

(51) International Patent Classification:
H04N 7/52 (2011.01)(21) International Application Number:
PCT/CN2012/086260(22) International Filing Date:
10 December 2012 (10.12.2012)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant (for all designated States except US): **INTEL CORPORATION** [US/US]; 2200 Mission College Blvd., Santa Clara, California 95054 (US).

(72) Inventors; and

(71) Applicants (for US only): **LI, Qiang** [CN/CN]; 8F, Raycom Infotech Park A No. 2, Kexueyuan South Road, Haidian District, Beijing 100080 (CN). **DU, Yangzhou** [CN/CN]; Room 502, Building 2-21-7, XinGangZhuangYuan, LiJiaQiao, Shunyi district, Beijing 101304 (CN). **LI, Wenlong** [CN/CN]; 4-502, Building 18, Tiantongyuan West 2 District, Changping district, Beijing 102218 (CN). **TONG, Xiaofeng** [CN/CN]; Room 1103, Building 1, Shuangyushu Xili, Haidian district, Beijing 100086 (CN). **HU, Wei** [CN/CN]; 8F, Raycom Infotech Park A, No. 2., Kexueyuan South Road, Haidian district, Beijing 100190 (CN). **XU, Lin** [CN/CN]; Floor 8, Raycom A, No. 2, Kexueyuan South Road, Beijing 100190 (CN). **ZHANG, Yimin** [CN/CN]; 11-1507 Taiyangyuan, Haidian district, Beijing 100098 (CN).(74) Agent: **CHINA PATENT AGENT (HK) LTD**; 22/F., Great Eagle Center, 23 Harbour Road, Wanchai, Hong Kong (CN).

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

Published:

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

(54) Title: METHOD, APPARATUS AND SYSTEM OF VIDEO AND AUDIO SHARING AMONG COMMUNICATION DEVICES

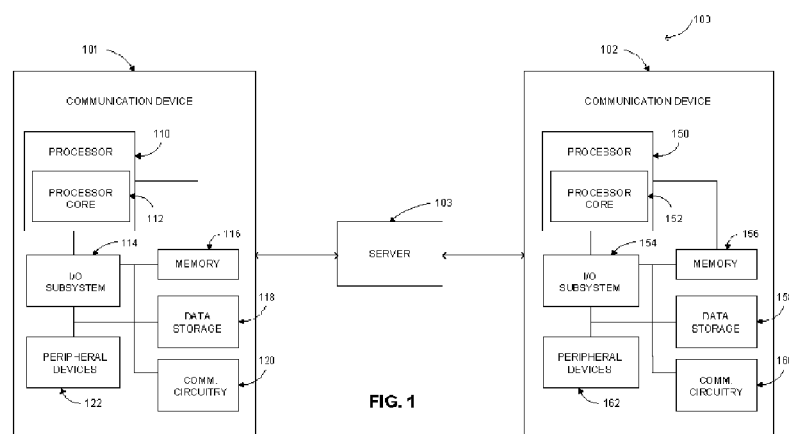


FIG. 1

(57) Abstract: A device, method and system of video and audio sharing among communication devices, may comprise a communication device for generating and sending a packet containing information related to the video and audio, and another communication device for receiving the packet and rendering the information related to the audio and video. In some embodiments, the communication device may comprise: an audio encoding module to encode a piece of audio into an audio bit stream; an avatar data extraction module to extract avatar data from a piece of video and generate an avatar data bit stream; and a synchronization module to generate synchronization information for synchronizing the audio bit stream with the avatar parameter stream. In some embodiments, the another communication device may comprise: an audio decoding module to decode an audio bit stream into decoded audio data; an Avatar animation module to animate an Avatar model based on an Avatar data bit stream to generate an animated Avatar model; and a synchronizing and rendering module to synchronize and render the decoded audio data and the animated Avatar model by utilizing the synchronization information.

METHOD, APPARATUS AND SYSTEM OF VIDEO AND AUDIO SHARING AMONG COMMUNICATION DEVICES

TECHNICAL FIELD

[0001] An embodiment of the present patent application relates to video and audio sharing, more specifically, to video and audio sharing using Avatar.

BACKGROUND

[0002] Wireless communication technologies provide people with more and more content-enriched information sharing on wireless networks. For example, people can communicate and share video and audio on-line (e.g., video conference), or off-line (e.g., instant message). However, high bandwidth consumption may significantly hinder a widespread use of the off-line video and audio sharing.

BRIEF DESCRIPTION OF THE DRAWINGS

[0003] The embodiment described herein is illustrated by way of example and not by way of limitation in the accompanying figures. For simplicity and clarity of illustration, elements illustrated in the figures are not necessarily drawn to scale. For example, the dimensions of some elements may be exaggerated relative to other elements for clarity. Further, where considered appropriate, reference labels have been repeated among the figures to indicate corresponding or analogous elements.

[0004] Fig.1 illustrates an embodiment of a system of sharing video and audio among two communication devices.

[0005] Fig. 2 illustrates an embodiment of an environment of one communication device for generating a packet containing the video and audio information to be shared with another communication device.

- [0006] Fig. 3 illustrates an embodiment of an environment of another communication device for receiving the packet and rendering the video and audio information.
- [0007] Fig. 4 illustrates an embodiment of a method of the one communication device for generating the packet containing the video and audio information to be shared with another communication device.
- [0008] Fig. 5 illustrates an embodiment of a method of another communication device for receiving the packet and rendering the video and audio information.

DETAILED DESCRIPTION

- [0009] While the concepts of the embodiment are susceptible to various modifications and alternative forms, specific exemplary embodiments thereof have been shown by way of example in the drawings and will herein be described in detail. It should be understood, however, that there is no intent to limit the concepts of the present disclosure to the particular forms disclosed, but on the contrary, the intention is to cover all modifications, equivalents, and alternatives consistent with the present disclosure and the appended claims.
- [0010] In the following description, numerous specific details such as logic implementations, opcodes, means to specify operands, resource partitioning/sharing/duplication implementations, types and interrelationships of system components, and logic partitioning/integration choices are set forth in order to provide a more thorough understanding of the present disclosure. It will be appreciated, however, by one skilled in the art that embodiments of the disclosure may be practiced without such specific details. In other instances, control structures, gate level circuits and full software instruction sequences have not been shown in detail in order not to obscure the embodiment. Those of

ordinary skill in the art, with the included descriptions, will be able to implement appropriate functionality without undue experimentation.

[0011] References in the specification to “one embodiment,” “an embodiment,” “an example embodiment,” etc., indicate that the embodiment described may include a particular feature, structure, or characteristic, but every embodiment may not necessarily include the particular feature, structure, or characteristic. Moreover, such phrases are not necessarily referring to the same embodiment. Further, when a particular feature, structure, or characteristic is described in connection with an embodiment, it is submitted that it is within the knowledge of one skilled in the art to effect such feature, structure, or characteristic in connection with other embodiments whether or not explicitly described.

[0012] Embodiments may be implemented in hardware, firmware, software, or any combination thereof. Embodiments implemented in a computer system may include one or more bus-based interconnects between components and/or one or more point-to-point interconnects between components. Embodiments may also be implemented as instructions carried by or stored on a transitory or non-transitory machine-readable (e.g., computer-readable) medium, which may be read and executed by one or more processors. A machine-readable medium may be embodied as any device, mechanism, or physical structure for storing or transmitting information in a form readable by a machine (e.g., a computing device). For example, a machine-readable medium may be embodied as read only memory (ROM); random access memory (RAM); magnetic disk storage media; optical storage media; flash memory devices; mini- or micro-SD cards, memory sticks, electrical signals, and others.

[0013] In the drawings, specific arrangements or orderings of schematic elements, such as those representing devices, modules, instruction blocks, and data elements, may be shown for ease of description. However, it should be understood by those skilled in the art that the

specific ordering or arrangement of the schematic elements in the drawings is not meant to imply that a particular order or sequence of processing, or separation of processes, is required. Further, the inclusion of a schematic element in a drawing is not meant to imply that such element is required in all embodiments or that the features represented by such element may not be included in or combined with other elements in some embodiments.

[0014] In general, schematic elements used to represent instruction blocks may be implemented using any suitable form of machine-readable instruction, such as software or firmware applications, programs, functions, modules, routines, processes, procedures, plug-ins, applets, widgets, code fragments and/or others, and that each such instruction may be implemented using any suitable programming language, library, application programming interface (API), and/or other software development tools. For example, some embodiments may be implemented using Java, C++, and/or other programming languages. Similarly, schematic elements used to represent data or information may be implemented using any suitable electronic arrangement or structure, such as a register, data store, table, record, array, index, hash, map, tree, list, graph, file (of any file type), folder, directory, database, and/or others.

[0015] Further, in the drawings, where connecting elements, such as solid or dashed lines or arrows, are used to illustrate a connection, relationship or association between or among two or more other schematic elements, the absence of any such connecting elements is not meant to imply that no connection, relationship or association can exist. In other words, some connections, relationships or associations between elements may not be shown in the drawings so as not to obscure the disclosure. In addition, for ease of illustration, a single connecting element may be used to represent multiple connections, relationships or associations between elements. For example, where a connecting element represents a communication of signals, data or instructions, it should be understood by those skilled in

the art that such element may represent one or multiple signal paths (e.g., a bus), as may be needed, to effect the communication.

[0016] Fig. 1 illustrates an embodiment of a system 100 of sharing video and audio between two communication devices 101 and 102 via a server 103. In an embodiment, the communication device 101 may be connected with the server 103 through a network, such as a wireless network or an internet. Similarly, the communication device 102 may also be connected with the server 103 through the network. The communication device 101 may be embodied as, without limitation, a smart phone, a cellular telephone, a tablet, a consumer electronic device, a laptop computer, a notebook computer, a desktop computer, a mobile computing system, a work station, a network appliance, a web appliance, a processor-based system, and/or any other communication device configured to generate a packet containing information related to the video and audio to be shared offline with the communication device 102. The packet may be embodied, without limitation, an instant message, a short message, or other data stream.

[0017] In an embodiment, the communication device 101 may include a processor 110, an I/O subsystem 114, a memory 116, a data storage 118, a communication circuitry 120, and one or more peripheral devices 122. In some embodiments, several of the foregoing components may be incorporated on a motherboard or main board of the communication device 101, while other components may be communicatively coupled to the motherboard via, for example, a peripheral port. Furthermore, it should be appreciated that the communication device 101 may include other components, sub-components, and devices commonly found in a communication and/or computing device, which are not illustrated in FIG. 1 for clarity of the description.

[0018] The processor 110 of the communication device 101 may be embodied as any type of processor capable of executing software/firmware, such as a microprocessor, digital signal

processor, microcontroller, or the like. The processor 110 is illustratively embodied as a single core processor having a processor core 112. However, in other embodiments, the processor 110 may be embodied as a multi-core processor having multiple processor cores 112. Additionally, the communication device 101 may include additional processors 110 having one or more processor cores 112.

[0019] The I/O subsystem 114 of the communication device 101 may be embodied as circuitry and/or components to facilitate input/output operations with the processor 110 and/or other components of the communication device 102. In some embodiments, the I/O subsystem 114 may be embodied as a memory controller hub (MCH or “northbridge”), an input/output controller hub (ICH or “southbridge”) or platform controller hub (PCH), and a firmware device. In such embodiments, the firmware device of the I/O subsystem 114 may be embodied as a memory device for storing Basic Input/Output System (BIOS) data and/or instructions and/or other information (e.g., a BIOS driver used during booting of the communication device 101). However, in other embodiments, I/O subsystems having other configurations may be used. For example, in some embodiments, the I/O subsystem 114 may be embodied as a platform controller hub (PCH). In such embodiments, the memory controller hub (MCH) may be incorporated in or otherwise associated with the processor 110, and the processor 110 may communicate directly with the memory 116 (as shown by the dashed line in FIG. 1). Additionally, in other embodiments, the I/O subsystem 114 may form a portion of a system-on-a-chip (SoC) and be incorporated, along with the processor 110 and other components of the communication device 101, on a single integrated circuit chip.

[0020] The processor 110 is communicatively coupled to the I/O subsystem 114 via a number of signal paths. These signal paths (and other signal paths illustrated in FIG. 1) may be embodied as any type of signal paths capable of facilitating communication between the

components of the communication device 101. For example, the signal paths may be embodied as any number of point-to-point links, wires, cables, light guides, printed circuit board traces, vias, bus, intervening devices, and/or the like.

[0021] The memory 116 of the communication device 101 may be embodied as or otherwise include one or more memory devices or data storage locations including, for example, dynamic random access memory devices (DRAM), synchronous dynamic random access memory devices (SDRAM), double-data rate synchronous dynamic random access memory device (DDR SDRAM), mask read-only memory (ROM) devices, erasable programmable ROM (EPROM), electrically erasable programmable ROM (EEPROM) devices, flash memory devices, and/or other volatile and/or non-volatile memory devices. The memory 116 is communicatively coupled to the I/O subsystem 114 via a number of signal paths. Although only a single memory device 116 is illustrated in FIG. 1, the communication device 101 may include additional memory devices in other embodiments. Various data and software may be stored in the memory 116. For example, one or more operating systems, applications, programs, libraries, and drivers that make up the software stack executed by the processor 110 may reside in memory 116 during execution.

[0022] The data storage 118 may be embodied as any type of device or devices configured for the short-term or long-term storage of data. For example, the data storage 118 may include any one or more memory devices and circuits, memory cards, hard disk drives, solid-state drives, or other data storage devices

[0023] The communication circuitry 120 of the communication device 101 may include any number of devices and circuitry for enabling communications between the communication device 101 and one or more devices or networks as discussed in more detail below. The communication circuitry 120 may be configured to use any one or more, or combination thereof, communication protocols to communicate, such as, for example, a cellular

communication protocol (e.g., Wideband Code Division Multiple Access (W-CDMA)), a wireless network communication protocol (e.g., Wi-Fi®, WiMAX), a wireless personal area network communication protocol (e.g., Bluetooth®), a wired network communication protocol (e.g., TCP/IP), and/or other communication protocols.

[0024] In some embodiments, the communication device 101 may also include one or more peripheral devices 122. Such peripheral devices 122 may include any number of additional input/output devices, interface devices, and/or other peripheral devices. For example, in some embodiments, the peripheral devices 122 may include a display, touch screen, graphics circuitry, keyboard, speaker system, and/or other input/output devices, interface devices, and/or peripheral devices.

[0025] The communication device 102, for receiving the packet and rendering the video and audio, may be substantially similar to the communication device 101 and include similar components, which have been identified in FIG. 1 with a common reference numbering scheme, e.g., a processor 150, a processor core 152, an I/O subsystem 154, a memory 156, a data storage 158, a communication circuitry 160 and one or more peripheral devices 162. As such, the description provided above of the components of the communication device 101 is equally applicable to those similar components of the communication device 102 and is not repeated herein so as not to obscure the present disclosure. Of course, it should be appreciated that in some embodiments the communication devices 101, 102 of the system 100 may be dissimilar to each other. For example, the communication devices 101 and 102 may be embodied as various types of communication devices different from each other (e.g., a smart phone, a tablet, a laptop computer, a notebook computer, or other communication device) and include components typically found in such corresponding communication devices.

- [0026] In order to lower bandwidth consumption while keeping reality of facial expression and/or motion of an object presenting in the video, the communication device 101 may extract Avatar data indicating the object's facial expression and/or motion from the video, and synchronization information may be inserted into Avatar data bit streams and/or audio bit streams before packing the packet. The communication device 101 may transmit the packet to the server 103 regardless if the communication device 102 is connected to the server or not.
- [0027] The communication device 102 may receive the packet from the server 103 regardless if the communication device 101 is connected to the server or not. With the help of the synchronization information, the communication device 102 may render the audio and an Avatar model animated based on the Avatar data to represent the scene where the video and audio are made at the communication device 101.
- [0028] Compared with transmitting the video through the instant message, the above-stated scheme may be useful to save the bandwidth resource with much less quality sacrifice, at least partly because of the offline Avatar data extraction, the Avatar animation, and the synchronization of the Avatar data bit streams and the audio bit streams. Further, the Avatar animation may help to keep secrecy if a message sender doesn't want to reveal his/her real image.
- [0029] Fig. 2 illustrates an embodiment of an environment 200 of the communication device 101 for generating the packet to be shared with the communication device 102. The illustrative environment 200 may include an audio recording module 201, a video capturing module 202, an audio encoding module 203, an Avatar data extraction module 204, a 2D/3D selection module 205, a synchronization module 206, a packing module 207, a transmission module 208, and others.

- [0030] In an embodiment, a user of the communication device 101 may instruct to send a message to the communication device 102, e.g., an instant message or a short message. The message may include a piece of audio recorded by the audio recording module 201 and a piece of video captured by the video capturing module 202, for example, the audio and video recorded and captured when the user speaks.
- [0031] The audio encoding module 203 may encode the recorded audio into a plurality of audio bit streams under various audio encoding schemes, such as MPEG, AAC, AC3, etc. The Avatar data extraction module 204 may extract Avatar data from the piece of video, based on Avatar technologies. The Avatar data may include Avatar parameters indicating facial expression and/or motions of the object presenting in the video, such as the user's head.
- [0032] The 2D/3D selection module 205 may select 2D or 3D dimensions for Avatar animation. The 2D/3D selection result may affect the Avatar parameters extracted by the Avatar data extraction module 204. For example, under 2D selection, the Avatar parameters may be embodied, without limitation, parameters related to x-axis translation, y-axis translation, in-plane rotation, and/or others. However, under 3D selection, the Avatar parameters may be embodied, without limitation, parameters related to out-of-plane rotation, z-axis translation, and/or others. It should be understood that 2D and 3D may have parameters in common, such as parameters related to motions of the object's mouth, nose, eyes, etc.
- [0033] The Avatar data extraction module 204 may further convert the Avatar data into one or more Avatar data bit streams using a data compression scheme, which may exploit its temporal and spatial redundancies.
- [0034] Compared with online video and audio communication and sharing, the offline messaging may allow the Avatar data extraction module 204 more time to extract the Avatar data from the captured video, which can help to improve the quality of Avatar animation at the communication device 102.

- [0035] The synchronization module 206 may generate synchronization information for synchronizing the audio bit streams with the Avatar data bit streams. The synchronization information may be embodied, without limitation, time markers (e.g., time stamps), synchronization symbols, and/or others. The synchronization module 206 may generate the synchronization information based on an audio sampling rate that the audio encoding module 203 samples the piece of audio before the encoding, and an Avatar sampling rate that the Avatar data extraction module 204 samples the piece of video before the extraction. The synchronization module 206 may further insert the synchronization information into the audio bit streams and the Avatar data bit streams.
- [0036] It should be understood that the time markers inserted into the audio bit streams may differ from the time markers inserted into the Avatar bit streams. In this case, the synchronization information may further include correlation information correlating the time markers in the audio bit streams and the time markers in the audio bit streams.
- [0037] The packing module 207 may pack the audio bit streams, the Avatar data bit streams and the synchronization information into one or more packets (e.g., the instant messages). During the packing, the audio bit streams, the Avatar data bit streams and the synchronization information may be compressed in order to further reduce the data size. In an embodiment, the packet may further include an identifier of the object presented in the video, in order to help the communication device 102 to retrieve an Avatar model corresponding to the object. In another embodiment, the communication device 101 may pack another packet containing the Avatar model and send the packet to the communication device 102, separately from the packet containing the audio and Avatar data bit streams. In an embodiment, the Avatar model may be a real model of the object, e.g., the user's head. In another embodiment, the Avatar model may be a virtual model of the object, such as a movie star's head or a cat's head.

- [0038] The transmission module 208 may transmit the packet to the server 103, regardless whether the communication device 102 is connected or not.
- [0039] Fig. 3 illustrates an embodiment of an environment 300 of the communication device 102 for receiving the packet and rendering the audio and the Avatar animation. The illustrative environment 300 may include a receiving module 301, an unpacking module 302, an audio decoding module 303, an Avatar animating module 304, a synchronizing and rendering module 305, an audio player 306, a video display 307, and/or other others.
- [0040] In an embodiment, the receiving module 301 may receive the packet containing the audio bit streams, the Avatar data bit streams, the synchronization information, the identifier of the object appearing in the video captured at the communication device 101, and/or others. The unpacking module 302 may unpack the packet into the audio bit streams, the Avatar data bit streams, the synchronization information, the object identifier, and/or others. During the unpacking, the packet may be uncompressed in order to obtain the above-stated data or information.
- [0041] The audio decoding module 303 may decode the audio bit streams into decoded audio data, under various audio decoding schemes, such as MPEG, AAC, AC3 and others. The Avatar animating module 304 may animate the Avatar model based on the Avatar data bit streams to generate an animated Avatar model. More particularly, the Avatar animating module 304 may retrieve the Avatar model corresponding to the object identifier included in the packet. In an embodiment, the Avatar model may be the real model of the object, such as the user's head. In another embodiment, the Avatar model may be the virtual model of the object, such as the cat's head. The Avatar animating module 304 may animate the Avatar model based on the Avatar data bit streams through mapping the Avatar data on the Avatar model. If the Avatar data bit streams include Avatar parameters indicating the facial expression and/or motions of the object, the Avatar animation module may map the facial expression

and/or motions on the Avatar model. For example, the Avatar animation module may transform motion animation parameters of sparse key-points of a source object model (e.g., a user's face) onto the target Avatar model and make the target Avatar model do the same animation, such as smile, surprise, yawn, etc.

[0042] Considering the Avatar model and the Avatar parameters may differ based on the 2D or 3D selection, the Avatar animation module 304 may retrieve and animate the Avatar model further based on the 2D/3D selection result, which may be sent from the communication device 101 via the packet.

[0043] The synchronizing and rendering module 305 may render the decoded audio data and the animated Avatar model, while synchronizing them with the synchronization information. For example, if the synchronization information is the timestamps inserted in the audio bit streams and the Avatar data bit streams, the synchronizing and rendering module 305 may render a part of the digital audio data decoded from the audio bit streams between the two timestamps, while rendering the Avatar model animated from the Avatar data bit streams between the same two timestamps.

[0044] It should be understood that the time stamps inserted into the audio bit streams may differ from the time stamps inserted into the Avatar bit streams. In this case, the synchronization information may further include the correlation information correlating the time stamps in the audio bit streams and the time stamps in the audio bit streams, which may be used by the synchronizing and rendering module 305 for the synchronization.

[0045] The audio player 306 and the video display 307 may play the rendered audio and Avatar video, so that a user of the communication device 102 may receive the message from the user of the communication device 101.

[0046] Fig. 4 illustrates an embodiment of a method of the communication device 101 for generating the packet containing the video and audio information to be shared with the

communication device 102. In block 401, the audio recording module 201 may record the piece of audio, and in block 402, the video capturing module 202 may capture the piece of video. For example, a recorder and a camera of the communication device 101 may record the audio and capture the video when the user of the communication device 101 speaks.

[0047] In block 403, the audio encoding module 203 may encode the piece of audio into the one or more audio bit streams under various audio encoding schemes, such as MPEG, AAC, AC3, or others. In block 404, the 2D/3D selection module 205 may select 2D or 3D for Avatar animation. Based on the 2D/3D selection result, in block 405, the Avatar data extraction module 204 may extract the Avatar data from the piece of video and convert the Avatar data into the one or more Avatar data bit streams. In an embodiment, the Avatar data may include the Avatar parameters indicating the facial expression and/or motions of the object presenting in the video, such as the user's head.

[0048] In block 406, the synchronization module 206 may generate the synchronization information for synchronizing the audio bit streams and the Avatar data bit streams. The synchronization information may be embodied, without limitation, time markers (e.g., time stamps), synchronization symbols, and/or others. In an embodiment, the synchronization module 206 may generate the synchronization information based on the audio sampling rate that the audio encoding module 203 samples the piece of audio before the encoding, and the Avatar sampling rate that the Avatar data extraction module 204 samples the piece of video before the extraction. The synchronization module 206 may further insert the synchronization information into the audio bit streams and the Avatar data bit streams.

[0049] It should be understood that the time markers may be inserted into the audio bit streams and may differ from the time markers inserted into the Avatar bit streams. In this case, the synchronization information may further include correlation information correlating the

time markers inserted in the audio bit streams and the time markers inserted in the audio bit streams.

[0050] In block 407, the packing module 207 may pack the audio bit streams, the Avatar data bit streams, and the synchronization information into the one or more packets, such as the instant messages. In an embodiment, the packet may include other information, such as the object identifier and/or 2D/3D selection result. In block 408, the transmission module 208 may transmit the packet to the server, regardless if the communication device 102 is connected to the server or not. In an embodiment, the packing module 207 may pack the Avatar model corresponding to the object into another packet, and transmit the packet separately to the server. In another embodiment, the Avatar model may be previously installed in the communication device 102. The Avatar model may differ based on the 2D/3D selection, such as 2D Avatar model or 3D Avatar model.

[0051] Fig. 5 illustrates an embodiment of a method of the communication device 102 for receiving the packet and rendering the video and audio information. In block 501, the receiving module 301 may receive the packet from the server, regardless if the communication device 101 is connected to the server or not.

[0052] In block 502, the unpacking module 302 may unpack the packet into the audio bit streams, the Avatar data bit streams, and the synchronization information. In an embodiment, the unpacking module 302 may obtain other information from the packet, such as the object identifier and/or the 2D/3D selection result. In block 503, the audio decoding module 303 may decode the audio bit streams into decoded audio data. In block 504, the Avatar animating module 304 may retrieve the Avatar model using the object identifier and/or the 2D/3D selection result.

[0053] In block 505, the Avatar animating module 304 may animate the Avatar model based on the Avatar data bit streams to generate the animated Avatar model, by mapping the Avatar data

on the Avatar model. For example, if the Avatar data includes the Avatar parameters indicating the facial expression and/or the motions of the object appearing in the video, the facial expression and/or the motions may be represented on the Avatar model. In block 506, the synchronizing and rendering module 305 may synchronize and render the decoded audio data and animated Avatar model by using the synchronization information. For example, if the synchronization information is the timestamps inserted in the audio bit streams and the Avatar data bit streams, the synchronizing and rendering module 305 may render a part of the digital audio data decoded from the audio bit streams between the two timestamps, meanwhile rendering the Avatar model animated from the Avatar data bit streams between the same two timestamps.

[0054] In block 507, the audio player 306 may play the audio and the video display 307 may play the Avatar video so that the user of the communication device 102 may receive the message from the user of the communication device 101.

[0055] While the disclosure has been illustrated and described in detail in the drawings and foregoing description, such an illustration and description is to be considered as exemplary and not restrictive in character, it being understood that only illustrative embodiments have been shown and described and that all changes and modifications consistent with the disclosure and recited claims are desired to be protected.

EXAMPLES

[0056] Illustrative examples of the devices, systems, and methods disclosed herein are provided below. An embodiment of the devices, systems, and methods may include any one or more, and any combination of, the examples described below.

[0057] In example 1, a communication device may comprise an audio encoding module to encode a piece of audio into an audio bit stream; an Avatar data extraction module to extract Avatar data from a piece of video and generate an Avatar data bit stream; and a synchronization

module to generate synchronization information for synchronizing the audio bit stream with the Avatar data bit stream.

- [0058] In example 2, the Avatar data according to the example 1, may include avatar parameters indicating at least one of movement and expression of an object presented in the video.
- [0059] In example 3, the synchronization information according to any of examples 1 and 2, may be a time marker inserted in the audio bit stream and the avatar data bit stream.
- [0060] In example 4, the synchronization information according to any of examples 1-3, may be generated based on a sampling rate of the audio bit stream and a sampling rate of the avatar data bit stream.
- [0061] In example 5, the communication device according to any of examples 1-4, may further comprise: a packaging module to package the audio bit stream, Avatar data bit stream and the synchronization information into a packet; and a transmission module to transmit the packet to a server, regardless if another communication device going to receive the packet from the server is connected or not.
- [0062] In example 6, the packet according to any of examples 1-5, may be an instant message.
- [0063] In example 7, the packet according to any of examples 1-6, may include an identifier of an object presented in the video.
- [0064] In example 8, a communication device may comprise an audio decoding module to decode an audio bit stream into decoded audio data; an Avatar animation module to animate an Avatar model based on an Avatar data bit stream to generate an animated Avatar model; and a synchronizing and rendering module to synchronize and render the decoded audio data and the animated Avatar model by utilizing the synchronization information.
- [0065] In example 9, the synchronization information according to example 8, may be a time marker inserted in the audio bit stream and the avatar data bit stream.

- [0066] In example 10, the communication device according to any of examples 8-9, may further comprise a receiving module to receive a packet from a server regardless if another communication device sending the packet is connected or not; and an unpacking module to unpack the packet into the audio bit stream, the avatar data bit stream and the synchronizing information.
- [0067] In example 11, the Avatar animation module according to any of examples 8-10, may further retrieve the avatar model corresponding to an object presented in a piece of video captured by another communication device.
- [0068] In example 12, the Avatar data according to any of examples 8-11, may include avatar parameters indicating at least one of movement and expression of an object presented in a piece of video captured by another communication device.
- [0069] In example 13, the packet according to any of examples 8-12, may further comprise an identifier to identify an object presented in a piece of video captured by another communication device.
- [0070] In example 14, the packet according to any of examples 8-13, may be an instant message.
- [0071] In example 15, a method comprises encoding a piece of audio into an audio bit stream; extracting avatar data from a piece of video to generate an avatar data bit stream; and generating synchronization information for synchronizing the audio bit stream with the avatar parameter stream.
- [0072] In example 16, the Avatar data according to example 15, includes Avatar parameters indicating at least one of movement and expression of an object presented in the video.
- [0073] In example 17, the synchronization information according to any of examples 15-16, is a time marker inserted in the audio bit stream and the avatar data bit stream.

- [0074] In example 18, the synchronization information according to any of examples 15-17, is generated based on a sampling rate of the audio bit stream and a sampling rate of the avatar data bit stream.
- [0075] In example 19, the method according to any of examples 15-18, further comprises a packaging module to package the audio bit stream, avatar data bit stream and the synchronization information into a packet; and a transmission module to transmit the packet to a server, regardless if another communication device going to receive the packet from the server is connected or not.
- [0076] In example 20, the packet according to any of examples 15-19, is an instant message.
- [0077] In example 21, the packet according to any of examples 15-20, further includes an identifier of an object presented in the video.
- [0078] In example 22, a method comprises decoding an audio bit stream into decoded audio data; animating an Avatar model based on an Avatar data bit stream to generate an animated Avatar model; and synchronizing and rendering the decoded audio data and the animated Avatar model by utilizing the synchronization information.
- [0079] In example 23, the synchronization information according to example 22, is a time marker inserted in the audio bit stream and the avatar data bit stream.
- [0080] In example 24, the method according to any of examples 22-23, comprises receiving a packet from a server regardless if another communication device sending the packet is connected or not; and unpacking the packet into the audio bit stream, the avatar data bit stream and the synchronizing information.
- [0081] In example 25, the avatar animation module according to any of examples 22-24, further retrieves the avatar model corresponding to an object presented in a piece of video captured by another communication device.

- [0082] In example 26, the avatar data according to any of examples 22-25, includes Avatar parameters indicating at least one of movement and expression of an object presented in a piece of video captured by another communication device.
- [0083] In example 27, the packet according to any of examples 22-26, further comprises an identifier to identify an object presented in a piece of video captured by another communication device.
- [0084] In example 28, the packet according to any of examples 22-27, is an instant message.
- [0085] In example 29, a machine-readable medium, comprising a plurality of instructions that in response to being executed result in a communication device to perform the method of any of examples 15-21.
- [0086] In example 30, a machine-readable medium, comprising a plurality of instructions that in response to being executed result in a communication device to perform the method of any of examples 22-28.

What is claimed is:

1. A communication device, comprising:

an audio encoding module to encode a piece of audio into an audio bit stream;

an avatar data extraction module to extract avatar data from a piece of video and generate an avatar data bit stream; and

a synchronization module to generate synchronization information for synchronizing the audio bit stream with the avatar parameter stream.
2. The communication device of claim 1, wherein the avatar data includes avatar parameters indicating at least one of movement and expression of an object presented in the video.
3. The communication device of claim 1, wherein the synchronization information is a time marker inserted in the audio bit stream and the avatar data bit stream.
4. The communication device of claim 1, wherein the synchronization information is generated based on a sampling rate of the audio bit stream and a sampling rate of the avatar data bit stream.
5. The communication device of claim 1, further comprises:

a packaging module to package the audio bit stream, avatar data bit stream and the synchronization information into a packet; and

a transmission module to transmit the packet to a server, regardless if another communication device going to receive the packet from the server is connected or not.

6. The communication device of claim 5, wherein the packet is an instant message.
7. The communication device of claim 5, wherein the packet is further to include an identifier of an object presented in the video.
8. A method, comprising:
 - encoding a piece of audio into an audio bit stream;
 - extracting avatar data from a piece of video to generate an avatar data bit stream; and
 - generating synchronization information for synchronizing the audio bit stream with the avatar parameter stream;
9. The method of claim 8, wherein the avatar data includes avatar parameters indicating at least one of movement and expression of an object presented in the video.
10. The method of claim 8, wherein the synchronization information is a time marker inserted in the audio bit stream and the avatar data bit stream.
11. The method of claim 8, wherein the synchronization information is generated based on a sampling rate of the audio bit stream and a sampling rate of the avatar data bit stream.
12. The method of claim 8, further comprising:
 - a packaging module to package the audio bit stream, avatar data bit stream and the synchronization information into a packet; and
 - a transmission module to transmit the packet to a server, regardless if another communication device going to receive the packet from the server is connected or not.
13. The method of claim 12, wherein the packet is an instant message.

14. The method of claim 12, wherein the packet is further to include an identifier of an object presented in the video.

15. A communication device, comprising:
an audio decoding module to decode an audio bit stream into decoded audio data;
an Avatar animation module to animate an Avatar model based on an Avatar data bit stream to generate an animated Avatar model; and
a synchronizing and rendering module to synchronize and render the decoded audio data and the animated Avatar model by utilizing the synchronization information.

16. The communication device of claim 15, wherein the synchronization information is a time marker inserted in the audio bit stream and the avatar data bit stream.

17. The communication device of claim 15, further comprising:
a receiving module to receive a packet from a server regardless if another communication device sending the packet is connected or not; and
an unpacking module to unpack the packet into the audio bit stream, the avatar data bit stream and the synchronizing information.

18. The communication device of claim 15, wherein the avatar animation module is further to retrieve the avatar model corresponding to an object presented in a piece of video captured by another communication device.

19. The communication device of claim 15, wherein the avatar data includes avatar parameters indicating at least one of movement and expression of an object presented in a piece of video captured by another communication device.

20. The communication device of claim 17, wherein the packet further comprises an identifier to identify an object presented in a piece of video captured by another communication device.

21. The communication device of claim 17, wherein the packet is an instant message.

22. A method, comprising:

decoding an audio bit stream into decoded audio data;

animating an Avatar model based on an Avatar data bit stream to generate an animated Avatar model; and

synchronizing and rendering the decoded audio data and the animated Avatar model by utilizing the synchronization information.

23. The method of claim 22, wherein the synchronization information is a time marker inserted in the audio bit stream and the avatar data bit stream.

24. The method of claim 22, further comprising:

receiving a packet from a server regardless if another communication device sending the packet is connected or not; and

unpacking the packet into the audio bit stream, the avatar data bit stream and the synchronizing information.

25. The method of claim 22, wherein the avatar animation module is further to retrieve the avatar model corresponding to an object presented in a piece of video captured by another communication device.

26. The method of claim 22, wherein the avatar data includes Avatar parameters indicating at least one of movement and expression of an object presented in a piece of video captured by another communication device.

27. The method of claim 24, wherein the packet further comprises an identifier to identify an object presented in a piece of video captured by another communication device.

28. The method of claim 24, wherein the packet is an instant message.

29. A machine-readable medium, comprising a plurality of instructions that in response to being executed result in a communication device to perform the method of any of claims 8-14.

30. A machine-readable medium, comprising a plurality of instructions that in response to being executed result in a communication device to perform the method of any of claims 22-28.

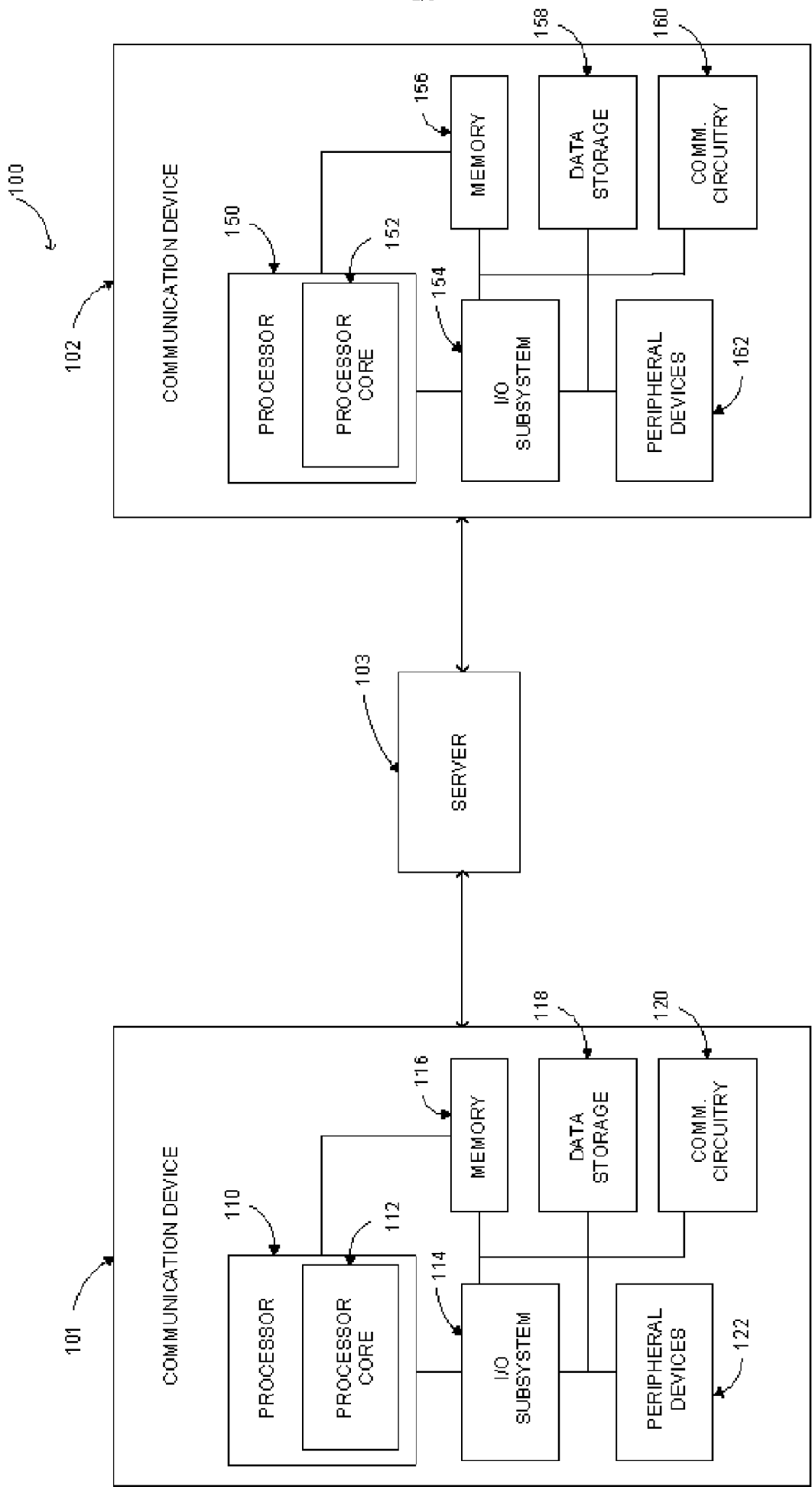


FIG. 1

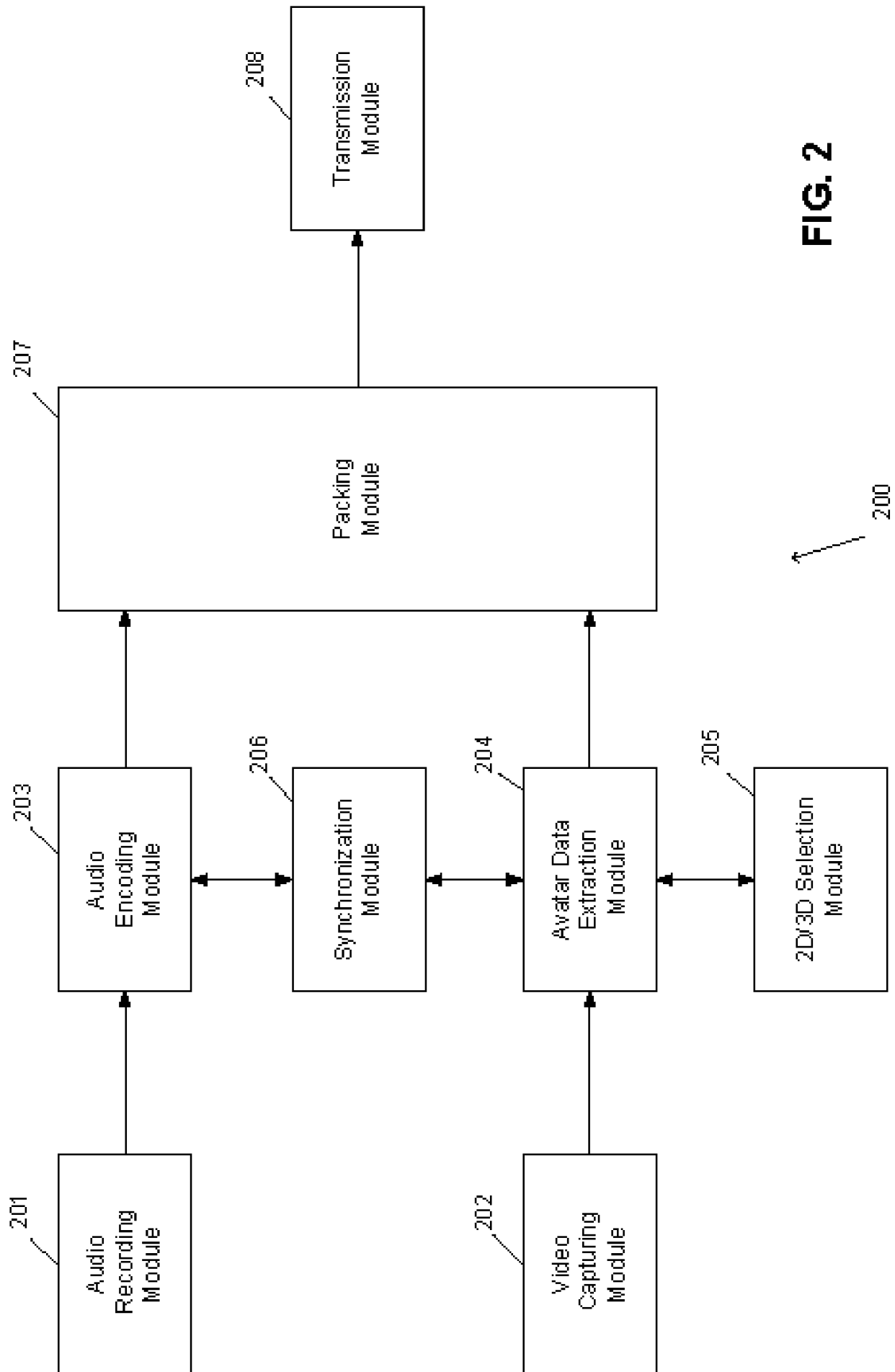


FIG. 2

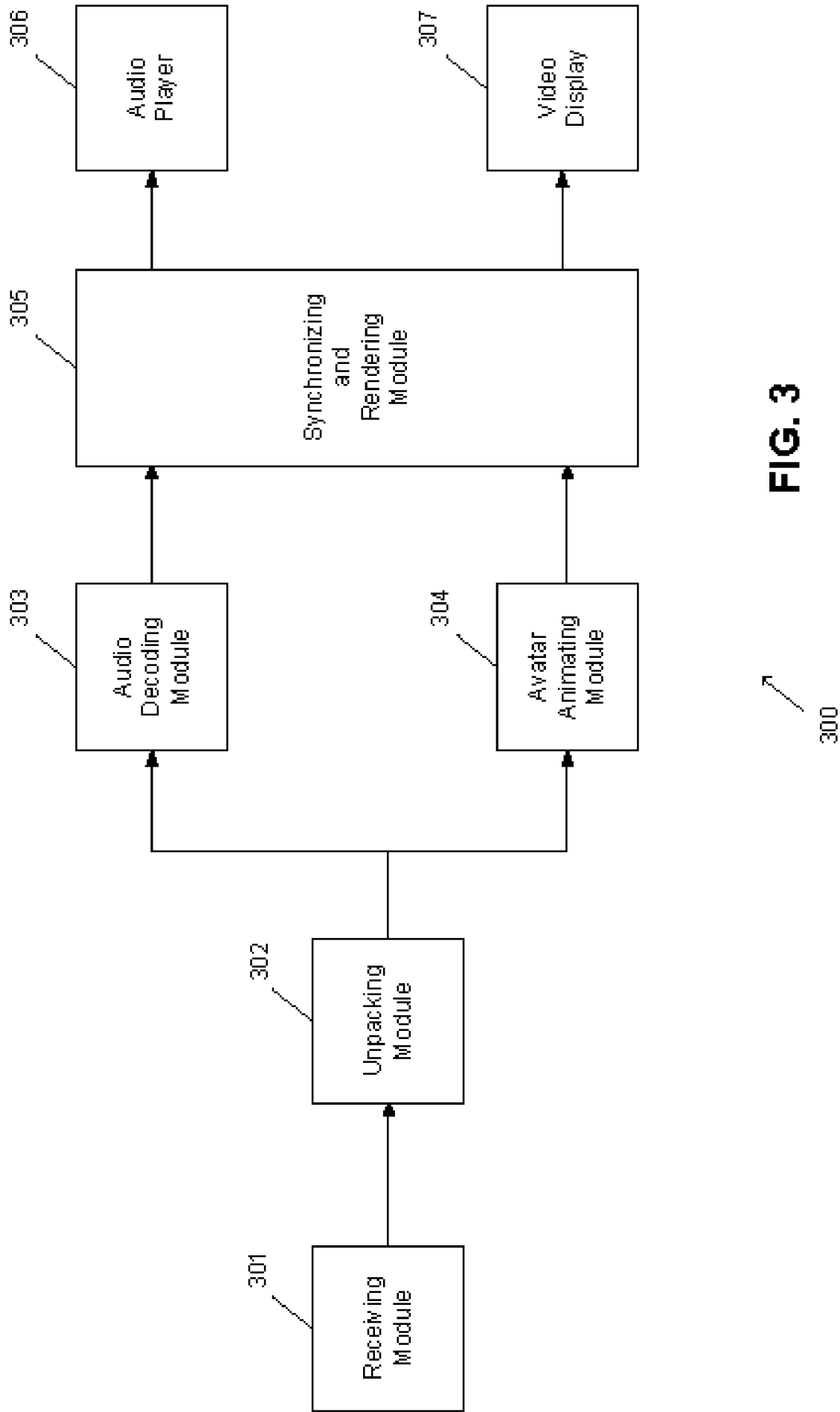
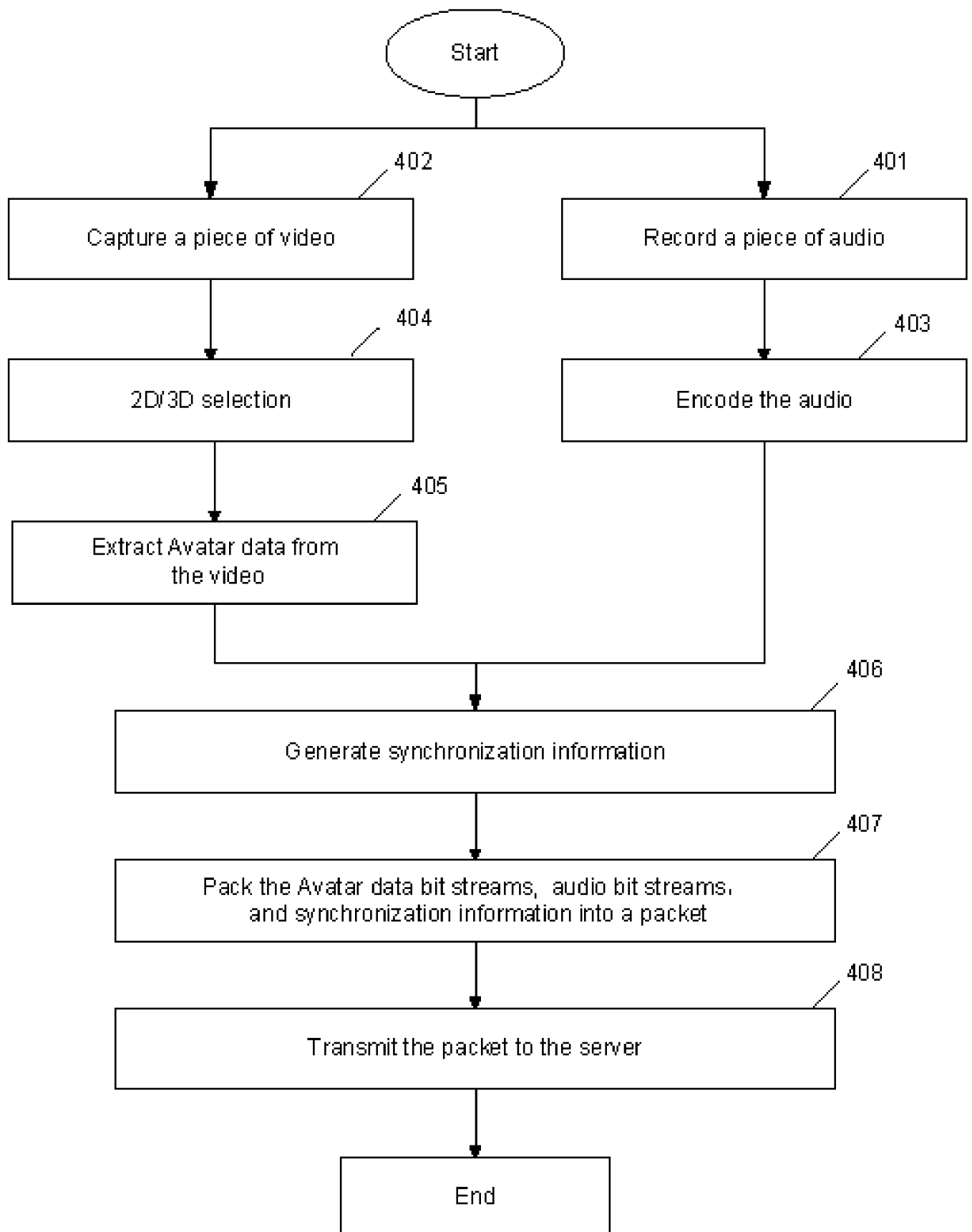
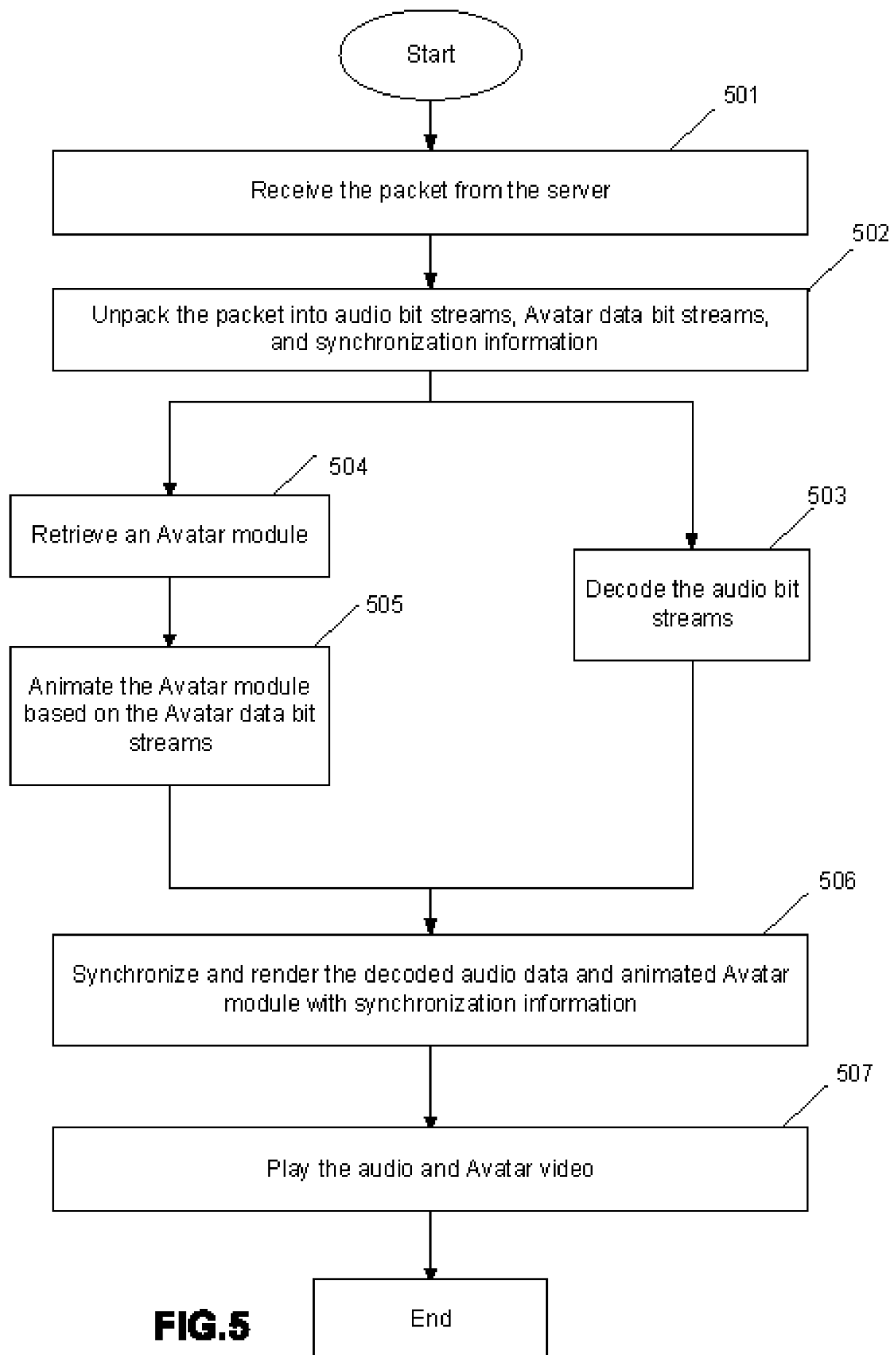


FIG. 3

**FIG.4**



INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2012/086260

A. CLASSIFICATION OF SUBJECT MATTER

H04N 7/52 (2011.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC: H04N

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

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

VEN, CNABS, CNTXT, CNKI: audio, video, bit stream, avatar, animate, extract, synchronization, time marker, code

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 1860504 A (KONINKL PHILIPS ELECTRONICS NV) 08 November 2006 (08.11.2006) description page 4, line 15 – page 6, line 3	1-30
A	CN 102663928 A (UNIV TIANJIN) 12 September 2012 (12.09.2012) the whole document	1-30
A	WO 2007110679 A3 (RADVISION LTD) 09 April 2009 (09.04.2009) the whole document	1-30

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:	“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
“A” document defining the general state of the art which is not considered to be of particular relevance	“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
“E” earlier application or patent but published on or after the international filing date	“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
“L” document which may throw doubts on priority claim (S) or which is cited to establish the publication date of another citation or other special reason (as specified)	“&” document member of the same patent family
“O” document referring to an oral disclosure, use, exhibition or other means	
“P” document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 02 September 2013 (02.09.2013)	Date of mailing of the international search report 19 Sep. 2013 (19.09.2013)
Name and mailing address of the ISA/CN The State Intellectual Property Office, the P.R.China 6 Xitucheng Rd., Jimen Bridge, Haidian District, Beijing, China 100088 Facsimile No. 86-10-62019451	Authorized officer PAN, Yun Telephone No. (86-10)62089148

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2012/086260

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 1860504 A	08.11.2006	US 2006290699 A1	28.12.2006
		US 7636662 B2	22.12.2009
		WO 2005031654 A1	07.04.2005
		EP 1671277 A1	21.06.2006
		KR 20060090687 A	14.08.2006
		JP 2007507784 T	29.03.2007
CN 102663928 A	12.09.2012	None	
WO 2007110679 A3	09.04.2009	WO 2007110679 A2	04.10.2007
		EP 1878228 A2	16.01.2008
		US 2006152577 A1	13.07.2006