

**(12) STANDARD PATENT
(19) AUSTRALIAN PATENT OFFICE**

(11) Application No. AU 2015236755 B2

(54) Title
Metadata for ducking control

(51) International Patent Classification(s)
G10L 19/008 (2013.01)

(21) Application No: **2015236755** (22) Date of Filing: **2015.02.26**

(87) WIPO No: **WO15/148046**

(30) Priority Data

(31) Number (32) Date (33) Country
61/970,284 **2014.03.25** **US**
14/613,203 **2015.02.03** **US**

(43) Publication Date: **2015.10.01**
(44) Accepted Journal Date: **2017.08.31**

(71) Applicant(s)
Apple Inc.

(72) Inventor(s)
Holman, Tomlinson M.;Baumgarte, Frank M.;Allamanche, Eric A.

(74) Agent / Attorney
FPA Patent Attorneys Pty Ltd, ANZ Tower 161 Castlereagh Street, Sydney, NSW, 2000, AU

(56) Related Art
WO 2012/138594 A1
WO 2011/044153 A1

(51) International Patent Classification:
G10L 19/008 (2013.01)

LAMANCHE, Eric A.; 1 Infinite Loop, Mail Stop 38-21IMG, Cupertino, California 95014 (US).

(21) International Application Number:
PCT/US2015/017830

(74) Agents: SCHELLER, James C. et al.; Blakely, sokoloff, taylor & Zafman LLP, 1279 Oakmead Parkway, Sunnyvale, California 94085 (US).

(22) International Filing Date:
26 February 2015 (26.02.2015)

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

(25) Filing Language: English

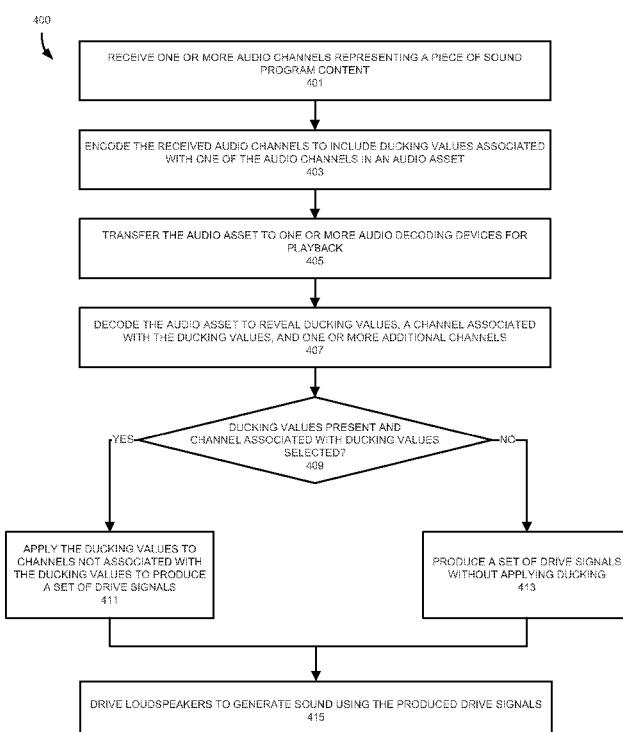
(26) Publication Language: English

(30) Priority Data:
61/970,284 25 March 2014 (25.03.2014) US
14/613,203 3 February 2015 (03.02.2015) US(71) Applicant: **APPLE INC.** [US/US]; 1 Infinite Loop, Cupertino, California 95014 (US).(72) Inventors: **HOLMAN, Tomlinson M.**; 1 Infinite Loop, Mail Stop 81-2PT, Cupertino, California 95014 (US). **BAUMGARTE, Frank M.**; 1 Infinite Loop, Mail Stop 38-21IMG, Cupertino, California 95014 (US). AL-

(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,

[Continued on next page]

(54) Title: METADATA FOR DUCKING CONTROL



(57) Abstract: An audio encoding device and an audio decoding device are described herein. The audio encoding device may examine a set of audio channels/channel groups representing a piece of sound program content and produce a set of ducking values to associate with one of the channels/channel groups. During playback of the piece of sound program content, the ducking values may be applied to all other channels/channel groups. Application of these ducking values may cause (1) the reduction in dynamic range of ducked channels/channel groups and/or (2) movement of channels/channel groups in the sound field. This ducking may improve intelligibility of audio in the non-ducked channel/channel group. For instance, a narration channel/channel group may be more clearly heard by listeners through the use of selective ducking of other channels/channel groups during playback.

FIG. 4



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))*

14 Jul 2017

2015236755

METADATA FOR DUCKING CONTROL

RELATED MATTERS

[0001] This application claims the benefit of the earlier filing date of U.S. provisional application no. 61/970,284, filed March 25, 2014.

FIELD

[0002] A system and method are described for adding ducking values into an audio asset representing a piece of sound program content. The ducking values may be associated with a particular channel/channel group and may be used for ducking every other channel/channel group in the piece of sound program content except for the associated channel/channel group. Other embodiments are also described.

BACKGROUND

[0003] An audio channel may be selectively toggled "on" and "off" by a listener. For example, a movie or television show may include a narration channel that may be selected by a user. The narration channel may provide audio descriptions of visual actions and activities taking place within a corresponding video. Accordingly, this narration channel may be beneficial to users that are visually impaired.

[0004] Although narration may allow a visually impaired user to better understand the visual actions and activities taking place in a video, directly playing this narration channel over other audio channels may compromise overall intelligibility of these channels. For example, main dialogue channels may interfere with narration channels when both of these channels are played simultaneously.

[0005] The approaches described in this section are approaches that could be pursued, but not necessarily approaches that have been previously conceived or pursued. Therefore, unless otherwise indicated, it should not be assumed that any of the approaches described in this section qualify as prior art merely by virtue of their inclusion in this section.

SUMMARY

[0005a] As used herein, except where the context requires otherwise, the term "comprise" and variations of the term, such as "comprising", "comprises" and "comprised", are not intended to exclude further additives, components, integers or steps.

[0005b] According to a first aspect of the invention there is provided a method for encoding audio content associated with video content, the method comprising: receiving a piece of sound program content comprising (i) a first channel group, object or stem that comprises visually

descriptive audio content that describes actions taking place in the video content, (ii) a second channel group, object or stem that comprises at least one of dialogue, music, and effects, and (iii) a third channel group, object or stem that comprises at least another one of dialogue, music, and effects; determining that the first channel group, object or stem in the piece of sound program content is to be emphasized during playback of the piece of sound program content; generating (i) a first set of ducking values to apply to the second channel group, object or stem and (ii) a second set of ducking values to apply to the third channel group, object or stem, based on activity within the first channel group, object or stem, wherein the first set of ducking values is different than the second set of ducking values such that when the first and second sets of ducking values are applied to the second and third channel groups, objects or stems, respectively, the second and third channel groups, objects or stems are deemphasized differently during playback of the piece of sound program content; and associating the first and second sets of ducking values with the first channel group, object or stem in an audio asset, wherein the audio asset includes (i) the first, second, and third channel groups, objects or stems and (ii) the first and second sets of ducking values that are associated with the first channel group, object or stem.

[0005c] According to a second aspect of the invention there is provided a method for playing back audio content associated with video content, the method comprising: receiving an audio asset representing a piece of sound program content comprising (i) a first channel group, object or stem that comprises visually descriptive audio content that describes actions taking place in the video content, (ii) a second channel group, object or stem that comprises at least one of dialogue, music, and effects, (iii) a third channel group, object or stem that comprises at least another one of dialogue, music, and effects than the second channel group, object or stem, and (iv) a first set of ducking values and a second set of ducking values associated with the first channel group, object or stem, wherein the first set of ducking values is different than the second set of ducking values and wherein the first and second sets of ducking values were generated based on activity in the first channel group, object or stem; extracting the first and second sets of ducking values along with the (i) first channel group, object or stem (ii) second channel group, object or stem, and (iii) third channel group, object or stem from the audio asset; and during playback of the piece of sound program content through a plurality of loudspeakers applying the first set of ducking values to the second channel group, object or stem; and applying the second set of ducking values to the third channel group, object or stem, wherein applying the first and second sets of ducking values deemphasizes the second channel group, object or stem differently than the third channel group, object or stem.

[0005d] According to a third aspect of the invention there is provided a non-transitory computer readable medium containing instructions for encoding audio content associated with video content, which when executed by a processor in an audio device cause the audio device to: receiving a piece of sound program content comprising (i) a first channel group, object or stem that comprises visually descriptive audio content that describes actions taking place in the video content, (ii) a second channel group, object or stem that comprises at least one of dialogue, music, and effects, and (iii) a third channel group, object or stem that comprises at least another one of dialogue, music, and effects; analyze the piece of sound program content to determine that the first channel group, object or stem in the piece of sound program content is to be emphasized during playback of the piece of sound program content; generate (i) a first set of ducking values to apply to the second channel group, object or stem and (ii) a second set of ducking values to apply to the third channel group, object or stem, based on activity within the first channel group, object or stem, wherein the first set of ducking values is different than the second set of ducking values such that when the first and second sets of ducking values are applied to the second and third channel groups, objects or stems, respectively, the second and third channel groups, objects or stems are deemphasized differently during playback in favor of the first channel group; and associate the ducking values with the first channel group, object or stem in an audio asset, wherein the audio asset includes (i) the first, second, and third channel groups, objects or stems and (ii) the first and second sets of ducking values that are associated with the first channel group, object or stem.

[0005e] According to a fourth aspect of the invention there is provided an audio device for playing back audio content associated with video content, the audio device comprising a hardware processor; and a memory unit storing instructions executed by the hardware processor that cause the audio device to: receive an audio asset representing a piece of sound program content comprising (i) a first channel group, object or stem that comprises visually descriptive audio content that describes actions taking place in the video content, (ii) a second channel group, object or stem that comprises at least one of dialogue, music, and effects, (iii) a third channel group, object or stem that comprises at least another one of dialogue, music, and effects than the second channel group, object or stem, and (iv) a first set of ducking values and a second set of ducking values associated with the first channel group, object or stem, wherein the first set of ducking values is different than the second set of ducking values and wherein the first and second sets of ducking values were generated based on activity in the first channel group, object or stem; extract the first and second sets of ducking values along with the (i) first channel group, object or stem, (ii) second channel group, object or stem, and (iii) third channel group, object or

stem from the audio asset; and during playback of the piece of sound program content through a plurality of loudspeakers apply the first set of ducking values to the second channel group, object or stem; and apply the second set of ducking values to the third channel group, object or stem, wherein applying the first and second sets of ducking values deemphasizes the second channel group, object or stem differently than the third channel group, object or stem.

[0006] An audio encoding device and an audio decoding device are described herein. The audio encoding device may examine a set of audio channels corresponding to a piece of sound program content and produce a set of ducking values to associate with one of the channels or a channel group. For example, the audio channels may include a narration channel that assists visually impaired users to understand actions taking place in a movie or television broadcast. For instance, the narration channel may include an audio description of movement of characters

in video or other actions that would not be evident to visually impaired users. The ducking values may be associated with this narration channel in an audio asset representing the encoded piece of sound program content.

[0007] The audio asset may be transferred to one or more audio decoding devices for playback through a set of loudspeakers. In one embodiment, the ducking values may reduce the dynamic range of all other channels apart from the narration channel in the piece of sound program content during playback. Accordingly, the narration channel may be emphasized through the reduction in range of the other channels being played back. In one embodiment, the ducking values may vary over time such that ducking is performed during periods in which activity is detected in the narration channel and not during periods of inactivity (e.g., voice/speech activity).

[0008] In some embodiments, ducking may be applied separately to audio channels/channel groups. For example, a dialogue channel may be more severely ducked than music/effects channels. This variability allows the improvement of intelligibility of a narration channel without the need to duck other channels that may not as harshly affect intelligibility of the narration channel.

[0009] In some embodiments, ducking values may also cause one or more channels not associated with the ducking values to be rendered at a different location within the sound field. For instance, ducking values may cause channels to be rendered by the surround loudspeakers during playback. Along with ducking and reducing the dynamic range of the channels, this modification of the rendering location may improve the intelligibility of another channel (e.g., a narration channel).

[0010] As described above, ducking values may be included in an encoded piece of sound program content and associated with a particular channel/channel group. During playback of the encoded piece of sound program content, the ducking values may be applied to all other channels/channel groups not associated with the ducking values. This ducking may improve intelligibility of audio in the non-ducked channel/channel group. For instance, as described above, a narration channel/channel group may be more clearly heard by listeners through the use of selective ducking of other channels/channel groups during playback.

[0011] The above summary does not include an exhaustive list of all aspects of the present invention. It is contemplated that the invention includes all systems and methods that can be practiced from all suitable combinations of the various aspects summarized above, as well as those disclosed in the Detailed Description below and particularly pointed out in the claims filed

with the application. Such combinations have particular advantages not specifically recited in the above summary.

BRIEF DESCRIPTION OF THE DRAWINGS

[0012] The embodiments of the invention are illustrated by way of example and not by way of limitation in the figures of the accompanying drawings in which like references indicate similar elements. It should be noted that references to "an" or "one" embodiment of the invention in this disclosure are not necessarily to the same embodiment, and they mean at least one.

[0013] **Figure 1** shows an audio system, including an audio encoding device and a set of audio decoding devices, according to one embodiment.

[0014] **Figure 2** shows a component diagram of the audio encoding device according to one embodiment.

[0015] **Figure 3** shows a component diagram of the audio decoding device according to one embodiment.

[0016] **Figure 4** shows a method according to one embodiment for (1) encoding a multi-channel piece of sound program content to include ducking values associated with a first channel/channel group and (2) playing back the encoded multi-channel piece of sound program content such that a second channel/channel group is ducked in favor of the first channel/channel group that is associated with the ducking values.

DETAILED DESCRIPTION

[0017] Several embodiments are described with reference to the appended drawings. While numerous details are set forth, it is understood that some embodiments of the invention may be practiced without these details. In other instances, well-known circuits, structures, and techniques have not been shown in detail so as not to obscure the understanding of this description.

[0018] **Figure 1** shows an audio system 100 according to one embodiment. The audio system 100 may include an audio encoding device 101 and a set of audio decoding devices 103₁-103_N. The audio encoding device 101 and the audio decoding devices 103₁-103_N may be communicatively coupled through the distributed network 105. In particular, the audio encoding device 101 may encode one or more pieces of sound program content and transfer the encoded data to one or more of the audio decoding devices 103₁-103_N over the distributed network 105.

[0019] In one embodiment, the audio encoding device 101 may encode a multi-channel piece of sound program content. As will be described in greater detail below, the encoded data

produced by the audio encoding device 101 may include metadata for separate channels or channel groups comprising the piece of sound program content. In particular, metadata for a particular channel/channel group may indicate that all other channels/channel groups should be “ducked” (i.e., reduced in intensity/volume) for a specified duration/period of time. This ducking may allow the channel/channel group that is not being ducked (i.e., channel/channel group associated with the ducking values) to be more clearly heard/understood by user/listeners.

[0020] Each element of the audio system 100 will now be described by way of example. In other embodiments, the audio system 100 may include more elements than those shown in **Figure 1** and described herein.

[0021] **Figure 2** shows a component diagram of the audio encoding device 101 according to one embodiment. The audio encoding device 101 may be any computing device that is capable of encoding a multi-channel piece of sound program content. For example, the audio encoding device 101 may be a laptop computer, a desktop computer, a computer server, a tablet computer, a gaming system, and/or a mobile device (e.g., cellular telephone or mobile media player). Each element of the audio encoding device 101 shown in **Figure 2** will now be described.

[0022] The audio encoding device 101 may include a main system processor 201 and a memory unit 203. The processor 201 and memory unit 203 are generically used here to refer to any suitable combination of programmable data processing components and data storage that conduct the operations needed to implement the various functions and operations of the audio encoding device 101. The processor 201 may be a special purpose processor such as an application-specific integrated circuit (ASIC), a general purpose microprocessor, a field-programmable gate array (FPGA), a digital signal controller, or a set of hardware logic structures (e.g., filters, arithmetic logic units, and dedicated state machines) while the memory unit 203 may refer to microelectronic, non-volatile random access memory.

[0023] An operating system may be stored in the memory unit 203, along with application programs specific to the various functions of the audio encoding device 101, which are to be run or executed by the processor 201 to perform the various functions of the audio encoding device 101. For example, the memory unit 203 may include an encoder 205, which in conjunction with other hardware and software elements of the audio encoding device 101, encodes a multi-channel piece of sound program content. As noted above and as will be described in further detail below, the encoded data produced by the encoder 205 may include metadata for separate channels or channel groups of the piece of sound program content. The metadata may indicate ducking values to be applied by one or more of the audio decoding devices 103₁-103_N to other channels/channel groups not associated with the metadata. In particular, as will be described in

greater detail below, the ducking values may be used by one or more of the audio decoding devices 103₁-103_N to duck other channels/channel groups to improve the intelligibility to users of the non-ducked channels/channel groups.

[0024] In one embodiment, the audio encoding device 101 may include a communications interface 207 for communicating with other components over one or more connections. For example, the communications interface 207 may be capable of communicating using Bluetooth, the IEEE 802.11x suite of standards, IEEE 802.3, cellular Global System for Mobile Communications (GSM) standards, cellular Code Division Multiple Access (CDMA) standards, and/or Long Term Evolution (LTE) standards. In one embodiment, the communications interface 207 facilitates the transmission/reception of video, audio, and/or other pieces of data over the distributed network 105. For example, the audio encoding device 101 may receive one or more pieces of sound program content to be encoded via the communications interface 207. As will be described in greater detail below, the pieces of sound program content may be encoded/processed and transmitted to one or more of the audio decoding devices 103₁-103_N for playback also via the communications interface 207.

[0025] Turning now to **Figure 3**, the audio decoding device 103₁ will be described. Although described in relation to the audio decoding device 103₁, each of the audio decoding devices 103₂-103_N may include similar or identical elements. The audio decoding device 103₁ may be any computing device that is capable of receiving, decoding, and playing back an encoded piece of sound program content. For example, the audio decoding device 103₁ may be a laptop computer, a desktop computer, a tablet computer, a set-top box, a multimedia player, a gaming system, and/or a mobile device (e.g., cellular telephone or mobile media player). As will be described in greater detail below, the audio decoding device 103₁ may receive an audio asset representing an encoded piece of sound program content from the audio encoding device 101. The audio decoding device 103₁ may read/extract metadata in the encoded piece of sound program content for a particular channel/channel group. The metadata may include ducking values to be applied to other channels/channel groups. Accordingly, as noted above, these channels/channel groups that are not associated with the ducking values may be ducked using the ducking values to improve the intelligibility of non-ducked channels/channel groups.

[0026] The audio decoding device 103₁ may include a main system processor 301 and a memory unit 303. Similar to the processor 201 and the memory unit 203, the processor 301 and the memory unit 303 are generically used here to refer to any suitable combination of programmable data processing components and data storage that conduct the operations needed to implement the various functions and operations of the audio decoding device 103₁. The

processor 301 may be a special purpose processor such as an ASIC, a general purpose microprocessor, a FPGA, a digital signal controller, or a set of hardware logic structures (e.g., filters, arithmetic logic units, and dedicated state machines) while the memory unit 303 may refer to microelectronic, non-volatile random access memory.

[0027] An operating system may be stored in the memory unit 303, along with application programs specific to the various functions of the audio decoding device 103₁, which are to be run or executed by the processor 301 to perform the various functions of the audio decoding device 103₁. For example, the memory unit 303 may include a decoder 305. The decoder 305, along with one or more other components of the audio decoding device 103₁, may extract/derive ducking values from an audio asset representing the encoded piece of sound program content. For example, the ducking values may have been derived through an analysis of a primary channel group and inserted with the audio asset. The decoder 305 may extract the ducking values or apply an algorithm to an encoded version of the ducking values within the mixed signal to produce the ducking values. The decoder 305 may thereafter apply these ducking values to one or more channels/channel groups of the encoded piece of sound program content. In particular, the ducking values may be associated in the encoded piece of sound program content with a first channel/channel group and the decoder 305 may apply the ducking values to a second channel/channel group that is distinct from the first channel/channel group. Accordingly, ducking performed by the decoder 305 is relative to other channel/channel groups that are not associated with the ducking values within the audio asset instead of the channel/channel group associated with the ducking values.

[0028] In one embodiment, the audio decoding device 103₁ may include one or more loudspeakers 309 for outputting sound based on the encoded piece of sound program content. The loudspeakers 309 may be any combination of full-range drivers, mid-range drivers, subwoofers, woofers, and tweeters. Each of the loudspeakers 309 may use a lightweight diaphragm, or cone, connected to a rigid basket, or frame, via a flexible suspension that constrains a coil of wire (e.g., a voice coil) to move axially through a cylindrical magnetic gap. When an electrical audio signal is applied to the voice coil, a magnetic field is created by the electric current in the voice coil, making it a variable electromagnet. The coil and the loudspeakers' 309 magnetic system interact, generating a mechanical force that causes the coil (and thus, the attached cone) to move back and forth, thereby reproducing sound under the control of the applied electrical audio signal coming from a source.

[0029] In one embodiment, the audio decoding device 103₁ may include a communications interface 307 for communicating with other components over one or more connections. For

example, the communications interface 307 may be capable of communicating using Bluetooth, the IEEE 802.11x suite of standards, IEEE 802.3, cellular Global System for Mobile Communications (GSM) standards, cellular Code Division Multiple Access (CDMA) standards, and/or Long Term Evolution (LTE) standards. In one embodiment, the communications interface 307 facilitates the transmission/reception of video, audio, and/or other pieces of data. For example, the audio decoding device 103₁ may receive an audio asset representing an encoded piece of sound program content from the audio encoding device 101 via the communications interface 307. This received audio asset may be decoded and played back by the audio decoding device 103₁ as will be described in greater detail below.

[0030] As noted above, the other audio decoding devices 103₂-103_N may be identical to the audio decoding device 103₁. In particular, the audio decoding devices 103₂-103_N may each include a processor 301, memory unit 303, decoder 305, communications interface 307, and one or more loudspeakers 309.

[0031] As described above, the audio encoding device 101 and the audio decoding devices 103₁-103_N may communicate over the distributed network 105. The distributed network 105 may be composed of any combination of networking devices, including switches, routers, controllers, access points, etc. The distributed network may operate using one or more standards/protocols, including Bluetooth, the IEEE 802.11x suite of standards, IEEE 802.3, cellular Global System for Mobile Communications (GSM) standards, cellular Code Division Multiple Access (CDMA) standards, and/or Long Term Evolution (LTE) standards.

[0032] Turning now to **Figure 4**, a method 400 will be described for (1) encoding a multi-channel piece of sound program content to include ducking values associated with a first channel/channel group and (2) playing back the encoded multi-channel piece of sound program content such that a second channel/channel group is ducked in favor of the first channel/channel group that is associated with the ducking values. Although the operations of the method 400 are described in relation to channels/channel groups, in other embodiments, ducking may be performed in a similar manner relative to sound objects or stems represented by the piece of sound program content.

[0033] Each operation of the method 400 may be performed by one or more components of the audio encoding device 101, one or more audio decoding devices 103₁-103_N, and/or another device. For example, as will be used in the description below, one or more of the encoder 205 of the audio encoding device 101 and the decoder 305 of the audio decoding device 103₁ may be used for performing the operations of the method 400. However, in other embodiments, different elements and devices may be used for performing the method 400. Although the

encoder 205 and the decoder 305 are described as residing in the memory units 203 and 303, respectively, in other embodiments, the encoder 205 and the decoder 305 may be implemented by one or more hardware structures, including filters, arithmetic logic units, and dedicated state machines.

[0034] Although the operations of the method 400 are shown and described in a particular order, in other embodiments the operations of the method 400 may be performed in a different order. For example, one or more of the operations may be performed concurrently or during overlapping time periods. Each operation of the method 400 will now be described below by way of example.

[0035] In one embodiment, the method 400 may commence at operation 401 with the receipt of a set of audio signals representing one or more channels for a piece of sound program content. For instance, the audio encoding device 101 may receive N channels of audio corresponding to a piece of sound program content (e.g., a musical composition or a soundtrack/audio track of a movie), where N is greater than or equal to two. For example, four channels of audio representing a piece of sound program content may be received at operation 401. In some embodiments, one of the received channels may correspond to a narration channel that assists visually impaired users to understand actions taking place in a movie or television broadcast. In these embodiments, the other channels may correspond to either spatial channels (e.g., left, right, and center audio channels) and/or represent objects/stems (e.g., dialogue, music, and effects audio objects/stems). The audio signals/channels may be received at operation 401 from an external system or device (e.g., an external computer or streaming audio service) via the communications interface 207. In other embodiments, the audio signals/channels may be stored locally on the audio encoding device 101 (e.g., stored in the memory unit 203) and retrieved at operation 401.

[0036] In some embodiments, the audio signals/channels received at operation 401 may be intended for reproduction simultaneously and in sync with a video. For example, as noted above, the audio signals/channels received at operation 401 may be an audio track for a movie or television program. In this embodiment, the audio signals/channels may be transmitted and received at operation 401 along with the corresponding video content or apart from the video content.

[0037] At operation 403, the channels received at operation 401 may be processed/encoded to generate an audio asset representing the piece of sound program content. In one embodiment, channels may be grouped and processed within these groups. Each channel group may include one or more individual channels that are associated together. The channel groups may be

analyzed together to determine Dynamic Range Control/Compression (DRC) gain values that are to be applied to each respective group of channels during decoding/playback of the encoded piece of sound program content. DRC reduces the volume of loud sounds or amplifies quiet sounds by narrowing or compressing an audio signal's dynamic range. The DRC gain values may be added as metadata to an audio asset representing the encoded piece of sound program content.

[0038] In one embodiment, ducking values may be determined for a single channel/channel group in the piece of sound program content. The ducking values represent the amount other channels/channel groups will be ducked (e.g., reduced in volume). For example, a piece of sound program content received at operation 401 may include four objects: (1) a narration object; (2) a dialogue object; (3) a music object; and (4) an effects object. In some embodiments, these four objects may be assigned to channel groups prior to processing (e.g., the narration object may be assigned to a first channel group, the dialogue object may be assigned to a second channel group, and the music and effects object may be both assigned to a third channel group). In this example, a set of ducking values may be generated and associated with the first channel group. The set of ducking values associated with the first channel group indicates the amount of attenuation to be applied to one or more of the other channel groups by decoders 305 in one or more of the audio decoding devices 103₁-103_N.

[0039] In some embodiments, individual ducking sequences may be associated with the original layout of the channels/channel groups and each possible downmix of the channels/channel groups. For instance, an identifier may be associated with the original layout of the channels/channel groups and the each possible downmix of the channels/channel groups. In this example, separate sequences of ducking values may be associated with each identifier. Accordingly, the appropriate sequence of ducking values may be applied to the channels/channel groups based on their layout/downmix arrangement.

[0040] In one embodiment, the ducking values may be applied to channel groups by narrowing/compressing the dynamic range of these channel groups. For example, the ducking values may indicate that other channel groups may be ducked by 10dB during playback by one or more of the audio decoding devices 103₁-103_N and may be applied similar to DRC gain values. This ducking may fluctuate over a specified time period. For instance, while detecting activity (e.g., voice or speech activity) in the first channel group during a first time period, operation 403 may set ducking values associated with the first channel group to a first level during this first time period. Conversely, upon failing to detect activity in the first channel group during a second time period, operation 403 may set ducking values associated with the first

channel group to a second level during this second time period. In this example, the first level may be greater than the second level such that the second and third channel groups are more severely ducked during periods of activity in the first channel group than during periods of inactivity. Accordingly, the first channel group may be emphasized during periods of activity while allowing the second and third channel groups to be heard and unchanged during periods in which the first channel group is relatively inactive.

[0041] As noted above, the ducking values may be applied similar to DRC gain values. In these embodiments, the ducking values may be less than or equal to zero dB such that application of ducking values causes no attenuation of channel groups (i.e., during periods of inactivity in the first channel group) or attenuation of channel groups. In some embodiments, application of ducking values may override other effects indicated in the audio asset, including “clipping” or “fading” of channels/channel groups.

[0042] In some embodiments, the ducking values may indicate application of different levels of ducking to each channel group. Using the three example channel groups discussed above, operation 403 may determine that the second channel group associated with dialogue may need to be more severely ducked during playback than the third channel group associated with music and effects. This more severe ducking of dialogue allows narration in the first channel group to reside in foreground audio along with music and effects in the third channel group while dialogue in the second channel group may be played just beneath (i.e., in the background). Since dialogue may more harshly interfere with the intelligibility of narration in comparison to music and effects, allowing the separate ducking of the second and third channel groups minimizes the amount of ducking while still promoting intelligibility of narration to users.

[0043] As described above, ducking values may be used to reduce the playback level of channels/channel groups. In some embodiments, ducking values may be extended to not only alter the level of playback, but to additionally or separately change the rendering location in the sound field of channels/channel groups. For instance, using the three example channel groups described above, the ducking values added to the audio asset at operation 403 may indicate that the rendering of the second channel group is to be moved from front loudspeakers 309 to surround loudspeakers 309 during playback by the audio decoding devices 103₁-103_N. This movement of channels/channel groups may be limited during a period of time. For example, movement of channels/channel groups may only be caused when activity is detected in the first channel group (e.g., the narration channel group). By moving the rendering location of channels/channel groups in the sound field, particular channels/channel groups of interest may be emphasized in relation to other channels/channel groups.

[0044] In one embodiment, the ducking values generated at operation 403 and inserted within the audio asset may be associated with a single channel/channel group. Accordingly, each audio asset includes a single channel/channel group that is not ducked during playback while all other channels/channel groups in the audio asset are ducked using the ducking values. In some embodiments, the ducking values generated at operation 403 may be associated with multiple channels/channel groups, but only one of these multiple channels/channel groups may be active for playback at a time. For instance, the audio asset may include multiple different narration channels/channel groups corresponding to different languages. In these embodiments, only one language may be selected for playback at a time (i.e., only one active channel/channel group amongst the different narration channel/channel groups). Accordingly, the ducking values are applied to all other active channel/channel groups not associated with the ducking values (i.e., ducking values not applied to active and inactive narration channel/channel groups).

[0045] In some embodiments, the audio asset may include multiple sets of ducking values that are each associated with a different channel/channel group. For instance, using the example above, the audio asset may include multiple different narration channels/channel groups corresponding to different languages with only one channel/channel group authorized to be active at a time. In this embodiment, each of the narration channels/channel groups may be associated with a different set of ducking values that will be applied to all other active channels/channel groups during playback.

[0046] Following generation of the encoded piece of sound program content, including ducking values, operation 405 may transfer the audio asset representing the piece of sound program content to one or more of the audio decoding devices 103₁-103_N. In one embodiment, transfer of the audio asset may be performed over the distributed network 105 using the communications interface 207 of the audio encoding device 101 and the communications interfaces 307 of the one or more of the audio decoding devices 103₁-103_N. In other embodiments, the audio asset may be transferred at operation 405 through other techniques (e.g., via a portable memory device).

[0047] At operation 407, the one or more of the audio decoding devices 103₁-103_N may decode the audio asset to reveal two or more channels/channel groups, ducking values associated within one of the channels/channel groups, and/or other audio information added at operation 403. As noted above, the ducking values may be used for (1) ducking particular channels/channel groups and/or (2) moving particular channels within the sound field.

[0048] At operation 409, the method 400 may determine if ducking values were present in the audio asset and whether the channel/channel group associated with the ducking values was

selected for playback (e.g., selected for playback by a user). At operation 413, when ducking values are not included in the audio asset or channels associated with the ducking values are not selected for playback, drive signals may be generated for the channels/channel groups selected for playback without ducking. These drive signals generated at operation 413 may be used for playback through the loudspeakers 309 at operation 415.

[0049] Conversely, at operation 411, when ducking values are included in the audio asset and the associated channel/channel group has been selected for playback, the decoded ducking values may be applied to non-associated channels/channel groups to produce a set of drive signals for playback through the loudspeakers 309 at operation 415. In particular, the ducking values may reduce the dynamic range of channels/channel groups not associated with the ducking values. In the three example channel groups described above, ducking values may be applied to the second and/or third channel groups such that audio from the first channel group remains un-ducked and consequently is more intelligible to users. In some embodiments, a scale factor may be applied to the ducking values before application at operation 411.

[0050] As described above in relation to operation 403, ducking values may also cause one or more channels/channel groups not associated with the ducking values to move within the sound field. For instance, at operation 411 ducking values may cause the second channel group, in the example above, to move into surround loudspeakers 309. Along with ducking the dynamic range of the second and/or third channel groups, this movement may improve the intelligibility of the first channel group.

[0051] As described above, ducking values may be included in an audio asset representing a piece of sound program content and associated with a particular channel/channel group. During playback of the encoded piece of sound program content, the ducking values may be applied to all other channels/channel groups. This ducking may improve intelligibility of audio in the non-ducked channel/channel group. For instance, a narration channel/channel group may be more clearly heard by users through the use of selective ducking of other channels/channel groups during playback.

[0052] As explained above, an embodiment of the invention may be an article of manufacture in which a machine-readable medium (such as microelectronic memory) has stored thereon instructions which program one or more data processing components (generically referred to here as a “processor”) to perform the operations described above. In other embodiments, some of these operations might be performed by specific hardware components that contain hardwired logic (e.g., dedicated digital filter blocks and state machines). Those

operations might alternatively be performed by any combination of programmed data processing components and fixed hardwired circuit components.

[0053] While certain embodiments have been described and shown in the accompanying drawings, it is to be understood that such embodiments are merely illustrative of and not restrictive on the broad invention, and that the invention is not limited to the specific constructions and arrangements shown and described, since various other modifications may occur to those of ordinary skill in the art. The description is thus to be regarded as illustrative instead of limiting.

CLAIMS

What is claimed is:

1. A method for encoding audio content associated with video content, the method comprising:

receiving a piece of sound program content comprising (i) a first channel group, object or stem that comprises visually descriptive audio content that describes actions taking place in the video content, (ii) a second channel group, object or stem that comprises at least one of dialogue, music, and effects, and (iii) a third channel group, object or stem that comprises at least another one of dialogue, music, and effects;

determining that the first channel group, object or stem in the piece of sound program content is to be emphasized during playback of the piece of sound program content;

generating (i) a first set of ducking values to apply to the second channel group, object or stem and (ii) a second set of ducking values to apply to the third channel group, object or stem, based on activity within the first channel group, object or stem, wherein the first set of ducking values is different than the second set of ducking values such that when the first and second sets of ducking values are applied to the second and third channel groups, objects or stems, respectively, the second and third channel groups, objects or stems are deemphasized differently during playback of the piece of sound program content; and

associating the first and second sets of ducking values with the first channel group, object or stem in an audio asset, wherein the audio asset includes (i) the first, second, and third channel groups, objects or stems and (ii) the first and second sets of ducking values that are associated with the first channel group, object or stem.

2. The method of claim 1, wherein the first and second sets of ducking values vary over time based on the activity within the first channel group, object or stem, such that during periods of high activity within the first channel group, object or stem, the first and second sets of ducking values provide more ducking of the second and third channel groups, objects or stems than during periods of low activity within the first channel group, object or stem.

3. The method of claim 1, wherein when the second channel group, object or stem comprises dialogue and the third channel group, object or stem comprises music and effects, the first and second sets of ducking values are generated such that the second channel group, object

or stem is to be deemphasized more than the third channel group, object or stem during playback of the piece of sound program content.

4. The method of claim 1, wherein the first and second sets of ducking values cause a dynamic range of the second and third channel groups, objects or stems to be reduced during playback.
5. The method of claim 1, wherein the first set of ducking values cause the second channel group, object or stem to move in a sound field during playback.
6. The method of claim 1, wherein the first set of ducking values are gain values less than or equal to zero decibels that when applied to the second channel group, object or stem, attenuate sound associated with the second channel group, object or stem.
7. The method of claim 5, wherein the activity comprises speech activity, and wherein the movement of the second channel group, object or stem is only when there is speech activity within the first channel group, object or stem.
8. A method for playing back audio content associated with video content, the method comprising:
 - receiving an audio asset representing a piece of sound program content comprising (i) a first channel group, object or stem that comprises visually descriptive audio content that describes actions taking place in the video content, (ii) a second channel group, object or stem that comprises at least one of dialogue, music, and effects, (iii) a third channel group, object or stem that comprises at least another one of dialogue, music, and effects than the second channel group, object or stem, and (iv) a first set of ducking values and a second set of ducking values associated with the first channel group, object or stem, wherein the first set of ducking values is different than the second set of ducking values and wherein the first and second sets of ducking values were generated based on activity in the first channel group, object or stem;
 - extracting the first and second sets of ducking values along with the (i) first channel group, object or stem (ii) second channel group, object or stem, and (iii) third channel group, object or stem from the audio asset; and
 - during playback of the piece of sound program content through a plurality of loudspeakers

applying the first set of ducking values to the second channel group, object or stem; and

applying the second set of ducking values to the third channel group, object or stem, wherein applying the first and second sets of ducking values deemphasizes the second channel group, object or stem differently than the third channel group, object or stem.

9. The method of claim 8, wherein application of the first and second sets of ducking values deemphasize by reducing the dynamic range of the (i) second channel group, object or stem and (ii) third channel group, object or stem during playback.

10. The method of claim 9, further comprising:

applying a scale factor to the first set or second set of ducking values prior to application of the first set or second set of ducking values.

11. The method of claim 8 further comprising producing a set of drive signals based on the first, second, and third channel groups, objects or stems to drive the plurality of loudspeakers to render sound in a sound field, wherein applying the first set of ducking values causes the rendering location of the second channel group, object or stem in the sound field to move to a different rendering location in the sound field during playback.

12. A non-transitory computer readable medium containing instructions for encoding audio content associated with video content, which when executed by a processor in an audio device cause the audio device to:

receiving a piece of sound program content comprising (i) a first channel group, object or stem that comprises visually descriptive audio content that describes actions taking place in the video content, (ii) a second channel group, object or stem that comprises at least one of dialogue, music, and effects, and (iii) a third channel group, object or stem that comprises at least another one of dialogue, music, and effects;

analyze the piece of sound program content to determine that the first channel group, object or stem in the piece of sound program content is to be emphasized during playback of the piece of sound program content;

generate (i) a first set of ducking values to apply to the second channel group, object or stem and (ii) a second set of ducking values to apply to the third channel group, object or stem, based on activity within the first channel group, object or stem, wherein the first set of ducking

values is different than the second set of ducking values such that when the first and second sets of ducking values are applied to the second and third channel groups, objects or stems, respectively, the second and third channel groups, objects or stems are deemphasized differently during playback in favor of the first channel group; and

associate the ducking values with the first channel group, object or stem in an audio asset, wherein the audio asset includes (i) the first, second, and third channel groups, objects or stems and (ii) the first and second sets of ducking values that are associated with the first channel group, object or stem.

13. The non-transitory computer readable medium of claim 12, wherein the first and second sets of ducking values vary over time based on the activity within the first channel group, object or stem, such that during periods of high activity within the first channel group, object or stem, the first and second sets of ducking values provide more ducking of the second and third channel groups, objects or stems than during periods of low activity within the first channel group, object or stem.

14. The non-transitory computer readable medium of claim 12, wherein when the second channel group, object or stem comprises dialogue and the third channel group, object or stem comprises music and effects, the first and second sets of ducking values are generated such that the second channel group, object or stem is to be deemphasized more than the third channel group, object or stem during playback of the piece of sound program content.

15. The non-transitory computer readable medium of claim 12, wherein the first and second sets of ducking values cause a dynamic range of the second and third channel groups, objects or stems to be reduced during playback.

16. The non-transitory computer readable medium of claim 12, wherein the first set of ducking values cause the second channel group, object or stem to move in a sound field during playback.

17. The non-transitory computer readable medium of claim 12, wherein the first set of ducking values are gain values less than or equal to zero decibels that when applied to the second channel group, object or stem, attenuate sound associated with the second channel group, object or stem.

14 Jul 2017
2015236755

1001863583

18. An audio device for playing back audio content associated with video content, the audio device comprising

a hardware processor; and

a memory unit storing instructions executed by the hardware processor that cause the audio device to:

receive an audio asset representing a piece of sound program content comprising

(i) a first channel group, object or stem that comprises visually descriptive audio content that describes actions taking place in the video content, (ii) a second channel group, object or stem that comprises at least one of dialogue, music, and effects, (iii) a third channel group, object or stem that comprises at least another one of dialogue, music, and effects than the second channel group, object or stem, and (iv) a first set of ducking values and a second set of ducking values associated with the first channel group, object or stem, wherein the first set of ducking values is different than the second set of ducking values and wherein the first and second sets of ducking values were generated based on activity in the first channel group, object or stem;

extract the first and second sets of ducking values along with the (i) first channel group, object or stem, (ii) second channel group, object or stem, and (iii) third channel group, object or stem from the audio asset; and

during playback of the piece of sound program content through a plurality of loudspeakers

apply the first set of ducking values to the second channel group, object or stem; and

apply the second set of ducking values to the third channel group, object or stem, wherein applying the first and second sets of ducking values deemphasizes the second channel group, object or stem differently than the third channel group, object or stem.

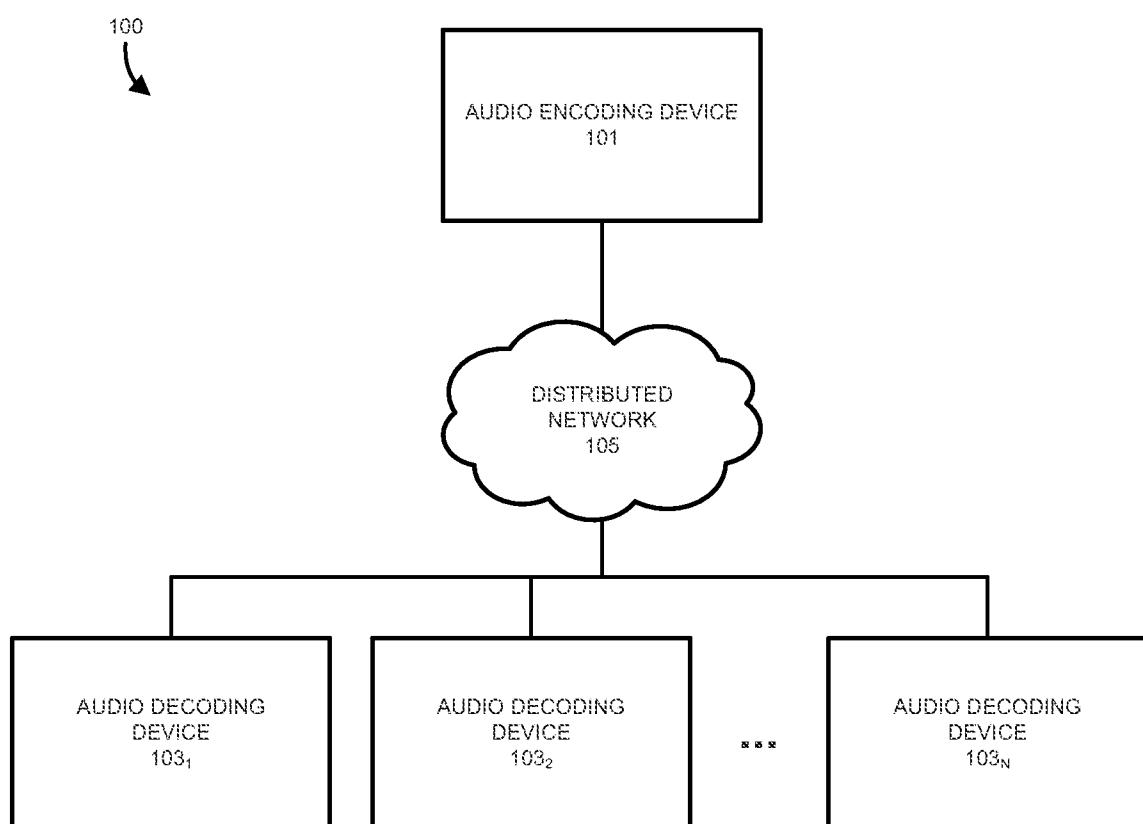
19. The audio device of claim 18, wherein application of the first and second sets of ducking values deemphasize by reducing the dynamic range of the (i) second channel group, object or stem and (ii) third channel group, object or stem during playback.

20. The audio device of claim 19, wherein the memory unit includes further instructions, which when executed by the hardware processor cause the audio device to:

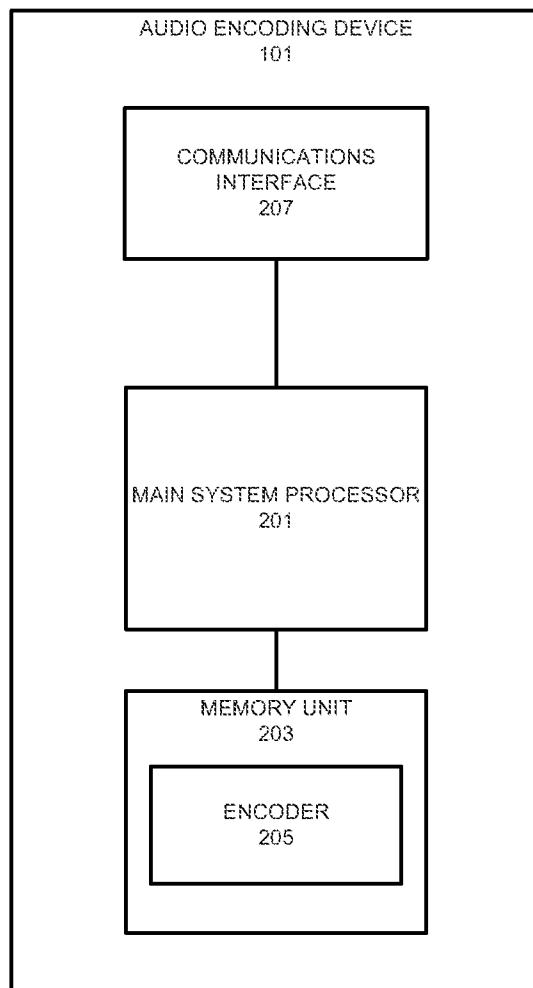
apply a scale factor to the first and second sets of ducking values prior to application of the first and second sets of ducking values.

21. The audio device of claim 18, wherein the memory unit includes further instructions, which when executed by the hardware processor cause the audio device to produce a set of drive signals based on the first, second, and third channel groups, objects, or stems to drive the plurality of loudspeakers to render sound in a sound field, wherein the instructions to apply the first set of ducking values causes the rendering location of the second channel group, object or stem to move to a different rendering location in the sound field during playback.

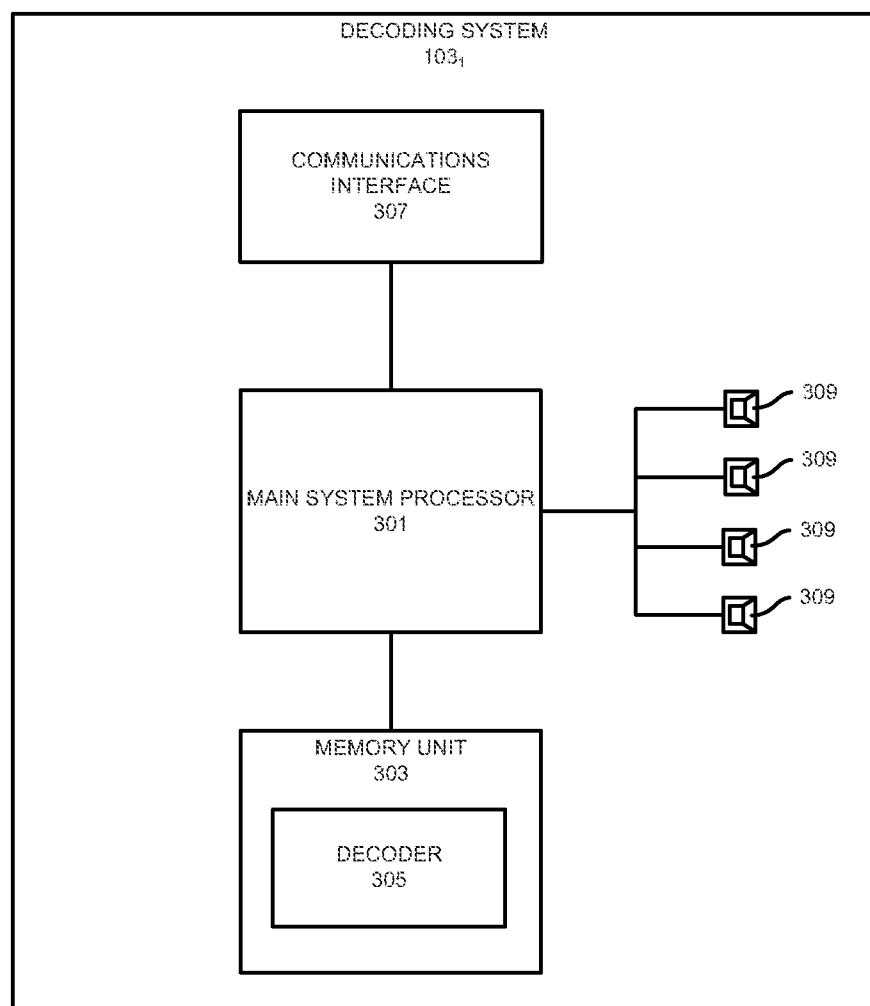
1/4

**FIG. 1**

2/4

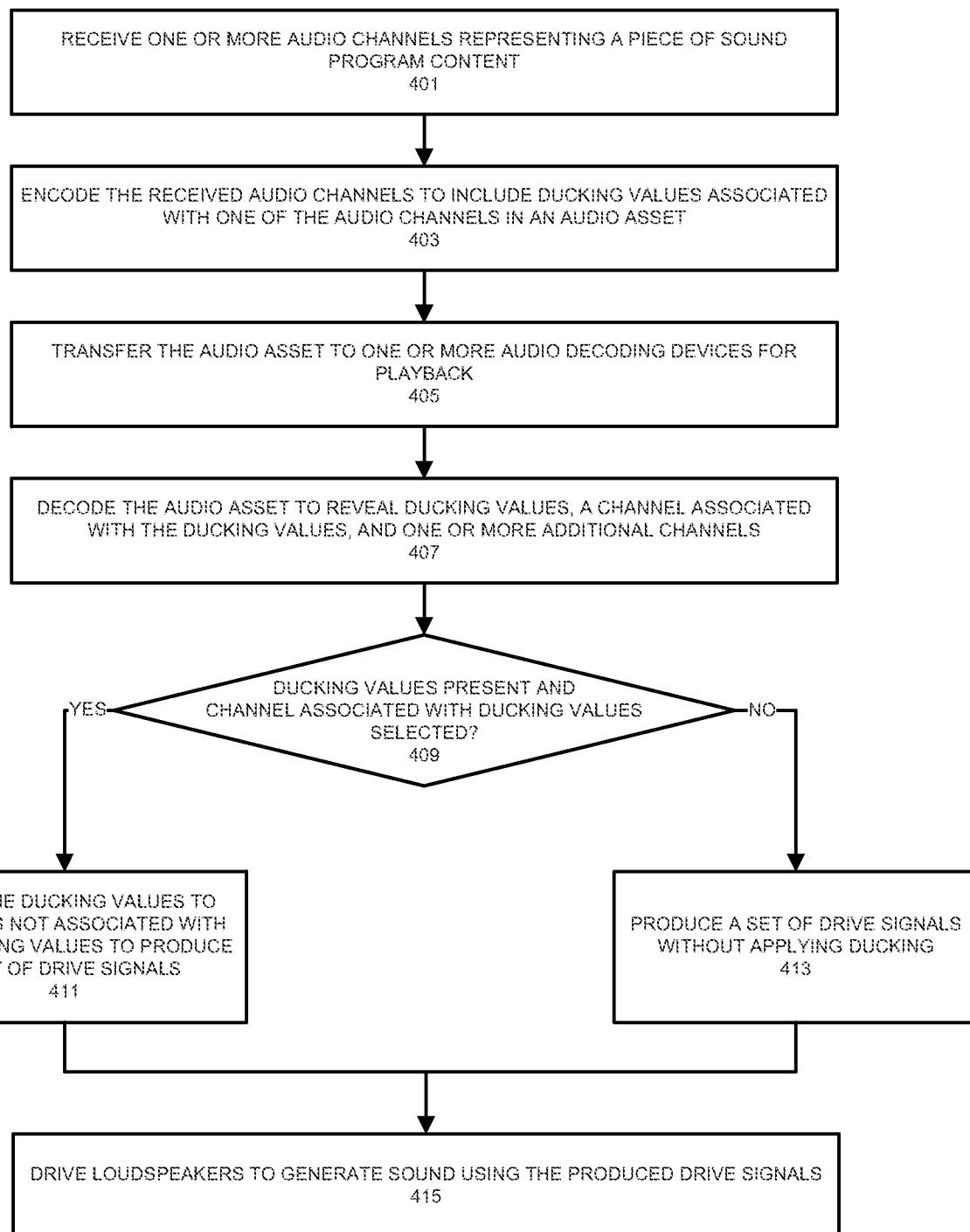
**FIG. 2**

3/4

**FIG. 3**

4/4

400

**FIG. 4**