



(51) International Patent Classification:

H04R 1/26 (2006.01) G06F 3/01 (2006.01)
H04R 5/02 (2006.01) G10L 21/0272 (2013.01)
H04S 7/00 (2006.01) G10L 21/10 (2013.01)

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))

(21) International Application Number:

PCT/FI2018/050289

(22) International Filing Date:

24 April 2018 (24.04.2018)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

1706474.2 24 April 2017 (24.04.2017) GB

(71) Applicant: NOKIA TECHNOLOGIES OY [FI/FI];
Karaportti 3, 02610 Espoo (FI).

(72) Inventors: LEPPÄNEN, Jussi; Mielikinkatu 26, 33580
Tampere (FI). ERONEN, Antti; Rintamäenkatu 13 A 6,
33820 Tampere (FI). PIHLAJAKUJA, Tapani; Heide-
hofinpolku 4 H 16, 01300 Vantaa (FI). LEHTINIEMI, Ar-
to; Lomarantatie 17, 33880 Lempäälä (FI).

(74) Agent: NOKIA TECHNOLOGIES OY et al.; Ari Aarnio,
IPR Department, Karakaari 7, 02610 Espoo (FI).

(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, DJ, DK, DM, DO,
DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,
HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP,
KR, KW, KZ, LA, LC, LK, LR, LS, 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, SA,
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, ST, 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,
KM, ML, MR, NE, SN, TD, TG).

(54) Title: SPATIAL AUDIO PROCESSING

(57) Abstract: A method comprising: allocating frequency sub-channels of an input audio signal to multiple spatial audio channels, each spatial audio channel for rendering at a location within a sound space; and responding to user input to cause a change in an allocation of frequency sub-channels of the input audio signal to multiple spatial audio channels; automatically changing the allocation of frequency sub-channels of the input audio signal to at least a one or more of the multiple spatial audio channels selected by the user input.



TITLE

Spatial audio processing.

TECHNOLOGICAL FIELD

- 5 Embodiments of the present invention relate to spatial audio processing. In particular, embodiments relate to providing a sound object with spatial extent.

BACKGROUND

- 10 Audio content may or may not be a part of other content. For example, multimedia content comprises a visual content and an audio content. The visual content and/or the audio content may be perceived live or they may be recorded and rendered.

- 15 For example, in an augmented reality application, at least part of the visual content is observed by a user via a see-through display while another part of the visual content is displayed on the see-through display. The audio content may be live or it may be rendered to a user.

In a virtual reality application, the visual content and the audio content are both rendered.

- 20 It may in some circumstances be desirable to control how a user perceives audio content.

BRIEF SUMMARY

- 25 According to various, but not necessarily all, embodiments of the invention there is provided a method comprising: allocating frequency sub-channels of an input audio signal to multiple spatial audio channels, each spatial audio channel for rendering at a location within a sound space; and responding to user input to cause a change in an allocation of frequency sub-channels of the input audio signal to multiple spatial audio channels; automatically changing the allocation of frequency sub-channels of the input audio signal to at least a one or more of the multiple spatial audio channels selected by the user input.

- 30 According to various, but not necessarily all, embodiments of the invention there is provided an apparatus comprising: means for allocating frequency sub-channels of an input audio signal to multiple spatial audio channels, each spatial audio channel for rendering at a location within a sound space; and
35 means for responding to user input to cause a change in an allocation of frequency sub-channels of the input audio signal to multiple spatial audio channels;

means for automatically changing the allocation of frequency sub-channels of the input audio signal to at least a one or more of the multiple spatial audio channels selected by the user input.

- 5 According to various, but not necessarily all, embodiments of the invention there is provided a computer program that when run on a processor enables: allocating frequency sub-channels of an input audio signal to multiple spatial audio channels, each spatial audio channel for rendering at a location within a sound space; and
- 10 responding to user input to cause a change in an allocation of frequency sub-channels of the input audio signal to multiple spatial audio channels; automatically changing the allocation of frequency sub-channels of the input audio signal to at least a one or more of the multiple spatial audio channels selected by the user input.

- 15 According to various, but not necessarily all, embodiments of the invention there is provided an apparatus comprising: at least one processor; and at least one memory including computer program code that the at least one memory and the computer program code configured to, with the at least one processor, cause the apparatus at least to perform: allocating frequency sub-channels of an input audio signal to multiple spatial audio channels, each spatial audio channel for rendering at a location within a sound space; responding to user input to cause a change
- 20 in an allocation of frequency sub-channels of the input audio signal to multiple spatial audio channels; automatically changing the allocation of frequency sub-channels of the input audio signal to at least a one or more of the multiple spatial audio channels selected by the user input.

- 25 According to various, but not necessarily all, embodiments of the invention there is provided examples as claimed in the appended claims.

BRIEF DESCRIPTION

- 30 For a better understanding of various examples that are useful for understanding the detailed description, reference will now be made by way of example only to the accompanying drawings in which:

- Figs 1A to 1D illustrates examples of a sound space comprising one or more sound objects; Figs. 2A to 2D illustrate examples of a recorded visual scene that respectively correspond with the sound space illustrated in Figs 1A to 1D;
- 35 Fig 3A illustrates an example of a controller and Fig 3B illustrates an example of a computer program;

Fig 4 illustrates an example of a spatial audio processing system comprising a spectral allocation module and a spatial allocation module;

Fig 5 illustrates an example of a method;

5 Figs 6A and 6B illustrate a visual scene presented via a user interface before (Fig 6A) and after (Fig 6B) re-allocation of frequency sub-channels of an input audio signal to one or more of multiple spatial audio channels selected by a user input;

Figs 7A and 7B illustrate an apparatus for providing the user interface of Figs 6A and 6B, respectively, using mediated reality;

10 Fig 8 illustrates an example of a system that is configured to perform an example of the method- allocating frequency sub-channels of the input audio signal to one or more of the multiple spatial audio channels in dependence upon a user input that selects the one or more spatial audio channels; and

15 Fig 9 illustrates an example of a method for controlling rendering of spatial audio and in particular controlling better perceptual rendering of a sound object that has a spatial extent, for example width.

DEFINITIONS

“artificial environment” may be something that has been recorded or generated.

20 “visual space” refers to fully or partially artificial environment that may be viewed, which may be three dimensional.

“visual scene” refers to a representation of the visual space viewed from a particular point of view within the visual space.

‘visual object’ is a visible object within a virtual visual scene.

25 “sound space” refers to an arrangement of sound sources in a three-dimensional space. A sound space may be defined in relation to recording sounds (a recorded sound space) and in relation to rendering sounds (a rendered sound space).

“sound scene” refers to a representation of the sound space listened to from a particular point of view within the sound space.

30 “sound object” refers to sound source that may be located within the sound space. A source sound object represents a sound source within the sound space. A recorded sound object represents sounds recorded at a particular microphone or position. A rendered sound object represents sounds rendered from a particular position.

35 “virtual space” may mean a visual space, a sound space or a combination of a visual space and corresponding sound space. In some examples, the virtual space may extend horizontally up to 360° and may extend vertically up to 180°

“virtual scene” may mean a visual scene, mean a sound scene or mean a combination of a visual scene and corresponding sound scene.

‘virtual object’ is an object within a virtual scene, it may be an artificial virtual object (e.g. a computer-generated virtual object) or it may be an image of a real object in a real space that is live or recorded. It may be a sound object and/or a visual object.

“Correspondence” or “corresponding” when used in relation to a sound space and a visual space means that the sound space and visual space are time and space aligned, that is they are the same space at the same time.

“Correspondence” or “corresponding” when used in relation to a sound scene and a visual scene means that the sound space and visual scene are corresponding and a notional listener whose point of view defines the sound scene and a notional viewer whose point of view defines the visual scene are at the same position and orientation, that is they have the same point of view.

“real space” refers to a real environment, which may be three dimensional. “Real visual scene” refers to a representation of the real space viewed from a particular point of view within the real space. ‘Real visual object’ is a visible object within a real visual scene.

The “visual space”, “visual scene” and visual object” may also be referred to as the “virtual visual space”, “virtual visual scene” and “virtual visual object” to clearly differentiate them from “real visual space”, “real visual scene” and “real visual object”

“mediated reality” in this document refers to a user visually experiencing a fully or partially artificial environment (a virtual space) as a virtual scene at least partially rendered by an apparatus to a user. The virtual scene is determined by a point of view within the virtual space. “augmented reality” in this document refers to a form of mediated reality in which a user experiences a partially artificial environment (a virtual space) as a virtual scene comprising a real scene of a physical real world environment (real space) supplemented by one or more visual or audio elements rendered by an apparatus to a user.

“virtual reality” in this document refers to a form of mediated reality in which a user experiences a fully artificial environment (a virtual visual space) as a virtual scene displayed by an apparatus to a user.

“perspective-mediated” as applied to mediated reality, augmented reality or virtual reality means that user actions determine the point of view within the virtual space, changing the virtual scene.

“first person perspective-mediated” as applied to mediated reality, augmented reality or virtual reality means perspective mediated with the additional constraint that the user’s real point of view determines the point of view within the virtual space.

“third person perspective-mediated” as applied to mediated reality, augmented reality or virtual reality means perspective mediated with the additional constraint that the user’s real point of view does not determine the point of view within the virtual space.

5 “user interactive” as applied to mediated reality, augmented reality or virtual reality means that user actions at least partially determine what happens within the virtual space.

“displaying” means providing in a form that is perceived visually (viewed) by the user.

“rendering” means providing in a form that is perceived by the user.

DETAILED DESCRIPTION

10 The following description describes methods, apparatuses and computer programs that control how audio content is perceived and, in particular, control a perceived size and/or position of a source of the audio content. In some, but not necessarily all examples, spatial audio rendering may be used to render sound sources as sound objects at particular positions within a sound space.

15 Automatic or user controlled editing of a sound space may occur by, for example, repositioning one or more sound objects or by changing sound characteristics of the sound objects such as a perceived lateral and/or vertical extent of the sound source.

20 Fig 1A illustrates an example of a sound space 10 comprising a sound object 12 within the sound space 10. The sound object 12 may be a sound object as recorded or it may be a sound object as rendered. It is possible, for example using spatial audio processing, to modify a sound object 12, for example to change its sound or positional characteristics. For example, a sound object can be modified to have a greater volume, to change its position within the
25 sound space 10 (Figs 1B & 1C) and/or to change its spatial extent within the sound space 10 (Fig 1D)

Fig 1B illustrates the sound space 10 before movement of the sound object 12 in the sound space 10. Fig 1C illustrates the same sound space 10 after movement of the sound object 12.

30 The sound object 12 may be a sound object as recorded and be positioned at the same position as a sound source of the sound object or it may be positioned independently of the sound source.

35 The position of a sound source may be tracked to render the sound object at the position of the sound source. This may be achieved, for example, when recording by placing a positioning

tag on the sound source. The position and the position changes of the sound source can then be recorded. The positions of the sound source may then be used to control a position of the sound object 12. This may be particularly suitable where an up-close microphone such as a boom microphone or a Lavalier microphone is used to record the sound source.

5

In other examples, the position of the sound source within the visual scene may be determined during recording of the sound source by using spatially diverse sound recording. An example of spatially diverse sound recording is using a microphone array. The phase differences between the sound recorded at the different, spatially diverse microphones, provides information that may be used to position the sound source using a beam forming equation. For example, time-difference-of-arrival (TDOA) based methods for sound source localization may be used.

10

15

The positions of the sound source may also be determined by post-production annotation. As another example, positions of sound sources may be determined using Bluetooth-based indoor positioning techniques, or visual analysis techniques, a radar, or any suitable automatic position tracking mechanism.

20

Fig 1D illustrates a sound space 10 after extension of the sound object 12 in the sound space 10. The sound space 10 of Fig. 1D differs from the sound space 10 of Fig. 1C in that the spatial extent of the sound object 12 has been increased so that the sound object has a greater breadth (greater width).

25

In some examples, a visual scene 20 may be rendered to a user that corresponds with the rendered sound space 10. The visual scene 20 may be the scene recorded at the same time the sound source that creates the sound object 12 is recorded.

30

Fig. 2A illustrates an example of a visual scene 20 that corresponds with the sound space 10. Correspondence in this sense means that there is a one-to-one mapping between the sound space 10 and the visual scene 20 such that a position in the sound space 10 has a corresponding position in the visual scene 20 and a position in the visual scene 20 has a corresponding position in the sound space 10. Corresponding also means that the coordinate system of the sound space 10 and the coordinate system of the visual scene 20 are in register such that an object is positioned as a sound object in the sound space and as a visual object in the visual scene at the same common position from the perspective of a user.

35

The sound space 10 and the visual scene 20 may be three-dimensional.

A portion of the visual scene 20 is associated with a position of visual content representing a sound source 22 within the visual scene 20. The position of the sound source 22 in the visual scene 20 corresponds with a position of the sound object 12 within the sound space 10.

In this example, but not necessarily all examples, the sound source 22 is an active sound source producing sound that is or can be heard by a user, for example via rendering or live, while the user is viewing the visual scene via the display 200.

In some examples, parts of the visual scene 20 are viewed through the display 200 (which would then need to be a see-through display). In other example, the visual scene 20 is rendered by the display 200.

In an augmented reality application, the display 200 is a see-through display and at least parts of the visual scene 20 is a real, live scene viewed through the see-through display 200. The sound source 22 may be a live sound source or it may be a sound source that is rendered to the user. This augmented reality implementation may, for example, be used for capturing an image or images of the visual scene 20 as a photograph or a video.

In another application, the visual scene 20 may be rendered to a user via the display 200, for example, at a location remote from where the visual scene 20 was recorded. This situation is similar to the situation commonly experienced when reviewing images via a television screen, a computer screen or a mediated/virtual/augmented reality headset. In these examples, the visual scene 20 is a rendered visual scene. The active sound source 22 produces rendered sound, unless it has been muted. This implementation may be particularly useful for editing a sound space by, for example, modifying characteristics of sound sources and/or moving sound sources within the visual scene 20.

Fig 2B illustrates a visual scene 20 corresponding to the sound space 10 of Fig 1B, before movement of the sound source 22 in the visual scene 20. Fig 2C illustrates the same visual scene 20 corresponding to the sound space 10 of Fig 1C, after movement of the sound source 22.

Fig 2D illustrates the visual scene 20 after extension of the sound object 12 in the corresponding sound space 10. While the sound space 10 of Fig. 1D differs from the sound

space 10 of Fig. 1C in that the spatial extent of the sound object 12 has been increased so that the sound object has a greater breadth, the visual scene 20 is not necessarily changed.

The above described methods may be performed using a controller. An example of a controller 400 is illustrated in Fig 3A.

Implementation of the controller 300 may be as controller circuitry. The controller 300 may be implemented in hardware alone, have certain aspects in software including firmware alone or can be a combination of hardware and software (including firmware).

As illustrated in Fig 3A the controller 300 may be implemented using instructions that enable hardware functionality, for example, by using executable instructions of a computer program 306 in a general-purpose or special-purpose processor 302 that may be stored on a computer readable storage medium (disk, memory etc.) to be executed by such a processor 302.

The processor 302 is configured to read from and write to the memory 304. The processor 302 may also comprise an output interface via which data and/or commands are output by the processor 302 and an input interface via which data and/or commands are input to the processor 302.

The memory 304 stores a computer program 306 comprising computer program instructions (computer program code) that controls the operation of the apparatus 300 when loaded into the processor 302. The computer program instructions, of the computer program 306, provide the logic and routines that enables the apparatus to perform the methods illustrated in the figures. The processor 302 by reading the memory 304 is able to load and execute the computer program 306.

The controller 300 may be part of an apparatus or system 320. The apparatus or system 320 may comprise one or more peripheral components 312. The display 200 is a peripheral component. Other examples of peripheral components may include: an audio output device or interface for rendering or enabling rendering of the sound space 10 to the user; a user input device for enabling a user to control one or more parameters of the method; a positioning system for positioning a sound source; an audio input device such as a microphone or microphone array for recording a sound source; an image input device such as a camera or plurality of cameras.

The apparatus or system 320 may be comprised in a headset for providing mediated reality.

5 The controller 300 may be configured as a sound rendering engine that is configured to control characteristics of a sound object 12 defined by sound content. For example, the rendering engine may be configured to control the volume of the sound content, a position of the sound object 12 for the sound content within the sound space 10, a spatial extent of new sound object 12 for the sound content within the sound space 10, and other characteristics of the sound content such as, for example, tone or pitch or spectrum or reverberation etc. The sound object may, for example, be rendered via an audio output device or interface. The sound content
10 may be received by the controller 300.

The sound rendering engine may, for example comprise a spatial audio processing system 50 that is configured to control the position and/or extent of a sound object 12 within a sound space 10. Fig 4, illustrates an example of a spatial audio processing system 50 comprising a
15 spectral allocation module 70 and a spatial allocation module 72. The spectral allocation module 70 takes frequency sub-channels 51 of a received input audio signal 113 and allocates them to multiple spatial audio channels 52 as allocated frequency sub-channels 53.

20 In some but not necessarily all examples, the input audio signal 113 comprises a monophonic source signal and comprises, is accompanied with or is associated with one or more spatial processing parameters defining a position and/or spatial extent of the sound source that will render the monophonic source signal.

Each spatial audio channel is for rendering at a different location within a sound space.

25 The spatial allocation module 72 achieves the correct spatial rendering of the spatial audio channels 52 by controlled mixing 74 of the different spatial audio channels 52 across different audio device channels 76 that are rendered by different audio output devices. In this example, there are four audio device channels one for front right (FR), one for front left (FL), one for rear
30 right (RR) and one for rear left (RL) but other configurations are possible. However in other examples, there may be more (e.g. 5.1 or 7.1 surround sound) or less (binaural) audio output devices.

35 The sound space 10 may be considered to be a collection of spatial audio channels 52 where each spatial audio channel 52 is a different direction. In some examples, the collection of spatial audio channels may be globally defined for all sound objects 12. In other examples,

the collection of spatial audio channels may be locally defined for each sound object 12. The collection of spatial audio channels may be fixed or may vary dynamically with time.

5 In some but not necessarily all examples, each spatial audio channel may be rendered as a single rendered sound source using amplitude panning signals 54, for example, using Vector Base Amplitude Panning (VBAP).

10 For example, in spherical polar co-ordinates the direction of the spatial audio channel S_{nm} may be represented by the couplet of polar angle ϑ_n and azimuthal angle Φ_m . Where ϑ_n is one polar angle in a set of N possible polar angles and Φ_m is one azimuthal angle in a set of M possible azimuthal angles.

15 A sound object 12 at position $\mathbf{z}$ may be associated with the spatial audio channel S_{nm} that is closest to $\text{Arg}(\mathbf{z})$.

If a sound object 12 is associated with a spatial audio channel S_{nm} then it is rendered as a point source.

20 A sound object 12 may however have spatial extent and be associated with a plurality of spatial audio channels. For example, a sound object 12 may be simultaneously rendered in a set of spatial audio channels $\{S\}$ defined by $\text{Arg}(\mathbf{z})$ and a spatial extent of the sound object 12. That set of spatial audio channels $\{S\}$ may, for example, include the set of spatial audio channels $S_{n'm'}$ for each value of n' between $n-\delta_n$ and $n+\delta_n$ and of m' between $n-\delta_m$ and $n+\delta_m$, where n and m define the spatial audio channel closest to $\text{Arg}(\mathbf{z})$ and δ_n and δ_m define in combination a spatial extent of the sound object 12. The value of δ_n defines a spatial extent in a polar direction and the value of δ_m defines a spatial extent in an azimuthal direction.

30 The number of spatial audio channels and their spatial relationship in the set of spatial audio channels $\{S\}$, allocated by the spatial allocation module 72 is dependent upon the desired spatial extent of the sound object 12.

35 A single sound object 12 may be simultaneously rendered in a set of spatial audio channels $\{S\}$ by decomposing the audio signal representing the sound object 12 into multiple different frequency sub-channels 51 and allocating each frequency sub-channel 51 to one of multiple spectrally-limited audio signals 53.

Each of the multiple spectrally-limited audio signals 53 may have one or more frequency sub-channels 51 allocated to it (as an allocated frequency sub-channel). Each frequency sub-channel 51 may be allocated to only one spectrally-limited audio signal 53 (as an allocated frequency sub-channel).

5

Each spectrally-limited audio signals 53 is allocated into the set of spatial audio channels {S} 52.

10

For example, each spectrally-limited audio signal 53 is allocated to one spatial audio channel 52 and each spatial audio channel 52 comprises only one spectrally-limited audio signal 53, that is, there is a one-to-one mapping between the spectrally-limited audio signals and the spatial audio channels at the interface between the spectral allocation module 70 and the spatial allocation module 72. In some but not necessarily all examples, each spectrally-limited audio signal may be rendered as a single sound source using amplitude panning by the spatial allocation module 72.

15

20

For example, if the set of spatial audio channels {S} comprised X channels, the audio signal 113 representing the sound object 12 would be separated into X different spectrally-limited audio signals 53 in different non-overlapping frequency bands each frequency band comprising one or more different frequency sub-channels 51 that may be contiguous and/or non-contiguous. In some but not necessarily all examples, there may be $N=2^n$ frequency sub-bands, for example, N may be 512 ($n=9$), 1024 ($n=10$), 2048 ($n=11$). This may be achieved using a filter bank comprising a selective band pass limited filter for each spectrally-limited audio signal 53 /spatial audio channel or, as illustrated in Fig 4, by using digital signal processing to distribute time-frequency bins to different spectrally-limited audio signals 53/spatial audio channels 52. Each of the X different spectrally-limited audio signals 53 in different non-overlapping frequency bands would be provided to only one of the set of spatial audio channels {S}. Each of the set of spatial audio channels {S} would comprise only one of the X different spectrally-limited audio signals in different non-overlapping frequency bands.

25

30

35

Where digital signal processing is used to distribute time-frequency bins to different spatial audio channels, then a short-term Fourier transform (STFT) may be used to transform from the time domain to the frequency domain, where selective filtering occurs for each frequency band. The different spectrally-limited audio signals 53 may be created using the same time period or different time periods for each STFT. The different spectrally-limited audio signals 53 may be created by selecting frequency sub-channels 51 of the same bandwidth (different

center frequencies) or different bandwidths. The different spatial audio channels {S} into which the spectrally-limited audio signals 53 are placed may be defined by a constant angular distribution e.g. the same solid angle ($\Delta\Omega = \sin\theta \cdot \Delta\theta \cdot \Delta\Phi$ in spherical coordinates) or by a non-homogenous angular distribution e.g. different solid angles.

5

An inverse transform 78 will be required to convert from the frequency to the time domain. In some examples, this may occur in the spectral allocation module 70 or the spatial allocation module 72 before mixing. In the example illustrated in Fig 4, the inverse transform 78 occurs for each audio device channel 76, after mixing 74, in the spatial allocation module 72.

10

Which frequency sub-channel 51 is allocated to which spectrally-limited audio signal 53/ spatial audio channel 52 in the set of spatial audio channels {S} may be controlled by allocation module 60. The allocation may be a quasi-random allocation or may be determined based on a set of predefined rules. In some but not necessarily all examples, the allocation module 60 is a programmable filter bank.

15

The predefined rules may, for example, constrain spatial-separation of spectrally-adjacent frequency sub-channels 51 to be above a threshold value. Thus frequency sub-channels 51 adjacent in frequency may be separated spatially so that they are not spatially adjacent. In some examples, effective spatial separation of the multiple frequency sub-channels 51 that are adjacent in frequency may be maximized.

20

The predefined rules may additionally or alternatively define how frequency sub-channels 51 are distributed amongst the spectrally-limited audio signals 53/ set of spatial audio channels {S} 52. For example, a low discrepancy sequence such as a Halton sequence, for example, may be used to quasi-randomly distribute the frequency sub-channels 51 amongst the spectrally-limited audio signals 53/ spatial audio channels {S} 52.

25

Which frequency sub-channel 51 is allocated to which spectrally-limited audio signal 53/ spatial audio channel 52 in the set of spatial audio channels {S} may be dynamically controlled. For example, the allocation of frequency sub-channels 51 of the input audio signal 113 to multiple spatial audio channels 52 may be automatically changed.

30

A user input 101 may be used to control at least part of the allocation of the frequency sub-channels 51 to the spectrally-limited audio signal 53/ spatial audio channel 52.

35

Fig 5 illustrates an example of a method 100 comprising: at block 102, allocating frequency sub-channels 51 of an input audio signal 113 to multiple spatial audio channels 52, each spatial audio channel 52 for rendering at a location within a sound space; at block 104, responding to user input 101 to cause a change in an allocation of frequency sub-channels 51 of the input audio signal 113 to multiple spatial audio channels 52; and at block 106, automatically changing the allocation of frequency sub-channels 51 of the input audio signal 113 to one or more of the multiple spatial audio channels 52 selected by the user input 101.

In some but not necessarily all examples, the input audio signal 113 represents only a single sound object (e.g. a single sound source).

The method 100 may be performed by the allocation module 60 and/or the controller 300. The method 100 may be used to improve the perceived spatial uniformity of a rendered spatially extended sound. As a consequence, the sound is heard as a uniform, spatially extended sound instead of as distinct audio components at distinct spatial positions.

The apparatus or controller 300 may therefore comprises: at least one processor 302; and at least one memory 304 including computer program code the at least one memory 304 and the computer program code configured to, with the at least one processor 302, cause the apparatus 300 at least to perform:

allocating frequency sub-channels 51 of an input audio signal 113 to multiple spatial audio channels 52 each spatial audio channel 52 for rendering at a location within a sound space

10;
responding to user input 101 to cause a change in an allocation of frequency sub-channels 51 of the input audio signal 113 to multiple spatial audio channels 52;

and automatically changing the allocation of frequency sub-channels 51 of the input audio signal 113 to one or more of the multiple spatial audio channels 52 selected by the user input 101.

Figs 6A and 6B illustrate a visual scene 20 before re-allocation of frequency sub-channels 51 of the input audio signal 113 to one or more of the multiple spatial audio channels 52 selected by the user input 101 (Fig 6A) and after re-allocation of frequency sub-channels 51 of the input audio signal 113 to one or more of the multiple spatial audio channels 52 selected by the user input 101 (Fig 6B).

The visual scene 20 is presented via a user interface 400. In some but not necessarily all examples, the user interface 400 is provided by apparatus or system 320 via peripherals 312 (see Fig 3A).

- 5 The user interface 400 renders a sound space 10 and a corresponding visual scene 20. In some but not necessarily all examples, the visual scene 20 is the scene recorded at the same time the sound source that creates the sound object 12 is recorded.

10 The sound space 10 comprises a laterally extended sound object 12. In this example, the sound object has an extended breadth (width).

The visual scene 20 comprises a visual indication 412 of a lateral extent of the sound object 12. In this example but not necessarily all examples, the visual indication is a horizontal bar.

- 15 The user interface 400 renders the multiple spatial audio channels 52 of the input audio signal 113 representing the sound object 12 at different locations within the sound space 10 while simultaneously displaying a visualisation 412 of an expected position and extent of the sound source 12.

20 As illustrated in Figs 6A, a first part 101A of the user input 101 selects a first portion of the sound space. The selection is indicated visually using a visual indication 414A in the visual scene 20. The first portion of the sound space 10 and the location of the visual indication 414A in the visual scene 20 correspond in space.

25 As illustrated in Fig 6B, a second part 101B of the user input 101 selects a second portion of the sound space. The selection is indicated visually using a visual indication 414B in the visual scene 20. The second portion of the sound space 10 and the location of the visual indication 414B in the visual scene 20 correspond in space.

30 In the particular illustrated use case, but not necessarily in all cases, a user hears that the spatial extent processing is slightly off- some notes of the guitar playing are heard towards the right at a right-hand region. The first part 101A of the user input 101 is a grabbing hand gesture in the right-hand region that needs adjusting. The second part 101B of the user input 101 is a moving hand gesture towards a left-hand region and a dropping hand gesture at the left-hand
35 region. This is a drag and drop gesture. The right-hand region is indicated by visual indication 414A. The left-hand region is indicated by visual indication 414B.

Referring back to Fig 5, the method 100, for this example, comprises:

at block 102, allocating frequency sub-channels 51 of an input audio signal 113 to multiple spatial audio channels 52, each spatial audio channel 52 for rendering at a location within a sound space 10;

at block 104, responding to user input 101 that selects a first portion of the sound space 10 and a second portion of the sound space 10 different to the first portion, to cause a change in an allocation of frequency sub-channels 51 of the input audio signal 113 to multiple spatial audio channels 52 by:

determining one or more first spatial audio channels 52 associated with the first portion of the sound space 10 and

determining one or more second spatial audio channels 52 associated with the second portion of the sound space 10,

at block 106, automatically changing the allocation of frequency sub-channels 51 of the input audio signal 113 to at least the determined first one or more spatial audio channels 52 associated with the first portion of the sound space 10, and

and to at least the determined second one or more spatial audio channels associated with the second portion of the sound space 10.

In some but not necessarily all examples, at least some of the frequency sub-channels 51 of the input audio signal 113 are exchanged between the determined first one or more spatial audio channels 52 and the determined second one or more spatial audio channels 52.

In a particular implementation, the allocation of frequency sub-channels 51 of the input audio signal 113 is changed for only the determined first one or more spatial audio channels 52 and the determined second one or more spatial audio channels 52. The allocation of frequency sub-channels 51 of the input audio signal 113 to other spatial audio channels 52 is not changed. In this example, frequency sub-channels 51 of the input audio signal 113 may only be exchanged between the determined first one or more spatial audio channels 52 and the determined second one or more spatial audio channels 52.

The processing of the audio input signal 113 may be modelled as dividing the input audio signal 113 simultaneously into possible spatial audio channels 52 (space-division) and also separately and orthogonally into possible frequency sub-channels 52 (frequency division) creating possible space-frequency tile spaces, each of which is defined by a combination of spatial audio channel 52 and frequency sub-channel 51. The allocation of space-frequency

tiles into the space-frequency tile spaces defined by the different combinations of spatial audio channel 52 and frequency sub-channel 51 is according to some rules: each frequency sub-channel 51 only has one allocated tile; each spatial channel 52 in use has one or more allocated tiles, each associated with a different frequency sub-channel 51. The method 100
5 re-configures the arrangement of allocated tiles. A tile allocated for a particular frequency sub-channel 51 may change location (change spatial channel 52). The tile spaces (tile sizes) may change by changing a size of a frequency sub-channel 51 and/or by changing a size of a spatial audio channel 52. The user input 101 selects at least some of the spatial audio channels 52, and the allocated tiles associated with those spatial audio channels 52 change
10 while respecting the rules. The size of the allocated tiles may change. The distribution of allocated tiles amongst the frequency sub-channels 51 within each spatial audio channel 52 may change.

There may be some limitation on how much change can occur over time. There may be a limit
15 on the number of tiles that change spatial audio channel 52, for example.

There may be a check to ensure that the re-allocation improves quality before it is implemented by for example measuring the shape of power spectra for the spatial audio channels 52.

20 Figs 7A and 7B illustrate an apparatus 420 for providing the user interface 400 using mediated reality such as augmented reality or virtual reality. In some but not necessarily all examples, the mediated reality is perspective-mediated, for example, it may be first person perspective-mediated, by head movement and/or gaze as illustrated, or third person perspective-mediated. In some but not necessarily all examples, the mediated reality is user interactive.

25 The rendered virtual space comprises the sound space 10 and a 'corresponding' visual scene 20 (corresponding visual space).

In Fig 7A, the apparatus 420 is a headset that provides the user interface 400 illustrated in Fig
30 6A, before re-allocation of frequency sub-channels 51 of the input audio signal 113 to one or more of the multiple spatial audio channels 52 selected by the user input 101.

In Fig 7B, the apparatus 420 is the same headset providing the user interface 400 illustrated in Fig 6A, after re- allocation of frequency sub-channels 51 of the input audio signal 113 to
35 one or more of the multiple spatial audio channels 52 selected by the user input 101.

The headset 420 may comprise (headphones) comprising a left ear loudspeaker 424L and a right ear loudspeaker 424R for rendering a sound scene to a user and/or may comprise a head-mounted display or displays 422 for rendering a visual scene to a user wearing the headset 420 on their head.

5

The apparatus 420 renders the multiple spatial audio channels 52 at different locations within the sound space 10 that corresponds to a visual space 20 presented to the user via user-interactive mediated reality. The user input 101 is a user action relative to the visual space and sound space that selects a corresponding portion of the sound space. As described with reference to Figs 6A, 6B the user input 101 may have a first part 101A that selects a first portion of the sound space and a second part 101B that selects a second portion of the sound space different to the first portion.

10

15

Fig 8 illustrates an example of a system 110 that is configured to perform an example of the method 100. The allocation of frequency sub-channels 51 of the input audio signal 113 to one or more of the multiple spatial audio channels 52 is dependent upon a user input 101 selecting the one or more spatial audio channels.

20

In this example, but not necessarily all examples, the system 110 is configured to automatically detect a sub-optimal allocation of frequency sub-channels 51 of the input audio signal 113 to one or more of the multiple spatial audio channels 52 selected by the user input 101, and in response to detecting a sub-optimal allocation, automatically uses a new allocation of frequency sub-channels 51 of the input audio signal 113 to at least one or more of the multiple spatial audio channels 52 selected by the user input 101.

25

The system 110 comprises a spatial extent synthesizer module 114 that changes an allocation of frequency sub-channels 51 of the input audio signal 113 to multiple spatial audio channels 52 to change a spatial extent of a sound object 12.

30

The allocation of frequency sub-channels 51 of the input audio signal 113 to multiple spatial audio channels 52 is defined by a distribution 119 provided by the distribution generator 118.

The distribution generator 118 generates the new distribution 119 for at least the one or more spatial audio channels 52 selected by the user input 101, in response to that user input 101.

35

An analyser module 116 is configured to automatically analyse at least one or more of the multiple spatial audio channels 52 selected by the user input 101 and to detect a sub-optimal allocation of frequency sub-channels 51 of the input audio signal 113 to the one or more of the multiple spatial audio channels 52 selected by the user input 101, and in response to
5 detecting a sub-optimal allocation, automatically controls the distribution generator 118 to define a new allocation 119 of frequency sub-channels 51 of the input audio signal to at least the one or more of the multiple spatial audio channels 52 selected by the user input 101. The new allocation 119 is used by spatial extent synthesizer module 114 to change the allocation of frequency sub-channels 51 of the input audio signal 113 to the spatial audio channels 52.

10 In some but not necessarily all examples, the system 110, is configured to automatically change the allocation of frequency sub-channels 51 of the input audio signal 113 to one or more of the multiple spatial audio channels 52 selected by the user input 101 based upon an analysis of the frequency sub-channels 51 of the input audio signal 113 allocated to the one
15 or more user-selected spatial audio channels 52 selected by the user input 101. The analysis may occur before or after re-allocation of the frequency sub-channels 51.

For example, in some but not necessarily all examples, frequency sub-channels 51 of the input audio signal 113 are re-allocated to the one or more user-selected spatial audio channels 52
20 selected by the user input 101, such that the frequency sub-channels 51 of the input audio signal 113 then allocated to the one or more user selected spatial audio channels 52 selected by the user input 101, have a frequency spectrum more similar to a reference frequency spectrum after re-allocation than before re-allocation.

25 For example, it may be desirable to ensure that harmonics are distributed amongst spatial audio channels 52. It may, for example be desirable to move an harmonic from one of the multiple spatial audio channel 52 selected by the user input 101 (e.g. the first one selected 101A at the first portion) to another different one of the multiple spatial audio channels 52 selected by the user input 101 (e.g. the second one selected 101B at the second portion).

30 For example, it may be desirable to ensure that certain frequency ranges are distributed amongst more than one spatial audio channel 52. It may, for example be desirable to split and move or move a frequency sub-channel 51 that has significant energy from one of the multiple spatial audio channels 52 selected by the user input 101 e.g. the first one selected 101A at
35 the first portion) to another different one of the multiple spatial audio channels 52 selected by the user input 101 (e.g. the second one selected 101B at the second portion).

For example, it may be desirable to ensure that the power spectrum of a user selected spatial audio channel 52 has a better match to a power spectrum of an input audio signal 113 and/or another user selected spatial audio channel 52.

5 In some but not necessarily all examples, the system 110, is configured to change the allocation of frequency sub-channels 51 of the input audio signal 113 to one or more of the multiple spatial audio channels 52 selected by the user input 101 in dependence upon one or more changes in a power spectrum of the input audio signal 113. In some but not necessarily all examples, the system 110, is configured to change the allocation of frequency sub-
10 channels 51 of the input audio signal 113 to the one or more of the multiple spatial audio channels 52 selected by the user input 101 to reduce deviation from a power spectrum of the input audio signal 113. This may prevent the power spectrum of each of the spatial audio channels from deviating significantly (e.g. more than a threshold value) from a power spectrum of the input audio signal 113.

15 In some but not necessarily all examples, the system 110 adjusts a current allocation of frequency sub-channels 51 of the input audio signal to at least the first one or more of the multiple spatial audio channels 52 selected by the user input 101 to reduce a cost function value for the current allocation of frequency sub-channels 51 of the input audio signal 113 to
20 multiple spatial audio channels 52. For example, in some but not necessarily all examples, the system 110 automatically determines a first cost function value for the current allocated frequency sub-channels 51 (based on a current allocation of frequency sub-channels of the input audio signal to multiple spatial audio channels 52) and automatically determines a second cost function value for putative allocated frequency sub-channels (based on a different
25 putative allocation of frequency sub-channels of the input audio signal to at least one or more of the multiple spatial audio channels 52 selected by the user input 101), and in response to determining the first cost function value is sufficiently greater than the second cost function value, makes the putative allocation of frequency sub-channels 51 of the input audio signal 113 to one or more of the multiple spatial audio channels 52 selected by the user input 101
30 part of the current allocation of frequency sub-channels 51 of the input audio signal 113 to spatial audio channels 52. The putative allocated frequency sub-channels therefore become current allocated frequency sub-channels.

The distribution generator module 118 may generate the putative allocation of frequency sub-
35 channels. The analyser module 116 may determine the cost function and compare the cost

function values making the decision to change the current allocation of frequency sub-channels of the input audio signal.

5 The cost function may compare one or more parameters of the current input signal 113 with one or more parameters of each of the different spatial audio channels 52.

10 The cost function may compare one or more parameters of a reference with one or more parameters of each of the one or more of the multiple spatial audio channels 52 selected by the user input 101. For example, as previously described it may be desirable to ensure that harmonics are distributed amongst more than one spatial audio channel and/or to ensure that certain frequency ranges are distributed amongst more than one spatial audio channel and/or ensure that the power spectrum of a user selected spatial audio channel has a better match to a power spectrum of an input audio signal and/or another user selected spatial audio channel.

15 The cost function may for example be based on different parameters such as, for example parameters $p(f)$ that vary with frequency f , such as amplitude or power spectral density or be based on cepstral analysis.

20 The cost function may for example be based on different combinations of parameters. It may, for example, comprise a function that averages a parameter over a range of frequencies, such as a moving mean calculation.

25 In some but not necessarily all examples, changing an allocation of frequency sub-channels of the input audio signal to spatial audio channels, comprises changing a definition of the frequency sub-channels and/or changing a distribution of frequency-sub channels 51 across spatial audio channels 52.

30 The frequency sub-channels 51 may for example each be defined by a center frequency and a bandwidth. The definition of a frequency sub-channel 51 may be changed by changing its center frequency and/or by changing its bandwidth. The definition of the frequency sub-channel 51 may for example occur subject to certain constraints such as the frequency sub-channels 51 do not overlap and/or that the frequency sub-channels 51 cover in combination certain frequency ranges.

35

In some but not necessarily all examples, a slowly varying part of the spectrum may be covered by fewer, wider frequency sub-channels 51 and more quickly varying part of spectrum may be covered by more, narrower frequency sub-channels 51.

- 5 In some but not necessarily all examples, the lower frequency part of the spectrum may be covered by narrower sub-channels 51 and the higher frequency part of the spectrum may be covered by wider frequency sub-channels 51.

10 The distribution of frequency-sub channels 51 across spatial audio channels 52 may be changed by changing the rules used to distribute frequency-sub channels 51 across spatial audio channels 52. This may correspond to block 102 and/or block 106 of method 100. The rules define how the spectrally-limited audio signals 53 are distributed amongst the set of spatial audio channels {S}. They may or may not include constraints concerning spatial separation of frequency sub-channels 51 that are adjacent in the frequency spectrum.

15 The distribution of frequency-sub channels across spatial audio channels 52 may be changed by changing one or more low-discrepancy sequences used for distribution of frequency-sub channels 51 across spatial audio channels 52.

- 20 A position or direction in the sound space 10 may be represented by one or more values derived from one or more low-discrepancy sequences. For example a point in two dimensions (x,y) (or $(|z|, \text{Arg}(z))$ in polar co-ordinates) may be determined from two low discrepancy sequences, one for x and one for y. For example a point in three dimensions (x,y, x), (or $(|z|, \text{Arg}(z))$ in polar co-ordinates) may be determined from three low discrepancy sequences, one
25 for x, one for y and one for z.

There are various different examples of low-discrepancy sequences. One example is a Halton sequence. A Halton sequence is defined by a base value and by a skip value. A new Halton sequence is a Halton sequence with a new base value and/or a new skip value. A new Halton
30 sequence may additionally or alternatively be created by scrambling a Halton sequence or leaping or changing leaping in a Halton sequence. Scrambling changes the order of a Halton sequence. Leaping results in certain values in the Halton sequence not being used.

35 The distribution of frequency-sub channels 51 across spatial audio channels 52 may be changed by changing one or more Halton sequences used for distribution of frequency-sub channels across spatial audio channels.

The parameters used for sequence generation, for example, base, skip, scrambling or leaping may be changed randomly or in a preset manner.

5 Fig 9 illustrates an example of a method 500 for controlling rendering of spatial audio and in particular controlling rendering of a sound object 12 that has a spatial extent, for example width in response to re-allocation of frequency sub-channels 51. The method may be used to improve the perceived spatial uniformity of a rendered spatially extended sound without undesirable transitional effects.

10

In the previous description, a first allocation of frequency sub-channels 51 of an input audio signal 113 to multiple spatial audio channels 52 is changed to a second allocation of frequency sub-channels of an input audio signal to multiple spatial audio channels 52. This can improve the perceived spatial uniformity of a rendered spatially extended sound. Distinct audio
15 components of the sound will not, as a consequence, be heard at distinct spatial positions and the sound will be heard as a uniform, spatially extended sound.

15

However, the second allocation of frequency sub-channels 51 may not be immediately used and there may be a gradual transition between the first allocation of frequency sub-channels
20 51 and the second allocation of frequency sub-channels 51.

20

In the example method 500, illustrated in Fig 9, at block 502 a first allocation of frequency sub-channels 51 is used to render a sound object 12, at block 506 the second allocation of frequency sub-channels 51 is used to render the sound object 12, and between blocks 502
25 and 506, at block 504, a transitional allocation of frequency sub-channels 51 is used to render the sound object 12.

25

For example, there may be a cross-fade from the first allocation to the second allocation. There may be an independently controlled cross-fade for each frequency sub-channel 51 such that
30 different frequency sub-channels cross-fade at different rates.

30

Referring back to the preceding examples, in some situations, additional processing may be required. For example, when the sound space 10 is rendered to a listener through a head-mounted audio output device, for example headphones or a headset using binaural audio
35 coding, it may be desirable for the rendered sound space to remain fixed in space when the listener turns their head in space. This means that the rendered sound space needs to be

35

rotated relative to the audio output device by the same amount in the opposite sense to the head rotation. The orientation of the rendered sound space tracks with the rotation of the listener's head so that the orientation of the rendered sound space remains fixed in space and does not move with the listener's head. The system uses a transfer function to perform a transformation T that rotates the sound objects 12 within the sound space. A head related transfer function (HRTF) interpolator may be used for rendering binaural audio. Vector Base Amplitude Panning (VBAP) may be used for rendering in loudspeaker format (e.g. 5.1) audio.

The distance of a sound object 12 from an origin at the user may be controlled by using a combination of direct and indirect processing of audio signals representing the sound object 12. The audio signals are passed in parallel through a "direct" path and one or more "indirect" paths before the outputs from the paths are mixed together. The direct path represents audio signals that appear, to a listener, to have been received directly from an audio source and an indirect (decorrelated) path represents audio signals that appear to a listener to have been received from an audio source via an indirect path such as a multipath or a reflected path or a refracted path. Modifying the relative gain between the direct path and the indirect paths, changes the perception of the distance D of the sound object 12 from the listener in the rendered sound space 10. Increasing the indirect path gain relative to the direct path gain increases the perception of distance. The decorrelated path may, for example, introduce a pre-delay of at least 2 ms.

In some but not necessarily all examples, to achieve a sound object with spatial extent (width and/or height and/or depth)_the spatial audio channels 52 are treated as spectrally distinct sound objects that are then positioned at suitable widths and/or heights and/or distances using known audio reproduction methods.

For example, in the case of loudspeaker sound reproduction amplitude panning can be used for positioning a spectrally distinct sound object in the width and/or height dimension, and distance attenuation by gain control and optionally direct to reverberant (indirect) ratio can be used to position spectrally distinct sound objects in the depth dimension.

For example, in case of binaural rendering, positioning in width and/or height dimension is obtained by selecting suitable head related transfer function (HRTF) filters (one for left ear, one for right ear) for each of the spectrally distinct sound objects depending on its position. A pair of HRTF filters model the path from a point in space to the listener's ears. The HRTF coefficient pairs are stored for all the possible directions of arrival for a sound. Similarly,

distance dimension of a spectrally distinct sound object is controlled by modelling distance attenuation with gain control and optionally direct to reverberant (indirect) ratio.

Thus, assuming that the sound rendering system supports width, then the width of a sound object may be controlled by the spatial allocation module 72. It achieves the correct spatial rendering of the spatial audio channels 52 by controlled mixing 74 of the different spatial audio channels 52 across different width-separated audio device channels 76 that are rendered by different audio output devices.

Thus assuming that the sound rendering system supports height, then the height of a sound object may be controlled in the same manner as a width of a sound object. The spatial allocation module 72 achieves the correct spatial rendering of the spatial audio channels 52 by controlled mixing 74 of the different spatial audio channels 52 across different height-separated audio device channels 76 that are rendered by different audio output devices.

Thus assuming that the sound rendering system supports depth, then the depth of a sound object may be controlled in the same manner as a width of a sound object. The spatial allocation module 72 achieves the correct spatial rendering of the spatial audio channels 52 by controlled mixing 74 of the different spatial audio channels 52 across different depth-separated audio device channels 76 that are rendered by different audio output devices. However, if that is not possible, the spatial allocation module 72 may achieve the correct spatial rendering of the spatial audio channels 52 by controlled mixing 74 of the different spatial audio channels 52 across different depth-separated spectrally distinct sound objects at different perception distances by modelling distance attenuation using gain control and optionally direct to reverberant (indirect) ratio.

It will therefore be appreciated that the extent of a sound object can be controlled widthwise and/or heightwise and/or depthwise.

Referring back to Figs 3A and 3B, the computer program 306 may arrive at the apparatus 300 via any suitable delivery mechanism 310. The delivery mechanism 310 may be, for example, a non-transitory computer-readable storage medium, a computer program product, a memory device, a record medium such as a compact disc read-only memory (CD-ROM) or digital versatile disc (DVD), an article of manufacture that tangibly embodies the computer program 306. The delivery mechanism may be a signal configured to reliably transfer the computer

program 306. The apparatus 300 may propagate or transmit the computer program 306 as a computer data signal.

Although the memory 304 is illustrated as a single component/circuitry it may be implemented as one or more separate components/circuitry some or all of which may be integrated/removable and/or may provide permanent/semi-permanent/ dynamic/cached storage.

Although the processor 302 is illustrated as a single component/circuitry it may be implemented as one or more separate components/circuitry some or all of which may be integrated/removable. The processor 302 may be a single core or multi-core processor.

References to 'computer-readable storage medium', 'computer program product', 'tangibly embodied computer program' etc. or a 'controller', 'computer', 'processor' etc. should be understood to encompass not only computers having different architectures such as single /multi- processor architectures and sequential (Von Neumann)/parallel architectures but also specialized circuits such as field-programmable gate arrays (FPGA), application specific circuits (ASIC), signal processing devices and other processing circuitry. References to computer program, instructions, code etc. should be understood to encompass software for a programmable processor or firmware such as, for example, the programmable content of a hardware device whether instructions for a processor, or configuration settings for a fixed-function device, gate array or programmable logic device etc.

As used in this application, the term 'circuitry' refers to all of the following:

(a) hardware-only circuit implementations (such as implementations in only analog and/or digital circuitry) and

(b) to combinations of circuits and software (and/or firmware), such as (as applicable): (i) to a combination of processor(s) or (ii) to portions of processor(s)/software (including digital signal processor(s)), software, and memory(ies) that work together to cause an apparatus, such as a mobile phone or server, to perform various functions and

(c) to circuits, such as a microprocessor(s) or a portion of a microprocessor(s), that require software or firmware for operation, even if the software or firmware is not physically present.

This definition of 'circuitry' applies to all uses of this term in this application, including in any claims. As a further example, as used in this application, the term "circuitry" would also cover an implementation of merely a processor (or multiple processors) or portion of a processor and its (or their) accompanying software and/or firmware. The term "circuitry" would also

cover, for example and if applicable to the particular claim element, a baseband integrated circuit or applications processor integrated circuit for a mobile phone or a similar integrated circuit in a server, a cellular network device, or other network device.

5 The blocks illustrated in the enclosed figures may represent steps in a method and/or sections of code in the computer program 306. The illustration of a particular order to the blocks does not necessarily imply that there is a required or preferred order for the blocks and the order and arrangement of the block may be varied. Furthermore, it may be possible for some blocks to be omitted.

10

Where a structural feature has been described, it may be replaced by means for performing one or more of the functions of the structural feature whether that function or those functions are explicitly or implicitly described.

15 The controller 300 may, for example be a module. 'Module' refers to a unit or apparatus that excludes certain parts/components that would be added by an end manufacturer or a user.

The term 'comprise' is used in this document with an inclusive not an exclusive meaning. That is any reference to X comprising Y indicates that X may comprise only one Y or may comprise
20 more than one Y. If it is intended to use 'comprise' with an exclusive meaning then it will be made clear in the context by referring to "comprising only one" or by using "consisting".

In this brief description, reference has been made to various examples. The description of features or functions in relation to an example indicates that those features or functions are
25 present in that example. The use of the term 'example' or 'for example' or 'may' in the text denotes, whether explicitly stated or not, that such features or functions are present in at least the described example, whether described as an example or not, and that they can be, but are not necessarily, present in some of or all other examples. Thus 'example', 'for example' or 'may' refers to a particular instance in a class of examples. A property of the instance can be
30 a property of only that instance or a property of the class or a property of a sub-class of the class that includes some but not all of the instances in the class. It is therefore implicitly disclosed that a feature described with reference to one example but not with reference to another example, can where possible be used in that other example but does not necessarily have to be used in that other example.

35

Although embodiments of the present invention have been described in the preceding paragraphs with reference to various examples, it should be appreciated that modifications to the examples given can be made without departing from the scope of the invention as claimed.

- 5 Features described in the preceding description may be used in combinations other than the combinations explicitly described.

Although functions have been described with reference to certain features, those functions may be performable by other features whether described or not.

10

Although features have been described with reference to certain embodiments, those features may also be present in other embodiments whether described or not.

15

Whilst endeavoring in the foregoing specification to draw attention to those features of the invention believed to be of particular importance it should be understood that the Applicant claims protection in respect of any patentable feature or combination of features hereinbefore referred to and/or shown in the drawings whether or not particular emphasis has been placed thereon.

20 I/we claim:

25

30

35

CLAIMS

1. A method comprising:

- 5 allocating frequency sub-channels of an input audio signal to multiple spatial audio channels, each spatial audio channel for rendering at a location within a sound space; and responding to user input to cause a change in an allocation of frequency sub-channels of the input audio signal to multiple spatial audio channels; automatically changing the allocation of frequency sub-channels of the input audio signal to at least a one or more of the multiple spatial audio channels selected by the user input.
- 10

2. A method as claimed in claim 1, wherein the user input selects a first portion of the sound space, the method comprising:

- 15 determining one or more first spatial audio channels associated with the first portion of the sound space, wherein automatically changing the allocation of frequency sub-channels of the input audio signal to at least one or more of the multiple spatial audio channels selected by the user input comprises automatically changing the allocation of frequency sub-channels of the input audio signal to at least the determined first one or more spatial audio channels associated with the first portion of the sound space,
- 20

3. A method as claimed in claim 2, wherein the user input selects a second portion of the sound space different to the first portion, the method comprising:

- 25 determining one or more second spatial audio channels associated with the second portion of the sound space, wherein automatically changing the allocation of frequency sub-channels of the input audio signal to at least one or more of the multiple spatial audio channels selected by the user input comprises automatically changing the allocation of frequency sub-channels of the input audio signal to at least the determined second one or more spatial audio channels associated with the second portion of the sound space.
- 30

4. A method as claimed in claim 3,

wherein automatically changing the allocation of frequency sub-channels of the input audio signal to at least the determined first one or more spatial audio channels and the determined second one or more spatial audio channels, comprises

changing the allocation of frequency sub-channels of the input audio signal to only the determined first one or more spatial audio channels and the determined second one or more spatial audio channels, and not changing the allocation of frequency sub-channels of the input audio signal to other spatial audio channels such that the allocation of frequency sub-channels of the input audio signal to the other spatial audio channels remains unchanged.

5. A method as claimed in any of claims 2 to 4, wherein the user input selects the first portion of the sound space, then the second portion of the sound space using a hand gesture.

6. A method as claimed in any of claims 1 to 5, wherein rendering the multiple spatial audio channels at different locations within a sound space while simultaneously displaying a visualisation of an expected position and extent of a sound source associated with the input audio signal.

7. A method as claimed in any of claims 1 to 6, wherein rendering the multiple spatial audio channels at different locations within a sound space that corresponds to a visual space presented to the user via mediated reality, wherein the user input is a user action relative to the visual space that selects a corresponding portion of the sound space.

8. A method as claimed in any of claims 1 to 7, wherein automatically changing the allocation of frequency sub-channels of the input audio signal to at least one or more of the multiple spatial audio channels selected by the user input comprises:
automatically changing the allocation of frequency sub-channels of the input audio signal to the at least one or more of the multiple spatial audio channels based upon an analysis of the frequency sub-channels of the input audio signal allocated to the one or more user-selected spatial audio channels selected by the user input.

9. A method as claimed in claim 8, wherein automatically changing the allocation of frequency sub-channels of the input audio signal to at least one or more of the multiple spatial audio channels based upon an analysis of the frequency sub-channels of the input audio signal allocated to one or more user-selected spatial audio channels selected by the user input, comprises re-allocating frequency sub-channels of the input audio signal to the one or more user-selected spatial audio channels selected by the user input, such that

the frequency sub-channels of the input audio signal then allocated to the one or more user selected spatial audio channels selected by the user input, have a frequency spectrum more similar to a reference frequency spectrum after re-allocation than before re-allocation.

5 10. A method as claimed in any of claims 1 to 9, wherein automatically changing the allocation of frequency sub-channels of the input audio signal to at least one or more of the multiple spatial audio channels selected by the user input comprises:
adjusting a current allocation of frequency sub-channels of the input audio signal to at least
the first one or more of the multiple spatial audio channels selected by the user input to reduce
10 a cost function value for the current allocation of frequency sub-channels of the input audio signal to multiple spatial audio channels.

11. A method as claimed in any of claims 1 to 10, wherein automatically changing the allocation of frequency sub-channels of the input audio signal to at least one or more of the
15 multiple spatial audio channels selected by the user input comprises automatically changing a definition of the frequency sub-channels and/or a distribution of frequency-sub channels across spatial channels.

12. A method as claimed in any of claims 1 to 11, wherein automatically changing the allocation of frequency sub-channels of the input audio signal to at least one or more of the
20 multiple spatial audio channels selected by the user input comprises automatically changing a distribution of frequency-sub channels across spatial channels without changing a definition of the frequency sub-channels.

25 13. A method as claimed in any of claims 1 to 12, wherein automatically changing the allocation of frequency sub-channels of the input audio signal to at least a first one or more of the multiple spatial audio channels selected by the user input, comprises changing a distribution of frequency-sub channels across spatial channels by changing one or more low-discrepancy sequences used for distribution.

30 14. A method comprising:
rendering multiple spatial audio channels at different locations within a sound space;
automatically applying a transition effect, when the allocation of frequency sub-channels of the input audio signal to multiple spatial audio channels is changed.

35

15. An apparatus comprising means for performing the method of any of claims 1 to 14 and/or a computer program that when run on a processor causes the method of any of claims 1 to 14.

5 16. An apparatus comprising: at least one processor; and at least one memory including computer program code the at least one memory and the computer program code configured to, with the at least one processor, causes the apparatus at least to: allocate frequency sub-channels of an input audio signal to multiple spatial audio channels, each spatial audio channel for rendering at a location within a sound space; respond to user input to cause a change in
10 an allocation of frequency sub-channels of the input audio signal to multiple spatial audio channels; automatically change the allocation of frequency sub-channels of the input audio signal to at least a one or more of the multiple spatial audio channels selected by the user input.

15 17. An apparatus as claimed in claim 16, wherein the user input selects a first portion of the sound space, the apparatus is caused to:
determine one or more first spatial audio channels associated with the first portion of the sound space, wherein the apparatus is cause to automatically change the allocation of frequency sub-channels of the input audio signal to at least one or more of the multiple spatial audio
20 channels selected by the user input is caused to automatically change the allocation of frequency sub-channels of the input audio signal to at least the determined first one or more spatial audio channels associated with the first portion of the sound space,

25 18. An apparatus as claimed in claim 17, wherein the user input selects a second portion of the sound space different to the first portion, the apparatus is caused to:
determine one or more second spatial audio channels associated with the second portion of the sound space, wherein the apparatus is caused to automatically change the allocation of frequency sub-channels of the input audio signal to at least one or more of the multiple spatial audio channels selected by the user input is further caused to automatically change the
30 allocation of frequency sub-channels of the input audio signal to at least the determined second one or more spatial audio channels associated with the second portion of the sound space.

35 19. An apparatus as claimed in claim 18, wherein the apparatus is caused to automatically change the allocation of frequency sub-channels of the input audio signal to at least the determined first one or more spatial audio channels and the determined second one or more

spatial audio channels, is further caused to change the allocation of frequency sub-channels of the input audio signal to only the determined first one or more spatial audio channels and the determined second one or more spatial audio channels, and is caused not to change the allocation of frequency sub-channels of the input audio signal to other spatial audio channels such that the allocation of frequency sub-channels of the input audio signal to the other spatial audio channels remains unchanged.

20. An apparatus as claimed in any of claims 17 to 19, wherein the user input selects the first portion of the sound space, then the second portion of the sound space using a hand gesture.

21. An apparatus as claimed in any of claims 16 to 20, wherein the multiple spatial audio channels are rendered at different locations within a sound space while the apparatus is caused to simultaneously display a visualisation of an expected position and extent of a sound source associated with the input audio signal.

22. An apparatus as claimed in any of claims 16 to 21, wherein the multiple spatial audio channels are rendered at different locations within a sound space that corresponds to a visual space presented to the user via mediated reality, wherein the user input is a user action relative to the visual space that selects a corresponding portion of the sound space.

23. An apparatus as claimed in any of claims 16 to 22, wherein the apparatus is caused to automatically change the allocation of frequency sub-channels of the input audio signal to at least one or more of the multiple spatial audio channels selected by the user input is further caused to: automatically change the allocation of frequency sub-channels of the input audio signal to the at least one or more of the multiple spatial audio channels based upon an analysis of the frequency sub-channels of the input audio signal allocated to the one or more user-selected spatial audio channels selected by the user input.

24. An apparatus as claimed in claim 23, wherein the apparatus is caused to automatically change the allocation of frequency sub-channels of the input audio signal to at least one or more of the multiple spatial audio channels based upon an analysis of the frequency sub-channels of the input audio signal allocated to one or more user-selected spatial audio channels selected by the user input, is further caused to re-allocate frequency sub-channels of the input audio signal to the one or more user-selected spatial audio channels selected by the user input, such that the frequency sub-channels of the input audio signal then allocated to the one or more user selected spatial audio channels selected by the user input, have a

frequency spectrum more similar to a reference frequency spectrum after re-allocation than before re-allocation.

5 25. An apparatus as claimed in any of claims 16 to 24, wherein the apparatus is caused to automatically change the allocation of frequency sub-channels of the input audio signal to at least one or more of the multiple spatial audio channels selected by the user input is further caused to: adjust a current allocation of frequency sub-channels of the input audio signal to at least the first one or more of the multiple spatial audio channels selected by the user input to reduce a cost function value for the current allocation of frequency sub-channels of the input
10 audio signal to multiple spatial audio channels.

26. An apparatus as claimed in any of claims 16 to 25, wherein the apparatus is caused to automatically change the allocation of frequency sub-channels of the input audio signal to at least one or more of the multiple spatial audio channels selected by the user input is further
15 caused to automatically change a definition of the frequency sub-channels and/or a distribution of frequency-sub channels across spatial channels.

27. An apparatus as claimed in any of claims 16 to 26, wherein the apparatus is caused to automatically change the allocation of frequency sub-channels of the input audio signal to at least one or more of the multiple spatial audio channels selected by the user input is further
20 caused to automatically change a distribution of frequency-sub channels across spatial channels without changing a definition of the frequency sub-channels.

28. An apparatus as claimed in any of claims 16 to 27, wherein the apparatus is caused to
25 automatically change the allocation of frequency sub-channels of the input audio signal to at least a first one or more of the multiple spatial audio channels selected by the user input, is further caused to change a distribution of frequency-sub channels across spatial channels by changing one or more low-discrepancy sequences used for distribution.

30 29. An apparatus comprising:
means for rendering multiple spatial audio channels at different locations within a sound space; and
means for automatically applying a transition effect, when the allocation of frequency sub-channels of the input audio signal to multiple spatial audio channels is changed.

35

1/6

SOUND SPACE

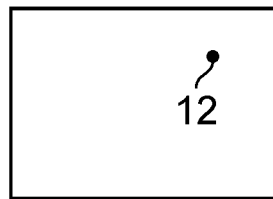


FIG. 1A

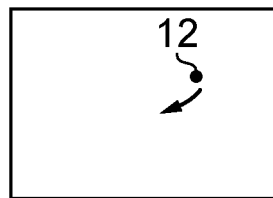


FIG. 1B

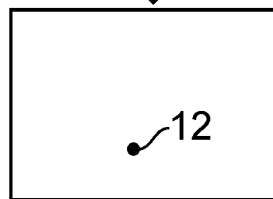


FIG. 1C

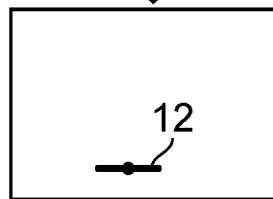


FIG. 1D

VISUAL SCENE

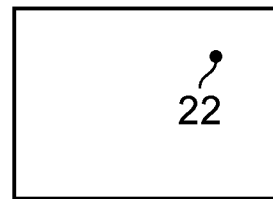


FIG. 2A

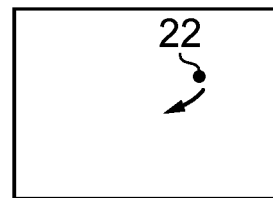


FIG. 2B

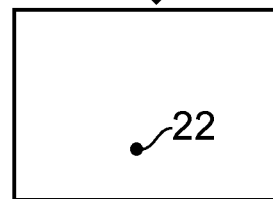


FIG. 2C

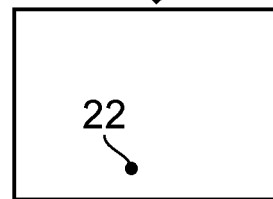


FIG. 2D

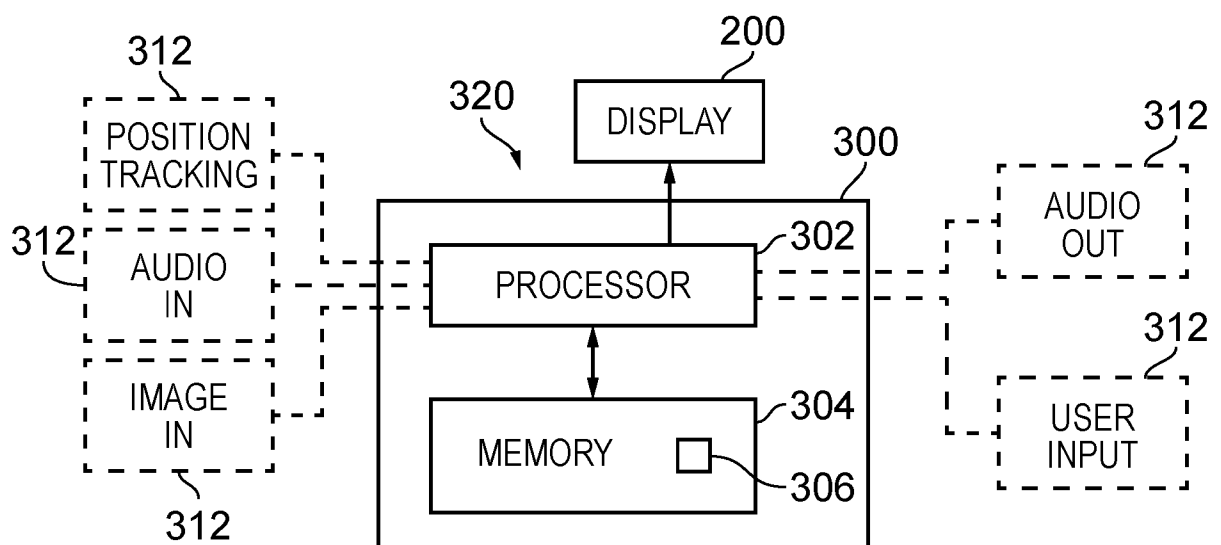


FIG. 3A

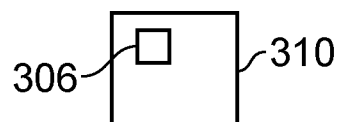


FIG. 3B

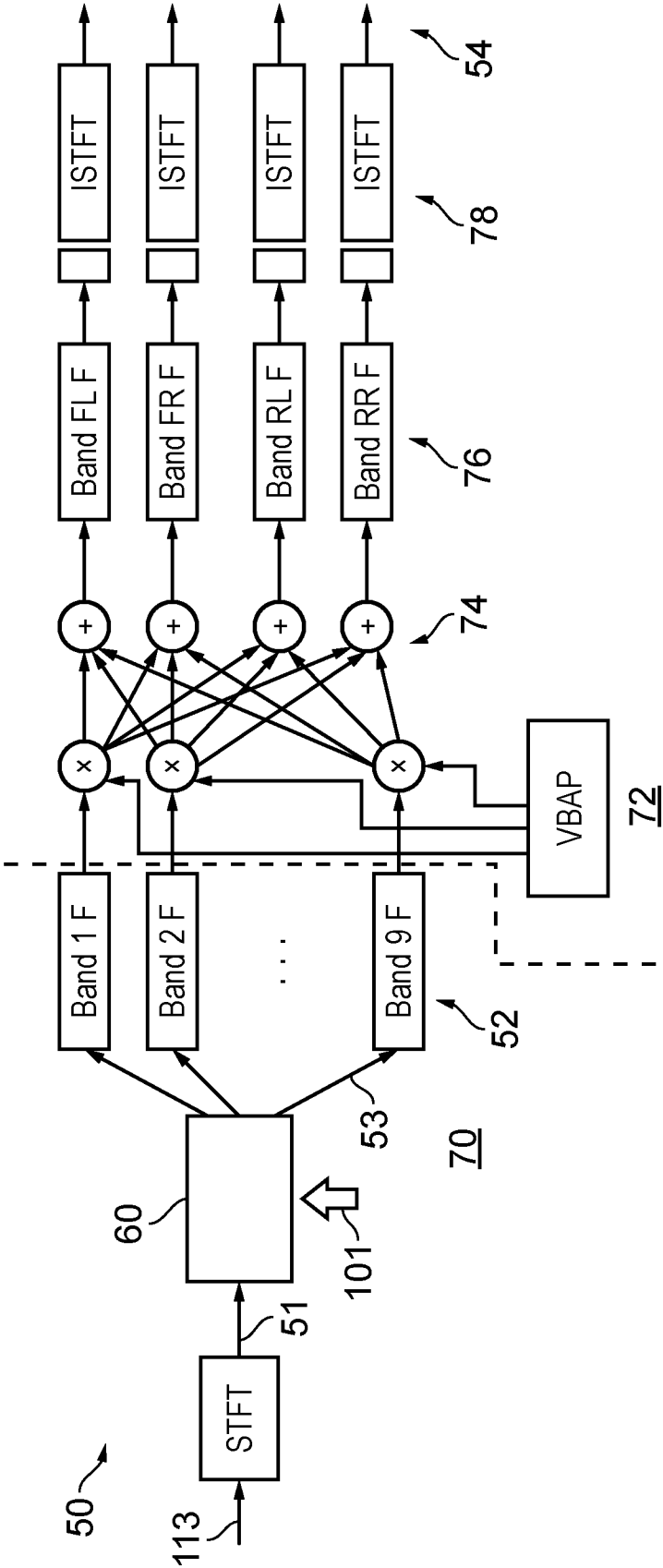


FIG. 4

3/6

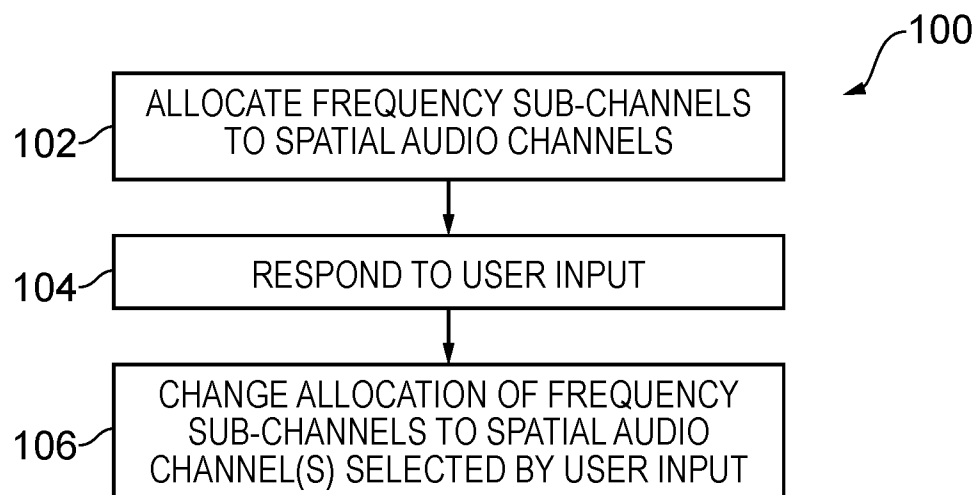
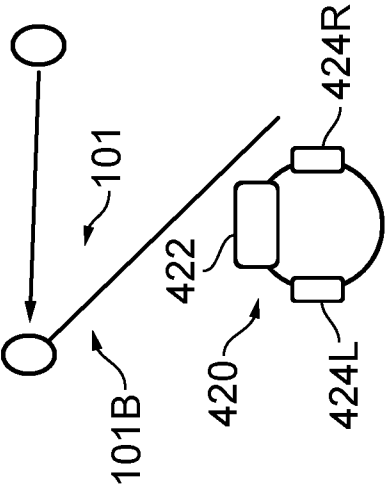
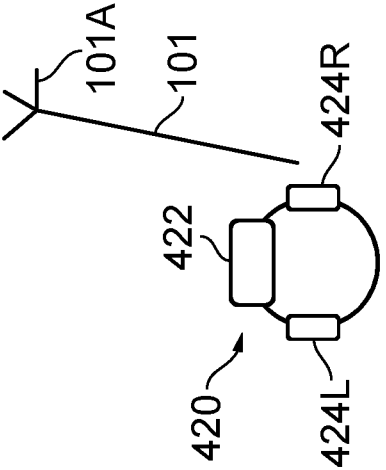
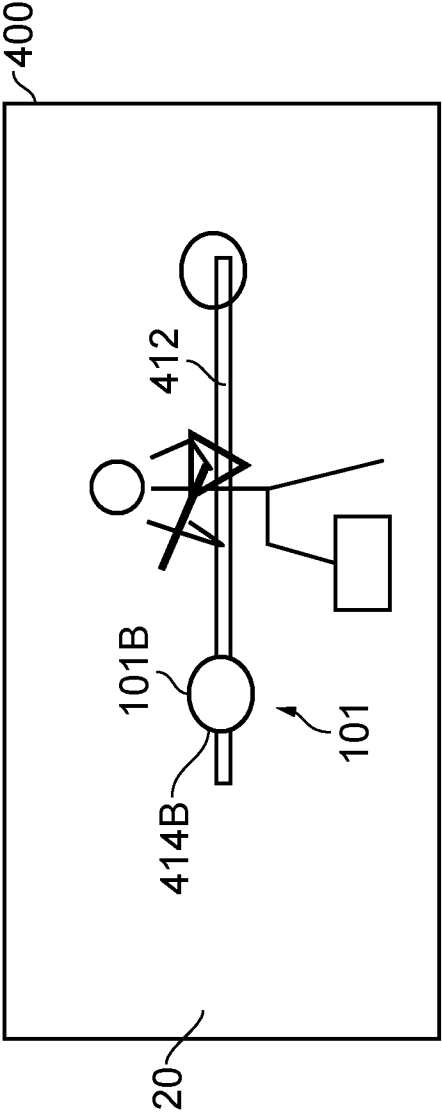
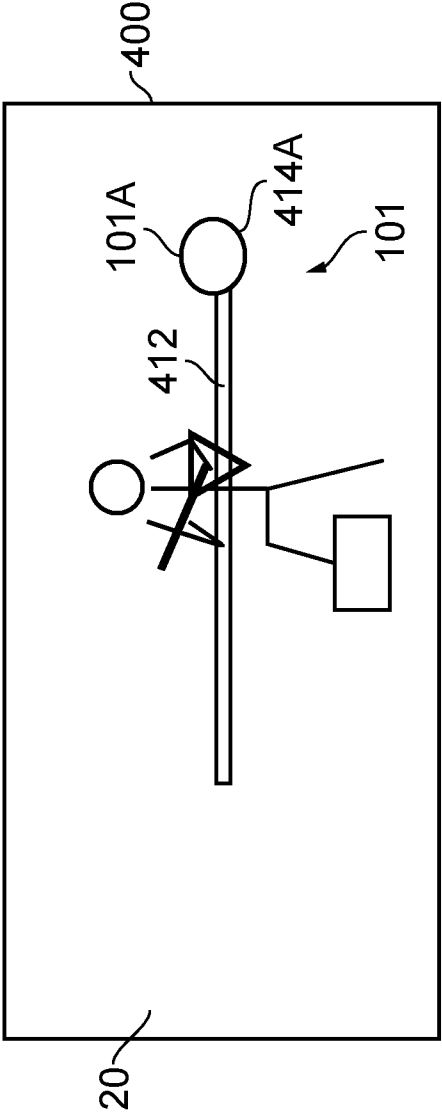


FIG. 5



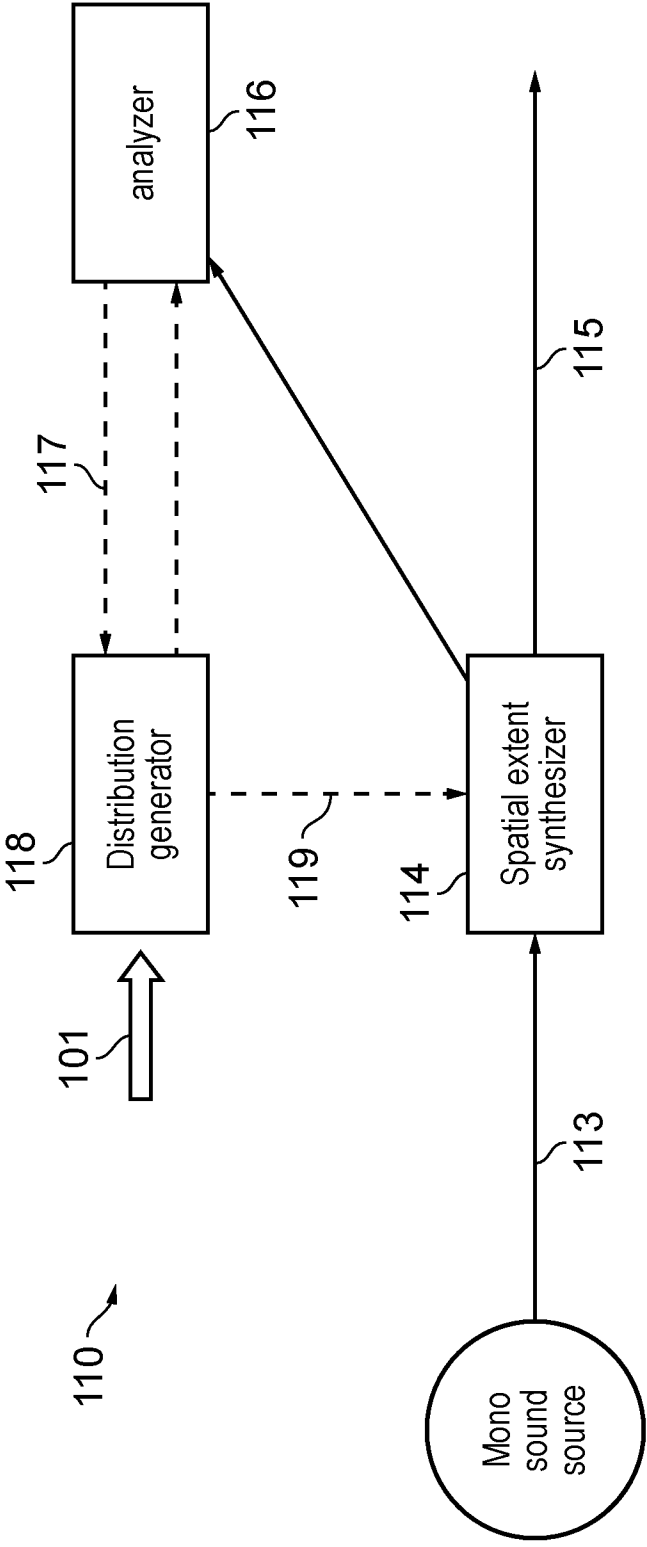


FIG. 8

6/6

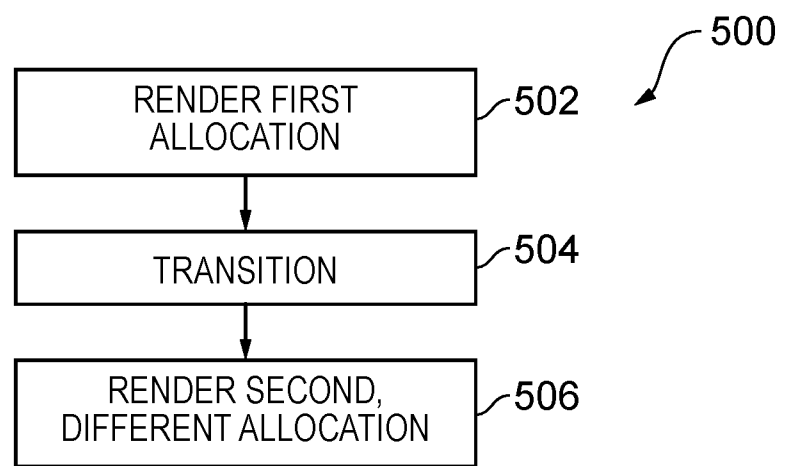


FIG. 9

INTERNATIONAL SEARCH REPORT

International application No.

PCT/FI2018/050289

A. CLASSIFICATION OF SUBJECT MATTER

See extra sheet

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)

IPC: G06F, H04R, H04S

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
FI, SE, NO, DK

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

EPODOC, EPO-Internal full-text databases, Full-text translation databases from Asian languages, WPIAP, XP3GPP, XPAIP, XPESP, XPETSI, XPI3E, XPIEE, XPIETF, XPIOP, XPIPCOM, XPJPEG, XPMISC, XPOAC, XPRD, XPTK, COMPDX, INSPEC, NPL, Internet, ESPACENET.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	GALLO E; TSINGOS N; LEMAITRE G 3D-audio matting, postediting, and rerendering from field recordings. In: Eurasip Journal on Advances in Signal Processing. Hindawi Publishing Corporation [online], 2007-02-26, [retrieved on 2018-07-02]. Retrieved from < https://asp-eurasipjournals.springeropen.com/articles/10.1155/2007/47970 >, <DOI:10.1155/2007/47970>	1-12, 14-27, 29
Y	abstract; sections 3-4, 6, 6.2, 7, 7.1, 7.2, 8.1, 8.3.1, 8.3.3, figures 4-5, 9-10, 13, 15-16.	13, 28



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:	"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
"A" document defining the general state of the art which is not considered to be of particular relevance	"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
"E" earlier application or patent but published on or after the international filing date	"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
"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)	"&" document member of the same patent family
"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	

Date of the actual completion of the international search

17 August 2018 (17.08.2018)

Date of mailing of the international search report

22 August 2018 (22.08.2018)

Name and mailing address of the ISA/FI
Finnish Patent and Registration Office
FI-00091 PRH, FINLAND

Facsimile No. +358 29 509 5328

Authorized officer

Harald Kaaja

Telephone No. +358 29 509 5000

INTERNATIONAL SEARCH REPORT

International application No.

PCT/FI2018/050289

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	PIHLAJAMÄKI, TAPANI; SANTALA, OLLI; PULKKI, VILLE Synthesis of Spatially Extended Virtual Source with Time-Frequency Decomposition of Mono Signals. In: J. Audio Eng. Soc.. [online], 2014-08-22, Vol. 62, No. 7/8, pages 467-484, [retrieved on 2018-08-15]. Retrieved from < http://www.aes.org/e-lib/browse.cfm?elib=17339 >. the whole document, in particular abstract, sections 1.1, 2.4.1-2.4.3, 2.6, 3.5, 5.2, figures 1-4, 6.	13, 28
A	US 2013195276 A1 (OJALA PASI [FI]) 01 August 2013 (01.08.2013) The whole document.	1-29
A	US 2013154930 A1 (XIANG PEI [US] et al.) 20 June 2013 (20.06.2013) The whole document.	1-29
A	US 2016299738 A1 (MÄKINEN TONI HENRIK [FI] et al.) 13 October 2016 (13.10.2016) The whole document.	1-29
A	POLITIS, A. ET AL. Parametric Spatial Audio Effects. In: Proc. of the 15th Int. Conference on Digital Audio Effects (DAFx-12), York, UK. [online], September 2012, [retrieved on 2018-08-17]. Retrieved from < https://www.dafx12.york.ac.uk/papers/dafx12_submission_22.pdf >, The whole document.	1-29

INTERNATIONAL SEARCH REPORT

International application No.

PCT/FI2018/050289

CLASSIFICATION OF SUBJECT MATTER

IPC

H04R 1/26 (2006.01)

H04R 5/02 (2006.01)

H04S 7/00 (2006.01)

G06F 3/01 (2006.01)

G10L 21/0272 (2013.01)

G10L 21/10 (2013.01)

INTERNATIONAL SEARCH REPORT
Information on Patent Family Members

International application No.
PCT/FI2018/050289

Patent document cited in search report	Publication date	Patent family members(s)	Publication date
US 2013195276 A1	01/08/2013	US 9584235 B2	28/02/2017
		CN 102656627 A	05/09/2012
		CN 102656627 B	30/04/2014
		EP 2513898 A1	24/10/2012
		EP 2513898 B1	13/08/2014
		KR 20120098883 A	05/09/2012
		KR 101450414 B1	14/10/2014
		TW 201135718 A	16/10/2011
		TW I490853 B	01/07/2015
		WO 2011072729 A1	23/06/2011
.....			
US 2013154930 A1	20/06/2013	US 9563278 B2	07/02/2017
		CN 103999021 A	20/08/2014
		CN 103999021 B	08/12/2017
		EP 2795432 A2	29/10/2014
		IN 973MUN2014 A	24/04/2015
		JP 2015506035 A	26/02/2015
		JP 6195843 B2	13/09/2017
		KR 20140107484 A	04/09/2014
		KR 101708101 B1	27/02/2017
		WO 2013095783 A2	27/06/2013
.....			
US 2016299738 A1	13/10/2016	EP 2982139 A1	10/02/2016
		WO 2014162171 A1	09/10/2014
.....			