

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

(11) Application No. **AU 2012359080 B2**

(54) Title  
**Managing playback of supplemental information**

(51) International Patent Classification(s)  
**H04M 11/00** (2006.01)                      **G06F 3/06** (2006.01)

(21) Application No: **2012359080**                      (22) Date of Filing: **2012.12.19**

(87) WIPO No: **WO13/096422**

(30) Priority Data

(31) Number	(32) Date	(33) Country
<b>13/332,157</b>	<b>2011.12.20</b>	<b>US</b>

(43) Publication Date: **2013.06.27**

(44) Accepted Journal Date: **2015.09.17**

(71) Applicant(s)  
**Audible, Inc.**

(72) Inventor(s)  
**Story, Guy A.;Goldstein, Douglas S.**

(74) Agent / Attorney  
**Freehills Patent Attorneys, ANZ Tower 161 Castlereagh Street, Sydney, NSW, 2000**

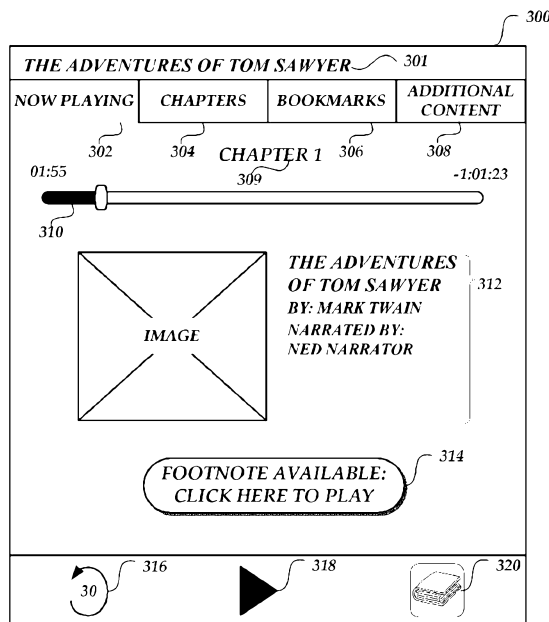
(56) Related Art  
**US 2009/0251440**  
**US 2010/0251386**



- (51) International Patent Classification: *H04M 11/00* (2006.01) *G06F 3/06* (2006.01)
- (21) International Application Number: PCT/US2012/070565
- (74) Agent: ANDERSON, Maria, Culic; Knobbe Martens Olson & Bear, LLP, 2040 Main Street, 14th Floor, Irvine, CA 92614 (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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,
- (22) International Filing Date: 19 December 2012 (19.12.2012)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 13/332,157 20 December 2011 (20.12.2011) US
- (71) Applicant: AUDIBLE, INC. [US/US]; 1 Washington Park - 16th Floor, Newark, NJ 07102 (US).
- (72) Inventors: STORY, Guy, A.; 1 Washington Park - 16th Floor, Newark, NJ 07102 (US). GOLDSTEIN, Douglas, S.; 1 Washington Park - 16th Floor, Newark, NJ 07102 (US).

[Continued on next page]

(54) Title: MANAGING PLAYBACK OF SUPPLEMENTAL INFORMATION



(57) Abstract: A computing device may provide an interface that enables a user to playback audio content including supplemental information. The supplemental information may correspond, for example, to endnotes, footnotes, glossaries, appendices, commentary from an author, editor, or other party, or to additional supplemental information. During playback of the audio content, a user may indicate that an item of supplemental information associated with a current playback position of the audio content should be played. In some embodiments, this indication may be in response to an indication to the user that supplemental information is available. The device may then playback the supplemental information associated with the current playback position.

Fig. 4A.

WO 2013/096422 A1

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG). **Published:**

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

**Declarations under Rule 4.17:**

- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))

## MANAGING PLAYBACK OF SUPPLEMENTAL INFORMATION

### BACKGROUND

[0001] Traditional printed books, electronic books or other printed media (whether in electronic or physical form) often contain a primary text and additional, supplementary information, such as footnotes, end notes, glossaries and appendices. These items of supplementary information often contain useful or interesting information, such as additional background or explanatory text regarding the primary text, external references, or commentary.

[0002] Frequently, printed media are converted into audio format. Generally, this may involve narrating and recording a reading of the printed medium. The resulting audio book or audio content may then be made available to users. Audio content may be played back on a wide variety of devices, such as notebook and tablet computers, mobile phones, personal music players, electronic book (“eBook”) readers, and other audio playback devices or devices with audio playback capability.

[0002A] Reference to any prior art in the specification is not an acknowledgment or suggestion that this prior art forms part of the common general knowledge in any jurisdiction or that this prior art could reasonably be expected to be combined with other pieces of prior art by a skilled person in the art.

### SUMMARY OF THE INVENTION

[0002B] According to a first aspect of the invention there is provided a system for output of supplemental audio information, the system including:

- one or more data stores that store audio content and one or more items of supplemental audio information, wherein each of the one or more items of supplemental audio information is associated with one or more playback positions within the audio content and a content category of a plurality of content categories;

- an input device that obtains an input of a user;

- a display device that presents information regarding the audio content;

- an output device that outputs the audio content; and

a processor in communication with the one or more data stores, the input device and the output device, the processor operative to:

receive a preference of the user specifying that an indication should be provided for an item of supplemental audio information associated with a preferred content category of the plurality of content categories;

cause output of the audio content via the output device;

determine that a first item of supplemental audio information of the one or more items of supplemental audio information is associated with a current playback position within the audio content;

identify, based at least in part on the first item of supplemental audio information, a first content category of the plurality of content categories;

determine that the first content category corresponds to the preferred content category;

cause output, via at least one of the output device and the display device, of the indication that the first item of supplemental audio information is available for playback;

obtain input from the user indicating that the first item of supplemental audio information should be played; and

cause output of the first item of supplemental audio information associated with the current playback position via the output device.

**[0002C]** According to a second aspect of the invention there is provided a system for outputting supplemental audio information, the system including:

one or more data stores that store audio content and one or more items of supplemental audio information, wherein each of the one or more items of supplemental audio information is associated with one or more playback positions within the at least one of the audio content and one of the one or more items of supplemental audio information;

a processor in communication with the data store, the processor operative to:

cause output of audio content via an output device;

determine that an item of supplemental audio information from the one or more items of supplemental audio information is associated with a current playback position within the audio content being output via the output device;

determine a user preference for a preferred content category of a plurality of content categories;

identify, based at least in part on the item of supplemental audio information, a first content category of the plurality of content categories;

determine that the first content category corresponds to the preferred content category;

based at least in part on the determination that the first content category corresponds to the preferred content category, cause output of an indication that supplemental audio information is available for playback;

obtain input from the user via an input device indicating that the supplemental audio information should be played;

cause output of the item of supplemental audio information via the output device.

**[0002D]** According to a third aspect of the invention there is provided a computer-implemented method for playback of supplemented audio information to a user, wherein the supplemented audio information includes a primary audio content and one or more items of supplemental audio information, and wherein each item of supplemental audio information is associated with at least one playback position within at least one of the primary audio content and one of the one or more items of supplemental audio information, the computer-implemented method including:

under control of one or more computing devices,

causing output of the primary audio content via an output device;

determining that a first item of supplemental audio information of the one or more items of supplemental audio information is associated with a current playback position within the primary audio content being output via the output device;

determining a user preference for outputting an indication during playback of the primary audio content that an item of supplemental audio information associated with a preferred content category is available for playback;

identifying, based at least in part on the first item of supplemental audio information, a first content category of the plurality of content categories;

determining that the first content category corresponds to the preferred content category;  
based at least in part on the user preference, causing output of the indication that the first item of supplemental audio information is available for playback;

obtaining input from the user via an input device, the input indicating that the first item of supplemental audio information should be played; and

causing output of the first item of supplemental audio information via the output device.

[0002E] According to a fourth aspect of the invention there is provided a computer readable, non-transitory storage medium having at least one computer-executable module including computer-executable instructions for providing supplemental audio information, the at least one computer-executable module including:

a supplemental audio information module that, when executed by a processor, causes the processor to:

cause output of a primary audio content, wherein the primary audio content is associated with one or more items of supplemental audio information, and wherein each of the one or more items of supplemental audio information is associated with one or more playback positions within the at least one of the audio content and one of the one or more items of supplemental audio information;

determine one or more items of supplemental audio information associated with a current playback position within the primary audio content;

determine a user preference for a preferred content category of a plurality of content categories;

identify, for at least one of the one or more items of supplemental audio information, a first content category of the plurality of content categories;

determine that the first content category corresponds to the preferred content category;

based at least in part on determining that the first content category corresponds to the first content category, cause output of an indication that supplemental audio information is available for playback;

obtain input indicating that the at least one supplemental audio information should be played; and  
cause output of the at least one item of supplemental audio information.

#### BRIEF DESCRIPTION OF THE DRAWINGS

**[0003]** The foregoing aspects and many of the attendant advantages of the present disclosure will become more readily appreciated as the same become better understood by reference to the following detailed description, when taken in conjunction with the accompanying drawings, wherein:

**[0004]** FIGURE 1 is a block diagram depicting an illustrative embodiment of a computing device configured to implement playback of supplemental audio information;

**[0005]** FIGURE 2 is a schematic block diagram depicting an illustrative operating environment in which a computing device of FIGURE 1 may interact with an electronic marketplace to acquire audio content and supplemental information;

**[0006]** FIGURE 3 is an illustrative graphical representation or visualization of audio content including supplemental information;

**[0007]** FIGURES 4A-4D depict illustrative user interfaces that may be used to facilitate playback of supplemental audio information; and



**[0008]** FIGURE 5 is a flow diagram depicting an illustrative routine for playback of supplemental audio information.

#### DETAILED DESCRIPTION

**[0009]** Generally described, aspects of the present disclosure relate to the output of supplemental audio information on a computing device. Specifically, systems and methods are disclosed for controlling the playback of audio content including one or more items of supplemental information, such as footnotes, endnotes, glossary information, or commentary. For example, a user may utilize a computing device such as a personal music player in order to obtain access to audio content including supplemental audio information. While listening to the audio content, the user may receive an indication that supplemental information is available and associated with recently played or upcoming audio content. This supplemental information may correspond to information contained within a printed text from which the audio content was created. For example, footnotes, endnotes, glossaries, and appendices may all constitute supplemental information. Further, supplemental information may correspond to other information, such as author or editor commentary, or commentary from other users who have purchased the audio content. Still further, supplemental information may correspond to commentary generated by contacts of a user. Such contacts may be associated with the user via an electronic marketplace used to retrieve audio content, through one or more social networking systems, or through other systems.

**[0010]** For example, embodiments of the present disclosure may be capable of communicating (over the Internet and/or another data network) with an external social networking system. The social networking system may host a social networking web site that provides functionality for users to set up social networking accounts (to become “members”), and to establish contact or “friend” relationships with other members. Examples of such social networking systems and web sites include Facebook™, MySpace™, Mixi™, and LinkedIn™. To establish a contact or “friend” relationship via such a system, one member typically initiates the transmission of an invitation (also sometimes referred to as a “friend request” or “connection request”) to another member by interacting with the social networking system. If the receiving member accepts this invitation (such as by clicking on an acceptance button or link), the social

networking system creates a database record or entry indicating that the two users are contacts of each other.

**[0011]** The social networking system may include a service interface, including an API (Application Programming Interface), which enables external systems to retrieve member profile information and to post information to members' profiles. Embodiments of the present disclosure may use this service interface to retrieve social networking profile information of particular members of the social networking system. For example, embodiments of the present disclosure may make API calls to the social networking system to request the social networking data of particular members of the social networking system. Further examples for determining contacts of a user are given in United States Patent Application No. 12/871,733, filed on August 30, 2010 and entitled "CUSTOMIZING CONTENT BASED ON SOCIAL NETWORK INFORMATION" which is hereby incorporated by reference in its entirety.

**[0012]** The supplemental information provided with an audio content may or may not be contained within a corresponding printed text. After receiving an indication that supplemental information is available, the user may enter a command indicating they wish to listen to the supplemental information. In some embodiments, the command may be received for some period before or after the point at which the supplemental information is available (e.g., the user may be given  $n$  seconds during which a command can be entered). This command may be obtained at the personal music player by any input, such as the press of a button, or, where the personal music player is equipped with a microphone or other audio input, speaking a command. After receiving the command, the personal music player may then output the supplemental information to the user. Illustratively, after playing the supplemental information, the personal music player may return to the audio content at the position at which the user entered the command, or a position near that at which the user entered the command.

**[0013]** As will be described below, various embodiments may be used exclusive to or in combination with the illustrative example described above. For example, in one embodiment, the personal music player may not output an indication that supplemental information is available, or may only indicate the availability of supplemental information that would be of particular interest the user. Preventing indication of every item of supplemental information may be beneficial, for example, where the amount of supplemental information is large or where

supplemental information is frequently available. Illustratively, a glossary or appendix of an audio content may be provided which contains explanatory information regarding characters, terms, locations, or entities within an audio content. It may not be desirable to indicate the presence of this supplemental information at every mention of those characters, terms, etc., within the audio content. Instead, the personal music player may be configured to receive a command to play supplemental information at any point within an audio content, and may search for supplemental information corresponding to a word recently recited in the audio content. Illustratively, if playback of an audio content mentions a character name with which the user is not familiar, the user may input a command to search for supplemental information. The personal music player may detect that the audio recently discussed the character, and locate corresponding supplemental information associated with the character. In this manner, a user listening to audio content may be provided with access to the same reference information given to a reader of text. In these embodiments, an error may be played if no supplemental information associated with the current playback position can be found.

**[0014]** In some embodiments, the audio content may be associated with corresponding textual content. For example, a personal music player may store an audio content and the corresponding text. In some instances, the corresponding text may be a book from which the audio content is created. In other instances, the text may be a transcript created from the audio content. For example, some embodiments or the present disclosure may include a companion identification module (not shown within FIGS. 1-9) for processing multiple content items (e.g., audio content and textual content) and determining whether such content items can be synchronized. The companion identification module may use a product identifier and/or metadata (e.g., title, author, edition, etc.) related to the content items to identify synchronizable content. For instance, such information can be used to determine that different formats/versions of the same or similar content, such as a given e-book and a given audio book, are companion content. In some implementations, the companion identification module may further process the content information to identify companion content by applying filtering, spell/grammar check, and any other suitable automated algorithms. In addition, the companion identification module can incorporate additional reference materials to identify a content match, for instance, by accessing a uniform resource identifier. Further, the companion identification module may

compare pairs or other combinations of content to determine whether each pair/combination is companion content, for example, by correlating each pair/combination of content in a common format. More specifically, if the content information corresponds to information related to the presentation of a specific representation of content (e.g., an audio book), the companion identification module may translate at least a portion of the content into a format corresponding a different version of the content (e.g., translating audio into a specific e-book format) and/or into a generic format (e.g., text). The translated content information can be correlated with other content information to determine whether each pair of content is companion content that can be synchronized. In other implementations, the companion identification module may prompt a user to enter data (e.g., title, author, edition, etc.) for one or more contents. The companion identification module may compare the data in order to identify companion content. In yet other implementations, the companion identification module may prompt a user to confirm that the first and second content are companion content.

**[0015]** Those skilled in the art will appreciate that even though a first content and a second content may be companion content, the first content may not correlate exactly to the second content. For example, the audio book of *The Adventures of Tom Sawyer* may not exhibit a one-to-one correspondence to the e-book of *The Adventures of Tom Sawyer* due to differences between the narration script of the audio book and the text of the e-book, differences between editions, differences in “front matter” (e.g., introduction to the audio production, table of contents, index, etc.) and/or “back matter,” differences between unabridged and abridged versions, differences in format of content, differences in device capabilities, etc. Due to the existence of such uncorrelated portions of potential companion content, the companion identification module can be configured to identify companion content amongst pairs/combination of content that include mismatching portions. Although more detail will be provided below with reference to correlating portions of an unabridged version of an e-book with an abridged version of the corresponding audio book, the concepts described may be applied to any application in which a first and second content include mismatching portions.

**[0016]** As noted above, when correlating an unabridged version of an e-book with an abridged version of the corresponding audio book, mismatches can result from a number of sources, such as deletions, modifications, insertion of new material, or any combination thereof.

Yet the abridged version of the e-book may include large portions of unmodified material, which may strongly correlate with the abridged version of the audio book. Accordingly, in some embodiments, the companion identification module may detect mismatches between an abridged version and an unabridged version of content using algorithms that solve “longest common sequence” problems, an implementation of which can be found, for example, in the UNIX “diff” operation. The companion identification module can then generate content synchronization information for the correlated portions of the different versions and provide the content synchronization information to the user’s computing device for use when synchronizing presentation of the abridged audio book with the unabridged e-book. More specifically, the companion identification module can correlate unmodified material similarly to how a full book and a full audio book are correlated. However, for passages of an abridged version with small-scale modifications, the companion identification module may correlate the modified material and generate an indicator that these portions include some mismatches. Although new and deleted material in the abridged version may not correlate with the full version, these uncorrelated passages and modified material can be handled so as to provide the most useful and natural user experience. This may include disabling a subset of synchronous presentation features and/or providing an indication of small scale mismatches. As a non-limiting example, when a user listens to the abridged audio book while viewing the unabridged e-book with synchronized highlighting of text, small scale modifications may be highlighted differently to indicate that audio playback does not completely match the written text. Similar techniques can be applied when correlating an unabridged audio book with a companion unabridged e-book, as some material may not match between these versions.

**[0017]** An illustrative method that can be implemented by the companion identification module for processing mismatches will now be described in general terms. First, potential companion content can be converted into a common format, such as text and/or phoneme. Then an attempt is made to match content at a small scale (e.g., at a word level) can be performed. Since missing or added passages can include larger portions (e.g., whole sentences or paragraphs) that do not match, a passage mismatch score indicating the frequency of mismatches can be computed for sentences, paragraphs, and/or other units of text. For example, the passage mismatch score may represent a number and/or percentage of word mismatches for a

larger portion of text (e.g., a sentence, paragraph, or page). Sequences of larger portions with similarly high mismatch frequencies compared with adjoining sequences (where “high” can be set empirically from analyzing correlation of known companion content) can be identified. These sequences may represent content that should not be synchronized with other content in a companion content pair. The high mismatch passages can be removed in either or both contents, and matching can be reattempted at a smaller scale (e.g., at the word level). Metadata can be generated based on the correlation(s) to determine alignment points for the content synchronization information. The metadata may also include information identifying reference points in at least one of the companion content at which a matching portion begins/ends. Further examples of synchronization of textual and audio content are given in United States Patent Application No. 13/070,313, filed on March 23, 2011 and entitled “SYSTEMS AND METHODS FOR SYNCHRONIZING DIGITAL CONTENT” which is hereby incorporated by reference in its entirety.

**[0018]** In these embodiments, the correlation of audio and text content may be used to provide additional functionality or to further enhance features described above. For example, as described above, when playback of an item of supplemental information has concluded, play may continue from a point at or near the point where the user issued a command to play the supplemental information. Illustratively, playback may resume at the exact point the command was received, or at some fixed time period prior to that point (e.g., 3 or 5 seconds). However, in embodiments where information regarding the text corresponding to the played audio is known, playback may resume at a point determined at least in part based on the corresponding text. For example, playback may resume at the point in the audio content just prior to the sentence spoken in which the user indicated supplemental information should be played. In other embodiments, playback may resume at the beginning of a paragraph, a word, or other unit of text. In these embodiments, an indicator of supplemental information may be suppressed for supplemental information that has already been played.

**[0019]** Correlations between audio and text may further be used to enhance the above embodiments. For example, in embodiments where a user may request supplemental information regarding any character, term, etc., by issuing a command to play supplemental information, the

personal music player may use text corresponding to recently played audio to search for such terms in a provided glossary or appendix.

**[0020]** In some embodiments, an item of supplemental information may be associated with multiple positions within a primary audio content. For example, supplemental information may be associated with a position within a chapter of an audio content, and also associated with the end of that chapter. In this manner, a user may have multiple opportunities to hear an item of supplemental information. Further, an item of supplemental information may be associated with a position within another item of supplemental information. For example, a first item of supplemental information may be associated with a position in a primary audio content, and a second item of supplemental information may be associated with a position in the first item of supplemental information. The second item of supplemental information may itself have one or more additional items of supplemental information associated with it. As such, multiple levels of supplemental information may be provided, each new level associated with a previous level or the primary audio content. In embodiments where associations of supplemental information would cause the supplemental information to be available multiple times within a playback of audio content, indication of the availability of that supplemental information may be suppressed after the first indication that the supplemental information is available. In other embodiments, indication of supplemental information may be suppressed after the supplemental information has been fully played.

**[0021]** In some embodiments, audio content and supplemental information may be stored within data storage of a playback device. In other embodiments, audio content and/or supplemental information may be stored remote from the playback device, such as on a remote server. Illustratively, the playback device may be configured to retrieve audio content and/or supplemental information from the remote server. In some embodiments, supplemental information associated with audio content may be retrieved at substantially the same time as the audio content. In other embodiments, a playback device may be configured to retrieve supplemental information periodically. For example, a playback device may query a remote server associated with audio content every  $n$  hours in order to determine whether new supplemental information is available. In further embodiments, a playback device may query a remote server for supplemental information associated with a currently played audio content. In

still other embodiments, a remote server may be configured to notify a playback device of available supplemental information.

**[0022]** In some embodiments, a user of a playback device may specify types of supplemental information which are desired. For example, a user may specify that supplemental information associated with the author (e.g., footnotes, glossaries, author commentary, etc.) should be played, while supplemental information associated with the publisher (e.g., editor commentary, etc.) should not be played. Further, a user may specify that supplemental information associated with contacts of the user should be played, while supplemental information associated with general users of an electronic marketplace from which the audio content was acquired should not be played.

**[0023]** In further embodiments, a user of a playback device may specify categories of supplemental information which are desired. For example, where audio content has been acquired from an electronic marketplace, the electronic marketplace may categorize items of supplemental information into one or more categories. Examples of such categories include, but are not limited to, “Top Rated,” “Funny,” “Insightful,” “Informative,” and “Interesting.” Illustratively, a user may specify that only supplemental information listed as “Top Rated” or “Funny” should be presented for playback, while other supplemental information should be excluded. In some embodiments, where a user has excluded some types of supplemental information and where that supplemental information is stored remotely from a playback device, it may not be necessary for the playback device to retrieve the remotely stored supplemental information.

**[0024]** Although the preceding description refers to a personal music player, any computing device capable of presenting audio content to a user may be used in accordance with the present disclosure. Such a computing device can include, but is not limited to, a laptop, personal computer, a tablet computer, personal digital assistant (PDA), hybrid PDA/mobile phone, mobile phone, electronic book reader, digital media player, integrated components for inclusion in computing devices, appliances, electronic devices for inclusion in vehicles or machinery, gaming devices, set top boxes, electronic devices for inclusion in televisions, and the like. These computing devices may be associated with any of a number of visual, tactile, or auditory output devices, and may be associated with a number of devices for user input,



including, but not limited to, keyboards, mice, trackballs, trackpads, joysticks, input tablets, trackpoints, touch screens, remote controls, game controllers, motion detectors and the like.

**[0025]** In addition, audio content can refer to any data containing audio information that can be directly or indirectly accessed by a user, including, but not limited to, multi-media data, digital video, audio data, electronic books (“eBooks”), electronic documents, electronic publications, computer-executable code, portions of the above, and the like. References to textual content or other visually displayed content should be understood to include any form of visual or tactile content, including text, images, charts, graphs, slides, maps, Braille, embossed images, or any other content capable of being displayed in a visual or tactile medium. Content may be stored on a computing device, may be generated by the computing device, or may be streamed across a network for display or output on the computing device. Moreover, content may be obtained from any of a number of sources, including a network content provider, a local data store, computer readable media, a content generation algorithm (e.g., a text-to-speech algorithm) running remotely or locally, or through user input (e.g., text entered by a user). Content may be obtained, stored, or delivered from any one or combination of sources as described above.

**[0026]** FIGURE 1 is a block diagram illustrating an embodiment of a computing device 100 configured to implement playback of content including supplemental information. The computing device 100 may have one or more processors 102 in communication with a network interface 104, a display interface 106, a computer readable medium drive 108, and an input/output device interface 110, all of which communicate with one another by way of a communication bus. The network interface 104 may provide connectivity to one or more networks or computing systems. The processor(s) 102 may thus receive information and instructions from other computing systems or services via a network. The processor(s) 102 may also communicate to and from memory 112 and further provide output information or receive input information via the display interface 106 and/or the input/output device interface 110. The input/output device interface 110 may accept input from one or more input devices 124, including, but not limited to, keyboards, mice, trackballs, trackpads, joysticks, input tablets, trackpoints, touch screens, remote controls, game controllers, heart rate monitors, velocity sensors, voltage or current sensors, motion detectors, transponders, global positioning systems,

radio frequency identification tags, or any other input device capable of obtaining a position or magnitude value from a user. The input/output interface may also provide output via one or more output devices 122, including, but not limited to, one or more speakers or any of a variety of digital or analog audio capable output ports, including, but not limited to, headphone jacks, ¼ inch jacks, XLR jacks, stereo jacks, Bluetooth links, RCA jacks, optical ports or USB ports, as described above. The display interface 106 may be associated with any number of visual or tactile interfaces incorporating any of a number of active or passive display technologies (e.g., electronic-ink, LCD, LED or OLED, CRT, projection, holographic imagery, three dimensional imaging systems, etc.) or technologies for the display of Braille or other tactile information.

[0027] Memory 112 may include computer program instructions that the processor(s) 102 executes in order to implement one or more embodiments. The memory 112 generally includes RAM, ROM and/or other persistent or non-transitory computer-readable storage media. Memory 112 may store a presentation module 114 for managing the output of information to a display and/or other output device(s) 122 via the display interface 106 and/or input/output interface 110. The memory 112 may further include a user control module 116 for managing and obtaining user input information received for one or more input device 124 via the input/output device interface 110. In one embodiment, the user control module 116 may additionally interpret user input information in order to initiate playback of supplemental information. Memory 112 may further store a supplemental information module 118. In one embodiment, the supplemental information module 118 may detect the presence of supplemental information associated with a recently played or upcoming item of audio content (e.g., output via the presentation module 114). The supplemental information module 118 may cause the presentation module 114 to output an indication that the supplemental information is available. This indication may correspond to any type of output possible via the output devices 122. For example, the indication may correspond to audio output via a speaker or headphone. This audio content may include a tone, bell, voice indication, or other sound indicating the presence and availability of additional content. In addition, the indication may correspond to visual output via display interface 106. Still further, the indication may correspond to a haptic indication, such as a vibration caused by a haptic feedback device included with the display interface 106 or otherwise provided.

**[0028]** In addition, the supplemental information module 118 may receive and interpret user input via the user control module 116 to determine whether to cause playback of supplemental information. When a command to play supplemental information is received, the supplemental information module 118 may cause playback of supplemental information associated with the current point of audio playback via the output device 122. Still further, the supplemental information module 118 may interpret commands received during the playback of supplemental information. As described above, supplemental information itself may be associated with one or more items of supplemental information. Illustratively, the supplemental information module 118 may interpret input received during playback of a first item of supplemental information to indicate a command to play a second item of supplemental information associated with the first supplemental information. In some embodiments, a user may issue a command to stop playback of an item of supplemental information. The supplemental information module 118 may interpret a received input as such a command and cause the presentation module 114 to return to playback for the previous item of content (i.e., the supplemental information or primary audio content played before playback of a current item of audio content).

**[0029]** FIGURE 2 is a schematic block diagram depicting an illustrative operating environment in which a computing device of FIGURE 1 may interact with an electronic marketplace 150 to acquire audio content and supplemental information. As illustrated in FIGURE 2, the operating environment includes one or more user computing devices 100, such as a computing device of FIGURE 1, in communication with the electronic marketplace 150 via a network 130.

**[0030]** Those skilled in the art will appreciate that the network 130 may be any wired network, wireless network or combination thereof. In addition, the network 130 may be a personal area network, local area network, wide area network, cable network, satellite network, cellular telephone network, or combination thereof. In the illustrated embodiment, the network 130 is the Internet. Protocols and components for communicating via the Internet or any of the other aforementioned types of communication networks are well known to those skilled in the art of computer communications and thus, need not be described in more detail herein.

**[0031]** Accordingly, a user, using his or her user computing device 100, may communicate with the electronic marketplace 150 regarding audio content. Supplemental information regarding audio content may also be made available by the electronic marketplace 150. In one embodiment a user, utilizing his or her computing device 100 may browse descriptions of audio content made available by the electronic marketplace 150. In another embodiment, a user, utilizing his or her computing device 100 may acquire desired audio content or supplemental information regarding audio content from the electronic marketplace 150.

**[0032]** The electronic marketplace 150 is illustrated in FIGURE 2 as operating in a distributed computing environment comprising several computer systems that are interconnected using one or more networks. More specifically, the electronic marketplace 150 may include a marketplace server 156, a content catalog 152, a supplemental information catalog 154, and a distributed computing environment 155 discussed in greater detail below. However, it may be appreciated by those skilled in the art that the electronic marketplace 150 may have fewer or greater components than are illustrated in FIGURE 1. In addition, the electronic marketplace 150 could include various Web services and/or peer-to-peer network configurations. Thus, the depiction of electronic marketplace 150 in FIGURE 1 should be taken as illustrative and not limiting to the present disclosure.

**[0033]** Any one or more of the marketplace server 156, the content catalog 152, and the supplemental information catalog 154 may be embodied in a plurality of components, each executing an instance of the respective marketplace server 156, content catalog 152, and supplemental information catalog 154. A server or other computing component implementing any one of the marketplace server 156, the content catalog 152, and the supplemental information catalog 154 may include a network interface, memory, processing unit, and computer readable medium drive, all of which may communicate with each other in any way of a communication bus. The network interface may provide connectivity over the network 130 and/or other networks or computer systems. The processing unit may communicate to and from memory containing program instructions that the processing unit executes in order to operate the respective marketplace server 156, content catalog 152, and supplemental information catalog 154. The memory may generally include RAM, ROM, other persistent and auxiliary memory, and/or any non-transitory computer-readable media.

**[0034]** In this illustrative example, the content catalog 152 and the supplemental information catalog 154 can be implemented by the distributed computing environment 155. In addition, in some embodiments, the marketplace server 156 or other components of the electronic marketplace 150 may be implemented by the distributed computing environment. In some embodiments, the entirety of the electronic marketplace 150 may be implemented by the distributed computing environment 150.

**[0035]** The distributed computing environment 155 may include a collection of rapidly provisioned and released computing resources hosted in connection with the electronic marketplace 150 or a third party. The computing resources may include a number of computing, networking and storage devices in communication with one another. In some embodiments, the computing devices may correspond to physical computing devices. In other embodiments, the computing devices may correspond to virtual machine instances implemented by one or more physical computing devices. In still other embodiments, computing devices may correspond to both virtual computing devices and physical computing devices.

**[0036]** For example, the distributed computing environment 155 may include a virtual local network or other virtual network between multiple remote computing nodes provided by creating an overlay network over one or more intermediate physical networks separating the multiple computing nodes. The overlay network may be implemented in various ways in various embodiments, such as without encapsulating communications by embedding virtual network address information for a virtual network in a larger physical network address space used for a networking protocol of the one or more intermediate physical networks. As one illustrative example, a virtual network may be implemented using 32-bit IPv4 (“Internet Protocol version 4”) network addresses, and those 32-bit virtual network addresses may be embedded as part of 128-bit IPv6 (“Internet Protocol version 6”) network addresses used by the one or more intermediate physical networks, such as by re-headering communication packets or other data transmissions (e.g., using Stateless IP/ICMP Translation, or SIIT), or otherwise modifying such data transmissions to translate them from a first networking protocol for which they are configured to a distinct second networking protocol. More generally, in some embodiments, an N-bit network address specified in accordance with a first network addressing protocol may be embedded as part of another M-bit network address specified in accordance with a second

network addressing protocol, with “N” and “M” being any integers that correspond to network addressing protocols. In addition, in at least some embodiments, an N-bit network address may be embedded in another network address using more or less than N bits of the other network address, such as if a group of N-bit network addresses of interest may be represented using a smaller number of bits (e.g., with L-bit labels or identifiers being mapped to particular N-bit network addresses and embedded in the other network addresses, where “L” is less than “N”). Additional details related to SIIT are available at “Request For Comments 2765—Stateless IP/ICMP Translation Algorithm”, February 2000, at <http://tools.ietf.org/html/rfc2765>, which is hereby incorporated by reference in its entirety. Various benefits may be obtained from embedding virtual network address information in physical network addresses for an underlying physical substrate network, including enabling an overlay of the virtual network on the physical substrate network without encapsulating communications or configuring networking devices of the physical substrate network. Furthermore, other information may similarly be embedded in the larger physical network address space for a communication between computing nodes in at least some embodiments and situations, such as an identifier specific to a particular virtual network that includes those computing nodes (e.g., a virtual network for an entity on whose behalf those computing nodes operate).

**[0037]** In addition, in at least some embodiments, the computing nodes between which communications are configured may be physical computing systems and/or may be virtual machines that are each hosted on one or more physical computing systems, and the communications may include transmissions of data (e.g., messages, packets, frames, streams, etc.) in various formats. Furthermore, in at least some embodiments, at least some of the computing nodes may be part of a program execution service (or “PES”) that executes multiple programs on behalf of multiple customers or other users of the service, such as a program execution service that uses multiple computing systems on multiple physical networks (e.g., multiple physical computing systems and networks within a data center). In at least some such embodiments, one or more system manager modules for the program execution service may track and/or manage which computing nodes belong to which virtual networks, such as based on the computing nodes executing programs on behalf of a customer or other entity, as discussed in greater detail below. Further examples of a distributed computing environment are given in

United States Patent No. 7,865,586, issued on January 4, 2011 and entitled “Configuring Communications Between Computing Nodes” which is hereby incorporated by reference in its entirety. A distributed computing environment may also be referred to as a cloud computing environment.

**[0038]** With further reference to FIGURE 2, illustrative components of the electronic marketplace 150 will now be discussed. The marketplace server 156 may enable browsing and acquisition of audio content and/or supplemental information relating to audio content available from the electronic marketplace 150. Further, the marketplace server 156 may transmit audio content and/or supplemental information to user computing devices 100.

**[0039]** The content catalog 152 may include information on audio content available from the electronic marketplace 150. The supplemental information catalog 154 may include supplemental information available from the electronic marketplace 150. Such supplemental information may include, by way of non-limiting example, supplemental information provided or generated by authors, editors, publishers, users of the electronic marketplace 150, or other third parties. Accordingly, the marketplace server 156 may obtain audio content information for audio content offered by the electronic marketplace 150, as well as supplemental information offered by the electronic marketplace 150, and may make such audio content and supplemental information available to a user from a single network resource, such as a Web site. A user may then acquire audio content and/or supplemental information from the electronic marketplace 150.

**[0040]** Illustratively, marketplace server 156 may generate one or more user interfaces through which a user utilizing a user computing device 100 or a distinct computing device, may browse audio content and/or supplemental information made available by the electronic marketplace 150, submit queries for matching audio content and/or supplemental information, view information and details regarding specific audio content and/or supplemental information, and acquire audio content and/or supplemental information.

**[0041]** After the user selects desired audio content and/or supplemental information from the electronic marketplace 150, the marketplace server 156 may facilitate the acquisition of the audio content and/or supplemental information. In this regard, the marketplace server 156 may receive payment information from the user computing device 100 or distinct computing

device. Further, the marketplace server 156 may, transmit the audio content and/or supplemental information to the user computing device 100.

[0042] In some embodiments, the marketplace server 156 may, subsequent to acquisition of an item of audio content, inform a user computing device 100 or newly available supplemental information which is associated with the audio content. In still more embodiments, the marketplace server 156 may enable streaming of audio content and/or supplemental information from the content catalog 152 or the supplemental information catalog 154 to a user computing device 100.

[0043] Though described herein with reference to catalogs associated with an electronic marketplace 150, in some embodiments, a playback device may obtain audio information or supplemental information from additional or alternative sources, such as third party content catalogs or supplemental information catalogs.

[0044] FIGURE 3 is an illustrative graphical representation or visualization of audio content including supplemental information. As shown in FIGURE 3, the audio content corresponds to the audio book "*The Adventures of Tom Sawyer*." The primary audio content 204 represents the content of the audio book excluding supplemental information. Illustratively, the primary audio content 204 can represent a visual map of audio content, such that the duration of the audio content 204 is displayed from left to right. As also shown in FIGURE 3, the primary audio content 204 is associated with supplemental information 206-214. Each item of supplemental information 206-214 is associated with one or more points  $X_1$ - $X_6$  in either the primary audio content 204 or another item of supplemental information 206-214. Though described herein with reference to specific points within the audio content 204, in some embodiments, supplemental information may be associated with a range of points within the audio content 204, or with a specified duration of audio content 204. As discussed above, supplemental information may include footnotes (supplemental information 206 and 210), editor commentary (supplemental information 208 and 212), or additional content (supplemental information 214). Additional content may correspond to other types of supplemental information described above, such as author commentary or commentary of other readers of the audio content.



**[0045]** An illustrative user interaction with the audio content 202 will now be described with reference to FIGURE 3. Illustratively, the audio content of the book 202 may be played by a computing device, such as device 100 of FIGURE 1, beginning at the left of 204 and proceeding to the right. At point  $X_1$ , the computing device 100 may indicate to the user that supplemental information 206 is available for playback. The user may input a command that the supplemental information 206 should be played. At this point, playback of the audio content 202 may temporarily cease, and playback of the supplemental information 206 may begin. After completion of playback of supplemental information 206, playback of the audio content 204 may resume. As discussed above, playback may resume at or near the point  $X_1$ , such as a set amount of time before  $X_1$  (e.g., 3 seconds), or the beginning of the sentence or paragraph containing  $X_1$ .

**[0046]** As playback proceeds to  $X_2$ , the device 100 can indicate the availability of supplemental information 208. As described above, such indication may correspond to a visual indication, audio indication, haptic indication, or any combination therein. As described above, the user may be given a period of time (e.g., 10 seconds) in which to command playback of supplemental information. In this example, if the user does not input a command to play supplemental information 208, playback of the audio content 202 continues.

**[0047]** At point  $X_3$ , the device 100 may indicate the availability of supplemental information 210, and receive a user command to play the content 210. As displayed in FIGURE 3, supplemental information 210 is itself associated with supplemental information 212. At the point  $X_6$  during playback of the supplemental information 210, the device 100 can indicate availability of supplemental information 212 and receive a command to play the content 212. During playback of supplemental information 212, the device 100 may receive a command from the user to stop playback of supplemental information 212. In some embodiments, this may cease playback of content 212 and resume playback of content 210 at or near the point  $X_6$ . In other embodiments, the command received during playback of supplemental information 212 may resume playback of the primary audio content 204 at or near the point  $X_3$ , the last point played of the audio content 204.

**[0048]** As playback of the primary audio content continues, a point  $X_4$  may be encountered which is associated with a previously played item of supplemental information 210. In some embodiments, if the supplemental information 210 has already been played, the device

100 may not indicate the presence of supplemental information 210 at X<sub>4</sub>. In other embodiments, an indication may be suppressed only if supplemental information 210 was played completely. In still other embodiments, an indication may always be played.

**[0049]** At point X<sub>5</sub>, the device 100 may indicate the presence of supplemental information 214. As described above, the user may input a command to the device 100 to indicate a desire to play supplemental information 214. If this input is received, supplemental information 214 is played. Otherwise, playback of the primary audio content 204 proceeds, until it reaches the end of audio content 204 or until user input is received which causes the audio content 204 to stop playing.

**[0050]** FIGURES 4A-4C depict an illustrative user interface 300 displayed by a computing device, such as computing device 100 of FIGURE 1, that allows playback and interaction with supplemented audio content. In these examples, the audio content illustratively corresponds to the audio content 204 of FIGURE 3. The title of the audio content 204 is displayed as *"The Adventures of Tom Sawyer"* 301. The user interface 300 contains a number of input controls 302-304, each of which may be selected by a user to display a different aspect of the user interface 300. As illustrated in FIGURE 4A, the input control 302, corresponding to "Now Playing" is currently selected. Further input controls 316-320 allow various inputs by the user, such as rewinding playback for a period of time with input control 316, starting and stopping playback with input control 318 (the display of which may alter between play and pause symbols depending on the playback state), and bookmarking a current position with input control 320. The interface includes audio content information 312, such as a graphic associated with the audio content, title, author, and narrator information, and a chapter indicator 309 that displays the current chapter of the audio content 204 that is ready for playback.

**[0051]** The interface 300 further includes content indicator 310 that indicates the content of the current of the chapter, as well as a progress indicator 311, which indicates the position of playback within the currently selected chapter. Illustratively, the position of the progress indicator 311 may correspond to position X<sub>1</sub> of FIGURE 3, which is associated with supplemental information 206. As such, the user interface 300 further includes an input control 314 which indicates the availability of the supplemental information 206. As described above, the input control 314 may be displayed for a period before or after the position X<sub>1</sub>, to

allow a user time to request playback of the supplemental information 206. After selection of the input control 314, playback of the audio content 204 may be temporarily ceased, and playback of the supplemental information 206 may begin.

**[0052]** FIGURE 4B depicts the illustrative user interface 300 during playback of an item of supplemental information, such as supplemental information 206. As shown in FIGURE 3B, content indicator 310 may be altered to indicate that playback of the primary audio content has been temporarily halted. Further, the user interface 300 may include a supplemental information title indicator 402 which describes the currently playing supplemental information. A supplemental information indicator 404 may be provided that displays the content and position within the currently playing supplemental information. The user interface 300 also includes an input control 406 which allows the user to stop playback of the current supplemental information and return to the primary audio content. As described above, playback of the primary audio content may resume at the point at which it ceased or a point nearby, such as the beginning of a paragraph or sentence, or a point some period (e.g., 3 seconds) prior. Though the user interface 300 of FIGURE 4B displays only input control 406 to return to the primary audio content, in some embodiments, an additional input control may be provided to play additional supplemental information associated with the currently playing supplemental information, as is described above.

**[0053]** FIGURE 4C depicts the illustrative user interface 300 after user selection of input control 308, which is configured to cause display of the supplemental information associated with the currently loaded audio content. In the current example, each input control 506-510 is selectable by a user to play the associated supplemental information independent of the primary audio content. In some embodiments, selection of an item of supplemental information via input controls 502-510 may cause the content to be played independent of the audio content. In other embodiments, selection of supplemental information via input controls 502-510 may cause the selected content to be played, and further cause playback of the audio content from a point with which the supplement content is associated. The user interface 300 of FIGURE 4C may further include an input control 520 which is selectable by a user to display a different aspect of the user interface 300, such as a portion of the user interface 300 which enables the user to specify preferences regarding supplemental information.

[0054] FIGURE 4D depicts the illustrative user interface 300 after user selection of input control 520, which is configured to cause display of a portion of the user interface 300 which enables a user to specify preferences regarding supplemental information. In some embodiments, such user preferences may be specific to the currently depicted audio content (e.g., “*The Adventures of Tom Sawyer*”). In other embodiments, user preferences may be specified for all audio content, or for specific sets of audio content.

[0055] The illustrative user interface 300 of FIGURE 4D contains user selectable input controls 552-558 which enable the user to specify various types of supplemental information which should be provided. For example, input control 552 corresponds to author provided supplemental information, which may include, for example, footnotes, endnotes, glossary information, or author commentary. Input control 554 corresponds to publisher provided supplemental information, which may include editor commentary or additional information provided by a publisher of audio content. Input controls 556 and 558 correspond to supplemental information associated with other users of the electronic marketplace 150. Specifically, input control 556 corresponds to users associated with the current user of the user interface 300. In some embodiments, these associated users may correspond to users designated as contacts or “friends” in the electronic marketplace 150. In other embodiments, these associated users may correspond to users designated as contacts or “friends” through external systems, such as one or more social networking systems in communication with the electronic marketplace 150. In addition, input control 558 corresponds to general users of the electronic marketplace 150 who are not necessarily designated as a contact or “friend” of the current user.

[0056] Illustratively, selection of one or more of the input controls 552-558 may cause the computing device 100 to retrieve supplemental information provided by the corresponding source (e.g., the author, the publisher, friends, or other users) during or prior to playback of audio content with which the supplemental information is associated. Supplemental information may be retrieved, for example, from the electronic marketplace 150 of FIGURE 2 via the network 130. In some embodiments, supplemental information may be retrieved from the electronic marketplace 150 regardless of user selection of input controls 552-558. For example, where supplemental information is retrieved prior to playback of corresponding audio content, retrieval of supplemental information regardless of user selection of input controls 552-558 may

enable a user to select new supplemental information for playback immediately, without waiting for the new supplemental information to be retrieved. In still more embodiments, user selection of one or more of the input controls 552-558 may enable or disable the availability of the associated supplemental information during playback of audio content. For example, de-selection of input control 552 may disable playback of supplemental content generated by the author of the audio content.

**[0057]** The user interface 300 as depicted in FIGURE 4D further includes input controls 560, which may enable a user to specify categories of desired supplemental information. For example, each item of supplemental information may be classified as one or more of “Top Rated,” “Funny,” “Insightful,” “Informative,” or “Interesting.” In some embodiments, such classification may be accomplished by the operator of the electronic marketplace 150. In further embodiments, such classification may be accomplished by users of the electronic marketplace 150. Illustratively, de-selection of one or more of the input controls 160 may disable playback of correspondingly tagged supplemental information. For example, if a particular item of supplemental information is classified by the electronic marketplace 150 as “funny,” and the user of the computing device 100 de-selects the input control 560 corresponding to “Funny,” then that particular item of supplemental content may not be available during playback of audio content. As a further example, if a particular item of supplemental information is classified by the electronic marketplace 150 as “top rated,” and the user of the computing device 100 selects the input control 560 corresponding to “Top Rated,” then that particular item of supplemental content may be made available during playback of audio content. In some embodiments, only specific types of supplemental information may be categorized. For example, supplemental information generated by contacts or general users may be categorized, while supplemental information generated by the author or publisher may not be. In these embodiments, input controls 560 may apply only to supplemental information which is categorized. In other embodiments, input controls 560 may apply to all supplemental information.

**[0058]** The user interface 300 as depicted in FIGURE 4D further includes input controls 562 and 563, which may enable a user to specify a limit to the amount of supplemental information provided during playback of audio content. For example, input control 562 may enable a user to specify that no more than three items of supplemental information should be

presented during any one minute of audio playback. As a further example, input control 563 may enable a user to specify a maximum duration of supplemental information that should be presented. In addition to limits per time period (e.g., minutes, hours, etc.), limits may, for example, be imposed over the course of a paragraph, page, chapter, book, or other measurement period.

**[0059]** After specifying one or more preferences regarding supplemental information, the user may return to portion of the user interface 300 displayed in FIGURE 4C by selection of input control 564.

**[0060]** Though not shown in FIGURE 4, in some embodiments, a user may provide preferences indicating one or more items, categories, or types of supplemental information that should be presented automatically, without output of an indication and without requiring user input. For example, a user may specify that all “top rated” items of supplemental information should be automatically presented, indicators should be provided for “funny” supplemental information, and no indicator should be provided for supplemental information only marked as “interesting.” As will be appreciated by those skilled in the art, such preferences may be combined to specify that any given item of supplemental information should be automatically presented, indicated for presentation or playback, or not indicated for presentation or playback.

**[0061]** FIGURE 5 is a flow diagram depicting an illustrative routine 600 for playback of supplemented audio content. The routine 600 may illustratively be implemented by the supplemental information module 118 of the computing device 100. The routine 600 begins at block 602, which causes the playback of a primary audio content, such as audio content 204. Playback may begin, for example, in response to user selection of input control 318 of FIGURES 4A-4C.

**[0062]** At block 604, the computing device 100 determines whether supplemental information is associated with a current position within the primary audio content. As described above, supplemental information may be made associated with a range of positions within a primary audio content (e.g., with a continuous 10 second range). If supplemental information is not available, playback continues at block 614, described below. If the current playback position is within such a range, and supplemental information is therefore available, the routine 600 proceeds to block 606, which outputs to the user an indication that supplemental information is

available. As described above, this indication may correspond to audio output by the device 100, such as a tone, bell, voice, or sound, to a visual output, such as the appearance of an input control on a display, or to haptic feedback, such as a vibration of the device 100. At block 608, the computing device 100 tests whether the user has entered a command to play the detected supplemental information. As described above, such a command may correspond to input via a display device or other input control, such as a physical button on the computing device 100 or an accessory connected to the computing device 100 (e.g., headphones). In some embodiments, the command may further correspond to a voice command from the user. If a command is not received, playback continues at block 614, described below. If a command is received, the routine 600 continues to block 610, which causes the playback of the supplemental information (and temporarily stops playback of the primary audio content). After playback of the supplemental information is completed, playback of the primary audio content resumes at block 614. Optionally, the routine 600 may also be configured to receive a user command at block 610 to cease the playback of the supplemental information and immediately resume playback of the primary audio content.

**[0063]** Additionally, as described above, some supplemental information may itself be associated with supplemental information. In these embodiments, additional instances of routine 600 may be executed at block 610, such that the user may indicate that secondary supplemental audio content should be played. As will be appreciated by one skilled in the art, each additional instance of block 610 may create an instance of routine 600, such that playback any configuration of supplemental information may be facilitated.

**[0064]** At block 614, playback of the primary audio content resumes. As discussed above, playback may be resumed at or near the point at which it was ceased. For example, playback may be resumed at a point prior to where playback was ceased, such as the beginning of a previous paragraph. At block 616, the routine 600 tests whether to end playback of the audio content. Playback may be ended, for example, in response to a user command or completion of an item of audio content. If playback is not ended, the routine continues at block 604, as described above. If playback is ended, the routine 600 may end.

**[0065]** It is to be understood that not necessarily all objects or advantages may be achieved in accordance with any particular embodiment described herein. Thus, for example,

those skilled in the art will recognize that certain embodiments may be configured to operate in a manner that achieves or optimizes one advantage or group of advantages as taught herein without necessarily achieving other objects or advantages as may be taught or suggested herein.

**[0066]** All of the processes described herein may be embodied in, and fully automated via, software code modules executed by one or more general purpose computers or processors, thus transforming the general purpose computers or processors into specifically configured devices. The code modules may be stored in any type of computer-readable medium or other computer storage device. Some or all the methods may alternatively be embodied in specialized computer hardware. In addition, the components referred to herein may be implemented in hardware, software, firmware or a combination thereof.

**[0067]** Conditional language such as, among others, "can," "could," "might" or "may," unless specifically stated otherwise, are otherwise understood within the context as used in general to convey that certain embodiments include, while other embodiments do not include, certain features, elements and/or steps. Thus, such conditional language is not generally intended to imply that features, elements and/or steps are in any way required for one or more embodiments or that one or more embodiments necessarily include logic for deciding, with or without user input or prompting, whether these features, elements and/or steps are included or are to be performed in any particular embodiment.

**[0068]** Conjunctive language such as the phrase "at least one of X, Y and Z," unless specifically stated otherwise, is otherwise understood with the context as used in general to convey that an item, term, etc. may be either X, Y or Z. Thus, such conjunctive language is not generally intended to imply that certain embodiments require at least one of X, at least one of Y and at least one of Z to each be present.

**[0069]** Any process descriptions, elements or blocks in the flow diagrams described herein and/or depicted in the attached figures should be understood as potentially representing modules, segments, or portions of code which include one or more executable instructions for implementing specific logical functions or elements in the process. Alternate implementations are included within the scope of the embodiments described herein in which elements or functions may be deleted, executed out of order from that shown, or discussed, including



substantially concurrently or in reverse order, depending on the functionality involved as would be understood by those skilled in the art.

[0070] It should be emphasized that many variations and modifications may be made to the above-described embodiments, the elements of which are to be understood as being among other acceptable examples. All such modifications and variations are intended to be included herein within the scope of this disclosure and protected by the following claims.

[0071] Clauses

1. A system for output of supplemental audio information, the system comprising:
  - one or more data stores that store audio content and one or more items of supplemental audio information, wherein each of the one or more items of supplemental audio information is associated with one or more playback positions within the audio content;
  - an input device that obtains an input of a user;
  - a display device that presents information regarding the audio content;
  - an output device that outputs the audio content; and
  - a processor in communication with the one or more data stores, the input device and the output device, the processor operative to:
    - cause output of the audio content via the output device;
    - determine that one of the one or more items of supplemental audio information is associated with a current playback position within the audio content;
    - cause output, via at least one of the output device and the display device, of an indication that supplemental audio information is available for playback;
    - obtain input from the user indicating that the determined item of supplemental audio information should be played; and
    - cause output of the determined item of supplemental audio information via the output device.
2. The system of Clause 1, wherein the processor is further operative to pause output of the audio content in response to the obtained input, and to resume output of the audio content after output of the determined item of supplemental audio information.

3. The system of Clause 2, wherein the output of the audio content is resumed at a point prior to the point at which the output of the audio content was paused.

4. The system of Clause 2, wherein the output of the audio content is resumed at the point at which the output of the audio content was paused.

5. The system of Clause 1, wherein the system comprises at least one of a personal music player, a personal computing device, a laptop, a tablet, an eBook reader, and a mobile phone.

6. The system of Clause 1, wherein the one or more items of supplemental audio information comprise at least one of footnotes, editor commentary, author commentary, user commentary, endnotes, glossary information, and appendices.

7. The system of Clause 1, wherein the determined item of supplemental audio information is associated with at least one information category.

8. The system of Clause 1, wherein the processor is further operative to obtain input from the user indicating a desired information category; and wherein the information category associated with the determined item of supplemental audio information corresponds to the desired information category.

9. The system of Clause 1, wherein the indication is at least one of a sound, a visual indication, and a vibration.

10. A computer-implemented method for playback of supplemented audio information to a user, wherein the supplemented audio information comprises a primary audio content and one or more items of supplemental audio information, and wherein each item of supplemental audio information is associated with at least one playback position within at least one of the primary audio content and one of the one or more items of supplemental audio information, the computer-implemented method comprising:

under control of one or more computing devices,

causing output of the primary audio content via an output device;

determining that one of the one or more items of supplemental audio information is associated with a current playback position within the primary audio content being output via the output device;

obtaining input from the user via an input device indicating that the determined item of supplemental audio information should be played; and

causing output of the determined item of supplemental audio information via the output device.

11. The computer-implemented method of Clause 10 further comprising receiving the determined item of supplemental audio information from a distinct computing device via a network.

12. The computer-implemented method of Clause 10, wherein the input is obtained in response to outputting an indication that the determined item of supplemental audio information is available for playback.

13. The computer-implemented method of Clause 10, wherein each of the one or more items of supplemental audio information is associated with a range of playback positions within the primary audio content.

14. The computer-implemented method of Clause 10, further comprising pausing output of the audio content in response to obtaining the input and resuming output of the audio content after completing output of the determined item of supplemental audio information.

15. The computer-implemented method of Clause 14, wherein resuming output of the audio content comprises resuming output of the audio content at a point prior to the point at which the output of the audio content was paused.

16. The computer-implemented method of Clause 15, wherein the point prior to the point at which the output of the audio content was paused is a specified duration prior to the point at which the output of the audio content was paused.

17. The computer-implemented method of Clause 15, wherein the point prior to the point at which the output of the audio content was paused is determined based at least in part on a text corresponding to the audio content.

18. The computer-implemented method of Clause 10, wherein the computer-implemented method is carried out by at least one of a personal music player, a personal computing device, a laptop, a tablet, an eBook reader, and a mobile phone.

19. The computer-implemented method of Clause 10, wherein a position within the determined item of supplemental audio information is associated one or more items of secondary supplemental audio information, the computer-implemented method further comprising:

determining that one of the one or more items of secondary supplemental audio information is associated with a current playback position within the determined item of supplemental audio information;

causing output of an indication that secondary supplemental audio information is available for playback;

obtaining input from the user indicating that the determined item of secondary supplemental audio information should be played; and

causing output of the determined item of secondary supplemental audio information.

20. A system for outputting supplemental audio information, the system comprising:

one or more data stores that store audio content and one or more items of supplemental audio information, wherein each of the one or more items of supplemental audio information is associated with one or more playback positions within the at least one of the audio content and one of the one or more items of supplemental audio information;

a processor in communication with the data store, the processor operative to:

cause output of audio content via an output device;

obtain input from the user via an input device indicating that the determined item of supplemental audio information should be played;

determine one of the one or more items of supplemental audio information associated with a current playback position within the audio content being output via the output device;

cause output of the determined item of supplemental audio information via the output device.

21. The system of Clause 20, wherein the processor is further operative to, in response to a determination of the item of supplemental audio information associated with a current playback position within the audio content, cause output of an indication that supplemental audio information is available for playback.

22. The system of Clause 20, wherein the processor is further operative to determine text corresponding to the current playback position within the audio content, and wherein the

determination of one of the one or more items of supplemental audio information associated with a current playback position within the audio content is based at least in part on the determined text corresponding to the current playback position.

23. The system of Clause 20, wherein at least one of the one or more items of supplemental audio information is associated with at least one information category.

24. A computer readable, non-transitory storage medium having at least one computer-executable module for providing supplemental audio information, the at least one computer-executable module comprising:

a supplemental audio information module operative to:

cause output of a primary audio content, wherein the primary audio content is associated with one or more items of supplemental audio information, and wherein each of the one or more items of supplemental audio information is associated with one or more playback positions within the at least one of the audio content and one of the one or more items of supplemental audio information;

obtain input indicating that the determined item of supplemental audio information should be played;

determine one or more items of supplemental audio information associated with a current playback position within the primary audio content; and

cause output of the determined item of supplemental audio information.

25. The computer readable, non-transitory storage medium of Clause 24, wherein the supplemental audio information module is further operative to, in response to a determination of the item of supplemental audio information associated with a current playback position within the primary audio content, cause output of an indication that supplemental audio information is available for playback.

26. The computer readable, non-transitory storage medium of Clause 24, wherein the one or more items of supplemental audio information comprise at least one of footnotes, editor commentary, author commentary, user commentary, endnotes, glossary information, and appendices.

27. The computer readable, non-transitory storage medium of Clause 24, wherein the supplemental audio information module is further operative to pause output of the primary audio

content in response to obtaining the input and resume output of the primary audio content after completing output of the determined item of supplemental audio information.

28. The computer readable, non-transitory storage medium of Clause 27, wherein resuming output of the primary audio content comprises resuming output of the primary audio content at a point prior to the point at which the output of the primary audio content was paused.

29. The computer readable, non-transitory storage medium of Clause 27, wherein the point prior to the point at which the output of the audio content was paused is determined based at least in part on a text corresponding to the primary audio content.

30. The computer readable, non-transitory storage medium of Clause 24, wherein the one or more items of supplemental audio information comprise commentary generated by a contact of the user.

31. The computer readable, non-transitory storage medium of Clause 30, wherein the contact comprises a contact associated with a provider of the primary audio content and a contact associated with a social networking system.

## WHAT IS CLAIMED IS:

1. A system for output of supplemental audio information, the system including:
  - one or more data stores that store audio content and one or more items of supplemental audio information, wherein each of the one or more items of supplemental audio information is associated with one or more playback positions within the audio content and a content category of a plurality of content categories;
  - an input device that obtains an input of a user;
  - a display device that presents information regarding the audio content;
  - an output device that outputs the audio content; and
  - a processor in communication with the one or more data stores, the input device, and the output device, the processor operative to:
    - receive a preference of the user specifying that an indication should be provided for an item of supplemental audio information associated with a preferred content category of the plurality of content categories;
    - cause output of the audio content via the output device;
    - determine that a first item of supplemental audio information of the one or more items of supplemental audio information is associated with a current playback position within the audio content;
    - identify, based at least in part on the first item of supplemental audio information, a first content category of the plurality of content categories;
    - determine that the first content category corresponds to the preferred content category;
    - cause output, via at least one of the output device and the display device, of the indication that the first item of supplemental audio information is available for playback;
    - obtain input from the user indicating that the first item of supplemental audio information should be played; and
    - cause output of the first item of supplemental audio information associated with the current playback position via the output device.

2. The system of Claim 1, wherein the processor is further operative to pause output of the audio content in response to the input obtained from the user, and to resume output of the audio content after output of the first item of supplemental audio information.
3. The system of Claim 2, wherein the output of the audio content is resumed at a point prior to the point at which the output of the audio content was paused.
4. The system of Claim 2, wherein the output of the audio content is resumed at the point at which the output of the audio content was paused.
5. The system of Claim 1, wherein the system includes at least one of a personal music player, a personal computing device, a laptop, a tablet, an eBook reader, and a mobile phone.
6. The system of Claim 1, wherein the first item of supplemental audio information includes at least one of a footnote, editor commentary, author commentary, user commentary, an endnote, glossary information, and an appendix.
7. The system of Claim 1, wherein the processor is further operative to obtain input from the user indicating the preference of the user.
8. The system of Claim 1, wherein the indication is at least one of a sound, a visual indication, and a vibration.
9. A system for outputting supplemental audio information, the system including:
  - one or more data stores that store audio content and one or more items of supplemental audio information, wherein each of the one or more items of supplemental audio information is associated with one or more playback positions within the at least one of the audio content and one of the one or more items of supplemental audio information;
  - a processor in communication with the data store, the processor operative to:
    - cause output of audio content via an output device;
    - determine that an item of supplemental audio information from the one or more items of supplemental audio information is associated with a current playback position within the audio content being output via the output device;



determine a user preference for a preferred content category of a plurality of content categories;

identify, based at least in part on the item of supplemental audio information, a first content category of the plurality of content categories;

determine that the first content category corresponds to the preferred content category;

based at least in part on the determination that the first content category corresponds to the preferred content category, cause output of an indication that supplemental audio information is available for playback;

obtain input from the user via an input device indicating that the supplemental audio information should be played;

cause output of the item of supplemental audio information via the output device.

10. The system of Claim 9, wherein the processor is further operative to determine text corresponding to the current playback position within the audio content, and wherein the determination of one of the one or more items of supplemental audio information associated with a current playback position within the audio content is based at least in part on the determined text corresponding to the current playback position.

11. The system of Claim 9, wherein the user preference further includes a maximum length of supplemental audio information preferred by a user, and wherein the processor is further operable to determine that a length of the item of supplemental audio information does not exceed the maximum length.

12. A computer-implemented method for playback of supplemented audio information to a user, wherein the supplemented audio information includes a primary audio content and one or more items of supplemental audio information, and wherein each item of supplemental audio information is associated with at least one playback position within at least one of the primary audio content and one of the one or more items of supplemental audio information, the computer-implemented method including:

under control of one or more computing devices,

causing output of the primary audio content via an output device;

determining that a first item of supplemental audio information of the one or more items of supplemental audio information is associated with a current playback position within the primary audio content being output via the output device;

determining a user preference for outputting an indication during playback of the primary audio content that an item of supplemental audio information associated with a preferred content category is available for playback;

identifying, based at least in part on the first item of supplemental audio information, a first content category of the plurality of content categories;

determining that the first content category corresponds to the preferred content category;

based at least in part on the user preference, causing output of the indication that the first item of supplemental audio information is available for playback;

obtaining input from the user via an input device, the input indicating that the first item of supplemental audio information should be played; and

causing output of the first item of supplemental audio information via the output device.

13. The computer-implemented method of Claim 12 further including receiving the first item of supplemental audio information from a distinct computing device via a network.

14. The computer-implemented method of Claim 12, wherein each of the one or more items of supplemental audio information is associated with a range of playback positions within the primary audio content.

15. The computer-implemented method of Claim 12, further including pausing output of the audio content in response to obtaining the input and resuming output of the audio content after completing output of the first item of supplemental audio information.

16. The computer-implemented method of Claim 15, wherein resuming output of the audio content includes resuming output of the audio content at a point prior to the point at which the output of the audio content was paused.

17. The computer-implemented method of Claim 16, wherein the point prior to the point at which the output of the audio content was paused is at a specified duration prior to the point at which the output of the audio content was paused.

18. The computer-implemented method of Claim 16, wherein the point prior to the point at which the output of the audio content was paused is determined based at least in part on a text corresponding to the audio content.

19. The computer-implemented method of Claim 12, wherein the computer-implemented method is carried out by at least one of a personal music player, a personal computing device, a laptop, a tablet, an eBook reader, and a mobile phone.

20. The computer-implemented method of Claim 12, wherein a position within the first item of supplemental audio information is associated with one or more items of secondary supplemental audio information, and wherein the computer-implemented method further includes:

determining that an item of secondary supplemental audio information from the one or more items of secondary supplemental audio information is associated with a current playback position within the first item of supplemental audio information;

causing output of an indication that secondary supplemental audio information is available for playback;

obtaining input from the user indicating that the item of secondary supplemental audio information should be played; and

causing output of the item of secondary supplemental audio information.

21. The computer-implemented method of Claim 12, wherein the user preference further includes a maximum number of indications desired to be output within a specified period of time, wherein the computer-implemented method further includes:

determining that less than the maximum number of indications has been output within the specified period of time, and wherein causing output of the indication that the first item of supplemental audio information is available for playback is responsive to determining that less than the maximum number of indications has been output within the specified period of time.

22. A computer readable, non-transitory storage medium having at least one computer-executable module including computer-executable instructions for providing supplemental audio information, the at least one computer-executable module including:

a supplemental audio information module that, when executed by a processor, causes the processor to:

cause output of a primary audio content, wherein the primary audio content is associated with one or more items of supplemental audio information, and wherein each of the one or more items of supplemental audio information is associated with one or more playback positions within the at least one of the audio content and one of the one or more items of supplemental audio information;

determine one or more items of supplemental audio information associated with a current playback position within the primary audio content;

determine a user preference for a preferred content category of a plurality of content categories;

identify, for at least one of the one or more items of supplemental audio information, a first content category of the plurality of content categories;

determine that the first content category corresponds to the preferred content category;

based at least in part on determining that the first content category corresponds to the first content category, cause output of an indication that supplemental audio information is available for playback;

obtain input indicating that the at least one supplemental audio information should be played; and

cause output of the at least one item of supplemental audio information.

23. The computer readable, non-transitory storage medium of Claim 22, wherein the content category corresponds to at least one of footnotes, editor commentary, author commentary, user commentary, endnotes, glossary information, and appendices.

24. The computer readable, non-transitory storage medium of Claim 22, wherein the supplemental audio information module is further operative to pause output of the primary audio content in response to obtaining the input and resume output of the primary audio content after completing output of the at least one item of supplemental audio information.

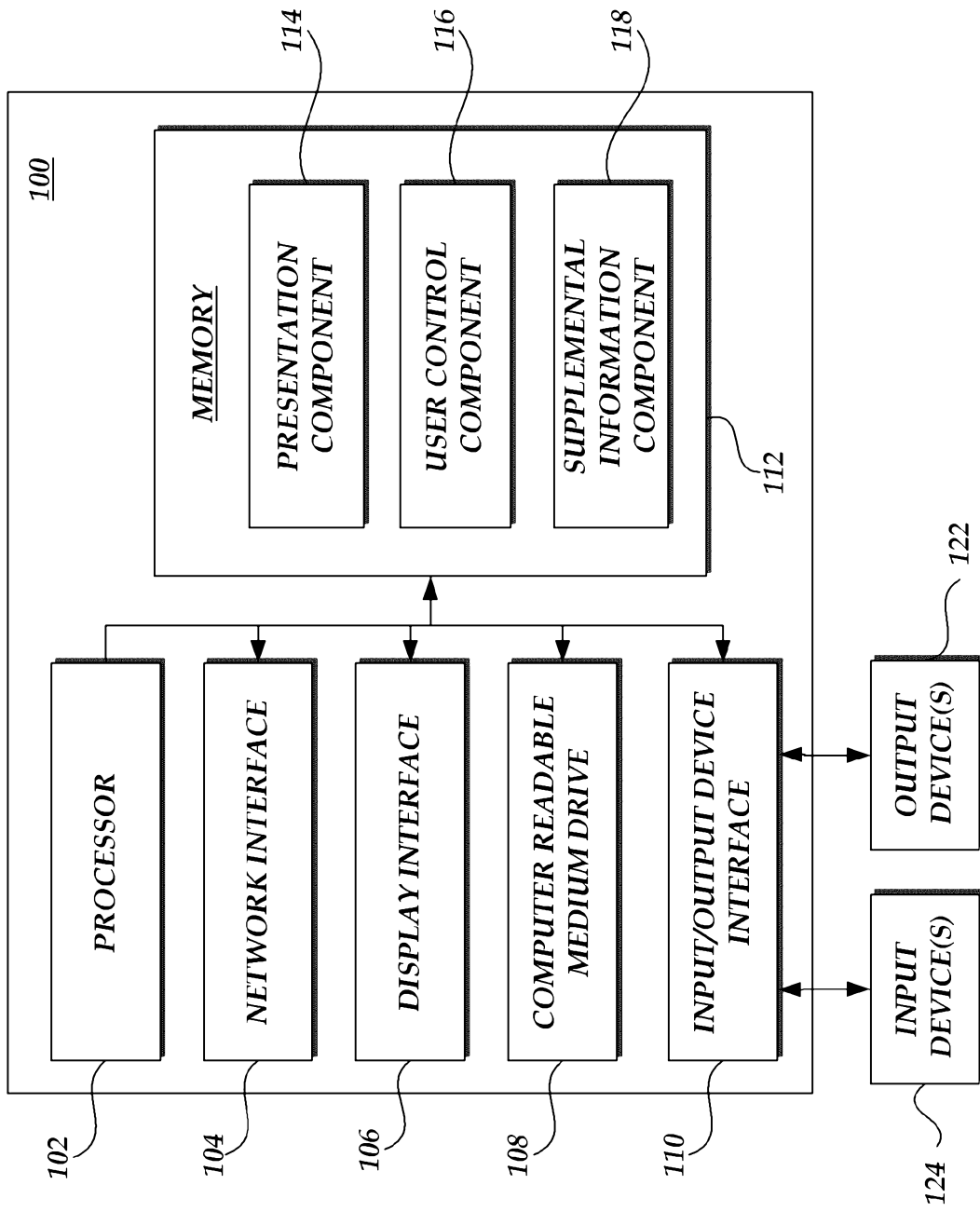
25. The computer readable, non-transitory storage medium of Claim 24, wherein resuming output of the primary audio content includes resuming output of the primary audio content at a point prior to the point at which the output of the primary audio content was paused.

26. The computer readable, non-transitory storage medium of Claim 24, wherein the point prior to the point at which the output of the audio content was paused is determined based at least in part on a text corresponding to the primary audio content.

27. The computer readable, non-transitory storage medium of Claim 22, wherein the at least one item of supplemental audio information includes commentary generated by a contact of the user.

28. The computer readable, non-transitory storage medium of Claim 27, wherein the contact includes at least one of a contact associated with a provider of the primary audio content and a contact associated with a social networking system.

2012359080 04 Aug 2015



**Fig. 1.**

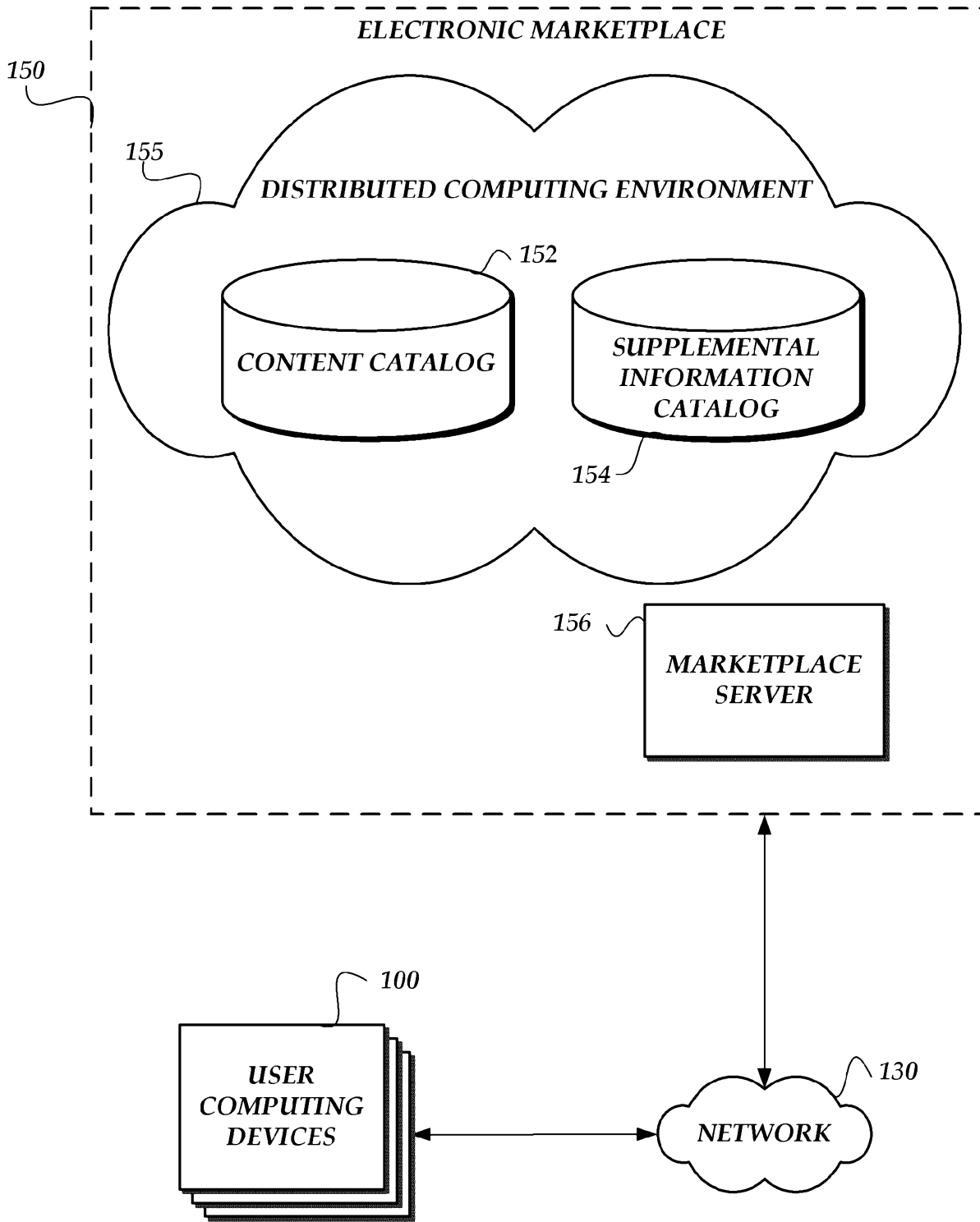


Fig.2.

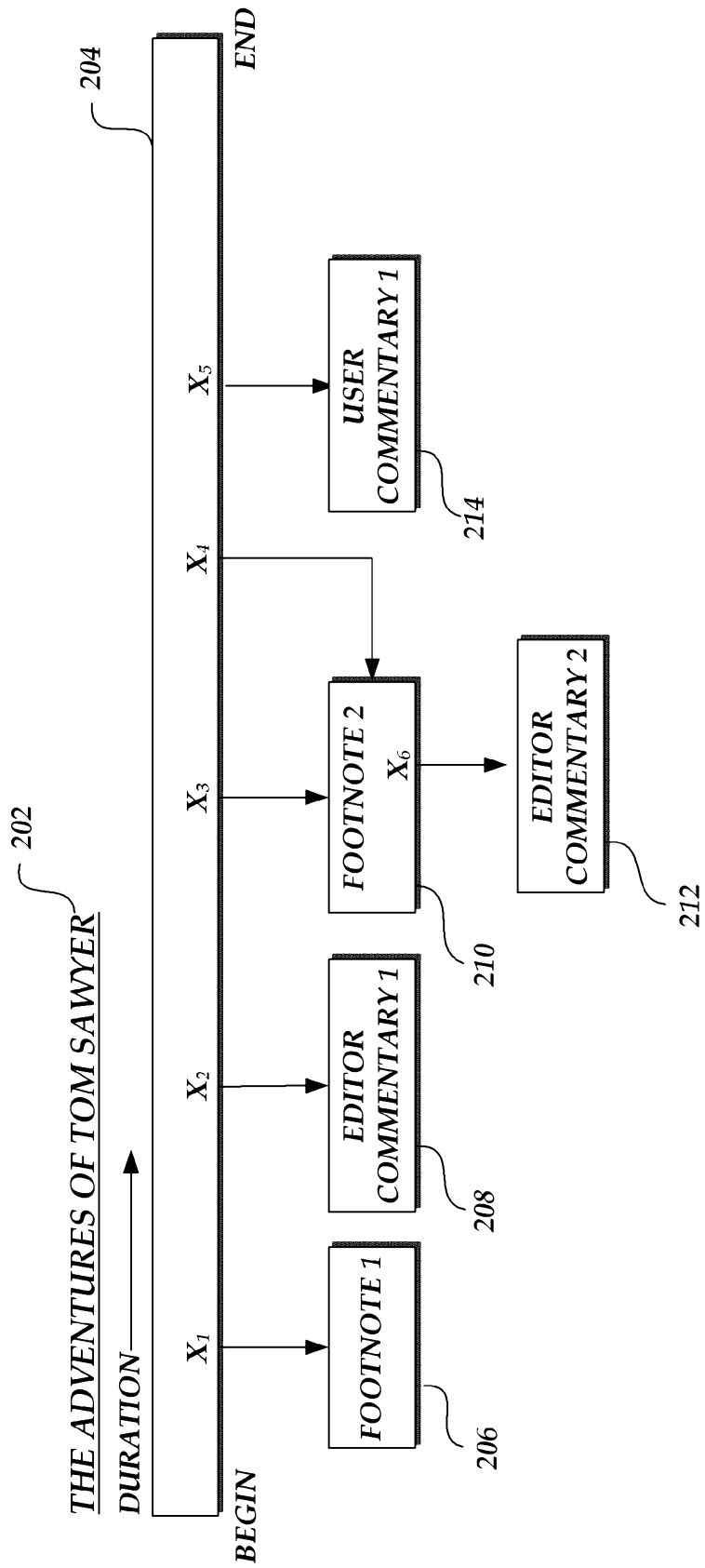
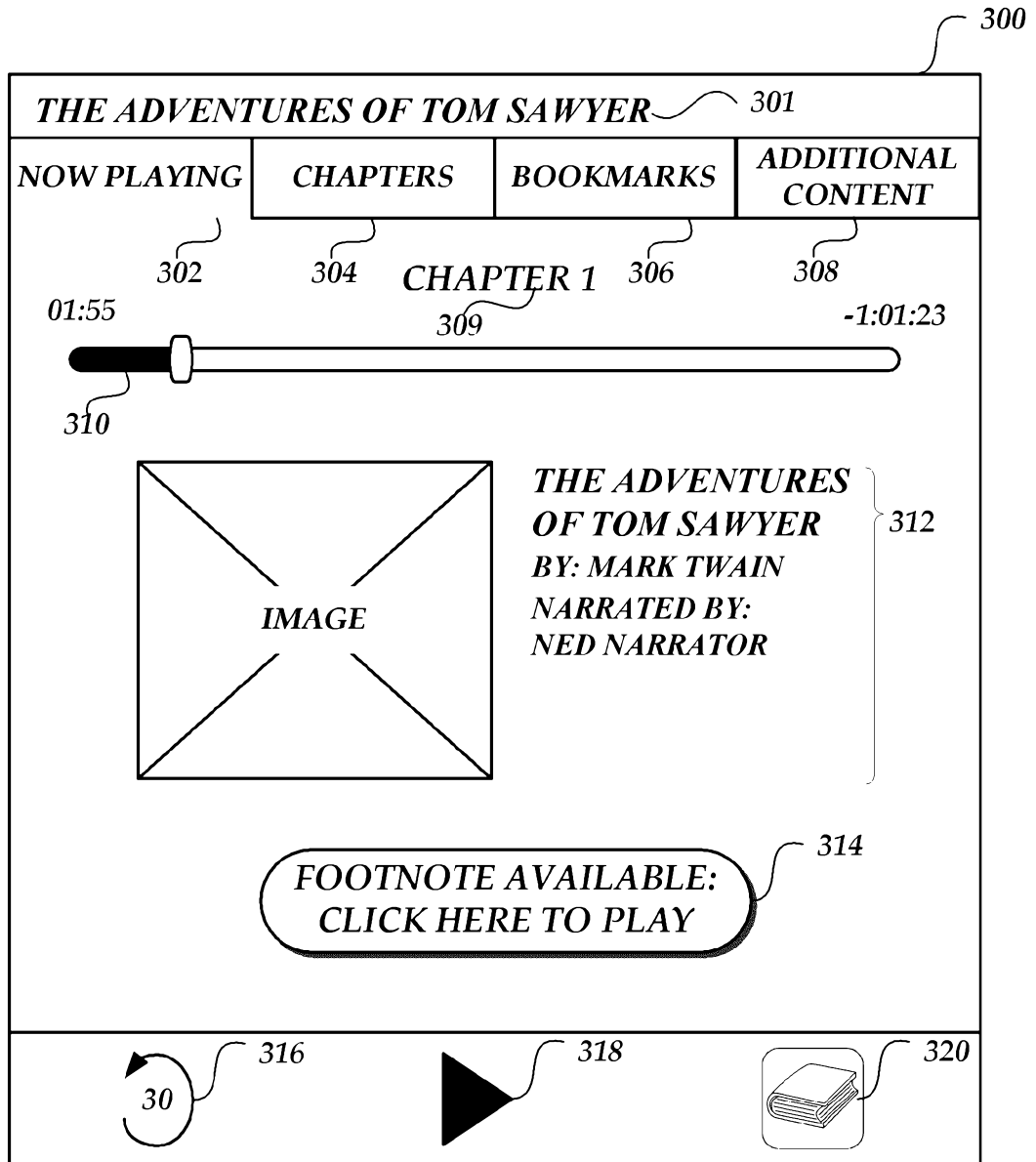
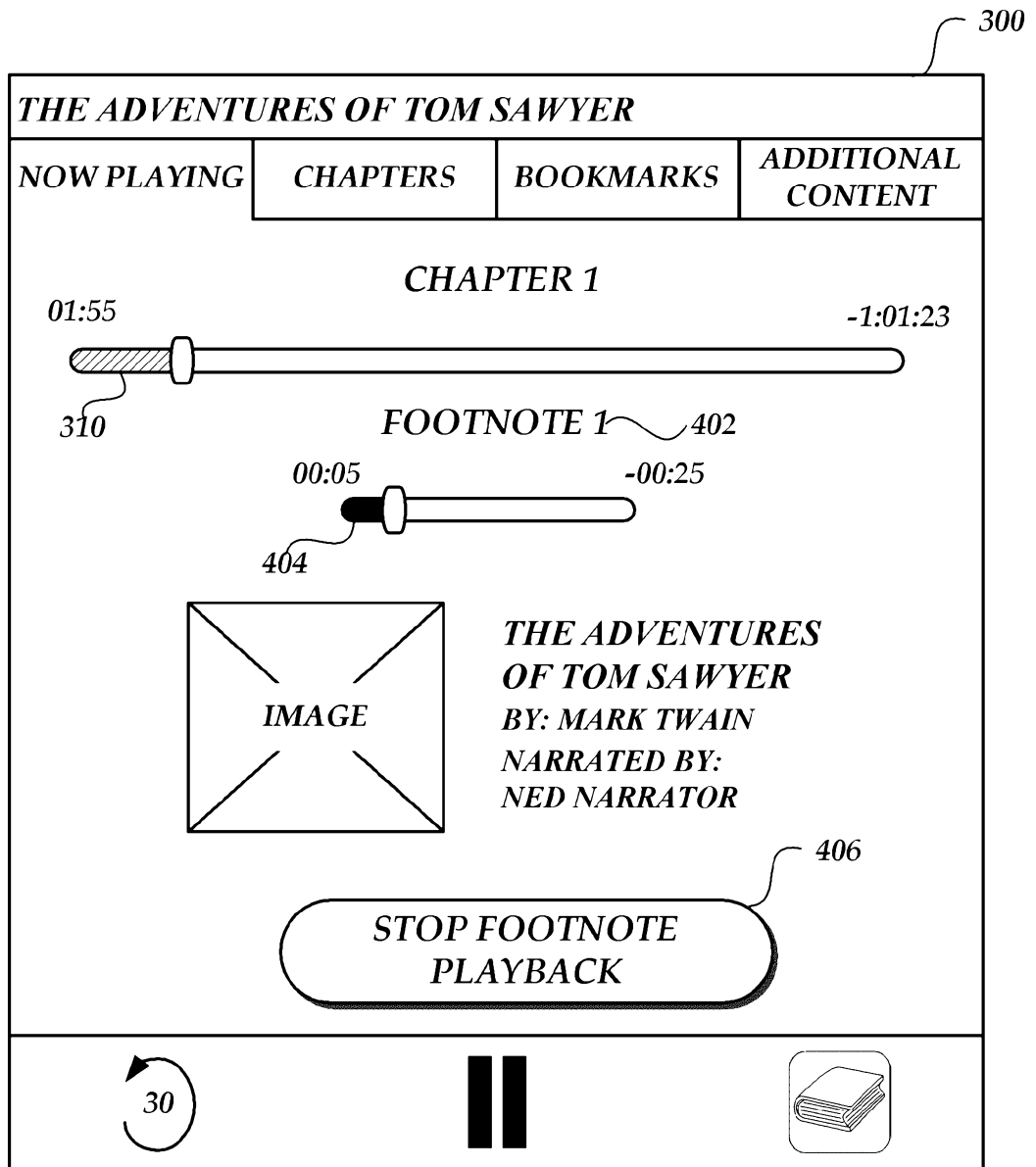


Fig.3.

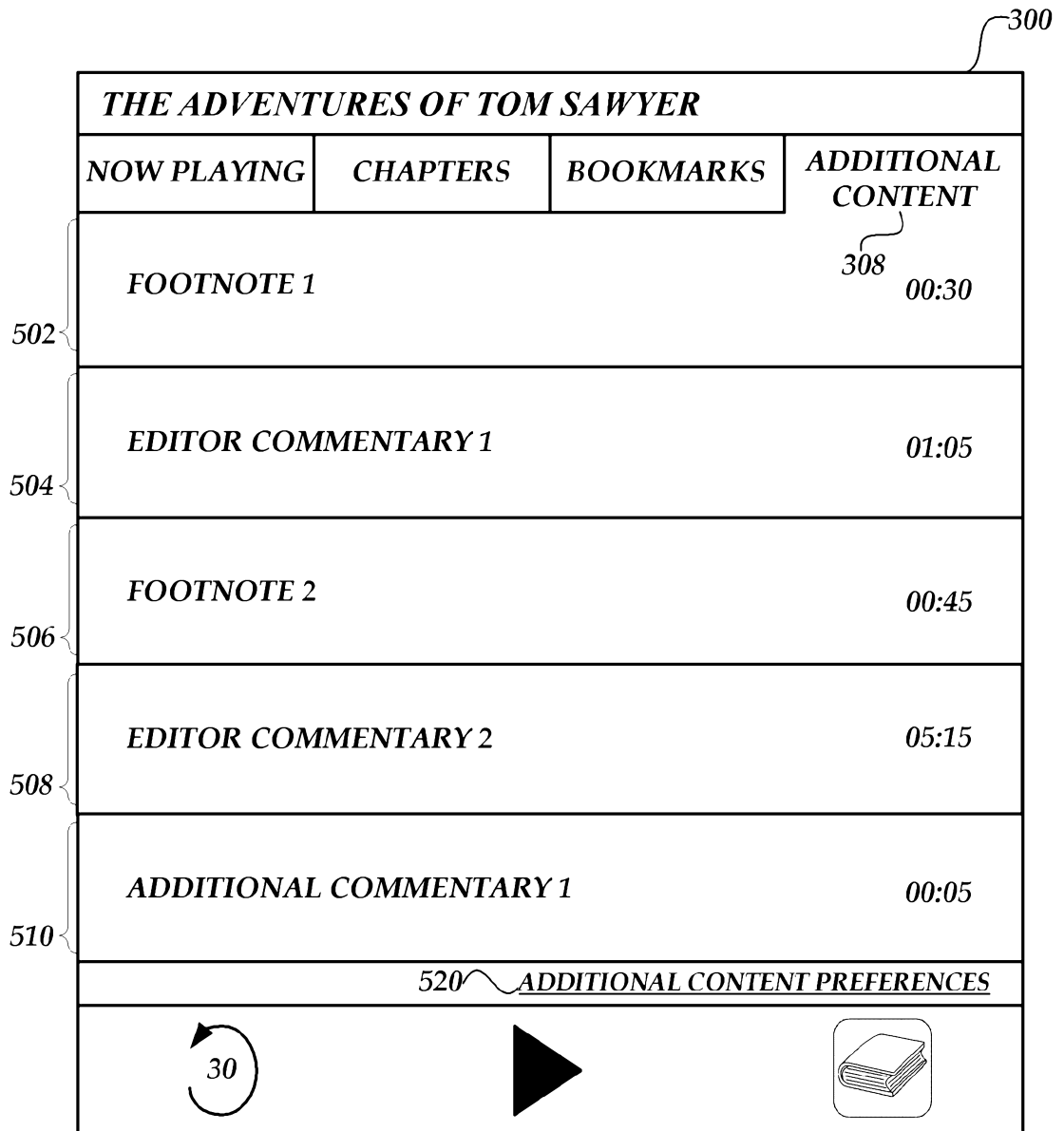




*Fig.4A.*



**Fig.4B.**



***Fig.4C.***

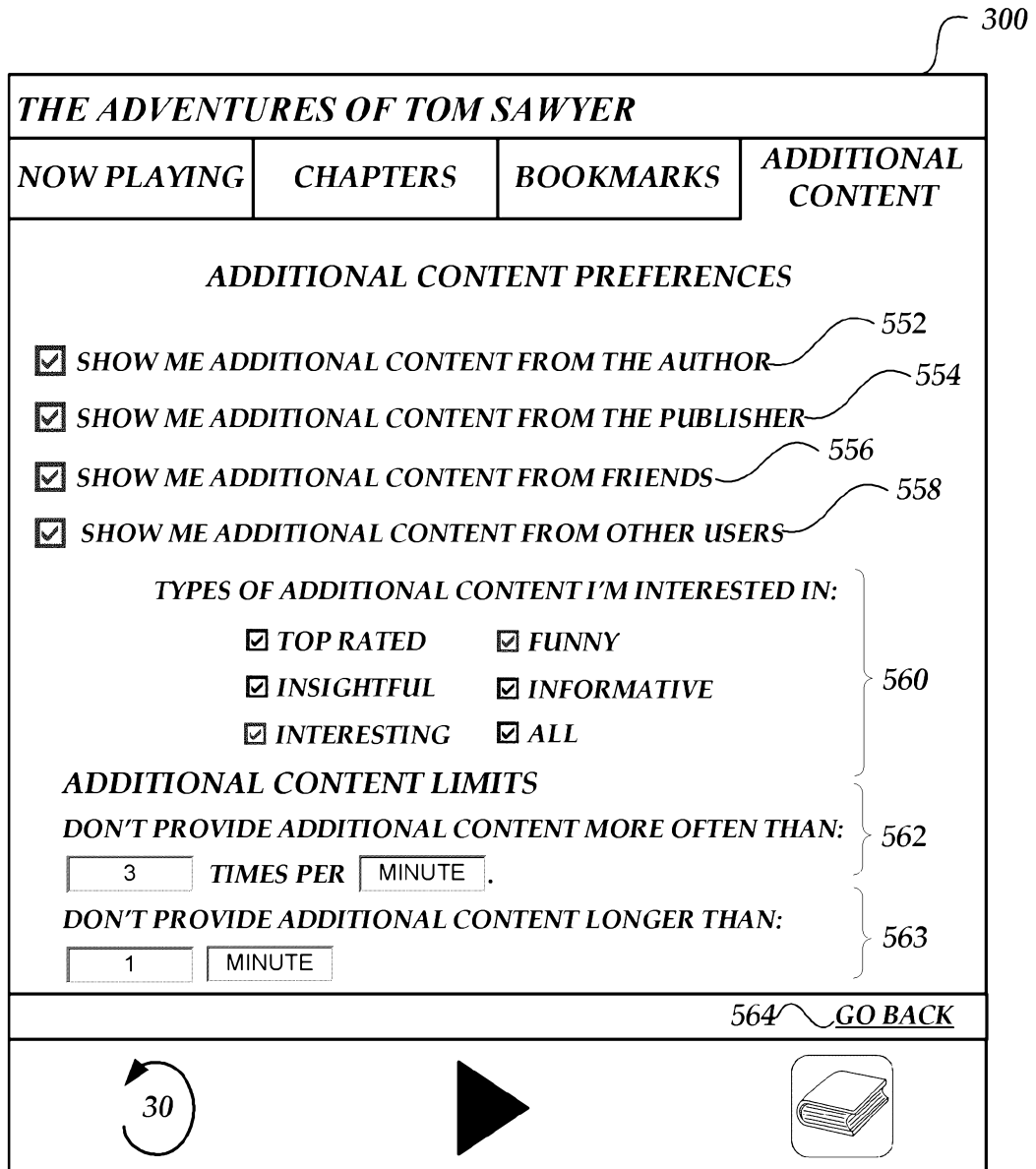


Fig.4D.

8/8

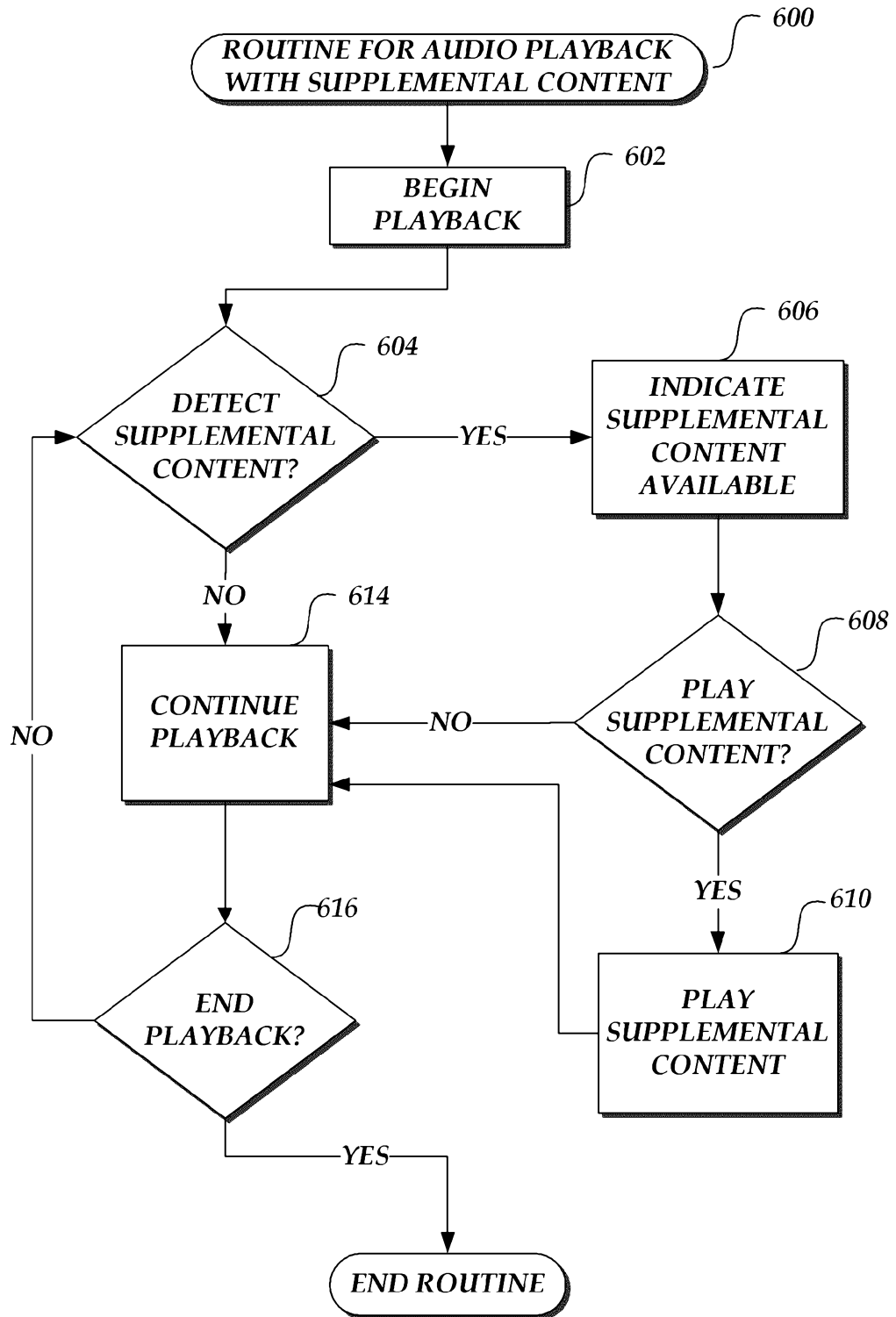


Fig.5.