



(51) International Patent Classification:

H04N 21/2187 (2011.01) H04N 21/2665 (2011.01)
H04N 21/234 (2011.01) H04N 21/437 (2011.01)
H04N 21/236 (2011.01) H04N 21/442 (2011.01)
H04N 21/24 (2011.01) H04N 21/658 (2011.01)
H04N 21/242 (2011.01) H04N 5/222 (2006.01)

(21) International Application Number:

PCT/US2021/043246

(22) International Filing Date:

26 July 2021 (26.07.2021)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

63/057,184 27 July 2020 (27.07.2020) US

(71) Applicant: DREAMSTAGE, INC. [US/US]; 88 Station Hill Road, Barrytown, New York 12507 (US).

(72) Inventor: CHASIN, Scott; 44 South Fairfax Street, Denver, Colorado 80220 (US).

(74) Agent: FREYER, Andrew J. et al.; 410 Seventeenth Street, Suite 2200, Denver, Colorado 80202 (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, 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))

(54) Title: SYSTEM AND METHOD FOR AGGREGATING AUDIOVISUAL CONTENT

(57) Abstract: Systems and methods described herein may operate a host device to convey a live performance to an audience member's electronic device and to provide performers an immersive experience that simulates a live performance in front of an audience. The host server may down-sample and/or modify real-time video from users and overlay the modified video feeds in an audience scene to mimic an audience. This audience video feed may be displayed to the performer at a host venue in real time. The performers may interact with the audience members and may receive audience feedback in real time. In some embodiments, the host server may aggregate or generate audience audio feeds to play-back to the performers. Additionally, the host venue may capture audio and video from the performers and send this data to the host server.

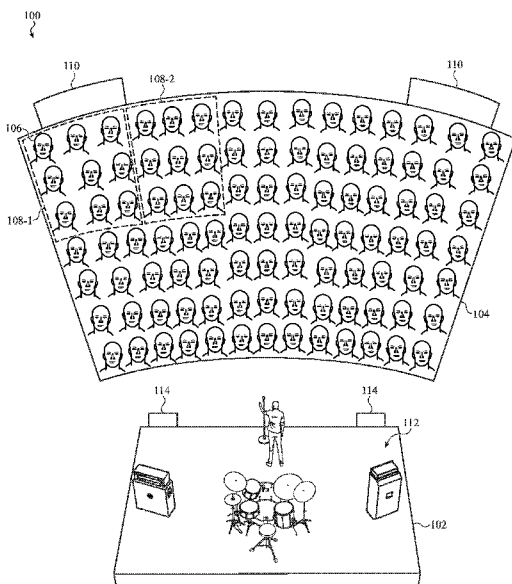


FIG. 1



SYSTEM AND METHOD FOR AGGREGATING AUDIOVISUAL CONTENT

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This Patent Cooperation Treaty patent application claims priority to U.S.
5 Provisional Patent Application No. 63/057,184, filed July 27, 2020, and titled "System and
Method for Aggregating Audiovisual Content," the contents of which are incorporated herein
by reference in their entirety.

TECHNICAL FIELD

[0002] The subject matter of this disclosure relates to systems and methods for remote
10 viewing of, and interaction with, live performances.

BACKGROUND

[0003] Live performances and presentations are unique experiences for both an
audience and performers or presenters. For an audience member, a live performance
provides a unique and memorable experience. For a performer/presenter, experiencing an
15 audience reacting to a performance is both validating and exciting. A conventional substitute
for a live performance (or presentation) is a live stream.

[0004] However, viewing a conventional live stream does not provide either (1) an
audience member with a comparable experience to a live performance, or (2) a performer
with a comparable experience of performing before a live audience.

BRIEF DESCRIPTION OF THE DRAWINGS

[0005] Reference will now be made to representative embodiments illustrated in the
accompanying figures. It should be understood that the following descriptions are not
intended to limit this disclosure to one included embodiment. To the contrary, the disclosure
provided herein is intended to cover alternatives, modifications, and equivalents as may be
25 included within the spirit and scope of the described embodiments, and as defined by the
appended claims.

[0006] The disclosure will be readily understood by the following detailed description in
conjunction with the accompanying drawings, wherein like reference numerals designate like
structural elements, and in which:

30 [0007] FIG. 1 depicts an example host venue;

- [0008] FIG. 2 depicts an example network environment with a host device in communication with multiple client devices;
- [0009] FIG. 3 depicts an example client device in communication with an example host device;
- 5 [0010] FIG. 4 depicts aggregator servers communicating with client devices and a host device;
- [0011] FIG. 5 depicts a host venue displaying an example audience video feed;
- [0012] FIG. 6 depicts a host venue displaying another example audience video feed;
- [0013] FIG. 7 depicts a host venue displaying another example audience video feed;
- 10 [0014] FIG. 8 depicts a host venue displaying another example audience video feed;
- [0015] FIG. 9 depicts a host venue displaying another example audience video feed with an audience member detail window;
- [0016] FIG. 10A depicts an example audience compositing process;
- [0017] FIG. 10B depicts an example composited audience;
- 15 [0018] FIGS. 11A-11E depict example audience video feeds of different resolutions;
- [0019] FIG. 12 depicts an example graphical user interface for a client device;
- [0020] FIG. 13 depicts an example process of generating composite audience audio; and
- [0021] FIG. 14 depicts an example process of generating synthesized audience audio.
- [0022] The use of the same or similar reference numerals in different figures indicates
20 similar, related, or identical items.
- [0023] Additionally, it should be understood that the proportions and dimensions (either relative or absolute) of the various features and elements (and collections and groupings thereof) and the boundaries, separations, and positional relationships presented therebetween, are provided in the accompanying figures merely to facilitate an
25 understanding of the various embodiments described herein and, accordingly, may not necessarily be presented or illustrated to scale, and are not intended to indicate any preference or requirement for an illustrated embodiment to the exclusion of embodiments described with reference thereto.

DETAILED DESCRIPTION

[0024] Live performances of the arts, such as music, theater, or the like, are unique experiences for both the audience and the performers. For the audience, part of the thrill of a live performance is proximity to the performers, and a dynamic, unpredictable experience.

5 For the performer, seeing and hearing the audience is a validating, exciting experience, and can charge performers to give a more exciting, energetic performance.

[0025] In some cases, however, it may not be feasible or convenient for audiences to attend performances live. For example, some individuals may want to attend a live performance, but not be able to travel to a location where a performance is taking place.

10 Also, individuals may not be able to attend due to capacity constraints of the venue. As another example, large public gatherings, such as for concerts, may be limited or prevented by health or safety concerns. However, as noted above, simply recording a performance is not a suitable alternative for either the audience or the performer.

[0026] Live performances may be streamed to audiences, but bandwidth utilization at a venue often presents a challenge for streaming providers. Streamed live performances that do not have an in-person attendance, or do not have large in-person attendance, do not, however, provide a suitable substitute experience for performers. As a result, live performances that are streamed without a typical audience in attendance often do not offer the same energy or unpredictability as in-person live performances. For example, performers do not exhibit the same authentic level of energy or excitement or crowd interaction because of the absence of live feedback and/or audience participation. Conventionally, in part for these reasons, streamed live performances typically draw low attendance and low audience engagement (which may correlate to future/repeat attendance, merchandise purchases, and so on).

25 **[0027]** Some conventional solutions to improve the experience for live performers of streamed live performances offer a very small number of audience members an opportunity to stream live video back to the performer at the venue of the performance. In such conventional solutions, a performer may interact to a certain extent with several audience members. However, due to extraordinary bandwidth requirements necessary to converge multiple video streams to a single venue, conventional solutions can only provide for a performer a live video audience of, at most, dozens of attendees. As known to a person of skill in the art, the greater the number of video streams (attendee video streams) converging to a particular single location, the greater the bandwidth requirement and on-site processing power is required. In addition to extraordinary bandwidth requirements, extraordinary on-site processing power requirements are also necessary to minimize latency, buffering, and so

35

on. Even in ideal network and processing circumstances, a performer does not experience real-time feedback from the several video attendees. As a result, conventional solutions are entirely incapable to provide a performer with an experience similar to performing in a sold-out venue in which tens of thousands of attendees are present.

5 **[0028]** Accordingly, described herein are systems and methods that provide a real-time, interactive experience for the audience, or for the performer, or both during live performances. For example, performers, such as a band, may put on a live performance in a theater, soundstage, or other host venue, and the performance is streamed (both video and audio) to multiple audience members who may view or listen to the live performance on the
10 device of their choice (e.g., a smartphone, tablet, notebook or desktop computer, television, or the like). Acoustic, haptic, and/or visual feedback may be provided to the performer in the theater, soundstage, or other host venue to provide the performer with the impression of performing at a venue with in-person attendees. In some cases, for example, a projection system or video screen can be provided in front of the band (e.g., in an audience seating
15 area) that depicts audience members. Due to stage lighting illuminating the performer and the distance separating the performer from the display screen/projection system, the individual audience members shown to the performer may be depicted at very low resolution. As a result, each individual low-resolution attendee image can be based on or otherwise extracted from or derived from a real-time video stream captured by an audience
20 member's personal electronic device (e.g., tablet, smartphone, and so on). In this manner, each individual event attendee's device may down-sample the video stream, and transmit to the venue a very low resolution image of the attendee, which in turn may be merged with other attendees' low-resolution images to assemble a simulated audience for the performer.

[0029] In other cases, bandwidth utilization may be further reduced by reducing a color
25 space of certain or all audience video streams. For example, audience members farther from the stage may be streamed in black and white at very low resolution (e.g., 10x10 pixels) whereas audience members closer to the stage may be streamed in a restricted or limited colors space at a higher resolution. Full color may not be required, as color may be difficult for a performer illuminated by stage lights to perceive.

30 **[0030]** In some embodiments, the merged audience video stream, including hundreds or thousands of individual real-time low resolution images, may be modified by a video processing pipeline, whether on-site or remote, to enhance or change motion and/or color thereof. For example, motion amplification techniques can be applied to the merged audience video stream to impart an effect of motion that can be synchronized in real time
35 with an audio stream captured at the venue. In this manner, a performer may observe the

audience video stream moving or changing in real time with music being performed at the venue, despite that individual video streams may be delayed or may have different latencies.

[0031] In other cases, the merged audience video stream may have a motion effect applied to it on-site that imparts a crowd-level motion effect, such as a linear progressive crowd motion (e.g., "the wave") that transits the crowd in a particular direction or that follows
5 a particular path, such as venue-front to venue-back, or stage-left to stage-right.

[0032] In other cases, colors can be selectively enhanced. For example, in some embodiments, white color may be enhanced in individual video streams in order to simulate an effect of an audience elevating lighters or cell phone flashes.

[0033] In further embodiments, crowd noise may be transmitted at the venue. The crowd noise may be simulated, and/or may be based, at least in part, on individual microphone signals from individual audience members. In some cases, different performance attendees may be selectively or randomly amplified, so that a performer may hear or perceive distinct voices among the audience in attendance. In many cases, as with video down-sampling and
10 resolution reduction as described above, audio may be transmitted from end-user devices in
15 a low-quality or reduced quality manner to reduce the bandwidth required thereof.

[0034] In many cases, video effects – such as those described above – can be synchronized with simulated and/or merged audio streams to further the experience for performer(s). For example, a motion video effect may result in an appearance that the
20 attendees are jumping in substantial unison. This visual effect may be paired and synchronized with a periodic change in envelope amplifying simulated or merged crowd noise. This effect can result in an appearance that an audience is jumping in unison.

[0035] In some cases, crowd noise may be directionally controlled or transmitted. For example, multiple loudspeakers may be positioned at different locations within a venue that
25 may be independently controlled to simulate different sections of an audience responding differently to different stimulus. For example, a performer may address a crowd in a particular section; in these examples, crowd noise can be selectively amplified in that area.

[0036] In some cases, different crowd noise effects can be generated which may be triggered automatically or manually (e.g., by an on-site or off-site producer) in response to an
30 action of a performer. For example, a performer may ask an audience to shout or respond only if they align with a particular demographic or sub-demographic of attendees (e.g., persons having birthdays, persons celebrating an event, particular gender identities, and so on).

[0037] In still other implementations, haptic effects can be provided to a stage as described herein. In these examples, a performer may see a crowd, may hear crowd noise, and may feel the crowd's response through vibrotactile or concussive haptic feedback provided to the stage floor.

5 [0038] In yet other examples, environmental properties of the venue can be controlled. For example, humidity may be increased to simulate a closed-environment venue with hundreds of attendees. In other cases, temperature and/or humidity and/or lighting conditions may be slowly changed over the course of a performance to simulate performing in a particular atmosphere. In other cases, fans, wind machines, and fog machines may be
10 used to simulate open or outdoor space. In some cases, fragrance or scents can be added to the venue to further enhance the performers' experience.

[0039] These foregoing examples are not exhaustive of the effects that may be provided to an audience video stream, an audience audio stream, and/or a venue itself, such as
15 described herein. For example, a venue as described herein can be operated and/or tailored to simulate a number of different venues of any suitable size. Environmental properties, video properties, motion properties, acoustic properties, audience reactions, crowd noise, humidity, temperature, scents, and so on may vary from venue to venue and thus may be varied according to a particular embodiment or implementation as described herein.

[0040] More broadly, to provide a more interactive experience, real-time video and/or
20 audio streams of the audience members may be displayed to the performers. In this way, the performers get to experience the audience in real-time, while the audience is watching and hearing the performance. To further increase immersion for the performers, the video streams of the audience members may be projected or otherwise displayed on a large (e.g., theater-sized) screen in front of the performers, simulating the experience of being in front of
25 a live audience. The video streams may be displayed, for example, through virtual reality headsets to provide the performer a more immersive and realistic experience. The video streams may show the audience member's video feed without modification, or it may be modified to show only portions of the audience member's video feed.

[0041] Audio of the audience members (or representing the audience members) may
30 also be presented or played back to the performers along with the video streams. For example, applause, cheers, and singing from the remote audience members may be streamed to the venue and presented or played back to the performers. The audio volume may, for example, be commensurate with the audience member's sound intensity. In other examples, the performers may adjust the volume in accordance with their preferences. The

audio of the audience members may be down-sampled prior to transmitting it to the performers.

[0042] In order to further customize the experience for both the audience members and the performers, various presentation parameters may be tailored based on factors like the audience size, the type or genre of performance, a selected venue size, or the like. For example, in the case of a rock concert, crowd noise may be presented to the band throughout the concert, while in the case of an orchestral performance, crowd noise may be muted or otherwise inaudible during the performances. As another example, the size of the individual audience member's video feeds (as shown to the performers) may be scaled based on the number of audience members viewing. Thus, a band playing to fewer people may see those people more closely, just as they would in a small venue with fewer audience members present. In another example, an audience member can tailor the proximity to the performers, including viewing angle, similar to choosing seats when purchasing tickets or choosing a viewing spot in a venue. The audience member may also choose to focus on a particular performer, for example. These and other aspects of the real-time, two-way remote performance system are described herein.

[0043] These foregoing and other embodiments are discussed below with reference to FIGs. 1-14. However, those skilled in the art will readily appreciate that the detailed description given herein with respect to these figures is for explanation only and should not be construed as limiting.

[0044] FIG. 1 illustrates an example host venue 100. The host venue 100 may be a location where a performance occurs. The host venue 100 may have a performance area 102, such as a stage, where the performance occurs. FIG. 1 illustrates the performance as a musical concert (where performers 112 are shown as a band), though this is merely for illustrative purposes, and the performance may be other types of live performances as well, such as theater productions (e.g., plays, musicals, etc.), sketch comedy, stand-up comedy, orchestral performances, dance performances, ballets, operas, speeches, presentations, panel discussions, or the like. In another example, the host venue 100 or the performance area 102 may include multiple locations. The multiple locations could show multiple aspects of a performance including, for example, a pop band, dancers, and an orchestra. The performance location 102 may be mobile.

[0045] The host venue 100 may also include an audio and video capture system 114. The audio and video capture system 114 (also referred to herein as an A/V capture system 114) is configured to capture audio and video of the performance for delivery to the audience members. The A/V capture system 114 may include one or more cameras, 3D cameras, one

or more microphones, and associated recording and processing equipment. For example, the A/V capture system 114 may include mixing boards, recording systems, computer systems for storing, processing, and transmitting recorded audio and video, and the like. The A/V capture system 114 may also include audience feedback features for the performers. For example, the A/V capture systems could cause the stage to vibrate to mimic audience members jumping during a live performance.

[0046] The host venue 100 may also include an audience presentation screen 104 on which images of an audience (including images of individual audience members) may be displayed. The audience presentation screen 104 may be sufficiently large to provide an immersive experience to the performers 112. For example, the audience presentation screen 104 may be about 100 feet wide and about 75 feet tall. Other dimensions (both larger and smaller) are also contemplated. The audience presentation screen 104 may be a single, continuous screen, or it may be formed of multiple smaller screens adjacent one another. In some cases, the audience presentation screen 104 may include discrete screens placed on and/or attached to theater seats in a physical theater environment. In other cases, the screens may surround the performance area 102. In some examples, the performers may have virtual reality headsets.

[0047] The audience presentation screen 104 may use any suitable type of display and/or projecting technology to display an audience (or other graphics) on the audience presentation screen 104. For example, the audience presentation screen 104 may be a projection screen on which an image of the audience is projected, using either a front or back projection system (which may include one projector or multiple projectors configured to cooperate to display a single audience video feed). As another example, the audience presentation screen 104 may be multiple discrete displays (e.g., LCD displays) arranged adjacent one another to form a larger screen. Separate video signals may be sent to each of the discrete displays to produce the image of the audience. In some cases, the separate video signals may correspond to different portions of an audience video feed (e.g., each discrete screen is configured to display a designated portion of a single video of an audience or representative of an audience). In some cases, the display may pan through the audience by displaying portions of the audience in the presentation screen 104. These portions of the audience could correspond, for example, on a performer's position in the performance area 102.

[0048] The audience presentation screen 104 may be flat or curved. In the case of a flat screen, it may be positioned substantially vertical or angled (e.g., with the top of the audience presentation screen 104 pitched towards or away from the performance area 102).

In the case of a curved screen, the bottom of the screen may be closer to the performance area 102, while the top of the screen may be further from the performance area 102. For the performers, a curved or angled presentation screen may provide a similar experience to being in a theater or auditorium, where balconies are set further back from the performance area 102 than the floor seats.

[0049] In some cases the host venue 100 is a theater with a seating area. In such cases, the audience presentation screen 104 may be positioned between the performance area 102 and the seating area. In this way, the audience presentation screen 104 may block the performer's view of the seating area (which may be empty), while also displaying the image of the audience in a manner that is similar to and/or representative of being in front of a live audience.

[0050] The host venue 100 may also include an audience audio output system 110. The audience audio output system 110 may include one or more speakers that output audience noise to the performers 112 during the performance. As described herein, the audience noise may include audio of audience members captured on the audience members' devices, synthesized audience noise (e.g., digitally generated), pre-recorded audience noise, or combinations thereof. Parameters of the audience noise may depend on factors such as the number of audience members, the presentation type, a virtual venue size or type, or the like. Audience noise and how it may be generated and presented is discussed in greater detail herein.

[0051] As described herein, the audience video feed may include videos of or corresponding to multiple audience members. For example, videos of the audience members may be captured in real-time, during the performance and while the audience members are viewing the performance, and sent to a server associated with the host venue 100. The videos of the audience members may be captured by the devices on which the audience members are viewing the performance. The videos may be combined into a single video feed that includes all or some of the audience members' videos, and the resulting single video feed may be presented on the audience presentation screen 104 for the performers 112 to view during their performance.

[0052] FIG. 1 shows the audience presentation screen 104. The audience presentation screen may include a plurality of audience video feeds 106. The audience video feeds may be presented in a tile or grid-like format, whereby each individual audience video feed 106 is displayed as a separate visual element. The audience video feeds may be combined into a single video feed (which may reduce the processing or projection complexity as compared to projecting or displaying discrete video feeds or streams for each audience member).

[0053] In some cases, audience members that are visible in the audience video feeds may be extracted or separated from their backgrounds and composited to form a more natural appearing audience (e.g., where the audience members overlap one another, are sitting in seats, standing in groups, or the like).

5 **[0054]** As used herein, video content (e.g., a video stream or video file) that corresponds to an audience member extracted or separated from a background portion of an audience video feed may be referred to as an isolated audience object. An isolated audience object may correspond to any portion of an individual in an audience video feed, such as the individual's head, head and shoulders, entire body, or any sub-portion thereof. In some
10 cases, multiple audience members in an audience video feed, may be extracted or separated from a background, such that multiple isolated audience objects may be produced from a single audience video feed. Compositing of isolated audience objects to produce a natural-looking audience is discussed further with respect to FIGS. 10A-10B. In some examples, an audience member's features not shown in the video feed, may be inserted or
15 replicated in the video feed to generate a more natural-looking audience.

[0055] In some cases, combining multiple audience video feeds to produce a single video feed of the audience may include resizing the audience video feeds (and/or the isolated audience objects) to produce a natural-appearing audience. For example, the video feeds of audience members that are closer to the bottom of the audience presentation screen 104
20 may be larger than those that are closer to the top of the audience presentation screen 104, thus producing the effect that the smaller audience members appear more distant. As another example, the resolution of the audience video feeds (and/or the isolated audience objects) may be higher for audience members that are closer to the bottom of the audience presentation screen 104 than those that are closer to the top of the audience presentation
25 screen 104, simulating the effect that more distant audience members are more difficult to see clearly.

[0056] The compositing process may also include showing the audience members in an audience environment that corresponds to a particular type of performance venue. For example, isolated audience objects may be composited into a theater environment, with
30 isolated audience objects shown seated in discrete seats. As another example, isolated audience objects may be composited into a field or stadium environment, simulating the audience of an outdoor music festival. The particular type of audience environment may be selected based on several factors, such as the performance type (e.g., pop music, classical music, plays, dance, spoken word, etc.), the number of audience members, a selection by

the performers (or other appropriate individual or entity), a size of the audience presentation screen 104, or the like.

[0057] FIG. 2 illustrates how audience devices may communicate with a host device via a network to provide the functionality described herein. For example, FIG. 2 shows audience members 202, 204, and 206, viewing and/or listening to a performance on their respective devices 203 (e.g., a computer), 205 (e.g., a mobile device such as a phone or tablet), 207 (e.g., a television). The devices 203, 205, 207 may include audio/video (A/V) output devices (e.g., display screens, speakers, etc.) to present the performance to the user. The devices 203, 205, 207 may also include client-side A/V capture systems that are configured to capture audio and/or video content of the audience members 202, 204, 206 while they are watching the performance. The devices may send the audience A/V content to a host device 212 via a network 210 (e.g., the Internet). Notably, the audience A/V content is sent to the host device 212 substantially in real time, while the audience members are consuming the live performance being delivered to their devices from the host device 212. This real-time exchange allows the performers to view and optionally hear the reactions of the audience members in real-time.

[0058] The host device 212 may be one or more server computers associated with a host venue. The host device 212 may receive performer audio/video (A/V) content 218 (e.g., from the A/V capture system 114) and transmit the performer A/V content 218 to the audience devices via the network 210. The host device 212 may cause the audience A/V content 216 (received from the audience device 203, 205, 207) to be presented to the performers in the host venue via an audience A/V presentation system 220, which may include the audience audio output system 110 and audience presentation screen 104, FIG. 1. In some examples, the audience A/V content 216 may also be transmitted to other audience members 204.

[0059] In some cases, the host device 212 processes the audience A/V content 216, or causes the audience A/V content to be processed, prior to being presented to the performers via the audience A/V presentation system 220. For example, the host device 212 (or other associated A/V processing device or system) may isolate and/or extract the portions of the audience video feeds that correspond to audience members (e.g., to produce isolated audience objects), and combine the isolated audience objects to form a virtual audience for display to the performers and/or other audience members.

[0060] FIG. 3 depicts a system 300 configured to facilitate real-time A/V exchange between performers and audience members. In this example embodiment, the system 300 includes a client device 302, and a host device 304. The client device 302 (which may be an embodiment of the audience devices 203, 205, 207 in FIG. 2) represents one of many

potential client devices that may communicate with the host device 304. For example, hundreds, thousands, or even more client devices may ultimately communicate with the host device 304 to send audience A/V content (e.g., audience video 308 and audience audio 310) to the host device 304 and to receive performer A/V content 314 from the host device 304 in
5 real-time.

[0061] The host device 304, which may be an embodiment of the host device 212, represents one or more server systems that are associated with a host venue. The host device 304 may capture and store performer A/V content (e.g., from microphones, depth sensing systems, and camera systems in the host venue), and may send performer A/V
10 content 314 (and optionally audience A/V content 312) to the client device 302.

[0062] With regard to the client device 302, a client application 301 instance may execute over one or more computing resources of the client device 302 (such as the resource allocation 303), and can receive performer A/V content 314, and optionally audience A/V content 312, from the host device 304, and present the performer A/V content 314 and
15 optional audience A/V content to a user (e.g., via a display, speakers, etc.). The client application 301 may also capture audience video 308 and optionally audience audio 310 (e.g., via a camera, microphone, etc.), and send the captured audience video and audio 308, 310 to the host device 304.

[0063] More specific to the foregoing, the client device 302 may include one or more
20 physical or virtual computing elements such as a processor, a working memory, and a persistent data store or persistent memory. In many embodiments, computer code defining an instance of the client application 301 can be stored at least in part in a persistent memory. Upon request, a processor of the client device 302 can be leveraged to access the persistent memory and load at least a portion of that computer code into the working
25 memory, thereby at least partially instantiating an instance of the client application 301.

[0064] The client application 301 can be configured to generate or otherwise render a graphical user interface that can be shown on a display of the client device 302. The graphical user interface, as noted above, can be configured to display any suitable information related to or associated with a performance, such as a live video feed of a
30 performance as the performance is taking place. The graphical user interface may also be configured to display a video of the audience. The graphical user interface may also share content, adjust settings, and provide other interactive options.

[0065] For example, the graphical user interface may display the same audience video that the performers see on the audience presentation screen 104 (in addition to or instead of

the live performance). The client application 301 may also be configured to output audio of the performance, and optionally audio of the audience, via speakers of or associated with the client device 302. With respect to the latter example, the client application 301 may output the audience audio that is presented to the performers. The audience audio that is sent to the client device 302 from the host device 304 may be mixed with the audio of the performance to provide an experience that is similar to being in the audience of a performance (e.g., where the audience member hears the audience as well as the performers). In some cases, the client application 301 may also allow a user to control aspects of the performance presentation. For example, the user may select whether or not they want to see or hear the audience in addition to the performance.

[0066] The client device 302 can be any suitable electronic device. In many embodiments, as noted above, the client device 302 is a mobile electronic device such as a smart phone, a tablet computer, or a laptop computing device. These are merely examples; any suitable computing device or computing resource may be configured to, in whole or in part, instantiate a client application as described herein, such as the client application 301.

[0067] With regard to the host device 304, the host device 304 may receive audio and video feeds from the A/V capture system 114 (FIG. 1), and at least temporarily cache or store the audio and video feeds in a database 306 or other memory store. The host device 304 may optionally process the audio and video feeds (e.g., to convert them to different format, to combine the audio and video feeds into a single container, to apply filtering, equalization, or other modifications, etc.). The host device 304 may also generate audience A/V content 312 for display to the performers and optionally to audience members.

[0068] The audience A/V content 312 may be generated using the received audience audio 310 and audience video 308, or it may be generated independently of the received audience audio 310 and audience video 308 (e.g., synthetically generated audio and/or video content). The audience A/V content 312 may be provided to the client device 302, and also to an audience A/V presentation system in a host venue (e.g., the audience presentation system 220, FIG. 2). The host device 304 may implement these and/or other processes using a host application 309. The host application 309 instance may execute over one or more computing resources of the host device 304 (such as the resource allocation 305).

[0069] The host device 304 may include one or more physical or virtual computing elements such as a processor, a working memory, signal processing, and a persistent data store or persistent memory. In many embodiments, computer code defining an instance of the host application 309 can be stored at least in part in a persistent memory. Upon request,

a processor of the host device 304 can be leveraged to access the persistent memory and load at least a portion of that computer code into the working memory, thereby at least partially instantiating an instance of the host application 309.

[0070] The host device 304 can be any suitable electronic device. In many embodiments, as noted above, the host device 304 is a server computer or multiple server computers. Each of the server computers may have all or some of the features of the host device as described herein. These are merely examples; any suitable computing device or computing resource may be configured to, in whole or in part, instantiate a host application as described herein, such as the host application 309.

[0071] While FIG. 3 illustrates an example of a single host device communicating with a single client device, it will be understood that the illustrated and described communications may occur between a host device (or multiple host devices) and multiple client devices. For example, an audience for a virtual performance as described herein may include hundreds or even many thousands of client devices, all of which may communicate with host devices in the manner shown and described with respect to FIG. 3.

[0072] In some cases, aggregator servers may be used to reduce the number of discrete connections the host device must maintain in order to send and receive performer and audience A/V content to and from client devices. FIG. 4 illustrates an example system 400 in which aggregator servers 402, 404 act as an intermediary between a host device 410 (which may be an embodiment of the host devices 212, 304) and client devices 405 (which may be embodiments of the audience devices 203, 205, 207, 302). In particular, a first aggregator server 402 may accept connections from a first group 406 of client devices 405, and a second aggregator server 404 may accept connections from a second group 408 of client devices 405. Of course, additional aggregator servers may also be implemented to accept connections from additional groups of client devices.

[0073] The aggregator servers 402, 404 may communicate with the host device 410 to receive performer A/V content (e.g., the performer A/V content 314, FIG. 3) and audience A/V content (e.g., the audience A/V content 312, FIG. 3). The aggregator servers 402, 404 may then send the received content to each client device over a discrete client-server connection.

[0074] The aggregator servers 402, 404 may also receive audience audio and audience video (e.g., audience audio 310 and audience video 308, FIG. 3) from each individual client over the discrete client-server connection. The aggregator servers 402, 404 may then provide the audience audio and audience video from their respective client devices to the

host device 410 via a single respective connection. In this way, the host device 410 need not maintain a discrete connection to each client device. In some embodiments, the aggregator servers 402, 404 may send or receive audience audio and/or video to intermediate aggregator servers or intermediate host devices.

5 [0075] In addition to consolidating the audio and video content from multiple client devices into a single communication channel to the host device, the aggregator servers 402, 404 may perform video and/or audio processing on the audience audio and audience video content received from the client devices. For example, the aggregator servers may process audience videos to produce isolated audience objects, and send only the isolated audience
10 objects to the host device. In some cases, the aggregator servers also composite isolated audience objects together to produce video content corresponding to a sub-portion of an audience video feed. The sub-portion of the audience video feed may be sent to the host device, which may further composite multiple sub-portions to form a single audience video feed (or project or otherwise display the multiple sub-portions on an audience presentation
15 screen 104 to produce the image of the audience). FIG. 1, for example, illustrates how an audience video presentation 108 may include multiple sub-portions (e.g., 108-1, 108-2).

[0076] In some cases, client devices 405 are assigned to a particular aggregator server based on the physical proximity to the aggregator server. For example, a client device 405 may be assigned to the aggregator server that is closest to that client device 405. In some
20 cases, client devices are assigned to aggregator servers based on other factors. For example, a client device 405 may be assigned to an aggregator server having a fastest connection to the client device, or to an aggregator server having a lowest connection latency to the client device, or the aggregator server having the smallest load or the highest capacity. Other factors may also be considered.

25 [0077] It is appreciated that the foregoing embodiments depicted in FIGS. 1-4 and the various alternatives thereof and variations thereto are presented, generally, for purposes of explanation, and to facilitate an understanding of various configurations and constructions of a system, such as described herein. However, it will be apparent to one skilled in the art that some of the specific details presented herein may not be required in order to practice a
30 particular described embodiment, or an equivalent thereof.

[0078] For example, each client device, host device, server, or service of FIGS. 1-4 may be implemented in a number of suitable ways. As illustrated, the client devices, host devices, and the aggregator servers each includes one or more purpose-configured components, which may be either software or hardware. More generally, it may be appreciated that the
35 various functions described herein of a client device, host device, and/or aggregator servers

can be performed by any suitable physical hardware, virtual machine, containerized machine, or any combination thereof. In some examples, the user devices may be connected to a cloud. In such examples, a substantial amount of processing A/V data may be performed using computing units located within the user's devices, or near the user's device in order to reduce the bandwidth requirements at the aggregator or at the host servers and to reduce the latency.

[0079] FIGS. 5-9 illustrate several examples of displaying audience video content to performers in a host venue. FIG. 5 illustrates a host venue 500 with a performance area 502 and an audience presentation screen 504. The performance area 502 and the audience presentation screen 504 may be embodiments of the performance area 102 and the audience presentation screen 104, and for brevity details of those items are not repeated here. FIG. 5 illustrates an embodiment in which audience video feeds 508 that are positioned higher on the audience presentation screen 504, thereby representing more distant audience members, are smaller than audience video feeds 506 that are positioned lower on the audience presentation screen 504. As noted above, this may provide the appearance, to the performers, of a natural-looking audience where more distant individuals appear smaller. In some cases, audience members may pay higher prices to have their video feed be larger and/or closer to the performers, just as conventional ticket prices may be higher for seats that are closer to the stage in a real-world venue. In other cases, an audience member's video feed may be larger depending on the quality of their video feed while smaller video feeds could correspond with audience members with more video lag.

[0080] FIG. 6 illustrates a host venue 600 with a performance area 602 and an audience presentation screen 604. The performance area 602 and the audience presentation screen 604 may be embodiments of the performance area 102 and the audience presentation screen 104, and for brevity details of those items are not repeated here. FIG. 6 illustrates an embodiment in which audience video feeds may have different resolutions. For example, audience video feeds that are closer to the bottom of the audience presentation screen 604 (and/or are closer to the performers) have a higher resolution than those that are closer to the top of the audience presentation screen 604 (and/or are further away from the performers). Audience video feed 606, for example, may be a highest resolution, in which the features of the audience member (e.g., facial features) may be visible and/or distinguishable. Audience video feed 608, which may have a lower resolution than the audience video feed 606, may be recognizable as a shape of a person, but may not be as visually distinct as the audience video feed 606. Audience video feed 610 may have a lower resolution than the audience video feeds 606, 608, and may not have a recognizable human

form. In this case, movements and/or colors of the audience member may be visible even if facial features or other shape characteristics are not.

[0081] The resolution of a given audience video feed may be selected in various ways. For example, a user may select how they wish to be presented to the performers. As another
5 example, audience video feeds that are subject to slower network connections may be down-sampled (e.g., converted to a lower resolution) to reduce the latency of the audience video feed, thereby ensuring that the audience video feed shows the audience members' reactions in substantially real-time to the performance. In one example, the video feed could be limited to a predetermined number of audience members, with rotating video feeds to
10 reduce the number of down-sampled feeds.

[0082] The resolution of the audience video feeds may be scaled according to their position on the audience presentation screen 604. For example, each "row" of audience video feeds may be associated with a particular resolution (with the resolution decreasing with increasing height of the row). As another example, multiple rows may be associated
15 with a same resolution. For example, a first third of the rows may be associated with a first resolution, a second third of the rows may be associated with a second resolution, and the remainder of the rows may be associated with a third resolution.

[0083] In cases where audience video feeds are down-sampled (e.g., based on connection speed, connection latency, audience member preference, etc.), the down-
20 sampling may occur on the client device. In this way, the latency may be further reduced, as a full-resolution video feed need not be sent from the client device. In cases where the client device communicates with an aggregator server, the aggregator server may perform the down-sampling. In this case, the processing load on the client device may be reduced, while the latency (and overall bandwidth requirements) may still be reduced. In some cases,
25 instead of or addition to down-sampling (e.g., reducing the resolution of audience video feeds), the audience video feeds may be decolorized or rendered in greyscale or black and white pixels. This may further reduce the size of the video feeds being sent, potentially decreasing latency. In some examples, a client device may not transmit any video feeds. This could depend on the system's capacity or the audience member's preference. In these
30 cases, the audience presentation screen 604 may show a non-distinct image, similar to the audience feed of 610 for example, as a placeholder for that audience member.

[0084] FIG. 7 illustrates a host venue 700 with a performance area 702 and an audience presentation screen 704. The performance area 702 and the audience presentation screen 704 may be embodiments of the performance area 102 and the audience presentation
35 screen 104, and for brevity details of those items are not repeated here. FIG. 7 illustrates an

embodiment in which fewer audience video feeds 706 are shown on the audience presentation screen 704, but each individual audience video feed is larger than those shown in other examples in which there are more audience members. The fewer number of audience video feeds in FIG. 7 may be due to fewer audience members viewing the performance, or fewer audience members having chosen to allow their video feeds to be displayed to the performers. In some cases, the size of the audience video feeds is scaled based on the number of audience video feeds available for display. Thus, fewer audience members will result in each audience video feed being larger.

[0085] In some cases, the number of audience members is limited based on a virtual venue size associated with a performance. For example, a band may choose to perform in a small venue, such as a bar or small theater. In such cases, the number of audience video feeds may be limited to what would be expected in the selected venue, and the sizes of those audience video feeds may be scaled up to substantially fill the audience presentation screen 704. In some cases, the number of audience members who are able to view the performance may be limited based on the virtual venue size. In such cases, video feeds for each audience member may be displayed on the audience presentation screen 704 (if they opt-in to being displayed to the performers). In other cases, the selected virtual venue size may define or limit the number of audience video feeds shown on the audience presentation screen 704, but more audience members may be able to view the performance. In some examples, the displayed video feed may swap the displayed audience members to show the performers more audience members without altering the size and resolution of the video feed.

[0086] FIG. 8 illustrates a host venue 800 with a performance area 802 and an audience presentation screen 804. The performance area 802 and the audience presentation screen 804 may be embodiments of the performance area 102 and the audience presentation screen 104, and for brevity details of those items are not repeated here. FIG. 8 illustrates an embodiment in which the audience video feeds of audience members who are associated with one another may be displayed together in a group 806. Audience members may be associated with one another (such that they are presented together in a group) based on any suitable criteria. For example, in some cases, tickets for a performance may be purchased in a group, and audience members using those tickets may be displayed in a group. As another example, users may opt-in to a group. For example, a ticket purchase interface may allow a purchaser to establish a group, such as "Family of Individual A." Other ticket purchasers may be able to select a group with which they want to be associated. An author of a group may be allowed to approve or decline requests to be associated with a group.

[0087] As yet another example, ticket purchasers who are linked to one another on other social networks may be grouped together or may be presented with an offer to be grouped together. For example, after purchasing a ticket to a performance, a user may be informed of which of their contacts on a social networking site have also purchased tickets to the performance, and given the opportunity to form or join a group with those contacts.

[0088] FIG. 9 illustrates a host venue 900 with a performance area 902 and an audience presentation screen 904. The performance area 902 and the audience presentation screen 904 may be embodiments of the performance area 102 and the audience presentation screen 104, and for brevity details of those items are not repeated here. As described herein, one of the advantages of displaying real-time video feeds of audience members to performers is that the experience of a live audience at a live performance may be produced in a virtual manner. To further enhance the experience, the performers may be able to interact with the audience video feeds. For example, performers (or individuals associated with a live performance, such as a producer) may be able to select individual audience video feeds in order to further interact with that audience member.

[0089] With reference to FIG. 9, a performer (or producer or other individual, or as a result of a random or other automated selection) may select a particular audience video feed 906. Selecting the audience video feed 906 may cause a detail window 910 to be displayed on the audience presentation screen 904. The detail window 910 may include a larger (and/or higher resolution) video feed 912 of the audience member, and optionally additional information 914 about the audience member (e.g., a virtual seat assignment, a name, a city, state, or region of residence, etc.). The performers may use the information in the detail window 910 to individually call out or recognize that audience member during the performance. In some cases, audience members may be randomly selected to have their video feeds presented in a detail window 910, or they may be selected as contest winners, or they may even pay to purchase such an experience.

[0090] Selection of a given audience video feed may be made in any suitable manner. In some cases, the host venue 900 may include gesture and/or motion recognition systems (including but not limited to cameras, LIDAR, dot projectors, motion capture processing systems, etc.) that can detect where a performer is pointing or gesturing, and thereby determine which audience video feed or group of audience video feeds (e.g., the group 806, FIG. 8) the performer is selecting. Once the audience video feed is identified, a detail window 910 may be displayed for the selected audience video feed. In other cases, a producer or other non-performing individual may select an audience video feed at an appropriate time. In yet other cases, audience video feeds are selected for presentation in a

detail window 910 automatically, based on detected locations in a performance (e.g., a particular part of a song where callouts often occur, between acts of a play, between songs or between a particular pair of adjacent songs in a set list, etc.).

[0091] As noted above, audience video feeds may be displayed on an audience presentation screen unmodified and in their entirety, or they may be down-sampled, cropped, subject to background deletion (e.g., extracting the video content corresponding to people), or the like. FIGS. 10A-10B illustrate how audience objects may be isolated from audience video feeds and how the isolated audience objects may be composited to form a natural-looking audience.

5 **[0092]** FIG. 10A illustrates a plurality of audience video feeds 1002, which may be provided to an audience object isolation service 1003. The audience object isolation service 1003 may process the audience video feeds 1002 to isolate audience objects from the audience video feeds 1002. For example, the audience object isolation service 1003 may determine which portions of a video feed correspond to audience members and which
15 portions correspond to background or other non-human objects in the video feed. The resulting isolated audience object (e.g., a video feed of just the isolated audience member(s)) may be provided to a compositing service 1004. In some cases, the audience object isolation service 1003 may perform other video processing operations on the isolated audience objects, such as changing the video resolution (e.g., down-sampling), color
20 removal, greyscale or black-and-white conversions, or the like.

[0093] The compositing service 1004 may combine the isolated audience objects into a single video feed that has the appearance of a live audience, without the superfluous background portions of the audience video feeds. The composited audience 1016 represents at least a portion of the composited audience video feed that may be produced
25 by the compositing service 1004.

[0094] Compositing the isolated audience objects may include overlapping different isolated audience objects to give the appearance that some audience members are in front of others, or differing video positioning to give the appearance of a difference in height. In some cases, the compositing service 1004 may also add features that are not present in the audience video feeds. For example, isolated audience objects may be composited into a theater environment, and optionally shown in theater seats, balconies, and the like. As
30 another example, a composited audience scene may include an image of a field, horizon, and sky, simulating the audience of an outdoor music festival. As another example, the audience may be in a futuristic setting like a spaceship. As used herein, the term "image"
35 may be used to refer to video images or still images. In the context of a video image, the

image need not be static (e.g., nonmoving), but may be any suitable graphical component of a video. For example, an image of a tree may show the tree moving in the wind. Unless otherwise indicated, the use of the term “image” does not limit the subject matter to still, static images.

5 [0095] The audience object isolation and compositing processes may occur in real-time or substantially real-time, such that the composited audience 1016 that is displayed to the performers on an audience presentation screen 1014 (FIG. 10B) shows the reactions of the audience members to the performance in substantially real-time. In some cases, the reactions of the audience are displayed back to the performers with a delay of less than 1
10 second, less than about 0.5 seconds, less than 0.25 seconds, or any other suitable value. Stated another way, the round-trip communication time to send performer A/V content to a client device, and receive and display audience A/V content to the performers (e.g., via the audience presentation screen) may be less than about 1 second, less than about 0.5 seconds, less than 0.25 seconds, or any other suitable value. In some cases, any
15 communication path from a host device to a client device, and back to the host device that exceeds a threshold round-trip time or latency value may be excluded from the compositing process (as well as any audience audio processing), thereby avoiding the presence of excessively delayed reactions to the performance.

[0096] In some cases, instead of or in addition to isolated audience objects extracted in
20 real-time from audience video feeds, the compositing service 1004 may include computer-generated audience objects. The computer-generated audience objects may be animated based on performer A/V content to synchronize or otherwise coordinate movements of the computer-generated audience objects to the performance.

[0097] The audience object isolation service 1003 and the compositing service 1004 may
25 be performed by any suitable devices. For example, the audience object isolation service 1003 may be performed by a client device, an aggregator server, a host server (and more particularly by resource allocations of the foregoing devices), or a different client device. The compositing service 1004 may be provided by an aggregator server or a host server (and more particularly by resource allocations of the foregoing devices). In some cases, the
30 audience object isolation service 1003 and/or the compositing service 1004 may be performed by other devices, such as an audience processing server, which may be positioned between the host device 410 and the aggregator servers 402, 404 in FIG. 4.

[0098] In some cases, the audience object isolation service 1003 and the compositing
35 service 1004 may include a content analysis operation where non-audience content or other objectionable or undesirable content may be excluded from the composited audience. For

example, video feeds with no human audience member may be rejected. As another example, video feeds depicting violence or nudity may be rejected. The content analysis operation may include human-based and/or automatic content analysis.

5 **[0099]** As described herein, video feeds of audience members may be down-sampled or be subjected to resolution reductions for various reasons, such as to decrease latency, decrease bandwidth requirements, or the like. FIGS. 11A-11E illustrate various examples of audience video feeds having various resolutions. The various image resolutions shown and described with respect to FIGS. 11A-11E may correspond to spatial- and/or pixel-based resolutions.

10 **[0100]** FIG. 11A depicts an audience video feed 1100 at a first resolution. In some cases, the audience video feed 1100 is a native resolution of an image capture device associated with a client device. In some cases, a remote performance system may have an upper limit on the resolution of audience video feeds, such that any native device resolutions that are greater than the upper limit may be down-sampled to the upper limit. This down-sampling
15 may occur on the client device, such as by a client application (e.g., the client application 301, FIG. 3), or it may occur on a host server or aggregator server, or any other suitable device.

[0101] FIG. 11B depicts an audience video feed 1102 that has been down-sampled to a lower resolution. As shown in FIG. 11B, the audience member is still distinguishable as a
20 person, though the features and details are not as distinct. Notably, the size of the video stream corresponding to the audience video feed 1102 may be less than that of the audience video feed 1100.

[0102] FIG. 11C depicts an audience video feed 1104 that has been down-sampled to an even lower resolution. As shown in FIG. 11C, the audience member has the general shape
25 of a person, but the size of the pixels eliminates most details. The size of the video stream corresponding to the audience video feed 1104 may be less than that of the audience video feeds 1100, 1102.

[0103] FIG. 11D depicts an audience video feed 1106 that has been down-sampled to a lower resolution than 1104. As shown in FIG. 11D, the audience member is represented by a
30 small number of large pixels, rendering the image of the audience member a block-like form. The size of the video stream corresponding to the audience video feed 1106 may be less than that of the audience video feeds 1100, 1102, and 1104.

[0104] FIG. 11E depicts an audience video feed 1108 that has been down-sampled to a lower resolution than 1006. In some embodiments, this resolution is one pixel. The pixel may move or change color or shade based on information in the original resolution audience video feed. Single-pixel audience video feeds may be used to represent distant audience members in a composited audience. In some cases, a large number of single-pixel audience video feeds (e.g., over 100 or over 1000 single-pixel video feeds arranged in a group) may appear as an area of flickering or undulating dots. Because the single-pixels audience video feeds may be based on actual audience video feeds, however, the group of single-pixel feeds may exhibit motion and/or color patterns and changes that react in real-time to the content of the performance. For example, they may be substantially still during an act of a play, but dynamically moving during an applause break.

[0105] The pixel resolutions of the audience video feeds 1100, 1102, 1104, 1106, and 1108 may be selected to be any suitable values. By way of nonlimiting examples, the audience video feed 1100 may have a resolution of around 1500 x 2000 pixels, the audience video feed 1102 may have a resolution of around 300 x 400 pixels, the audience video feed 1104 may have a resolution of around 90 x 120 pixels, the audience video feed 1106 may have a resolution of around 15 x 20 pixels, and the audience video feed 1108 may have a resolution of 1 x 1 pixel. Of course, other pixel resolutions are also contemplated.

[0106] In some cases, instead of or in addition to video resolution down-sampling, video feeds are rendered in black-and-white or greyscale, or otherwise subject to color modifications prior to being composited to form a composited audience video feed. This may help in rendering a more uniform composited audience, and may further reduce latency and bandwidth requirements.

[0107] As described herein, a client application (e.g., the client application 301, FIG. 3) may display, on a display of the client device, both the performance and at least a portion of the audience video feed that is also being displayed to the performers. FIG. 12 illustrates an example user interface 1200 showing both a performance 1202 and an audience video feed 1204. The user interface 1200 uses a picture-in-picture style of display, where the audience video feed 1204 is shown in a separate window or frame within the video feed of the performance 1202, though other display graphical techniques or styles may also be used. The viewer may be able to select whether or not the audience video feed 1204 is displayed during the performance, and may be able to selectively toggle the visibility of the audience video feed 1204 at his or her discretion.

[0108] As noted above, audience audio may be presented to the performers before, during, and/or after a performance. The audience audio that is presented to the performers

may be generated in various ways, including by using actual audience audio feeds sent from the client devices, by generating synthesized audience audio, or a combination thereof.

[0109] FIG. 13 illustrates an example method 1300 for generating and presenting audience audio based on audio feeds sent from client devices. The method may be performed by a host device, such as the host devices 212, 304, or by any other suitable device, such as an audio processing server that is communicatively coupled to the host device 212.

[0110] At operation 1302, audience audio feeds are received from client devices. The audio feeds may be recordings that are generated by the client devices (e.g., the audience devices 203, 205, 207, 302) and sent to the host device via the Internet or other network connection. The received audio feeds may be subject to audio or video processing steps, either by the client devices or by a receiving device (e.g., the host device, an aggregator device, another user's client device, and so on). Such processing steps may include, without limitation, volume normalization, feature extraction, filtering, equalization, compression, or the like.

[0111] At operation 1304, composite audience audio is generated. The composite audience audio may be generated by combining all or a subset of the received audio feeds. In some cases, only audio feeds that are received over a low-latency connection are included in the composite audience audio, thereby ensuring that unduly delayed reactions or audience sounds are not included in the composite audience audio. The composite audience audio may also be limited to a certain number of audience members if, for example, adding additional audience members would not audibly change the composite audience audio.

[0112] In some cases, synthesized audio content is also included in the composite audience audio. For example, if there are not enough audio feeds to produce a sound that is representative of a target audience size, synthesized audio content (e.g., digitally rendered audio) may be added to the audio mix to produce a desired sound profile. The synthesized audio content may be synthesized based at least in part on the received audience audio feeds. Thus, for example, if the received audio feeds include clapping at a particular time, the synthesized audio content may also resemble clapping; if the received audio feeds include cheering at a particular time, the synthesized audio content may also resemble cheering; if the received audio feeds include singing at a particular time, the synthesized audio content may also resemble singing. In this way, the operation of generating the composite audience audio using synthesized audio content may use the actual audience's reactions as a key or guide to determine the nature of the synthesized audio content. In

some cases, the synthesized audio content may be generated by duplicating and then modifying received audience audio feeds.

5 [0113] The operation of generating the composite audience audio 1304 may include applying effects, filters, duplicating the audience member's audio, or otherwise modifying the composite audience audio to suit a particular performance. For example, an echo or reverb effect may be added to the audience audio if the virtual performance venue is a large hall, while a muting effect may be added to the audience audio if the virtual performance is in a small concert venue. Background noise may also be added, such as the noise of a club or bar, if the virtual venue is a club or bar.

10 [0114] At operation 1306, the composite audience audio may be output to the performers. For example, the composite audience audio may be output via the audience audio output system 110 (FIG. 1). The composite audience audio may also be sent to client devices. In such cases, the composite audience audio may be mixed with the performance audio to provide a viewing experience that resembles a live performance (e.g., where the audience members can hear the audience noise in addition to the performance sound).

15 [0115] The composite audience audio may also facilitate direct interaction between the performers and the audience. For example, a band may call for the audience to sing along to a well-known song or chorus (or the audience may simply sing along on their own initiative), and the audience members' singing will be captured by their devices and included in the composite audience audio for the band to hear while they are performing, just as they would in a live performance setting.

20 [0116] FIG. 14 illustrates an example method 1400 for generating and coordinating synthesized audio during a performance. The method may be performed by a host device, such as the host devices 212, 304, or by any other suitable device, such as an audio processing server that is communicatively coupled to the host device 212.

25 [0117] At operation 1402, audience audio is generated. The audience audio may be generated by using pre-recorded or computer-generated audience sounds to produce audience audio that is suitable for the performance and the virtual venue. The operation 1402 of generating the audience audio may include applying effects, filters, or otherwise modifying the pre-recorded or computer-generated audience sounds to suit a particular performance. For example, an echo or reverb effect may be added to the audience audio if the virtual performance venue is a large hall, while a muting effect may be added to the audience audio if the virtual performance is in a small concert venue. Background noise may also be added, such as the noise of a club or bar, if the virtual venue is a club or bar.

[0118] At operation 1404, the audience audio is coordinated with the performance. For example, a human operator may monitor the performance, and coordinate the audience audio based on the actual performance. For example, the human operator may adjust the volume or change another aspect of the synthesized audience audio when appropriate for the given performance. For example, the audience audio may be increased in volume (or enthusiasm, number of voices present, etc.) when an introduction to a well-known song begins. Similarly, the audience audio may be decreased in volume, or even eliminated, when an orchestral performance begins. As another example, the audience audio may be increased in volume (or enthusiasm, number of voices present, etc.) in anticipation of a portion of a song or portion of a song, at an applause break in a performance, or the like.

[0119] Another example of an aspect of the audience audio that may be modified or selected, in real time, includes the type of audience noise. For example, an operator may select whether cheers, shouts, applause, laughter, or other sounds are included in and/or are predominant in the audience audio. In this way, synthesized audience audio may be generated that is contextually relevant to the performance, and is coordinated with the performance in real-time.

[0120] In some cases, all or some of the synthesized audience audio is generated using an automated process (which may be implemented by a computer system). In some cases, the computer system is configured to recognize certain cues, keywords, musical notes or phrases, or other signals in the performance, and, in response, select a certain type of audio content. In some cases, the computer system uses a machine learning model trained on a corpus of performance recordings to determine the content and other parameters for the synthesized audience audio. In such cases, the corpus used to train the model may include recordings of past live performances. The content of the corpus may be limited to a certain category of recordings. For example, if a rock band is performing, the corpus may be recordings of live rock concerts, or recordings of live concerts of the same band that is performing, or recordings of live concerts of multiple musical genres. As another example, if a play is being performed, the corpus may include live recordings of that same play. As yet another example, if a stand-up comedian is performing, the corpus may include live recordings of that same stand-up comedy set.

[0121] At operation 1406, the synthesized audience audio may be output to the performers. For example, the synthesized audience audio may be output via the audience audio output system 110 (FIG. 1). The synthesized audience audio may also be sent to client devices. In such cases, the synthesized audience audio may be mixed with the performance audio to provide a viewing experience that resembles a live performance (e.g.,

where the audience members can hear the audience noise in addition to the performance sound).

[0122] The foregoing description, for purposes of explanation, used specific nomenclature to provide a thorough understanding of the described embodiments. However, 5 it will be apparent to one skilled in the art that the specific details are not required in order to practice the described embodiments. Thus, the foregoing descriptions of the specific embodiments described herein are presented for purposes of illustration and description. They are not targeted to be exhaustive or to limit the embodiments to the precise forms disclosed. It will be apparent to one of ordinary skill in the art that many modifications and 10 variations are possible in view of the above teachings. Also, when used herein to refer to positions of components, the terms above, below, over, under, left, or right (or other similar relative position terms), do not necessarily refer to an absolute position relative to an external reference, but instead refer to the relative position of components within the figure being referred to. Similarly, horizontal and vertical orientations may be understood as 15 relative to the orientation of the components within the figure being referred to, unless an absolute horizontal or vertical orientation is indicated.

[0123] The foregoing embodiments are not exhaustive of the uses of a system as described herein. In particular, in other configurations, a system as described herein can be leveraged to, without limitation: provide an option to display advertisements or other inserted 20 content in between, or throughout, a performance or presentation; provide an option to filter audience/crowd noise, show or hide crowd flags or banners, show or hide audience member comments, etc. available depending on the type of performance (e.g. rock concert, chamber orchestra, educational presentation, and so on); merge multiple virtual venues' video streams together such that multiple performers need not be physically collocated (an 25 audience video streams projected before each artist at each broadcast location may be the same or a different audience merged stream); provide, in a lecture or educational presentation setting, a mechanism for a presenter to see both the student / attendee and that student's workspace (e.g., in a classroom, a teacher can view a student's screen and provide feedback via markup that shows up on their screen, or taking control of that 30 student's device; e.g. in a cooking class, an instructor can view a person's progress and give feedback); provide a system for leveraging an audience member's personal electronic device(s) to provide nonvisual feedback, such as haptic feedback (e.g., trigger a vibration or tactile response in response to crowd noise increasing); provide a means of subtle interaction between a presenter and an attendee by leveraging the attendee's electronic 35 device for personal communications between the presenter and the attendee; (e.g., in a classroom setting, a teacher can nudge a student's devices to make sure the student is

paying attention; e.g., in a dramatic point of a theater performance, a user's device can vibrate to add intensity); provide an interface option to purchase merchandise or food / beverages (by communication with and integration with an audience-member-local delivery or courier service); provide an overlay over a merged crowd image shown/projected to a presenter, the overlay including an audience member's banner or other audience member message on physical or virtual media, animations (e.g., confetti, pyrotechnics, crowd interaction objects like inflatable balls to be volleyed by audience members, synchronized crowd interactions such as the wave or other synchronized motion, fireworks, simulated weather, simulated lighters or cellphone flashlights, simulated air quality or particulates such as smoke and so on); provide a system in which audience members can display a personalized banner or a flag and performer and other attendees can see that display; meter crowd reactions through a sliding scale of like / dislike, funny / not, and so on; provide a system to receive and display comments from a crowd to be voted on, most popular comments can be displayed to a performer(s) (e.g., a political debate can be guided by democratically selected audience questions); provide a software instance to compile a user's movements or dance moves and overlay in a larger or on-stage area next to the performer (e.g., a performer and users can see others moving in this space); provide a system or interface providing real-time close captioning, sign language, translations, and so on; provide an option to overlay a sign-language interpreter over performance video; provide a system for displaying lyrics or other content based on real time action to audience members (all or some); overlay one or more video elements over a stream provided to particular end users (e.g., the performer's stage can be programmed with lights, fireworks, a particular setting or environment or venue), such that the performance itself (including crowd noise and feedback) can vary from audience member to audience member; facilitate audience members capturing images or video or screenshots or other image / screenshot to share in social media; provide a means for an audience member selecting a particular camera angle or set of camera angles to view or cycle through including non-stage camera angles (e.g., backstage, side-stage, artist view, and so on); notify certain audience members that a particular song or particular moment is expected to occur shortly; and so on. A person of skill in the art may readily appreciate that these embodiments are not exhaustive; many configurations of a system described herein can be implemented.

[0124] An embodiment described herein relates to a method for operating a host service. Such host service can be configured to combine multiple video streams to generate an audience stream. The host service receives, from a client device, a video stream that has an associated resolution and a timestamp. The client device is configured to be operated by a user. The host service then generates a reduced resolution video stream with a lower

resolution than its first resolution and the timestamp. The host service may also receive a second video stream from a second client device. Such video stream can include a second resolution and a second timestamp. After receiving the second video stream, the host service may generate a second reduced resolution video stream by reducing the second video's resolution. This reduced resolution may be equivalent to the first reduced resolution video stream. In some examples, the host server receives hundreds or thousands of video streams where the resolution of each video stream is reduced to a predetermined amount.

[0125] Thereafter, the host service may generate an audience video stream by locating a first subject within a first frame of the first reduced resolution video stream, cropping at least one frame of the first reduced resolution video stream to center the first subject to generate a first cropped reduced resolution video stream, locating a second subject within a first frame of the second reduced resolution video stream, cropping at least one frame of the second reduced resolution video stream to center the second subject to generate a second cropped reduced resolution video stream, overlaying the first cropped reduced resolution video stream to a first position within an audience scene, overlaying the second cropped reduced resolution video stream to a second position within the audience scene, and generating the audience video stream from the audience scene. In the audience video stream, the first timestamp and the second timestamp are substantially similar, in some embodiments. Once generated, the audience stream can be transmitted from the host service to a venue server.

[0126] In some examples, the host service receives a performer video stream from a host venue. The performer video stream can include audio, video, and a timestamp associated with the performance. Afterwards, the host service may generate at least two reduced bitrate performer video streams by converting the performer video stream from a first bitrate to a second bitrate. Also, the host service may repackage each reduced bitrate performer video stream from a first format to a plurality of end-user formats. The host service may also be configured to transmit the reduced bitrate performer video stream to the first client device and to the second client device.

[0127] In other embodiments, the host venue captures the performer video stream. In this example, a sound processing device may be communicably connected to at least one camera and at least one microphone. The sound processing device can be configured to compress and encode the performer video stream. Simultaneously, the host venue may display the audience stream to the performer. In this example, timestamp of the performer video stream is substantially similar as the timestamp associated with the audience stream.

[0128] In another example, the host service receives from multiple of client devices multiple streams, each stream including a resolution and a timestamp. The host service may

attribute to each of the video streams a location in an audience scene and a corresponding reduced resolution. Afterwards, the host service can generate multiple reduced resolution video streams by reducing the resolution of each of the video streams to its corresponding resolution. Upon reducing each video stream, the host service may generate an audience video stream by locating each subject within a first frame of each of the reduced resolution video streams, cropping at least one frame of each of the reduced resolution video streams to center each subject in each frame to generate a plurality of cropped reduced resolution video streams, overlaying each cropped reduced resolution video stream to the location in the audience scene corresponding to the reduced resolution of the reduced resolution video stream, generating the audience video stream from the audience scene, and transmitting the audience stream from the host service to a venue server.

[0129] In some examples, the host server may have a plurality of intermediate servers and a plurality of aggregator servers. The host server can be configured to transmit the audience stream to the first client device and to the second client device. In other examples, the host service may remove the background from a subject in at least one frame of the reduced resolution video in order to generate the cropped reduced resolution video.

[0130] An embodiment described herein relates to a system configured to generate and transmit audience video stream content to a performance venue. Such system may have a host server that includes a network resource, a memory allocation storing at least one executable asset, and a processor allocation that may be configured to access the memory allocation to load the executable asset to instantiate an instance of an intermediate server application. This intermediate server application can be configured to communicably couple to a first client device via the network resource; communicably couple to a second client device via the network resource; receive, via the network resource of the host server, a first video stream at a first resolution and at a first timestamp from the first client device; receive, via the network resource of the host server, a second video stream at a second resolution and at a second timestamp from the second client device; generate a first reduced resolution video stream upon reducing the first resolution of the first video stream to a third resolution lower than the first resolution; and generate a second reduced resolution video stream upon reducing the second resolution of the second video stream to the third resolution, where the third resolution is lower than the second resolution. In some examples, the system generates an audience video stream which includes a first subject of the first reduced resolution video stream, a second subject of the second reduced resolution video stream, and an audience view. Each first subject and second subject are located within a first frame of each reduced resolution video, and may be a portion of the reduced resolution video streams. The first subject and second subject can be positioned relative each other within the audience view.

Each cropped reduced resolution video may include a timestamp. Such timestamps are substantially similar to a timestamp associated with the audience video stream. The audience video stream can be transmitted from the host service to a venue server. In one example, the processor allocation is communicatively coupled to a host venue. Such
5 processor allocation may also be configured to receive, via the network resource of the host server, a performer video stream from the host venue, convert the performer video stream from a first format to a second format, and transmit via the network resource the performer video from the host service to the client devices.

[0131] In another example, the aggregator server may also be configured to generate a
10 partial audience video stream. The partial audience video stream can include a first local subject, a second local subject, a partial audience view, and a timestamp. The first local subject of a first locally reduced resolution video stream can be located within a first frame of the locally reduced resolution video stream. The second local subject of a second locally reduced resolution video stream can be located within a second frame of the second locally
15 reduced resolution video stream. Both the first and the second local subjects may be a portion of each of the locally reduced resolution video streams. The first local subject and the second local subject can be positioned within the partial audience view. The timestamp of the partial audience video stream may be substantially similar to the timestamp of the first local video stream and of the second local video stream. In addition, the system may
20 transmit the partial audience video stream to the host server. In another example, a client application instance may be configured to execute at least one computing resource of the client device and display the performer video stream.

CLAIMS

What is claimed is:

1. A method for operating a host service that is configured to combine a plurality of video streams in an audience stream, the method comprising:

5 receiving, at the host service, from a first client device, a first video stream comprising a first resolution and a first timestamp;

generating, at the host service, a first reduced resolution video stream comprising a third resolution and the first timestamp, wherein the third resolution is lower than the first resolution;

10 receiving, at the host service from a second client device, a second video stream comprising a second resolution and a second timestamp;

generating, at the host service, a second reduced resolution video stream comprising the third resolution and the second timestamp, wherein the third resolution is lower than the second resolution;

15 generating an audience video stream by;

locating a first subject within a first frame of the first reduced resolution video stream;

cropping at least one frame of the first reduced resolution video stream to center the first subject to generate a first cropped reduced resolution video stream;

20 locating a second subject within a first frame of the second reduced resolution video stream;

cropping at least one frame of the second reduced resolution video stream to center the second subject to generate a second cropped reduced resolution video stream;

25 overlaying the first cropped reduced resolution video stream to a first position within an audience scene;

overlaying the second cropped reduced resolution video stream to a second position within the audience scene, wherein the first timestamp and the second timestamp are substantially similar; and

30 generating the audience video stream from the audience scene; and transmitting the audience stream from the host service to a venue server.

2. The method of claim 1 further comprising:

receiving, at the host service, a performer video stream from a host venue, the performer video stream comprising audio, video, and a performer timestamp;

35 generating at least two reduced bitrate performer video streams by converting the performer video stream from a first bitrate to a second bitrate;

repackaging each reduced bitrate performer video stream from a first format to a plurality of end-user formats; and

transmitting each reduced bitrate performer video stream to the first client device and to the second client device.

5 3. The method of claim 2, further comprising:
 capturing the performer video stream at the host venue by communicably connecting
 at least one camera and at least one microphone to a sound processing device;
 compressing and encoding the performer video stream; and
 displaying the audience stream to a performer at the host venue, wherein the
10 performer timestamp and an audience stream timestamp are substantially similar.

 4. The method of claim 1 further comprising:
 Receiving, at the host service from a plurality of client devices, the plurality of video
 streams, each video stream comprising a resolution and a timestamp;
 attributing to each of the plurality of video streams a location in the audience scene
15 and a corresponding reduced resolution;
 generating a plurality of reduced resolution video streams by reducing the resolution
 of each of the plurality of video streams to the corresponding reduced resolution;
 generating an audience video stream by;
 locating each subject within a first frame of each of the plurality of reduced
20 resolution video streams;
 cropping at least one frame of each of the plurality of reduced resolution video
 streams to center each subject in each frame to generate a plurality of cropped reduced
 resolution video streams;
 overlaying each cropped reduced resolution video stream to the location in
25 the audience scene corresponding to a reduced resolution of the plurality of reduced
 resolution video streams; and
 generating the audience video stream from the audience scene; and
 transmitting the audience stream from the host service to the venue server.

30 5. The method of claim 1, wherein the host service further comprises a plurality
of intermediate servers and a plurality of aggregator servers.

 6. The method of claim 1, further comprising:
 transmitting the audience stream to the first client device and to the second client
device.

 7. The method of claim 1, further comprising:

removing, from at least one frame of the first reduced resolution video stream, a first background from the first subject on a frame of the first reduced resolution video stream to generate the first cropped reduced resolution video stream; and

removing, from at least one frame of the second reduced resolution video stream, a second background from the second subject on the frame of the second reduced resolution video stream to generate the second cropped reduced resolution video stream.

8. A system configured to generate and transmit audience video stream content to a performance venue, the system comprising:

a host server comprising:

10 a network resource;

a memory allocation storing at least one executable asset;

a processor allocation configured to access the memory allocation to load the at least one executable asset to instantiate an instance of an intermediate server application, the intermediate server application configured to:

15 communicably couple to a first client device via the network resource;

communicably couple to a second client device via the network resource;

receive, via the network resource of the host server, a first video stream at a first resolution and at a first timestamp from the first client device;

20 receive, via the network resource of the host server, a second video stream at a second resolution and at a second timestamp from the second client device;

generate a first reduced resolution video stream upon reducing the first resolution of the first video stream to a third resolution lower than the first resolution;

25 generate a second reduced resolution video stream upon reducing the second resolution of the second video stream to the third resolution, the third resolution lower than the second resolution;

generate an audience video stream, the audience video stream comprising:

30 a first subject of the first reduced resolution video stream within a first frame of the first reduced resolution video stream, wherein the first subject is a portion of the first reduced resolution video stream;

a second subject of the second reduced resolution video stream within a second frame of the second reduced resolution video stream, wherein the second subject is a portion of the second reduced resolution video stream;

35 an audience view, wherein the first subject is positioned relative to the second subject; and

a timestamp of the audience video stream that is substantially similar to a timestamp of the first video stream and of the second video stream; and
transmit the audience video stream from the host server to a venue server.

5 9. The system of claim 8 wherein:
 the processor allocation communicatively couples to a host venue; the processor allocation further configured to:
 receive, via the network resource of the host server, a performer video stream from the host venue;
10 convert the performer video stream from a first format to a second format; and
 transmit via the network resource the performer video stream from the host server to the first and the second client devices.

15 10. The system of claim 9 wherein:
 the processor allocation is further configured to:
 store, via the memory allocation, the performer video stream.

20 11. The system of claim 8 further comprising:
 a plurality of aggregator servers, wherein the host server is communicatively coupled to an aggregator server; and
 each aggregator server comprises:
 an aggregator network resource;
 an aggregator memory allocation storing at least one executable asset;
 an aggregator processor allocation configured to access the memory allocation to load the at least one executable asset to instantiate an instance of an
25 intermediate server application, the intermediate server application configured to:
 communicably couple to a local client device via the network resource;
 receive, via the network resource of the host server, a local video stream having a resolution and a timestamp from the local client device;
 generate a locally reduced resolution video stream upon reducing the
30 resolution of the local video stream to a third resolution lower than the first resolution; and
 transmit the locally reduced resolution video stream to the host server.

35 12. The system of claim 11 wherein:
 the aggregator server is further configured to:
 generate a partial audience video stream, the partial audience video stream comprising:

a first local subject of a first locally reduced resolution video stream within a first frame of the locally reduced resolution video stream, wherein the first local subject is a portion of the locally reduced resolution video stream;

5 a second local subject of a second locally reduced resolution video stream within a second frame of the second locally reduced resolution video stream, wherein the second local subject is a portion of the second locally reduced resolution video stream;

a partial audience view, wherein the first local subject is positioned relative to the second local subject; and

10 a timestamp of the partial audience video stream that is substantially similar to the timestamp of the first locally reduced resolution video stream and of the second locally reduced resolution video stream; and

transmit the partial audience video stream to the host server.

13. The system of claim 9 further comprising:

15 a client application instance configured to execute at least one computing resource of a client device and display the performer video stream.

14. A system configured to operate a host venue, the system comprising:

an audio-visual capture system comprising a camera, a microphone, a memory allocation, and a processor allocation; wherein

20 the camera and the microphone are arranged to capture, from a performance area, a performance video stream; and

the processor allocation is configured to access the memory allocation to load an executable asset to instantiate an instance of an application, the application configured to:

25 encode the performer video stream;

compress the performer video stream; and

transmit the performer video stream to a host server; and

an audience presentation screen communicably coupled to the host server, the audience presentation screen configured to display an audience video stream, the audience video stream comprising:

30 a first cropped reduced resolution video stream comprising an audience resolution and a first timestamp;

a second cropped reduced resolution video stream comprising the audience resolution and a second timestamp; and

35 an overlay in an audience view, generated and transmitted via a host server, of the first cropped reduced resolution video stream relative to the second cropped

reduced resolution video stream, wherein the first timestamp and the second timestamp are substantially similar.

15. The system of claim 14 wherein:
a horizontal distance between a bottom portion of the audience presentation screen
5 and the performance area is less than a horizontal distance between a top portion of the
audience presentation screen and the performance area.

16. The system of claim 15 wherein:
the resolution of a cropped reduced resolution stream decreases as a distance
between the performance area and a location of the cropped reduced resolution stream
10 within the audience video stream increases.

17. The system of claim 14 further comprising:
an audience output system comprising at least one speaker and a motor system,
upon meeting a noise criteria, the at least one speaker is configured to output
an audience feedback sound; and
15 upon meeting a predetermined criteria, the motor system is configured to
output to the performance area a vibration.

18. The system of claim 17 wherein:
the audience feedback sound comprises a pre-recorded audience noise.

19. The system of claim 17 wherein:
20 the audience output system further comprises an executable asset configured to:
upon a performer selection of a first user within the audience video stream,
retrieve, from the host server, an original video stream;
increase, within the audience video stream, the audience resolution of the first
user cropped reduced resolution video stream from a third resolution to a first resolution; and
25 display, on the audience video stream, the original video stream of the first
user.

20. The system of claim 14 wherein:
a maximum number of users within the audience video stream is substantially
equivalent to a maximum capacity of a predetermined physical venue.

30

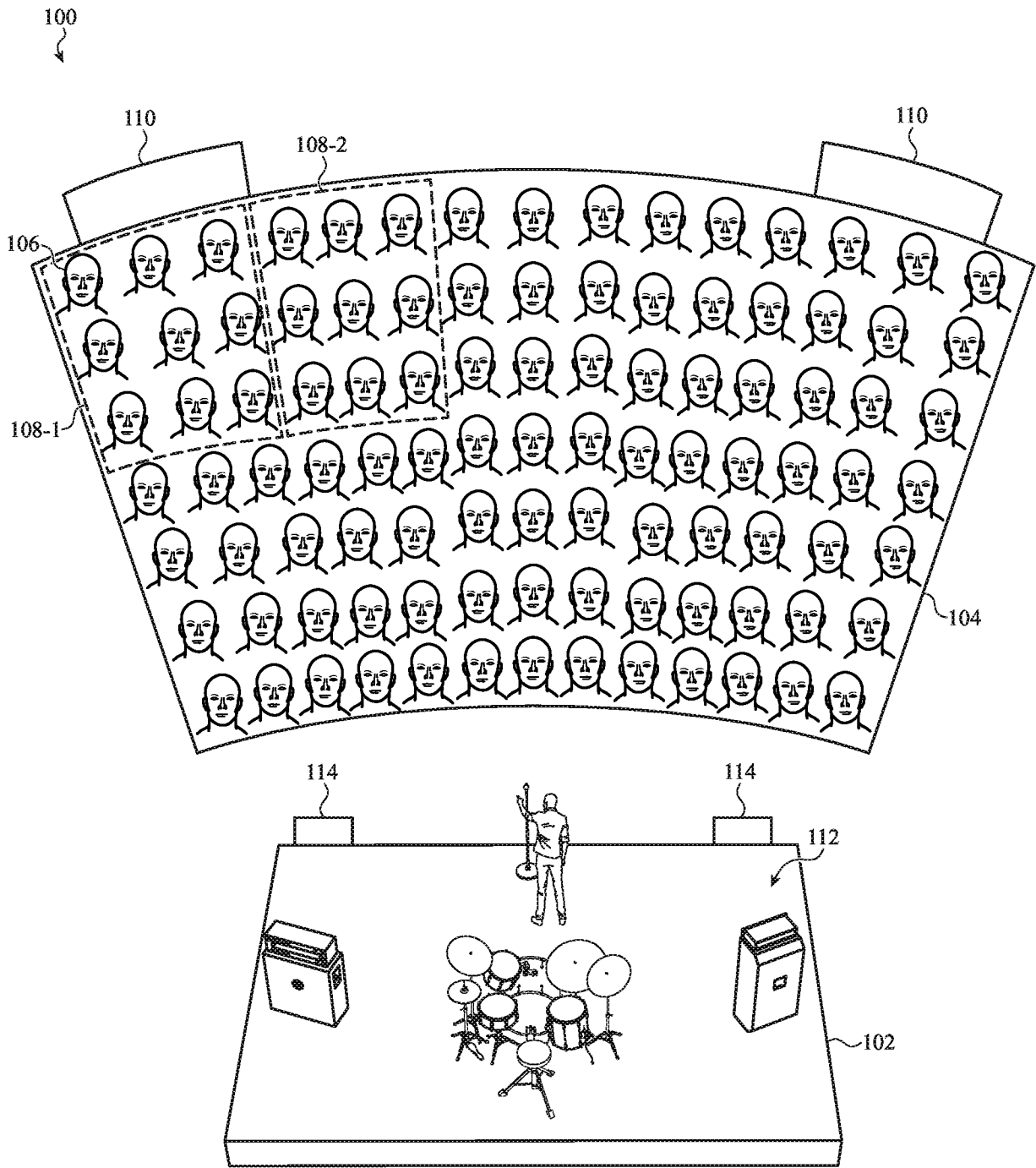


FIG. 1

2/14

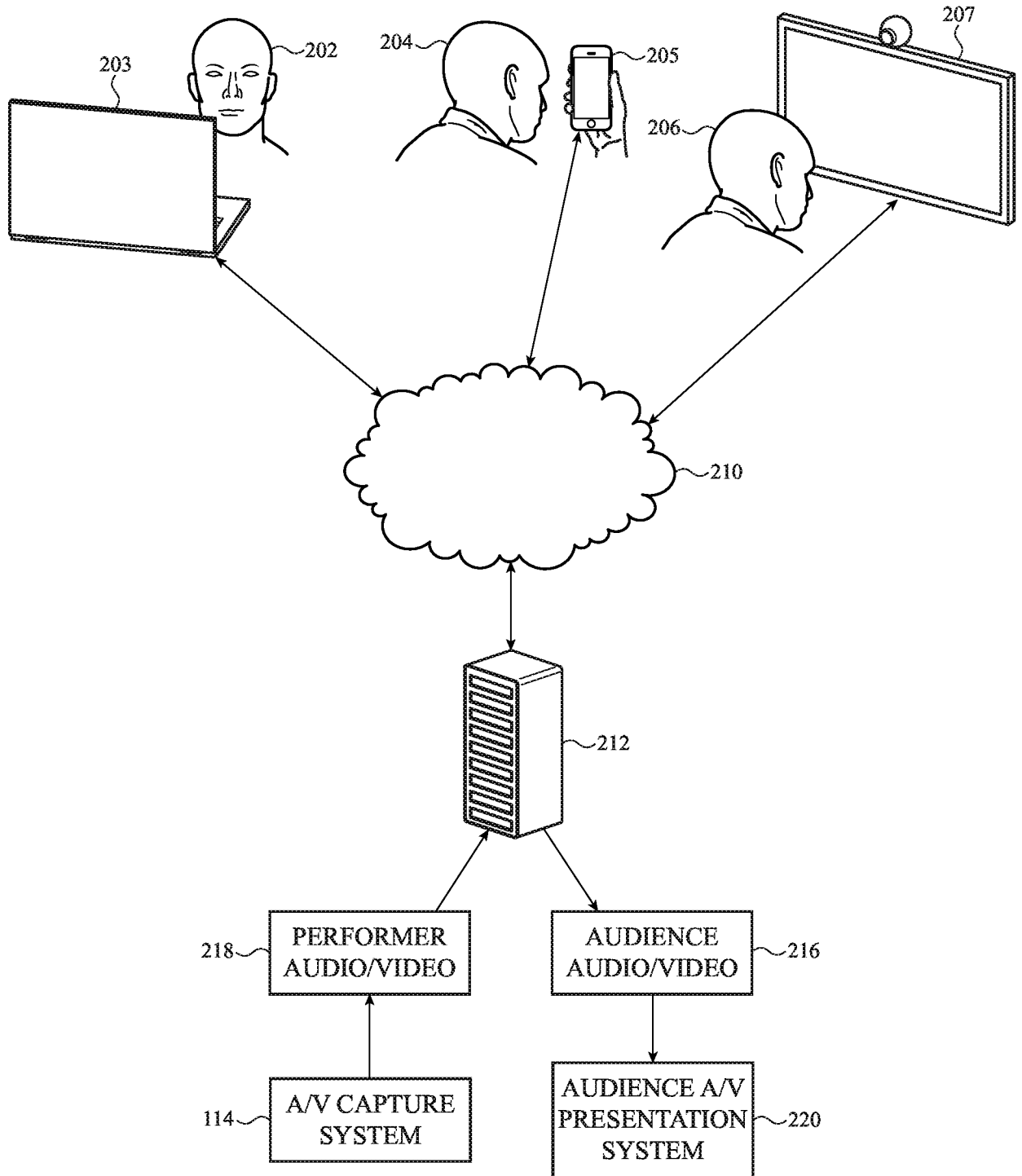


FIG. 2

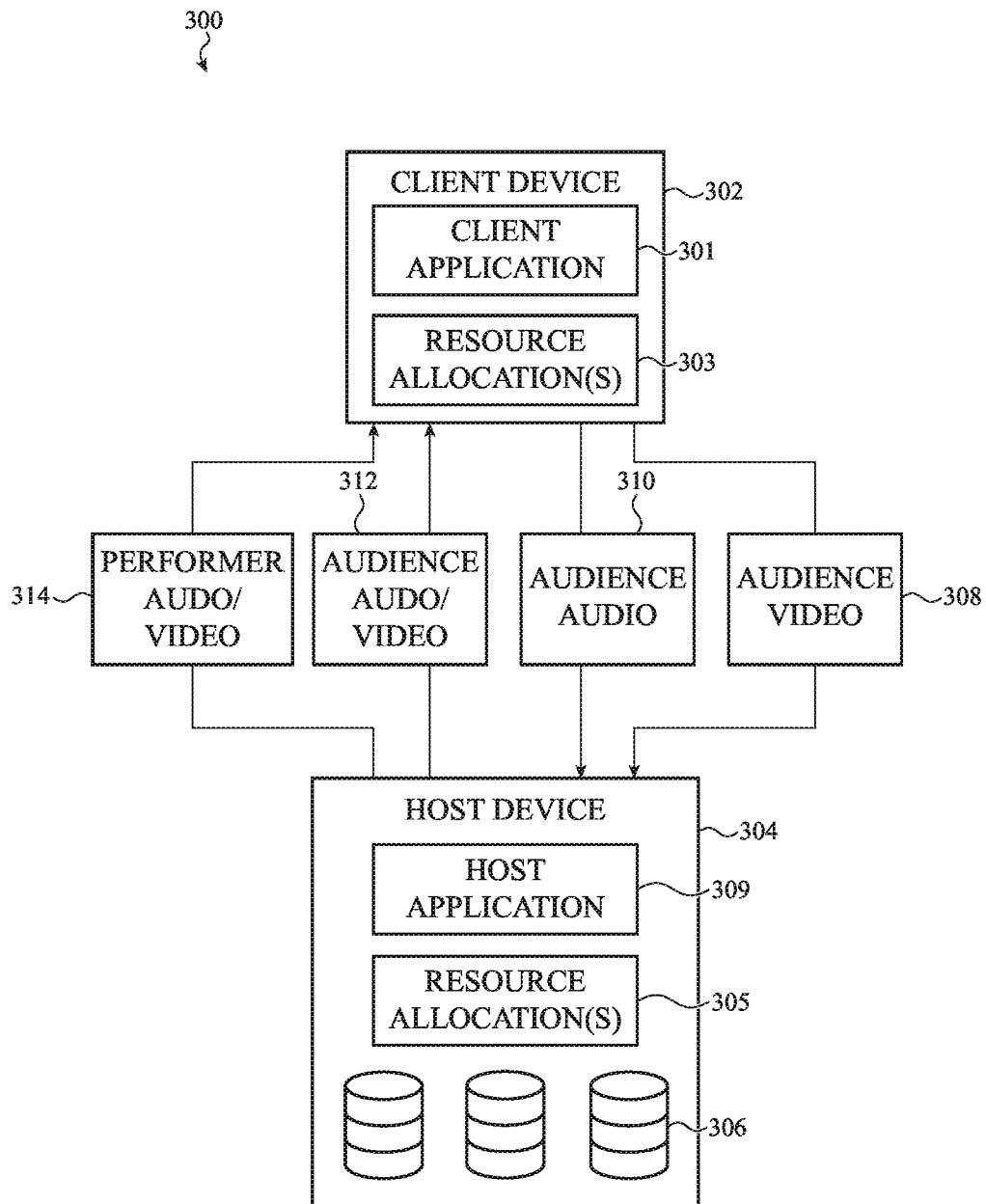


FIG. 3

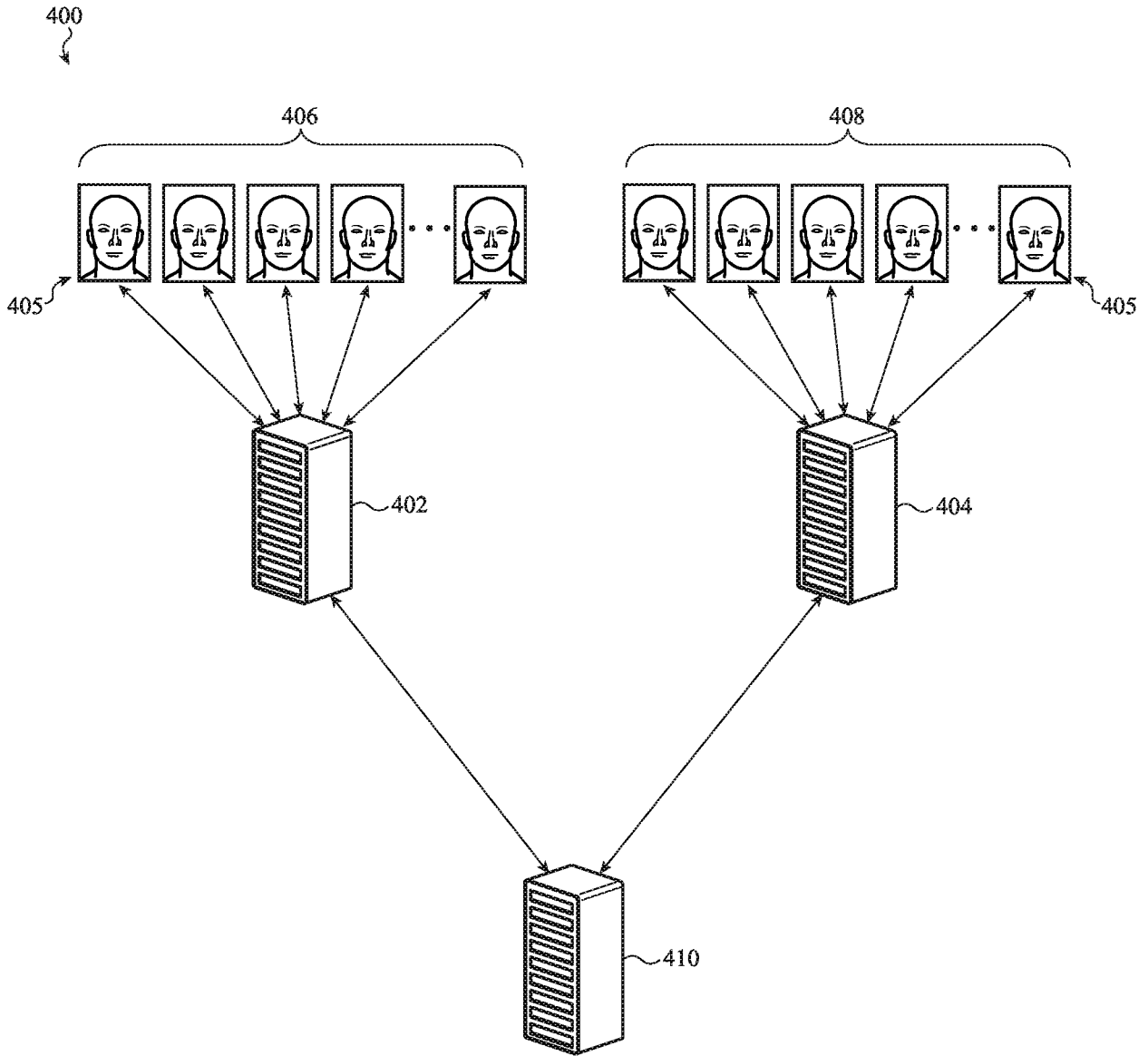


FIG. 4

500

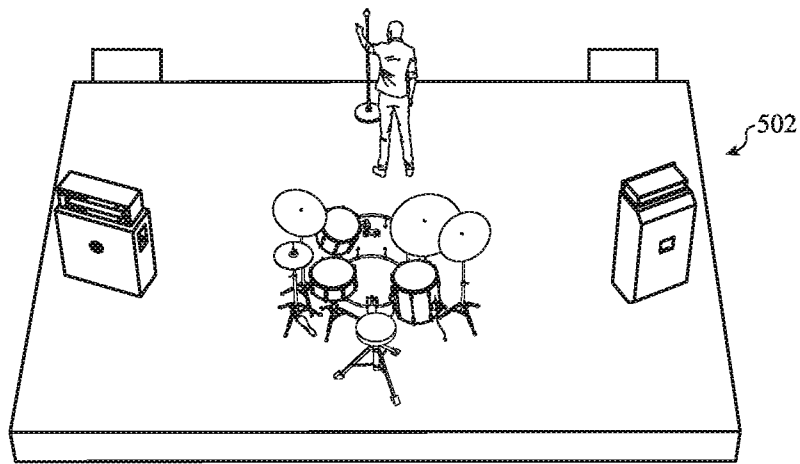
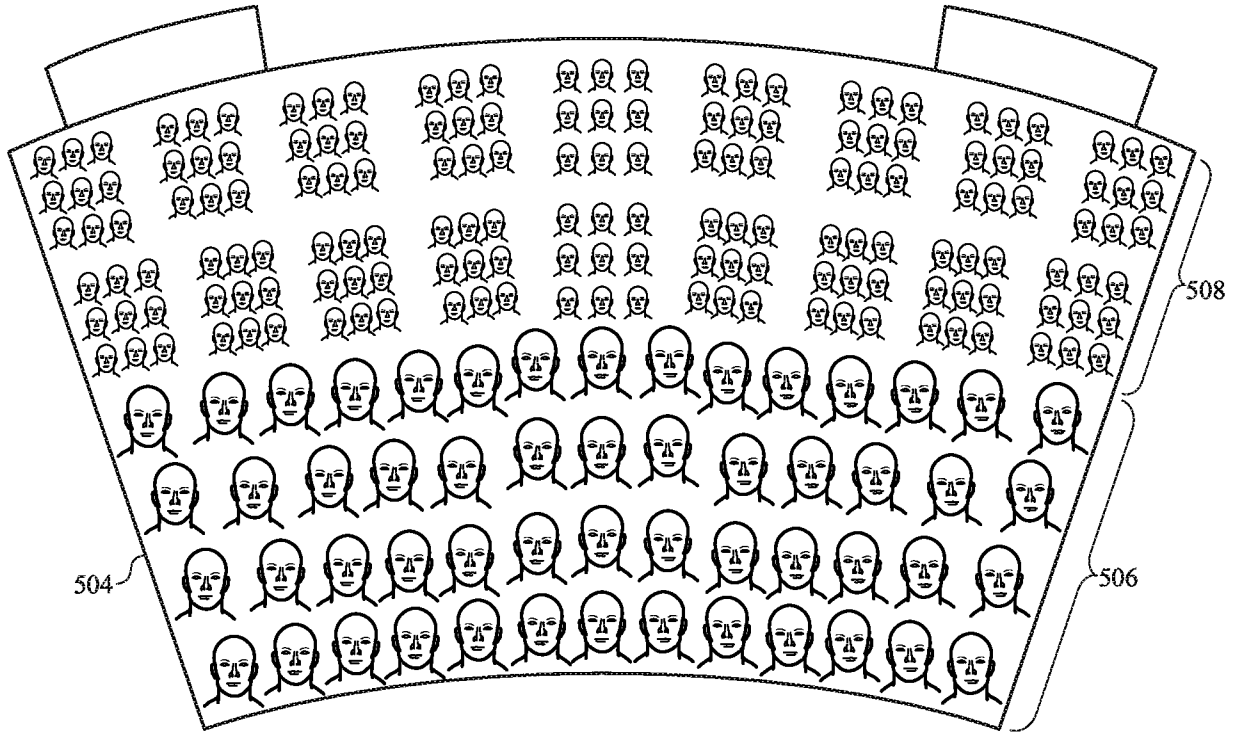


FIG. 5

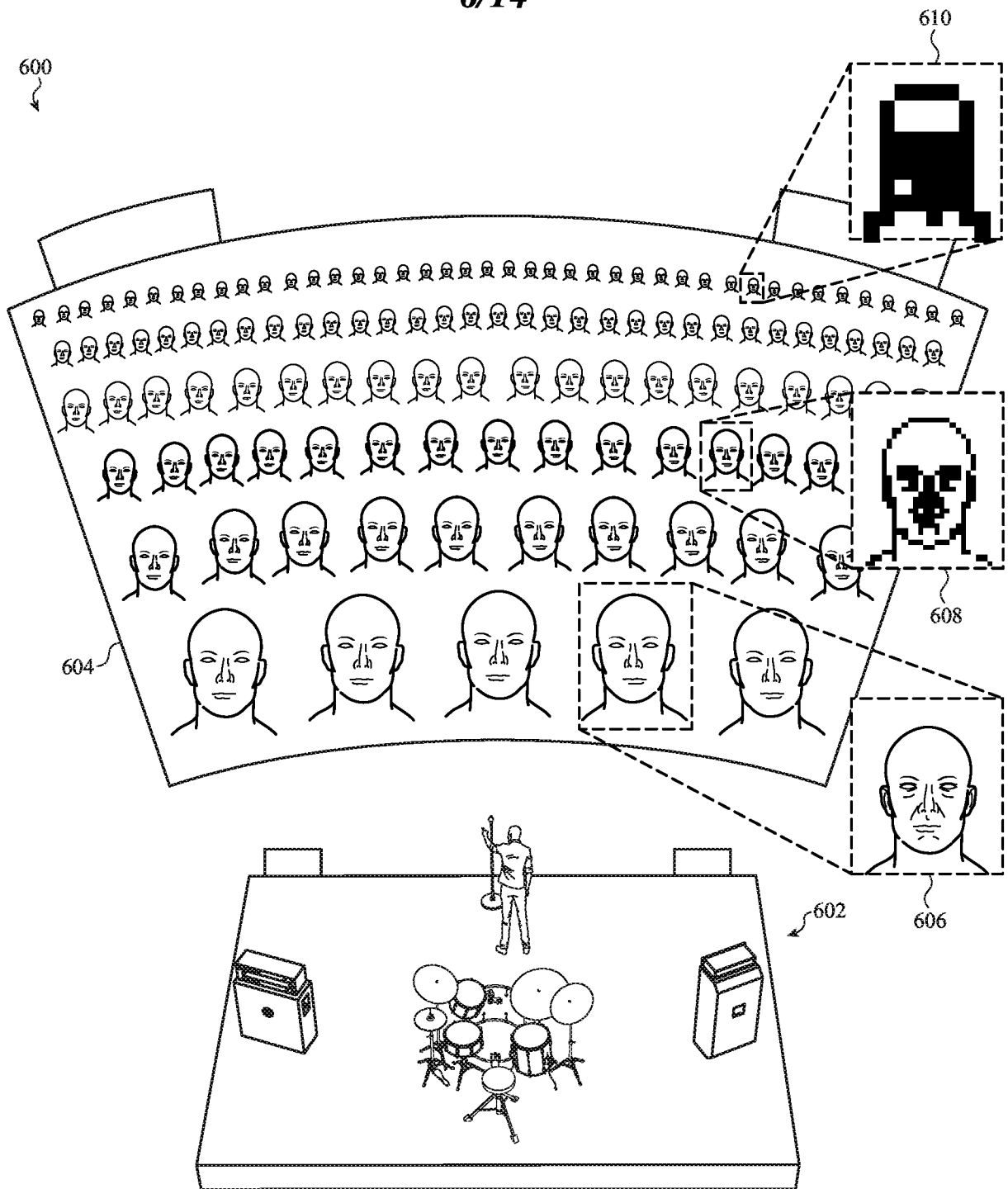


FIG. 6

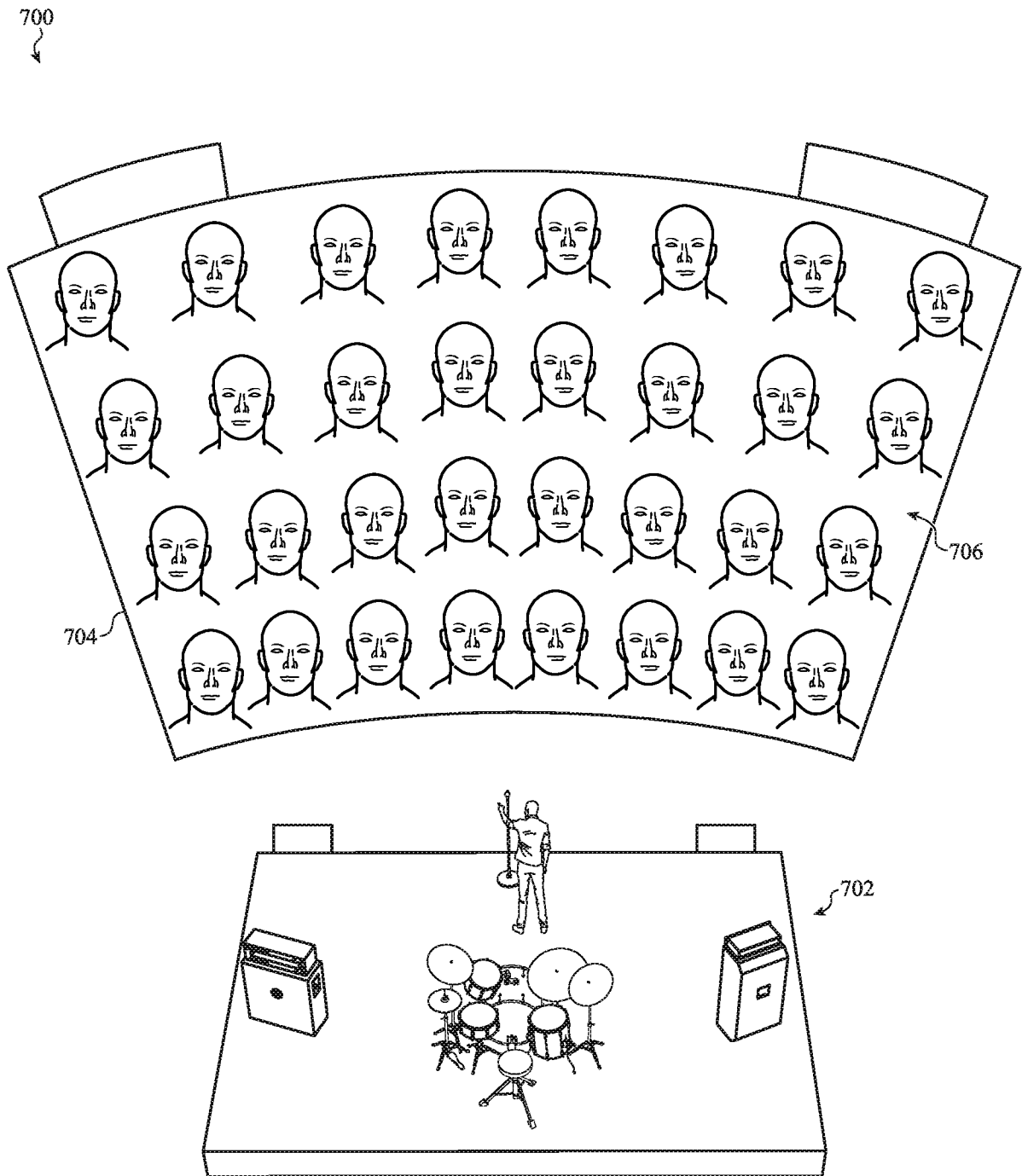


FIG. 7

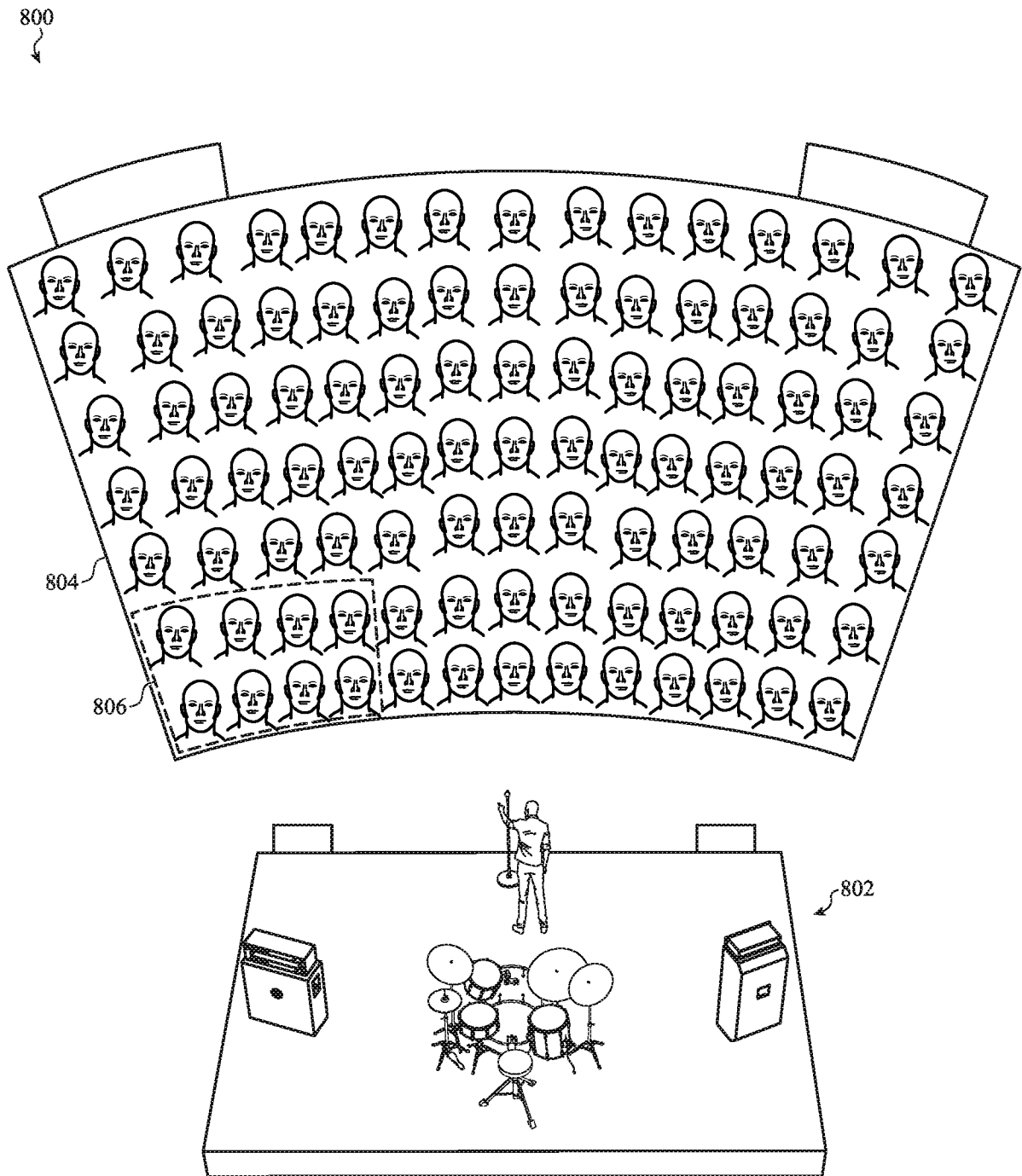


FIG. 8

9/14

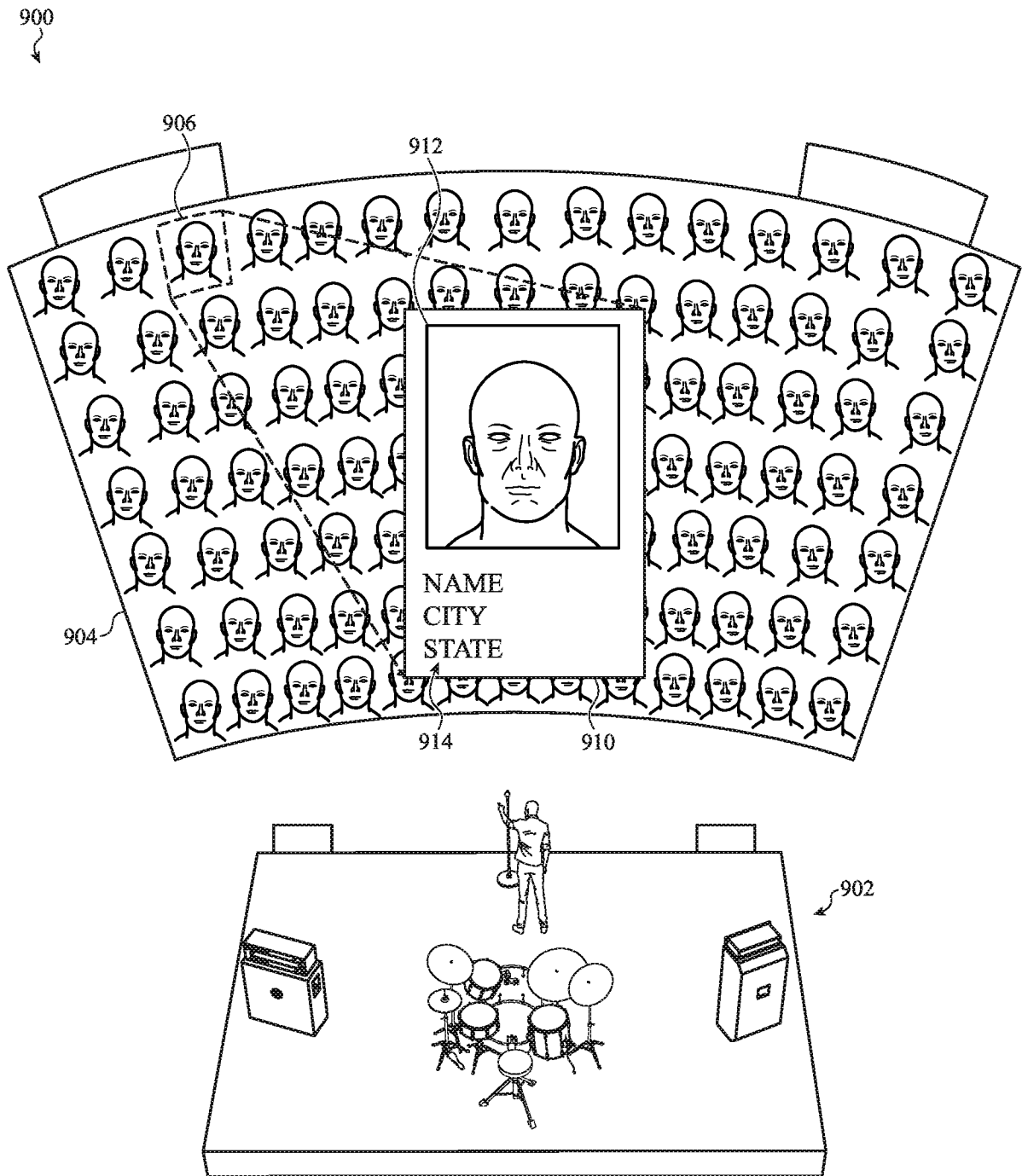


FIG. 9

10/14

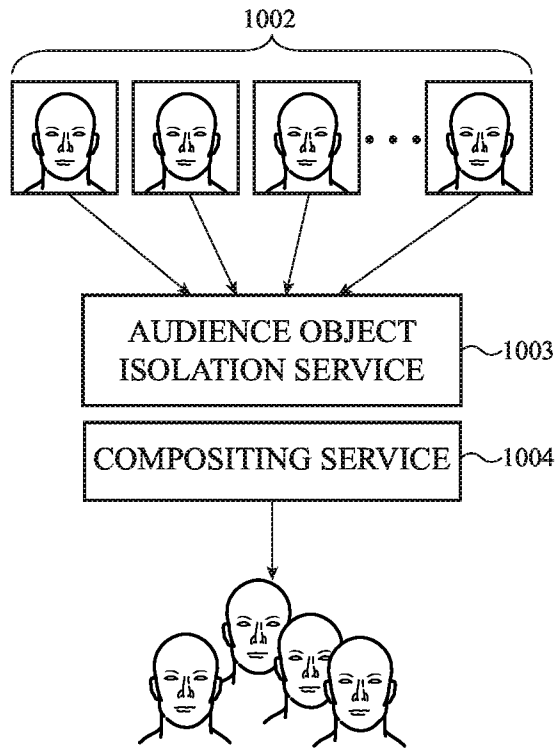


FIG. 10A

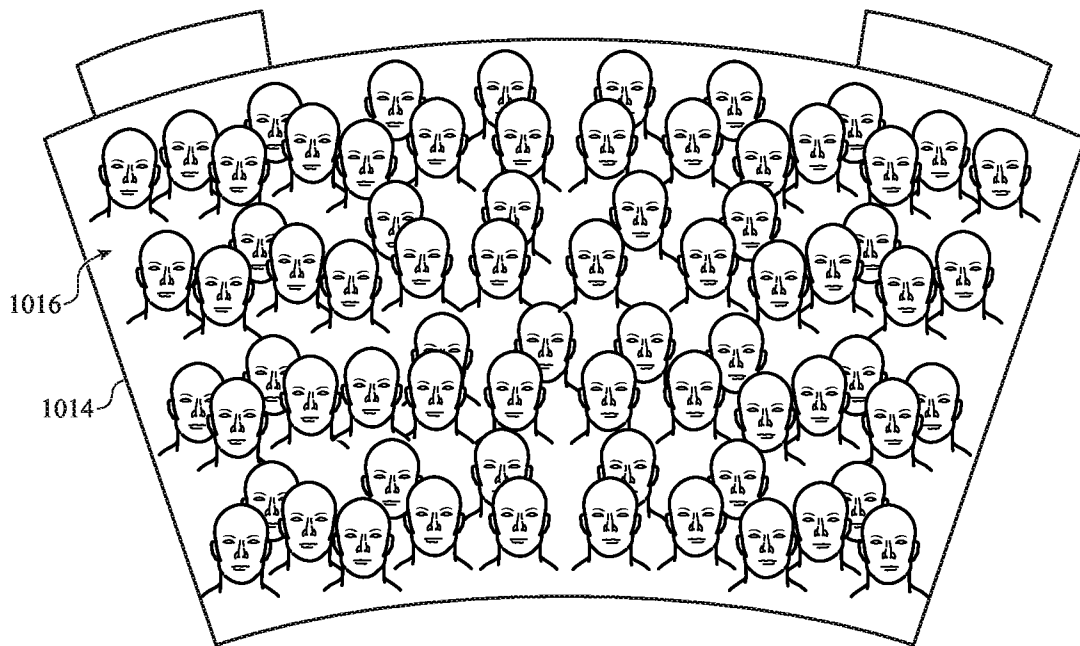


FIG. 10B

11/14

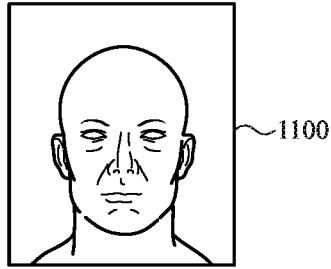


FIG. 11A

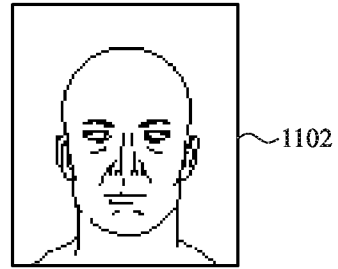


FIG. 11B

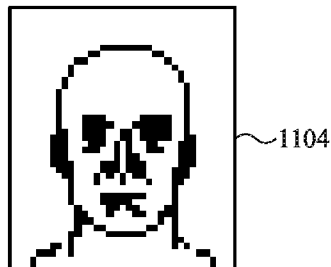


FIG. 11C

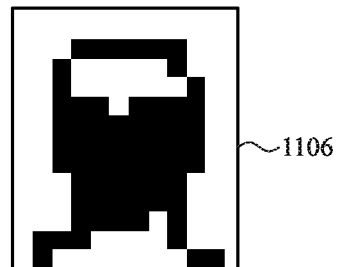


FIG. 11D

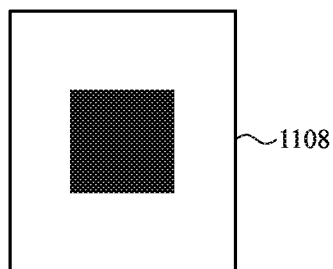


FIG. 11E

12/14

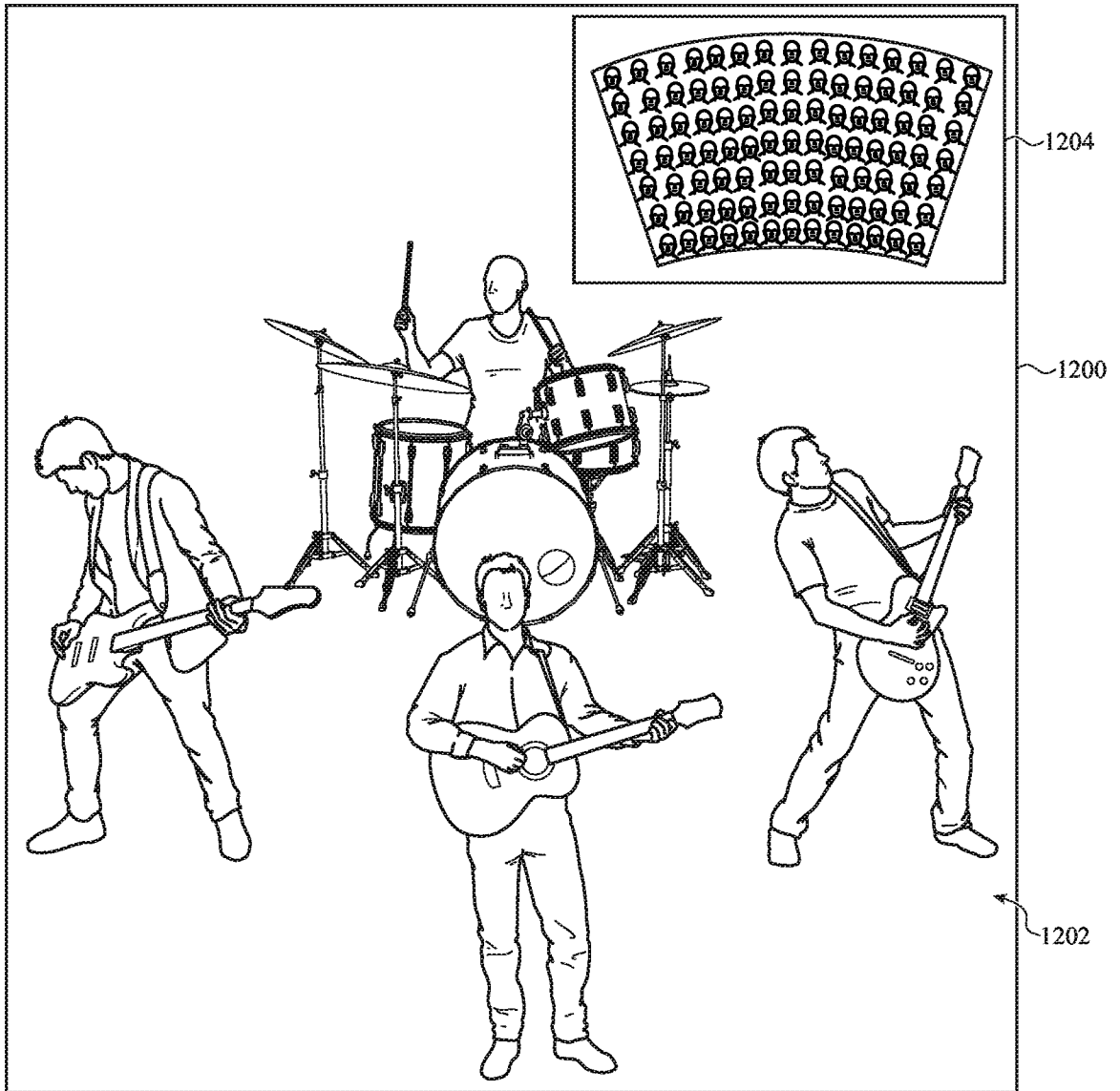


FIG. 12

13/14

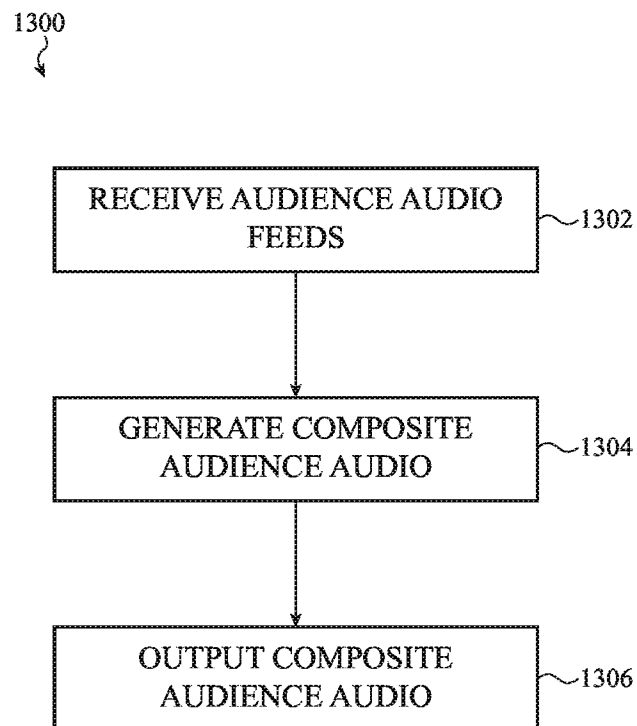


FIG. 13

14/14

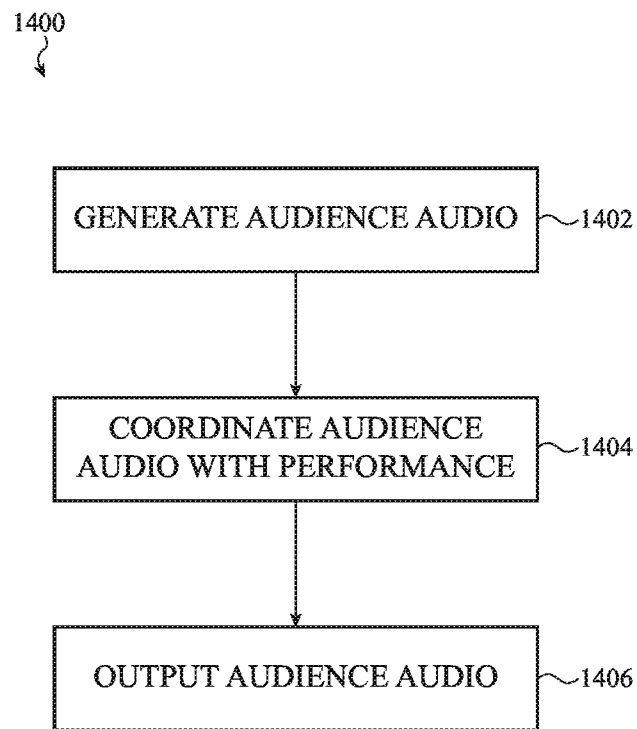


FIG. 14

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2021/043246

| | | | | | |
|---|-------------|------------|------------|------------|------------|
| A. CLASSIFICATION OF SUBJECT MATTER | | | | | |
| INV. | H04N21/2187 | H04N21/234 | H04N21/236 | H04N21/24 | H04N21/242 |
| | H04N21/2665 | H04N21/437 | H04N21/442 | H04N21/658 | H04N5/222 |
| 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) H04N |
| Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched |
| Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) 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. |
| Y | US 2015/244987 A1 (DELEGUE GERARD [FR] ET AL) 27 August 2015 (2015-08-27) paragraphs [0042], [0007], [0057]; figure 1 paragraph [0060]; figures 2, 3,4 paragraphs [0017], [0023] paragraph [0050] - paragraph [0052]; figure 2 | 1-20 |
| Y | US 2009/067349 A1 (GLUECKMAN ALAN JAY [US] ET AL) 12 March 2009 (2009-03-12) paragraphs [0021], [0055]; figure 2 | 5,8-13 |
| A | | 1-4,6,7, 14-20 |
| | ----- -/-- | |

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 | "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 |
| "E" earlier application or patent but published on or after the international filing date | "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 |
| "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) | "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 |
| "O" document referring to an oral disclosure, use, exhibition or other means | "&" document member of the same patent family |
| "P" document published prior to the international filing date but later than the priority date claimed | |

| | |
|--|--|
| Date of the actual completion of the international search 8 September 2021 | Date of mailing of the international search report 20/09/2021 |
| 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 Ong, Bee |

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2021/043246

| C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT | | |
|--|--|--------------------------------|
| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
| Y A | <p>US 2014/007147 A1 (ANDERSON GLEN J [US]) 2 January 2014 (2014-01-02) paragraph [0021]; figure 1 paragraphs [0043], [0044], [0057]; figure 5 paragraphs [0060], [0067] paragraph [0023] paragraphs [0017], [0043], [0056], [0023] paragraph [0047] - paragraph [0048] paragraph [0059]; figure 3 & 4 paragraph [0034]</p> <p style="text-align: center;">-----</p> | <p>1-7, 13-20 8-12</p> |

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2021/043246

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|--|------------------|-------------------------|------------------|
| US 2015244987 | A1 | 27-08-2015 | CN 104685858 A |
| | | | EP 2713593 A1 |
| | | | KR 20150048821 A |
| | | | US 2015244987 A1 |
| | | | WO 2014048686 A1 |
| ----- | | | |
| US 2009067349 | A1 | 12-03-2009 | NONE |
| ----- | | | |
| US 2014007147 | A1 | 02-01-2014 | EP 2867849 A1 |
| | | | US 2014007147 A1 |
| | | | WO 2014003965 A1 |
| ----- | | | |