



- (51) **International Patent Classification:**
H04S 5/00 (2006.01) *H04R 5/04* (2006.01)
- (21) **International Application Number:**
PCT/CA2017/050050
- (22) **International Filing Date:**
16 January 2017 (16.01.2017)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
62/279,140 15 January 2016 (15.01.2016) US
- (72) **Inventor; and**
- (71) **Applicant : GODFREY, Michael** [CA/CA]; 99 Burnside Drive, Toronto, Ontario M6G 2M9 (CA).
- (74) **Agent: GILBERT'S LLP;** 77 King Street West, Suite 2010, P.O. Box 301, Toronto Dominion Centre, Toronto, Ontario M5K 1K2 (CA).
- (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, 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).

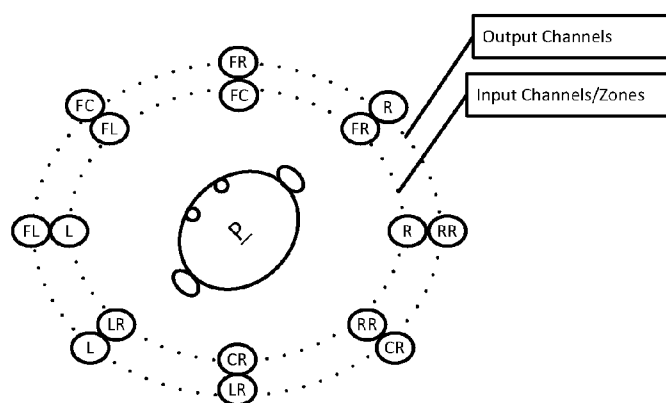
Declarations under Rule 4.17:

- as to the identity of the inventor (Rule 4.17(i))
- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))
- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))
- of inventorship (Rule 4.17(iv))

Published:

- with international search report (Art. 21(3))

(54) **Title:** METHOD AND SYSTEM FOR AUTOMATICALLY DETERMINING A POSITIONAL THREE DIMENSIONAL OUTPUT OF AUDIO INFORMATION BASED ON A USER'S ORIENTATION WITHIN AN ARTIFICIAL IMMERSIVE ENVIRONMENT

**FIG. 3B**

(57) **Abstract:** A method, system and computer program product for automatically determining a positional three dimensional output of audio information based on a user's orientation within an artificial immersive environment are provided. The method includes, based on a fixed geometric relationship between a plurality of input audio channels, defining a plurality of unique geometric audio zones in the artificial immersive environment, each geometric audio zone uniquely characterized by a subset of one or more of the input audio channels; receiving data corresponding to the user's orientation within the artificial immersive environment; automatically, for each of a plurality of output channels: based on the user's orientation with respect to the unique geometric audio zones, identifying a respective subset and producing a mix profile for the one or more input audio channels in the subset; and outputting data defining the subset and its corresponding mix profile.

METHOD AND SYSTEM FOR AUTOMATICALLY DETERMINING A POSITIONAL THREE DIMENSIONAL OUTPUT OF AUDIO INFORMATION BASED ON A USER'S ORIENTATION WITHIN AN ARTIFICIAL IMMERSIVE ENVIRONMENT

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This application claims priority under 35 U.S.C. 119(e) from United States Provisional Patent Application Serial No. 62/279,140 filed on January 15, 2016.

FIELD OF INVENTION

[0001] The present specification relates generally to audio processing, and more specifically to a method and system for automatically determining a positional three dimensional output of audio information based on a user's orientation within an artificial immersive environment.

BACKGROUND OF INVENTION

[0002] It is well known in the art that virtual reality, otherwise known as artificial reality, provides users with an immersive and interactive audiovisual experience and may be applied to any form of entertainment, broadcasting or communication.

[0003] Generally, the user is immersed into a virtual reality environment when real-world stimuli are replaced with electronically generated stimuli. This usually involves the use of audiovisual devices that provide audiovisual content, which may be based on any one or a combination of pre-recorded content, real-time content, or digitally simulated content.

[0004] Additionally, audiovisual devices typically provide "spherical" three-dimensional visual simulations surrounding the user, which is synchronized to three-dimensional audio simulations. However, current three-dimensional audio simulation is very limited and often lacks synchronicity with the three-dimensional video simulation.

[0005] The human auditory and vestibular systems, which provide the brain with the ability to determine sound sources in a three dimensional environment, are extremely sensitive. As such, any lack of synchronicity between the three-dimensional visual and audio simulations in a virtual reality environment may cause a sense of disorientation in the user, resulting in headaches, nausea, vertigo, and ultimately, a lack of immersion. For instance, rotating 180 degrees from a forward facing viewpoint to a backward facing viewpoint while providing the same audio output perspective without the audio naturally following the viewers movements would be confusing and disorientating to the user as it would present the audio to the viewer backwards in relation to the visual material.

[0006] It would be desirable to provide users with an improved technology for virtual reality experiences that are comprised of spatially coordinated audio and visual information. Accordingly, there remains a need for improvements in the art.

SUMMARY OF INVENTION

[0007] In accordance with an aspect, there is provided a method for automatically determining a positional three dimensional output of audio information based on a user's orientation within an artificial immersive environment, the method comprising: based on a fixed geometric relationship between a plurality of input audio channels, defining a plurality of unique geometric audio zones in the artificial immersive environment, each geometric audio zone uniquely characterized by a subset of one or more of the input audio channels; receiving data corresponding to the user's orientation within the artificial immersive environment; automatically, for each of a plurality of output channels: based on the user's orientation with respect to the unique geometric audio zones, identifying a respective subset and producing a mix profile for the one or more input audio channels in the subset; and outputting data defining the subset and its corresponding mix profile.

[0008] In accordance with another aspect, there is provided a system for automatically determining a positional three dimensional output of audio information based on a user's orientation within an artificial immersive environment, the system comprising: an input interface receiving data corresponding to the user's orientation within the artificial immersive environment; processing structure defining a plurality of unique geometric audio zones in the artificial immersive environment based on a fixed geometric relationship between a plurality of input audio channels, each geometric audio zone uniquely characterized by a subset of one or more of the input audio channels, the processing structure automatically, for each of a plurality of output channels, identifying a respective subset and establishing a mix profile for the one or more input audio channels in the subset based on the user's orientation with respect to the unique geometric audio zones; and an output interface outputting data defining the subset and its corresponding mix profile.

[0009] In accordance with another aspect, there is provided a processor readable medium embodying a computer program for automatically determining a positional three dimensional output of audio information based on a user's orientation within an artificial immersive environment, the computer program comprising: program code for, based on a fixed geometric relationship between a plurality of input audio channels, defining a plurality of unique geometric audio zones in the artificial immersive environment, each geometric audio zone uniquely characterized by a subset of one or more of the input audio channels; program code for receiving data corresponding to the user's orientation within the artificial immersive environment; and program code for automatically, for each of a plurality of output

channels: based on the user's orientation with respect to the unique geometric audio zones, identifying a respective subset and producing a mix profile for the one or more input audio channels in the subset; and outputting data defining the subset and its corresponding mix profile.

BRIEF DESCRIPTION OF DRAWINGS

[0010] Embodiments of the invention will now be described with reference to the appended drawings in which:

[0011] Figure 1 is a flowchart depicting steps in a method, according to an embodiment;

[0012] Figure 2 is a schematic diagram of a computing system according to an embodiment;

[0013] Figures 3A, 3B and 3C show top views of different user orientations resulting in different audio outputs for multiple audio output channels, according to an embodiment of the invention;

[0014] Figure 4 shows an example of a fixed geometric relationship between a plurality of input audio channels, defined unique geometric audio zones, and a user orientation corresponding to a particular geometric audio zone;

[0015] Figures 5A, 5B and 5C show examples of calculating weights for a mix profile based on a degree of geometrical alignment of the user orientation with input audio channels for the corresponding geometric audio zone; and

[0016] Figure 6 is a schematic diagram of a system for automatically determining a positional three dimensional output of audio information based on a user's orientation within an artificial immersive environment, according to an embodiment.

DETAILED DESCRIPTION OF EMBODIMENTS

[0017] The detailed embodiments of the present invention are disclosed herein. It should be understood, however, that the disclosed embodiments are merely exemplary of the invention, which may be embodiments in various forms. Therefore, the details disclosed herein are not to be interpreted as limiting, but merely as a basis for teaching one skilled in the art about how to make and/or use the invention.

[0018] As will be appreciated based upon the following disclosure, the present invention relates to the provision of spatially accurate three-dimensional binaural audio information, which is coordinated with the user's orientation and, in embodiments, location in an artificial immersive environment such as a virtual reality or artificial reality environment. Generally, as the user orients and, in embodiments, positions, him or herself within the artificial immersive environment, the audio output orientation will be matched similarly to the video output orientation.

[0019] Figure 1 is a flowchart depicting steps in a process 90 for automatically determining a positional three dimensional output of audio information based on a user's orientation within an artificial immersive environment, according to an embodiment.

[0020] In this embodiment, process 90 is executed on one or more systems such as special purpose computing system 1000 shown in Figure 2. Computing system 1000 may also be specially configured with software applications and hardware components to enable a user to author, edit and/or play media such as digital video and audio, as well as to encode, decode and/or transcode the digital video and corresponding audio from and into various formats such as MP4, AVI, MOV, WEBM and using a selected compression algorithm such as H.264 or H. 265 and according to various selected parameters, thereby to compress, decompress, view and/or manipulate the digital video and audio as desired for a particular application, media player, or platform. Computing system 1000 may also be configured to enable an author or editor to form multiple copies of a particular digital video, each encoded with a respective bitrate, to facilitate streaming of the same digital video to various downstream users who may have different or time-varying capacities to stream it through adaptive bitrate streaming.

[0021] Computing system 1000 includes a bus 1010 or other communication mechanism for communicating information, and a processor 1018 coupled with the bus 1010 for processing the information. The computing system 1000 also includes a main memory 1004, such as a random access memory (RAM) or other dynamic storage device (e.g., dynamic RAM (DRAM), static RAM (SRAM), and synchronous DRAM (SDRAM)), coupled to the bus 1010 for storing information and instructions to be executed by processor 1018. In addition, the main memory 1004 may be used for storing temporary variables or other intermediate information during the execution of instructions by the processor 1018. Processor 1018 may include memory structures such as registers for storing such temporary variables or other intermediate information during execution of instructions. The computing system 1000 further includes a read only memory (ROM) 1006 or other static storage device (e.g., programmable ROM (PROM), erasable PROM (EPROM), and electrically erasable PROM (EEPROM)) coupled to the bus 1010 for storing static information and instructions for the processor 1018.

[0022] The computing system 1000 also includes a disk controller 1008 coupled to the bus 1010 to control one or more storage devices for storing information and instructions, such as a magnetic hard disk 1022 and/or a solid state drive (SSD) and/or a flash drive, and a removable media drive 1024 (e.g., solid state drive such as USB key or external hard drive, floppy disk drive, read-only compact disc drive, read/write compact disc drive, compact disc jukebox, tape drive, and removable magneto-optical drive). The storage devices may be added to the computing system 1000 using an appropriate device interface (e.g., Serial ATA (SATA), peripheral component interconnect (PCI), small computing system interface

(SCSI), integrated device electronics (IDE), enhanced-IDE (E-IDE), direct memory access (DMA), ultra-DMA, as well as cloud-based device interfaces).

[0023] The computing system 1000 may also include special purpose logic devices (e.g., application specific integrated circuits (ASICs)) or configurable logic devices (e.g., simple programmable logic devices (SPLDs), complex programmable logic devices (CPLDs), and field programmable gate arrays (FPGAs)).

[0024] The computing system 1000 also includes a display controller 1002 coupled to the bus 1010 to control a display 1012, such as an LED (light emitting diode) screen, organic LED (OLED) screen, liquid crystal display (LCD) screen or some other device suitable for displaying information to a computer user, or for controlling data to an external display system. In this embodiment, display controller 1002 incorporates a dedicated graphics processing unit (GPU) for processing mainly graphics-intensive or other highly-parallel operations. Such operations may include rendering by applying texturing, shading and the like to wireframe objects including polygons such as spheres and cubes thereby to relieve processor 1018 of having to undertake such intensive operations at the expense of overall performance of computing system 1000. The GPU may incorporate dedicated graphics memory for storing data generated during its operations, and includes a frame buffer RAM memory for storing processing results as bitmaps to be used to activate pixels of display 1012. The GPU may be instructed to undertake various operations by applications running on computing system 1000 using a graphics-directed application programming interface (API) such as OpenGL, Direct3D and the like.

[0025] The computing system 1000 includes input/output devices, such as a head mounted display (HMD) having an external display system and associated audio headphones 1016, for interacting with a computer user and providing information such as orientation and/or position information to the processor 1018. Various input/output devices may be employed, such as those that provide data to the computing system via wires or wirelessly, such as gesture detectors including infrared detectors, gyroscopes, accelerometers, radar/sonar and the like.

[0026] The computing system 1000 performs a portion or all of the processing steps discussed herein in response to the processor 1018 and/or GPU of display controller 1002 executing one or more sequences of one or more instructions contained in a memory, such as the main memory 1004. Such instructions may be read into the main memory 1004 from another processor readable medium, such as a hard disk 1022 or a removable media drive 1024. One or more processors in a multi-processing arrangement such as computing system 1000 having both a central processing unit and one or more graphics processing unit may also be employed to execute the sequences of instructions contained in main memory 1004 or in dedicated graphics memory of the GPU. In alternative embodiments, hard-wired circuitry may be used in place of or in combination with software instructions.

[0027] As stated above, the computing system 1000 includes at least one processor readable medium or memory for holding instructions programmed according to the teachings of the invention and for containing data structures, tables, records, or other data described herein. Examples of processor readable media are solid state devices (SSD), flash-based drives, compact discs, hard disks, floppy disks, tape, magneto-optical disks, PROMs (EPROM, EEPROM, flash EPROM), DRAM, SRAM, SDRAM, or any other magnetic medium, compact discs (e.g., CD-ROM), or any other optical medium, punch cards, paper tape, or other physical medium with patterns of holes, a carrier wave (described below), or any other medium from which a computer can read.

[0028] Stored on any one or on a combination of processor readable media, is software for controlling the computing system 1000, for driving a device or devices to perform the functions discussed herein, and for enabling the computing system 1000 to interact with a human user (e.g., digital video author/editor/user). Such software may include, but is not limited to, device drivers, operating systems, development tools, and applications software. Such processor readable media further includes the computer program product for performing all or a portion (if processing is distributed) of the processing performed discussed herein.

[0029] The computer code devices discussed herein may be any interpretable or executable code mechanism, including but not limited to scripts, interpretable programs, dynamic link libraries (DLLs), Java classes, and complete executable programs. Moreover, parts of the processing of the present invention may be distributed for better performance, reliability, and/or cost.

[0030] A processor readable medium providing instructions to a processor 1018 may take many forms, including but not limited to, non-volatile media, volatile media, and transmission media. Non-volatile media includes, for example, optical, magnetic disks, and magneto-optical disks, such as the hard disk 1022 or the removable media drive 1024. Volatile media includes dynamic memory, such as the main memory 1004. Transmission media includes coaxial cables, copper wire and fiber optics, including the wires that make up the bus 1010. Transmission media also may also take the form of acoustic or light waves, such as those generated during radio wave and infrared data communications using various communications protocols.

[0031] Various forms of processor readable media may be involved in carrying out one or more sequences of one or more instructions to processor 1018 for execution. For example, the instructions may initially be carried on a magnetic disk of a remote computer or in combination with a processor such as a processor resident on HMD 1014. The remote computer can load the instructions for implementing all or a portion of the present invention remotely and send the instructions or processing results over a wired or wireless connection. A modem local to the computing system 1000 may receive data via wired Ethernet or wirelessly via WiFi and place data on the bus 1010. The bus 1010 carries the data to the main memory

1004, from which the processor 1018 retrieves and executes the instructions. The instructions received by the main memory 1004 may optionally be stored on storage device 1022 or 1024 either before or after execution by processor 1018.

[0032] The computing system 1000 also includes a communication interface 1020 coupled to the bus 1010. The communication interface 1020 provides a two-way data communication coupling to a network link that is connected to, for example, a local area network (LAN) 1500, or to another communications network 2000 such as the Internet. For example, the communication interface 1020 may be a network interface card to attach to any packet switched LAN. As another example, the communication interface 1020 may be an asymmetrical digital subscriber line (ADSL) card, an integrated services digital network (ISDN) card or a modem to provide a data communication connection to a corresponding type of communications line. Wireless links may also be implemented. In any such implementation, the communication interface 1020 sends and receives electrical, electromagnetic or optical signals that carry digital data streams representing various types of information.

[0033] The network link typically provides data communication through one or more networks to other data devices, including without limitation to enable the flow of electronic information. For example, the network link may provide a connection to another computer through a local network 1500 (e.g., a LAN) or through equipment operated by a service provider, which provides communication services through a communications network 2000. The local network 1500 and the communications network 2000 use, for example, electrical, electromagnetic, or optical signals that carry digital data streams, and the associated physical layer (e.g., CAT 5 cable, coaxial cable, optical fiber, etc). The signals through the various networks and the signals on the network link and through the communication interface 1020, which carry the digital data to and from the computing system 1000, may be implemented in baseband signals, or carrier wave based signals. The baseband signals convey the digital data as unmodulated electrical pulses that are descriptive of a stream of digital data bits, where the term "bits" is to be construed broadly to mean symbol, where each symbol conveys at least one or more information bits. The digital data may also be used to modulate a carrier wave, such as with amplitude, phase and/or frequency shift keyed signals that are propagated over a conductive media, or transmitted as electromagnetic waves through a propagation medium. Thus, the digital data may be sent as unmodulated baseband data through a "wired" communication channel and/or sent within a predetermined frequency band, different than baseband, by modulating a carrier wave. The computing system 1000 can transmit and receive data, including program code, through the network(s) 1500 and 2000, the network link and the communication interface 1020. Moreover, the network link may provide a connection through a LAN 1500 to a mobile device 1300 such as a personal digital assistant (PDA) laptop computer, or cellular telephone.

[0034] Computing system 1000 may be provisioned with or be in communication with live broadcast/streaming equipment that receives and transmits, in near real-time, a stream of digital video and audio content captured in near real-time from a particular live event.

[0035] Alternative configurations of computing system, such as those that are not interacted with directly by a human user through a graphical or text user interface, may be used to implement process 90. For example, for live-streaming and broadcasting applications, a hardware-based processing system may be employed that also executes process 90 as described herein.

[0036] The electronic data store implemented in the database or other data structures described herein may be one or more of a table, an array, a database, a structured data file, an XML file, or some other functional data store, such as hard disk 1022 or removable media 1024.

[0037] In this embodiment, during process 90, based on a fixed geometric relationship between multiple input audio channels, multiple unique geometric audio zones in the artificial immersive environment are defined (step 100). Each geometric audio zone is characterized by a subset of one or more of the input audio channels. The process 90 proceeds with receiving data corresponding to the user's orientation within the artificial immersive environment (step 200). Automatically, for each of a plurality of output channels, a respective subset is identified (step 300) and a mix profile for the one or more input audio channels in the subset is produced (step 400). Pursuant to this, data defining the subset and its corresponding mix profile are outputted (step 500) for use downstream as will be described.

[0038] In this embodiment, the receiving, identifying, producing and outputting are conducted continuously during use of the artificial immersive environment. That is, as the user is continuously navigating the artificial immersive environment, the data defining the subset and its corresponding mix profile are continuously being updated and outputted according to the user's orientation. In an alternative embodiment, processing resources can be efficiently managed such that, only if the user has changed his or her orientation in the artificial immersive environment are the identifying, producing and outputting conducted.

[0039] In this embodiment, the fixed geometric relationship corresponds to the fixed geometric relationship between positions and orientations of microphones in a particular array configuration with respect to an origin. An example configuration is described in United States Patent No. 5,778,083 to Godfrey, the contents of which are incorporated herein by reference. In the '083 patent, for example, a generally football-shaped frame supports eight (8) microphones distributed laterally around the generally spherical cross-sectioned frame (center front, front left, left, rear left, center rear, rear right, right, and front right) as well as both top and bottom microphones at the top and bottom, respectively, of the frame. The microphones are each oriented such that their diaphragms are oriented outwardly from the frame. In the '083 patent, the lateral microphones have a hypercardioid pickup pattern while the top and bottom

microphones each have a hemispherical pickup pattern, but the positions and orientations of the microphones on the frame form a fixed geometric relationship.

[0040] It is important for the listening experience that the fixed geometric relationship be preserved in some way from capture to playback, or that the fixed geometric relationship at least is synthesized very carefully in a manner that provides spatial integrity. Maintaining or synthesizing the integrity of the spatial positioning of the audio sources provides a very visceral auditory “core” to the listener. When the user re-orient himself or herself in the artificial immersive environment, it is this auditory core that should re-orient accordingly thereby to re-orient all of the corresponding audio information together as one. If no re-orientation of audio information is done, then the orientation of the audio information presented to the user can be perceptually misaligned - severely so in some cases - with the orientation of the video information. For example, a user could see an airplane approaching from the front right, but hear it as though it was approaching from the rear left. Still further, if only a portion of the audio information is re-oriented so as to be inconsistent with the fixed geometric relationship in connection with other audio information, the user may still feel disoriented or, at the least, will not be receiving the visceral benefits of the auditory information being oriented in unison to preserve the auditory core.

[0041] It will be understood that alternative capture techniques, such as various multi-channel surround sound capture techniques conducted according to a standardized methodology thereby defining a standardized fixed geometric relationship, and/or tetrahedral or multiple binaural array configurations may be employed. However, again, it is important to be able to maintain the integrity of the spatial positioning between audio capture devices (i.e. microphones) in a fixed geometric relationship as the user changes his or her orientation within the artificial immersive environment.

[0042] Data defining the fixed geometric relationship may be encoded along with the input audio information thereby to enable downstream processing to recreate the fixed geometric relationship after decoding.

[0043] In the event that object-based audio, produced post-capture, is to be used as the basis for providing audio information to a user, it would be advantageous to re-process the object-based audio for core integrity, thereby to ensure that the channel assignments are correct spatially (according to a suitable fixed geometric relationship) and to ensure that the correct spatial assignments are maintained during re-orientations. As such, it may be advantageous to first translate such incoming object-based audio into individual channels, to lock the spatial positioning and orientation of the audio channels with respect to each other according to an appropriate fixed geometric relationship, and to thereafter to conduct the identifying and producing as described above using the translated individual channels.

[0044] In this embodiment, producing each mix profile includes calculating a contribution of each of the one or more input audio channels in the subset to the output channel. In the event that a subset

includes only one input audio channel, then the calculation is straightforward: the mix profile specifies simply that the one input audio channel contributes 100% to the corresponding output audio channel.

[0045] Table 1 shows an embodiment of a data structure, in this embodiment a two-dimensional matrix, which is appropriate for embodiments where a mix profile is to specify simply that one input audio channel contributes 100% to its corresponding output channel. In this embodiment, the two-dimensional matrix stores a sequence of input audio channels according to multiple different user orientations. The sequence itself, constant through the various possible rotational positions available in the matrix, represents the fixed geometric relationship between audio channels distributed laterally such as those described above in the '083 patent.

[0046] For example, as shown in Figure 3A, when the user is oriented to face in a direction corresponding to the front center (FC) input audio channel, the front left (FL), left (L), left rear (LR), center rear (CR), right rear (RR), right (R) and front right (FR) input audio channels are spatially positioned and oriented in sequence around an ellipse in a counterclockwise direction. In this embodiment, the geometric audio zones are defined as the FC zone, the FL zone, the L zone, the LR zone, the CR zone, the RR zone, the R zone and the FR zone. Accordingly, in this embodiment, the subset characterizing the FC zone includes only the FC input audio channel, the subset characterizing the FL zone includes only the FL audio input channel, the subset characterizing the L zone includes only the L audio input channel, the subset characterizing the LR zone includes only the LR input audio channel, the subset characterizing the CR zone includes only the CR input audio channel, the subset characterizing the RR zone includes only the RR input audio channel, the subset characterizing the R zone includes only the R input audio channel, and the subset characterizing the FR zone includes only the FR input audio channel.

[0047] As shown in Figure 3B, the user has re-oriented to face in a direction corresponding to the FL zone. This orientation information is employed to, for each output audio channel, identify a respective subset and produce a mix profile. As such, in this example, the FL zone corresponds to the user's orientation so the FL input audio channel is now to contribute 100% to the FC output audio channel, the L input audio channel is now to contribute 100% to the FL output audio channel, the LR input audio channel is now to contribute 100% to the L output audio channel, and so forth.

[0048] Similarly, as shown in Figure 3C, should the user re-orient himself to face in a direction corresponding to the L zone, the L input audio channel is now to contribute 100% to the FC output audio channel, the RL input audio channel is now to contribute 100% to the FL output audio channel, the R input audio channel is now to contribute 100% to the RL output audio channel, and so forth.

[0049] As such, in this embodiment, as the user changes his yaw about a Z-axis running through an origin of the fixed geometric relationship to face certain predefined positions specified in the matrix, the audio core sequence of FC-FL-L-LR-CR-RR-R-FR “rotates” as a unit according to the change in yaw.

Output Channel Sequence (Counter-clockwise)	Direction User Faces (Orientation)							
	Front Center (FC)	Front Left (FL)	Left (L)	Left Rear (LR)	Center Rear (CR)	Right Rear (RR)	Right (R)	Front Right (FR)
	FC	FL	L	LR	CR	RR	R	FR
	FL	L	LR	CR	RR	R	FR	CF
	L	LR	CR	RR	R	FR	FC	FL
	LR	CR	RR	R	FR	FC	FL	L
	CR	RR	R	FR	FC	FL	L	LR
	RR	R	FR	FC	FL	L	LR	CR
	R	FR	FC	FL	L	LR	CR	RR
	FR	FC	FL	L	LR	CR	RR	R

Table 1

[0050] The embodiment described above in connection with Table 1 is illustrative of principles of the invention, where one input audio channel contributes fully (weighted to 100%) to its corresponding output audio channel. The example embodiment provides for one-to-one re-assignment (or routing) of audio input channels to particular output audio channels for a discrete number of positions (8 positions) corresponding to spatial locations of audio channels, based on the orientation of a user in an artificial immersive environment. Due to the preserved sequence, as described above, the audio core is preserved as the user re-orient in the artificial immersive environment.

[0051] In this embodiment, any Top or Bottom input audio channel is to be simply routed to the Top or Bottom output audio channels accordingly. That is, in the straightforward example above, only certain changes in yaw of a user’s orientation, and not changes in pitch or roll, result in re-orientation of the audio information. However, in a practical system, it would be valuable to provide audio orientation changes when a user changed pitch or roll. For this reason, a third dimension added to the two-dimensional matrix depicted above would enable encoding of the corresponding fixed geometrical structure that account for top and bottom channels in various ways could be used. As would be understood however, unless the matrix is very large, there is likely to be orientations that would not have exact predefined counterparts in the matrix. As a result, in a matrix implementation additional processing may be provided to derive output audio information from multiple input audio channels rather than

simply allocate a full input channel to a full output channel, and to smooth changes between zones specified in the matrix so that the resultant audio information does not present clicks or sharp changes across zone boundaries.

[0052] However, alternatives are contemplated that account for more sophisticated processes for producing subsets of more than one input audio channel and producing corresponding mix profiles using more than one input audio channels for orientations that are not predefined as in Table 1 or a three-dimensional variant of it. As an example, using a minimal number of channels to define zones, Figure 4 shows a configuration for audio capture that preserves a fixed geometric relationship between audio input channels T (Top), FC, FL, L, RL and CR as well as those not shown in Figure 4: audio input channels RR, R, FR and B (Bottom). In this embodiment, a plurality of unique geometric zones is each defined to be characterized by a subset of three (3) input audio channels. In particular, in this embodiment the zones are defined as shown in Table 2, below.

Zone	Subset of Input Audio Channels
A	T-FC-FL
B	T-FL-L
C	T-L-RL
D	T-RL-CR
E (Not Shown)	T-CR-RR
F (Not Shown)	T-RR-R
G (Not Shown)	T-R-FR
H (Not Shown)	T-FR-FC
I (Not Shown)	B-FC-FL
J (Not Shown)	B-FL-L
K (Not Shown)	B-L-RL
L (Not Shown)	B-RL-CR
M (Not Shown)	B-CR-RR
N (Not Shown)	B-RR-R
O (Not Shown)	B-R-FR
P (Not Shown)	B-FR-FC

Table 2

[0053] In Figure 4, an Orientation is shown that represents the orientation of the user in the artificial immersive environment. It will be noted that, despite the appearance as drawn in Figure 4, the origin for the Orientation is not the same location as the L input audio channel. Rather, in this embodiment, the origin corresponds with the geometric center about which the geometric audio zones are positioned.

[0054] As shown, the Orientation in Figure 4 does not align with any of spatial locations of the input channels T, FC, FL, L, RL, CR, RR, R, FR or B. As such, in this embodiment, the mix profile for each output audio channel will include non-zero values corresponding to each of the input audio channels for the corresponding subset, with the values summing to 100%.

[0055] For example, in Figure 4, the Orientation corresponds with Zone A. As such, the subset of input audio channels that are to be combined to produce the output audio channel for FC is T-FC-FL. While the corresponding mix profile could simply be an even mix of all three input audio channels (33%, 33%, 33%), it may be advantageous to further refine the mix profile for improved realism according to the proximity of Orientation to each of the three input audio channels. In the embodiment shown in Figure 4, while there is no alignment between the Orientation and any one of input audio channels T, FC and FL, Orientation differs least in alignment in Zone A from input audio channel FC, and differs most in alignment in Zone A from input audio channel T. Its alignment with input audio channel FL in Zone A falls somewhere between its degree of alignment with input audio channels FC and T. As such, a geometric calculation could be conducted that produces a mix profile weighting the relative contribution of an input audio channel higher the closer aligned the Orientation is to that input audio channel in the corresponding zone (that is, the closer the point of intersection of Orientation on the surface forming Zone A is to the points corresponding to the input audio channels). As such, in the example Orientation shown in Figure 4, the mix profile corresponding to (T, FC, FL) might be produced something akin to the following: (20%, 50%, 30%).

[0056] With the geometric audio zone subset and mix profile for the CF output audio channel according to the Orientation shown in Figure 4 having been produced, a very similar process can be conducted for each of the FL, L, RL, CR, RR, R, FR, Top and Bottom output audio channels. Advantageously, it will be understood that the zone corresponding to each of these output audio channels can be derived using straightforward geometrical calculations because the fixed geometric relationship between the channels has been maintained and the Orientation of the user is known. Similarly, the respective mix profiles for each output audio channel will be the same as that produced for the CF output audio channel as described above, except the subset of input audio channels will be different for each output audio channel.

[0057] It should be understood that the technique for weighting the various contributions to an output signal for an output channel may vary based on audio principles such as differences in pickup patterns of microphones (such as hypercardioid versus hemispherical), differences in perceptions of different frequencies, and differences in human perception of sound approaching the ears laterally versus from above or below, as well as other factors. However, these factors could also be codified or otherwise

processed and accounted for in the relative weights in the mix profiles as the user navigates in the artificial geometric environment.

[0058] It will be understood that other fixed geometric relationships can be similarly codified, and that subsets of input audio channels for unique geometric audio zones and corresponding mix profiles could implicate just two input audio channels in certain instances, or more than three input audio channels.

[0059] With the data defining the subset and its corresponding mix profile having been produced, actual audio information can be sent to each output audio channel from the one or more input audio channels in its corresponding subset according to the corresponding mix profile. This may be done in practice by presenting the subsets and mix profiles to a mixer/blender, which can use the subsets and mix profiles to define various buses from the input audio channels and the respective levels of the input audio channels for each bus, and routing each bus to respective output audio channels. Such a mixer/blender may be implemented as a software module, may be part of a module conducting the steps described herein, or may be a hardware-implemented special-purpose mixer/blender.

[0060] The audio channel outputs from the from the mixer/blender, containing audio information produced as described above, may then be directed to a respective one of a plurality of corresponding audio output devices, where each of the audio output devices is positioned and oriented with respect to a user position according to the fixed geometric relationship.

[0061] Alternatively (or in some combination), the audio channel outputs may be directed to an audio virtualizer. An audio virtualizer generally “spreads” a plurality of output channels across a software- or hardware-implemented geometric structure such as the inner walls of a sphere or the inner walls of a sound booth or the inner walls of a concert hall (as desired), such that each of a plurality of locations (typically many more than there are actual individual output audio channels) on the surface of the geometric structure can be notionally considered a point source of audio. In this way, the virtualized audio can more readily be presented for output on downstream audio output devices of various configurations instead of the particular configuration used to capture the audio information in the first place.

[0062] In one embodiment useful for virtual reality and/or 360 video and/or other applications that lend themselves to personal use of headphones, the virtualized audio information may be directed to the inputs of both left and right head related transfer functions (HRTFs), either implemented in software or hardware or a combination thereof, the outputs of which are conveyed to respective earpieces of headphones or earbuds or the like, or other combinations of left and right audio output devices directed at the user’s ears.

[0063] In addition to receiving orientation information corresponding to up to three degrees of freedom (3-DOF) of the user in the artificial immersive environment, additional information about user's navigation in the artificial immersive environment may be available. For example, an HMD system for virtual reality gaming and the like may offer information in six degrees of freedom (6-DOF), including information about translation in X, Y, and Z dimensions in addition to the information about yaw, pitch and roll described above. As such, in an embodiment, in addition to receiving data corresponding to the user's orientation (yaw, pitch, roll) within the artificial immersive environment, data is received corresponding to the user's position within the artificial immersive environment. In response to the position data, beamforming of one or more of the output audio channels is applied or modified in order to enable a user to receive more sound in the direction of change of the user's position. This would enable a user to not only cock their head in an effort to orient one of their ears to the direction of a particular sound, but to "get closer" to the sound by moving his or her head towards the sound. In an implementation, therefore, amplitude of one or more output audio channels corresponding to a direction of change in the user's position is increased, and amplitude of one or more output audio channels corresponding to the opposite of the direction of change in the user's position is decreased. This feature may be enabled in an implementation of an artificial immersive environment where 6-DOF is tracked by user equipment (such as an HMD or accompanying device) but may simply be disabled in an implementation of an artificial immersive environment where only 3-DOF is tracked by the user equipment.

[0064] The methods described herein may be implemented where input audio channels have never been encoded, but are more likely to be implemented where input audio channels have been decoded from one or more sources of either pre-recorded audio information or streaming audio information. Furthermore, the methods described herein may be implemented where one or more of the input audio channels have been synthesized rather than captured using an audio sensor, and the synthesized one or more channels have been oriented according to a fixed geometric relationship with each other and with any other input channels.

[0065] Figure 6 is schematic diagram of a system 2000 for automatically determining a positional three dimensional output of audio information based on a user's orientation within an artificial immersive environment. System 2000 may be implemented in software, hardware or a combination of both and sharing or using resources provided by system 1000 in Figure 2. The system 2000 includes an input interface 2002 receiving data corresponding to the user's orientation within the artificial immersive environment. Processing structure is configured to defining a plurality of unique geometric audio zones 2004 in the artificial immersive environment based on a fixed geometric relationship between a plurality of input audio channels 2500. As described above, each geometric audio zone is uniquely characterized

by a subset of one or more of the input audio channels 2500. The processing structure is also configured to automatically, for each of a plurality of output audio channels 2600, identify a respective subset 2006 and establish a mix profile 2008 for the one or more input audio channels 2500 in the subset 2006 based on the user's orientation with respect to the unique geometric audio zones. An output interface 2010 outputs data 2012 defining the subset and its corresponding mix profile for each output audio channel.

[0066] The data 2012 may be routed to a mixer/blender 2014 that also receives input audio information via the audio input channels 2500. The mixer/blender produces output audio information 2600 with the input audio information and the data 2012. The output audio information can be conveyed to one or more downstream systems, such as a speaker system 2700 or a virtualizer/spatializer 2800. In this embodiment, outputs of the virtualizer/spatializer 2800 may be fed to two head related transfer functions 2900R and 2900L, the outputs of each of which are, in turn, conveyed to respective speakers of headphones.

[0067] In this embodiment, the input interface 2002 also optionally receives data corresponding to the user's position within the artificial immersive environment; and provides that position data to a beamformer 2009 to apply or modifying beamforming for one or more of the output audio channels in response to changes in the user's position.

[0068] The present invention may be embodied in other specific forms without departing from the spirit or essential characteristics thereof. Certain adaptations and modifications of the invention will be obvious to those skilled in the art. Therefore, the presently discussed embodiments are considered to be illustrative and not restrictive, the scope of the invention being indicated by the appended claims rather than the foregoing description and all changes which come within the meaning and range of equivalency of the claims are therefore intended to be embraced therein.

What is claimed is:

1. A processor-implemented method for automatically determining a positional three dimensional output of audio information based on a user's orientation within an artificial immersive environment, the method comprising:

based on a fixed geometric relationship between a plurality of input audio channels, defining a plurality of unique geometric audio zones in the artificial immersive environment, each geometric audio zone uniquely characterized by a subset of one or more of the input audio channels;

receiving data corresponding to the user's orientation within the artificial immersive environment;

automatically, for each of a plurality of output channels:

based on the user's orientation with respect to the unique geometric audio zones, identifying a respective subset and producing a mix profile for the one or more input audio channels in the subset; and

outputting data defining the subset and its corresponding mix profile.

2. The method of claim 1, wherein the receiving, identifying, producing and outputting are conducted continuously during use of the artificial immersive environment.

3. The method of claim 1, further comprising:

conducting the identifying, producing and outputting of a different subset and mix profile only in the event that the user's orientation within the artificial immersive environment has changed.

4. The method of claim 1, wherein producing each mix profile comprises calculating a contribution of each of the one or more input audio channels in the subset to the output channel.

5. The method of claim 4, wherein calculating a contribution comprises:

for each of the one or more input audio channels in the subset:

establishing a weight for a signal based at least in part on a degree that the user's orientation aligns with the position of the input audio channel in the fixed geometric relationship.

6. The method of claim 1, wherein the fixed geometric relationship is stored in a data structure.

7. The method of claim 6, wherein the data structure is a matrix.

8. The method of claim 7, wherein the matrix codifies a sequence of input audio channels according to a plurality of predefined user orientations.
9. The method of claim 8, wherein each subset includes only one input audio channel.
10. The method of claim 9, wherein the mix profile for each of the output audio channels comprises a 100% contribution of the one input audio channel in the corresponding subset.
11. The method of claim 1, wherein a plurality of the subsets includes at least two input audio channels and the corresponding mix profiles comprise a ratio of contributions of the at least two input audio channels.
12. The method of claim 1, wherein the fixed geometric relationship is a standardized geometric relationship according to a number of input audio channels.
13. The method of claim 12, wherein data defining the fixed geometric relationship is encoded with input audio information.
14. The method of claim 1, further comprising:
sending to each output audio channel audio information from the one or more input audio channels in the corresponding subset according to the mix profile.
15. The method of claim 14, further comprising:
directing audio information from each output audio channel to a virtualizer.
16. The method of claim 14, further comprising:
directing audio information from each output audio channel to a respective one of a plurality of audio output devices, each of the output audio devices positioned and oriented with respect to a user position according to the fixed geometric relationship.
17. The method of claim 15, further comprising:
directing output audio information from the virtualizer to the inputs of each of left and right head related transfer functions (HRTFs); and

conveying the outputs of the HRTFs to respective earpieces of headphones.

18. The method of claim 1, further comprising:

receiving data corresponding to the user's position within the artificial immersive environment;

and

applying or modifying beamforming to one or more of the output audio channels in response to changes in the user's position.

19. The method of claim 18, wherein applying or modifying beamforming comprises:

increasing the amplitude of one or more output audio channels corresponding to a direction of the change in user's position;

decreasing the amplitude of one or more output audio channels corresponding to the opposite of the direction of the change in user's position.

20. The method of claim 1, wherein the input audio channels are decoded from one or more sources of pre-recorded audio information.

21. The method of claim 1, wherein the input audio channels are decoded from one or more sources of streaming audio information.

21. The method of claim 1, further comprising:

prior to defining the plurality of unique geometric audio zones, synthesizing the input audio channels and the fixed geometric relationship.

22. A system for automatically determining a positional three dimensional output of audio information based on a user's orientation within an artificial immersive environment, the system comprising:

an input interface receiving data corresponding to the user's orientation within the artificial immersive environment;

processing structure defining a plurality of unique geometric audio zones in the artificial immersive environment based on a fixed geometric relationship between a plurality of input audio channels, each geometric audio zone uniquely characterized by a subset of one or more of the input audio channels,

the processing structure automatically, for each of a plurality of output channels, identifying a respective subset and establishing a mix profile for the one or more input audio channels in the subset based on the user's orientation with respect to the unique geometric audio zones; and

an output interface outputting data defining the subset and its corresponding mix profile.

23. A processor readable medium embodying a computer program for automatically determining a positional three dimensional output of audio information based on a user's orientation within an artificial immersive environment, the computer program comprising:

program code for, based on a fixed geometric relationship between a plurality of input audio channels, defining a plurality of unique geometric audio zones in the artificial immersive environment, each geometric audio zone uniquely characterized by a subset of one or more of the input audio channels;

program code for receiving data corresponding to the user's orientation within the artificial immersive environment; and

program code for automatically, for each of a plurality of output channels:

based on the user's orientation with respect to the unique geometric audio zones, identifying a respective subset and producing a mix profile for the one or more input audio channels in the subset; and

outputting data defining the subset and its corresponding mix profile.

1/6

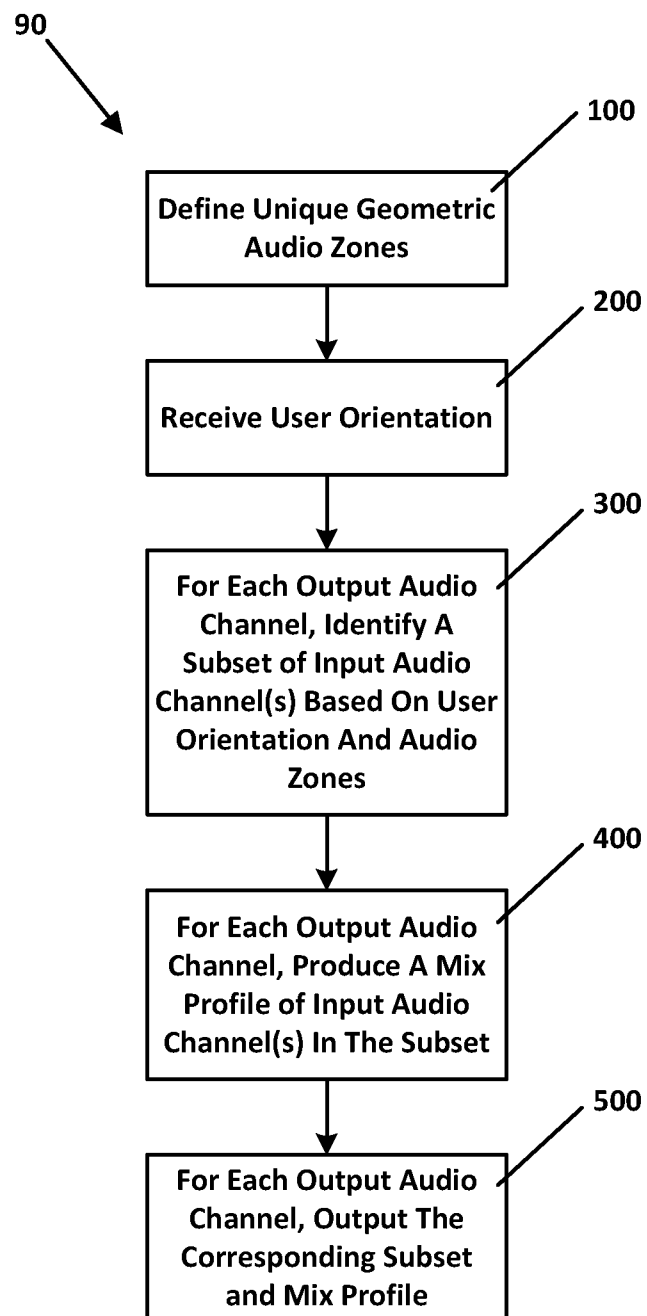


FIG. 1

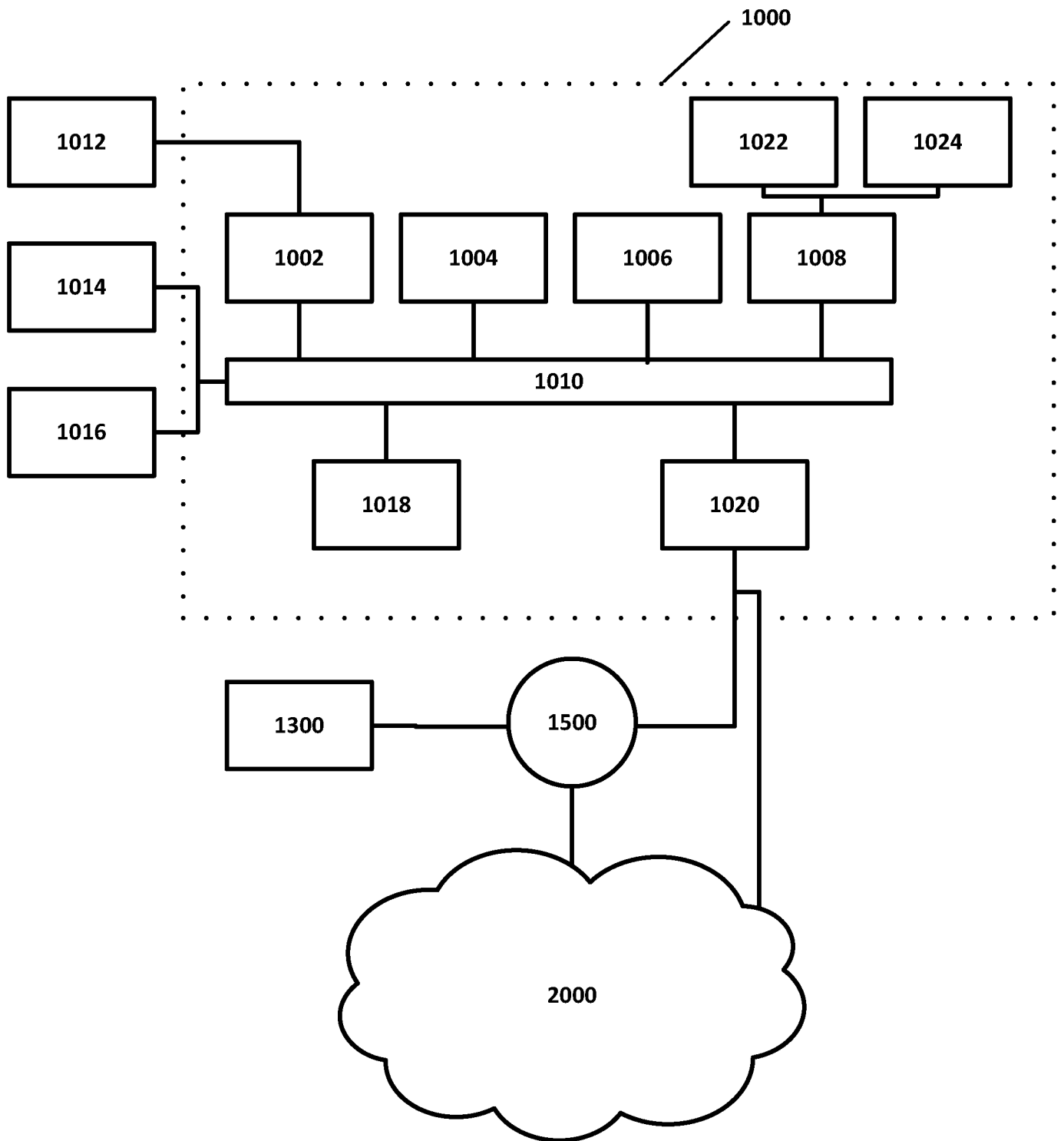


FIG. 2

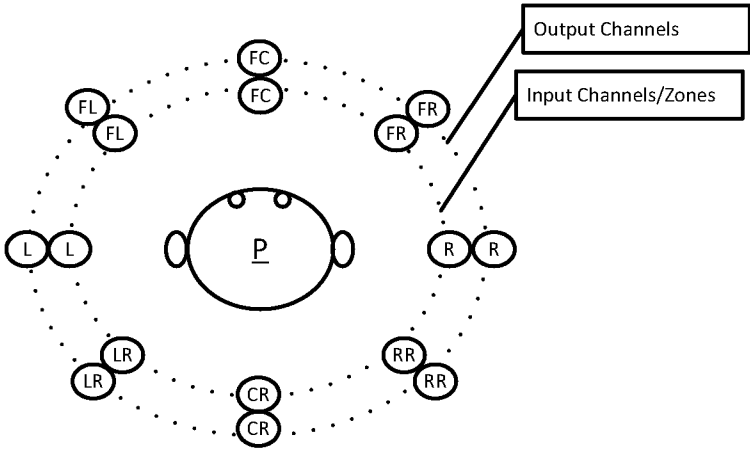


FIG. 3A

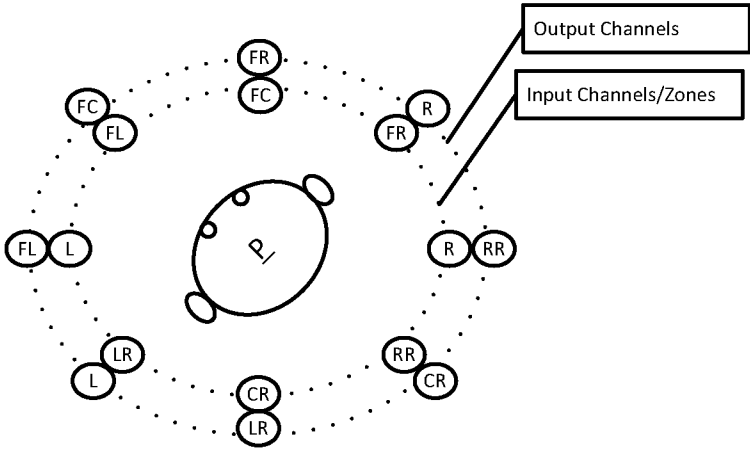


FIG. 3B

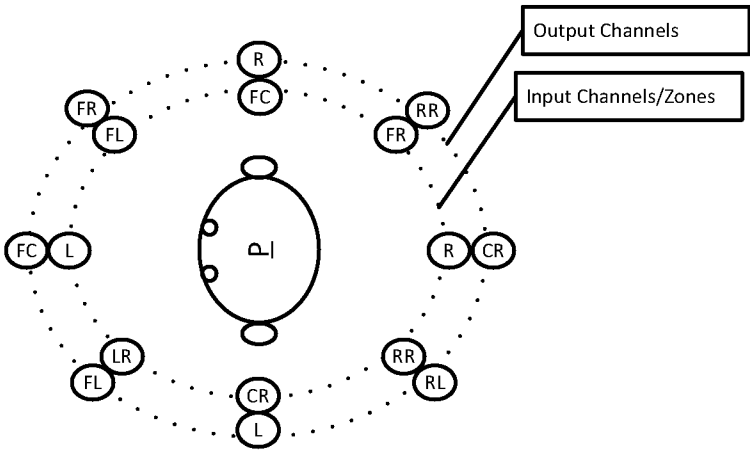


FIG. 3C

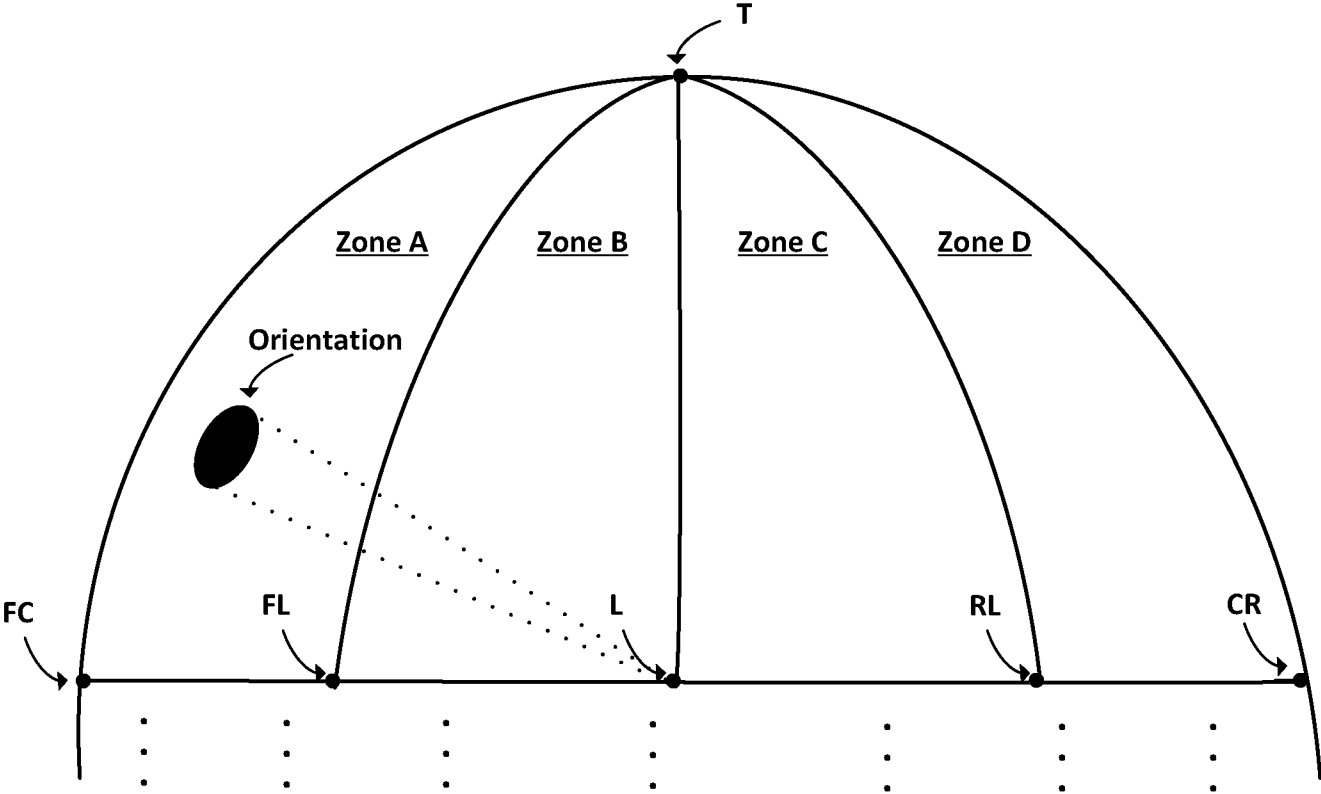


FIG. 4

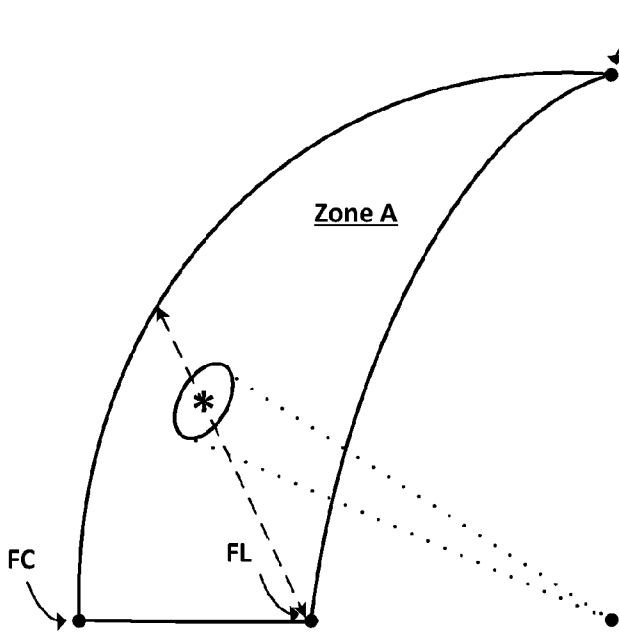


FIG. 5A

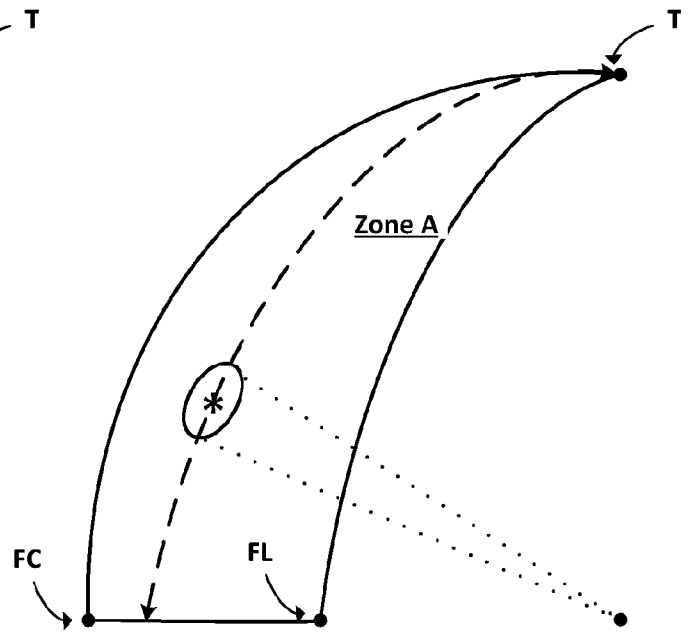


FIG. 5B

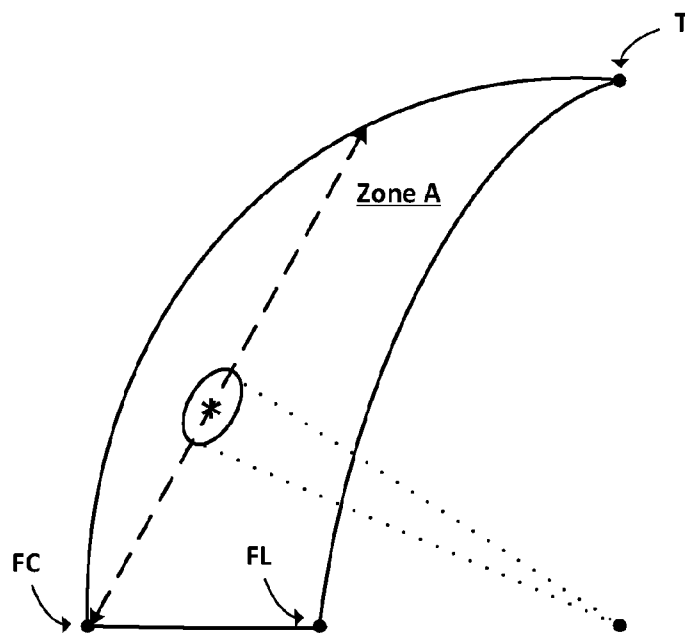


FIG. 5C

6/6

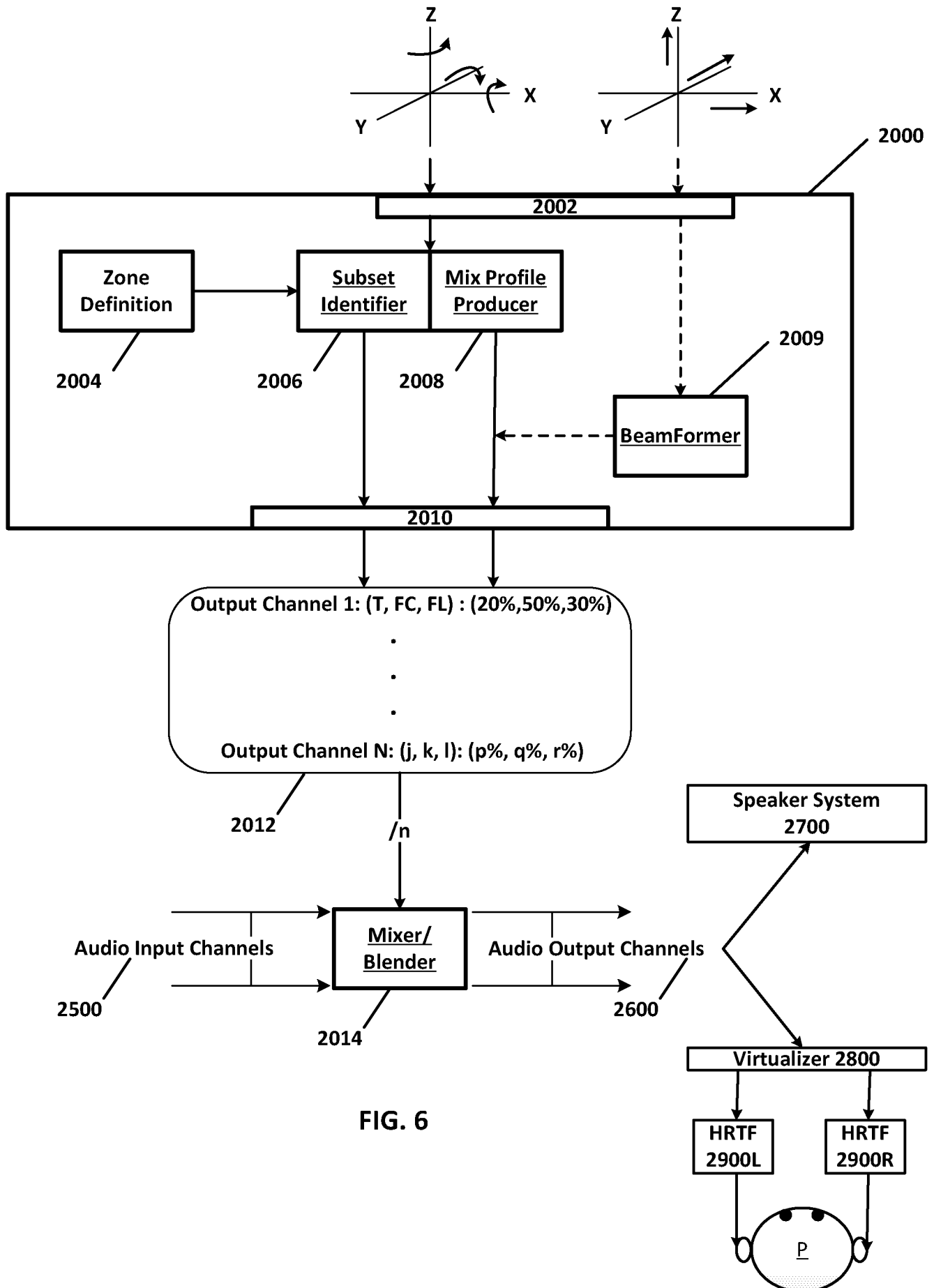


FIG. 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CA2017/050050

A. CLASSIFICATION OF SUBJECT MATTER
IPC: **H04S 5/00** (2006.01) , **H04R 5/04** (2006.01)

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: H04S 5/00 (2006.01), H04R 5/04 (2006.01)

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

Electronic database(s) consulted during the international search (name of database(s) and, where practicable, search terms used)
Databases: Questel Orbit, Canadian Patent Database
Keywords: three dimensional, audio, user orientation/position, geometric, zone, subset, mix profile

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 8 767 968 B2 (Flaks et al.) 1 July 2014 (01-07-2014) *entire document*	1-23
A	EP 1 296 155 B1 (Willins et al.) 26 March 2003 (26-03-2003) *entire document*	1-23
A	WO 2010-140088 A1 (De Bruijn) 9 December 2010 (09-12-2010) *entire document*	1-23

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* "A" "E" "L" "O" "P"	Special categories of cited documents: document defining the general state of the art which is not considered to be of particular relevance earlier application or patent but published on or after the international filing date 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 referring to an oral disclosure, use, exhibition or other means document published prior to the international filing date but later than the priority date claimed	"T" "X" "Y" "&"	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 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 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
08 March 2017 (08-03-2017)

Date of mailing of the international search report
23 March 2017 (23-03-2017)

Name and mailing address of the ISA/CA
Canadian Intellectual Property Office
Place du Portage I, C114 - 1st Floor, Box PCT
50 Victoria Street
Gatineau, Quebec K1A 0C9
Facsimile No.: 819-953-2476

Authorized officer

S. Khan (819) 639-8498

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CA2017/050050

Patent Document Cited in Search Report	Publication Date	Patent Family Member(s)	Publication Date
US8767968B2	01 July 2014 (01-07-2014)	US2012093320A1 CN102413414A CN102413414B US2013208898A1 US9522330B2 US2013208897A1 US2013208899A1 US2013208900A1 US2013208926A1	19 April 2012 (19-04-2012) 11 April 2012 (11-04-2012) 25 March 2015 (25-03-2015) 15 August 2013 (15-08-2013) 20 December 2016 (20-12-2016) 15 August 2013 (15-08-2013) 15 August 2013 (15-08-2013) 15 August 2013 (15-08-2013) 15 August 2013 (15-08-2013)
EP1296155A1	26 March 2003 (26-03-2003)	EP1296155A1 EP1296155B1 DE60216209D1 DE60216209T2 EP1746434A2 EP1746434A3 JP2003215243A US2003061001A1 US6684176B2	26 March 2003 (26-03-2003) 22 November 2006 (22-11-2006) 04 January 2007 (04-01-2007) 13 September 2007 (13-09-2007) 24 January 2007 (24-01-2007) 09 July 2008 (09-07-2008) 30 July 2003 (30-07-2003) 27 March 2003 (27-03-2003) 27 January 2004 (27-01-2004)
WO2010140088A1	09 December 2010 (09-12-2010)	WO2010140088A1 CN102461214A CN102461214B EP2438770A1 JP2012529213A JP5572701B2 KR20120026591A KR101659954B1 RU2011154335A RU2543937C2 US2012075957A1 US9332371B2	09 December 2010 (09-12-2010) 16 May 2012 (16-05-2012) 30 September 2015 (30-09-2015) 11 April 2012 (11-04-2012) 15 November 2012 (15-11-2012) 13 August 2014 (13-08-2014) 19 March 2012 (19-03-2012) 26 September 2016 (26-09-2016) 20 July 2013 (20-07-2013) 10 March 2015 (10-03-2015) 29 March 2012 (29-03-2012) 03 May 2016 (03-05-2016)