



(51) International Patent Classification:

G11B 27/031 (2006.01) G06F 16/68 (2019.01)

G11B 27/10 (2006.01)

(21) International Application Number:

PCT/US2023/022164

(22) International Filing Date:

13 May 2023 (13.05.2023)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

17/746,873 17 May 2022 (17.05.2022) US

(71) Applicant: META PLATFORMS, INC. [US/US]: 1601 Willow Road, Menlo Park, California 94025 (US).

(72) Inventors: ORTEGA ESTRADA, Salvael; 1601 Willow Road, Menlo Park, California 94025 (US). LOVITT, Andrew; 1601 Willow Road, Menlo Park, California 94025 (US).

(74) Agent: COLBY, Steven et al.; Rimon Law, 420 West Main Street, Suite 101B, Boise, Idaho 83702 (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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, 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))

(54) Title: SYSTEM AND METHOD FOR INCORPORATING AUDIO INTO AUDIOVISUAL CONTENT

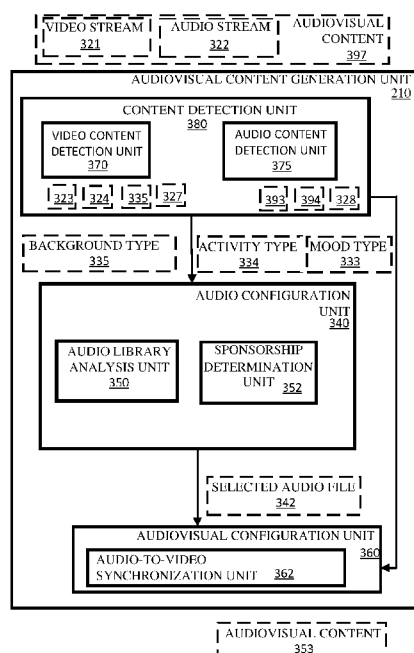


FIG. 3

(57) Abstract: In some embodiments, a method includes detecting an incorporation attribute of audiovisual content; analyzing an audio library to determine an audio file that maps to the incorporation attribute; selecting the audio file from the audio library that maps to the incorporation attribute; incorporating the selected audio file into the audiovisual content to generate egocentric audiovisual content; and providing the egocentric audiovisual content to a user for audiovisual consumption. In some embodiments of the method, the incorporation attribute is at least one of a mood of a target of the audiovisual content, an activity of the target of the audiovisual content, and a background of the audiovisual content.

SYSTEM AND METHOD FOR INCORPORATING AUDIO INTO AUDIOVISUAL CONTENT

FIELD

[0001] The present disclosure relates to editing audiovisual content and, in particular, to incorporating audio into audiovisual content.

BACKGROUND

[0002] The background description provided herein is for the purpose of generally presenting the context of the disclosure. Work of the presently named inventor(s), to the extent it is described in this background section, as well as aspects of the description that may not otherwise qualify as prior art at the time of filing, are neither expressly nor impliedly admitted as prior art against the present disclosure.

[0003] Due to the ease of accessibility and hands-free characteristics of smart glasses, smart glasses are becoming ubiquitous for the generation of audiovisual content for social networks and other social platforms. While off-the-shelf video editing capabilities that are used to generate the audiovisual content for smart glasses have advanced substantially, creators and users of the smart glasses still have substantial issues when editing audiovisual content. For example, adding music or audio effects to video using the smart glasses is often cumbersome and inefficient due to the number of steps and processing power required to add the music and audio effects to the videos. Furthermore, current techniques used to add music or audio effects to video do not appropriately rectify the above issues since the current techniques generally require the manual editing of the audiovisual content.

[0004] The present disclosure seeks to address, at least in part, any or all of the drawbacks and disadvantages described above.

SUMMARY

[0005] According to a first aspect of the present disclosure there is provided a method comprising: detecting an incorporation attribute of audiovisual content; analyzing an audio library to determine an audio file that maps to the incorporation attribute; selecting the audio file from the audio library that maps to the incorporation attribute; incorporating the selected audio file into the audiovisual content to generate egocentric audiovisual content; and providing the egocentric audiovisual content to a user for audiovisual consumption.

[0006] In some embodiments of the method, the incorporation attribute may be at least one of a mood of a target of the audiovisual content, an activity of the target of the audiovisual content, and a background of the audiovisual content.

[0007] In some embodiments of the method, the audiovisual content may include video

1 content of a video stream and audio content of an audio stream.

2 [0008] In some embodiments, the method may further comprise synchronizing the selected
3 audio file to the audiovisual content based upon an audio file marker placed in the audio file.

4 [0009] In some embodiments, the method may further comprise mapping audio files in the
5 audio library to at least one of an activity type, a mood type, a background type, and a
6 sponsorship type.

7 [0010] In some embodiments of the method, an audio configuration unit may select the audio
8 file based upon the mapping of the audio file to the incorporation attribute.

9 [0011] In some embodiments of the method, an audio library analysis unit may analyze the
10 audio library to determine which audio file in the audio library maps to the incorporation
11 attribute of the audiovisual content.

12 [0012] In some embodiments of the method, synchronizing the audio file to a video stream of
13 the audiovisual content may be dependent upon a video stream timing marker being placed in
14 the video stream of the audiovisual content.

15 [0013] In some embodiments of the method, synchronizing the audio file to the video stream
16 may be dependent upon an audio stream time marker being placed in the audio stream.

17 [0014] According to a second aspect of the present disclosure there is provided a system
18 comprising: a video recorder; an audio recorder coupled to the video recorder; and a
19 controller coupled to the video recorder and the audio recorder, wherein the controller is
20 configured to: detect an incorporation attribute of audiovisual content; analyze an audio
21 library to determine an audio file that maps to the incorporation attribute; select the audio file
22 from the audio library that maps to the incorporation attribute; incorporate the selected audio
23 file into the audiovisual content to generate egocentric audiovisual content; and provide the
24 egocentric audiovisual content to a user for audiovisual consumption.

25 [0015] In some embodiments of the system, the incorporation attribute may be at least one of
26 a mood of a target of the audiovisual content, an activity of the target of the audiovisual
27 content, and a background of the audiovisual content.

28 [0016] In some embodiments of the system, the audiovisual content may include video
29 content of a video stream and audio content of an audio stream.

30 [0017] In some embodiments of the system, the selected audio file may be synchronized to
31 the audiovisual content based upon an audio file marker placed in the audio file.

32 [0018] In some embodiments of the system, audio files in the audio library may be mapped to
33 at least one of an activity type, a mood type, a background type, and a sponsorship type.

34 [0019] In some embodiments of the system, an audio configuration unit may select the audio

1 file based upon the mapping of the audio file to the incorporation attribute.

2 [0020] In some embodiments of the system, an audio library analysis unit may analyze the
3 audio library to determine which audio file in the audio library maps to the incorporation
4 attribute of the audiovisual content.

5 [0021] In some embodiments of the system, the audio file may be synchronized to a video
6 stream of the audiovisual content using a video stream timing marker that is placed in the
7 video stream of the audiovisual content.

8 [0022] In some embodiments of the system, the audio file that may be synchronized to a
9 video stream of the audiovisual content using an audio stream time marker that may be
10 placed in the audio stream of the audiovisual content.

11 [0023] According to a third aspect of the present disclosure there is provided a non-transitory
12 computer-readable medium storing instructions that, when executed by one or more
13 processors, cause the one or more processors to perform operations including: detecting an
14 incorporation attribute of audiovisual content; analyzing an audio library to determine a first
15 audio file that maps to the incorporation attribute of the audiovisual content; selecting the
16 first audio file from the audio library that maps to the incorporation attribute; incorporating
17 the selected first audio file into the audiovisual content to generate egocentric audiovisual
18 content; uploading the egocentric audiovisual content to a social network for view by a
19 viewer of the social network; and modifying the egocentric audiovisual content to include a
20 second audio file selected using preferences of the viewer of the egocentric audiovisual
21 content.

22 [0024] In some embodiments of the non-transitory computer-readable medium, the second
23 audio file selected by the viewer of the egocentric audiovisual content may replace the first
24 audio file selected from the audio library that maps to the incorporation attribute.

25 [0025] It will be appreciated that any features described herein as being suitable for
26 incorporation into one or more aspects or embodiments of the present disclosure are intended
27 to be generalizable across any and all aspects and embodiments of the present disclosure.

28 Other aspects of the present disclosure can be understood by those skilled in the art in light of
29 the description, the claims, and the drawings of the present disclosure. The foregoing general
30 description and the following detailed description are exemplary and explanatory only and
31 are not restrictive of the claims.

32 BRIEF DESCRIPTION OF THE DRAWINGS

33 [0026] Embodiments of the present disclosure will now be described, by way of example
34 only, and with reference to the accompanying drawings, in which:

1 [0027] FIG. 1 is a diagram of a headset in accordance with one or more embodiments of the
2 present disclosure;

3 [0028] FIG. 2 is a block diagram of an example environment in accordance with one or more
4 embodiments of the present disclosure;

5 [0029] FIG. 3 is an audiovisual content generation unit in accordance with one or more
6 embodiments of the present disclosure;

7 [0030] FIG. 4 is flowchart diagram illustrating a method for generating audiovisual content
8 in accordance with one or more embodiments of the present disclosure; and

9 [0031] FIG. 5 is a block diagram of a computing environment according to an example
10 implementation of the present disclosure.

11 DETAILED DESCRIPTION

12 [0032] FIG. 1 illustrates a diagram of a headset 100 in accordance with one or more
13 embodiments. In some embodiments, the headset 100 includes a display 105, a video camera
14 180, a microphone 190, a controller 170, a speaker 120, and a frame 110. In some
15 embodiments, the frame 110 enables the headset 100 to be worn on a face of a user of the
16 headset 100 and houses the components of the headset 100. In one embodiment, the headset
17 100 may be a head mounted display (HMD). In another embodiment, the headset 100 may be
18 a near eye display (NED). In some embodiments, the headset 100 may be worn on the face of
19 the user such that audiovisual content generated using the audiovisual content generation
20 features described herein is presented using the headset 100. In some embodiments,
21 audiovisual content may include audio content and visual content that is presented to the user
22 via the speaker 120 and the display 105, respectively. In some embodiments, the audiovisual
23 content generated by the headset 100 may be provided to other electronic display devices for
24 display to additional users or creators of audiovisual content.

25 [0033] In some embodiments, the display 105 presents the visual content of the audiovisual
26 content to the user of the headset 100. In some embodiments, the visual content may be video
27 content recorded by the video camera 180 of the headset 100 or video content recorded by a
28 video camera external to the headset 100. In some embodiments, the display 105 may be an
29 electronic display element, such as a liquid crystal display (LCD), an organic light emitting
30 diode (OLED) display, a quantum organic light emitting diode (QOLED) display, a
31 transparent organic light emitting diode (TOLED) display, some other display, or some
32 combination thereof. In some embodiments, the display 105 may be backlit. In some
33 embodiments, the display 105 may include one or more lenses configured to augment the
34 video observed by the user while wearing the headset 100.

1 [0034] In some embodiments, the speaker 120 presents the audio content of the audiovisual
2 content to the user of the headset 100. In some embodiments, the audio content may include
3 audio, such as music, advertisements, and other audio effects, incorporated (either
4 automatically or with human assistance) into the audio presented to the user based on the
5 context of the video content and audio content recorded by the headset 100 or other external
6 device. In some embodiments, the audio content is incorporated into audio content recorded
7 by the headset 100 such that the user does not have to manually edit the audio content to add
8 audio content to the video content. In some embodiments, the audiovisual content generated
9 by the headset 100 is presented to the user of the headset 100.

10 [0035] FIG. 2 illustrates an environment 200 in accordance with some embodiments. In some
11 embodiments, the environment 200 includes a server 220, a network 225, and headset 100 of
12 FIG. 1. In some embodiments, the headset 100 includes display 105, video camera 180,
13 microphone 190, speaker 120, and controller 170. In some embodiments, the controller 170
14 of headset 100 includes an audio processor 242, an image processor 244, an audiovisual
15 content generation unit 210, and memory 215. In some embodiments, the image processor
16 244 is a processor configured to recognize and detect content within a video stream provided
17 to the image processor 244. In some embodiments, the audio processor 242 is a processor
18 configured to recognize and detect content within an audio stream provided to the audio
19 processor 242. In some embodiments, the audio processor 242 and the image processor 244
20 are configured to recognize or detect an activity type of a target (e.g., an individual or
21 individuals) of the audiovisual content, a mood type of the target of the audiovisual content,
22 and background type of the video content of the audiovisual content (described further herein
23 with reference to FIG. 3). In some embodiments, audiovisual content generation unit 210 is
24 processing logic (e.g., hardware, software, or a combination of both) configured to generate
25 audiovisual content using the content detection capabilities provided by audio processor 242
26 and image processor 244 to select and incorporate audio files (e.g., music, advertisements,
27 and other audio effects) into audiovisual content presented to the user of headset 100.

28 [0036] In some embodiments, the server 220 includes an audio library 251 that includes a
29 plurality of audio files 278. In some embodiments, the audio files 278 are audio files, such as,
30 music files, advertisements, or other audio effects that are incorporated into audiovisual
31 content that is presented to the user of headset 100. In some embodiments, the audio files 278
32 include attribution identification tags that are configured to be used to identify the audio files
33 for use in incorporation of the audio files into the audiovisual content generated by the
34 audiovisual content generation unit 210 (described further in detail herein).

1 [0037] In some embodiments, the attribution identification tags include, for example, an
2 activity type 281, a mood type 282, a background type 283, and a sponsor indicator 284. In
3 some embodiments, the activity type 281 is an indicator of the type of activity the audio file
4 may be associated with, the mood type 282 is an indicator of the type of mood the audio file
5 may be associated with, the background type 283 is an indicator of the type of background
6 the audio file may be associated with, and the sponsor indicator 284 is an indicator of a
7 sponsor the audio file may be associated with (e.g., for advertising or other purposes).

8 [0038] In some embodiments, the activity type 281 may be, for example, a running activity, a
9 walking activity, a climbing activity, a construction activity, a sports related activity, a
10 relationship activity, or any other type of activity capable of being recorded and recognized
11 by audiovisual content generation unit 210. In some embodiments, the mood type 282 may
12 be, for example, a bad mood type, a happy mood type, a sad mood type, an upset mood type,
13 a lonely mood type, a loving mood type, or any other type of mood capable of being
14 recognized by audiovisual content generation unit 210. In some embodiments, the
15 background type 283 may be, for example, a construction background type, a mountainous
16 background type, an athletic event background type, a bedroom background type, a living
17 room background type, a kitchen background type, or any other background type capable of
18 being recognized by audiovisual content generation unit 210. In some embodiments, the
19 sponsor indicator 284 may be, for example, a corporation name, an institution name, a name
20 of a product, a musician, or any other sponsor or name of a sponsor that may request the
21 audio file be used to sponsor or promote a product, music, or service in the audiovisual
22 content generated by audiovisual content generation unit 210.

23 [0039] In some embodiments, the headset 100 is configured to record a video stream and an
24 accompanying audio stream using video camera 180 and microphone 190 (or receive the
25 video stream and accompanying audio stream from an external electronic device). In some
26 embodiments, the audiovisual content generation unit 210 of the headset 100 receives the
27 video stream and audio stream and utilizes the image processor 244 and/or audio processor
28 242 to determine the background of the video content and the activity and mood of a target in
29 the received video stream and audio stream. In some embodiments, the audiovisual content
30 generation unit 210 utilizes the activity, mood, and background discerned from the video
31 stream and audio stream to select an audio file from the audio library 251 that maps to an
32 activity type (e.g., activity type 281), a mood type (e.g., mood type 282), a background type
33 (e.g., background type 283), or a sponsor indicator (sponsor indicator 284) associated with an
34 audio file (e.g., audio file 279) in audio files 278. In some embodiments, the selected audio

1 file is incorporated into audiovisual content that is presented to a user for audiovisual
2 consumption (e.g., viewed on display 105 and listened to using speaker 120 of headset 100).
3 In some embodiments, the incorporation of the selected audio file into the audiovisual content
4 is an improvement over other video editing and audio editing techniques in that no input from
5 the user of the headset 100 is required to edit the audio incorporated into the audiovisual
6 content. The generation of the audiovisual content is described further in detail with reference
7 to FIG. 3 and FIG. 4.

8 [0040] FIG. 3 illustrates an audiovisual content generation unit 210 in accordance with some
9 embodiments. In some embodiments, the audiovisual content generation unit 210 includes a
10 content detection unit 380, an audio configuration unit 340 and an audiovisual configuration
11 unit 360. In some embodiments, the content detection unit 380 includes a video content
12 detection unit 370 and an audio content detection unit 375. In some embodiments, the audio
13 configuration unit 340 includes an audio library analysis unit 350, and a sponsorship
14 determination unit 352. In some embodiments, the audiovisual configuration unit 360
15 includes an audio-to-video synchronization unit 362. In some embodiments, the content
16 detection unit 380, the audio configuration unit 340 and the audiovisual configuration unit
17 360 are collectively configured to select and incorporate audio files, such as, for example,
18 music, advertisements, and other audio effects, into audiovisual content presented to display
19 using headset 100.

20 [0041] In some embodiments, the content detection unit 380 is software configured to
21 determine a background type of video content of a video stream 321 and a mood type and
22 activity type of a target or targets of the video stream 321 and an audio stream 322 provided
23 to audiovisual content generation unit 210. In some embodiments, a target may be, for
24 example, an individual or individuals that are the focus of the audiovisual content 397. In
25 some embodiments, as stated previously, the content detection unit 380 includes the video
26 content detection unit 370 and the audio content detection unit 375. In some embodiments,
27 the video content detection unit 370 is software configured to identify a mood type 323 of the
28 target in video content represented by a video stream 321, an activity type 324 of the target of
29 the video content represented by the video stream 321, and a background type 335 of the
30 video content represented by the video stream 321. In some embodiments, the audio content
31 detection unit 375 is software configured to analyze an audio stream 322 and determine a
32 mood type 393 of the target in the audio content represented by the audio stream 322 and an
33 activity type 394 of the target of the audio content represented by the audio stream 322. In
34 some embodiments, the audio content detection unit 375 is software executed by audio

processor 242 and the video content detection unit 370 is software executed by image processor 244. In some embodiments, content detection unit 380 is configured to select a mood type 333 and an activity type 334 from the mood type 323, the activity type 324, the mood type 323 and the activity type 324 output by the video content detection unit 370 and the audio content detection unit 375. In some embodiments, the content detection unit 380 selects the mood type 333 and the activity type 334 based upon the probability that the selected mood type 333 and the selected activity type 334 represent the accurate mood type and activity type of the target of the video content.

[0042] In some embodiments, in operation, content detection unit 380 receives video stream 321 and audio stream 322 (e.g., audiovisual content 397) recorded by, for example, video camera 180 and microphone 190 and commences the process of analyzing the video content of the video stream 321 and the audio content of audio stream 322 to determine which audio file of audio files 278 to incorporate into the audiovisual content 397. In some embodiments, the video stream 321 and audio stream 322 may be received from, for example, server 220, memory 215, or directly from video camera 180 and microphone 190. In some embodiments, video content detection unit 370 analyzes the video stream 321 and determines an activity type 324 of the target of the video content represented by the video stream 321, a mood type 323 of the target in the video content represented by the video stream 321, and the background type 335 of the video content represented by the video stream 321.

[0043] In some embodiments, the audio content detection unit 375 analyzes the audio stream 322 and determines a mood type 393 of the target in the audio content represented in the audio stream 322 and an activity type 394 of the target of the audio content represented in the audio stream 322. In some embodiments, after determining the mood type 323, activity type 324, mood type 393, and activity type 394, content detection unit 380 selects from the mood type 323 and the activity type 324 and the mood type 323 and the activity type 324, the mood type and activity type as mood type 333 and activity type 334. In some embodiment, after content detection unit 380 determines the activity type 334 and the mood type 333 of the target of the audiovisual content 397 and the background type 335 of the video content, the content detection unit 380 provides the activity type 334, the mood type 333, and the background type 335 to audio configuration unit 340.

[0044] In some embodiments, audio configuration unit 340 receives the activity type 334, the mood type 333, and the background type 335 from content detection unit 380 and commences the process of analyzing the activity type 334, the mood type 333, and the background type 335 in order for audiovisual content generation unit 210 to incorporate

1 audio, such as audio file 279, into the audiovisual content 397. In some embodiments, audio
2 library analysis unit 350 of audio configuration unit 340 receives the mood type 333, the
3 activity type 334, and the background type 335 and analyzes audio files 278 in audio library
4 251 to ascertain an audio file that maps to the mood type 333, the activity type 334, and the
5 background type 335 provided by content detection unit 380.

6 [0045] In some embodiments, for example, when content detection unit 380 determines that
7 the activity type 334 is a running activity type, the mood type 333 is a happy mood type, and
8 the background type 335 is a gymnasium background type, the audio library analysis unit 350
9 selects an audio file 279 that maps to the running activity type, the happy mood type, and the
10 gymnasium background type from the audio library 251. For example, audio library analysis
11 unit 350 conducts an analysis of audio library 251 and determines that an audio file 279
12 titled, “Gonna Fly Now” from the motion picture Rocky has an activity type, a mood type,
13 and a background type that is equivalent to the mood type 333, the activity type 334, and/or
14 the background type 335 provided by content detection unit 380. In some embodiments, the
15 audio library analysis unit 350 selects the audio file 279 that maps to the running activity
16 type, the happy mood type, and the gymnasium background type from the audio library 251.
17 In some embodiments, audio configuration unit 340 provides the selected audio file 342 (e.g.,
18 audio file 279) to audiovisual configuration unit 360.

19 [0046] In some embodiments, audiovisual configuration unit 360 receives the selected audio
20 file 342 and synchronizes the selected audio file 342 to video stream 321 using audio-to-
21 video synchronization unit 362. In some embodiments, audio-to-video synchronization unit
22 362 is software configured to synchronize the selected audio file 342 with the video stream 321
23 such that the audio representing a particular mood or activity plays in the background of the
24 video content of video stream 321. In some embodiments, for example, in order to align a
25 detected activity, mood, and/or background of the video stream 321 with the relevant activity,
26 mood, and/or background of the selected audio file 342, a video stream timing marker 327 and
27 audio stream timing marker 328 are placed by, for example, content detection unit 380, at the
28 location of the detected activity, mood, sponsor, and/or background in the video stream 321 and
29 audio stream 322. Likewise, in some embodiments, an audio file marker 285 is placed in the
30 audio files 278 (e.g., audio file 279) at the location of an activity, sponsor, mood, and/or
31 background, of the selected audio file 342 (e.g., audio file 279). In some embodiments, in order
32 to place the audio file marker 285 in the audio files 278, audiovisual content generation unit
33 210 scans the audio files 278 and determines the locations in the audio files 278 where relevant
34 moods, activities, and/or a product of a sponsor are present. In some embodiments, the video

1 stream timing marker 327 is a marker placed in the video stream 321 that is indicative of the
2 timing and type of activity and/or mood present at the location of the video stream 321. In some
3 embodiments, the audio stream timing marker 328 is a marker placed in the audio stream 322
4 that is indicative of the timing and type of activity and/or mood present at the location of the
5 audio stream 322.

6 [0047] In some embodiments, with further reference to FIG. 3, after selecting the audio file
7 from audio files 278 (e.g., selected audio file 342) that maps to the video stream 321, audio-to-
8 video synchronization unit 362 aligns the video stream timing marker 327 with the matching
9 audio file marker 285 of the selected audio file 342 such that the corresponding selected audio
10 file 342 is played at the video streaming time marker 327 location of the video stream 321.

11 Similarly, in some embodiments, after selecting the audio file from audio files 278 that maps to
12 the audio stream 322, audio-to-video synchronization unit 362 aligns the audio stream timing
13 marker 328 with a matching audio file marker 285 of the selected audio file 342 such that the
14 corresponding selected audio file 342 is played at the audio stream timing marker 328 location
15 of the audio stream 322.

16 [0048] In some embodiments, after synchronizing the video stream timing marker 327 with
17 the matching audio file marker 285 of the selected audio file 342 or the audio stream time
18 marker 328 with the matching audio file marker 285, audiovisual configuration unit 360
19 provides the synchronized selected audiovisual content 353 (e.g., selected audio file 342 and
20 video stream 321) for audiovisual consumption in headset 100. In some embodiments, the
21 synchronized audiovisual content 353 may be provided to other electronic devices, such as a
22 personal computer, tablet, or the like, for view in other displays.

23 [0049] FIG. 4 is a process flow diagram illustrating a method 400 for incorporating audio
24 into audiovisual content in accordance with some embodiments. In some embodiments,
25 method 400 may be performed using processing logic that includes hardware, software, or a
26 combination thereof. In some embodiments, the hardware may include, for example,
27 programmable logic, decision making logic, dedicated logic, application-specific integrated
28 circuits (ASIC), etc. In some embodiments, the processing logic refers to one or more
29 elements of an electronic device, such as, for example, headset 100 or server 220 of FIG. 1
30 and FIG. 2. In some embodiments, operations of method 400 may be implemented in an order
31 different than described herein or shown in FIG. 4. In some embodiments, method 400 may
32 have additional or less operations not shown herein.

33 [0050] In some embodiments, at block 410, content detection unit 380 receives a video
34 stream 321 recorded using a video camera (e.g., video camera 180) and an audio stream 322

1 recorded using a microphone (e.g., microphone 190). In some embodiments, content
2 detection unit 380 may receive the video stream 321 and corresponding audio stream 322
3 from server 220, memory 215, or a database.

4 [0051] In some embodiments, at block 420, the content detection unit 380 analyzes the video
5 stream 321 and audio stream 322 and determines the activity type 334 of the target of the
6 video content represented by the video stream 321 and audio stream 322, the mood type 333
7 of the target in the video content represented by the video stream 321 and audio stream 322,
8 and the background type 335 of the video content represented by the video stream 321. In
9 some embodiments, content detection unit 380 also determines the video stream timing
10 marker 327 (indicative of the location of the activity and mood in the video stream 321) and the
11 audio stream timing marker 328 (indicative of the location of the activity and mood in the audio
12 stream 322). In some embodiments, as stated previously, content detection unit 380 uses
13 image processing provided by image processor 244 to detect the activity type, the mood type,
14 and the background type of the video stream 321, and to generate the video stream timing
15 marker 327. In some embodiments, content detection unit 380 uses audio processing provided
16 by audio processor 242 to detect the activity type, the mood type, and the background type of
17 the audio stream 322, and to generate the audio stream timing marker 328.

18 [0052] In some embodiments, at block 430, based on the detected the mood type 333, the
19 activity type 334, and the background type 335, audio configuration unit 340 analyzes audio
20 files 278 in audio library 251 to determine the audio file to incorporate into the audiovisual
21 content 397 represented by video stream 321 and audio stream 322.

22 [0053] In some embodiments, at block 440, audio library analysis unit 350 selects an audio
23 file from audio files 278 as the selected audio file 342 that maps to the mood type 333, the
24 activity type 334, and the background type 335 detected by content detection unit 380. In
25 some embodiments, the selected audio file 342 is provided to audiovisual configuration unit
26 360.

27 [0054] In some embodiments, at block 450, the audiovisual configuration unit 360
28 synchronizes the selected audio file 342 to the video content represented by video stream
29 321. In some embodiments, at block 460, the synchronized selected audio file 342 (e.g., audio
30 file 279) and video stream 321 are presented as audiovisual content 353 to the display 105
31 and speaker 120 for view and audio listening pleasure by the wearer of the headset 100.

32 [0055] In one or more embodiments, a user may record video and accompanying audio of a
33 target, such as a person or animal, running happily in a park and provide the video and
34 accompanying audio as audiovisual content 397 to audiovisual content generation unit 210.

1 In some embodiments, content detection unit 380 of audiovisual content generation unit 210
2 analyzes the audiovisual content 397 and determines that the activity type 334 being
3 performed by the target is a running activity, the mood type 333 of the target is a happy
4 mood, and that the background type 335 of the audiovisual content 397 is a park background.
5 In some embodiments, content detection unit 380 provides the activity type 334 (e.g., running
6 activity), the mood type 333 (e.g., happy mood), and the background type 335 (e.g., park
7 background) to audio configuration unit 340. In some embodiments, the audio library
8 analysis unit 350 of audio configuration unit 340 receives the activity type 334, the mood
9 type 333 and the background type 335 and analyzes the audio files 278 in the audio library
10 251 to ascertain an audio file or audio files that map to the activity type 334, the mood type
11 333 and the background type 335. In some embodiments, audio library analysis unit 350
12 analyzes the audio files 278 in the audio library 251 by comparing the activity type 334 (e.g.,
13 running activity), the mood type 333 (e.g., happy mood), and the background type 335 with
14 the activity type (e.g., activity type 281), the mood type (e.g., mood type 282), and the
15 background type (e.g., background type 283) associated with each audio file of audio files
16 278. In some embodiments, based upon the analysis, audio library analysis unit 350 selects a
17 plurality of audio files from audio library 251 that map to the running activity, the happy
18 mood, and the park background. In some embodiments, the audio configuration unit 340
19 provides the titles of the plurality of audio files to display 105 for view by the user of headset
20 100, to allow the user to select the desired audio file to play in the background of the video
21 content of video stream 321 (audiovisual content 397). In some embodiments, the user views
22 the titles of the audio files presented on display 105 and selects the desired audio file, such as,
23 for example, an ironic clown car song, using an audio command via microphone 190 or other
24 input selection medium (such as, e.g., a hand gesture in a virtual or augmented reality
25 system). In some embodiments, the selected audio file 342 is provided to audiovisual
26 configuration unit 360, which synchronizes the selected audio file 342 with the video content
27 of video stream 321 to generate audiovisual content 353.

28 [0056] In one or more embodiments, a user may record video and accompanying audio of a
29 target hammering a nail into a bedroom wall, hitting their thumb, and screaming in pain. In
30 some embodiments, the user provides the video as video stream 321 and accompanying audio
31 as audio stream 322 (e.g., audiovisual content 397) to audiovisual content generation unit
32 210. In some embodiments, content detection unit 380 of audiovisual content generation unit
33 210 analyzes the audiovisual content 397 and determines that the activity type 334 being
34 performed by the target is a carpentering activity, the mood type 333 of the target is an upset

1 mood (e.g., from the audio stream 322), and the background type 335 of the video is a
2 bedroom background. In some embodiments, content detection unit 380 provides the activity
3 type 334 (e.g., carpentering activity), the mood type 333 (e.g., upset mood), and the
4 background type 335 (e.g., bedroom background) to audio configuration unit 340. In some
5 embodiments, the audio library analysis unit 350 of audio configuration unit 340 receives the
6 activity type 334, the mood type 333 and the background type 335 and analyzes the audio
7 files 278 in the audio library 251. In some embodiments, the audio library analysis unit 350
8 analyzes the audio files 278 by comparing the activity type 334 (e.g., carpentering activity),
9 the mood type 333 (e.g., upset mood), and the background type 335 (e.g., bedroom
10 background) with the activity type (e.g., activity type 281), the mood type (e.g., mood type
11 282), and the background type (e.g., background type 283) associated with each audio file of
12 audio files 278. In some embodiments, based on the analysis, audio library analysis unit 350
13 selects an audio file that has the same the activity type, mood type, and/or background type,
14 such as, for example, the song “So You Had a Bad Day” by Daniel Powter. In some
15 embodiments, the selected audio file 342 is provided to audiovisual configuration unit 360,
16 which synchronizes the selected audio file 342 with the video content of video stream 321 to
17 generate audiovisual content 353. In some embodiments, the audiovisual configuration unit
18 360 synchronizes the selected audio file 342 with video content of the video stream 321
19 and/or audio content of the audio stream 322 such that the chorus of the selected audio file
20 342 starts or plays when the target performs the activity (e.g., hits a thumb), experiences the
21 mood (e.g., painful mood), or displays a sponsor (e.g., BRAND of hammer) in the video
22 content or audio content. In some embodiments, the audiovisual configuration unit 360 is
23 configured to edit or crop the music of the selected audio file 342 such that the music plays
24 only during the activity (e.g., the audio file plays when the activity starts and stops playing
25 when the activity ends). In some embodiments, the audiovisual configuration unit 360 is
26 configured to edit the audio content of audio stream 322 such that specific words or verbiage
27 may filtered or muted, such as, for example, a swear word. In some embodiments, only the
28 audio selected (e.g., selected audio file 342) is played in the background of the video content
29 of video stream 321 when a specific word is muted. In some embodiments, the edited version
30 of the audiovisual content (e.g., audiovisual content 353) is uploaded automatically to
31 memory 215 of headset 100. In some embodiments, the synchronized selected audio file 342
32 and the video stream 321 are provided as audiovisual content 353 for audiovisual
33 consumption using headset 100.
34 [0057] In one or more embodiments, a user records audiovisual content 397 (e.g., a video and

1 audio) of a target running into a wall. In some embodiments, the user selects an audio file
2 from audio library 251 to incorporate into the audiovisual content 397 that is a heavy metal
3 song that includes a swear word. In some embodiments, headset 100 determines that the
4 audiovisual content is to be viewed on display 105 by a child (due to, for example, a swear
5 word filter being implemented on the headset 100). In some embodiments, based the audio
6 configuration unit 340 recognizing that a swear word filter has been initiated, the audio
7 configuration unit 340 selects a replacement audio file with a matching background type 335,
8 activity type 334, and mood type 333, but no swear words, and synchronizes the replacement
9 audio file with the video content of video stream 321 to create audiovisual content 353. In
10 some embodiments, using the audiovisual content generation unit 210, the audio file with
11 swear words is replaced as the accompanying audio to the video stream 321 with an audio file
12 without swear words that is similar to the original audio file selected by the user (e.g., a
13 similar song selected without swear words based on the analysis by audio library analysis unit
14 350).

15 [0058] In one or more embodiments, the headset 100 is configured to distribute audiovisual
16 content 353 to other members of a social network associated with a user of the headset 100
17 via network 225. For example, in some embodiments, based on an analysis of video content
18 of video streams associated with the other members in the social network, the headset 100
19 may select an audio file 279 that maps to the activity type 334, the mood type 333, and the
20 background type 335 liked by another member or members of the social network. In some
21 embodiments, the audiovisual content 353 is provided via network 225 to the other members
22 of the social network associated with the user.

23 [0059] In one or more embodiments, instead of the audio configuration unit 340 selecting the
24 selected audio file 342, the audio configuration unit 340 may provide a plurality of audio files
25 that meet criteria for selection (e.g., the activity type 334, the mood type 333, and/or
26 background type 335), and present the plurality of audio files to the user of the headset 100
27 for selection. In some embodiments, the user selects the desired audio file from the plurality
28 of audio files based on the preference of the viewer of the audiovisual content or the creator
29 of the audiovisual content. In some embodiments, audiovisual content generation unit 210
30 provides selected audio file 342 as background audio for the audiovisual content 353.

31 [0060] In some embodiments, the audio configuration unit 340 is configured to mine the
32 selection options to improve recommendations based on, for example, a population for which
33 the audio files are selected. In some embodiments, an audio file is selected based on, for
34 example, a market of the creator of the audiovisual content or a market of the viewer of the

1 audiovisual content. In some embodiments, the audio configuration unit 340 is configured to
2 promote audio files (e.g., songs and options for purchase) based on, for examples, heuristics
3 or promotions provided by, for example, a company or other commercial entity, an owner of
4 the audio files, a creator of the audio files (e.g., musician, producer).

5 [0061] In some embodiments, the audio configuration unit 340 is configured to perform the
6 operations herein automatically, without direct input from the creator or viewer of the
7 audiovisual content.

8 [0062] In some embodiments, the audio configuration unit 340 is configured to modify or
9 change the play rate of the audiovisual content (e.g., video or song or both the video and
10 song) to synchronize the audio file 279 to the images in the video content (e.g., align the
11 music with the images in the video).

12 [0063] In some embodiments, video content detection unit 370 is configured to detect tempo
13 patterns in the video content of the video stream or the audio content detection unit 375 and
14 use the detected tempo patterns to select an audio file 279 that matches with the tempo
15 patterns of the video content or audio content. In some embodiments, for example, when a song
16 detected by content detection unit 380 in the audiovisual content 397 has a tempo of 120 beats
17 per minute, audio library analysis unit 350 may select a replacement song (having similar
18 incorporation attributes) as selected audio file 342 with a tempo of 120 beats per minute.

19 [0064] In one or more embodiments, a corporation or sponsor may request that a promotional
20 video utilize a specific audio file (e.g., a specific song) from audio files 278 to promote a
21 specific product. In some embodiments, the company may record the promotional video of a
22 target with a product wherein the target performs a specific activity with the product. In some
23 embodiments, the content detection unit 380 of audiovisual content generation unit 210 is
24 configured to detect product specific audio or video content (e.g., what the target is talking
25 about or how the target is using the product) in the received promotional video. In some
26 embodiments, audiovisual content generation unit 210 selects, recommends or automatically
27 adds an audio file of a selected song or jingle to the audiovisual content. In some
28 embodiments, the sponsorship determination unit 352 is configured to determine the sponsor
29 that is requesting sponsorship using, for example, the promotional video. In some
30 embodiments, the video and corresponding audio is automatically tagged as promotional
31 audiovisual content and uploaded to the memory 215 of headset 100 for promotional
32 purposes and view on display 105.

33 [0065] In some embodiments, a song writer or musician may request that a song written by
34 the song writer or played by the musician be used for audiovisual content that includes

1 specific activity types or background types, such as, for example, a sunny background and or
2 a beach background. In some embodiments, audio library analysis unit 350 selects a plurality
3 of audio files 279 for selection by the user of the headset 100. In some embodiments, while
4 presenting the audio files 279 that map to the video content of video stream 321, audio
5 configuration unit 340 promotes an audio file, e.g., audio file 279, associated with the
6 songwriter or musician such that the song associated with the songwriter or musician is more
7 likely to be selected by the user of the headset 100.

8 [0066] In some embodiments, a plurality of videos that include audio files from audio files
9 278 associated with a musician (via ownership or the like) may be used to create a feed
10 accessible to the musician or others via a link or a website. In some embodiments, when, for
11 example, an individual with ownership authority of audio files of audio files 278 requests that
12 audio files be played only in an enumerated set of countries, the audiovisual content
13 generation unit 210 may only play the audio files in the enumerated set of countries. In some
14 embodiments, the audiovisual content generation unit 210 assesses the location of the user of
15 the headset 100 and applies the audio files (e.g., songs) to videos when a user of the headset
16 100 is in the country or specific set of countries and opens a video with approved content based
17 on, for example, a ranking (e.g., high ranking) in an audio file ranking system.

18 [0067] In one or more embodiments, during the synchronization process, the audiovisual
19 configuration unit 360 is configured to utilize the audio-to-video synchronization unit 362 to
20 adjust the timing of the music represented by the audio file 279 based on the content of the
21 video or audio.

22 [0068] In one or more embodiments, the audio library analysis unit 350 is configured to
23 classify the audio files 278 using automatic music library tagging which may automatically
24 classify songs by genre, a user's genre preferences, a tempo of the audio file, lyrics in the
25 audio file, a mood of the audio file, and/or a popularity of the audio file. In some embodiments,
26 the audio library analysis unit 350 is configured to classify audio effects provided by the audio
27 library using a name of the audio effect or an associated search query of the audio effect. In
28 some embodiments, the audio effects may be used to generate the audiovisual content 353.

29 [0069] In one or more embodiments, the audiovisual content generation unit 210 is
30 configured to utilize a historical analysis to generate the audiovisual content 353. In some
31 embodiments, a user history identifies a user's previous selections of music or audio effect
32 recommendations and music listening preferences. In some embodiments, a network history
33 identifies a social network's previous selections of music or audio effects for similar captured
34 content. In some embodiments, the user history and/or network history is utilized to select the

1 audio file from audio files 278 and generate the audiovisual content 353.

2 [0070] In one or more embodiments, with further reference to FIG. 3, the content detection
3 unit 380 is configured to perform a captured audiovisual content analysis, which includes a
4 visual scene understanding, an acoustic scene understanding, a context understanding, and
5 error correction of the audiovisual content 397. In some embodiments, the visual scene
6 understanding of the captured audiovisual content 397 includes in-frame, e.g., activities and
7 actions of objects and people, including their classification, repetitive patterns (e.g., tempo,
8 beats), and relationships to each other, progression curve of events and activities, and place
9 and event classifications. In some embodiments, the acoustic scene understanding of the
10 captured audiovisual content 397 includes classification of events, people, places, and
11 background noise, and speech characteristics (such as, e.g., timbre, tone, intonation, and
12 mental state). In some embodiments, the context understanding of the captured audiovisual
13 content 397 includes an understanding of device permissions, such as, for example, for the
14 user's location or social context (e.g., the relationship of the audiovisual content 397 to a
15 user's contacts and events (e.g., 'Family Vacation in Mexico,' 'Dinner With Friends,' or 'My
16 Birthday party')). In some embodiments, the error correction may be utilized to compare and
17 match visual, acoustic, and context parameters to improve accuracy of the collective
18 understanding of the content.

19 [0071] Various operations described herein can be implemented on computer systems. FIG.
20 5 shows a block diagram of a representative computing system 514 usable to implement the
21 present disclosure. In some embodiments, the operations of headset 100 or the server 220
22 of FIG. 2 is implemented by the computing system 514. Computing system 514 can be
23 implemented, for example, as a consumer device such as a smartphone, other mobile phone,
24 tablet computer, wearable computing device (e.g., smart watch, eyeglasses, head mounted
25 display), desktop computer, laptop computer, or implemented with distributed computing
26 devices. The computing system 514 can be implemented to provide VR, AR, MR experience.
27 In some embodiments, the computing system 514 can include conventional computer
28 components such as processors 516, storage device 518, network interface 520, user input
29 device 522, and user output device 524.

30 [0072] Network interface 520 can provide a connection to a wide area network (e.g., the
31 Internet) to which WAN interface of a remote server system is also connected. Network
32 interface 520 can include a wired interface (e.g., Ethernet) and/or a wireless interface
33 implementing various RF data communication standards such as Wi-Fi, Bluetooth, or cellular
34 data network standards (e.g., 3G, 4G, 5G, 60 GHz, LTE, etc.).

[0073] User input device 522 can include any device (or devices) via which a user can provide signals to computing system 514; computing system 514 can interpret the signals as indicative of particular user requests or information. User input device 522 can include any or all of a keyboard, touch pad, touch screen, mouse or other pointing device, scroll wheel, click wheel, dial, button, switch, keypad, microphone, sensors (e.g., a motion sensor, an eye tracking sensor, etc.), and so on.

[0074] User output device 524 can include any device via which computing system 514 can provide information to a user. For example, user output device 524 can include a display to display images generated by or delivered to computing system 514. The display can incorporate various image generation technologies, e.g., a liquid crystal display (LCD), light-emitting diode (LED) including organic light-emitting diodes (OLED), projection system, cathode ray tube (CRT), or the like, together with supporting electronics (e.g., digital-to-analog or analog-to-digital converters, signal processors, or the like). A device such as a touchscreen that function as both input and output device can be used. Output devices 524 can be provided in addition to or instead of a display. Examples include indicator lights, speakers, tactile “display” devices, printers, and so on.

[0075] Some implementations include electronic components, such as microprocessors, storage and memory that store computer program instructions in a computer readable storage medium. Many of the features described in this specification can be implemented as processes that are specified as a set of program instructions encoded on a computer readable storage medium. When these program instructions are executed by one or more processors, they cause the processors to perform various operation indicated in the program instructions. Examples of program instructions or computer code include machine code, such as is produced by a compiler, and files including higher-level code that are executed by a computer, an electronic component, or a microprocessor using an interpreter. Through suitable programming, processor 516 can provide various functionality for computing system 514, including any of the functionality described herein as being performed by a server or client, or other functionality associated with message management services.

[0076] It will be appreciated that computing system 514 is illustrative and that variations and modifications are possible. Computer systems used in connection with the present disclosure can have other capabilities not specifically described here. Further, while computing system 514 is described with reference to particular blocks, it is to be understood that these blocks are defined for convenience of description and are not intended to imply a particular physical arrangement of component parts. For instance, different blocks can be located in the

1 same facility, in the same server rack, or on the same motherboard. Further, the blocks need
2 not correspond to physically distinct components. Blocks can be configured to perform
3 various operations, e.g., by programming a processor or providing appropriate control
4 circuitry, and various blocks might or might not be reconfigurable depending on how the
5 initial configuration is obtained. Implementations of the present disclosure can be realized in
6 a variety of apparatus including electronic devices implemented using any combination of
7 circuitry and software.

8 [0077] Having now described some illustrative implementations, it is apparent that the
9 foregoing is illustrative and not limiting, having been presented by way of example. In
10 particular, although many of the examples presented herein involve specific combinations of
11 method acts or system elements, those acts and those elements can be combined in other
12 ways to accomplish the same objectives. Acts, elements and features discussed in connection
13 with one implementation are not intended to be excluded from a similar role in other
14 implementations or implementations.

15 [0078] The hardware and data processing components used to implement the various
16 processes, operations, illustrative logics, logical blocks, modules and circuits described in
17 connection with the embodiments disclosed herein may be implemented or performed with a
18 general purpose single- or multi-chip processor, a digital signal processor (DSP), an
19 application specific integrated circuit (ASIC), a field programmable gate array (FPGA), or
20 other programmable logic device, discrete gate or transistor logic, discrete hardware
21 components, or any combination thereof designed to perform the functions described herein.
22 A general purpose processor may be a microprocessor, or, any conventional processor,
23 controller, microcontroller, or state machine. A processor also may be implemented as a
24 combination of computing devices, such as a combination of a DSP and a microprocessor, a
25 plurality of microprocessors, one or more microprocessors in conjunction with a DSP core, or
26 any other such configuration. In some embodiments, particular processes and methods may
27 be performed by circuitry that is specific to a given function. The memory (e.g., memory,
28 memory unit, storage device, etc.) may include one or more devices (e.g., RAM, ROM, Flash
29 memory, hard disk storage, etc.) for storing data and/or computer code for completing or
30 facilitating the various processes, layers and modules described in the present disclosure. The
31 memory may be or include volatile memory or non-volatile memory, and may include
32 database components, object code components, script components, or any other type of
33 information structure for supporting the various activities and information structures
34 described in the present disclosure. According to an exemplary embodiment, the memory is

1 communicably connected to the processor via a processing circuit and includes computer
2 code for executing (e.g., by the processing circuit and/or the processor) the one or more
3 processes described herein.

4 [0079] The present disclosure contemplates methods, systems and program products on any
5 machine-readable media for accomplishing various operations. The embodiments of the
6 present disclosure may be implemented using existing computer processors, or by a special
7 purpose computer processor for an appropriate system, incorporated for this or another
8 purpose, or by a hardwired system. Embodiments within the scope of the present disclosure
9 include program products comprising machine-readable media for carrying or having
10 machine-executable instructions or data structures stored thereon. Such machine-readable
11 media can be any available media that can be accessed by a general purpose or special
12 purpose computer or other machine with a processor. By way of example, such machine-
13 readable media can comprise RAM, ROM, EPROM, EEPROM, or other optical disk storage,
14 magnetic disk storage or other magnetic storage devices, or any other medium which can be
15 used to carry or store desired program code in the form of machine-executable instructions or
16 data structures and which can be accessed by a general purpose or special purpose computer
17 or other machine with a processor. Combinations of the above are also included within the
18 scope of machine-readable media. Machine-executable instructions include, for example,
19 instructions and data which cause a general purpose computer, special purpose computer, or
20 special purpose processing machines to perform a certain function or group of functions.

21 [0080] The phraseology and terminology used herein is for the purpose of description and
22 should not be regarded as limiting. The use of “including” “comprising” “having”
23 “containing” “involving” “characterized by” “characterized in that” and variations thereof
24 herein, is meant to encompass the items listed thereafter, equivalents thereof, and additional
25 items, as well as alternate implementations consisting of the items listed thereafter
26 exclusively. In one implementation, the systems and methods described herein consist of one,
27 each combination of more than one, or all of the described elements, acts, or components.

28 [0081] Any references to implementations or elements or acts of the systems and methods
29 herein referred to in the singular can also embrace implementations including a plurality of
30 these elements, and any references in plural to any implementation or element or act herein
31 can also embrace implementations including only a single element. References in the singular
32 or plural form are not intended to limit the presently disclosed systems or methods, their
33 components, acts, or elements to single or plural configurations. References to any act or
34 element being based on any information, act or element can include implementations where

1 the act or element is based at least in part on any information, act, or element.
2 [0082] Any implementation disclosed herein can be combined with any other implementation
3 or embodiment, and references to “an implementation,” “some implementations,” “one
4 implementation” or the like are not necessarily mutually exclusive and are intended to
5 indicate that a particular feature, structure, or characteristic described in connection with the
6 implementation can be included in at least one implementation or embodiment. Such terms as
7 used herein are not necessarily all referring to the same implementation. Any implementation
8 can be combined with any other implementation, inclusively or exclusively, in any manner
9 consistent with the aspects and implementations disclosed herein.

1 WHAT IS CLAIMED IS:

- 2 1. A method, comprising:
 - 3 detecting an incorporation attribute of audiovisual content;
 - 4 analyzing an audio library to determine an audio file that maps to the incorporation
 - 5 attribute;
 - 6 selecting the audio file from the audio library that maps to the incorporation attribute;
 - 7 incorporating the selected audio file into the audiovisual content to generate
 - 8 egocentric audiovisual content; and
 - 9 providing the egocentric audiovisual content to a user for audiovisual consumption.
- 10 2. The method of claim 1, wherein:
 - 11 the incorporation attribute is at least one of a mood of a target of the audiovisual
 - 12 content, an activity of the target of the audiovisual content, and a background
 - 13 of the audiovisual content.
- 14 3. The method of claim 1 or 2, wherein:
 - 15 the audiovisual content includes video content of a video stream and audio content of
 - 16 an audio stream.
- 17 4. The method of any one of the preceding claims, further comprising:
 - 18 synchronizing the selected audio file to the audiovisual content based upon an audio
 - 19 file marker placed in the audio file.
- 20 5. The method of any one of the preceding claims, further comprising:
 - 21 mapping audio files in the audio library to at least one of an activity type, a mood
 - 22 type, a background type, and a sponsorship type; preferably wherein:
 - 23 an audio configuration unit selects the audio file based upon the mapping of the audio
 - 24 file to the incorporation attribute; and preferably wherein:
 - 25 an audio library analysis unit analyzes the audio library to determine which audio file
 - 26 in the audio library maps to the incorporation attribute of the audiovisual
 - 27 content.
- 28 6. The method of any one of the preceding claims, wherein:
 - 29 synchronizing the audio file to a video stream of the audiovisual content is dependent
 - 30 upon a video stream timing marker being placed in the video stream of the
 - 31 audiovisual content; and/or wherein:
 - 32 synchronizing the audio file to the video stream is dependent upon an audio stream
 - 33 time marker being placed in the audio stream.
- 34 7. A system, comprising:

- 1 a video recorder;
2 an audio recorder coupled to the video recorder; and
3 a controller coupled to the video recorder and the audio recorder, wherein the
4 controller is configured to:
5 detect an incorporation attribute of audiovisual content;
6 analyze an audio library to determine an audio file that maps to the
7 incorporation attribute;
8 select the audio file from the audio library that maps to the incorporation
9 attribute;
10 incorporate the selected audio file into the audiovisual content to generate
11 egocentric audiovisual content; and
12 provide the egocentric audiovisual content to a user for audiovisual
13 consumption.
- 14 8. The system of claim 7, wherein:
15 the incorporation attribute is at least one of a mood of a target of the audiovisual
16 content, an activity of the target of the audiovisual content, and a background
17 of the audiovisual content.
- 18 9. The system of claim 7 or 8, wherein:
19 the audiovisual content includes video content of a video stream and audio content of
20 an audio stream.
- 21 10. The system of any one of claims 7 to 9, wherein:
22 the selected audio file is synchronized to the audiovisual content based upon an audio
23 file marker placed in the audio file.
- 24 11. The system of any one of claims 7 to 10, wherein:
25 audio files in the audio library are mapped to at least one of an activity type, a mood
26 type, a background type, and a sponsorship type; preferably wherein:
27 an audio configuration unit selects the audio file based upon the mapping of the audio
28 file to the incorporation attribute; and preferably wherein:
29 an audio library analysis unit analyzes the audio library to determine which audio file
30 in the audio library maps to the incorporation attribute of the audiovisual
31 content.
- 32 12. The system of any one of claims 7 to 11, wherein:
33 the audio file is synchronized to a video stream of the audiovisual content using a
34 video stream timing marker that is placed in the video stream of the

1 audiovisual content.

2 13. The system of claim 12, wherein:

3 the audio file that is synchronized to a video stream of the audiovisual content using
4 an audio stream time marker that is placed in the audio stream of the
5 audiovisual content.

6 14. A non-transitory computer-readable medium storing instructions that, when executed
7 by one or more processors, cause the one or more processors to perform operations
8 comprising:

9 detecting an incorporation attribute of audiovisual content;
10 analyzing an audio library to determine a first audio file that maps to the incorporation
11 attribute of the audiovisual content;
12 selecting the first audio file from the audio library that maps to the incorporation
13 attribute;
14 incorporating the selected first audio file into the audiovisual content to generate
15 egocentric audiovisual content;
16 uploading the egocentric audiovisual content to a social network for view by a viewer
17 of the social network; and
18 modifying the egocentric audiovisual content to include a second audio file selected
19 using preferences of the viewer of the egocentric audiovisual content.

20 15. The non-transitory computer-readable medium of claim 14, wherein:

21 the second audio file selected by the viewer of the egocentric audiovisual content
22 replaces the first audio file selected from the audio library that maps to the
23 incorporation attribute.

24

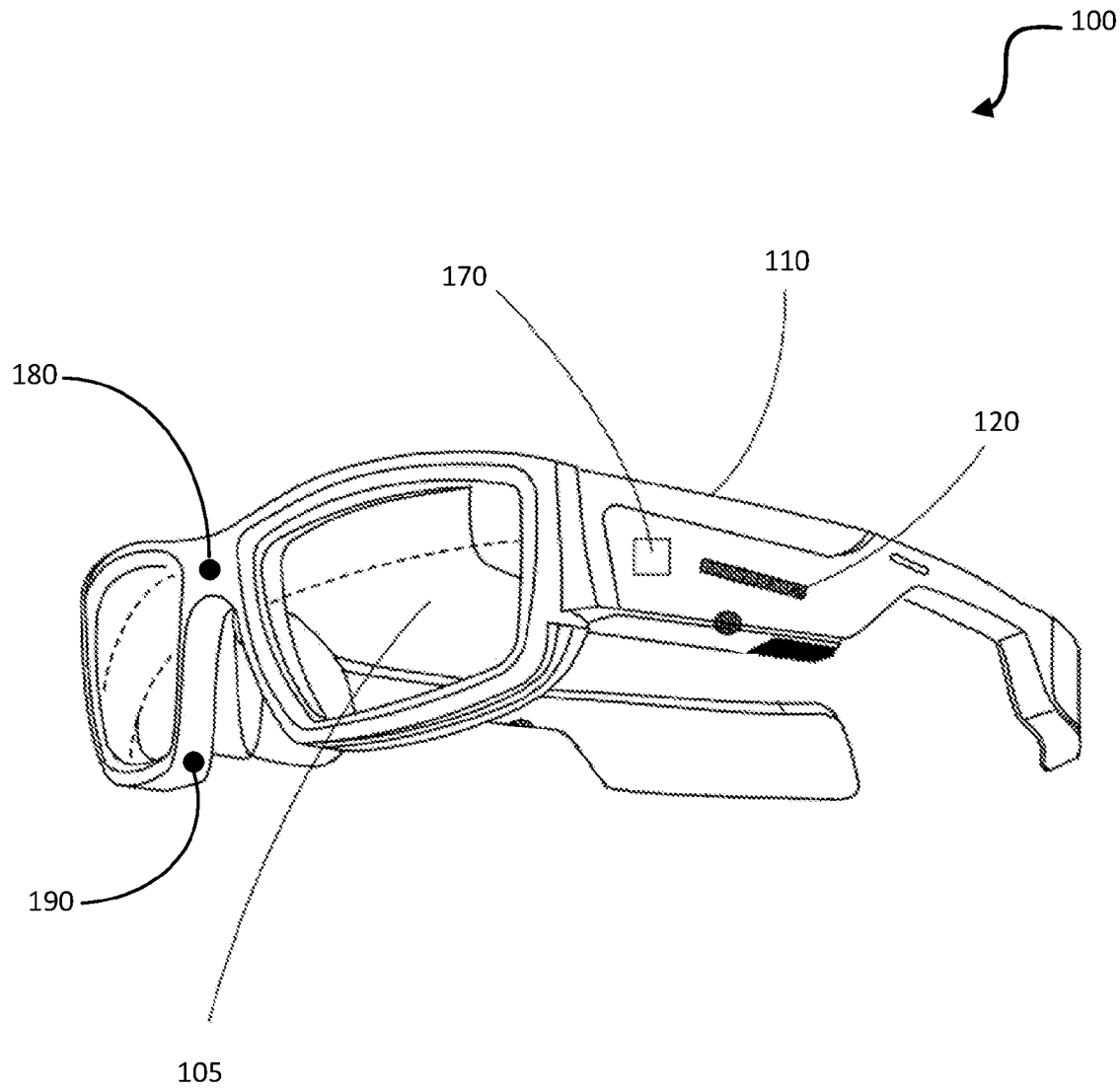
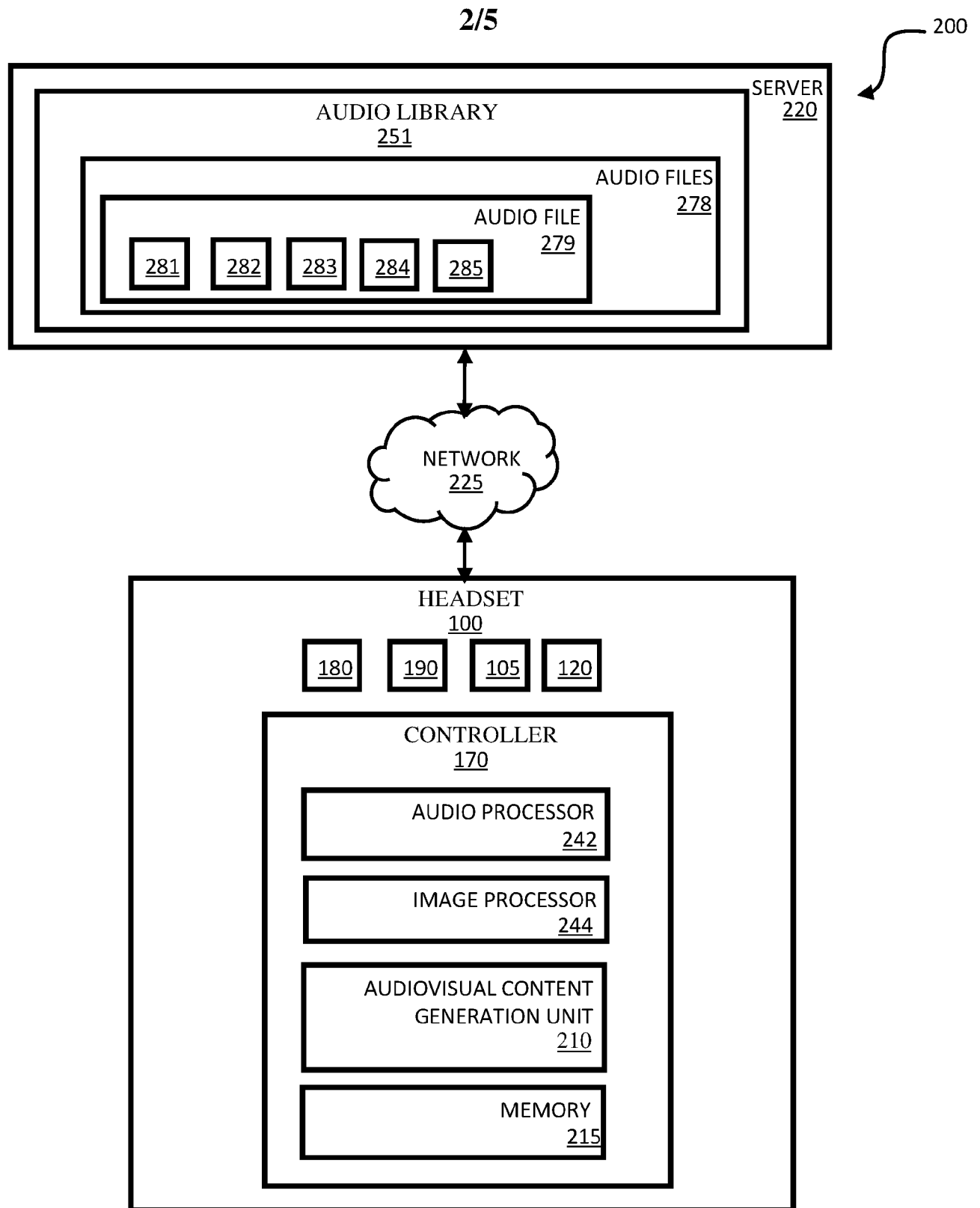


FIG. 1

**FIG. 2**

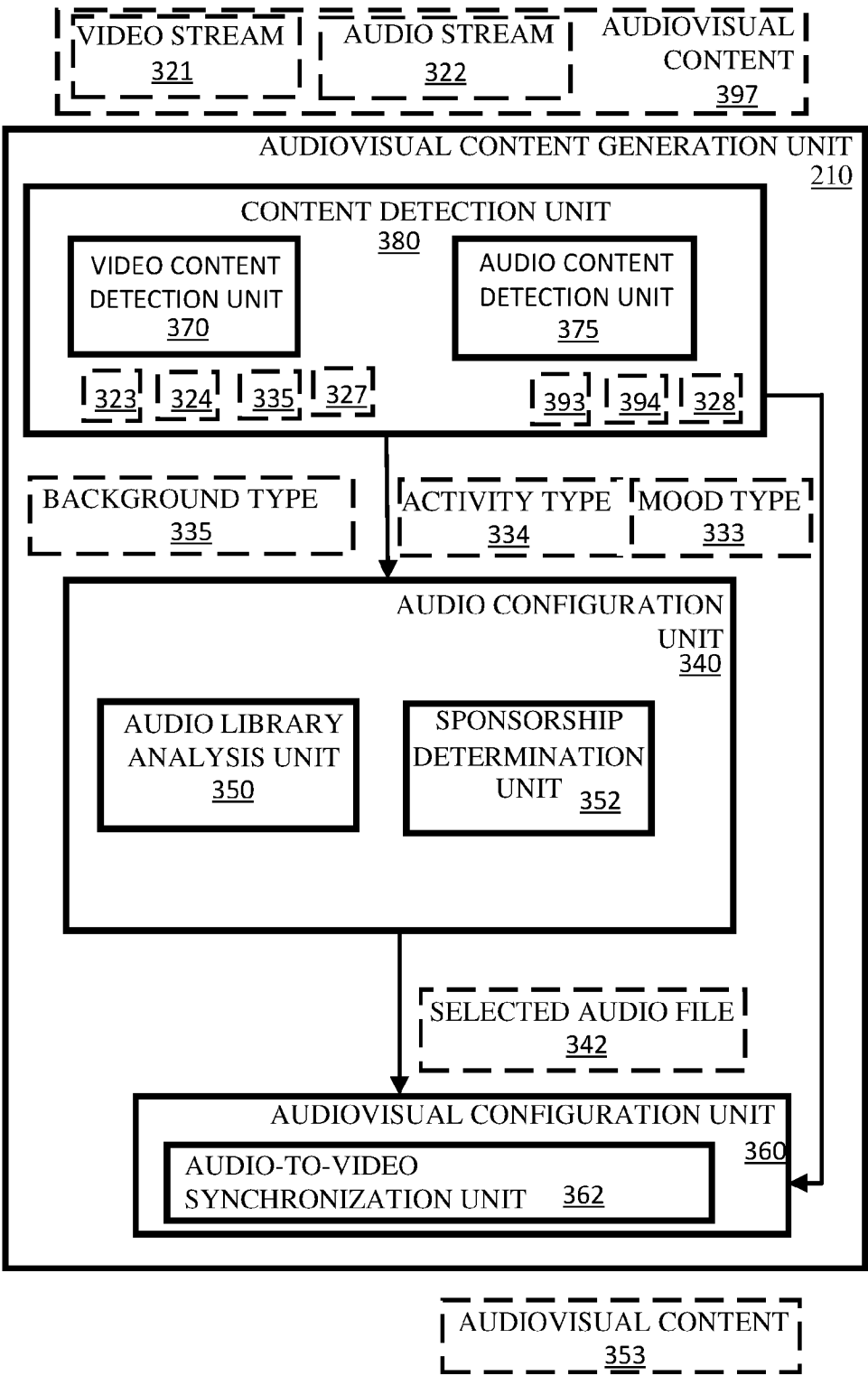
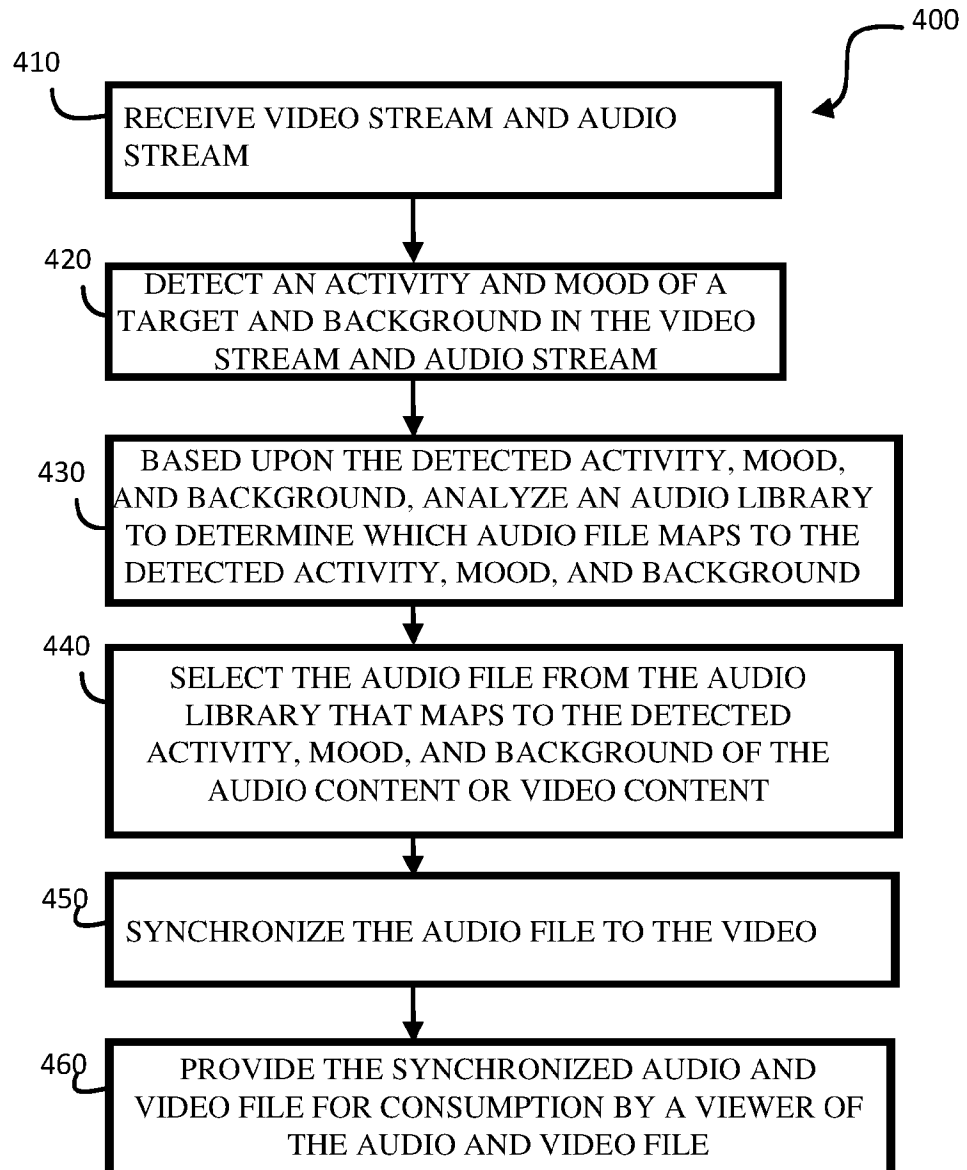


FIG. 3

4/5

**FIG. 4**

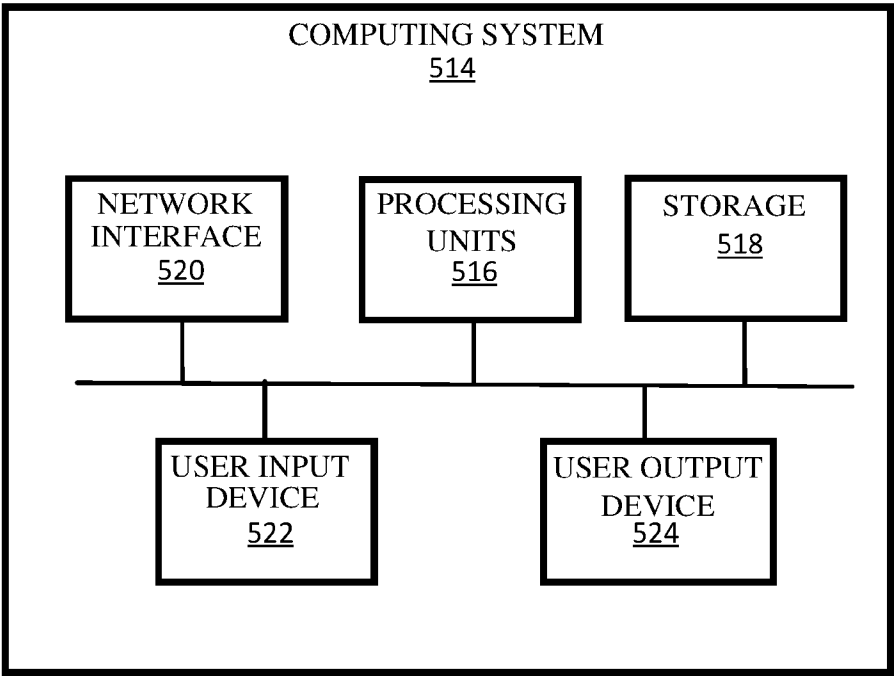


FIG. 5

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2023/022164

A. CLASSIFICATION OF SUBJECT MATTER INV. G11B27/031 G11B27/10 G06F16/68 ADD.		
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) G06F G11B		
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) EPO-Internal, WPI Data		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	RAJIV RATN SHAH ET AL: "ADVISOR", MULTIMEDIA, ACM, 2 PENN PLAZA, SUITE 701 NEW YORK NY 10121-0701 USA, 3 November 2014 (2014-11-03), pages 607-616, XP058058684, DOI: 10.1145/2647868.2654919 ISBN: 978-1-4503-3063-3	1-3, 5, 7-9, 11, 14, 15
A	the whole document	4, 6, 10, 12, 13
X	----- WO 2022/005442 A1 (PONOCHEVNYI NAZAR YURIEVYCH [UA]) 6 January 2022 (2022-01-06)	1-10, 12, 13
A	abstract	14, 15
	-/--	
☒ 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 "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		Date of mailing of the international search report
19 July 2023		27/07/2023
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016		Authorized officer Maetz, Arnaud

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2023/022164

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P A,P	& EP 4 178 206 A1 (HARMIX INC [US]) 10 May 2023 (2023-05-10) abstract paragraphs [0045], [0049], [0055]; figures 2,9 -----	1-10,12, 13 14,15
X	US 2006/122842 A1 (HERBERGER TILMAN [DE] ET AL) 8 June 2006 (2006-06-08) paragraphs [0011], [0014], [0041] - [0045], [0057] - [0058], [0067], [0068], [0072], [0077] -----	1-15

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2023/022164

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2022005442 A1	06-01-2022	CA 3184814 A1	06-01-2022
		EP 4178206 A1	10-05-2023
		WO 2022005442 A1	06-01-2022
US 2006122842 A1	08-06-2006	EP 1666967 A1	07-06-2006
		US 2006122842 A1	08-06-2006