



(12) **DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION**

(13) **A1**

(86) Date de dépôt PCT/PCT Filing Date: 2017/12/04
(87) Date publication PCT/PCT Publication Date: 2018/06/14
(85) Entrée phase nationale/National Entry: 2019/05/29
(86) N° demande PCT/PCT Application No.: US 2017/064540
(87) N° publication PCT/PCT Publication No.: 2018/106605
(30) Priorité/Priority: 2016/12/05 (US62/430,268)

(51) Cl.Int./Int.Cl. *G06F 17/00* (2019.01)
(71) Demandeur/Applicant:
MAGIC LEAP, INC., US
(72) Inventeurs/Inventors:
SANGER, GEORGE A., US;
SCHMIDT, BRIAN LLOYD, US;
TAJIK, ANASTASIA ANDREYEVNA, US;
O'GARA, TERRY MICHAEL, US;
SCHUMWAY, DAVID MATTHEW, US;
HOWARTH, ALAN, US...
(74) Agent: RICHES, MCKENZIE & HERBERT LLP

(54) Titre : TECHNIQUES DE CAPTURE AUDIO REPARTIE POUR DES SYSTEMES DE REALITE VIRTUELLE (VR), DE REALITE AUGMENTEE (AR) ET DE REALITE MIXTE (RM)
(54) Title: DISTRIBUTED AUDIO CAPTURING TECHNIQUES FOR VIRTUAL REALITY (VR), AUGMENTED REALITY (AR), AND MIXED REALITY (MR) SYSTEMS

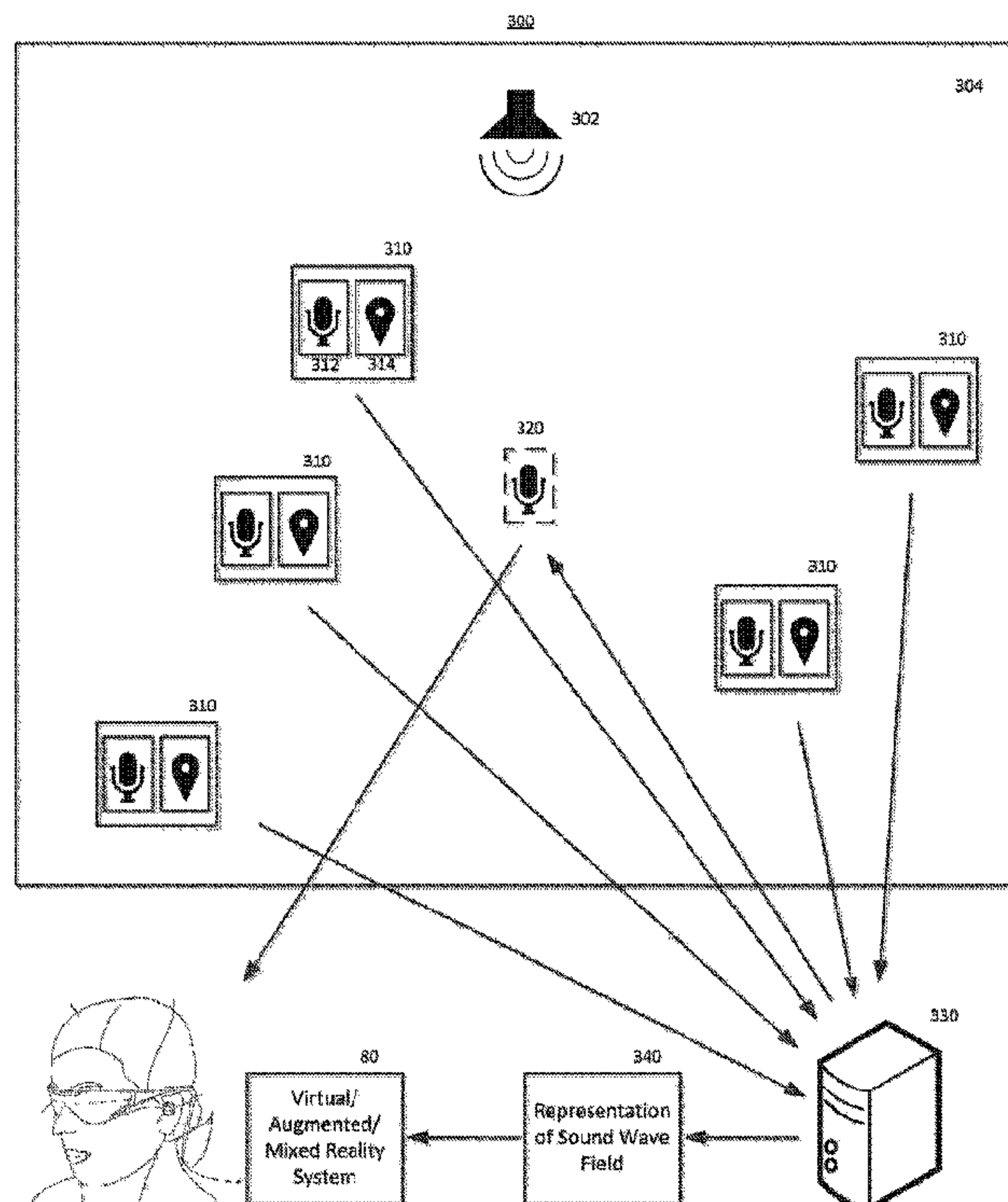


FIG. 3

(57) Abrégé/Abstract:

Systems and methods for capturing audio which can be used in applications such as virtual reality, augmented reality, and mixed reality systems. Some systems may include a plurality of distributed monitoring devices in an environment, each having a

(57) **Abrégé(suite)/Abstract(continued):**

microphone and a location tracking unit. The system can capture audio signals while also capturing location tracking signals which indicate the locations of the monitoring devices over time during capture of the audio signals. The system can generate a representation of at least a portion of a sound wave field in the environment based on the audio signals and the location tracking signals. The system may also determine one or more acoustic properties of the environment based on the audio signals and the location tracking signals.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization

International Bureau

(43) International Publication Date
14 June 2018 (14.06.2018)



(10) International Publication Number

WO 2018/106605 A1

(51) International Patent Classification:

G06F 17/00 (2006.01)

(21) International Application Number:

PCT/US2017/064540

(22) International Filing Date:

04 December 2017 (04.12.2017)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/430,268

05 December 2016 (05.12.2016) US

(71) Applicant: MAGIC LEAP, INC. [US/US]; 7500 W. Sunrise Blvd., Plantation, FL 33322 (US).

(72) Inventors: SANGER, George, A.; 7500 W. Sunrise Blvd., Plantation, FL 33322 (US). **SCHMIDT, Brian, Lloyd;** 7500 W. Sunrise Blvd., Plantation, FL 33322 (US). **TAJIK, Anastasia, Andreyevna;** 7500 W. Sunrise Blvd., Plantation, FL 33322 (US). **O'GARA, Terry, Michael;** 7500 W. Sunrise Blvd., Plantation, FL 33322 (US). **SCHUMWAY, David, Matthew;** 7500 W. Sunrise Blvd., Plantation, FL 33322 (US). **HOWARTH, Alan;** 7500 W. Sunrise Blvd., Plantation, FL 33322 (US).

(74) Agent: ALTMAN, Daniel, E.; Knobbe, Martens, Olson & Bear LLP, 2040 Main Street 14th Floor, Irvine, CA 92614 (US).

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

(54) Title: DISTRIBUTED AUDIO CAPTURING TECHNIQUES FOR VIRTUAL REALITY (VR), AUGMENTED REALITY (AR), AND MIXED REALITY (MR) SYSTEMS

(57) Abstract: Systems and methods for capturing audio which can be used in applications such as virtual reality, augmented reality, and mixed reality systems. Some systems may include a plurality of distributed monitoring devices in an environment, each having a microphone and a location tracking unit. The system can capture audio signals while also capturing location tracking signals which indicate the locations of the monitoring devices over time during capture of the audio signals. The system can generate a representation of at least a portion of a sound wave field in the environment based on the audio signals and the location tracking signals. The system may also determine one or more acoustic properties of the environment based on the audio signals and the location tracking signals.

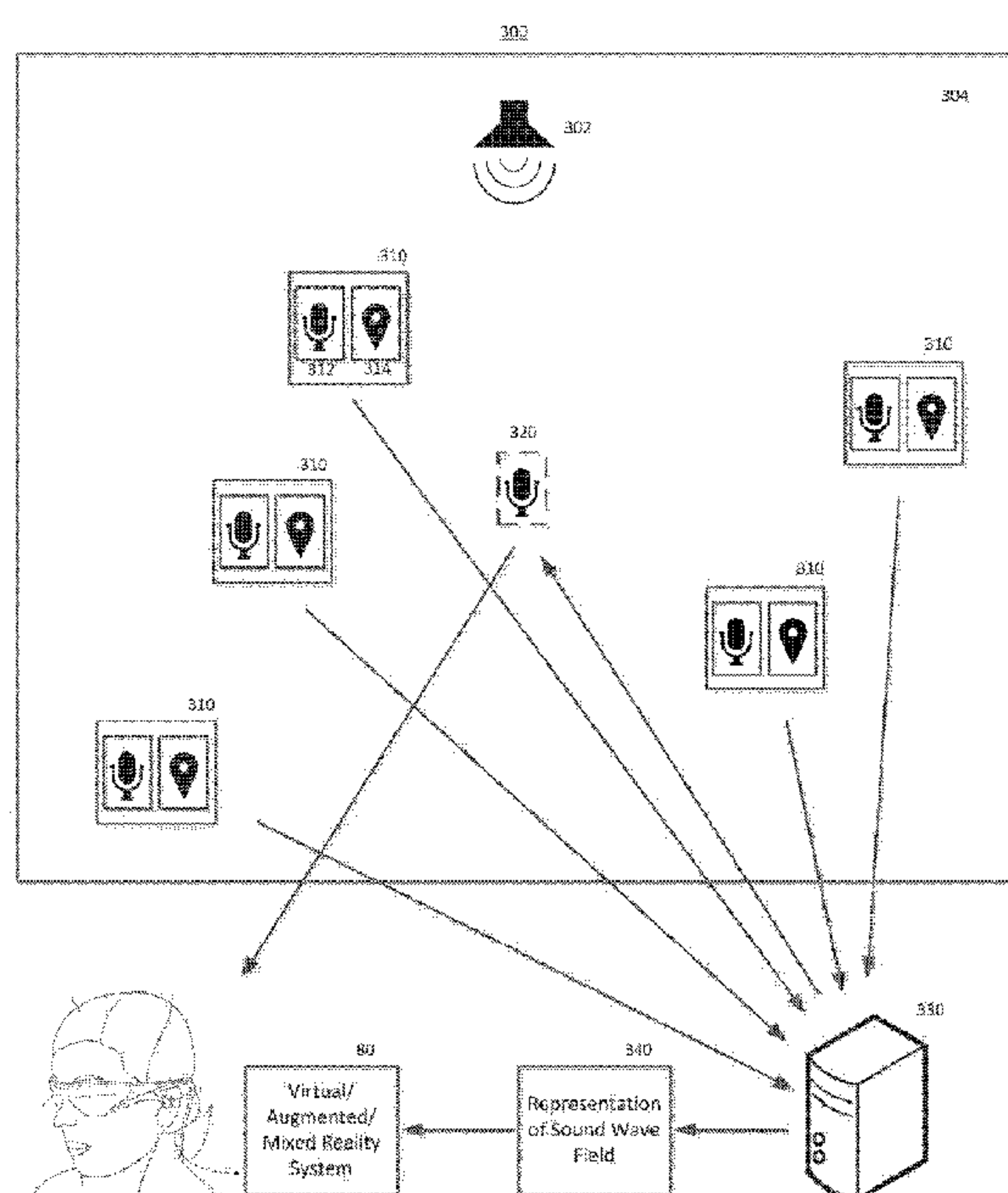


FIG. 3

WO 2018/106605 A1

Published:

— *with international search report (Art. 21(3))*

DISTRIBUTED AUDIO CAPTURING TECHNIQUES FOR VIRTUAL REALITY (VR), AUGMENTED REALITY (AR), AND MIXED REALITY (MR) SYSTEMS

INCORPORATION BY REFERENCE TO ANY PRIORITY APPLICATIONS

[0001] Any and all applications for which a foreign or domestic priority claim is identified in the Application Data Sheet as filed with the present application are hereby incorporated by reference under 37 CFR 1.57. Namely, this application claims priority to U.S. Provisional Patent Application No. 62/430,268, filed December 5, 2016, and entitled “DISTRIBUTED AUDIO CAPTURING TECHNIQUES FOR VIRTUAL REALITY (VR), AUGMENTED REALITY (AR), AND MIXED REALITY (MR) SYSTEMS,” the entirety of which is hereby incorporated by reference herein.

BACKGROUND

Field

[0002] This disclosure relates to distributed audio capturing techniques which can be used in applications such as virtual reality, augmented reality, and mixed reality systems.

Description of the Related Art

[0003] Modern computing and display technologies have facilitated the development of virtual reality, augmented reality, and mixed reality systems. Virtual reality, or “VR,” systems create a simulated environment for a user to experience. This can be done by presenting computer-generated imagery to the user through a head-mounted display. This imagery creates a sensory experience which immerses the user in the simulated environment. A virtual reality scenario typically involves presentation of only computer-generated imagery rather than also including actual real-world imagery.

[0004] Augmented reality systems generally supplement a real-world environment with simulated elements. For example, augmented reality, or “AR,” systems may provide a user with a view of the surrounding real-world environment via a head-mounted display. However, computer-generated imagery can also be presented on the display to enhance the real-world environment. This computer-generated imagery can include elements which are contextually-related to the real-world environment. Such elements can include simulated text, images, objects, etc. Mixed reality, or “MR,” systems

also introduce simulated objects into a real-world environment, but these objects typically feature a greater degree of interactivity than in AR systems.

[0005] Figure 1 depicts an example AR/MR scene 1 where a user sees a real-world park setting 6 featuring people, trees, buildings in the background, and a concrete platform 20. In addition to these items, computer-generated imagery is also presented to the user. The computer-generated imagery can include, for example, a robot statue 10 standing upon the real-world platform 20, and a cartoon-like avatar character 2 flying by which seems to be a personification of a bumble bee, even though these elements 2, 10 are not actually present in the real-world environment.

[0006] It can be challenging to produce VR/AR/MR technology that facilitates a natural-feeling, convincing presentation of virtual imagery elements. But audio can help make VR/AR/MR experiences more immersive. Thus, there is a need for improved audio techniques for these types of systems.

SUMMARY

[0007] In some embodiments, a system comprises: a plurality of distributed monitoring devices, each monitoring device comprising at least one microphone and a location tracking unit, wherein the monitoring devices are configured to capture a plurality of audio signals from a sound source and to capture a plurality of location tracking signals which respectively indicate the locations of the monitoring devices over time during capture of the plurality of audio signals; and a processor configured to receive the plurality of audio signals and the plurality of location tracking signals, the processor being further configured to generate a representation of at least a portion of a sound wave field created by the sound source based on the audio signals and the location tracking signals.

[0008] In some embodiments, a device comprises: a processor configured to carry out a method comprising receiving, from a plurality of distributed monitoring devices, a plurality of audio signals captured from a sound source; receiving, from the plurality of monitoring devices, a plurality of location tracking signals, the plurality of location tracking signals respectively indicating the locations of the monitoring devices over time during capture of the plurality of audio signals; generating a representation of at least a portion of a sound wave field created by the sound source based on the audio signals and the location tracking signals; and a memory to store the audio signals and the location tracking signals.

[0009] In some embodiments, a method comprises: receiving, from a plurality of distributed monitoring devices, a plurality of audio signals captured from a sound source; receiving, from the plurality of monitoring devices, a plurality of location tracking signals, the plurality of location tracking signals respectively indicating the locations of the monitoring devices over time during capture of the plurality of audio signals; generating a representation of at least a portion of a sound wave field created by the sound source based on the audio signals and the location tracking signals.

[0010] In some embodiments, a system comprises: a plurality of distributed monitoring devices, each monitoring device comprising at least one microphone and a location tracking unit, wherein the monitoring devices are configured to capture a plurality of audio signals in an environment and to capture a plurality of location tracking signals which respectively indicate the locations of the monitoring devices over time during capture of the plurality of audio signals; and a processor configured to receive the plurality of audio signals and the plurality of location tracking signals, the processor being further configured to determine one or more acoustic properties of the environment based on the audio signals and the location tracking signals.

[0011] In some embodiments, a device comprises: a processor configured to carry out a method comprising receiving, from a plurality of distributed monitoring devices, a plurality of audio signals captured in an environment; receiving, from the plurality of monitoring devices, a plurality of location tracking signals, the plurality of location tracking signals respectively indicating the locations of the monitoring devices over time during capture of the plurality of audio signals; determining one or more acoustic properties of the environment based on the audio signals and the location tracking signals; and a memory to store the audio signals and the location tracking signals.

[0012] In some embodiments, a method comprises: receiving, from a plurality of distributed monitoring devices, a plurality of audio signals captured in an environment; receiving, from the plurality of monitoring devices, a plurality of location tracking signals, the plurality of location tracking signals respectively indicating the locations of the monitoring devices over time during capture of the plurality of audio signals; and determining one or more acoustic properties of the environment based on the audio signals and the location tracking signals.

[0013] In some embodiments, a system comprises: a plurality of distributed video cameras located about the periphery of a space so as to capture a plurality of videos of a central portion of the space from a plurality of different viewpoints; a plurality of distributed microphones located about the periphery of the space so as to capture a plurality of audio signals during the capture of the plurality of videos; and a processor configured to receive the plurality of videos, the plurality of audio signals, and location information about the position of each microphone within the space, the processor being further configured to generate a representation of at least a portion of a sound wave field for the space based on the audio signals and the location information.

[0014] In some embodiments, a device comprises: a processor configured to carry out a method comprising receiving, from a plurality of distributed video cameras, a plurality of videos of a scene captured from a plurality of viewpoints; receiving, from a plurality of distributed microphones, a plurality of audio signals captured during the capture of the plurality of videos; receiving location information about the positions of the plurality of microphones; and generating a representation of at least a portion of a sound wave field based on the audio signals and the location information; and a memory to store the audio signals and the location tracking signals.

[0015] In some embodiments, a method comprises: receiving, from a plurality of distributed video cameras, a plurality of videos of a scene captured from a plurality of viewpoints; receiving, from a plurality of distributed microphones, a plurality of audio signals captured during the capture of the plurality of videos; receiving location information about the positions of the plurality of microphones; and generating a representation of at least a portion of a sound wave field based on the audio signals and the location information.

BRIEF DESCRIPTION OF THE DRAWINGS

[0016] Figure 1 illustrates a user's view of an augmented/mixed reality scene using an example AR/MR system.

[0017] Figure 2 shows an example VR/AR/MR system.

[0018] Figure 3 illustrates a system for using a plurality of distributed devices to create a representation of a sound wave field.

[0019] Figure 4 is a flowchart which illustrates an example embodiment of a method of operation of the system shown in Figure 3 for creating a sound wave field.

[0020] Figure 5 illustrates a web-based system for using a plurality of user devices to create a representation of a sound wave field for an event.

[0021] Figure 6 is a flowchart which illustrates an example embodiment of operation of the web-based system shown in Figure 5 for creating a sound wave field of an event.

[0022] Figure 7 illustrates an example embodiment of a system which can be used to determine acoustic properties of an environment.

[0023] Figure 8 is a flowchart which illustrates an example embodiment of a method for using the system shown in Figure 7 to determine one or more acoustic properties of an environment.

[0024] Figure 9 illustrates an example system for performing volumetric video capture.

[0025] Figure 10 illustrates an example system for capturing audio during volumetric video capture.

[0026] Figure 11 is a flow chart which shows an example method for using the system shown in Figure 10 to capture audio for a volumetric video.

DETAILED DESCRIPTION

[0027] Figure 2 shows an example virtual/augmented/mixed reality system 80. The virtual/augmented/mixed reality system 80 includes a display 62, and various mechanical and electronic modules and systems to support the functioning of that display 62. The display 62 may be coupled to a frame 64, which is wearable by a user 60 and which is configured to position the display 62 in front of the eyes of the user 60. In some embodiments, a speaker 66 is coupled to the frame 64 and positioned adjacent the ear canal of the user (in some embodiments, another speaker, not shown, is positioned adjacent the other ear canal of the user to provide for stereo/shapeable sound control). The display 62 is operatively coupled, such as by a wired or wireless connection 68, to a local data processing module 70 which may be mounted in a variety of configurations, such as attached to the frame 64, attached to a helmet or hat worn by the user, embedded in headphones, or otherwise removably attached to the user 60 (e.g., in a backpack-style configuration, in a belt-coupling style configuration, etc.).

[0028] The local processing and data module 70 may include a processor, as well as digital memory, such as non-volatile memory (e.g., flash memory), both of which may be utilized to assist in the processing and storing of data. This includes data captured from local sensors provided as part of the system 80, such as image monitoring devices (e.g., cameras), microphones, inertial measurement units, accelerometers, compasses, GPS units, radio devices, and/or gyros. The local sensors may be operatively coupled to the frame 64 or otherwise attached to the user 60. Alternatively, or additionally, sensor data may be acquired and/or processed using a remote processing module 72 and/or remote data repository 74, possibly for passage to the display 62 and/or speaker 66 after such processing or retrieval. In some embodiments, the local processing and data module 70 processes and/or stores data captured from remote sensors, such as those in the audio/location monitoring devices 310 shown in Figure 3, as discussed herein. The local processing and data module 70 may be operatively coupled by communication links (76, 78), such as via a wired or wireless communication links, to the remote processing module 72 and remote data repository 74 such that these remote modules (72, 74) are operatively coupled to each other and available as resources to the local processing and data module 70. In some embodiments, the remote data repository 74 may be available through the internet or other networking configuration in a “cloud” resource configuration.

SOUND WAVE FIELD CAPTURE AND USAGE IN VR, AR, AND MR SYSTEMS

[0029] This section relates to using audio recordings from multiple distributed devices to create a representation of at least a portion of a sound wave field which can be used in applications such as virtual reality (VR), augmented reality (AR), and mixed reality (MR) systems.

[0030] Sounds result from pressure variations in a medium such as air. These pressure variations are generated by vibrations at a sound source. The vibrations from the sound source then propagate through the medium as longitudinal waves. These waves are made up of alternating regions of compression (increased pressure) and rarefaction (reduced pressure) in the medium.

[0031] Various quantities can be used to characterize the sound at a point in space. These can include, for example, pressure values, vibration amplitudes, frequencies, or other quantities. A sound wave field generally consists of a collection of one or more such

sound-defining quantities at various points in space and/or various points in time. For example, a sound wave field can consist of a measurement or other characterization of the sound present at each point on a spatial grid at various points in time. Typically, the spatial grid of a sound wave field consists of regularly spaced points and the measurements of the sound are taken at regular intervals of time. But the spatial and/or temporal resolution of the sound wave field can vary depending on the application. Certain models of the sound wave field, such as representation by a set of point sources, can be evaluated at arbitrary locations specified by floating point coordinates and not tied to a predefined grid.

[0032] A sound wave field can include a near field region relatively close to the sound source and a far field region beyond the near field region. The sound wave field can be made up of sound waves which propagate freely from the source without obstruction and of waves that reflect from objects within the region or from the boundaries of the region.

[0033] Figure 3 illustrates a system 300 for using a plurality of distributed devices 310 to create a representation of a sound wave field 340. In some embodiments, the system 300 can be used to provide audio for a VR/AR/MR system 80, as discussed further herein. As shown in Figure 3, a sound source 302 projects sound into an environment 304. The sound source 302 can represent, for example, a performer, an instrument, an audio speaker, or any other source of sound. The environment 304 can be any indoor or outdoor space including, for example, a concert hall, an amphitheater, a conference room, etc. Although only a single sound source 302 is illustrated, the environment 304 can include multiple sound sources. And the multiple sound sources can be distributed throughout the environment 304 in any manner.

[0034] The system 300 includes a plurality of distributed audio and/or location monitoring devices 310. Each of these devices can be physically distinct and can operate independently. The monitoring devices 310 can be mobile (e.g., carried by a person) and can be spaced apart in a distributed manner throughout the environment 304. There need not be any fixed relative spatial relationship between the monitoring devices 310. Indeed, as the monitoring devices 310 are independently mobile, the spatial relationship between the various devices 310 can vary over time. Although five monitoring devices 300 are illustrated, any number of monitoring devices can be used. Further, although Figure 3 is a two-dimensional drawing and therefore shows the monitoring devices 300 as being

distributed in two dimensions, they can also be distributed throughout all three dimensions of the environment 304.

[0035] Each monitoring device 310 includes at least one microphone 312. The microphones 312 can be, for example, isotropic or directional. Useable microphone pickup patterns can include, for example, cardioid, hyper cardioid, and supercardioid. The microphones 312 can be used by the monitoring devices 310 to capture audio signals by transducing sounds from one or more sound sources 302 into electrical signals. In some embodiments, the monitoring devices 310 each include a single microphone and record monaural audio. But in other embodiments the monitoring devices 310 can include multiple microphones and can capture, for example, stereo audio. Multiple microphones 312 can be used to determine the angle-of-arrival of sound waves at each monitoring device 310.

[0036] Although not illustrated, the monitoring devices 310 can also each include a processor and a storage device for locally recording the audio signal picked up by the microphone 312. Alternatively and/or additionally, each monitoring device 310 can include a transmitter (e.g., a wireless transmitter) to allow captured sound to be digitally encoded and transmitted in real-time to one or more remote systems or devices (e.g., processor 330). Upon receipt at a remote system or device, the captured sound can be used to update a stored model of the acoustic properties of the space in which the sound was captured, or it can be used to create a realistic facsimile of the captured sound in a VR/AR/MR experience, as discussed further herein.

[0037] Each monitoring device 310 also includes a location tracking unit 314. The location tracking unit 314 can be used to track the location of the monitoring device 310 within the environment 304. Each location tracking unit 314 can express the location of its corresponding monitoring device 310 in an absolute sense or in a relative sense (e.g., with respect to one or more other components of the system 300). In some embodiments, each location tracking unit 314 creates a location tracking signal, which can indicate the location of the monitoring device 310 as a function of time. For example, a location tracking signal could include a series of spatial coordinates indicating where the monitoring device 310 was located at regular intervals of time.

[0038] In some embodiments, the location tracking units 314 directly measure location. One example of such a location tracking unit 314 is a Global Positioning System

(GPS). In other embodiments, the location tracking units 314 indirectly measure location. For example, these types of units may infer location based on other measurements or signals. An example of this type of location tracking unit 314 is one which analyzes imagery from a camera to extract features which provide location cues. Monitoring devices 310 can also include audio emitters (e.g., speakers) or radio emitters. Audio or radio signals can be exchanged between monitoring devices and multilateration and/or triangulation can be used to determine the relative locations of the monitoring devices 310.

[0039] The location tracking units 314 may also measure and track not just the locations of the monitoring devices 310 but also their spatial orientations using, for example, gyroscopes, accelerometers, and/or other sensors. In some embodiments, the location tracking units 314 can combine data from multiple types of sensors in order to determine the location and/or orientation of the monitoring devices 310.

[0040] The monitoring devices 310 can be, for example, smart phones, tablet computers, laptop computers, etc. (as shown in Figure 5). Such devices are advantageous because they are ubiquitous and often have microphones, GPS units, cameras, gyroscopes, accelerometers, and other sensors built in. The monitoring devices 310 may also be wearable devices, such as VR/AR/MR systems 80.

[0041] The system 300 shown in Figure 3 also includes a processor 330. The processor 330 can be communicatively coupled with the plurality of distributed monitoring devices 310. This is illustrated by the arrows from the monitoring devices 310 to the processor 330, which represent communication links between the respective monitoring devices 310 and the processor 330. The communication links can be wired or wireless according to any communication standard or interface. The communication links between the respective monitoring devices 310 and the processor 330 can be used to download audio and location tracking signals to the processor 330. In some embodiments, the processor 330 can be part of the VR/AR/MR system 80 shown in Figure 1. For example, the processor 330 could be the local processing module 70 or the remote processing module 72.

[0042] The processor 330 includes an interface which can be used to receive the respective captured audio signals and location tracking signals from the monitoring devices 310. The audio signals and location tracking signals can be uploaded to the processor 330 in real time as they are captured, or they can be stored locally by the monitoring devices 310

and uploaded after completion of capture for some time interval or for some events, etc. The processor 330 can be a general purpose or specialized computer and can include volatile and/or non-volatile memory/storage for processing and storing the audio signals and the location tracking signals from the plurality of distributed audio monitoring devices 310. The operation of the system 300 will now be discussed with respect to Figure 4.

[0043] Figure 4 is a flowchart which illustrates an example embodiment of a method 400 of operation of the system 300 shown in Figure 3. At blocks 410a and 410b, which are carried out concurrently, the monitoring devices 310 capture audio signals from the sound source 302 at multiple distributed locations throughout the environment 304 while also tracking their respective locations. Each audio signal may typically be a digital signal made up of a plurality of sound measurements taken at different points in time, though analog audio signals can also be used. Each location tracking signal may also typically be a digital signal which includes a plurality of location measurements taken at different points in time. The resulting audio signals and location tracking signals from the monitoring devices 310 can both be appropriately time stamped so that each interval of audio recording can be associated with a specific location within the environment 304. In some embodiments, sound samples and location samples are synchronously taken at regular intervals in time, though this is not required.

[0044] At block 420, the processor 330 receives the audio signals and the tracking signals from the distributed monitoring devices 310. The signals can be uploaded from the monitoring devices 310 on command or automatically at specific times or intervals. Based on timestamp data in the audio and location tracking signals, the processor 330 can synchronize the various audio and location tracking signals received from the plurality of monitoring devices 310.

[0045] At block 430, the processor 330 analyzes the audio signals and tracking signals to generate a representation of at least a portion of the sound wave field within the environment 304. In some embodiments, the environment 304 is divided into a grid of spatial points and the sound wave field includes one or more values (e.g., sound measurements) per spatial point which characterize the sound at that spatial point at a particular point in time or over a period of time. Thus, the data for each spatial point on the grid can include a time series of values which characterize the sound at that spatial point over

time. (The spatial and time resolution of the sound wave field can vary depending upon the application, the number of monitoring devices 310, the time resolution of the location tracking signals, etc.)

[0046] In general, the distributed monitoring devices 310 only perform actual measurements of the sound wave field at a subset of locations on the grid of points in the environment 304. In addition, as the monitoring devices 310 are mobile, the specific subset of spatial points represented with actual sound measurements at each moment in time can vary. Thus, the processor 330 can use various techniques to estimate the sound wave field for the remaining spatial points and times so as to approximate the missing information. For example, the sound wave field can be approximately reproduced by simulating a set of point sources of sound where each point source in the set corresponds in location to a particular one of the monitoring devices and outputs audio that was captured by the particular one of the monitoring devices. In addition, multilateration, triangulation or other localization methods based on the audio segments received at the monitoring devices 310 can be used to determine coordinates of sound sources and then a representation of the sound wave field that is included in virtual content can include audio segments emanating from the determined coordinates (i.e., a multiple point source model). Although the sound wave field may comprise a large number of spatial points, it should be understood that the processor 330 need not necessarily calculate the entire sound wave field but rather can calculate only a portion of it, as needed based on the application. For example, the processor 330 may only calculate the sound wave field for a specific spatial point of interest. This process can be performed iteratively as the spatial point of interest changes.

[0047] The processor 330 can also perform sound localization to determine the location(s) of, and/or the direction(s) toward, one or more sound sources 302 within the environment 304. Sound localization can be done according to a number of techniques, including the following (and combinations of the same): comparison of the respective times of arrival of certain identified sounds at different locations in the environment 304; comparison of the respective magnitudes of certain identified sounds at different locations in the environment 304; comparison of the magnitudes and/or phases of certain frequency components of certain identified sounds at different locations in the environment 304. In some embodiments, the processor 330 can compute the cross correlation between audio

signals received at different monitoring devices 310 in order to determine the Time Difference of Arrival (TDOA) and then use multilateration to determine the location of the audio source(s). Triangulation may also be used. The processor 330 can also extract audio from an isolated sound source. A time offset corresponding to the TDOA for each monitoring device from a particular audio source can be subtracted from each corresponding audio track captured by a set of the monitoring devices in order to synchronize the audio content from the particular source before summing audio tracks in order to amplify the particular source. The extracted audio can be used in a VR/AR/MR environment, as discussed herein.

[0048] The processor 330 can also perform transforms on the sound wave field as a whole. For example, by applying a stored source elevation, azimuth, and distance (θ , ϕ , r) dependent Head Related Transfer Functions (HRTF), the processor 330 can modify captured audio for output through left and right speaker channels for any position and orientation relative to the sound source in a virtual coordinate system. Additionally, the processor 330 can apply rotational transforms to the sound wave field. In addition, since the processor 330 can extract audio from a particular sound source 302 within the environment, that source can be placed and/or moved to any location within a modeled environment by using three dimensional audio processing.

[0049] Once the processor 330 has calculated a representation of the sound wave field 340, it can be used to estimate the audio signal which would have been detected by a microphone at any desired location within the sound wave field. For example, Figure 3 illustrates a virtual microphone 320. The virtual microphone 320 is not a hardware device which captures actual measurements of the sound wave field at the location of the virtual microphone 320. Instead, the virtual microphone 320 is a simulated construct which can be placed at any location within the environment 304. Using the representation of the sound wave field 340 within the environment 304, the processor 330 can determine a simulated audio signal which is an estimate of the audio signal which would have been detected by a physical microphone located at the position of the virtual microphone 320. This can be done by, for example, determining the grid point in the sound wave field nearest to the location of the virtual microphone for which sound data is available and then associating that sound data with the virtual microphone. In other embodiments, the simulated audio signal from the

virtual microphone 320 can be determined by, for example, interpolating between audio signals from multiple grid points in the vicinity of the virtual microphone. The virtual microphone 320 can be moved about the environment 304 (e.g., using a software control interface) to any location at any time. Accordingly, the process of associating sound data with the virtual microphone 320 based on its current location can be repeated iteratively over time as the virtual microphone moves.

[0050] The method 400 can continue on to blocks 440-460. In these blocks, the representation of the sound wave field 340 can be provided to a VR/AR/MR system 80, as shown in Figure 3. As already discussed, the VR/AR/MR system 80 can be used to provide a simulated experience within a virtual environment or an augmented/mixed reality experience within an actual environment. In the case of a virtual reality experience, the sound wave field 340, which has been collected from a real world environment 304, can be transferred or mapped to a simulated virtual environment. In the case of an augmented and/or mixed reality experience, the sound wave field 340 can be transferred or mapped from one real world environment 304 to another.

[0051] Whether the environment experienced by the user is an actual environment or a virtual one, at block 440 of Figure 4, the VR/AR/MR system 80 can determine the location and/or orientation of the user within the virtual or actual environment as the user moves around within the environment. Based on the location and/or orientation of the user within the virtual or actual environment, the VR/AR/MR system 80 (or the processor 330) can associate the location of the user with a point in the representation of the sound wave field 340.

[0052] At block 450 of Figure 4, the VR/AR/MR reality system 80 (or the processor 330) can generate a simulated audio signal that corresponds to the location and/or orientation of the user within the sound wave field. For example, as discussed herein, one or more virtual microphones 320 can be positioned at the location of the user and the system 80 (or the processor 330) can use the representation of the sound wave field 340 in order to simulate the audio signal which would have been detected by an actual microphone at that location.

[0053] At block 460, the simulated audio signal from a virtual microphone 320 is provided to the user of the VR/AR/MR system 80 via, for example, headphones worn by the

user. Of course, the user of the VR/AR/MR reality system 80 can move about within the environment. Therefore, blocks 440-460 can be repeated iteratively as the position and/or orientation of the user within the sound wave field changes. In this way, the system 300 can be used to provide a realistic audio experience to the user of the VR/AR/MR system 80 as if he or she were actually present at any point within the environment 304 and could move about through it.

[0054] Figure 5 illustrates a web-based system 500 for using a plurality of user devices 510 to create a representation of a sound wave field for an event. The system 500 includes a plurality of user devices 510 for capturing audio at an event, such as a concert. The user devices 510 are, for example, smart phones, tablet computers, laptop computers, etc. belonging to attendees of the event. Similar to the audio/location monitoring devices 310 discussed with respect to Figure 3, the user devices 510 in Figure 5 each include at least one microphone and a location tracking unit, such as GPS. The system also includes a web-based computer server 530 which is communicatively coupled to the user devices 510 via the Internet. Operation of the system 400 is discussed with respect to Figure 6.

[0055] Figure 6 is a flowchart which illustrates an example embodiment of operation of the web-based system shown in Figure 5 for creating a sound wave field of an event. At block 610, the computer server 530 provides a mobile device application for download by users. The mobile device application is one which, when installed on a smartphone or other user device, allows users to register for events and to capture audio signals and location tracking signals during the event. Although Figure 6 shows that the computer server 530 offers the mobile device application for download, the application could also be provided for download on other servers, such as third party application stores.

[0056] At block 620, users download the application to their devices 510 and install it. The application can provide a list of events where it can be used to help create a sound wave field of the event. The users select and register for an event at which they will be in attendance.

[0057] At block 630, during the event, the application allows users to capture audio from their seats and/or as they move about through the venue. The application also creates a location tracking signal using, for example, the device's built-in GPS. The operation of the devices 410, including the capturing of audio and location tracking signals,

can be as described herein with respect to the operation of the audio/location monitoring devices 310.

[0058] At block 640, users' devices upload their captured audio signals and location tracking signals to the computer server 530 via the Internet. The computer server 530 then processes the audio signals and location tracking signals in order to generate a representation of a sound wave field for the event. This processing can be done as described herein with respect to the operation of the processor 330.

[0059] Finally, at block 660, the computer server 530 offers simulated audio signals (e.g., from selectively positioned virtual microphones) to users for download. The audio signal from a virtual microphone can be created from the sound wave field for the event using the techniques discussed herein. Users can select the position of the virtual microphone via, for example, a web-based interface. In this way, attendees of the event can use the mobile application to experience audio from the event from different locations within the venue and with different perspectives. The application therefore enhances the experience of attendees at a concert or other event.

[0060] While the computer server 530 may calculate a sound wave field for the event, as just discussed, other embodiments may use different techniques for allowing users to experience audio from a variety of locations at the event venue. For example, depending upon the density of registered users at the event, the audio signal from a virtual microphone may simply correspond to the audio signal captured by the registered user nearest the location of the virtual microphone. As the position of the virtual microphone changes, or as the nearest registered user varies due to movements of the registered users during the event, the audio from the virtual microphone can be synthesized by cross-fading from the audio signal captured by one registered user to the audio signal captured by another registered user.

DETERMINATION OF ENVIRONMENTAL ACOUSTIC INFORMATION USING VR, AR, AND MR SYSTEMS

[0061] As already discussed, VR, AR, and MR systems use a display 62 to present virtual imagery to a user 60, including simulated text, images, and objects, in a virtual or real world environment. In order for the virtual imagery to be realistic, it is often accompanied by sound effects and other audio. This audio can be made more realistic if the acoustic properties of the environment are known. For example, if the location and type of

acoustic reflectors present in the environment are known, then appropriate audio processing can be performed to add reverb or other effects so as to make the audio sound more convincingly real.

[0062] But in the case of AR and MR systems in particular, it can be difficult to determine the acoustic properties of the real world environment where the simulated experience is occurring. Without knowledge of the acoustic properties of the environment, including the type, location, size, etc. of acoustic reflectors and absorbers such as walls, floors, ceilings, and objects, it can be difficult to apply appropriate audio processing to provide a realistic audio environment. For example, without knowledge of the acoustic characteristics of the environment, it can be difficult to realistically add spatialization to simulated objects so as to make their sound effects seem authentic in that environment. There is thus a need for improved techniques for determining acoustic characteristics of an environment so that such acoustic characteristics can be employed in the acoustic models and audio processing used in VR/AR/MR systems.

[0063] Figure 7 illustrates an example embodiment of a system 700 which can be used to determine acoustic properties of an environment 704. As shown in Figure 7, four users 60a, 60b, 60c, and 60d are present in the environment 704. The environment 704 can be, for example, a real world environment being used to host an AR or MR experience. Each user 60 has an associated device 80a, 80b, 80c, and 80d. In some embodiments, these devices are VR/AR/MR systems 80 that the respective users 60 are wearing. These systems 80 can each include a microphone 712 and a location tracking unit 714. The VR/AR/MR systems 80 can also include other sensors, including cameras, gyroscopes, accelerometers, and audio speakers.

[0064] The system 700 also includes a processor 730 which is communicatively coupled to the VR/AR/MR systems 80. In some embodiments, the processor 730 is a separate device from the VR/AR/MR systems 80, while in others the processor 730 is a component of one of these systems.

[0065] The microphone 712 of each VR/AR/MR system 80 can be used to capture audio of sound sources in the environment 704. The captured sounds can include both known source sounds which have not been significantly affected by the acoustic properties of the environment 704 and environment-altered versions of the source sounds

after they have been affected by the acoustic properties of environment. Among these are spoken words and other sounds made by the users 60, sounds emitted by any of the VR/AR/MR systems 80, and sounds from other sound sources which may be present in the environment 704.

[0066] Meanwhile, the location tracking units 714 can be used to determine the location of each user 60 within the environment 704 while these audio recordings are being made. In addition, sensors such as gyroscopes and accelerometers can be used to determine the orientation of the users 60 while speaking and/or the orientation of the VR/AR/MR systems 80 when they emit or capture sounds. The audio signals and the location tracking signals can be sent to the processor 730 for analysis. The operation of the system 700 will now be described with respect to Figure 8.

[0067] Figure 8 is a flowchart which illustrates an example embodiment of a method 800 for using the system 700 shown in Figure 7 to determine one or more acoustic properties of an environment 704. The method 800 begins at blocks 810a and 810b, which are carried out concurrently. In these blocks, the VR/AR/MR systems 80 capture audio signals at multiple distributed locations throughout the environment 704 while also tracking their respective locations and/or orientations. Once again, each audio signal may typically be a digital signal made up of a plurality of sound measurements taken at different points in time, though analog audio signals can also be used. Each location tracking signal may also typically be a digital signal which includes a plurality of location and/or orientation measurements taken at different points in time. The resulting audio signals and location tracking signals from the VR/AR/MR systems 80 can both be appropriately time stamped so that each interval of audio recording can be associated with a specific location within the environment 704. In some embodiments, sound samples and location samples are synchronously taken at regular intervals in time, though this is not required.

[0068] For the processing described later with respect to block 830, it can be advantageous to have an audio copy of at least two types of sounds: 1) known source sounds which are either known a priori or are captured prior to the source sound having been significantly affected by the acoustics of the environment 704; and 2) environment-altered sounds which are captured after having been significantly affected by the acoustics of the environment 704.

[0069] In some embodiments, one or more of the VR/AR/MR systems 80 can be used to emit a known source sound from an audio speaker, such as an acoustic impulse or one or more acoustic tones (e.g., a frequency sweep of tones within the range of about 20 Hz to about 20 kHz, which is approximately the normal range of human hearing). If the system 80a is used to emit a known source sound, then the microphones of the remaining systems 80b, 80c, and 80d can be used to acquire the corresponding environment-altered sounds. Acoustic impulses and frequency sweeps can be advantageous because they can be used to characterize the acoustic frequency response of the environment 704 for a wide range of frequencies, including the entire range of frequencies which are audible to the human ear. But sounds outside the normal range of human hearing can also be used. For example, ultrasonic frequencies can be emitted by the VR/AR/MR systems 80 and used to characterize one or more acoustic and/or spatial properties of the environment 704.

[0070] As an alternative to using known source sounds emitted by the VR/AR/MR systems 80 themselves, captured audio of spoken words or other sounds made by one or more of the users 60 can also be used as known source sounds. This can be done by using a user's own microphone to capture his or her utterances. For example, the microphone 712a of the VR/AR/MR system 80a corresponding to user 60a can be used to capture audio of him or her speaking. Because the sounds from user 60a are captured by his or her own microphone 712a before being significantly affected by acoustic reflectors and/or absorbers in the environment 704, these recordings by the user's own microphone can be considered and used as known source sound recordings. The same can be done for the other users 60b, 60c, and 60d using their respective microphones 712b, 712c, and 712d. Of course, some processing can be performed on these audio signals to compensate for differences between a user's actual utterances and the audio signal that is picked up by his or her microphone. (Such differences can be caused by effects such as a user's microphone 712a not being directly located within the path of sound waves emitted from the user's mouth.) Meanwhile, the utterances from one user can be captured by the microphones of other users to obtain environment-altered versions of the utterances. For example, the utterances of user 60a can be captured by the respective VR/AR/MR systems 80b, 80c, and 80d of the remaining users 60b, 60c, and 60d and these recordings can be used as the environment-altered sounds.

[0071] In this way, utterances from the users 60 can be used to determine the acoustic frequency response and other characteristics of the environment 704, as discussed further herein. While any given utterance from a user may not include diverse enough frequency content to fully characterize the frequency response of the environment 704 across the entire range of human hearing, the system 700 can build up the frequency response of the environment iteratively over time as utterances with new frequency content are made by the users 60.

[0072] In addition to using sounds to determine acoustic characteristics such as the frequency response of the environment 704, they can also be used to determine information about the spatial characteristics of the environment 704. Such spatial information may include, for example, the location, size, and/or reflective/absorptive properties of features within the environment. This can be accomplished because the location tracking units 714 within the VR/AR/MR systems 80 can also measure the orientation of the users 60 when making utterances or the orientation of the systems 80 when emitting or capturing sounds. As already mentioned, this can be accomplished using gyroscopes, accelerometers, or other sensors built into the wearable VR/AR/MR systems 80. Because the orientation of the users 60 and VR/AR/MR systems 80 can be measured, the direction of propagation of any particular known source sound or environment-altered sound can be determined. This information can be processed using sonar techniques to determine characteristics about the environment 704, including sizes, shapes, locations, and/or other characteristics of acoustic reflectors and absorbers within the environment.

[0073] At block 820, the processor 730 receives the audio signals and the tracking signals from the VR/AR/MR systems 80. The signals can be uploaded on command or automatically at specific times or intervals. Based on timestamp data in the audio and location tracking signals, the processor 730 can synchronize the various audio and location tracking signals received from the VR/AR/MR systems 80.

[0074] At block 830, the processor 730 analyzes the audio signals and tracking signals to determine one or more acoustic properties of the environment 704. This can be done, for example, by identifying one or more known source sounds from the audio signals. The known source sounds may have been emitted at a variety of times from a variety of locations within the environment 704 and in a variety of directions. The times can be

determined from timestamp data in the audio signals, while the locations and directions can be determined from the location tracking signals.

[0075] The processor 730 may also identify and associate one or more environment-altered sounds with each known source sound. The processor 730 can then compare each known source sound with its counterpart environment-altered sound(s). By analyzing differences in frequency content, phase, time of arrival, etc., the processor 730 can determine one or more acoustic properties of the environment 730 based on the effect of the environment on the known source sounds. The processor 730 can also use sonar processing techniques to determine spatial information about the locations, sizes, shapes, and characteristics of objects or surfaces within the environment 704.

[0076] At block 840, the processor 730 can transmit the determined acoustic properties of the environment 704 back to the VR/AR/MR systems 80. These acoustic properties can include the acoustic reflective/absorptive properties of the environment, the sizes, locations, and shapes of objects within the space, etc. Because there are multiple monitoring devices, certain of those devices will be closer to each sound source and will therefore likely be able to obtain a purer recording of the original source. Other monitoring devices at different locations will capture sound with varying degrees of reverberation added. By comparing such signals the character of the reverberant properties (e.g., a frequency dependent reverberation decay time) of the environment can be assessed and stored for future use in generating more realistic virtual sound sources. The frequency dependent reverberation time can be stored for multiple positions of monitoring devices and interpolation can be used to obtain values for other positions.

[0077] Then, at block 850, the VR/AR/MR systems 80 can use the acoustic properties of the environment 704 to enhance the audio signals played to the users 60 during VR/AR/MR experiences. The acoustic properties can be used to enhance sound effects which accompany virtual objects which are displayed to the users 60. For example the frequency dependent reverberation corresponding to a position of user of the VR/AR/MR system 80 can be applied to virtual sound sources output through the VR/AR/MR system 80.

AUDIO CAPTURE FOR VOLUMETRIC VIDEOS

[0078] Distributed audio/location monitoring devices of the type described herein can also be used to capture audio for volumetric videos. Figure 9 illustrates an example

system 900 for performing volumetric video capture. The system 900 is located in an environment 904, which is typically a green screen room. A green screen room is a room with a central space 970 surrounded by green screens of the type used in chroma key compositing, which is a conventional post-production video processing technique for compositing images or videos based on their color content.

[0079] The system 900 includes a plurality of video cameras 980 set up at different viewpoints around the perimeter of the green screen room 904. Each of the video cameras 980 is aimed at the central portion 970 of the green screen room 904 where the scene that is to be filmed is acted out. As the scene is acted out, the video cameras 980 film it from a discrete number of viewpoints spanning a 360° range around the scene. The videos from these cameras 980 can later be mathematically combined by a processor 930 to simulate video imagery which would have been captured by a video camera located at any desired viewpoint within the environment 904, including viewpoints between those which were actually filmed by the cameras 980.

[0080] This type of volumetric video can be effectively used in VR/AR/MR systems because it can permit users of these systems to experience the filmed scene from any vantage point. The user can move in the virtual space around the scene and experience it as if its subject were actually present before the user. Thus, volumetric video offers the possibility of providing a very immersive VR/AR/MR experience.

[0081] But one difficulty with volumetric video is that it can be hard to effectively capture high-quality audio during this type of filming process. This is because typical audio capture techniques which might employ boom microphones or lavalier microphones worn by the actors might not be feasible because it may not be possible to effectively hide these microphones from the cameras 1080 given that the scene is filmed from many different viewpoints. There is thus a need for improved techniques for capturing audio during the filming of volumetric video.

[0082] Figure 10 illustrates an example system 1000 for capturing audio during volumetric video capture. As in Figure 9, the system 1000 is located in an environment 1004, which may typically be a green screen room. The system 1000 also includes a number of video cameras 1080 which are located at different viewpoints around the green screen

room 1004 and are aimed at the center portion 1070 of the room where a scene is to be acted out.

[0083] The system 1000 also includes a number of distributed microphones 1012 which are likewise spread out around the perimeter of the room 1004. The microphones 1012 can be located between the video cameras 1080 (as illustrated), they can be co-located with the video cameras, or they can have any other desired configuration. Figure 10 shows that the microphones 1012 are set up to provide full 360° coverage of the central portion 1070 of the room 1004. For example, the microphones 1012 may be placed at least every 45° around the periphery of the room 1004, or at least every 30°, or at least every 10°, or at least every 5°. Although not illustrated in the two-dimensional drawing of Figure 10, the microphones 1012 can also be set up to provide three-dimensional coverage. For example, the microphones 1012 could be placed at several discrete locations about an imaginary hemisphere which encloses the space where the scene is acted out. The operation of the system 1000 will now be described with respect to Figure 11.

[0084] Figure 11 is a flow chart which shows an example method 1100 for using the system 1000 shown in Figure 10 to capture audio for a volumetric video. At block 1110a, a scene is acted out in the green screen room 1004 and the volumetric video is captured by the cameras 1080 from multiple different viewpoints. Simultaneously, the microphones 1012 likewise capture audio of the scene from a variety of vantage points. The recorded audio signals from each of these microphones 1012 can be provided to a processor 1030 along with the video signals from each of the video cameras 1080, as shown at block 1120.

[0085] Each of the audio signals from the respective microphones 1012 can be tagged with location information which indicates the position of the microphone 1012 within the green screen room 1004. At block 1110b, this position information can be determined manually or automatically using location tracking units of the sort described herein. For example, each microphone 1012 can be provided in a monitoring device along with a location tracking unit that can provide data to the processor 1030 regarding the position of the microphone 1012 within the room 1004.

[0086] At block 1130, the processor performs the processing required to generate the volumetric video. Accordingly, the processor can generate simulated video which

estimates the scene as it would have been filmed by a camera located at any specified viewpoint. At block 1140, the processor analyzes the audio signals from the microphones 1012 to generate a representation of the sound wave field within the environment 1104, as described elsewhere herein. Using the sound wave field, the processor can estimate any audio signal as it would have been captured by a microphone located at any desired point within the environment 1104. This capability allows the flexibility to effectively and virtually specify microphone placement for the volumetric video after it has already been filmed.

[0087] In some embodiments, the sound wave field can be mapped to a VR/AR/MR environment and can be used to provide audio for a VR/AR/MR system 80. Just as the viewpoint for the volumetric video can be altered based upon the current viewpoint of a user within a virtual environment, so too can the audio. In some embodiments, the audio listening point can be moved in conjunction with the video viewpoint as the user moves about within the virtual space. In this way, the user can experience a very realistic reproduction of the scene.

Example Embodiments

[0088] A system comprising: a plurality of distributed monitoring devices, each monitoring device comprising at least one microphone and a location tracking unit, wherein the monitoring devices are configured to capture a plurality of audio signals from a sound source and to capture a plurality of location tracking signals which respectively indicate the locations of the monitoring devices over time during capture of the plurality of audio signals; and a processor configured to receive the plurality of audio signals and the plurality of location tracking signals, the processor being further configured to generate a representation of at least a portion of a sound wave field created by the sound source based on the audio signals and the location tracking signals.

[0089] The system of the preceding embodiment, wherein there is an unknown relative spatial relationship between the plurality of distributed monitoring devices.

[0090] The system of any of the preceding embodiments, wherein the plurality of distributed monitoring devices are mobile.

[0091] The system of any of the preceding embodiments, wherein the location tracking unit comprises a Global Positioning System (GPS).

[0092] The system of any of the preceding embodiments, wherein the representation of the sound wave field comprises sound values at each of a plurality of spatial points on a grid for a plurality of times.

[0093] The system of any of the preceding embodiments, wherein the processor is further configured to determine the location of the sound source.

[0094] The system of any of the preceding embodiments, wherein the processor is further configured to map the sound wave field to a virtual, augmented, or mixed reality environment.

[0095] The system of any of the preceding embodiments, wherein, using the representation of the sound wave field, the processor is further configured to determine a virtual audio signal at a selected location within the sound wave field, the virtual audio signal estimating an audio signal which would have been detected by a microphone at the selected location.

[0096] The system of any of the preceding embodiments, wherein the location is selected based on the location of a user of a virtual, augmented, or mixed reality system within a virtual or augmented reality environment.

[0097] A device comprising: a processor configured to carry out a method comprising receiving, from a plurality of distributed monitoring devices, a plurality of audio signals captured from a sound source; receiving, from the plurality of monitoring devices, a plurality of location tracking signals, the plurality of location tracking signals respectively indicating the locations of the monitoring devices over time during capture of the plurality of audio signals; generating a representation of at least a portion of a sound wave field created by the sound source based on the audio signals and the location tracking signals; and a memory to store the audio signals and the location tracking signals.

[0098] The device of the preceding embodiment, wherein there is an unknown relative spatial relationship between the plurality of distributed monitoring devices.

[0099] The device of any of the preceding embodiments, wherein the plurality of distributed monitoring devices are mobile.

[0100] The device of any of the preceding embodiments, wherein the representation of the sound wave field comprises sound values at each of a plurality of spatial points on a grid for a plurality of times.

[0101] The device of any of the preceding embodiments, wherein the processor is further configured to determine the location of the sound source.

[0102] The device of any of the preceding embodiments, wherein the processor is further configured to map the sound wave field to a virtual, augmented, or mixed reality environment.

[0103] The device of any of the preceding embodiments, wherein, using the representation of the sound wave field, the processor is further configured to determine a virtual audio signal at a selected location within the sound wave field, the virtual audio signal estimating an audio signal which would have been detected by a microphone at the selected location.

[0104] The device of any of the preceding embodiments, wherein the location is selected based on the location of a user of a virtual, augmented, or mixed reality system within a virtual or augmented reality environment.

[0105] A method comprising: receiving, from a plurality of distributed monitoring devices, a plurality of audio signals captured from a sound source; receiving, from the plurality of monitoring devices, a plurality of location tracking signals, the plurality of location tracking signals respectively indicating the locations of the monitoring devices over time during capture of the plurality of audio signals; generating a representation of at least a portion of a sound wave field created by the sound source based on the audio signals and the location tracking signals.

[0106] The method of the preceding embodiment, wherein there is an unknown relative spatial relationship between the plurality of distributed monitoring devices.

[0107] The method of any of the preceding embodiments, wherein the plurality of distributed monitoring devices are mobile.

[0108] The method of any of the preceding embodiments, wherein the representation of the sound wave field comprises sound values at each of a plurality of spatial points on a grid for a plurality of times.

[0109] The method of any of the preceding embodiments, further comprising determining the location of the sound source.

[0110] The method of any of the preceding embodiments, further comprising mapping the sound wave field to a virtual, augmented, or mixed reality environment.

[0111] The method of any of the preceding embodiments, further comprising, using the representation of the sound wave field, determining a virtual audio signal at a selected location within the sound wave field, the virtual audio signal estimating an audio signal which would have been detected by a microphone at the selected location.

[0112] The method of any of the preceding embodiments, wherein the location is selected based on the location of a user of a virtual, augmented, or mixed reality system within a virtual or augmented reality environment.

[0113] A system comprising: a plurality of distributed monitoring devices, each monitoring device comprising at least one microphone and a location tracking unit, wherein the monitoring devices are configured to capture a plurality of audio signals in an environment and to capture a plurality of location tracking signals which respectively indicate the locations of the monitoring devices over time during capture of the plurality of audio signals; and a processor configured to receive the plurality of audio signals and the plurality of location tracking signals, the processor being further configured to determine one or more acoustic properties of the environment based on the audio signals and the location tracking signals.

[0114] The system of the preceding embodiment, wherein the one or more acoustic properties comprise acoustic reflectance or absorption in the environment, or the acoustic frequency response of the environment.

[0115] The system of any of the preceding embodiments, wherein there is an unknown relative spatial relationship between the plurality of distributed monitoring devices.

[0116] The system of any of the preceding embodiments, wherein the plurality of distributed monitoring devices are mobile.

[0117] The system of any of the preceding embodiments, wherein the location tracking unit comprises a Global Positioning System (GPS).

[0118] The system of any of the preceding embodiments, wherein the location tracking signals also comprise information about the respective orientations of the monitoring devices.

[0119] The system of any of the preceding embodiments, wherein the plurality of distributed monitoring devices comprise virtual reality, augmented reality, or mixed reality systems.

[0120] The system of any of the preceding embodiments, wherein the processor is further configured to identify a known source sound within the plurality of audio signals.

[0121] The system of any of the preceding embodiments, wherein the known source sound comprises a sound played by one of the virtual reality, augmented reality, or mixed reality systems.

[0122] The system of any of the preceding embodiments, wherein the known source sound comprises an acoustic impulse or a sweep of acoustic tones.

[0123] The system of any of the preceding embodiments, wherein the known source sound comprises an utterance of a user captured by a virtual reality, augmented reality, or mixed reality system worn by the user.

[0124] The system of any of the preceding embodiments, wherein the processor is further configured to identify and associate one or more environment-altered sounds with the known source sound.

[0125] The system of any of the preceding embodiments, wherein the processor is further configured to send the one or more acoustic properties of the environment to the plurality of virtual reality, augmented reality, or mixed reality systems.

[0126] The system of any of the preceding embodiments, wherein the plurality of virtual reality, augmented reality, or mixed reality systems are configured to use the one or more acoustic properties to enhance audio played to a user during a virtual reality, augmented reality, or mixed reality experience.

[0127] A device comprising: a processor configured to carry out a method comprising receiving, from a plurality of distributed monitoring devices, a plurality of audio signals captured in an environment; receiving, from the plurality of monitoring devices, a plurality of location tracking signals, the plurality of location tracking signals respectively indicating the locations of the monitoring devices over time during capture of the plurality of audio signals; determining one or more acoustic properties of the environment based on the audio signals and the location tracking signals; and a memory to store the audio signals and the location tracking signals.

[0128] The device of the preceding embodiment, wherein the one or more acoustic properties comprise acoustic reflectance or absorption in the environment, or the acoustic frequency response of the environment.

[0129] The device of any of the preceding embodiments, wherein the location tracking signals also comprise information about the respective orientations of the monitoring devices.

[0130] The device of any of the preceding embodiments, wherein the plurality of distributed monitoring devices comprise virtual reality, augmented reality, or mixed reality systems.

[0131] The device of any of the preceding embodiments, wherein the processor is further configured to identify a known source sound within the plurality of audio signals.

[0132] The device of any of the preceding embodiments, wherein the known source sound comprises a sound played by one of the virtual reality, augmented reality, or mixed reality systems.

[0133] The device of any of the preceding embodiments, wherein the known source sound comprises an acoustic impulse or a sweep of acoustic tones.

[0134] The device of any of the preceding embodiments, wherein the known source sound comprises an utterance of a user captured by a virtual reality, augmented reality, or mixed reality system worn by the user.

[0135] The device of any of the preceding embodiments, wherein the processor is further configured to identify and associate one or more environment-altered sounds with the known source sound.

[0136] The device of any of the preceding embodiments, wherein the processor is further configured to send the one or more acoustic properties of the environment to the plurality of virtual reality, augmented reality, or mixed reality systems.

[0137] A method comprising: receiving, from a plurality of distributed monitoring devices, a plurality of audio signals captured in an environment; receiving, from the plurality of monitoring devices, a plurality of location tracking signals, the plurality of location tracking signals respectively indicating the locations of the monitoring devices over time during capture of the plurality of audio signals; and determining one or more acoustic properties of the environment based on the audio signals and the location tracking signals.

[0138] The method of the preceding embodiment, wherein the one or more acoustic properties comprise acoustic reflectance or absorption in the environment, or the acoustic frequency response of the environment.

[0139] The method of any of the preceding embodiments, wherein the location tracking signals also comprise information about the respective orientations of the monitoring devices.

[0140] The method of any of the preceding embodiments, wherein the plurality of distributed monitoring devices comprise virtual reality, augmented reality, or mixed reality systems.

[0141] The method of any of the preceding embodiments, further comprising identifying a known source sound within the plurality of audio signals.

[0142] The method of any of the preceding embodiments, wherein the known source sound comprises a sound played by one of the virtual reality, augmented reality, or mixed reality systems.

[0143] The method of any of the preceding embodiments, wherein the known source sound comprises an acoustic impulse or a sweep of acoustic tones.

[0144] The method of any of the preceding embodiments, wherein the known source sound comprises an utterance of a user captured by a virtual reality, augmented reality, or mixed reality system worn by the user.

[0145] The method of any of the preceding embodiments, further comprising identifying and associating one or more environment-altered sounds with the known source sound.

[0146] The method of any of the preceding embodiments, further comprising sending the one or more acoustic properties of the environment to the plurality of virtual reality, augmented reality, or mixed reality systems.

[0147] A system comprising: a plurality of distributed video cameras located about the periphery of a space so as to capture a plurality of videos of a central portion of the space from a plurality of different viewpoints; a plurality of distributed microphones located about the periphery of the space so as to capture a plurality of audio signals during the capture of the plurality of videos; and a processor configured to receive the plurality of videos, the plurality of audio signals, and location information about the position of each microphone within the space, the processor being further configured to generate a representation of at least a portion of a sound wave field for the space based on the audio signals and the location information.

[0148] The system of the preceding embodiment, wherein the plurality of microphones are spaced apart to provide 360° of the space.

[0149] The system of any of the preceding embodiments, wherein the representation of the sound wave field comprises sound values at each of a plurality of spatial points on a grid for a plurality of times.

[0150] The system of any of the preceding embodiments, wherein the processor is further configured to map the sound wave field to a virtual, augmented, or mixed reality environment.

[0151] The system of any of the preceding embodiments, wherein, using the representation of the sound wave field, the processor is further configured to determine a virtual audio signal at a selected location within the sound wave field, the virtual audio signal estimating an audio signal which would have been detected by a microphone at the selected location.

[0152] The system of any of the preceding embodiments, wherein the location is selected based on the location of a user of a virtual, augmented, or mixed reality system within a virtual or augmented reality environment.

[0153] A device comprising: a processor configured to carry out a method comprising receiving, from a plurality of distributed video cameras, a plurality of videos of a scene captured from a plurality of viewpoints; receiving, from a plurality of distributed microphones, a plurality of audio signals captured during the capture of the plurality of videos; receiving location information about the positions of the plurality of microphones; and generating a representation of at least a portion of a sound wave field based on the audio signals and the location information; and a memory to store the audio signals and the location tracking signals.

[0154] The system of the preceding embodiment, wherein the plurality of microphones are spaced apart to provide 360° of the space.

[0155] The system of any of the preceding embodiments, wherein the representation of the sound wave field comprises sound values at each of a plurality of spatial points on a grid for a plurality of times.

[0156] The system of any of the preceding embodiments, wherein the processor is further configured to map the sound wave field to a virtual, augmented, or mixed reality environment.

[0157] The system of any of the preceding embodiments, wherein, using the representation of the sound wave field, the processor is further configured to determine a virtual audio signal at a selected location within the sound wave field, the virtual audio signal estimating an audio signal which would have been detected by a microphone at the selected location.

[0158] The system of any of the preceding embodiments, wherein the location is selected based on the location of a user of a virtual, augmented, or mixed reality system within a virtual or augmented reality environment.

[0159] A method comprising: receiving, from a plurality of distributed video cameras, a plurality of videos of a scene captured from a plurality of viewpoints; receiving, from a plurality of distributed microphones, a plurality of audio signals captured during the capture of the plurality of videos; receiving location information about the positions of the plurality of microphones; and generating a representation of at least a portion of a sound wave field based on the audio signals and the location information.

[0160] The method of the preceding embodiment, wherein the plurality of microphones are spaced apart to provide 360° of the space.

[0161] The method of any of the preceding embodiments, wherein the representation of the sound wave field comprises sound values at each of a plurality of spatial points on a grid for a plurality of times.

[0162] The method of any of the preceding embodiments, further comprising mapping the sound wave field to a virtual, augmented, or mixed reality environment.

[0163] The method of any of the preceding embodiments, further comprising, using the representation of the sound wave field, determining a virtual audio signal at a selected location within the sound wave field, the virtual audio signal estimating an audio signal which would have been detected by a microphone at the selected location.

[0164] The method of any of the preceding embodiments, wherein the location is selected based on the location of a user of a virtual, augmented, or mixed reality system within a virtual or augmented reality environment.

Conclusion

[0165] For purposes of summarizing the disclosure, certain aspects, advantages and features of the invention have been described herein. It is to be understood that not necessarily all such advantages may be achieved in accordance with any particular embodiment of the invention. Thus, the invention may be embodied or carried out in a manner that achieves or optimizes one advantage or group of advantages as taught herein without necessarily achieving other advantages as may be taught or suggested herein.

[0166] Embodiments have been described in connection with the accompanying drawings. However, it should be understood that the figures are not drawn to scale. Distances, angles, etc. are merely illustrative and do not necessarily bear an exact relationship to actual dimensions and layout of the devices illustrated. In addition, the foregoing embodiments have been described at a level of detail to allow one of ordinary skill in the art to make and use the devices, systems, methods, etc. described herein. A wide variety of variation is possible. Components, elements, and/or steps may be altered, added, removed, or rearranged.

[0167] The devices and methods described herein can advantageously be at least partially implemented using, for example, computer software, hardware, firmware, or any combination of software, hardware, and firmware. Software modules can comprise computer executable code, stored in a computer's memory, for performing the functions described herein. In some embodiments, computer-executable code is executed by one or more general purpose computers. However, a skilled artisan will appreciate, in light of this disclosure, that any module that can be implemented using software to be executed on a general purpose computer can also be implemented using a different combination of hardware, software, or firmware. For example, such a module can be implemented completely in hardware using a combination of integrated circuits. Alternatively or additionally, such a module can be implemented completely or partially using specialized computers designed to perform the particular functions described herein rather than by general purpose computers. In addition, where methods are described that are, or could be, at least in part carried out by computer software, it should be understood that such methods can be provided on non-transitory computer-readable media (e.g., optical disks such as CDs or DVDs, hard disk drives, flash

WO 2018/106605

PCT/US2017/064540

memories, diskettes, or the like) that, when read by a computer or other processing device, cause it to carry out the method.

[0168] While certain embodiments have been explicitly described, other embodiments will become apparent to those of ordinary skill in the art based on this disclosure.

WHAT IS CLAIMED IS:

1. A system comprising:
 - a plurality of distributed monitoring devices, each monitoring device comprising at least one microphone and a location tracking unit, wherein the monitoring devices are configured to capture a plurality of audio signals from a sound source and to capture a plurality of location tracking signals which respectively indicate the locations of the monitoring devices over time during capture of the plurality of audio signals; and
 - a processor configured to receive the plurality of audio signals and the plurality of location tracking signals, the processor being further configured to generate a representation of at least a portion of a sound wave field created by the sound source based on the audio signals and the location tracking signals.
2. The system of Claim 1, wherein there is an unknown relative spatial relationship between the plurality of distributed monitoring devices.
3. The system of Claim 2, wherein the plurality of distributed monitoring devices are mobile.
4. The system of Claim 1, wherein the location tracking unit comprises a Global Positioning System (GPS).
5. The system of Claim 1, wherein the representation of the sound wave field comprises sound values at each of a plurality of spatial points on a grid for a plurality of times.
6. The system of Claim 1, wherein the processor is further configured to determine the location of the sound source.
7. The system of Claim 1, wherein the processor is further configured to map the sound wave field to a virtual, augmented, or mixed reality environment.
8. The system of Claim 1, wherein, using the representation of the sound wave field, the processor is further configured to determine a virtual audio signal at a selected location within the sound wave field, the virtual audio signal estimating an audio signal which would have been detected by a microphone at the selected location.

9. The system of Claim 8, wherein the location is selected based on the location of a user of a virtual, augmented, or mixed reality system within a virtual or augmented reality environment.

10. A device comprising:

a processor configured to carry out a method comprising

receiving, from a plurality of distributed monitoring devices, a plurality of audio signals captured from a sound source;

receiving, from the plurality of monitoring devices, a plurality of location tracking signals, the plurality of location tracking signals respectively indicating the locations of the monitoring devices over time during capture of the plurality of audio signals;

generating a representation of at least a portion of a sound wave field created by the sound source based on the audio signals and the location tracking signals; and

a memory to store the audio signals and the location tracking signals.

11. The device of Claim 10, wherein there is an unknown relative spatial relationship between the plurality of distributed monitoring devices.

12. The device of Claim 11, wherein the plurality of distributed monitoring devices are mobile.

13. The device of Claim 10, wherein the representation of the sound wave field comprises sound values at each of a plurality of spatial points on a grid for a plurality of times.

14. The device of Claim 10, wherein the processor is further configured to determine the location of the sound source.

15. The device of Claim 10, wherein the processor is further configured to map the sound wave field to a virtual, augmented, or mixed reality environment.

16. The device of Claim 10, wherein, using the representation of the sound wave field, the processor is further configured to determine a virtual audio signal at a selected location within the sound wave field, the virtual audio signal estimating an audio signal which would have been detected by a microphone at the selected location.

17. The device of Claim 16, wherein the location is selected based on the location of a user of a virtual, augmented, or mixed reality system within a virtual or augmented reality environment.

18. A method comprising:

receiving, from a plurality of distributed monitoring devices, a plurality of audio signals captured from a sound source;

receiving, from the plurality of monitoring devices, a plurality of location tracking signals, the plurality of location tracking signals respectively indicating the locations of the monitoring devices over time during capture of the plurality of audio signals;

generating a representation of at least a portion of a sound wave field created by the sound source based on the audio signals and the location tracking signals.

19. The method of Claim 18, wherein there is an unknown relative spatial relationship between the plurality of distributed monitoring devices.

20. The method of Claim 19, wherein the plurality of distributed monitoring devices are mobile.

21. The method of Claim 18, wherein the representation of the sound wave field comprises sound values at each of a plurality of spatial points on a grid for a plurality of times.

22. The method of Claim 18, further comprising determining the location of the sound source.

23. The method of Claim 18, further comprising mapping the sound wave field to a virtual, augmented, or mixed reality environment.

24. The method of Claim 18, further comprising, using the representation of the sound wave field, determining a virtual audio signal at a selected location within the sound wave field, the virtual audio signal estimating an audio signal which would have been detected by a microphone at the selected location.

25. The method of Claim 24, wherein the location is selected based on the location of a user of a virtual, augmented, or mixed reality system within a virtual or augmented reality environment.

ABSTRACT

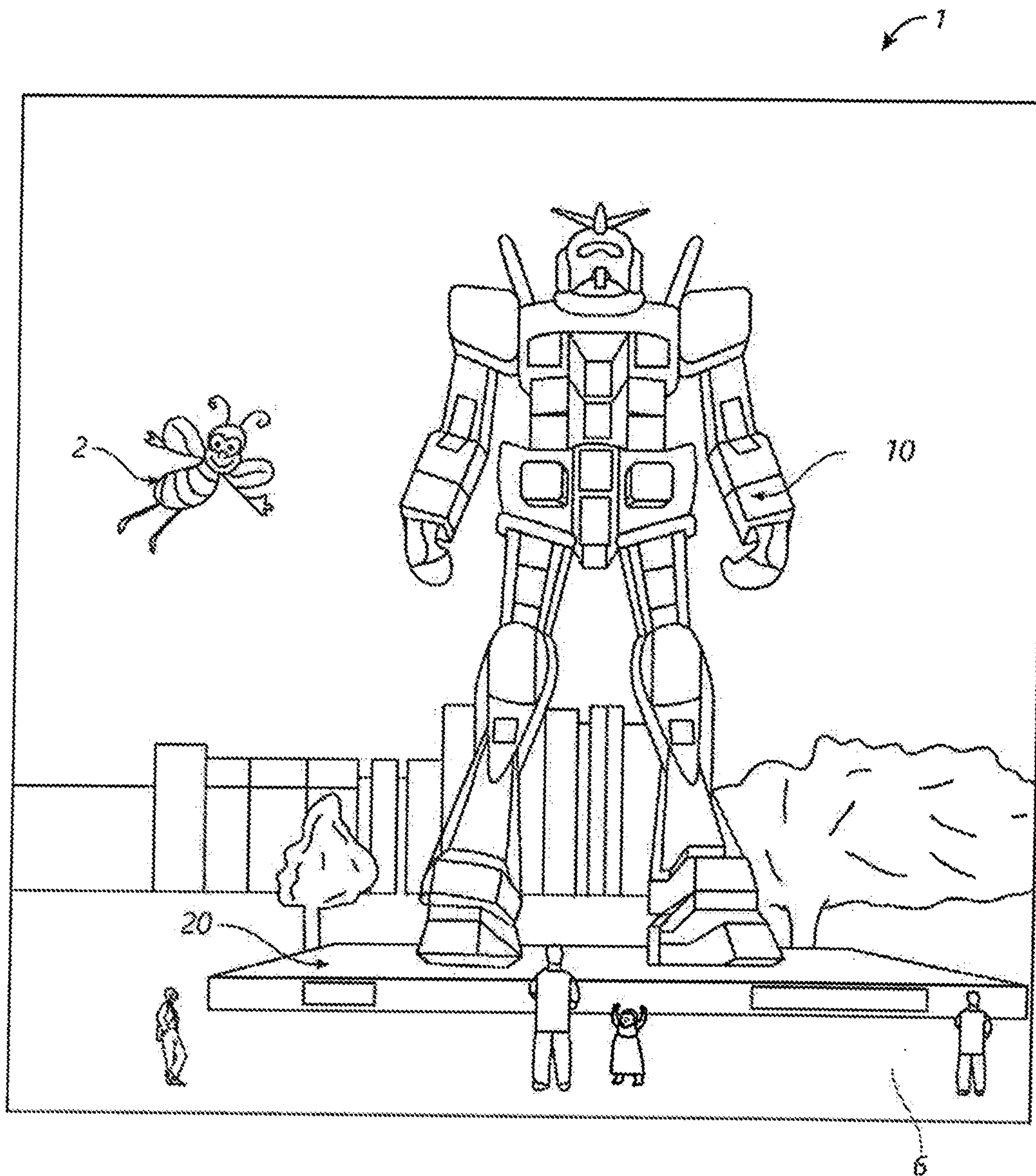


FIG. 1

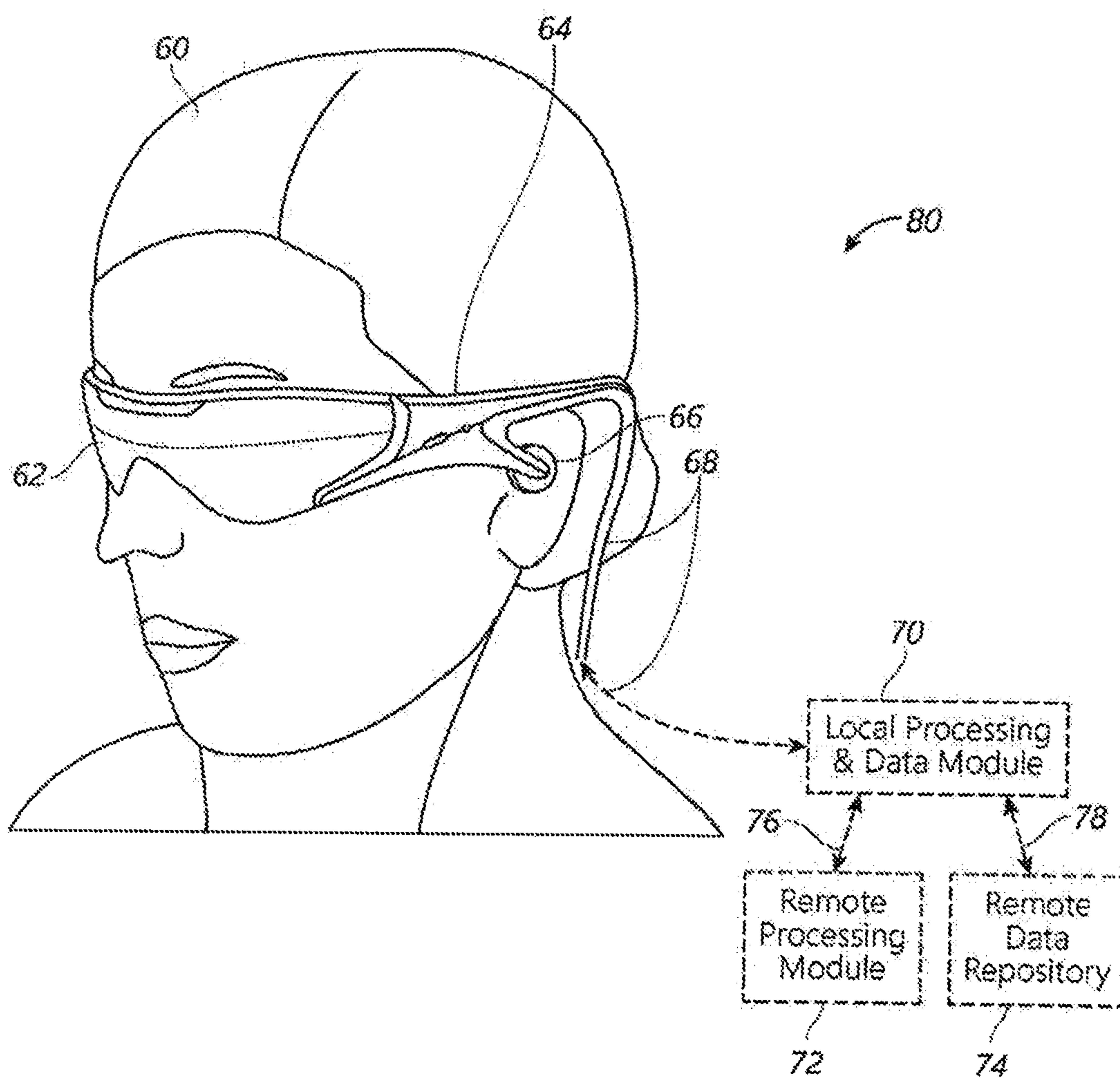


FIG. 2

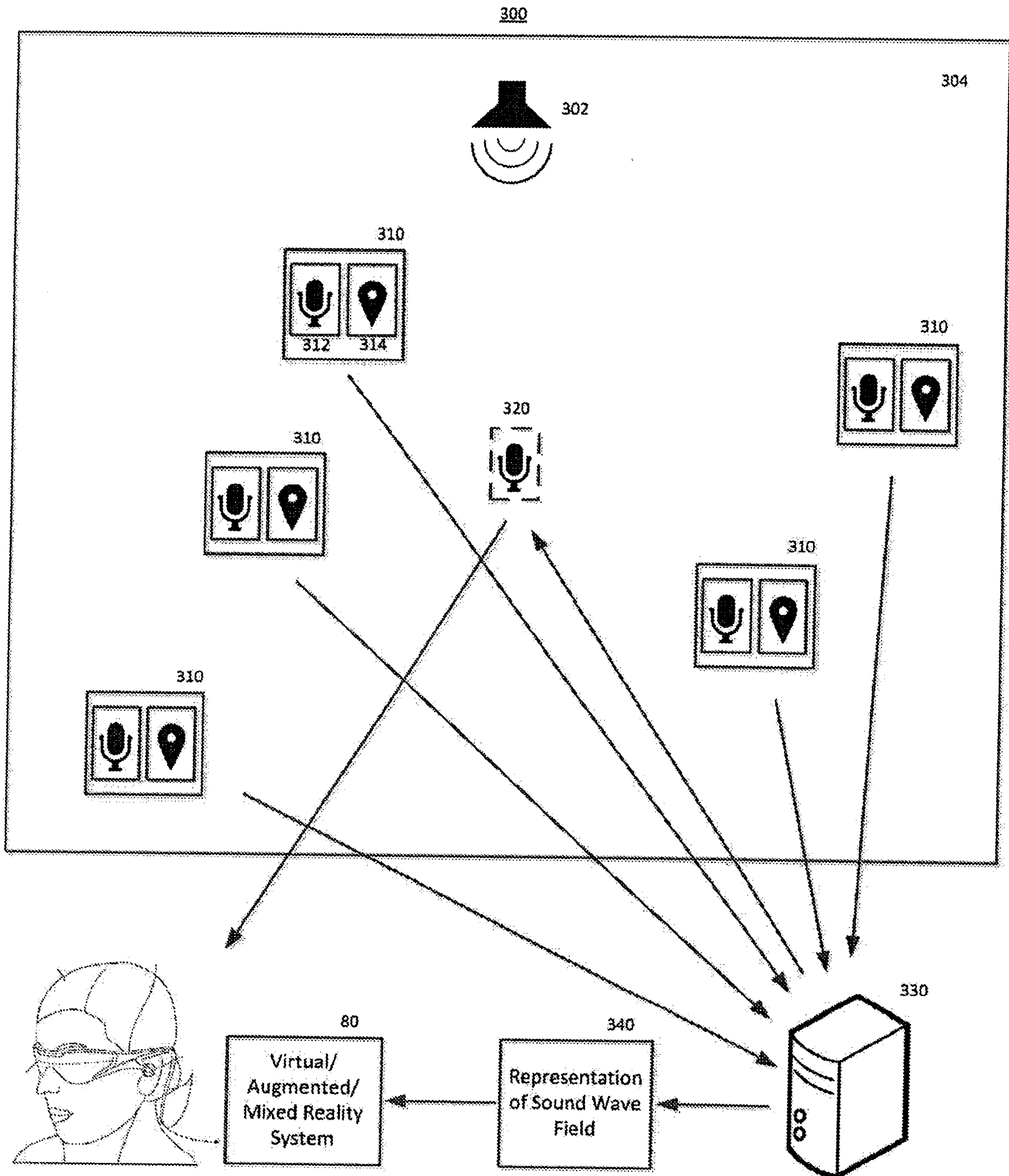


FIG. 3

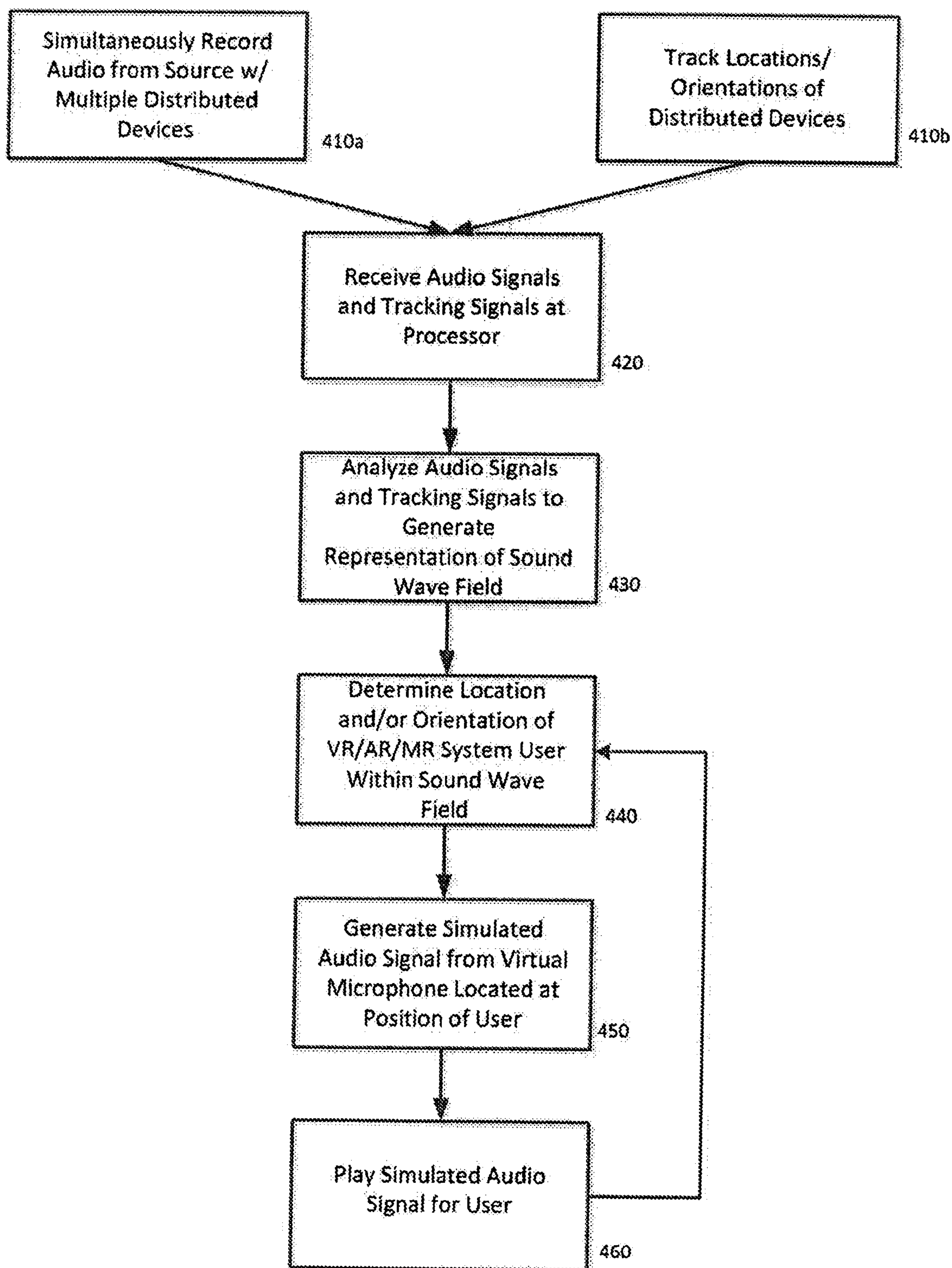
4/11
400

FIG. 4

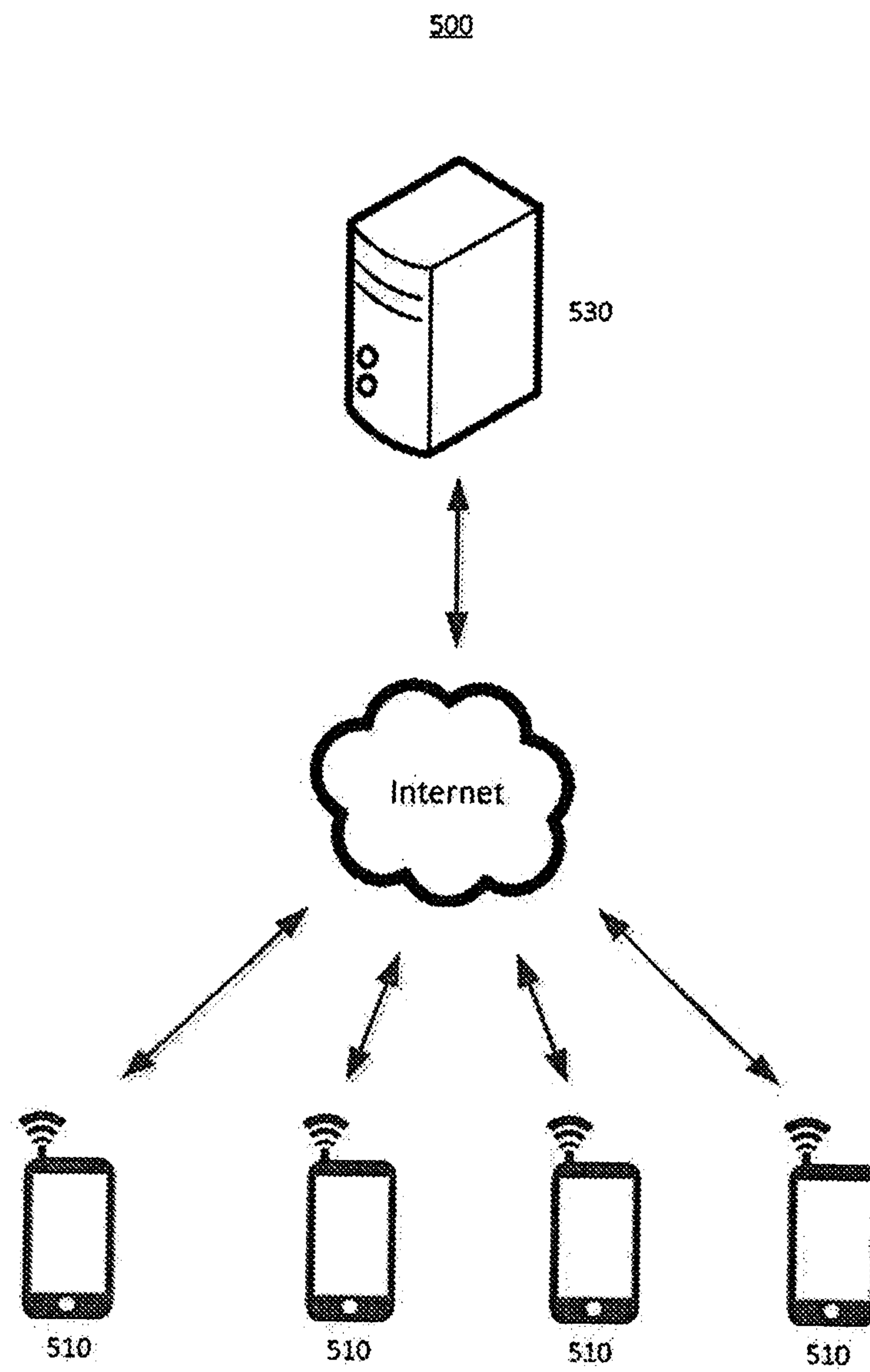


FIG. 5

WO 2018/106605

PCT/US2017/064540

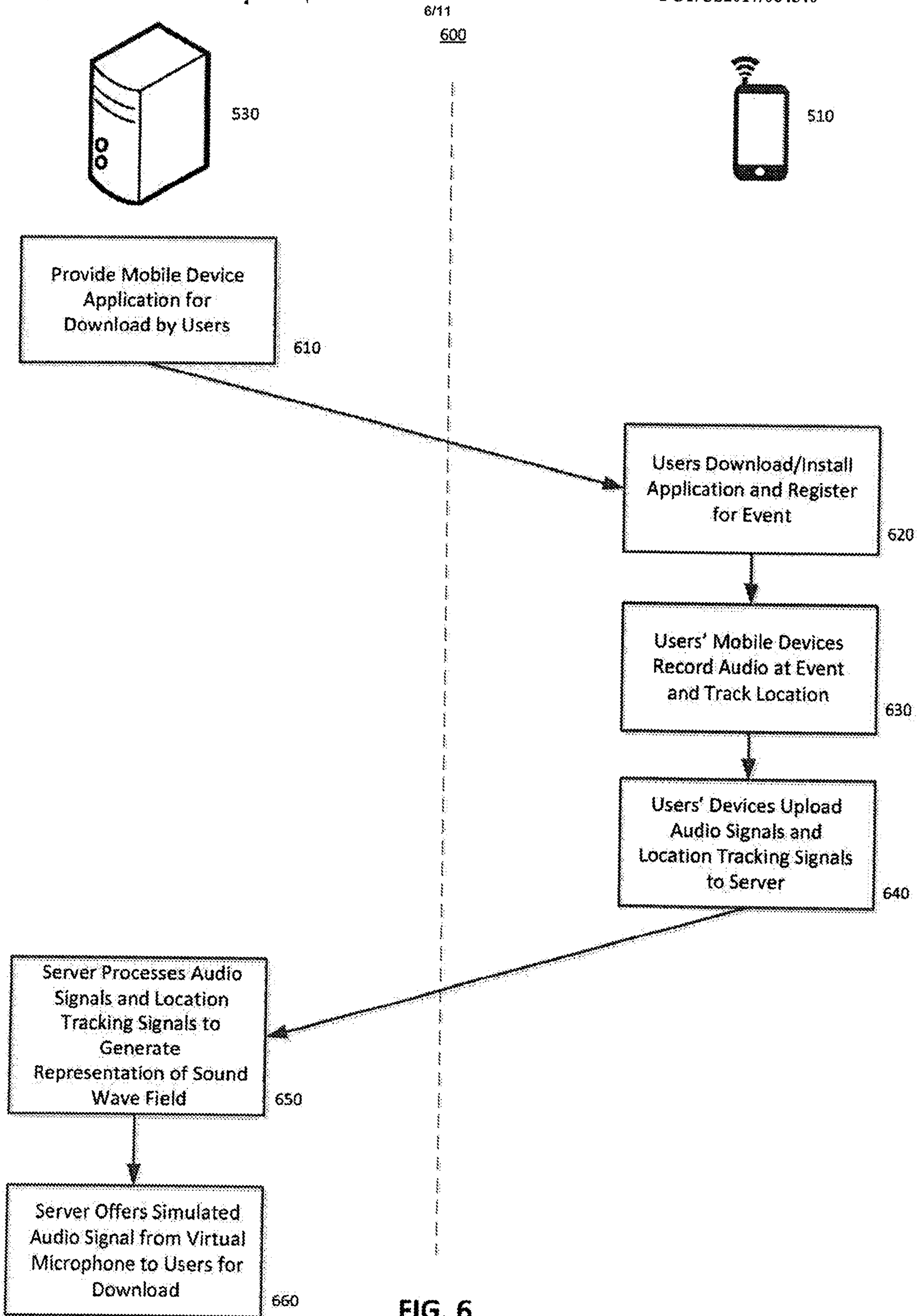


FIG. 6

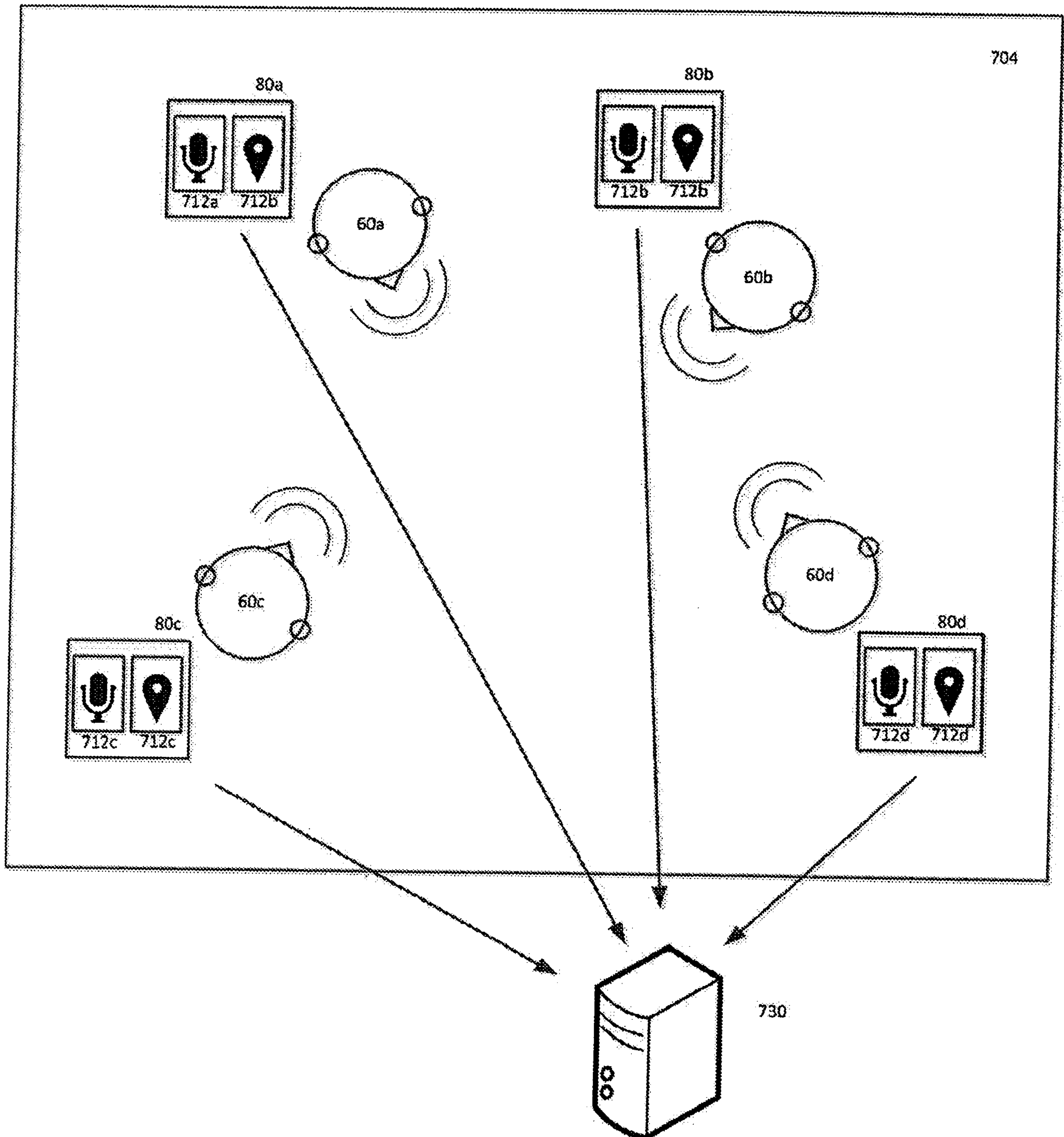


FIG. 7

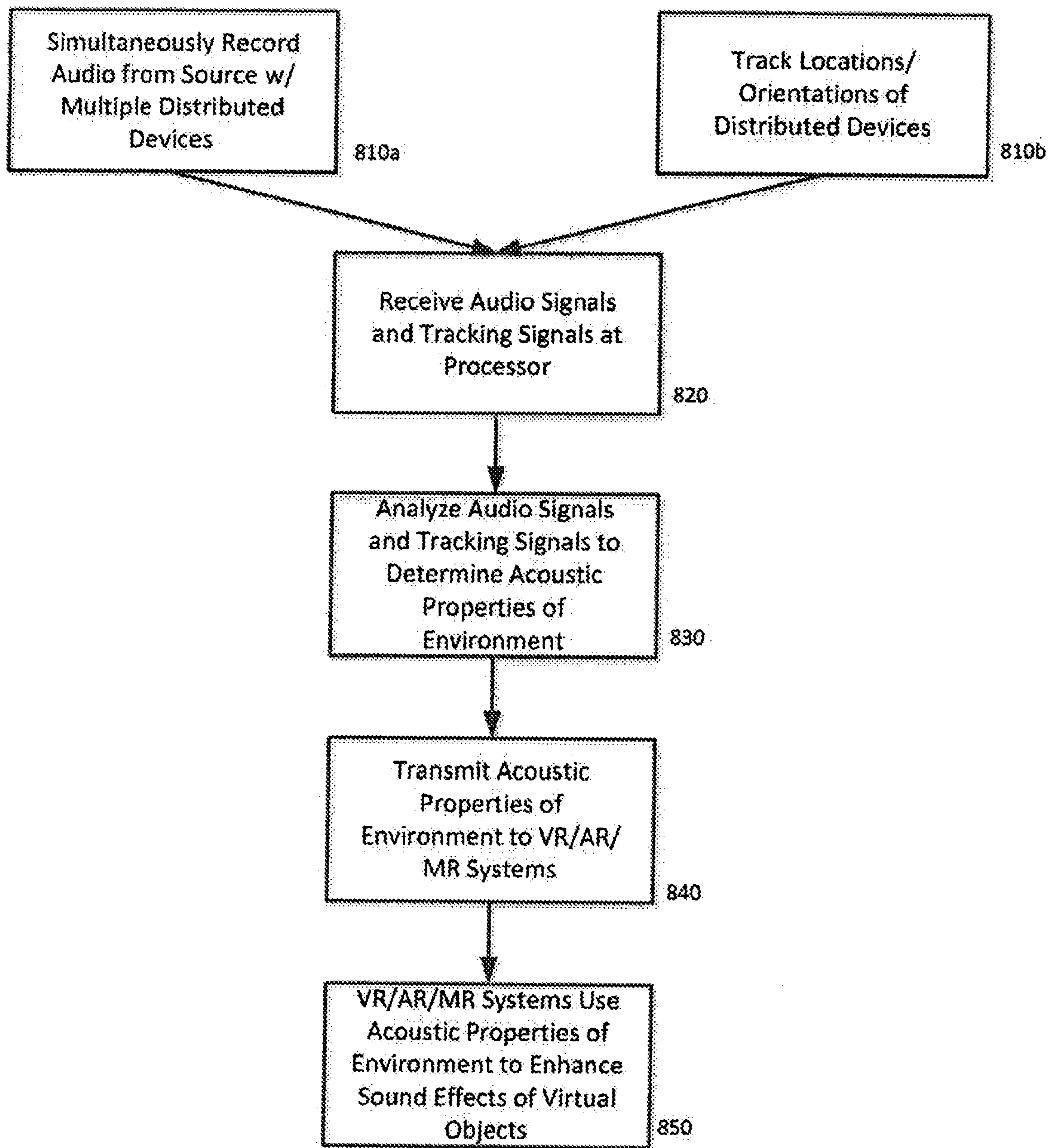
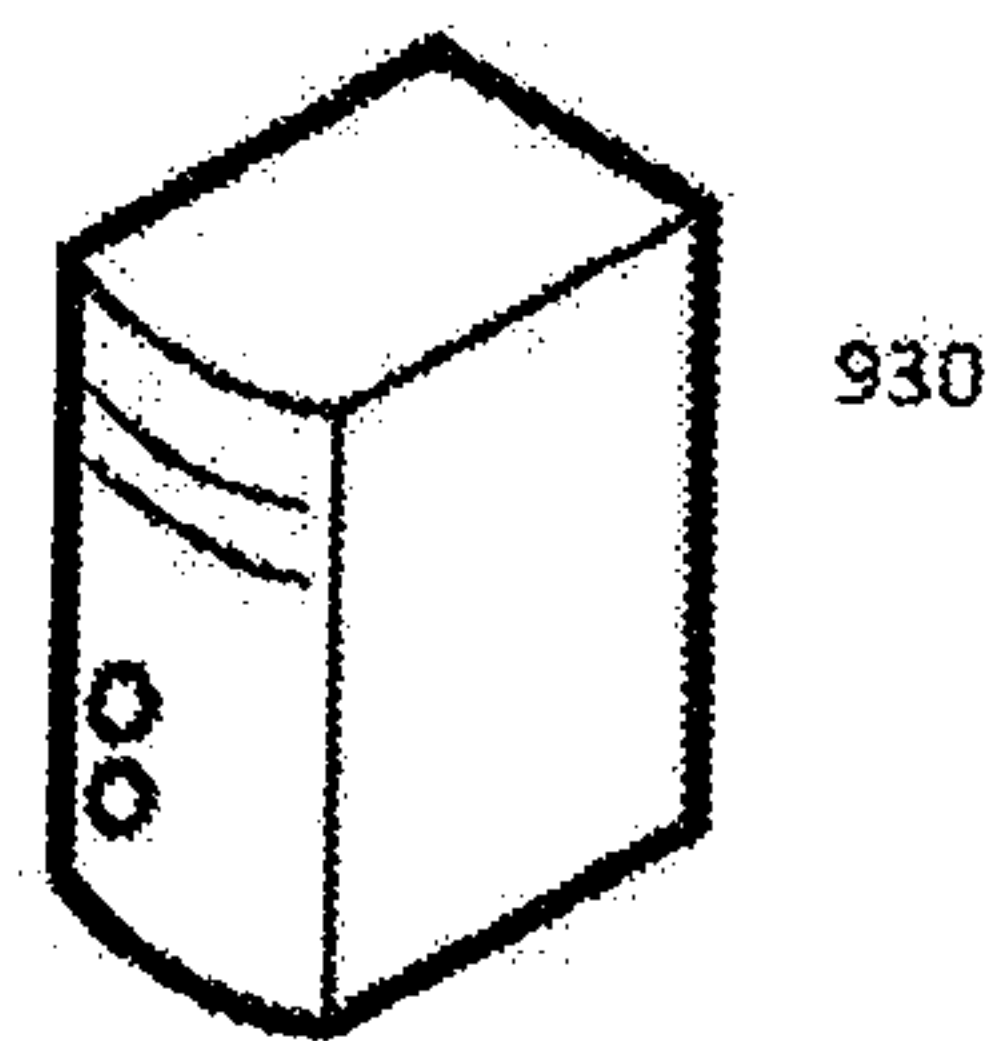
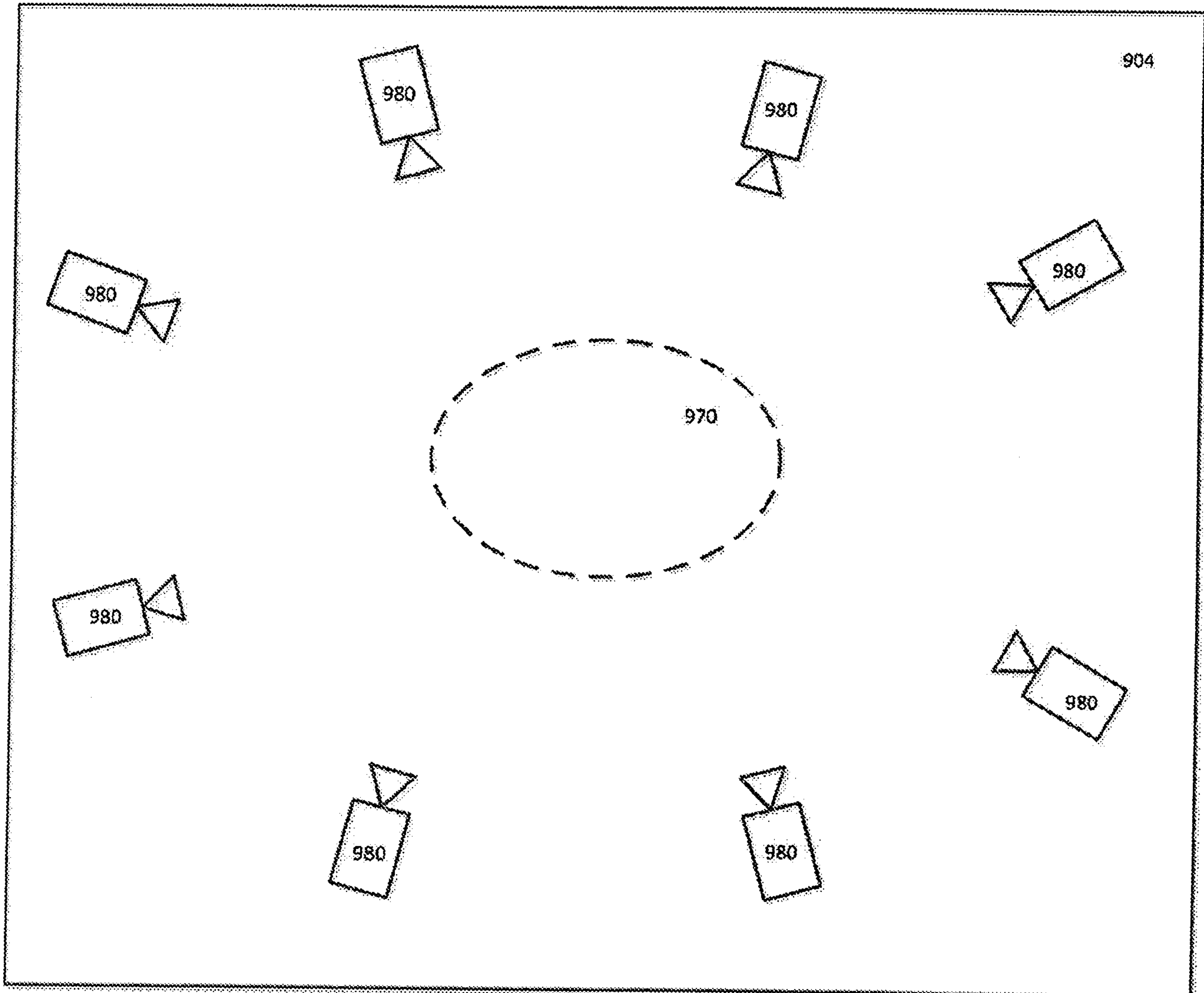
8/11
800

FIG. 8

900**FIG. 9**

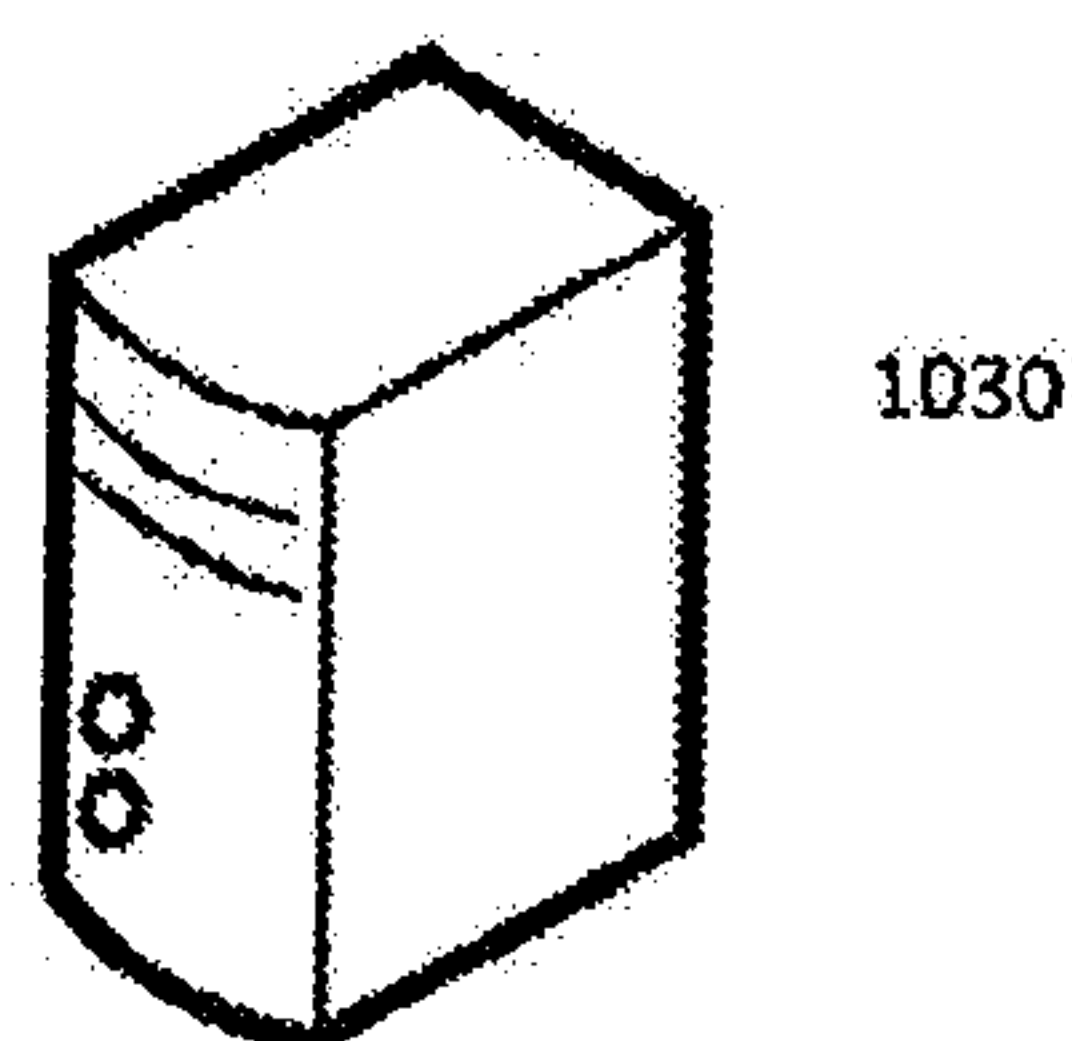
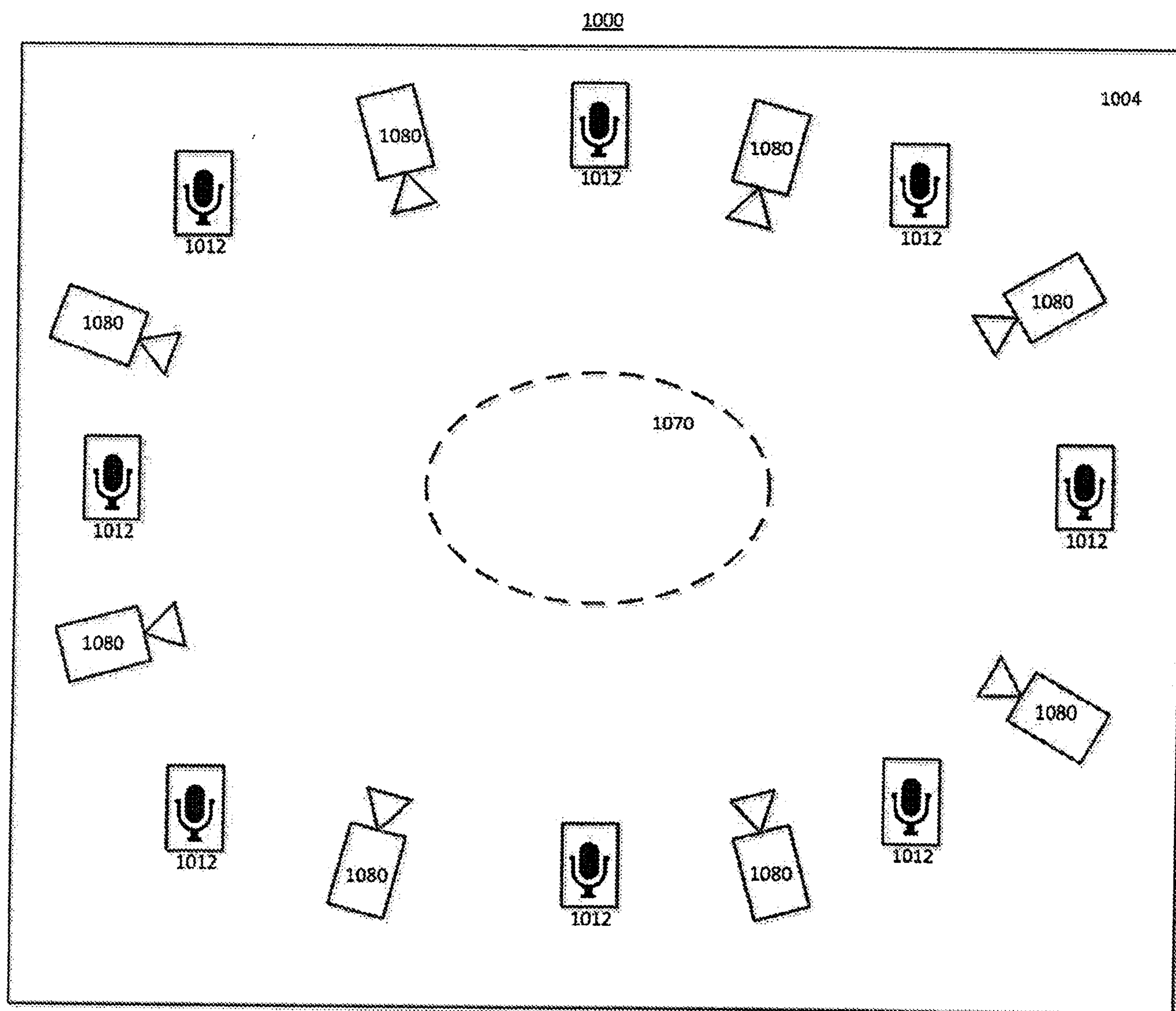
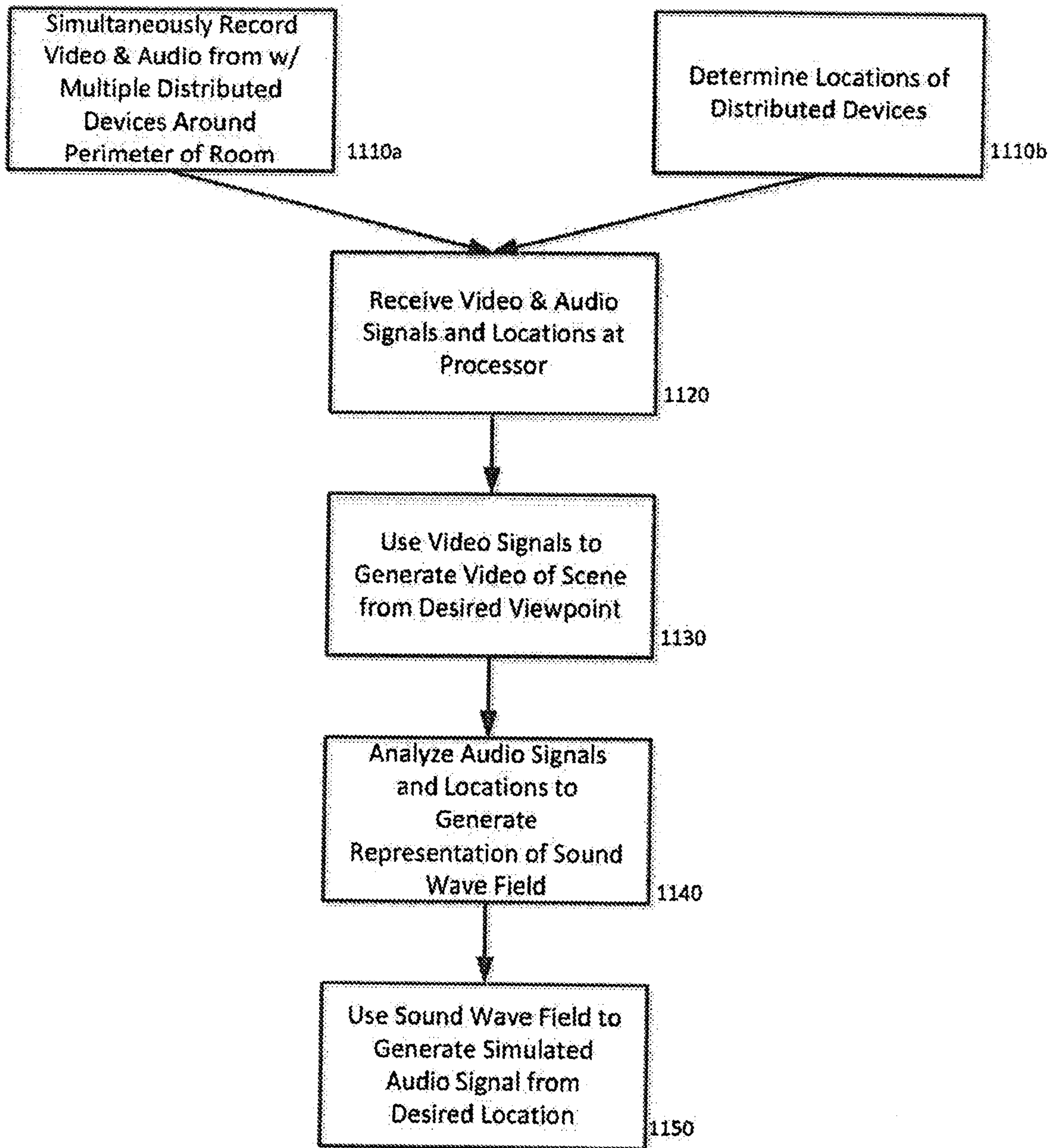


FIG. 10

11/11
1100**FIG. 11**

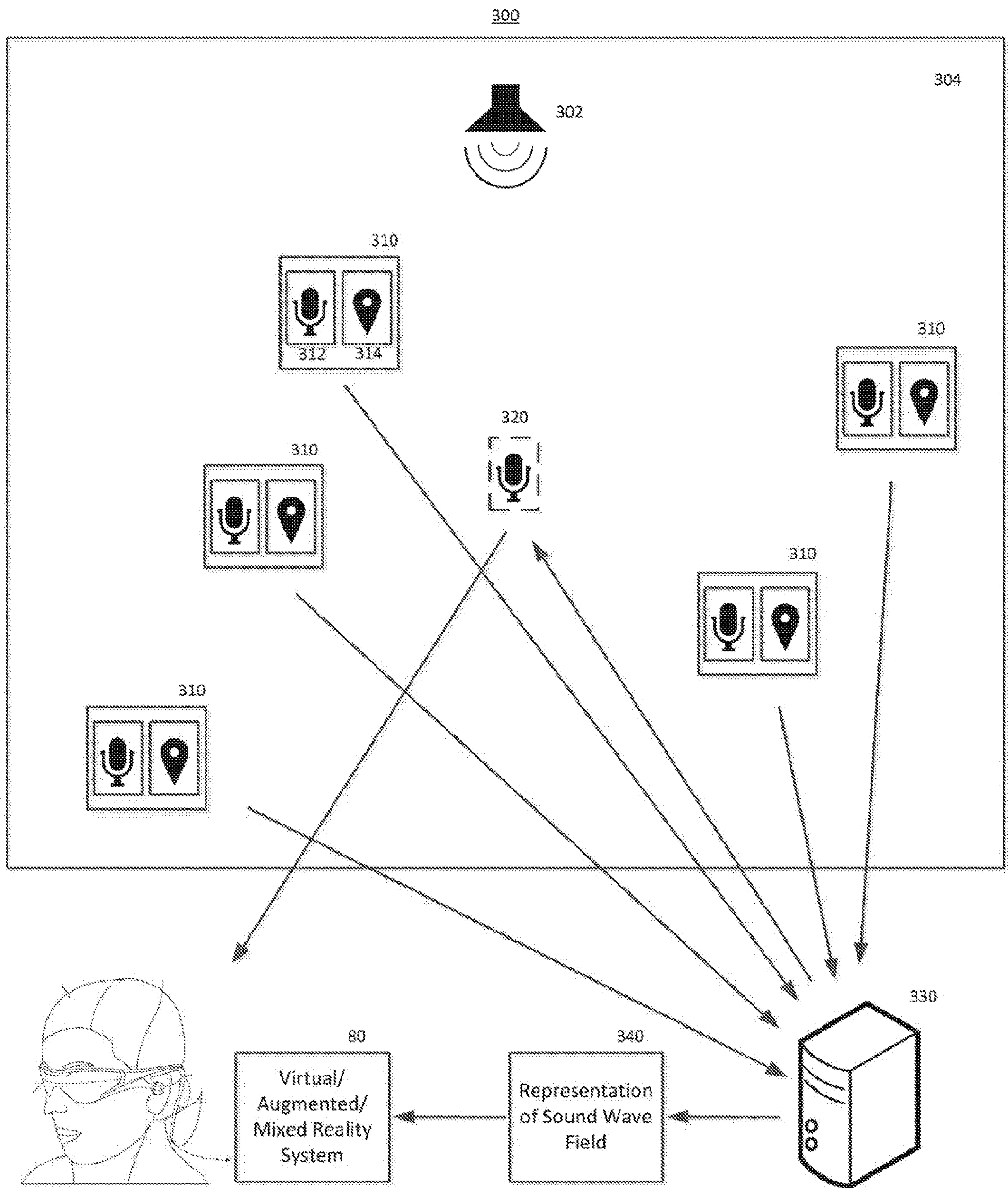


FIG. 3