



(51) International Patent Classification:
G10L 25/00 (2013.01)

(21) International Application Number:
PCT/US2021/050823

(22) International Filing Date:
17 September 2021 (17.09.2021)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/086,402 01 October 2020 (01.10.2020) US

(71) Applicant: ARRIS ENTERPRISES LLC [US/US]; 3871
Lakeview Drive, Suwanee, GA 30024 (US).

(72) Inventors: ELCOCK, Albert Fitzgerald; 1475 Glen-
brook Lane, West Chester, PA 19380 (US). DEL SOR-
DO, Christopher S.; 229 Heatherfield Drive, Souderton,

PA 18964 (US). HARDT, Charles R.; 1175 Whitehawk
Trail, Lawrenceville, GA 30043 (US).

(74) Agent: MARLEY, Robert P. et al.; Botos Churchill IP
Law LLP, 430 Mountain Ave., Suite 401, New Providence,
NJ 07974 (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, IT, 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, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,

(54) Title: CONTROLLING A MEDIA DEVICE TO PROVIDE AN IMPROVED SONIC ENVIRONMENT FOR THE RECEPTION
OF A VOICE COMMAND

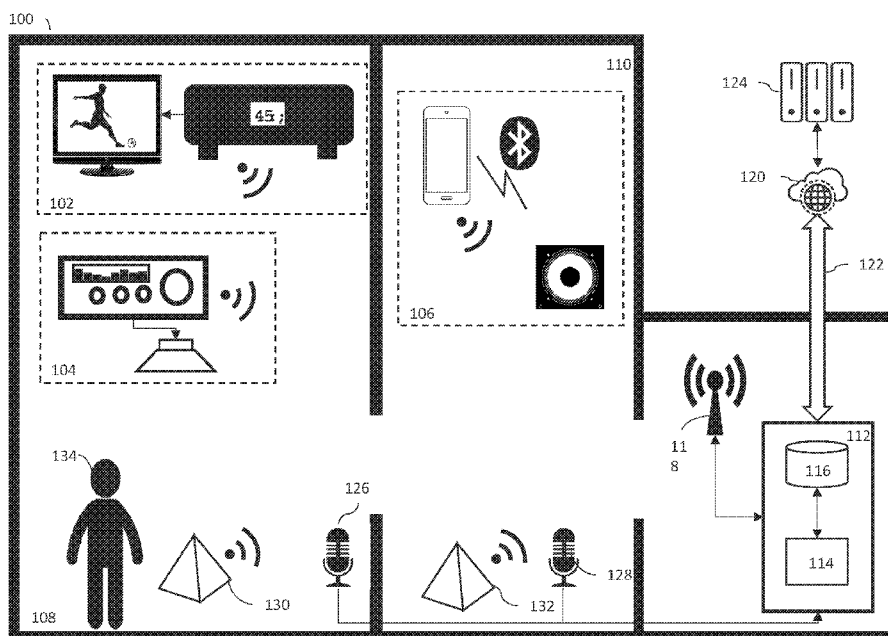


FIG. 1A

(57) Abstract: A system and method for controlling media systems identified as co-located with one or more voice-command responsive devices. The disclosed technology is responsive to predetermined activation and/or deactivation phrases so as to control the audio output of the identified media systems and provide an improved sonic environment for a voice command to be properly received and interpreted by the voice-command responsive device(s). The activation phrase for the disclosed system and method can be the same one utilized by the co-located voice-command responsive devices, thereby providing a user with an intuitive and substantially immediate means of improving the sonic environment prior to issuing one or more voice commands.

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).

Published:

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

CONTROLLING A MEDIA DEVICE TO PROVIDE AN IMPROVED SONIC ENVIRONMENT FOR THE RECEPTION OF A VOICE COMMAND

BACKGROUND OF THE INVENTION

[0001] In any given residential environment, it is not uncommon to have multiple media devices in a single room or living space. Such devices would include a multimedia gateway device (such as a digital set-top box), a television, an audio system adapted to play streaming or recorded audio, a gaming system, tablets, smartphones, etc. Each of these devices is quite likely linked (either wirelessly or wired) to a home network and capable of producing an independent audio output.

[0002] Given the proliferation of digital assistants and other voice-responsive systems and appliances, both as stand-alone devices and as integral components of other devices, it is likely that a one or more voice-responsive systems (will be situated in that same room as the aforementioned media and audio devices. All such voice-responsive systems would typically be linked to the home network and programmed to respond to verbal commands that follow the recitation of a particular activation phrase (“Alexa”, “Hey, Google”, etc.). The use of an activation phrase minimizes the chance of normal conversation being mistakenly interpreted as a command. However, given the presumed proximity of voice-responsive systems to or more of the media devices, there exists a significant chance that a verbal command would be interfered with, or entirely eclipsed by, the audio output of one or more co-located audio devices.

[0003] Although a user could preemptively pause the device programming or lower/mute the volume of one or more of the devices that were producing sound prior to the issuance of a voice command, it would likely prove inconvenient, if not entirely impractical. It is doubtful that a user would embrace the chore of manually pausing or adjusting one or more media devices every time he or she wished to voice a command to a digital assistant or other voice-responsive device. This is especially true given that one of the attributes of such devices is their inobtrusive and seamless use. Uttering a digital assistant command is ideally something that can be done without causing an interruption in a daily routine or a particular task or activity the user might be engaged in. Having to pause/adjust one or more devices, utter a command, and then resume/readjust those same devices is a task no user is likely to enjoy. Consequently, it would be advantageous to provide for a system and method whereby programming would be paused or volume level lowered/muted on one or more active

networked media devices prior to the issuance of any spoken voice-responsive system commands. Ideally, the system and method would restore the networked media devices to their previous operational states once the transaction with the voice-responsive device was completed.

BRIEF SUMMARY OF THE INVENTION

[0004] A system and method for controlling media systems identified as co-located with one or more voice-command responsive devices. The disclosed technology is responsive to predetermined activation and/or deactivation phrases so as to control the audio output of the identified media systems and provide an improved sonic environment for a voice command to be properly received and interpreted by the voice-command responsive device(s). The activation phrase for the disclosed system and method can be the same one utilized by the co-located voice-command responsive devices, thereby providing a user with an intuitive and substantially immediate means of improving the sonic environment prior to issuing one or more voice commands.

BRIEF DESCRIPTION OF THE DRAWINGS

[0005] The aspects, and advantages of the present invention will become better understood with regard to the following description, appended claims, and accompanying drawings in which:

[0006] FIG. 1A is a functional block diagram of a preferred embodiment of a system adapted to provide an improved sonic environment for voice commands depicting a user in a first location.

[0007] FIG. 1B is a functional block diagram of a preferred embodiment of a system adapted to provide an improved sonic environment for voice commands depicting a user in a second location.

[0008] FIG. 2 is a flow diagram of a process supported by the systems of FIGs. 1A and 1B.

DETAILED DESCRIPTION

[0009] FIG. 1A is a functional diagram of a first preferred embodiment of a system (100) for controlling the programming/volume level of networked media devices in response to a digital assistant recognizing a command phrase. As shown, three media systems (102, 104 and 106), each capable of producing an audio output. Systems 102 and 104 are situated

within living room 108, and system 106 is situated in foyer 110. System 102 comprises a set-top box connected to a television; system 106 is an audio system; and system 104 is a smartphone linked to a Bluetooth[®] speaker. Also shown in FIG. 1A is gateway appliance (“GWA”) 112. A cable modem or router would be examples of such an appliance. As shown, the GWA 112 includes processor 114 and memory 116, and is linked to local network wireless transceiver 118. An example of one such transceiver would be an IEEE802.11 Wi-Fi transceiver. GWA 112 is also shown to be linked to external network 120 via broadband connection 122. External network 120 provides GWA 112 with a link to multi-service operator (“MSO”) headend 124. GWA 112 would typically be located in the same residence as the media systems, but could be situated in a different room or on a different floor from those systems. Living room microphone 126 is situated within living room 108, and foyer microphone 128 is situated in foyer 110. Both of these microphones are connected to GWA 112. Although the connections are depicted as wired, they could also be wireless. In addition, there could be multiple other, similarly connected microphones throughout the residence; each situated in a particular room and/or associated with a particular digital assistant. GWA 112 is also wirelessly linked to each of the media systems (102, 104 and 106) via the local network supported by wireless transceiver 118.

[0010] Also shown in FIG. 1A are digital assistants 130 and 132. Each of these digital assistants are wirelessly linked to GWA 112 via the local network supported by wireless transceiver 118. Digital assistants 130 and 132 are programmed to respond to a variety of verbal commands and requests after a user (134) recites the activation phrase “Attention Assistant”.

[0011] Processor 112 is programmed to analyze the audio received from microphones 126 and 128, and in particular recognize audio consistent with a user reciting a GWA activation phrase and a GWA deactivation phrase. This GWA activation phrase would ideally be the same activation phrase or phrases utilized by the digital assistants within the residence (“Attention Assistant”). The GWA deactivation phrase could be any phrase that would not likely be used with great regularity in normal conversation, such as “That will be all, Assistant”. Processor 112 is adapted to enable a user to specify particular GWA activation and deactivation phrases by interfacing with GWA 112 via a networked device such as a computer, tablet or smartphone. Processor 112 then stores the GWA activation and deactivation phrases in memory 114.

[0012] The process that system 100 executes with a user in living room 108 is depicted in the flow diagram of FIG. 2. As shown, in steps 202 and 204 the processor

determines if a signal from living room microphone 126 consistent with user 134 having spoken the words “Attention Assistant” has been received and analyzed. If so, processor 114 checks the stored information in memory 116 to determine if any media devices have been identified as being co-located living room microphone 126 (and digital assistant 130) (see step 206). The information stored in memory 116 that can be accessed by processor 114 is represented in Table A, below:

Microphone ID	Associated GWA Activation Phrase	Associated GWA Deactivation Phrase	Associated Media Systems	Activation Command	Completion Command
Living Room	Attention Assistant	That will be all Assistant	Set-Top Box/TV	Pause Program	Resume Program
			Audio System	Set volume to 10% of maximum	Restore volume to previous level
Foyer	Attention Assistant	That will be all Assistant	Smartphone/Bluetooth® Speaker	None	None

If no such media devices are identified, the process reverts back to step 202. However, if there are identified co-located media devices, processor 114 then queries memory 114 to determine if the user has stored activation commands for the identified media systems (step 208). If not, the process reverts to step 202. However, if associated activation commands have been stored in memory 114, as is the case for media systems 102 and 104, instructs GWA 112 to generate and transmit the associated activation commands to the media devices via the local wireless network (step 210). As shown in Table A, this would result in sending a command to pause any program being viewed to media system 102, and command to lower the volume level to 10% of maximum to media system 104. The process then continues with step 202.

[0013] Once the transmitted commands are received by media systems 102 and 104, user 134 will have a much quieter environment in which to recite one or more voice commands for digital assistant 130. When user 134 has complete his or her verbal transaction with digital assistant 130, they would say “That will be all, Assistant” if they desired the various media systems to be returned to their previous states.

[0014] If user 134 does recite “That will be all, Assistant”, processor 114 would again analyze the signal received from living room microphone 126, first checking to determine if the received signal is consistent with the phrase “Attention Assistant” has been spoken (steps 202 and 204). That test would fail and the process would continue with step 212 wherein

processor 114 would analyze the received signals to determine if they were consistent with the with the phrase “That will be all, Assistant” having been spoken. If that test fails, the process reverts back to step 202. However, if the conditional is answered in the affirmative, processor 114 determines if any media systems have been identified as associated with foyer microphone 128 (and digital assistant 132) (step 214). If there are no such associated devices the process reverts to step 202. However, as shown in Table A, media systems 102 and 104 have been identified as associated with foyer microphone 128, and the process would continue with step 216. In step 216 the processor determines if the user has stored deactivation commands for the identified media systems in memory 116. (step 216). As shown in table A, there deactivation commands associated with media systems 102 and 104. Accordingly, processor 114 instructs GWA 112 to generate and transmit the associated deactivation commands to media devices 102 and 104 via the local wireless network (step 218). This would result in sending a command to resume playing any previously paused program to media system 102, and a command to raise the volume to its previous level to media system 104. The process then continues with step 202.

[0015] If user 134 were to enter foyer 110, as shown in FIG. 1B, and speak the phrase “Attention Assistant”, digital assistant 132 and GWA 112 would react in much the same manner as was outlined above with respect to the living room scenario. However, as shown in Table A, the media system identified as being co-located with digital assistant 132 (smartphone and Bluetooth[®] speaker) has no activation or deactivation command associated with it. Consequently, the processor would determine that a signal from foyer microphone 128 consistent with user 134 having spoken the words “Attention Assistant” had been received (steps 202 and 204). The processor would then determine that media system 106 was identified as being co-located with foyer microphone 128 (and digital assistant 132) (see step 206). However, in step 208 processor 114 would determine that memory 116 stored no associated activation commands for media system 106 (see Table A). Having found no specified activation commands associated with media system 106, the process would revert to step 202 and processor 114 would continue to analyze any signals received from microphones 126 and 128 (steps 204 and 212). No commands would be sent to media system 106 and user 134 would not find the sonic environment within foyer 110 to be any more or less conducive to the reception of voice commands by digital assistant 132 that it was prior to the recitation of the phrase “Attention Assistant”.

[0016] As shown in FIGs. 1A and 1B, GWA 112 is connected, either directly or indirectly, to external network 120 and MSO headend 124. It will be understood that all or

part of the above describe functionality of GWA 112 could be performed by an offsite server or processing means located either at the MSO headend or otherwise connected via the external network. Such offsite processing/storage is commonly employed in support of digital assistants and its implementation and usage is well known in the art.

[0017] Although the invention herein has been described with reference to particular embodiments, it is to be understood that these embodiments are merely illustrative of the principles and applications of the present invention. For example, any of the links or connections described as being wireless could be accomplished via wired means (such as Ethernet). The converse holds true for any links or connections described as being wired – They could be accomplished via wireless means without departing from the scope of the disclosed invention. The wireless network supporting the disclosed invention could be any such network capable of supporting the necessary connectivity, such as complying with IEEE 802.11 Wi-Fi, IEEE 802.15 ZigBee, and/or Bluetooth® standards. The microphones utilized by the disclosed system to capture user commands need not be stand-alone devices. The system could utilize microphones embedded in any device or system connected to the local network, such as those found in tablets, smartphones, televisions, computers, set-top boxes, etc. to monitor user voice commands, including the microphone or microphones integral to digital assistants or other voice-responsive devices. Although the above embodiments specify that the voice-responsive device is a digital assistant, the disclosed system and method could be utilized to provide an improved sonic environment for any voice-responsive device or system. Additionally, the media systems that could be controlled via the disclosed technology are in no way limited to those specifically referenced above, nor are the commands that could be communicated to the media systems limited to those mentioned above. For example, the GWA could send a command to a media system instructing it to turn off or turn on, or tune to a particular station or channel, or access a particular source. Media systems could include, but are not limited to: computing systems, computer monitors, gaming systems, radios, tablets, smartphones, audio systems, televisions, projectors, set-top boxes, etc. All of the above variations and reasonable extensions therefrom could be implemented and practiced without departing from the spirit and scope of the present invention as defined by the appended claims.

CLAIMS

1. A system for controlling at least one media device comprising:
 - at least one voice-responsive device activated by an activation phrase and situated in a first location;
 - at least one microphone adapted to capture sounds in the first location;
 - at least one media system situated in the first location;
 - at least one memory adapted to store information indicative of:
 - the activation phrase;
 - the identity and location of the at least one microphone;
 - the identity and location of the at least one voice-responsive device;
 - the identity and location of the at least one media system; and
 - at least one command applicable to the at least one media system;
 - at least one network linking the at least one voice responsive device, the at least one microphone; the at least one media system; the at least one memory, and at least one processor, wherein the at least one processor is adapted to:
 - receive information indicative of sounds captured by the least one microphone in the first location;
 - detect a pattern within the received information consistent with the activation phrase having been spoken;
 - retrieve from the at least one memory, based at least in part upon the detection the pattern:
 - the identity and location of the at least one microphone;
 - the identity of the at least one media system; and
 - the at least one command; and
 - communicate the at least one command to the at least one media device.
2. The system of claim 1 wherein the at least one voice-responsive device comprises a digital assistant.
3. The system of claim 1 wherein the at least one command comprises an instruction to for the at least one media device to execute at least one of the following actions:
 - adjust audio volume;

set audio volume to a predetermined level;
pause a video program;
pause an audio program;
mute audio;
tune to a particular channel;
access a particular source;
power on; and
power off.

4. The system of claim 1 wherein at least one media device comprises at least one of the following:

a television;
a radio;
a gaming system;
a computing system;
a computer monitor;
a tablet;
a smartphone;
a projector;
an audio system;
a set-top box; and
a digital assistant.

5. The system of claim 1, wherein the at least one network comprises at least one of:

an 802.11 network;
an 802.15 network;
a Bluetooth network; and
an Ethernet network.

6. The system of claim 1, wherein:

the at least one memory is further adapted to store information indicative of a deactivation phrase; and

the at least one processor is further adapted to:

detect a pattern within the received information consistent with the deactivation phrase having been spoken;

retrieve from the at least one memory, based at least in part upon the detection the pattern consistent with the deactivation phase having been spoken:

the identity and location of the at least one microphone;

the identity of the at least one media system; and

the at least one command; and

communicate the at least one command to the at least one media device.

7. The system of claim 6 wherein the at least one voice-responsive device comprises a digital assistant.

8. The system of claim 6 wherein the at least one command comprises an instruction to for the at least one media device to execute at least one of the following actions:

adjust audio volume;

set audio volume to a predetermined level;

pause a video program;

pause an audio program;

mute audio;

tune to a particular channel;

access a particular source;

power on; and

power off.

9. The system of claim 6 wherein at least one media device comprises at least one of the following:

a television;

a radio;

a gaming system;

a computing system;

a computer monitor;

a tablet;

a smartphone;

a projector;

an audio system;

a set-top box; and

a digital assistant.

10. The system of claim 6 wherein the at least one network comprises at least one of:

- an 802.11 network;
- an 802.15 network;
- a Bluetooth network; and
- an Ethernet network.

11. A method for controlling at least one media device comprising in a system comprising:

at least one voice-responsive device activated by an activation phrase and situated in a first location;

at least one microphone adapted to capture sounds in the first location;

at least one media system situated in the first location;

at least one memory adapted to store information indicative of:

- the activation phrase;
- the identity and location of the at least one microphone;
- the identity and location of the at least one voice-responsive device;
- the identity and location of the at least one media system; and
- at least one command applicable to the at least one media system;

at least one network linking the at least one-voice responsive device, the at least one microphone; the at least one media system; the at least one memory;

the method comprising the steps of:

receiving information indicative of sounds captured by the least one microphone in the first location;

detecting a pattern within the received information consistent with the activation phrase having been spoken;

retrieving from the at least one memory, based at least in part upon the detection the pattern:

- the identity and location of the at least one microphone;
- the identity of the at least one media system; and
- the at least one command; and

communicating the at least one command to the at least one media device.

12. The method of claim 11 wherein the at least one voice-responsive device comprises a digital assistant.

13. The method of claim 11 wherein the at least one command comprises an instruction to for the at least one media device to execute at least one of the following actions:

- adjust audio volume;
- set audio volume to a predetermined level;
- pause a video program;
- pause an audio program;
- mute audio;
- tune to a particular channel;
- access a particular source;
- power on; and
- power off.

14. The method of claim 11 wherein at least one media device comprises at least one of the following:

- a television;
- a radio;
- a gaming system;
- a computing system;
- a computer monitor;
- a tablet;
- a smartphone;
- a projector;
- an audio system;
- a set-top box; and
- a digital assistant.

15. The method of claim 11 wherein the at least one network comprises at least one of:

- an 802.11 network;
- an 802.15 network;
- a Bluetooth network; and
- an Ethernet network.

16. The method of claim 11 wherein:

the at least one memory is further adapted to store information indicative of a deactivation phrase; and

further comprising the steps of:

detecting a pattern within the received information consistent with the deactivation phrase having been spoken;

retrieving from the at least one memory, based at least in part upon the detection the pattern consistent with the deactivation phase having been spoken:

the identity and location of the at least one microphone;

the identity of the at least one media system; and

the at least one command; and

communicating the at least one command to the at least one media device.

17. The method of claim 16 wherein the at least one voice-responsive device comprises a digital assistant.

18. The method of claim 16 wherein the at least one command comprises an instruction to for the at least one media device to execute at least one of the following actions:

adjust audio volume;

set audio volume to a predetermined level;

pause a video program;

pause an audio program;

mute audio;

tune to a particular channel;

access a particular source;

power on; and

power off.

19. The method of claim 16 wherein at least one media device comprises at least one of the following:

a television;

a radio;

a gaming system;

- a computing system;
- a computer monitor;
- a tablet;
- a smartphone;
- a projector;
- an audio system;
- a set-top box; and
- a digital assistant.

20. The method of claim 16 wherein the at least one network comprises at least one of:

- an 802.11 network;
- an 802.15 network;
- a Bluetooth network; and
- an Ethernet network.

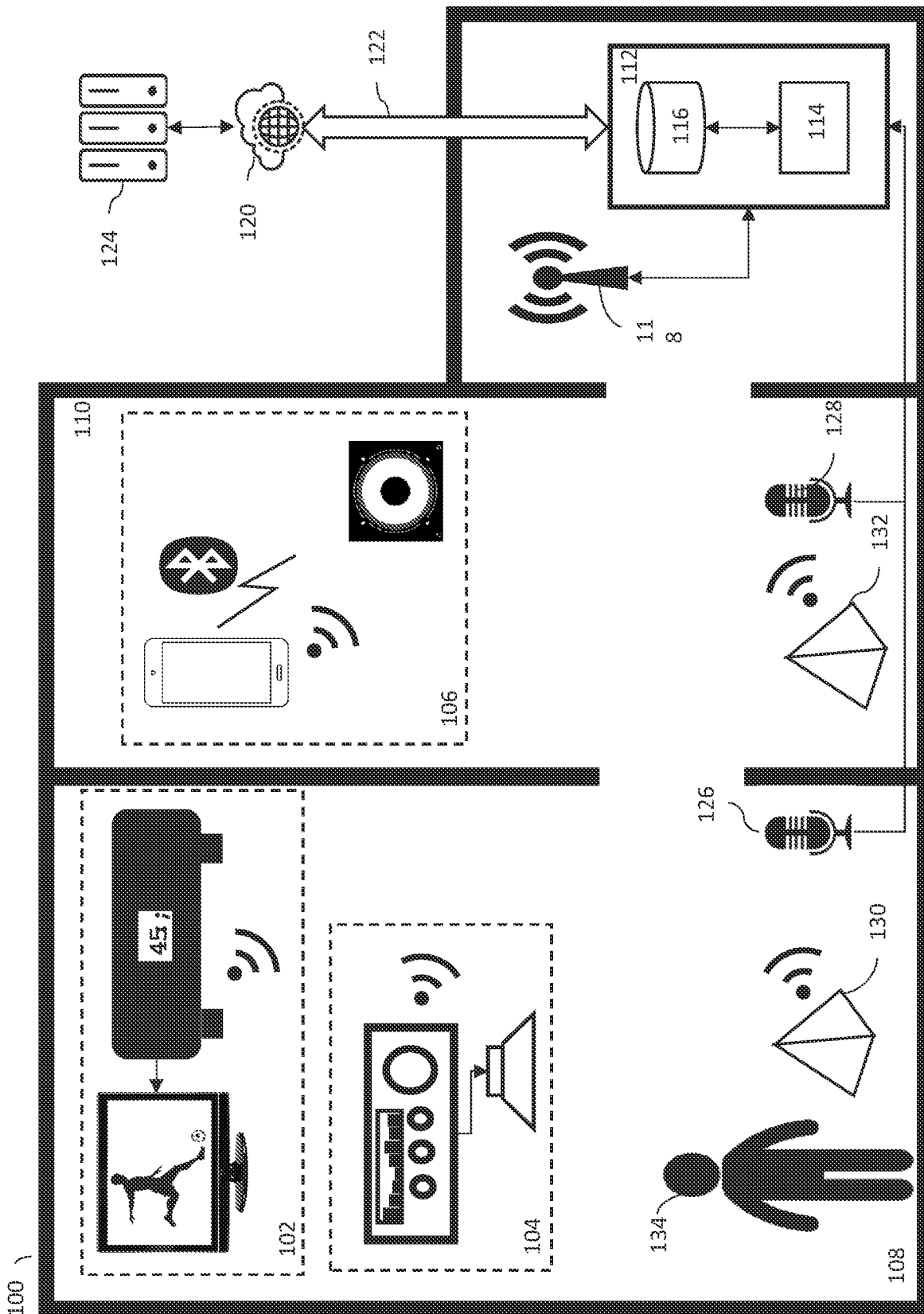


FIG. 1A

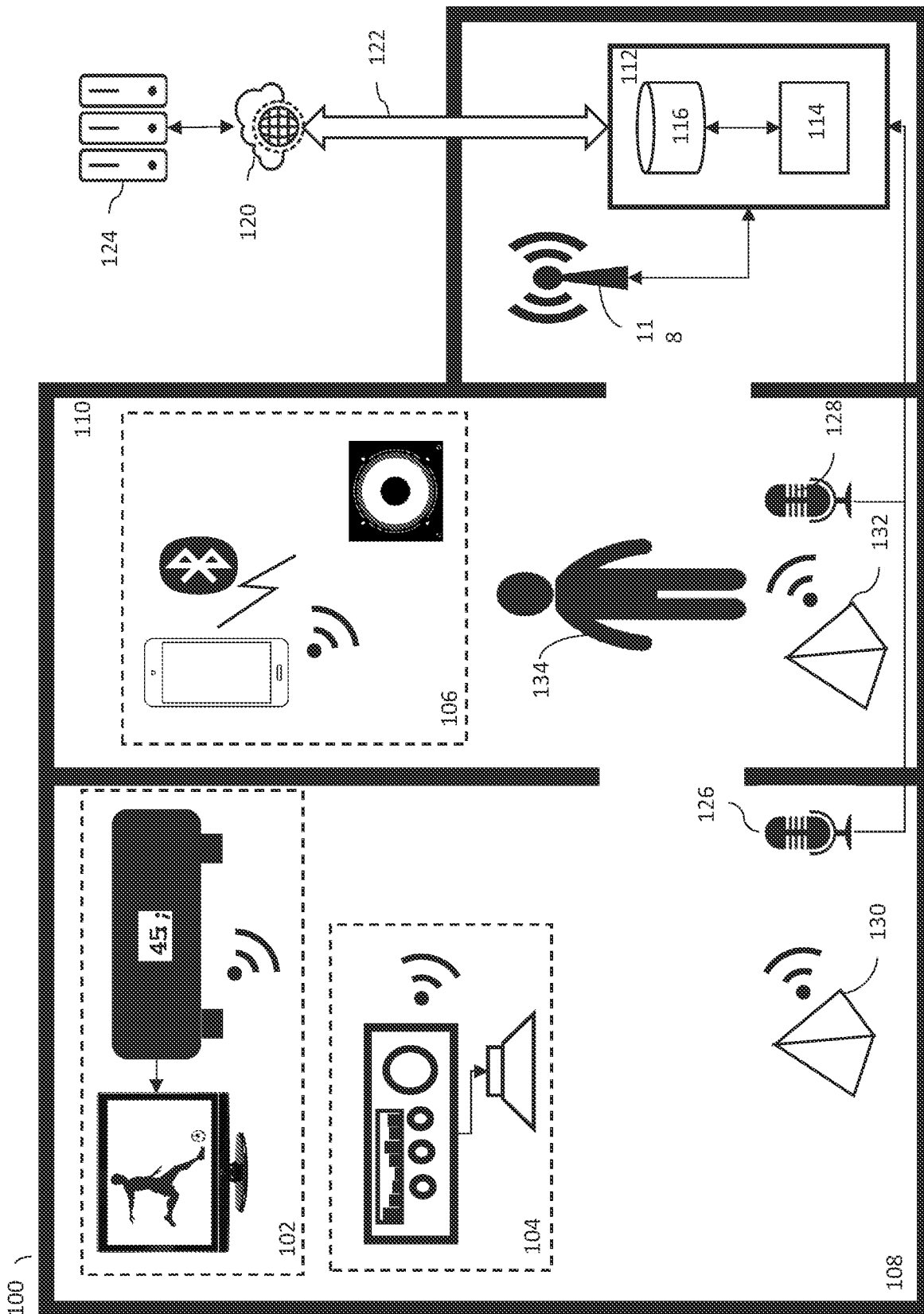


FIG. 1B

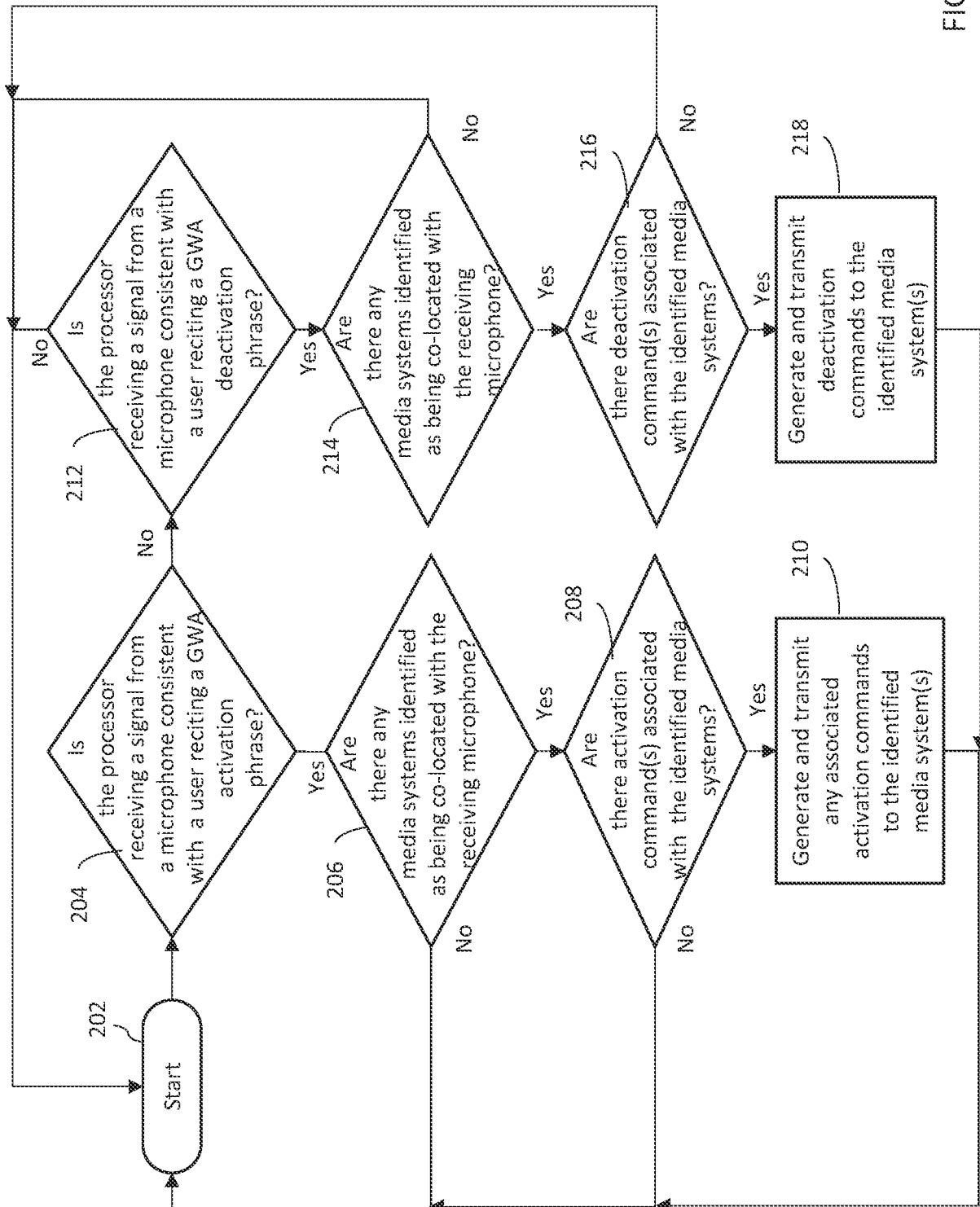


FIG. 2

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 21/50823

A. CLASSIFICATION OF SUBJECT MATTER

IPC - G10L 25/00 (2021.01)

CPC - G10L 15/265, G10L 15/22, G06F 3/16, H04M 2201/40, G10L 15/30, H04M 3/4938, G10L 19/12, G10L 19/173, G10L 19/10, G10L 19/0018, G10L 19/24

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)

See Search History document

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

See Search History document

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

See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X ---- Y	US 2019/0362608 A1 (GOOGLE LLC) 28 November 2019 (28.11.2019) entire document, especially para [0113], [0111].	1-2, 6-7, 11-12, 16-17 ----- 3-5, 8-10, 13-15, 18-20
Y	US 2019/0172467 A1 (Apple Inc.) 06 June 2019 (06.06.2019) entire document, especially para [0007].	3-5, 8-10, 13-15, 18-20

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

* Special categories of cited documents:

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

"D" document cited by the applicant in the international application

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search

14 November 2021 (14.11.2021)

Date of mailing of the international search report

DEC 27 2021

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Kari Rodriguez

Telephone No. PCT Helpdesk: 571-272-4300