding (SAC) reduces the number of channels by exploiting human spatial hearing. In binaural cue coding (BCC), inter-channel level differences (ICLD), inter-channel time differences (ICTD), and inter-channel correlation (ICC) are extracted as spatial parameters. Techniques such as parametric stereo (PS), MPEG Surround (MPS), and spatial audio
30 object coding (SAOC), may also utilize other parameters such as inter-channel predictability and utilize signal processing techniques such as

decorrelators. These spatial parameters are transmitted together with the downmixed signals and the residual signals. At the decoder, multichannel signals are reconstructed and reproduced, ideally creating the same 3D perceptual effect as the original multichannel signals, which is called the perceptual transparency.

The great benefit of these perceptual-based coders is that they can achieve bit rates as low as 3 kbps for transmitting spatial parameters only, as in the case of MPS. This can be achieved because each time-frequency bin is assumed to be occupied by one channel audio signal. However, this assumption causes a major drawback in that it is not capable of reproducing high quality audio for low-correlated signals.

Other techniques also exist that are based on extracting the source location information from multiple channels, thereby simplifying the representation of the spatial scene. These techniques calculate direction vectors representing the directional composition of a scene. At the decoder side, virtual sources are created from the downmixed signal at positions given by the direction vectors. This approach gives the coder the capability to reproduce a different number of output channels from the number of input channels. Examples of this technique are spatial audio scene coding (SASC) and directional audio coding (DirAC). DirAC is slightly different in that it incorporates a recording system using a microphone array. Therefore, the direction vectors in this coder are calculated from the microphone signals.

Squeezing or mapping the auditory space from 360° into 60° has also been proposed for reducing the number of channels to be transmitted. Such a mapping relies on estimating virtual sources, via the inverse amplitude panning technique applied between pairs of input channels. In order to provide a stereo downmixed signal, subsequent panning is

applied into estimated virtual sources. This work, known as Spatial Squeeze Surround Audio Coding (S³AC), is unique in that it basically does not need to transmit additional side information. The major drawback of S³AC is that it introduces the ambiguity of sound positions
5 caused by representing audio channels by a set of virtual sources.

Summary of the Invention

The present invention provides a system for coding multi-channel audio signals for transmission to a decoder. The system may comprise an input
10 arranged to receive audio signals. The system may comprise a trial coding signal generator arranged to output a trial coding signal. The system may comprise a model of the decoder, which may be arranged to receive, as an input, the trial coding signal and to synthesize, as an output, a trial re-synthesized audio signal. The system may comprise optimization means
15 arranged to compare the re-synthesized audio signal with the received audio signal, optionally thereby to determine the form of an optimized coding signal. The system may be arranged to transmit the optimized coding signal.

20 The trial coding signal generator may include an encoder arranged to calculate the trial coding signal from the received audio signal. Alternatively, or in addition the trial coding signal generator may be arranged to retrieve the trial coding signal from data stored in memory.

25 The optimization means may be arranged to calculate the optimized coding signal from the trial coding signal based on its comparison of the re-synthesized audio signal with the received audio signal. Alternatively, or in addition, the optimization means may be arranged to select the optimized coding signal from a plurality of trial coding signals.

The encoder may comprises a plurality of encoder sub-units arranged in a tree structure, and the system is arranged to calculate a modified output for each of the encoder sub-units. At least one of the encoder subunits may be arranged to receive two inputs and output a downmixed signal and
5 at least one residual signal. It may output one residual signal for each input. However at least one of the subunits may be arranged to output only one residual signal for coding both of its inputs. This may be selected from the two residual signals corresponding to the two inputs, or it may be calculated from those two residual signals, for example as an
10 average or weighted average.

The optimization means may comprises a plurality of modification sub-units each arranged to receive a sub-unit output from one of the encoder sub-units and to output a modified sub-unit output.

15

The system may further comprise a quantizer arranged to quantize the coding signal. The optimization means may further comprise an inverse quantizer arranged to re-generate the coding signal from the output of the quantizer.

20

The coding signal may comprise at least one of a downmixed signal, a spatial parameter, and a residual signal. The optimization means may be arranged to optimize the downmixed signal, the spatial parameter, and the residual signal, or any one or more of them.

25

The optimization means may be arranged to optimize the downmixed signal, the spatial parameter, and the residual signal simultaneously and/or in an iterative sequential procedure.

30 The system may further comprise an audio coder arranged to code the downmixed signal. The optimization means may be arranged to receive

the downmixed signal via the audio coder, and the optimization means may further comprise an audio decoder arranged to decode the output of the audio coder.

- 5 The system may further comprise a model of at least part of a transmission channel between the system and the decoder.

The present invention further provides a system for decoding audio signals the system comprising any one or more of the following: a
10 decoder arranged to receive coded audio signals and to re-synthesize the audio signals and output the re-synthesized audio signals to processing means arranged to perform at least one of mixing and rendering of the re-synthesized signals; and optimization means comprising a model of the processing means and arranged to receive the re-synthesized audio signals
15 and the coded audio signals, to process the re-synthesized audio signals and the coded audio signals, and to calculate modified re-synthesized audio signals.

The known systems described can all be classified as open loop systems,
20 including the latest SAOC technique that promises flexible and interactive rendering of audio objects. The major drawback of open-loop systems is that there is error introduced by quantizing spatial parameters and coding of downmixed signals. Hence, when using such open-loop systems, it is more difficult to reach even perceptual lossless quality.

25

In order to minimize quantization errors and improve further the reconstructed audio quality, a closed-loop system can be implemented according to the invention. Some embodiments therefore use an analysis-by-synthesis (AbS) framework, which takes advantage of a close loop
30 system, for enhancing the quality of multichannel audio reproduction. Many methods are suitable to be implemented within this framework.

Transparent quality of reconstructed audio can be achieved since feedback mechanism will naturally attempt to minimize the error resulted from the quantization and the coding processes. As an example, the AbS framework can be applied on the tree-structured two-to-one (TTO) encoder of the MPS standard.

Preferred embodiments of the present invention will now be described by way of example only with reference to the accompanying drawings.

10 **Brief Description of the Drawings**

Figure 1 is a diagram of a multichannel encoder-decoder which forms part of an embodiment of the invention;

15 **Figure 2** is a diagram of a TTO encoder and OTT decoder pair which forms part of an embodiment of the invention;

Figure 3 is a diagram of a tree-structure multi-channel encoder which forms part of an embodiment of the invention;

20

Figure 4 is a diagram of a spatial audio object coding and decoding system which forms part of an embodiment of the invention;

Figure 5 is a diagram of a spatial audio object coding and decoding system with audio object extraction which forms part of an embodiment of the invention;

25

Figure 6 is a graph showing head-related impulse responses in a binaural rendering system forming part of an embodiment of the invention;

30

Figure 7 is a diagram of a two-channel rendering system which forms part of an embodiment of the invention;

Figure 8 is a diagram of a wave field synthesis rendering system which
5 forms part of an embodiment of the invention;

Figure 9 is a functional diagram of an analysis by synthesis system which forms the basis of an embodiment of the invention;

10 **Figure 10** is a diagram of a coding system according to an embodiment of the invention;

Figure 10a is a diagram of a coding system according to a further embodiment of the invention;

15

Figure 11 is a diagram of a coding system according to a further embodiment of the invention;

Figure 12 is a diagram of a coding system according to a further
20 embodiment of the invention;

Figure 13 is a diagram of a multi-channel tree structure encoder according to an embodiment of the invention;

25 **Figure 14** is a diagram of the residual signal recalculation block of a five channel coding system similar to that of Figure 12;

Figure 15 is a diagram of a coding system according to a further embodiment of the invention;

30

Figure 16 is a graph showing segmental signal to noise ratios for various coding systems;

Figure 17 is a graph showing variation of segmental signal to noise ratio for a coding system according to an embodiment of the invention and a conventional coding system; and

Figure 18 is a diagram of a decoding system according to an embodiment of the invention.

10

MPEG Surround Architecture

Referring to Figure 1 a transmitter in the form of a spatial audio encoder 1 comprises an MPEG Surround encoder 2 arranged to receive a multichannel audio input 4, and to output downmixed signals 6 and spatial parameters 8. A further audio encoder 10 is arranged to receive the downmixed audio signals and encode them. The encoded audio signals 12 and the spatial parameters 8 are then input to a multiplexer 14 for transmission to a receiver in the form of a decoder 15. The decoder 15 comprises a de-multiplexer 16 arranged to extract the spatial parameters 8 and the encoded downmixed signals 12 from the received signal, an audio decoder 17 arranged to decode the received downmixed signal, and an MPEG Surround decoder 18 arranged to re-synthesise each channel of the multi-channel input from the downmixed signals and spatial parameters.

MPEG Surround therefore works by extracting the spatial parameters 8 that define the relationship between the different channels of the input signals, and downmixing multiple audio signals into either one channel (mono audio) or dual channel (stereo audio) signals 6. As shown in Figure 1 the downmixed signals are subsequently compressed by the audio encoder 10 and then transmitted accompanied by spatial parameters as side information. MPEG surround (MPS) basically extracts three spatial

parameters: channel level differences (CLD), inter-channel coherences (ICCs), and channel prediction coefficients (CPCs). Any receiver system which cannot handle multichannel audio can simply remove this side information and just render the downmixed signals. This provides the
 5 coder backward compatibility which is important for implementation in various legacy systems. For high quality reproduction, the low frequency component of the residual signal is also transmitted.

There are two pairs of elementary building blocks on MPEG Surround
 10 (MPS) encoder which are two-to-one (TTO) encoder - one-to-two (OTT) decoder pairs and three-to-two (TTT) encoder two-to-three (TTT) decoder pairs. CLDs, ICCs, and the residual signal are extracted from the TTO encoder, whereas CPCs, ICCs and the residual signal are calculated from the TTT encoder. The whole encoding process in the encoder and
 15 decoding process in the decoder is built up by combining several TTOs and TTTs, or OTTs and TTTs in a tree-structure. This tree structure forms the basis of some embodiments of the invention. Details of the other parameters and the other tree-structures which can be used in other embodiments of the present invention are described in [J. Breebaart and
 20 G. Hotho and J. Koppens and E. Schuijers and W. Oomen and S. van de Par. Background, Concepts, and Architecture for the Recent MPEG Surround Standard on Multichannel Audio Compression. *J. Audio Eng. Soc.*, 55:331-351, 2007.].

25 Referring to Figure 2 a two-to-one encoder 20 is arranged to receive two input channels 21, and to output a single downmixed audio output 22 and spatial parameters 23 in the form of CLDs, ICCs, and the residual signal. A two-to-one decoder 25 is arranged to receive a single audio signal input 26 and the spatial parameters 27, and to output two audio signals on
 30 respective channels 28. The encoder 20 therefore converts two input channels 21 into one downmixed output channel 22, and extracts CLD and

ICC as spatial parameters in a parameter band which can be a subband or a group of subbands. A parameter band can consist of one or more subbands which share a common spatial parameter (CLD and ICC). The more the number of subbands in a parameter band, the fewer the spatial parameters to be transmitted. Conversely, the decoder resynthesizes two channels from one channel by utilizing the spatial parameters input to it.

The CLD, denoted as C^b , is defined as the ratio between the energies e of the signals x_1 and x_2 in the first and second channels for a parameter band b :

$$C^b = \frac{e_{x_1}^b}{e_{x_2}^b}, \quad (1)$$

Where the energy is calculated as $e_{x_k}^b = \sum_n x_k^b[n] x_k^{b*}[n]$, $k \in \{1, 2\}$.

For transmission, the logarithmic values of the CLDs are preferred.

The second parameter ICC, denoted as I , is determined by,

$$I^b = \text{Re} \left\{ \frac{\sum_n x_1^b[n] x_2^{b*}[n]}{\sqrt{e_{x_1}^b e_{x_2}^b}} \right\}. \quad (2)$$

This parameter describes the degree of correlation between the input channels x_1 and x_2 .

One set of CLD and ICC parameters are calculated for each parameter band. However, the downmixed signal and the residual signal are calculated for each subband, s .

The downmixed signal $y[n]$ is a weighted sum of the input channels in each subband s ,

$$y^s[n] = \frac{x_1^s[n] + x_2^s[n]}{a^s + b^s}, \quad (3)$$

5

where, the constants a and b represent the energy preservation constraint calculated as,

$$(a^s + b^s)^2 = \frac{e_{x_1}^s + e_{x_2}^s + 2I\sqrt{e_{x_1}^s e_{x_2}^s}}{e_{x_1}^s + e_{x_2}^s}. \quad (4)$$

10 The residual signal $r[n]$ in each subband s is determined from the following decomposition:

$$x_1^s[n] = a^s y^s[n] + r^s[n], \quad (5)$$

$$x_2^s[n] = b^s y^s[n] - r^s[n]. \quad (6)$$

15 The downmixed signal is further coded by the mono audio coder 10. Spatial parameters (CLD, ICC) as well as the residual signal are quantized and then transmitted to the decoder15.

At the decoder side, both audio signals are recreated by estimating a and

20 b as follows:

$$\hat{a} = X \cos(A + B) \quad (7)$$

$$\hat{b} = Y \cos(A - B) \quad (8)$$

where,

$$X = \sqrt{\frac{\hat{C}}{1 + \hat{C}}} \quad (9)$$

$$25 \quad Y = \sqrt{\frac{1}{1 + \hat{C}}} \quad (10)$$

$$A = \frac{1}{2} \arccos(\hat{I}) \quad (11)$$

$$B = \tan \left[- \left(\frac{X-Y}{X+Y} \right) \arctan(A) \right]. \quad (12)$$

Hence, both signals can be reconstructed as

5

$$\hat{x}_1[n] = \hat{a}\hat{y}[n] + \hat{r}[n], \quad (13)$$

$$\hat{x}_2[n] = \hat{b}\hat{y}[n] - \hat{r}[n]. \quad (14)$$

10 The symbols \hat{C} , \hat{I} , and $\hat{r}[n]$ are the quantized-values of the C , I and $r[n]$, respectively, and $\hat{y}[n]$ is the resynthesized downmixed signal. The indexes for the subbands and parameter bands have been ignored for notation simplicity.

15 Referring to Figure 3, the tree-structured encoder 2 comprises a number of TTO encoders 31 and is arranged for downmixing K-channel audio into one channel. The encoder further comprises a set of analysis filter banks (not shown) arranged to receive as inputs the signals on the K channels, and then to break down each of these inputs into frequency subbands (there can be 32, 64 or 71 subbands depending on the type of the implemented filterbank). For each frequency subband, the outputs from the filterbanks for the different input channels are paired together, x_1 and x_2 , x_3 and x_4 etc. Each of these signal pairs is input to a respective one of the TTO encoders 31. The outputs from these encoders are then grouped in pairs and input to further TTO encoders, and this is repeated until the number of audio signals is reduced to one. A set of spatial parameters are extracted in each TTO encoder $E_{p,q}$ where $p \in \{1, 2, \dots, P\}$ and $q \in \{1, 2, \dots, Q\}$ representing TTO indexes with q representing the layer in the tree structure and p indicating the encoder within the layer. For

20

25

simplicity, the structure of TTO encoders and decoders is assumed to be symmetric which means that $K = 2^Q = 2P$. The final downmixed audio output y_{dmix} is input via a synthesis filter bank to the audio encoder 10.

- 5 Using (3), the intermediate downmixed signal $y_{p,q}[n]$ output by any encoder $E_{p,q}$ in the tree structure can be written as

$$y_{p,q}[n] = \frac{y_{2p-1,q+1}[n] + y_{2p,q+1}[n]}{a_{p,q} + b_{p,q}}, \quad (15)$$

where $y_{2p-1,q+1}[n]$ and $y_{2p,q+1}[n]$ are the intermediate downmixed signals.

10

For $q = Q$, the intermediate downmixed signals are calculated from the input signals as

$$y_{p,Q}[n] = \frac{x_{2p-1}[n] + x_{2p}[n]}{a_{p,Q} + b_{p,Q}}. \quad (16)$$

- 15 At the decoder side, the audio signals can be estimated using (13) and (14) as

$$\hat{x}_{k-1}[n] = \hat{a}_{p,Q} \hat{y}_{p,Q} + \hat{r}_{p,Q}, \quad (17)$$

$$\hat{x}_k[n] = \hat{b}_{p,Q} \hat{y}_{p,Q} - \hat{r}_{p,Q}, \quad (18)$$

20

where $p = k/2$ and $\hat{y}_{p,Q}$ is the estimated intermediate downmixed signal. For arbitrary TTO decoder $D_{p,q}$, the estimated intermediate downmixed signal can be represented as

$$25 \quad \hat{y}_{p-1,q}[n] = \hat{a}_{p/2,q-1} \hat{y}_{p/2,q-1} + \hat{r}_{p/2,q-1}, \quad (19)$$

$$\hat{y}_{p,q}[n] = \hat{b}_{p/2,q-1} \hat{y}_{p/2,q-1} - \hat{r}_{p/2,q-1}, \quad (20)$$

where p should be an even number. For $p=2$ and $q=2$, (19) and (20) become

$$5 \quad \hat{y}_{1,2}[n] = \hat{a}_{1,1}\hat{y}_{1,1} + \hat{r}_{1,1}, \quad (21)$$

$$\hat{y}_{2,2}[n] = \hat{b}_{1,1}\hat{y}_{1,1} - \hat{r}_{1,1}, \quad (22)$$

where $\hat{y}_{1,1}$ is the estimated downmixed signal.

10 It will be appreciated that, in the systems of Figures 2 and 3, the system that reproduces or renders the sound scene needs to mirror the recording system that recorded it. 3D audio is essential to create the aural spatial awareness required for high quality presence and immersiveness. Spatial fidelity becomes especially important when 3D audio accompanies the 3D
15 video as discrepancies between the perceived aural and visual environments become more noticeable. There are two main challenges in achieving spatial audiovisual fidelity: perceived quality of spatial audio degrades when the recording system does not match the reproduction system; and immersiveness degrades when aural cues conflict with visual
20 cues. Whether it is a live production or a post production, recomposing a natural scene and rendering its spatial audio in a way that is agnostic of the reproduction system requires access to individual audio objects. This also makes it possible to preserve spatial synchronicity of the captured audiovisual scenes so that perceived positions of sound objects match with
25 the visual ones.

Spatial audio object coding (SAOC) is designed to offer such flexibility. Referring to Figure 4, an SAOC system 40 is similar to an MPEG Surround system, comprising a coder 41 which creates mono or stereo
30 downmixed signals 42 and side information 43 according to the inter

object relationships, such as the object level differences (OLD) and inter-object cross coherence (IOC). An audio encoder 41a is arranged to encode the downmixed signals 42 and a multiplexer 41b is arranged to multiplex the side information 43 and the encoded downmixed signals 42
5 for transmission to the receiver 44. The receiver 44 includes a demultiplexer 45a arranged to extract the side information and the downmixed signals from the received multiplexed signal, and an audio decoder 45b arranged to decode the downmixed signal. A decoder 45 receives the demultiplexed and decoded signals and creates the estimated
10 audio objects 46 from the downmixed signals 42 and the side information 43. The objects 46 can then be remixed according to the user preferences on the scene composition and rendered on a number of channels 48 in a format suited to the loudspeaker setup, using a mixer/renderer 47. Such a system has obvious advantages in broadcasting where the content
15 producers have no way of predicting the loudspeaker systems at the receivers. SAOC also makes it possible for users to remove vocals for karaoke, or remove background music and other unwanted sounds for improving intelligibility of speech.

20 For reduced computational complexity, the decoding and rendering blocks 45, 47 can be combined. This combined block utilizes a SAOC transcoder and a downmixed signal preprocessor. According to the user interaction and rendering system selection, new side information and downmixed signals are produced in MPS format. Therefore, the decoder 45 of the
25 SAOC systems includes an MPS decoder.

The block diagram in Figure 4 assumes that separate audio signals, for each of the separate audio objects that are combined to form the audio scene, are available. Extraction of separate audio signals for separate
30 audio objects can be achieved, for example, as described in WO2009/050487. An example of an audio system having object extraction

capability is shown in Figure 5. A natural scene comprising a number of separate objects 50 is recorded using an array of microphones 51. The signals 52 from the microphones are processed by an object extraction and encoding block 53 which is arranged to extract from them individual audio signals for the individual objects 50 together with scene description data which describes the relative positions of the objects within the scene. The object audio signals and scene description information 54 are transmitted over a transmission channel 59 and then input to a decoding and scene recomposition block 55 which also receives object manipulation data 56 allowing manipulation of the objects within the scene and rendering selection data 57 indicative of the rendering system 58 to be used to render the audio scene. The decoding and scene recomposition block 55 outputs audio signals arranged to regenerate the recorded audio scene on the selected rendering system 58, which may, in various embodiments of the invention, be a stereo, binaural, WFS or 5.1 or other multichannel system, as described in more detail below.

Binaural reproduction is the generation of two signals suitable for listening with headphones and involves the use of head related transfer functions (HRTFs). HRTFs are the direction dependent acoustic filters that model the frequency response of the pinna, head and torso of a listener. HRTFs also subsume the inter-aural time delay (ITD) and inter-aural intensity difference (IID) information, which are the main cues for localisation of sound sources.

25

HRTFs are measured for both the left and right ears of human subjects (for individualized HRTFs) or mannequins (for generic HRTFs). The procedure involves recording an impulsive source with microphones positioned at the ears of a listener or a mannequin. The impulsive source is placed at various directions at a fixed radius from the listener. This gives the time-domain representation of the HRTFs, called head-related

30

impulse responses (HRIRs). Figure 6 shows examples of the left and right ear HRIRs for 60° azimuth and 0° elevation source direction measured on a KEMAR dummy head. For this direction, the sound waves arrive earlier and stronger to the right ear than the left ear creating the ITD and IID as is clearly visible in the Figure 6. Once the HRTF dataset is available, a virtual sound source can be rendered anywhere in 3D space by convolving it with the left and right ear HRTFs of the corresponding direction and playing the resulting signals through headphones.

- 10 For binaural reproduction, the decomposed audio objects are multiplied with the HRTFs in the frequency domain and summed to obtain the left ear and the right ear binaural signals, b_L and b_R , respectively:

$$b_L(\omega, t) = \sum_{i=1}^N a_i p_D(\phi_i, \omega, t) h_L(\phi_i, \omega) + p_A(\omega, t) h_A(\omega, t) \quad (23)$$

15
$$b_R(\omega, t) = \sum_{i=1}^N a_i p_D(\phi_i, \omega, t) h_R(\phi_i, \omega) + p_A(\omega, t) h_A(\omega, t) \quad (24)$$

Where

ω is frequency

t is time

a_i is the amplitude of the i^{th} object, with $0 \leq a_i \leq 1$ and $\sum_i a_i = 1$

- 20 ϕ is the direction of the i^{th} object determined by the user or the scene description information, with $0 < \phi_i \leq 2\pi$

h_L and h_R are the left ear and right ear HRTFs

$h_A(\omega)$ is found by averaging all left and right ear HRTFs.

- p_D is the decomposed audio object at direction ϕ calculated as described in WO2009/050487.

p_A is the ambient sounds which is found by subtracting the sum of all decomposed audio objects from the audio recording of the scene made with an omnidirectional microphone.

This process can also be applied using HRTF filters. However, this usually requires high order filters due to the arbitrary shapes of the HRTFs. In order to decrease the computational complexity, which is essential for rendering binaural audio on mobile devices, HRTFs can be smoothed before filter design, leading to reduced order filters without
 5 loss of perceptual accuracy.

Head movements can also be incorporated in this model if a head tracker is available. When the head rotates α degrees in the horizontal plane
 10 along its axis, the extracted audio object signals in Eqs (23) and (24) are replaced with

$$\check{p}_D(\phi, \omega, t) = p_D(\phi - \alpha, \omega, t). \quad (25)$$

15 Since the processing is done for each time-frequency block, compensation for head movements can easily be included in virtual reality and gaming applications.

Multichannel sound systems are another type of rendering system and
 20 usually have more than two loudspeakers surrounding the listener. The most commonly found multichannel system is the 5.1-channel system, which is mainly an extension of stereo. It uses three channels in the front and two channels at the sides towards the back. There is also another channel (.1) with limited bandwidth for low frequency effects.

25

In order to render extracted audio objects at various directions on a multichannel system, amplitude panning techniques can be utilized, such as the vector based amplitude panning (VBAP). An audio object can be positioned at any point between a pair of loudspeakers by reproducing the
 30 audio object signals at different amplitudes according to the stereophonic law of sines. For a listener facing the front as shown in Figure 7, the

following relationship holds between the front left 72 and front right 74 channel signals and the perceived direction of the audio object.

$$\frac{\sin \theta_v}{\sin \theta} = \frac{g_L - g_R}{g_L + g_R} \quad (26)$$

5

where $0^\circ < \theta < 90^\circ$, $-\theta_v \leq \theta \leq \theta_v$ and g_L and g_R are the left channel and right channel gains, which can be calculated using the additional condition that $g_L^2 + g_R^2 = 1$.

- 10 For reproduction of M audio objects, each audio object is assigned a pair of loudspeakers, so that the object direction is between the loudspeaker pair. Assuming there are K audio objects assigned to a pair, then the channel signals of this pair according to Figure 7 are found as

$$s_1(\omega, t) = \sum_{k=1}^K a_k p_D(\phi_k, \omega, t) g_L(k), \quad (27)$$

15

$$s_2(\omega, t) = \sum_{k=1}^K a_k p_D(\phi_k, \omega, t) g_R(k), \quad (28)$$

where $0 \leq a_k \leq 1$, $\sum_i a_k = 1$ and $0 < \phi_i \leq 2\pi$ are the new amplitude and direction of the k^{th} object determined by the user or the scene description.

20

This weighting and summing is repeated for each pair. However, note that if no pair can be found to yield the vector summation of an audio object, then that object can not be reproduced accurately. Ambient sounds can be reproduced via the loudspeakers at the back or distributed equally
25 between all loudspeakers.

Wave field synthesis (WFS) is a further approach to sound reproduction which aims to create the exact sound field in the environment without any limitation on the listening position. It is based on the Huygen's principle which says that the wavefront of a sound source can be generated with a
 5 distribution of secondary sources along the wavefront. An array of loudspeakers 80 is used as secondary sources as shown in Figure 8.

For WFS reproduction each extracted audio object can be reproduced as a virtual source 82 in the desired location, by generating sound having
 10 wavefronts 84 which simulate those that would be produced from the virtual source 82. The listeners 86 then perceive each audio object to be at the respective virtual source location. However, most commercial WFS systems can only reproduce a certain number of virtual sources due to the computational complexity of calculating filters for each source and for
 15 each loudspeaker. Therefore, if the number of audio objects exceeds the capacity of the system, compound audio objects may be formed before reproduction. Similar to the multichannel reproduction, the ambient sounds are distributed between all loudspeakers. This type of rendering is advantageous in cinemas and other large performance spaces since there
 20 is no sweet-spot for listening.

Analysis-by-synthesis (AbS) technique is a framework which has been used for encoding of speech signals [C. G. Bell and others. Reduction of Speech Spectra by Analysis-by-Synthesis Techniques. *J. Acoust. Soc. Am.*, 33(12):1725-1736, 1961] and determining the excitation signal on an
 25 LPC-based speech coder [B. S. Atal and J. Remde. A New Model of LPC Excitation for Producing Natural-Sounding Speech at Low Bit Rates. *Proc. IEEE Intl. Conf. Acoust. Speech, Signal Process.*, Paris, France, 1982]. Since then, many other speech coders have been proposed using
 30 this framework, such as the code-excited linear prediction (CELP) and regular pulse excitation long term prediction (RPE-LTP) [A. M.

Kondoz. *Digital Speech: Coding for Low Bit Rate Communication Systems* - John Wiley Ltd, 2nd edition, 2004; and P. Kroon and W. B. Kleijn - *Speech Coding and Synthesis*, chapter Linear-prediction based Analysis-by-synthesis coding, pages 79-120. Elsevier, Amsterdam/The

5 Netherlands, 1995]. Some embodiments of the present invention apply the principle of analysis-by-synthesis (AbS) to spatial audio coding systems such as those described above. Referring to Figure 9, in an AbS system 90, a model 91 which is able to synthesize signals from a set of parameters is defined. These parameters are usually variable in order to
10 produce the best matched synthesized signal. The system 90 is arranged to receive as an input an observed signal 92, and includes an error minimization block 93 which is arranged to compare the synthesized signal 94 with the original observed signal 92, and then generate improved model parameters 95 which are selected so as to minimize the
15 error, or difference, between the observed signals and the synthesized signals. These improved parameters are fed back into the system model 91 so as to obtain the optimum synthesized signals. The model 91 is typically a model of at least a part of the system that will be used to synthesize the signals. In an AbS coder according to the invention it can
20 be a model of the decoder which will be used to decode the coded signals. Alternatively, or in addition, it can model the channel over which the coded signal will be transmitted to the decoder. If there are other parts of the system which produce errors in the communication, then these can be included in the model as well.

25

In some embodiments of the invention the AbS framework is implemented on an MPS system. These embodiments have a number of advantages. Firstly, the structure of the SAOC is quite similar to the MPS structure. However, currently, the application of MPS is more widespread than
30 SAOC. Secondly, the decoder of the SAOC also includes an MPS decoder as a subblock.

In order to take full advantage from an AbS framework on MPS, single and multiple optimum parameter can be found in the error minimization mechanism. Details of these configurations are described below.

5 Finding Single Optimum Parameter

Using a single feedback mechanism is the simplest form of the AbS framework. Either the downmixed signal, the spatial parameters or the residual signal can be chosen as the variable parameter to be optimized. The optimum signals or other parameters can then be determined by
 10 minimizing an error criterion, which is usually mean-squared error (mse), between the observed and synthesized signals.

Referring to Figure 10, in an encoder 100 according to one embodiment a SAC encoder 101 corresponding to the encoder 1 of Figure 1 is used to
 15 generate from original audio signals 106 a downmixed signal 108 and spatial parameters 109 and the residual signal if it is provided. These are then input to an SAC decoder 102 which is used as the system model, and the downmixed signals used as the variable model parameters. An error minimisation block 103 receives the synthesized audio signals 104 from
 20 the decoder 102 and compares them with the original audio signals that are also input to the SAC encoder. The error minimization block 103 is arranged to calculate improved downmixed signals which are then fed back to the SAC encoder 101 to replace those it originally output. The encoder 101 is then arranged to transmit the optimized downmixed signal
 25 together with the residuals, CLD and ICC generated by the SAC encoder 101. When these are received by an SAC decoder, which will correspond to the model SAC decoder 102, the SAC decoder will synthesize the audio signals which will be as close as possible to the original audio signals input to the SAC encoder.

In a modification to the embodiment of Figure 10 the error minimization block 103 can be arranged to send back to the SAC encoder just a correction factor giving an indication of how the downmixed signals need to be modified so as to optimized it, and the SAC encoder is then
5 arranged to generate the optimized downmixed signal.

In a further modification where all the spatial parameters and residual signals are optimized as well, the same method is performed in order to find the optimum spatial parameter and the optimum residual signal.

10

Referring to Figure 10a, an encoder 100a according to a further embodiment of the invention comprises a trial signal generator 101a which has memory 105a in which are stored an indexed set of parameters which in this case are a set of trial quantized downmixed signals. The
15 trial signal generator 101a is arranged to output the set of trial quantized downmixed signals 108a from memory. A model of an SAC decoder 102a is arranged to synthesize audio signals 104a from each of the trial quantized downmixed signals. An error minimization block 103a is arranged to receive the trial synthesized audio signals and compare each
20 of them with them with the original audio signal 110a to identify which of the synthesized audio signals 104a is closest to the original. In this way the encoder 100a is arranged to select the optimum quantized downmixed signals from the set of stored possible values in such a way that the difference between synthesized audio signals and the original audio
25 signals is minimized. The encoder 100a can be arranged to transmit the optimized downmixed signal to a decoder, but in this embodiment the encoder 100a is arranged to transmit the index of each sample (or block of samples) of the optimum downmixed signals together with the index of the quantized spatial parameters generated by the SAC encoder. When
30 these are received by an SAC decoder, which will correspond to the model SAC decoder 102a, the SAC decoder is arranged to have the

indexed set of quantized downmixed signals stored in memory and to use the optimum quantized downmixed signal, as identified by the index it receives from the encoder, to synthesize the audio signals which will be as close as possible to the original audio signals input to the SAC encoder
5 100a. The same method can be performed in order to find the optimum spatial parameter and the optimum residual signal. Any encoder may be arranged to optimize any one of these parameters or any two, or all three.

In some systems information is available describing the channel over
10 which the coded signals are to be transmitted. If this is the case, then the model, as well as modelling the decoder, can also include a model of the channel over which the coded signals are to be transmitted. This allows the system to correct, not only for distortion caused by the decoding process, but also for distortion that will arise during transmission over
15 the channel.

For example, in one embodiment of the invention a system like that of Figure 5 is provided, in which the object extraction and coding block 53 includes the encoder 100 of Figure 10. The encoder in a further
20 embodiment also includes a model of the transmission channel 59. The model of the transmission channel may be arranged to be updated to reflect changing conditions on the channel.

In a further embodiment of the invention the tree-structured TTO, as used
25 in MPEG Surround, is implemented within the analysis-by-synthesis framework, to form an analysis by synthesis tree-structured two-to-one encoder (AbS TS-TTO). This has the same structure as the embodiment of Figure 10. The tree-structured TTO decoder is implemented as a model for reconstructing multichannel audio signals so that equations (17) and
30 (18) become the formulas of the model. The purpose of the feedback mechanism is to minimize the errors introduced by the coding processes.

The error signals that are the differences between the original audio signals $x_k[i]$ and the synthesized signals $\hat{x}_k[i]$, where i is index of signal sample in time domain, can be written in vector form as

$$\mathbf{e}_k = \mathbf{x}_k - \hat{\mathbf{x}}_k, \quad (29)$$

- 5 If error minimization is performed in subband domain then the original audio signals $x_k[n]$ and the synthesized audio signals $\hat{x}_k[n]$ where n is index of signal sample in time-frequency domain, are compared after each signal is converted into time-frequency domain by a set of filterbanks.

10 In subband domain, by substituting (17) and (18) into (29), the errors can be represented as

$$\mathbf{e}_{k-1} = \mathbf{x}_{k-1} - \hat{a}_{p,Q} \hat{\mathbf{y}}_{p,Q} - \hat{\mathbf{r}}_{p,Q}, \quad (30)$$

$$\mathbf{e}_k = \mathbf{x}_k - \hat{b}_{p,Q} \hat{\mathbf{y}}_{p,Q} + \hat{\mathbf{r}}_{p,Q}. \quad (31)$$

The mean-squared error for each channel can then be represented as

$$15 \quad \text{mse}_{k-1} = (\mathbf{x}_{k-1} - \hat{a}_{p,Q} \hat{\mathbf{y}}_{p,Q} - \hat{\mathbf{r}}_{p,Q}) \mathbf{e}_{k-1}^T, \quad (32)$$

$$\text{mse}_k = (\mathbf{x}_k - \hat{b}_{p,Q} \hat{\mathbf{y}}_{p,Q} + \hat{\mathbf{r}}_{p,Q}) \mathbf{e}_k^T. \quad (33)$$

where T denotes transpose.

- 20 The minimum mse is obtained by requiring each component of the errors to be orthogonal to the error transpose. These are described as follows.

The input signal \mathbf{x}_k has to be orthogonal to the error's transpose \mathbf{e}^T (i.e.,

$$\mathbf{x}_{k-1} \mathbf{e}_{k-1}^T = 0 \text{ and } \mathbf{x}_k \mathbf{e}_k^T = 0), \text{ hence}$$

25

$$\mathbf{x}_{k-1} \mathbf{e}_{k-1}^T = \mathbf{x}_{k-1} \mathbf{x}_{k-1}^T - \hat{a}_{p,Q} \mathbf{x}_{k-1} \hat{\mathbf{y}}_{p,Q}^T - \mathbf{x}_{k-1} \hat{\mathbf{r}}_{p,Q}^T = 0, \quad (34)$$

$$\mathbf{x}_k \mathbf{e}_k^T = \mathbf{x}_k \mathbf{x}_k^T - \hat{b}_{p,Q} \mathbf{x}_k \hat{\mathbf{y}}_{p,Q}^T + \mathbf{x}_k \hat{\mathbf{r}}_{p,Q}^T = 0. \quad (35)$$

Therefore

$$\hat{\mathbf{r}}_{p,Q}^1 = \mathbf{x}_{k-1} - \hat{a}_{p,Q} \hat{\mathbf{y}}_{p,Q}, \quad (36)$$

$$\hat{\mathbf{r}}_{p,Q}^2 = \hat{b}_{p,Q} \hat{\mathbf{y}}_{p,Q} - \mathbf{x}_k, \quad (37)$$

where each residual signal is indexed assuming that they have different values.

5

The downmixed signals $\hat{\mathbf{y}}_{p,Q}$ has to be orthogonal to the errors transpose \mathbf{e}^T (i.e., $\hat{\mathbf{y}}_{p,Q} \mathbf{e}_{k-1}^T = 0$ and $\hat{\mathbf{y}}_{p,Q} \mathbf{e}_k^T = 0$), therefore

$$\hat{\mathbf{y}}_{p,Q} \mathbf{e}_{k-1}^T = \hat{\mathbf{y}}_{p,Q} \mathbf{x}_{k-1}^T - \hat{a}_{p,Q} \hat{\mathbf{y}}_{p,Q} \hat{\mathbf{y}}_{p,Q}^T - \hat{\mathbf{y}}_{p,Q} \hat{\mathbf{r}}_{p,Q}^T = 0, \quad (38)$$

$$10 \quad \hat{\mathbf{y}}_{p,Q} \mathbf{e}_k^T = \hat{\mathbf{y}}_{p,Q} \mathbf{x}_k^T - \hat{b}_{p,Q} \hat{\mathbf{y}}_{p,Q} \hat{\mathbf{y}}_{p,Q}^T + \hat{\mathbf{y}}_{p,Q} \hat{\mathbf{r}}_{p,Q}^T = 0 \quad (39)$$

These equations can be simplified as

$$\mathbf{x}_{k-1} = \hat{a}_{p,Q} \hat{\mathbf{y}}_{p,Q} + \hat{\mathbf{r}}_{p,Q}^1, \quad (40)$$

$$15 \quad \mathbf{x}_k = \hat{b}_{p,Q} \hat{\mathbf{y}}_{p,Q} - \hat{\mathbf{r}}_{p,Q}^2, \quad (41)$$

where they are actually the same equations as (36) and (37).

The residual signal $\hat{\mathbf{r}}_{p,Q}$ has to be orthogonal to the error \mathbf{e}^T (i.e., $\hat{\mathbf{r}}_{p,Q} \mathbf{e}_{k-1}^T = 0$ and $\hat{\mathbf{r}}_{p,Q} \mathbf{e}_k^T = 0$), hence

20

$$\hat{\mathbf{r}}_{p,Q} \mathbf{e}_{k-1}^T = \hat{\mathbf{r}}_{p,Q} \mathbf{x}_{k-1}^T - \hat{a}_{p,Q} \hat{\mathbf{r}}_{p,Q} \hat{\mathbf{y}}_{p,Q}^T - \hat{\mathbf{r}}_{p,Q} \hat{\mathbf{r}}_{p,Q}^T = 0, \quad (42)$$

$$\hat{\mathbf{r}}_{p,Q} \mathbf{e}_k^T = \hat{\mathbf{r}}_{p,Q} \mathbf{x}_k^T - \hat{b}_{p,Q} \hat{\mathbf{r}}_{p,Q} \hat{\mathbf{y}}_{p,Q}^T + \hat{\mathbf{r}}_{p,Q} \hat{\mathbf{r}}_{p,Q}^T = 0. \quad (43)$$

Therefore

$$\hat{\mathbf{r}}_{p,Q}^1 = \mathbf{x}_{k-1} - \hat{a}_{p,Q} \hat{\mathbf{y}}_{p,Q}, \quad (44)$$

$$25 \quad \hat{\mathbf{r}}_{p,Q}^2 = \hat{b}_{p,Q} \hat{\mathbf{y}}_{p,Q} - \mathbf{x}_k, \quad (45)$$

where the results are also exactly the same as obtained from the first and the second components.

Finding The Optimum Spatial Parameter (Single)

In the system of Figure 10a, based on equations (36) and (37), the optimum spatial parameter can be chosen by a trial and error procedure. The difference between the original and the synthesized audio signals in time domain can be used as criterion. In practice, the equation below can be used as criterion to find the optimum spatial parameter in time-frequency domain instead of reconstructing the audio signals and performing error minimization in time domain :

$$Q_1 = \hat{\mathbf{r}}_{p,Q}^1 - \mathbf{x}_{k-1} + \hat{a}_{p,Q} \hat{\mathbf{y}}_{p,Q},$$

$$Q_2 = \hat{\mathbf{r}}_{p,Q}^2 - \hat{b}_{p,Q} \hat{\mathbf{y}}_{p,Q} + \mathbf{x}_k,$$

In addition, by assuming that both residual signals are equal, the optimum spatial parameter can also be determined in time-frequency domain by,

$$Q_3 = \mathbf{x}_{k-1} + \mathbf{x}_k - (\hat{a}_{p,Q} + \hat{b}_{p,Q}) \hat{\mathbf{y}}_{p,Q}.$$

The optimum spatial parameter is determined by requiring Q_1 , Q_2 and/or Q_3 is minimum.

DSR TS-TTO: Using The Optimum Downmixed Signal (As the single optimized parameter)

In a system similar to that of Figure 10, from (36) and (37), the optimum downmixed signals are obtained by assuming both residual signals are exactly the same. Therefore,

$$\hat{\mathbf{y}}_{p,Q} = \frac{\mathbf{x}_{k-1} + \mathbf{x}_k}{\hat{a}_{p,Q} + \hat{b}_{p,Q}}. \quad (46)$$

By substituting (46) in (30) and (31) and simplifying the expressions, the errors can be represented as

$$\mathbf{e}_{k-1} = -\mathbf{e}_{k-1} = \frac{\hat{a}_{p,Q}(\mathbf{x}_k + \hat{\mathbf{r}}_{p,Q}) - \hat{b}_{p,Q}(\mathbf{x}_{k-1} - \hat{\mathbf{r}}_{p,Q})}{\hat{a}_{p,Q} + \hat{b}_{p,Q}}, \quad (47)$$

where each channel has the same error but in different phase.

In a subband, the samples of the optimum downmixed signals can be written as

$$y_{p,Q}[n] = \frac{x_{k-1}[n] + x_k[n]}{\hat{a}_{p,Q} + \hat{b}_{p,Q}}, \quad (48)$$

and generalized for any p and q as

$$5 \quad y_{p,q}[n] = \frac{y_{2p-1,q+1}[n] + y_{2p,q+1}[n]}{\hat{a}_{p,q} + \hat{b}_{p,q}}, \quad (49)$$

where the tree structure is assumed to be symmetric.

The equations in (48) and (49) can be used in an algorithm for downmixed signal recalculation (DSR) which does not need to use comparison of the system inputs and outputs and error minimization. The downmixed signals are modified based on equation (46) or (48) instead of performed a trial and error procedure to find the optimum downmixed samples. After they are modified, the downmixed signals are coded further by a core coder. The new downmixed signals are then transmitted replacing the downmixed signals from the original TS-TTO encoder.

Referring to Figure 11, for more than two channels ($K > 2$), an encoder comprises a tree structure TTO encoder 112 and a tree structure downmixed signal recalculate block 111. The TS TTO encoder 112 receives the original audio signal from each of the channels and outputs an original downmixed signal y , residual signals r and spatial parameters S . Quantizers 117, 118 are arranged to quantize the residual signals r and spatial parameters S and those quantized signals are combined by an adder 116 and transmitted as side information. The quantized spatial parameters are input via an inverse quantizer 119, which re-generates the spatial parameter, to the DSR block 111. An algorithm in the DSR block 111 performing equation (49) is associated with each TTO encoder in the original TS-TTO 112. The new intermediate downmixed signals 113 generated by these algorithms are fed back into the original TS-TTO 112,

replacing the original intermediate downmixed signal y , and used by the next TTO encoder in the tree structure. The final optimized downmixed signals are output by the DSR block 111 via a core coder 115 for transmission with the side information.

5

RSR TS-TTO: Using The Optimum Residual Signal (As the single optimized parameter)

Referring to Figure 12, for more than two channels ($K > 2$), in one embodiment, an encoder comprises a tree structure TTO encoder 122 and a tree structure residual signal recalculation block 121. The TS TTO encoder 122 receives the original audio signal from each of the channels and outputs an original downmixed signal y , residual signals r and spatial parameters S . A quantizer 127 is arranged to quantize the spatial parameters S . The quantized spatial parameters are input via an inverse quantizer 129, which re-generates the spatial parameter, to the RSR block 121. The downmixed signals y are coded by a core coder 123 which outputs coded downmixed signals for transmission. The coded downmixed signals are also input via a decoder 128 to the RSR block 121. The original audio signals from the input channels as well as the intermediate downmixed signals from the TTO encoder 122 are also input to the RSR block 121. The optimum residual signals given in equation (36) and (37) are calculated in the RSR block 121 from the inputs that block receives.. The final optimized residual signals are output by the RSR block 121 and quantized by a quantizer 126. Those quantized signals are combined by an adder 125 with the quantized spatial parameters and transmitted as side information.

For any single TTO encoder element, if it is assumed that the first and the second residual signals are different, two residual signals can be transmitted. For transmitting only a single optimum residual signal for

30

both inputs, a variety of approximation methods can be performed. An approximation can be done by scaling each residual signal as

$$\mathbf{r}_{p,Q}^{new} = c_1 \hat{\mathbf{r}}_{p,Q}^1 + c_2 \hat{\mathbf{r}}_{p,Q}^2,$$

where $0 \leq c_1 \leq 1$, $0 \leq c_2 \leq 1$ and $c_1 + c_2 = 1$.

- 5 The scale assigning can be done based on the priority of each channel.
In the case of $c_1 = c_2 = 0.5$, the optimum residual signal is a simple averaging as

$$\begin{aligned} \mathbf{r}_{p,Q}^{new} &= \frac{\hat{\mathbf{r}}_{p,Q}^1 + \hat{\mathbf{r}}_{p,Q}^2}{2}, \\ &= \frac{\mathbf{x}_{k-1} - \mathbf{x}_k - (\hat{a}_{p,Q} - \hat{b}_{p,Q})\hat{\mathbf{y}}_{p,Q}}{2}, \end{aligned} \quad (50)$$

- By substituting (50) into (30) and (31), both channels in a TTO encoder
10 will have the equal error as

$$\mathbf{e}_{k-1} = \mathbf{e}_k = \frac{\mathbf{x}_{k-1} + \mathbf{x}_k - (\hat{a}_{p,Q} + \hat{b}_{p,Q})\hat{\mathbf{y}}_{p,Q}}{2}. \quad (51)$$

and consequently the equal mean-squared error ($\text{mse} = \mathbf{e}_{k-1} \mathbf{e}_{k-1}^T = \mathbf{e}_k \mathbf{e}_k^T$).

In time-frequency domain, the new residual signal can be written as

$$r_{p,Q}^{new}[n] = \frac{x_{k-1}[n] - x_k[n] - (\hat{a}_{p,Q} - \hat{b}_{p,Q})\hat{y}_{p,Q}[n]}{2}. \quad (52)$$

- 15 In general, for each p and q , the new residual signal is represented as

$$r_{p,q}^{new}[n] = \frac{y_{2p,q+1}[n] - y_{2p-1,q+1}[n] - (\hat{a}_{p,q} - \hat{b}_{p,q})\hat{y}_{p,q}[n]}{2}. \quad (53)$$

- The equations in (52) and (53) can be used in an algorithm for residual
signal recalculation (RSR). The new residual signal is obtained by
20 performing equations (52) and (52) without any requirement for a trial
and error mechanism. After residual signal recalculation the residual
signal is converted into MDCT domain and then quantized with the same
way as performed in AAC coder.

Another approximation can be used that is by choosing a residual signal from the two residual signals $\hat{\mathbf{r}}_{p,Q}^1$ and $\hat{\mathbf{r}}_{p,Q}^2$. A signal which has maximum energy can be chosen.

5

Five-Channel RSR TS-TTO

Referring to Figure 13, in 5-channel TS-TTO 130, there are four TTO encoders E_0 to E_3 so that four residual signal recalculation algorithms are required and performed in a tree structure. Referring to Figure 14, the RSR function is performed by four RSR blocks R_0 to R_3 , each RSR block R_0 to R_3 is associated with one of the TTO encoders E_0 to E_3 , and is arranged to receive as inputs the inverse quantized spatial parameters S_j from the corresponding TTO encoder E_j and the two audio signals. Each RSR block as can be seen in Figure 14 is arranged to update and output the recalculated residual signal. Using the average residual signal algorithm (52), four new residual signals are calculated as,

$$r_{new,0}[n] = \frac{y_1[n] - x_5[n] + (\hat{b}_0 - \hat{a}_0)\hat{y}_0[n]}{2}, \quad (54)$$

$$r_{new,1}[n] = \frac{y_2[n] - y_3[n] + (\hat{b}_1 - \hat{a}_1)\hat{y}_1[n]}{2}, \quad (55)$$

$$20 \quad r_{new,2}[n] = \frac{x_1[n] - x_2[n] + (\hat{b}_2 - \hat{a}_2)\hat{y}_2[n]}{2}, \quad (56)$$

$$r_{new,3}[n] = \frac{x_3[n] - x_4[n] + (\hat{b}_3 - \hat{a}_3)\hat{y}_3[n]}{2}. \quad (57)$$

Four new residual signals $r_{new,j}$, where $j \in \{0,1,2,3\}$ represents the TTO index, are then output by the RSR blocks to replace the old signals r_j and input to the multiplexer, which multiplexes them with the spatial parameters so that they can be transmitted, with the downmixed signal y_0 .

25

The use of a feedback-type of AbS system in an audio system can enhance the accuracy of sound reproduction. Where a tree structure or other encoding system is formed from a number of encoder sub-units, such as TTO encoders, the use of a plurality of feedback sub-units each associated with one or more of the decoder sub-units can further enhance the sound reproduction.

Finding Multiple Optimum Parameter : Fully AbS TS-TTO

10

In order to take full advantage of the close-loop system, in one embodiment as shown in Figure 15, all signals and parameters (the downmixed signals, the spatial parameters and the residual signals) are optimized at the same time (simultaneously). In this embodiment the basic components are the same as the embodiment of Figure 10 with corresponding parts indicated by the same number increased by 50. In this embodiment the error minimization block is arranged to perform recalculation of the downmixed signals, spatial parameters and residual signals. As the error minimization procedure becomes more complex, in order to reduce the complexity, an iterative sequential procedure can be performed to optimize the different parameters.

In a modification to this embodiment, any one or all of the parameters can be optimized by the trial and error method described above rather than by recalculation. If all of the parameters are optimized by the trial and error method, then the recalculation method is not used at all. If one or more parameters are optimized by the trial and error method and one or more parameters are optimized by the recalculation method, then the system is a hybrid system using both methods.

30

Experimental Results

In order to demonstrate the effectiveness of some embodiments of the invention, experiments were carried out. The results are described and then analysed below. The implemented analysis-by-synthesis system is arranged so as to attempt to reduce the error introduced to the reconstructed audio signals. This means that the embodiments described are trying to make the output waveform or the reproduced sound similar to the input waveform, for example of the originally recorded sound. For this reason an objective performance metric, rather than a subjective one has been found more suitable to evaluate the performance of the embodiments of the invention, particularly for comparing it to the conventional open-loop system. Due to its simplicity signal-to-noise ratio (SNR) measurement is chosen.

In terms of multichannel audio coding, we define SNR as,

$$\text{SNR} = 10 \log_{10} \left[\frac{\sum_n (x_k[n])^2}{\sum_n (x_k[n] - \hat{x}_k[n])^2} \right], \quad (58)$$

where $x_k[n]$ and $\hat{x}_k[n]$ are the original and reconstructed audio signals, respectively for the k^{th} channel and for a time frame length L samples where n is the sample index within the time frame and $n = 0, \dots, L-1$, and $k \in \{1, 2\}$ and $k \in \{1, 2, 3, 4, 5\}$ for the 2-channel and the 5-channel encoders, respectively. SNR is then averaged for all frames, which is called as the segmental SNR, segSNR_k .

The Experiment Setup

A number of low-correlated and high-correlated audio signals sampled at 44100 Hz have been provided for the experiments. For 5-channel input, low-correlated signals consisted of individual audio objects that were

female and male speech, cello, trumpet and percussion music, while high-correlated signals consisted of the Left (L), Right (R), Centre (C), Left surround (Ls), and Right surround (Rs) channels of 5.1 recordings containing panned mixtures of the individual audio objects. For 2-channel
 5 input, low-correlated signals consisted of the female speech and cello while high-correlated signals used only L and R channels.

Each of these signals were fed into a hybrid filter bank decomposing the signal into 71 subbands. Segmentation and overlap-add windowing were
 10 performed in the subband domain. In this experiment, each audio segment consisted of 32 subband samples which is equivalent to the effective length of 1024 time domain samples. In calculating CLDs and ICCs 20 parameter bands were used.

15 For experiments using quantized spatial parameters, CLDs and ICCs, were quantized as in MPS using 5 and 3 bits by non-uniform quantization, respectively as follows

$$\begin{aligned} \text{CLD} &= [150, 45, 40, 35, 30, 25, 22, 19, 16, 13, 10, 8, 6, 4, 2, 0, -2, - \\ 20 \quad &4, -6, -8, -10, -13, -16, -19, -22, -25, -30, -35, -40, -45, -150], \\ \text{ICC} &= [1, 0.937, 0.84118, 0.60092, 0.36764, 0, -0.589, -0.99]. \end{aligned}$$

Furthermore, for coding the downmixed signal AAC was used at the bit rates of 64, 80, 96, 112, 128, 144 and 160 kbps. The residual signals
 25 were coded with the bit rates starting from 16 to 160 kbps. To grade the performance of the close-loop AbS TS-TTO audio coders, an open-loop audio coder (TS-TTO) has also been implemented.

In order to determine how much the close-loop AbS system reduces the
 30 error introduced by the quantization and coding processes, a system transmitting unquantised spatial parameters (denoted as TS-TTO-ContSP)

and another system transmitting unquantised downmixed signal, residual signal and spatial parameters (denoted as unquantised TS-TTO) were included in the analysis.

- 5 The upper bound of the segSNR that can be achieved by the close-loop AbS system can be found by transmitting both unquantised residual signals from each TTO in the subband domain (denoted as unquantised RSR TS-TTO). For this system, the error between the input and output signals in each frame occasionally reaches the zero value. For these
- 10 cases, the segSNR was limited to be 80 dB.

Comparison of SegSNR results for open-loop and close-loop systems

- 15 The experiment in this section is aimed at demonstrating that the DSR algorithm is able to improve the segSNR achieved by the TS-TTO. In this experiment, full waveform reconstruction was performed by transmitting all residual signals from each TTO encoder at 160 kbps.
- 20 Table 1 shows the results for the 2-channel coder for low-correlated and high-correlated inputs. The average segSNR measured on DSR TS-TTO is 32.30 dB, which is 4.41 dB higher than that of the conventional open-loop TS-TTO. It clearly indicates that the DSR algorithm applied on TS-TTO can minimize the error introduced by the quantization process of the
- 25 spatial parameters.

Table 2 shows the results for the 5-channel coder. The overall average segSNR for the DSR TS-TTO is 30.49 dB, which is 6.13 dB higher than that of the conventional open-loop TS-TTO.

Table 1: Average SegSNR (dB) of 2 Channel Audio Coders

Type of Input	Channel	AbS TS-TTO 1	AbS TS-TTO 2	AbS TS- TTO 3
Low-correlated signals	1	24.52	30.92	31.08
	2	33.99	37.23	37.35
Average (for low-correlated signals)		29.26	34.08	34.22
High-correlated signals	L	28.83	31.55	35.66
	R	24.21	29.50	34.18
Average (for high-correlated signals)		26.52	30.52	34.92
Overall average		27.89	32.30	34.57

- 5 In order to evaluate the effect of the AbS system when using TTO in a tree-structure, the segSNR measured for the TS-TTO-ContSP can be observed. The TS-TTO can not be used because the error contributed by the quantization of the spatial parameter and the error contributed by the tree-structure of the TTO can not be distinguished. As can be seen in
- 10 Table 1, when 2-channel audio coder is used, the average segSNR measured for the TS-TTO-ConstSP for low-correlated signals is 34.22 dB which is not much different than 34.92 dB measured for high-correlated signals. However, this is not the case when 5-channel audio coder is used. As shown in Table 2, the segSNR measured for 5-channel TS-TTO-
- 15 ContSP for low-correlated signals is 30.27 dB which is 4.04 dB lower than that for high-correlated signals. This indicates that the tree-structure of TTO used in TS-TTO also introduces error when low-correlated signals are used as input although this is not the case for high-correlated signals. The average segSNRs of DSR TS-TTO and TS-TTO-ContSP for low-

correlated signals in Table 2 show that the DSR algorithm can minimize the error introduced by the spatial parameter quantizer, although it is not able to minimize the error introduced by the use of TTO in a tree-structure.

5

Table 2 also presents that the SNR improvement achieved by the DSR TS-TTO when high-correlated signals are used is 6.79 dB which is 2.79 dB higher than that achieved by the 2-channel DSR TS-TTO for high-correlated signals as in Table 1. These results indicate that the DSR
10 algorithm is able to maintain the segSNR, while the TS-TTO's segSNR decreases when more channels are transmitted.

For a closer look into the SNR improvement, the segSNR for several audio frames for one of the channels of the low-correlated signals has
15 been plotted in Fig. 16. It shows the segSNRs of the conventional TS-TTO and the DSR TS-TTO systems for the same bit rate. For comparison, the SegSNR of the unquantised TS-TTO was shown for which all of the signals and parameters are unquantised. The unquantised RSR TS-TTO was also plotted as the upper bound of the segSNR which
20 can be achieved by the AbS system. This figure shows that significant SNR improvement can be achieved with the DSR algorithm, although there is still a high margin before reaching the upper bound of the AbS system. This is because the DSR TS-TTO is performing an AbS loop where only the spatial quantizer is included. The SNR improvement
25 would be higher if the downmixed and the residual signal coding processes were also included within the loop.

Table 2: Average SegSNR (dB) of 5 Channel Audio Coders

Type of Input	Channel	TS-TTO	DSR TS-TTO	TS-TTO-ContSP
Low-correlated signals	1	21.86	26.47	26.92
	2	21.72	26.21	26.56
	3	27.19	31.72	32.19
	4	28.61	32.05	32.32
	5	22.53	32.83	33.39
Average (for low-correlated signals)		24.38	29.85	30.27
High-correlated signals	L	26.38	32.06	35.31
	R	25.49	31.27	34.62
	C	20.64	28.63	33.12
	Ls	22.25	31.98	34.20
	Rs	26.92	31.70	34.32
Average (for high-correlated signals)		24.34	31.13	34.31
Overall average		24.36	30.49	32.29

5

Performance comparison for various bit rates

In this section, performance of the 5-channel DSR TS-TTO audio coder is evaluated for various bit rates. In this experiment, lower bit-rates were achieved by limiting the bandwidth of the transmitted residual signal. The spatial parameter resolution was kept constant on 20 parameter bands and

10

the time resolution was also fixed at the effective frame length of 1024 samples. At the decoder side, the decorrelator was not performed as it provides a synthetic residual signal particularly when all residual signals are not transmitted from the encoder side.

5

The results are given in Fig.17. The segSNR in dB is plotted against the bit rate in kbps. The result of this experiment shows that the proposed framework outperforms the open-loop system for all bit rates. The highest SNR improvement is achieved when all the residual signals are transmitted. However, it decreases as the bit rate allocated for the residual signal becomes lower. From the test results one can conclude that the proposed tree-structured TTO applied within the analysis-by-synthesis framework is suitable for all bit-rate applications.

15 Referring to Figure 18 according to a further embodiment of the invention a decoding system is arranged to receive downmixed signals and spatial parameters and includes a decoder 170 arranged to re-synthesize audio signals 171 from those inputs. The system further comprises an error minimization module 174 which is arranged to receive the originally
20 received and the re-synthesized downmixed signal and spatial parameters, and the resynthesized audio signals 171. The error minimization module 174 is then arranged to calculate, from these inputs, optimized audio signals that are optimized to compensate for the effects of the decoder and the mixer/renderer, and to output these optimized audio signals for
25 rendering.

It will be appreciated that embodiments of the present invention can include spatial audio coding techniques which provide high-fidelity spatial audio object coding. Capturing, extracting and rendering audio objects
30 can be achieved addressing the whole chain from production to consumption of 3D audio. Systems based on analysis-by-synthesis

techniques can be provided which are applicable to MPEG Surround and SAOC. Comparisons with the conventional open-loop MPEG Surround system using objective metrics indicate that embodiments of the invention can significantly improve the quality of reconstructed audio signals.

Claims

1. A system for coding multi-channel audio signals for transmission to a decoder, the system comprising: an input arranged to receive an audio signal; a trial coding signal generator arranged to output a trial coding signal; a model of the decoder arranged to receive, as an input, the trial coding signal and to synthesize, as an output, a trial re-synthesized audio signal; and optimization means arranged to compare the re-synthesized audio signal with the received audio signal thereby to determine the form of an optimized coding signal, and to transmit the optimized coding signal.
2. A system according to claim 1 wherein the trial coding signal generator includes an encoder arranged to calculate the trial coding signal from the received audio signal.
3. A system according to claim 1 wherein the trial coding signal generator is arranged to retrieve the trial coding signal from memory.
4. A system according any foregoing claim wherein the optimization means is arranged to calculate the optimized coding signal from the trial coding signal based on its comparison the re-synthesized audio signal with the received audio signal.
5. A system according to claim 3 wherein the optimization means is arranged to select the optimized coding signal from a plurality of trial coding signals.
6. A system according to claim 2 wherein the encoder comprises a plurality of encoder sub-units arranged in a tree structure, and the system is arranged to calculate a modified output for each of the encoder sub-units.

7. A system according to claim 6 wherein at least one of the encoder subunits is arranged to receive two inputs and output a downmixed signal and at least one residual signal.

5

8. A system according to claim 7 wherein said at least one of the subunits is arranged to output only one residual signal for coding both inputs.

10 9. A system according to any of claims 6 to 8 wherein the optimization means comprises a plurality of modification sub-units each arranged to receive a sub-unit output from one of the encoder sub-units and to output a modified sub-unit output.

15 10. A system according to claim 2 further comprising a quantizer arranged to quantize the coding signal, wherein the optimization means further comprises an inverse quantizer arranged to re-generate the coding signal from the output of the quantizer.

20 11. A system according to any foregoing claim wherein the coding signal comprises at least one of a downmixed signal, a spatial parameter, and a residual signal.

12. A system according to claim 3 wherein the optimization means is
25 arranged to find the optimum spatial parameter by minimization of both of the formulas:

$$Q_1 = \hat{\mathbf{r}}_{p,Q}^1 - \mathbf{x}_{k-1} + \hat{a}_{p,Q} \hat{\mathbf{y}}_{p,Q},$$

$$Q_2 = \hat{\mathbf{r}}_{p,Q}^2 - \hat{b}_{p,Q} \hat{\mathbf{y}}_{p,Q} + \mathbf{x}_k,$$

or, of the formula:

30
$$Q_3 = \mathbf{x}_{k-1} + \mathbf{x}_k - (\hat{a}_{p,Q} + \hat{b}_{p,Q}) \hat{\mathbf{y}}_{p,Q}.$$

where

$\hat{\mathbf{r}}_{p,Q}^1$ and $\hat{\mathbf{r}}_{p,Q}^2$ is the first and the second new/modified residual signal from one of the system sub-units

\mathbf{x}_{k-1} and \mathbf{x}_k is the input signals in the corresponding TTO encoder

5 $\hat{\mathbf{y}}_{p,Q}$ is the resynthesized output (intermediate downmixed signal) in the corresponding TTO encoder

$\hat{a}_{p,q}$ is the estimated constant at the decoder side specific to the sub-unit p,q

$\hat{b}_{p,q}$ is the estimated constant at the decoder specific to the sub-unit p,q.

10

13. A system according to claim 4 wherein the optimization means is arranged to calculate the modified downmixed signal using a formula of the form:

$$y_{p,q}[n] = \frac{y_{2p-1,q+1}[n] + y_{2p,q+1}[n]}{\hat{a}_{p,q} + \hat{b}_{p,q}},$$

15 Where

$y_{p,q}[n]$ is the downmixed signal from one of the system sub-units

n is the index of the time-frequency (subband) sample

q represents the layer in the tree structure,

p represents the encoder sub-unit in the layer,

20 $y_{2p,q+1}[n]$ and $y_{2p,q+1}[n]$ are the input signals in the corresponding TTO encoder

$\hat{a}_{p,q}$ is a constant specific to the sub-unit p,q

$\hat{b}_{p,q}$ is a constant specific to the sub-unit p,q

25 14. A system according to claim 4 wherein the coding signal is a residual signal and the optimization means is arranged to calculate the modified residual signal using a formula of the form:

$$\mathbf{r}_{p,Q}^{new} = c_1 \hat{\mathbf{r}}_{p,Q}^1 + c_2 \hat{\mathbf{r}}_{p,Q}^2,$$

where $0 \leq c_1 \leq 1$, $0 \leq c_2 \leq 1$ and $c_1 + c_2 = 1$. The new modified residual signals $\hat{\mathbf{r}}_{p,q}^1$ and $\hat{\mathbf{r}}_{p,q}^2$ are as defined in claim 7.

In case of $c_1 + c_2 = 0.5$, this formula is applied:

$$r_{p,q}^{new}[n] = \frac{y_{2p,q+1}[n] - y_{2p-1,q+1}[n] - (\hat{a}_{p,q} - \hat{b}_{p,q})\hat{y}_{p,q}[n]}{2}$$

5 Where

$r_{p,q}^{new}[n]$ is the modified residual signal for each sub-unit, with p, q indicating the sub-unit within the tree structure

n is the index of the time frame

q represents the layer in the tree structure,

10 p represents the encoder sub-unit in the layer,

$y_{2p,q+1}[n]$ and $y_{2p-1,q+1}[n]$ are the input signals in the corresponding TTO encoder

$\hat{a}_{p,q}$ is a constant specific to the sub-unit p,q

$\hat{b}_{p,q}$ is a constant specific to the sub-unit p,q

15 $y_{p,q}[n]$ is the downmixed signal from one of the system sub-units

15. A system according to claim 14 wherein the optimization means is arranged to optimize the downmixed signal, the spatial parameter, and the residual signal.

20

16. A system according to claim 15 wherein the optimization means is arranged to optimize the downmixed signal, the spatial parameter, and the residual signal simultaneously and/or in an iterative sequential procedure.

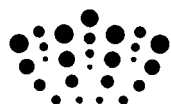
25 17. A system according to any of claims 9, 15 or 16 further comprising an audio coder arranged to code the downmixed signal, wherein the optimization means is arranged to receive the downmixed signal via the audio coder, and the optimization means further comprises an audio decoder arranged to decode the output of the audio coder.

18. A system according to any foregoing claim further comprising a model of at least part of a transmission channel between the system and the decoder.

5

19. A system for decoding audio signals the system comprising: a decoder arranged to receive coded audio signals and to re-synthesize the audio signals and output the re-synthesized audio signals to processing means arranged to perform at least one of mixing and rendering of the re-synthesized signals; and optimization means comprising a model of the processing means and arranged to receive the re-synthesized audio signals and the coded audio signals, to process the re-synthesized audio signals and the coded audio signals, and to calculate modified re-synthesized audio signals.

15



Application No: GB1020087.1

Examiner: Mrs Hannah Sylvester

Claims searched: 1-19

Date of search: 28 March 2011

Patents Act 1977: Search Report under Section 17

Documents considered to be relevant:

Category	Relevant to claims	Identity of document and passage or figure of particular relevance
A	-	WO2008/078973 A1 (BEACK SEUNG-KWON [KR] et al)
A	-	US2002/090113 A1 (RHOADS GEOFFREY B [US])
A	-	WO2008/039039 A1 (YOON SUNG YONG [KR] et al)

Categories:

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.

Field of Search:

Search of GB, EP, WO & US patent documents classified in the following areas of the UKC^X :

Worldwide search of patent documents classified in the following areas of the IPC

G10L

The following online and other databases have been used in the preparation of this search report

WPI EPODOC TXTE

International Classification:

Subclass	Subgroup	Valid From
None		