Title: INTERACTIVE SPHERICAL GRAPHICAL INTERFACE FOR MANIPULATION AND PLACEMENT OF AUDIO-OBJECTS WITH AMBISONIC RENDERING.

Abstract: The present invention discloses an audio data processing and a 3D graphical user interface with: Virtual spatial representation of sound sources as objects; manipulation of the objects in the virtual space resulting in a modification of the location and characteristics of the real sound to produce an active soundstage, thanks to a mapping between real sound characteristics and virtual object properties. The virtual active soundstage environment is seamlessly selected through a grid of multiple virtual soundstage environments. The method may be applied to other variables than sound. The method comprises inputting a data element and displaying the data element in a visual format; the data element can then be manipulated in the visual format using predefined gestures to create a modified data element, output aurally as a 3D spatial sound (for example spherical surround or ambisonic (TM) sound).
The present invention relates to a data processing and comprehension method comprising a method of inputting data in a first format and outputting the data aurally. In particular, the invention relates to a method and practice of electronic sound/physical modelling synthesis; whereby the sound is synthesised by interlinked visual and physical variables, which may or may not be in 3-dimensions (3-D), and may or may not be in direct relationship with 3D sound modelling via a 3D graphical user interface (GUI).

One type of musical performance involves playing back recorded music and sounds in a new "mix." The performer is commonly called a disk jockey or DJ. A disk jockey relies on many tools to create a musical performance. For example, vinyl records may be manipulated on multiple turntables, or multiple compact disks (CDs) may be manipulated, to create a variety of sound effects. The disk jockey may vary the speed, pitch, direction, and sequence of playback and may apply a variety of effects to the played back music.

More recently disk jockeys have been using tools that manipulate audio data stored in data files on computers. Some of these tools have been developed based on the idea of computerizing what is usually done with turntables and vinyl records and CD players. However, the flexibility of the computer enables new tools and techniques to be explored for use in musical performances.

Another type of musical performance involves composing, synthesising, sequencing, or ordering, individual digital/electronic or recorded musical parts (varying in length from single beats to entire pieces of music), over time, in a live presentation (set, show, concert, event, etc.). The performer is commonly called an Electronic Dance Music artist (EDM artist). An EDM artist has access to all the same control parameters available to DJs, with more freedom and control over the creation of the individual tracks being played, known as electronic compositions.
Another type of musical performance involves individual musical parts being performed by musicians playing (commonly traditional) musical instruments by following transcribed sheet music. The sheet music contains musical information known as a composition, which is written by a composer. The performance of a composition is led and directed by, what is commonly called, a conductor. In the context of an entire presentation, a DJ, EDM artist, or conductor makes decisions on the sequence of pieces of music over the course of the show. In this context, the artistic value is more relative to curation, than creation.

In the creation of an individual musical piece, a composer transcribes musical parts, which are scored to represent where/when they will be played in the overall musical piece. Each piece (or section) is given a tempo, time signature, and a musical key. Each part is written in sync with the overall tempo, time signature, and key of the piece, so that each individual part fits with the rhythmic, harmonic, and melodic structure of the overall piece. In a classical performance context, the transcript of the compositional parts would be distributed to each of the separate instrumentalists, and a conductor would instruct each instrumentalist where to sit (soundstaging), when and how to play, and when to stop. Each conductor adds his own personal artistry to each composition. In the practices of live popular music, the role of the conductor is omitted. The composition and parts are rehearsed and usually performed, or improvised, without formal direction or use of sheet music (transcripts). In the practices of electronic music production, the separate instrumental parts are arranged on a timeline sequence displayed via a graphics engine, usually on a 2-dimensional graphical user interface. This sequence is similar to classical sheet music. The sequence is first given a set tempo and time signature, and then parts are added to the timeline, usually in individual rows, to compose a piece of music. This piece is usually exported as a digital media file, which is then presented, usually by a DJ, as a piece of recorded music. There is no conductor involvement. In the practices of live electronic production, parts are arranged in 2-dimensional rows and columns. However, an overall piece of music is not pre-composed. Individual parts are composed, but not pre-sequenced to complete an entire piece of music. The parts are stored as individual loops, which are
automatically analysed and processed to match any tempo and time signature set by the composer. These loops are sequenced in real-time during a presentation of music. This is similar to a conductor of music, without the physical placement/positioning of instrumental parts on stage (known as soundstaging).

Soundstaging is the act of creating, or creating the perception, of a soundstage in the presentation of a piece, or pieces, of music. In a traditional context, soundstaging is the arrangement of instrumentalists on stage, in front of a conductor, and/or audience, and how the audience perceives the arrangement. This perception is usually created by the physical placement of the instruments i.e. left or right, rows back or forth, and sometimes height or depth (if they are situated in the orchestra pit). Modern technology allows more control over the delivery of the soundstage via systems comprising of microphones, mixing desks and loudspeakers. Using individual volume, equalisation, and panning controls, as well as, special sound effects (such as Reverb), these systems allow for live control of the overall mix of the sound, as perceived by the audience.

A similar principle is applied to recorded music and electronic music production, where tracks are mixed down to project the desired soundstage to the listener. For example, for a pop track, the vocalist will be at the front/centre of the soundstage, whereas, for a house/techno track, the kick drum will be at the front/centre of the soundstage.

In comparison to traditional soundstages, mixing the soundstage of a track using modern technology is mechanical and linear, and less intuitive to both the mix engineer and recipient(s) of the sound(s), due to the use of faders and knobs to control specific parameters, as opposed to the physical placement of instrumentalists.

The creation of sound can be organic or synthetic. Organic sound production involves traditional instruments made from organic material i.e. wood and metal. Synthetic sound production involves electronic instruments, or synthesisers, made
from electronic components and computer processors. Sounds produced by
traditional musical instruments are embodied by the shape, function, and material
of the instrumental object, and are recognised equally by their visual form.

Electronic synthesis can reform the previous proposition, but limitations are
defined by current music theories, and although sampling is used to add variation,
digitally processed sound is not embodied or connected to any physical context.
Users with less information, prior knowledge, and experience, create sounds that
lack musical refinement due to the required knowledge transfer being unavailable.

Such a knowledge transfer is given with a traditional background in music theory,
which most audio platforms are fundamentally based on. The traditional approach
affords audible and visual refinement of sound, providing multi-dimensional
feedback, so that an audience might also identify and relate to the music being
produced.

Currently available software might require a level of knowledge that would off put
the general audience; this might be, for example, developing an understanding of the
GUI, reduced sensory feedback, unintuitive and arbitrary function, and lack of
natural context. Current technology restricts output through alienation and
exclusivity, for example, 2-dimensional motion input devices (i.e. computer mouse),
text entry interface (computer keyboard), arbitrary 2-dimension GUIs, midi
keyboard controllers, or even keyboards on a synthesis platform – there are
variables on the latter, but these are not all that fluid.

Physical modelling synthesis realises a degree of natural context by providing sets
of real-world variables, such as, material of instrument to synthesis, and type of
gesture used to strike said instrument. This data then translates into relative
synthesised sound. This method is limited to predefined scenarios and known
sounds, and lacks the provision for synthesising original complex
frequencies/sounds.
In addition, current technologies are limited in the type of data they can use as an input for mixing and manipulating sound outputs, and rely heavily on using sound inputs. One example of using visual representations to manipulate pre-recorded sound is US6,898,291. Various parameters of sound (frequency, volume for example) are mapped to parameters of a visual image displayed in a three-dimensional environment. Changing the parameters of the visual image (size, location, texture, for example) change the corresponding parameter of the pre-recorded sound. Other types of inputs are just not envisaged, and therefore data outputs are limited to being aural, and merely being variations of the sounds input to such a system.

The present invention aims to address these problems by providing a data comprehension method, comprising of inputting at least a first data element; displaying the first data element in a visual format; Manipulating the first data element in the visual format using pre-defined gestures to create a modified data element; and outputting the modified data element aurally.

One advantage of the present invention is to be able to be virtually present within the audio mix visually, and navigate around the mix in real-time, at a real-world scale, as if the user is a conductor and the graphical user interface is a soundstage. Another advantage is that rather than using pre-recorded sound, sounds are synthesized from the visual data element, creating the possibility of using a wide variety of inputs.

Preferably, outputting the modified data element aurally comprises outputting as an output sound and presenting the output sound ambisonically.

Preferably the method further comprises the steps of assigning a first object to the first data element; displaying the first data element visually in the form of the first object, such that the pre-defined gestures manipulate the object; and assigning an aural component to the first object.
 Preferably, when more than a first data element is input, assigning an object to each data element and assigning an aural component to each object.

 Preferably, when the object is manipulated, the aural component varies in accordance with the manipulation.

 Preferably the first data element is modified within an environment data element and the method further comprises the steps of: displaying the environment data element in a visual format; and manipulating the environment data element in the visual format using pre-defined gestures to create a modified first data element.

 Preferably the environment data element is selected from the group of: gravitational, atmospheric, meteorological, ambient, terrestrial, celestial, astrological, cosmological, or technological variables.

 Preferably, the method further comprises the steps of: assigning a first area to the environment in which the first data element exists; displaying the environment visually in the form of the first area, such that the first area is the active area, or soundstage; and when more than a first area is input, assigning an inactive area to further areas to create global environment having a navigable grid such that manipulating a position within the grid modifies the active area.

 Preferably, the first data element represents a physical, optical, tactile, odorous or aural variable.

 Preferably, the modified data element is output aurally via at least one sound emitting device.

 Preferably, the first data element represents a sound, and the modified data element or environment forms part of a soundstage.
The visual format may be in 3-D, in which case the gestures are enacted in 3-D. Alternatively, the visual format is in 2-D.

Preferably, the visual format is displayed on a touch-sensitive screen device. Alternatively, the visual format is displayed on a screen, and the first data element is manipulated via a remote touch sensitive device.

Preferably, the object is rendered in the form of voxels, and the gesture influences the voxel nodes.

The present invention also provides a computer program product adapted to perform the above method.

The present invention also provides a computer-readable storage medium or data carrier comprising the computer program.

The present invention will now be described by way of example only, and with reference to the accompanying drawings, in which:

Figure 1 is a flow chart illustrating the method of the present invention;

Figure 2 illustrates how the object position along the Z-axis determines the intensity (loudness) of the sound output;

Figure 3 illustrates how the object position along the X-axis determines the stereo width (panning) left and right of the sound output;

Figure 4 shows the outer side partitions of the 3D environment used to isolate object(s) within a partitioned area to a specific sound output;

Figure 5 illustrates grouped together soundstages, with objects set within each environment and the current active soundstage in the centre;

Figure 6 illustrates map navigation and how the current active soundstage changes dependent on location data;

Figure 7 illustrates an active soundstage can be anywhere on the map;

Figure 8a shows waveforms A and B in phase;

Figure 8b shows waveform B half a cycle forward of waveform A;
Figure 8c shows waveform B half a cycle behind waveform A;
Figure 8d shows waveforms A and B completely out of phase;
Figure 9a shows a sine wave waveform cycle;
Figure 9b shows a square wave waveform cycle;
Figure 9c shows a triangle wave waveform cycle;
Figure 9d shows a sawtooth wave waveform cycle;
Figure 10a shows a 3D spherical object;
Figure 10b shows a 3D cubic object;
Figure 10c shows a 3D conal object;
Figure 10d shows a 3D pyramidal object;
Figure 11a shows a cube that has almost been transformed into a sphere;
Figure 11b shows a cube that has been rounded at the edges;
Figure 11c shows a cube that has been indented on each of the sides from the centre;
Figure 11d shows a cube that has been fully rounded on all vertical sides to form a cylinder;
Figure 11e shows a cube that has been tilted to one side to form a parallelogram shape;
Figure 11f shows a cube that has had one side tilted to form a trapezium shape prism;
Figure 11g shows a cube that has been formed into a hexagonal prism;
Figure 11h shows a cube that has been formed into a multi-sided diamond-like shape;
Figure 11i shows a cube that has been rounded at the edges and corrugated on all sides;
Figure 11j shows a cube that has been formed into a curved pyramid;
Figure 11k shows a cube that has been formed into an indented cone;
Figure 12a shows an equally combined sine waveform and square waveform;
Figure 12 shows a square waveform with rounded corners;
Figure 12c shows an indented square wave;
Figure 12d shows a square wave with a curved top;
Figure 12e shows a square wave that is ramped at the start;
Figure 12f shows a square wave that is sloped to end;
Figure 12g shows a square wave that is ramped at start and end;
Figure 12h shows a square wave that has been ramped at the start and sloped to end;
Figure 12i shows a square wave with ripples at the peak, which may produce a damped oscillation;
Figure 12j shows a square wave that has a ramped curve at the start to peak and a sloped end;
Figure 12k shows a square wave that has a ramped curve from the start to the end;
5 Figure 13 shows two different complex objects;
Figure 14 shows an object and how an object's [3D] dimensions relate to the Cartesian coordinate system;
Figure 15 shows different combinations of cross-section slices of different object surface layer textures;
10 Figure 16a shows a monophonic texture pattern traditionally notated for sheet music;
Figure 16b shows a heterophonic texture pattern traditionally notated for sheet music;
Figure 16c shows a homophonic texture pattern traditionally notated for sheet music;
15 Figure 16d shows a homorhythmic texture pattern traditionally notated for sheet music;
Figure 16e shows a polyphonic texture pattern traditionally notated for sheet music;
20 Figure 17 shows a cross-section slice of an object, which displays each of the layers within the object and their relative distances between each layer;
Figure 18 shows a standard mobile media device, including built in camera, microphone, and light sensor;
Figure 19 shows a diagram of the magnetic interaction relationship between objects; and
25 Figure 20 shows a diagram of an example speaker array completely surrounding a dance floor from all angles/directions.
The present invention provides a data processing and comprehension method, where different types of variables can be converted into sounds, enabling the user to manipulate or process the data to comprehend the data more easily. The present invention utilises the input of at least a first data element, for example, a colour or a sound, and displays this first data element in a visual format, for example, as an image on the screen of a smartphone. This visual display allows the manipulation of the first data element, using pre-defined gestures, for example, by pinching or swiping the image, to create a modified data element. For example, a sound may be displayed visually as a coloured square. The square may be manipulated by changing its size, orientation, or the brightness or contrast of its colour. Once the modification is complete, the modified data element is output aurally. This may be instantaneous or sequentially. This may be through a loudspeaker, a pair of headphones or other sound device. The step of displaying the sound as a coloured square involves assigning a first object to the first data element, so that the first data element is displayed visually in the form of the first object. This first object has a specific aural component assigned to it. For example, a yellow square will always output as the same tone. Unlike prior art systems rather than manipulating pre-recorded sounds the user manipulates a pre-assigned visual element, such as a shape, resulting in the output of an aural element.

When more than a first data element is input, an object is assigned to each data element, and an aural component is assigned to each object. However, many data elements may be assigned to any one object, if desired. When the object is manipulated, the aural component assigned to that object varies in accordance with the manipulation. For example, increasing the size of a yellow square increases the volume of the assigned tone. The modified data element is output aurally via at least one sound outputting device. The outputting of the data as sound enables the user to comprehend the data by means of an ambisonic environment.

The first data element may represent any desirable quantity, for example, a physical, optical, tactile, odorous or aural variable. However, the first data element may
represent a sound, and the modified data element forms part of or is used within the soundstage.

The visual format may be in 3-D, such that the gestures are enacted in 3-D. This greatly enhances the manipulation possibilities, and creates a composition environment in 3-D. Alternatively, the visual format is in 2-D, for example, displayed on a touch-sensitive screen if the visual format is displayed on a screen, and the first data element could be manipulated via a remote touch sensitive device.

The generated and output sound may also be presented in 3D, commonly referred to as ambisonic. The term ambisonic is used to mean all 3D Sound/Periphonic Sound Spatialisation techniques, and includes, but is not limited to, technologies such as Distance-Based Amplitude Panning (DbAP), Vector-Based Amplitude Panning (VbAP), Higher Order Ambisonics, Holophony, Wave Field Synthesis, Virtual Microphone Control (ViMic), and 3-Dimensional Highly Directive Virtual Microphone (3D-VMS).

Typical gestures used in the control of mobile smart devices (such as tap, double tap, swipe, pinch, slide and zoom) can be assigned to data, and used to manipulate data to produce a variety of sounds. For this to happen, data is rendered in the form of voxels, and the gesture influences the voxel nodes. Rendering sounds (and other data) as voxels creates objects that can be manipulated. The data and application flow required to allow the manipulation of the voxel nodes is shown in Figure 1. The term voxels is used to mean all values and units of 3D graphical space, including, but not limited to texels, and any combination such values and units, which may or may not include 2D values and units, such as, but not limited to pixels.

Figure 1 is a flow chart illustrating the method of the present invention. In the flow chart, solid arrows represent the flow within the application, and hollow arrows represent the flow of data. To begin with, a user decides to run the application 1001, and accesses a Start menu 1002. At this point, the user chooses between a Start File Menu 1004 and a Start Settings Menu 1006, depending upon whether it is
desired to start an AV (audio visual) project or create settings for a project. If the Start File Menu is chosen, the user can chose to either construct a new project 1008 or load an existing project 1010. If constructing a new project, an AV project is assigned at 1012, checking current AV project memory data at 1014 and loading the main menu to initialise the current project at 1016. If an existing project is chosen, this is loaded at step 1018 from a search of filenames in the hard disk 1020, the data deserialised at 1022 and fed back to step 1012 where the AV project is assigned, and steps 1014 and 1016 repeated. Once the Main Menu is loaded at step 1024 this can be used to select tools relating to the data display environment at step 1026, sound objects at step 1028 as well as the ingame menu at step 1030 and the settings menu at step 1032. The ingame menu 1030 enables the user to reach the file menu, to save or create new AV projects once sounds have been manipulated. The environment tools accessed at step 1026 can be used, for example, to focus gestures and therefore movements into a brush area on the surface or in the region where gestures will be made. By representing the data at this point in the form of voxels, it is possible to make the voxel nodes appear more solid (at step 1034), less solid (at step 1036) or smoother (at step 1038). Object tools can be used to create new tools or vary the effect a gesture has on a particular display, which can be saved relative to individual projects (at step 1040) or in general (at step 1042). When outputting data aurally in the form of sounds or tones, it may be that a certain amount of reverb is applied (step 1044), or other effects. These steps and variations thereof will become more apparent from the many detailed examples below. Initially the invention will be described in terms of the input data being sound, and the environment in which the sounds are manipulated being represented by soundstages.

Figure 2 shows how object positioning along the Z-axis determines the intensity (loudness) of the sound output. This is a key manner in which the output of the data aurally is managed. Intensity may or may not also be presented as volume, gain, loudness, magnitude, sound pressure level, velocity, and any other relative sound level variable.
Figure 3 shows object positioning along the X-axis, which determines the stereo width (panning) left and right of the sound output. Again, this can be varied to give a wider range of aural data output formats. Stereo height (vertical panning) is applied to object placement along the Y-axis. The Y-axis could or could not also be applied to sound equalisation (EQ) – position coordinates above the centre line apply higher EQ values, whilst position coordinates below the centre line apply lower EQ values. The object size determines how much of an EQ bandwidth is applied i.e. a larger object would apply a larger EQ bandwidth. The frequency bandwidth an object doesn’t occupy will either be reduced or removed. In order to apply two separate bandwidth EQs to a sound (with a cut between), a duplicate object would need to be created, and both objects separately positioned in the respective spaces to apply the desired EQs.

Figure 4 shows the outer side partitions of the 3D environment used to isolate object(s) within a partitioned area to a specific sound output (see Figure 20 below). Further partitioning could be made to include the separate EQ bands if desired (or in another/EQ mode). The size and amount of partitions may be changed. The output assignment may be changed. The central un-isolated space uses all outputs, similar to surround sound stereo, which will present the output sound relative to the placement of the object within the un-isolated environment. Objects can be automated to follow defined paths to generate synchronous ambisonic panning effects.

Figure 5 shows grouped together soundstages, with objects set within each environment and the current active soundstage in the centre. This grouping forms a navigable map (as shown in Figure 6). Figure 6 shows map navigation and the current active soundstage change dependent on location data.

Figure 7 shows that the active soundstage can be anywhere on the map, irrespective of territorial boundaries. All objects outside of the active soundstage area are stopped. All objects within the active soundstage area are played.
Figure 8 shows two waveforms, A and B. In Figure 8a the first waveform relationship shows both A and B in perfect phase, Figure 8b shows B forward half a cycle, Figure 8c shows B behind half a cycle, and Figure 8d shows both A and B completely out of phase.

Figure 9a shows a sine wave waveform cycle, Figure 9b shows a square wave waveform cycle, Figure 9c shows a triangle wave waveform cycle and Figure 9d shows a sawtooth wave waveform cycle.

Figure 10a shows a 3D spherical object, Figure 10b shows a 3D cubic object, Figure 10c shows a 3D conal object, and Figure 10d shows a 3D pyramidal object. A) If an object is spherical/ellipsoidal, the waveform will be sinusoidal; B) If an object is cubic, the waveform will be a square wave; C) If an object is equilateral or isosceles prismatic and/or has rounded sides, then the waveform will be a triangle wave; D) If the shape is right-angle prismatic and/or has flat sides, then the waveform will be sawtooth.

Figure 11 shows a group of objects that originally started out as cubes and have been changed in various ways. A cube is used as a base shape to serve as an example only. Figure 11a shows a cube that has almost been transformed into a sphere. Figure 11b shows a cube that has been rounded at the edges. Figure 11c shows a cube that has been indented on each of the sides from the centre. Figure 11d shows a cube that has been fully rounded on all vertical sides to form a cylinder. Figure 11e shows a cube that has been tilted to one side to form a parallelogram shape. Figure 11f shows a cube that has had one side tilted to form a trapezium shape prism. Figure 11g shows a cube that has been formed into a hexagonal prism. Figure 11h shows a cube that has been formed into a multi-sided diamond-like shape. Figure 11i shows a cube that has been rounded at the edges and corrugated on all sides. Figure 11j shows a cube that has been formed into a curved pyramid. Figure 11k shows a cube that has been formed into an indented cone.
Figure 12 shows the waveform shapes created by such shape manipulation, as described in Figure 11 – as in Figure 11 a cube is used as an example to explain shape manipulation, hereby a square wave is used as the base waveform shape to demonstrate the relationship between the two. Therefore, the following figures are all modified square waves, for example purposes only, even though some may more closely resemble that of sine, triangle, or sawtooth waves. Figure 12a shows an equally combined sine waveform and square waveform, which may produce a large amount of high frequency loss. Figure 12b shows a square waveform with rounded corners, which may produce a small amount of high frequency loss. Figure 12c shows an indented square wave, which may produce an amount of low frequency loss. Figure 13d shows a square wave with a curved top, which may produce a low frequency boost (accentuated fundamental). Figure 12e shows a square wave that is ramped at the start, which may produce total high frequency loss. Figure 12f shows a square wave that is sloped to end, which may produce a low frequency phase shift. Figure 12g shows a square wave that is ramped at start and end, which may produce both a high and a low frequency loss. Figure 12h shows a square wave that has been ramped at the start and sloped to end, which may produce a high frequency loss and low frequency phase shift. Figure 12i shows a square wave with ripples at the peak, which may produce a damped oscillation. Figure 12j shows a square wave that has a ramped curve at the start to peak and a sloped end, which may produce high frequency loss and low frequency phase shift. Figure 12k shows a square wave that has a ramped curve from the start to the end, which may produce low frequency loss and low frequency phase shift.

Figure 13 shows two different complex objects. Such complex objects may be included in a scene, included in a soundstage, or imported to the system via any of the aforementioned hardware input devices, including virtual reality headset spatial awareness to include objects present in the user’s real environment, computer vision analysis of 2D or 3D images. Figure 13a shows an animated motor vehicle. A system analysis of this object for sound production would include; object recognition = one car and no background, monochromatic colour – majorly range is grey, black, white; smooth reflective surface; rounded, curved shapes. The output
sound production for this image may be relative to one (layer) atonal sine wave with added white noise and reverb. Other visual effect analysis may be included (that cannot be represented in the diagram), such as, brightness, which would also apply relative effect to sound output. Figure 13b shows an animated (pale green) house. A system analysis of this object for sound production would include; object recognition = house, door, windows; colour range = red, green; multiple surfaces, majority non-reflective and rough; straight square and triangular shapes. The output sound production for this image may be relative to a complex sound made up of two (layers) square and triangle wave frequencies at different musical tones – ##C1, A#3##. Other visual effect analysis may be included (that cannot be represented in the diagram), such as, brightness, which would also apply relative effect to sound output.

Figure 14 shows an object and how an object's [3D] dimensions relate to the Cartesian coordinate system. X-axis relates to object Width; Y-axis relates to object Height; Z-axis relates to object Depth. The centre point of the Cartesian coordinate marker remains the centre of the object – plus (+) and minus (-) lengths of each axis can be changed independently from this centre point.

Figure 15 shows different combinations of cross-section slices of different object surface layer textures. These diagrams describe the relation of object surface layer texturing combinations in a defined proximity of one another and the possible resulting compositional texturing. a: Monophony: If there is only one object and the surface is set to default (no texture) then the sound/music will be monophonic i.e. remain the way it is originally notated by the composer. b: Heterophony 1: If there is only one object layer and the surface is set to have texture then the sound/music will be heterophonic. c: Heterophony 2: If there is more than one object and the surface of one object is set to have no texture, but the other objects are set to have different textures, then the sound/music may become heterophonic. d: Homophony: If there is more than one object and all surface layers are set to have no texture, then the sound/music may become homophonic. e: Homorhythm: If there is more than one object and all surface layers are set to have texture, but all
the textures are the same, then the sound/music may become homorhythmic. If there is more than one object and all surface layers are set to have texture, but all the textures are different, then the sound/music may become polyphonic.

Figure 16a shows a monophonic texture pattern traditionally notated for sheet music. Monophony consists of melody without accompanying harmony. Hence, this is created by one untextured master layer. Figure 16b shows a heterophonic texture pattern traditionally notated for sheet music. Heterophony is characterized by the simultaneous variation of a single melodic line. Such a texture can be regarded as a kind of complex monophony in which there is only one basic melody, but realized at the same time in multiple voices, each of which plays the melody differently, either in a different rhythm or tempo, or with various embellishments and elaborations. Hence, this is created by, either, (1) one textured master layer, or, (2) an untextured master layer being accompanied by layers of varying textures. Figure 16c shows a homophonic texture pattern traditionally notated for sheet music. In homophony two or more parts move together in harmony, the relationship between them creating chords. Hence, this is created by multiple untextured (monophonic) layers. Figure 16d shows a homorhythmic texture pattern traditionally notated for sheet music. Homorhythm is similar to homophony, however, is only rhythmically similar, not harmonically. Hence, this is created by multiple textured layers, but all textures being similar. Figure 16e shows a polyphonic texture pattern traditionally notated for sheet music. Polyphony consists of two or more simultaneous lines, all of independent melody. Hence, this is created by multiple textured layers, with all textures being different.

Figure 17 shows a cross-section slice of an object, which displays each of the layers within the object and their relative distances between each layer. The centre distance is the size of the core. The distance between each layer may be changed. The distance between each layer determines the audible level of each layer in relation to each layer present in the object. The surface layer is always the master level. If any two layers are close together, their levels will be equal, but the further
apart from the surface layer, the lower the level will be in relation. The order of the layers may or may not be changed accordingly for mixing purposes.

Figure 18 shows a standard mobile media device, including built in camera, microphone, and light sensor. Many mobile devices also feature GPS location services, Bluetooth, Wi-Fi, and accelerometers. Such I/Os and I/O data may be used as control data in the generation of sound in such a method and processes, as described within this system. For example, a live feed of facial tracking to recognise a busy environment, location data to determine urban or rural landscapes, light sensitivity to control various effects, etc.

Other sensors may be used as inputs to control the data element(s). This may include, but is not limited to, computer vision, kinetic/motion, gesture, and biofeedback - namely such sensors may include, but are not limited to, Leap Motion, which uses infrared LEDs for precise 3D hand and finger recognition and tracking; Google Soli, which uses radar technology that can track sub-millimetre hand and finger motions at high speed and accuracy; Microsoft Kinect/Softkinetic DepthSense, which uses Time-of-Flight CMOS sensors for 3D full-body recognition and tracking; Myo, which uses mechanomyogram (MMG) sensors that observe muscle activity in the forearm(s) for hand and finger movement and gesture recognition and tracking.

Figure 19 shows a diagram of the magnetic interaction relationship between objects. Objects may be given magnetic properties of varying strengths, either with a plus (+) or a minus (-) polarization. When objects are in proximity of such fields, this determines whether they are attracted to or repelled away from one another. The size of such proximal area is determined by the strength of the magnetic field applied to any object.

Figure 20 shows a diagram of an example speaker array completely surrounding a dance floor from all angles/directions. The polygons represent speakers that may each be full speakers systems (providing all frequency ranges, including high, mid,
low, sub), or may be separated by EQ ranges i.e. high ranges, mid ranges, low ranges, sub ranges.

These and other features of the invention will now be illustrated by the following examples.

Example 1: Ambisonic Soundstage

In this example, the present invention provides a system for sound mixing and sequencing in 3-dimensions (3D), in real-time, whereby sound sources are visually displayed via a graphics engine in 3D, within a 3D environment.

The processes described herein refers to 3-dimensions, but may or may not be represented in 2-dimensions, digitally displayed and modified digitally via a sound and graphics engine, or calculated linearly through transcription. The 3-dimensional environment herein is described using a Y-up, left-handed (Cartesian) coordinate system, but could be applied to any similar method.

Sensor data may or may not be used to either control data, or as input data. This may or may not be camera data, microphone data, location data, biometric data, or uploaded multimedia. Non-musical data from such sensors and/or media, such as, motion, gesture, position, orientation, location, ambient noise, face tracking, heart rate, colour analysis, contrast, etc. may or may not be used as input data for the invention to generate further manipulation.

The entire environment is a virtual soundstage. Each object's position within an active soundstage can be moved around, which is output in real-time via a sound system. Manipulation of object position can control sound output panning and intensity. Positioning of objects in the environment can control sound frequency equalisation. The outer sides of the 3D environment can be partitioned to isolate object(s) within a partitioned area to a specific sound output or EQ. An example output arrangement may represent a dome of tiered speaker arrays/rings
surrounding a dance floor or concert hall. This allows the user to mix objects within isolated sources and points of the soundstage.

Each virtual soundstage represents real-world environmental conditions, which can be controlled. These conditions affect variables of the sound, as they would in reality. Such variables include, but are not limited to atmospheric pressure, temperature, and humidity.

The size of each soundstage space can be controlled. The size of the space controls a reverb effect that is applied to each object within the space. A larger space will apply a wide reverb, whereas a smaller space will apply a narrow reverb. The overall perceived effect on the sound will be dependent on each object position. Each space can also be set as open or closed, which sets the reverb to open or closed reverb. The type of reverb applied is representative of natural reverb that would occur in a similar real-world environment.

Completed soundstages can be grouped together to form an extended environment map, or grid, which can be navigated in an upward, downward, sideways, or diagonal direction, like a virtual open world. As the user navigates around the map, the current active soundstage changes dependent on the user’s current location. Through this navigation, the user can sequence entire presentations of music.

Example 2: Visual and Physical Modelling Synthesis

The invention may be used as a process, method, and/or system for visual-physical modelling synthesis of complex sounds using frequencies of visual light and properties of physical matter, and natural or unnatural user and environmental conditions in 2-dimensions and/or 3-dimensions, whereby the visual, matter, and shape data embodies the sound to be produced (similar to traditional instruments).

Visual and physical data, such as hue, brightness, weight, and shape, are analysed and used by the system’s algorithm to process and synthesise sound, using non-instrumental/non-musical values to create objects for physical modelling synthesis.
The data processing described herein refers to 3-dimensions, but may or may not be represented in 2-dimensions, digitally displayed and modified digitally via a sound and graphics engine, or calculated linearly through transcription. Sensor or media data may or may not be used to either control data or as input data. This may include, but not limited to, motion and gesture data, camera data, microphone data, location data, biometric data, electroencephalography (EEG) data, mechanomyogram (MMG) data, uploaded multimedia, and/or any combination of the aforementioned. Non-musical data received by such sensors and/or media, such as, motion, gesture, position, orientation, location, ambient noise, face tracking, heart rate, colour analysis, contrast, etc. may or may not be used as input data for the invention to generate audio-visual synthesis output and/or for further manipulation. This could, for example, range from controlling the system via a motion sensor by hand within a 3D virtual reality (VR) world, or analysing a photographic image or live camera feed to generate sound.

Direct visual and physical sound modelling relationships are discussed below:
<table>
<thead>
<tr>
<th>VISUAL LIGHT</th>
<th>SOUND</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colour/hue</td>
<td>Note/tone/semitone +octave</td>
</tr>
<tr>
<td>Volume</td>
<td>Amplitude</td>
</tr>
<tr>
<td>Layers</td>
<td>Additive synthesis</td>
</tr>
<tr>
<td>Brightness/luma</td>
<td>White noise</td>
</tr>
<tr>
<td>Contrast</td>
<td>Phase</td>
</tr>
<tr>
<td>Tint</td>
<td>High pass filter</td>
</tr>
<tr>
<td>Shade</td>
<td>Low pass filter</td>
</tr>
<tr>
<td>Saturation</td>
<td>Accuracy/syncopation</td>
</tr>
<tr>
<td>Invert</td>
<td>Frequency inversion</td>
</tr>
<tr>
<td>Definition/pixel rate</td>
<td>Fidelity/bit rate</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PHYSICS</th>
<th>SOUND</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>Velocity, tremolo/amplitude modulation</td>
</tr>
<tr>
<td>Density</td>
<td>Oscillation/intensity</td>
</tr>
<tr>
<td>State</td>
<td>Vibrato/frequency modulation</td>
</tr>
<tr>
<td>Shape</td>
<td>Waveform shape</td>
</tr>
<tr>
<td>Texture</td>
<td>Harmonic-rhythmic-melodic texture</td>
</tr>
<tr>
<td>Width (R/L)</td>
<td>Stereo width</td>
</tr>
<tr>
<td>Height (U/D)</td>
<td>Stereo height, decay (U), release (D)</td>
</tr>
<tr>
<td>Depth/length (F/B)</td>
<td>Length of sustain</td>
</tr>
<tr>
<td>Reflection</td>
<td>Reverb</td>
</tr>
</tbody>
</table>
i) Light-to-Sound Synthesis

Sound is synthesised from variables of light frequencies and visual effects. These are broken down into individual parameters of visual light frequency, which translate and control variables that synthesise audible sound frequencies. More than one frequency can be used; each individual frequency is hereby called a layer(s). Each layer adds a signal/tone that can be individually edited and controlled as described throughout this document. This is similar to additive synthesis. These layers are interconnected and serve as part of a whole object.

Colour/Hue sets the actual frequency of the base tone. This may or may not be more or less the range shown below, or may or may not be more or less a full 256 RGB or CMYK colour range. Ranges of colours larger, or smaller, than what is shown below will be spread, or condensed, between the colour bands shown below, however, such bands remain within their colour of the spectrum i.e. in a colour range of 256 colours, all sound frequency between the musical note range F3 to A3 would still remain within the yellow spectrum, but in this instance Yellow would be divided into different tones of yellow.

<table>
<thead>
<tr>
<th>Colour</th>
<th>Colour Frequency (THz)</th>
<th>Musical Note Range</th>
<th>Octave</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>None</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Infrared</td>
<td>3-400</td>
<td>C0-B0</td>
<td>0</td>
</tr>
<tr>
<td>Red</td>
<td>400-484</td>
<td>C1-A#2/Bflat2</td>
<td>1-2</td>
</tr>
<tr>
<td>Orange</td>
<td>484-508</td>
<td>A#2/Bflat2-E3</td>
<td>2-3</td>
</tr>
<tr>
<td>Yellow</td>
<td>508-526</td>
<td>F3-A3</td>
<td>3</td>
</tr>
<tr>
<td>Green</td>
<td>526-606</td>
<td>A#3/Bflat3-G5</td>
<td>3-5</td>
</tr>
<tr>
<td>Blue</td>
<td>606-668</td>
<td>G5-B6</td>
<td>5-6</td>
</tr>
<tr>
<td>Indigo</td>
<td>668-715</td>
<td>B6-C8</td>
<td>6-8</td>
</tr>
<tr>
<td>Violet</td>
<td>715-789</td>
<td>C8-B8</td>
<td>8</td>
</tr>
<tr>
<td>Ultraviolet</td>
<td>789-3000</td>
<td>C9-D#10/Eflat10</td>
<td>9</td>
</tr>
<tr>
<td>White</td>
<td>All</td>
<td>All</td>
<td>All</td>
</tr>
</tbody>
</table>
Non-colours, black and white, and non-visual colours, infrared and ultraviolet, are included to represent disparate sounds. White contains all possible frequencies, which represents white noise. Black has no frequency and represents atonal sounds, such as, but not limited to drum hits. Infrared contains very low-end frequencies down to frequencies below human hearing. Ultraviolet contains extreme high-end frequencies, up to those above human hearing.

The volume of a colour (of a predefined range 0–maximum) determines the amplitude of a layer (between a predefined range 0–maximum). The visual Brightness/Luma of a colour (of a predefined range 0–maximum) determines the amount of additional white noise (between a predefined range 0–maximum) added to a layer, or object. The visual tint effect of a colour (of a predefined range 0–maximum) determines an amount of High Pass Filter (HPF) applied (between a predefined range 0–maximum) to a layer, or object. The maximum amount of effect applied will fully kill the audible sound so that nothing can be heard. The visual Shade effect of a colour (of a predefined range 0–maximum) determines the amount of Low Pass Filter (LPF) applied (between a predefined range 0–maximum) to a layer, or object. The maximum amount of effect applied will fully kill the audible sound so that nothing can be heard. The visual saturation effect of a colour (of a predefined range 0–maximum) determines the amount of Band Pass Filter (BPF) applied (between a predefined range 0–maximum) to a layer, or object. The maximum amount of effect applied will fully kill the audible sound so that nothing can be heard. The visual contrast effect parameter (of a predefined range +100% and -100% of 0) controls the phase relationship between two or more layers of an object, by moving their cycle forwards, or backwards. If more than one layer is used, this can be used to move each layer in and out of phase. The visual sharpness/acutece effect parameter (of a predefined range +100% and -100% of 0) determines the accuracy of a layer in relation to the musical time (tempo). When set at 0, the layer will be in sync with the tempo; when set at +100%, the layer will be a full musical beat ahead of the tempo; when set at -100%, the layer will be a full musical beat behind the tempo – the difference between the percentage range
respectively represents such time difference offsets i.e. +50% would set the layer ahead of the tempo by half a beat, whereas, -25% would set the layer behind a quarter of a beat. This could be viewed as syncopation (delay and anticipation).

The visual invert effect inverts all layers of an object. This flips each layer to a polar opposite frequency between 20HZ and 20,000HZ – the centre point for such mirroring is 9990HZ. This is similar to frequency inversion.

Visual definition (pixel rate) controls the fidelity (bit rate) of a layer, or object. Higher pixel rate and better aspect ratios provide higher audio bit rates, and vice versa. Lowering the definition to blur or distort the visual applies degrees of distortion to a layer, or object.

The distance between layers may or may not be controlled, which controls the audio level (not to be confused with volume/amplitude) relationship between each of the layers present in any one object.

ii) 3D Waveshape Modelling

The shape of an object directly controls the audio waveform shape. The basic/fundamental shapes and waveform shapes are:

<table>
<thead>
<tr>
<th>Sine/sinusoidal wave</th>
<th>Curved/rounded (spherical)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Square wave</td>
<td>Square (cuboidal)</td>
</tr>
<tr>
<td>Triangle wave</td>
<td>Triangular, even (prismatic)</td>
</tr>
<tr>
<td>Sawtooth wave</td>
<td>Triangular, edged (prismatic)</td>
</tr>
</tbody>
</table>

These shapes are the basis to how shape data of objects translate to the output sound/waveform. This is not limited to basic shapes, and may or may not include 3D shapes. Other considerations may include polyhedrons, superquadrics, superellipsoids, supertoroids, and fractaloids. Any complex shape is generally a combination of basic shapes. Similarly, more complex waveform shapes are formed.
from combinations of the basic waveform shapes. If such shapes are used, then respective complex waveform shapes will be generated.

Other known systems provide the option of selecting basic pre-defined 2D shapes to set and output a corresponding predefined basic audio waveform shape. The present invention provides fluid modelling of 3D animated shapes, which are user-defined and can be infinitely graphically complex, and include motion/animation (change shape over time), which are instantaneously calculated, set/modulated, and output as a corresponding variable complex audio waveform shape.

iii) Physical Timbre Modelling

The system provides a method where physical properties of matter determine the timbre aspects of sound. Physical properties can be assigned to objects that translate as properties of sound.

The physical weight data of an object (between a predefined range 0–maximum) determines the speed of attack of the amplitude envelope (velocity) and amount of amplitude modulation (AM)/tremolo applied (between 0–maximum).

The physical state data of an object (between a predefined range 0–maximum) determines the amount of frequency modulation (FM)/vibrato applied (between 0–maximum). If an object(s) is used in a live environment/active soundstage, such as that described in Ambisonic Soundstage above, where a temperature control is applied, this can change the state of an object i.e. heating above a set melting/boiling point may melt a solid object into a liquid state, or boil a liquid object into a gas state, or cooling below a set freezing point may freeze a liquid into a solid state.

The physical density data of an object determines (between a predefined range 0–maximum) the amount of oscillation/intensity applied to AM and FM (between 0–maximum). The physical width data of an object (between a predefined range 0–maximum) determines the stereo width of the sound (between 0–maximum). This
is measured both left and right of the centre point of an object. Distance measurements left and right may or may not be adjusted independently.

The physical height data of an object (between a predefined range 0–maximum) determines the stereo height of the sound, and the length of decay and release of the amplitude envelope (between 0–maximum). This is measured both up and down of the centre point of an object. The upward measurement determines the length of decay, while the downward measurement determines the length of release. Distance measurements both up and down may or may not be adjusted independently, although only by the overall height measurement of an object affects stereo height.

The physical depth (or length) data of an object (between a predefined range 0–maximum) determines the length of sustain of the amplitude envelope (between 0–maximum). This is measured both back and fourth of the centre point of an object. Distance measurements both backward and forward may or may not be adjusted independently, although only the overall distance measurement affects the length of sustain.

Other first data elements may include, but is not limited to, viscosity, half-life, buoyancy, or absorbency.

iv) Physical Texture Composition

The physical/visual surface texture of layer(s)/object(s) determines the added (musical) harmonic, rhythmic, and melodic texture of objects in close proximity to one another. Such textures may include, but not limited to, flat/smooth (no texture), or grainy, course, or rough. The number and/or combination of object/layer surface textures within specific vicinity determine such added musical texture. This proximity can be user defined and is also directly affected by magnetism and other naturally occurring relationships between objects and conditions of the environment.
The amount/level of visual texturing applied to layer(s)/object(s) may or may not be increased or decreased, and such levelling may or may not have an affect on the level of complexity and texturing added to the sound/music. This may or may not also include contrapuntal texturing, such as, but not limited to, canon and fugue.

v) Image-to-Sound Effects

Other visual effects may be used to affect objects, which may or may not be a combination of the aforementioned variables. These may or may not include, and not be limited to, reflection, strobe/pulsate, and hollowness. These may or may not apply such sound effects as, but not limited to, flanger, phaser, chorus, reverb, transform, panning, and echo. Flanger, phaser, and chorus are all applied through layering and controlling layers in different ways. If a delayed layer is added to an original (master) layer with a continuously variable delay (usually smaller than 10ms), then one layer is slowed, so that it becomes out-of-phase with its partner, and produces a phasing effect, this creates a Flanger effect. Moving layers in and out of phase with the master makes the phasing effect appear to slide up the frequency spectrum. This phasing up-and-down the register can be performed rhythmically.

If a layer(s) is split, and a portion is filtered with an all-pass filter to produce a phase-shift, and then the unfiltered and filtered signals are mixed, this produces a Phaser effect. If a delayed layer is added to a master layer with a constant delay, and the delay is short, in order not to be perceived as echo, but above 5ms to be audible – If the delay is too short, it will destructively interfere with the un-delayed layer, and create a flanging effect – the delayed signals will be slightly pitch shifted to more realistically convey the effect of multiple voices, this produces a Chorus effect.

The reflection of a surface of a layer determines the amount, and type, of reverberation applied to a layer, or an object. Such surface type effects may or may not include, but not limited to, diffusion, emission, specular, and opacity. Such reverb sound effect parameters may or may not include, but not limited to, size, outdoor/indoor, and delay. A strobe/pulsate light effect applied to an audio
transform will affect the sound at different intervals depending on the rate of strobe. This could be applied to a beat ratio of 1:1, 1:2, 1:4, 1:18, or 1:16. The Hollowness of the core an object determines an amount of audio DSP effect, echo, which is applied – more echo is applied to objects that are hollower.

5

vi) Physics

A magnetism value can be applied to objects. This determines how objects may interact in proximity to other objects, if used in a live environment/active soundstage (such as described in Ambisonic Soundstage). Polarisation would mean objects would either attract or repel. The strength of Magnetism applied would determine the force and distance such an effect would have on neighbouring objects.

The general environment represents a [virtual] physical environment, to which real-world laws, physics, variables, and/or conditions called environment data elements can be set, which manipulate first data elements/objects within the environment. An environment data element represents either an environmental condition that defines the physical environment a first data element/object may be positioned in, or an environmental force/condition that acts upon a first data element/object and/or the physical environment simultaneously, for example, this may be a terrestrial environment data element, such as the shape or state of the landscape, the first data element(s) is in, or a meteorological environment data element, such as a rain shower, that acts upon the first data element within the environment. The environmental data element may be displayed in a visual format, such that manipulating the environment data element in the visual format using pre-defined gestures creates a modified first data element/object. This is analogous with the method outlined above in Example 1, where the data elements are modified using pre-determined gestures whilst the data element is in a visual format. Environment data elements may be manipulated to affect objects within the environment, which, for example, may be selected from the group of; gravitational, atmospheric, meteorological, ambient, terrestrial, celestial, astronomical, astrological, cosmological, or technological variables.
The environment can also be used to create a navigable map or grid, either by dividing the global environment into separate soundstage areas, or by generating new soundstage areas, which can be connected to the existing environment, enabling the user to move around the environment/areas, in an upward, downward, sideward, or diagonal direction, such that the global environment is a 3D open world. To achieve this, the method of the present invention involves assigning a first area to the environment. Next is displaying the first area visually. This causes the first area to become the active soundstage, or the active area. When more than a first area is created or input, it is possible to assign inactive areas to the other areas. Using both the connected active area and inactive areas creates a navigable map, or grid, and is collectively known as the global environment. Generally, only one soundstage environment area is active at any time, which is determined by the user's location within the global environment. The user perspective is generally at front-centre of the main/selected/active soundstage – a similar perspective to that of the position of a conductor, or mix engineer, of an orchestra, or stage, in a live performance situation. As the user moves around the map/grid, the active soundstage area changes in response. The active area can be modified either by changing position within the navigable map or grid, or by increasing or decreasing the size of the scope of the active soundstage/area; Increasing the size of scope activates more of the global environment, and creates less inactivity – Decreasing the size of the scope deactivates more of the global environment, and creates more inactivity. The use of active and inactive soundstages, analogous to the concept of the active and inactive areas, is described above with reference to Figures 5 and 6.

The connected areas represented and described are in 2D and relative to a real-world 2D map or atlas of a 3D planet, such that any navigation off the represented 2D map in any direction would place the location point at a polar opposite position on the map.

A gravitational environment data element value may be applied to the global environment, soundstage areas, or specified sections of a soundstage. The gravity value would affect objects within the selected area(s) as it would naturally in a real-
world environment, dependant on the properties of the object(s). For example, this may be, but not limited to, gravity becoming lighter and object(s) floating, or gravity becoming heavier and object(s) dropping – objects may or may not float out of the audible area of the environment, or, may rest or bounce off the floor depending on the gravity value applied and the material value given to the object and the floor of the environment. Other environmental values may affect such behaviours, for example, but not limited to, the state and material of the environment itself.

An ambient temperature environment data element value may be applied to the global environment, soundstage areas, or specified sections of a soundstage. The temperature value would affect objects within the selected area(s) as it would naturally in a real-world environment, dependent on the properties of the object(s). For example, this may be, but not limited to, temperature decreasing and object(s) condensing or solidifying, or temperature increasing and object(s) melting or evaporating. Other environmental values may affect such behaviours, for example, but not limited to, the state and material of the environment itself.

Other atmospheric environment data elements may include, but not limited to; pressure, humidity, and atmospheric electricity, or meteorological environment data elements, including weather conditions that may modify the ambient and atmospheric environment data elements. Whereby the object(s) and output sound(s) will be naturally affected relative to such first data elements adopting such environment data element changes as they would in a real-world environment.

Further environment data elements may or may not also include natural and unnatural phenomenon, such as, but not limited to; technological, such as, destructive weapons (i.e. electromagnetic pulse, bombs) and mechanical devices; terrestrial, such as, natural disasters; celestial, astrological, or astronomical, such as, supernova explosions/implosions, lunar changes, planetary alignments, ultraviolet radiation, solar flares. A planet, sun, or moon may be selected as a environment data element, so that the atmosphere closely resembles that observed through scientific study; cosmological, such as, molecular scientific experiments, which may
include colliding atoms and mimicking real-world results by colliding objects and sounds.

Different areas of the global environment or soundstages may be partitioned or user-defined and assigned different environment data elements.

Odours may be manipulated, for example, by assigning visual elements to scents, aromas or chemical representations of smells.

Software mechanics like artificial intelligence (AI) characteristics and behaviours may be applied to objects. Such characteristics may include proximal defence and attack response mechanisms – AV outputs would be defined by the combination of present audio and visual variables and the interaction between two or more objects being in proximity/contact. This may also include applying predefined paths for objects to follow, as described in Fig. 4. Other such AI may be relative to environmental changes, such as temperature and humidity, which may change or fluctuate over a predefined time or event.

Animation, motion graphics, and or visual effects may be applied to Objects, which may include changes to any of the aforementioned definable variables of the Object, over a pre-defined time, subsequently applying automation to the aural output.

The described environment refers to standard conditions, such as, but not limited to, room temperature and pressure, and Earth’s gravitational force and oxygen-rich gaseous environment. However, the environment atmosphere can be fully modified, for example, but not limited to, any of the aforementioned properties/conditions, as well as, state and matter – solid(s), liquid(s), gas(es) – which may include, but not limited to, natural or unnatural/real or imaginary elements/materials (i.e. red mercury). Similarly, objects residing in the environment are affected by such changes as they would be affected in nature.
A computer program product adapted to perform the above method, or a computer-readable storage medium or data carrier comprising the program may be provided.

These and other embodiments will be apparent to those skilled in the art.
CLAIMS

1. Data comprehension method, comprising:
   Inputting at least a first data element;
   Displaying the first data element in a visual format;
   Manipulating the first data element in the visual format using pre-defined gestures to create a modified data element; and
   Outputting the modified data element aurally.

2. Method of claim 1, wherein outputting the modified data element aurally comprises outputting as an output sound and presenting the output sound ambisonically.

3. Method of claim 1 or 2, further comprising the steps of:
   Assigning a first object to the first data element;
   Displaying the first data element visually in the form of the first object, such that the pre-defined gestures manipulate the object; and
   Assigning an aural component to the first object.

4. Method of claim 3, wherein when more than a first data element is input, assigning an object to each data element and assigning an aural component to each object.

5. Method of claims 2, 3 or 4, wherein when the object is manipulated, the aural component varies in accordance with the manipulation.

6. Method of any preceding claim, wherein the first data element represents a physical, optical, tactile, odorous or aural variable.

7. Method of any preceding claim, wherein the first data element is modified within an environment data element, and wherein the method further comprises the steps of:
Displaying the environment data element in a visual format; and
Manipulating the environment data element in the visual format using pre-defined gestures to create a modified first data element.

5 8 Method of claim 7, wherein the environment data element is selected from the group of: gravitational, atmospheric, meteorological, ambient, terrestrial, celestial, astronomical, astrological, cosmological, or technological variables.

9. Method of claims 7 or 8, further comprising the steps of:
Assigning a first area to the environment in which the first data element exists;
Displaying the environment visually in the form of the first area, such that the first area is the active area, or soundstage; and
When more than a first area is input, assigning an inactive area to a further area to create a global environment having a navigable grid such that manipulating a position within the grid or the scope of the active area within the grid modifies the active area; and
Outputting data elements within the active area aurally.

10. Method of any preceding claim, wherein the modified data element is output aurally via at least one sound emitting device.

11. Method of any of claims 1 to 6, wherein the first data element represents a sound, and the modified data element forms part of a soundstage.

12. Method of any preceding claim, wherein the visual format is in 3-D.

13. Method of claim 12, wherein the gestures are enacted in 3-D.

14. Method of any of claims 1 to 12, wherein the visual format is in 2-D.
15. Method of claim 14, wherein the visual format is displayed on a touch-sensitive screen device.

16. Method of claim 14, wherein the visual format is displayed on a screen, and the first data element is manipulated via a remote touch sensitive device.

17. Method of any of claims 8 to 16, wherein the object or environment is rendered in the form of voxels, and the gesture influences the voxel nodes.

18. A computer program product adapted to perform the method of any of claims 1 to 17.

19. A computer-readable storage medium or data carrier comprising the program of claim 18.
FIG. 15

a.

b. {REMOVED}

c.

d.

e.

f.
FIG. 16

a.

b.

Oboe

Cello

C.

Oboe

Organ

Cello
**INTERNATIONAL SEARCH REPORT**

**INTERNATIONAL APPLICATION NO.**
PCT/GB2015/053355

**A. CLASSIFICATION OF SUBJECT MATTER**

**INV.** H04S7/00

**ADD.** G10H1/00

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)

G10H H04H H04S G11B G10L G06F A63F H04R

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

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

EPO-Internal, WPI Data

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>US 2014/119581 A1 (TSINGOS NICOLAS R [US] ET AL) 1 May 2014 (2014-05-01) based on metadata, mapping audio reproduction to the reproduction speakers of the actual reproduction environment; claim 55 dynamic object blobbing; claim 61 metadata associated with the audio data object; claim 68 the plurality of speaker zones correspond to virtual speakers of a virtual surround sound environment.; claim 71; figures 1,4A time delay; paragraph [0077] visual in 2D; paragraph [0083] - paragraph [0084]</td>
<td>1-8, 10-19 9</td>
</tr>
</tbody>
</table>

Further documents are listed in the continuation of Box C.

See patent family annex.

---

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

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

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

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

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

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

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

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

"&" document member of the same patent family

**Date of the actual completion of the international search**

12 February 2016

**Date of mailing of the international search report**

19/02/2016

**Name and mailing address of the ISA/**

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

**Authorized officer**

Glasser, Jean-Marc

Form PCT/ISA210 (second sheet) (April 2005)
<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>mixing console; paragraph [0088] Speaker zones. User selects an active zone; paragraph [0193] touch screen &amp; gesture; paragraph [0198]</td>
<td>1,3-8, 10-19</td>
</tr>
<tr>
<td>Y</td>
<td>Audio mixing; paragraph [0002] visual and audio representation of the mix.; paragraph [0009] a unique visual image of a definable shape and/or color for each of the one or more of audio files; paragraph [0008]; figures 5-11 visual representation of a reverb effect; manipulation of visual; paragraphs [0057], [0059] touch screen; paragraph [0047] audio rendering; &quot;play&quot;; paragraph [0051]; figure 12 gesture; paragraph [0054]</td>
<td>2,9</td>
</tr>
<tr>
<td>Y</td>
<td>US 6 898 291 B2 (GIBSON DAVID A [US]) 24 May 2005 (2005-05-24) cited in the application Each of the audio signal visual images has visual characteristics that correspond to the audio characteristics of the corresponding audio signa; column 1, line 8 - line 50</td>
<td>1-19</td>
</tr>
<tr>
<td>Y</td>
<td>EP 2 450 880 A1 (THOMSON LICENSING [FR]) 9 May 2012 (2012-05-09) Artistic sound field representations &amp; Ambisonics signal; paragraph [0026] acoustic situation in a cinema, reverb; paragraph [0005] B-Format; paragraph [0007] Spherical coordinates of the source position as HOA encoding vector. Ambisonic playback; paragraph [0109] region of interest (sources outside or within the listening area); paragraph [0076]</td>
<td>1-19</td>
</tr>
<tr>
<td>Category</td>
<td>Citation of document, with indication, where appropriate, of the relevant passages</td>
<td>Relevant to claim No.</td>
</tr>
<tr>
<td>----------</td>
<td>---------------------------------------------------------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td></td>
<td>coordinate system; paragraph [0007]; figures 9-11 interface is directly related to the perception of the reproduced sound by the listener; page 16 perceptual factors are manipulated by sliders; paragraph [5.1.2] Localization of the sound source; paragraph [05.2] panning; paragraph [04.3]</td>
<td>9</td>
</tr>
<tr>
<td>Y</td>
<td>CN 101 175 337 A (SONY CORP [JP]) 7 May 2008 (2008-05-07) abstract &amp; EP 1 919 259 A2 (SONY CORP) 7 May 2008 (2008-05-07) virtual sound sources at positions surrounding a listener with a plurality of loudspeakers; paragraph [0003]; figure 1 output to corresponding loudspeaker so that sound image is localized on center of a multi-display; paragraph [0103] - paragraph [0105]; claims 6-8; figures 6, 7 reverberation effect; claim 11 Focusing the sound image, and providing reverberation and delay. Sound image localized at the center regardless of the number of displays; paragraph [0226] zoom; paragraphs [0127], [0130]; figures 10, 11</td>
<td>9</td>
</tr>
<tr>
<td>Category</td>
<td>Citation of document, with indication, where appropriate, of the relevant passages</td>
<td>Relevant to claim No.</td>
</tr>
<tr>
<td>----------</td>
<td>---------------------------------------------------------------------------------------------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Y</td>
<td>mixed-down tracks may be manipulated for final output; paragraph [0028] Manipulation of the graphical form corresponds to the resulting edited audio recording; paragraph [0008] audio editing and mixing as audio sculpting; paragraph [0007]</td>
<td>2,7-11, 14-17</td>
</tr>
<tr>
<td>Patent document cited in search report</td>
<td>Publication date</td>
<td>Patent family member(s)</td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>----------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CA 2837894 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CN 103650535 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EP 2727381 A2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JP 2014520491 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JP 2016007048 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>KR 20140017684 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>KR 20150018645 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RU 2015109613 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TW 201316791 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 2014119581 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 2016037280 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WO 2013006330 A2</td>
</tr>
<tr>
<td>US 2011271186 A1</td>
<td>03-11-2011</td>
<td>US 2011271186 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WO 2011136852 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 2004240686 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CN 103250207 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EP 2450880 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EP 2636036 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HK 1189297 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JP 2013545391 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>KR 20140000240 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PT 2636036 E</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 2013216070 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WO 2012059385 A1</td>
</tr>
<tr>
<td>CN 101175337 A</td>
<td>07-05-2008</td>
<td>CN 101175337 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EP 1919259 A2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JP 4979319 B2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JP 2008109209 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>KR 20080036531 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 2008165992 A1</td>
</tr>
<tr>
<td>US 2008229200 A1</td>
<td>18-09-2008</td>
<td>NONE</td>
</tr>
</tbody>
</table>