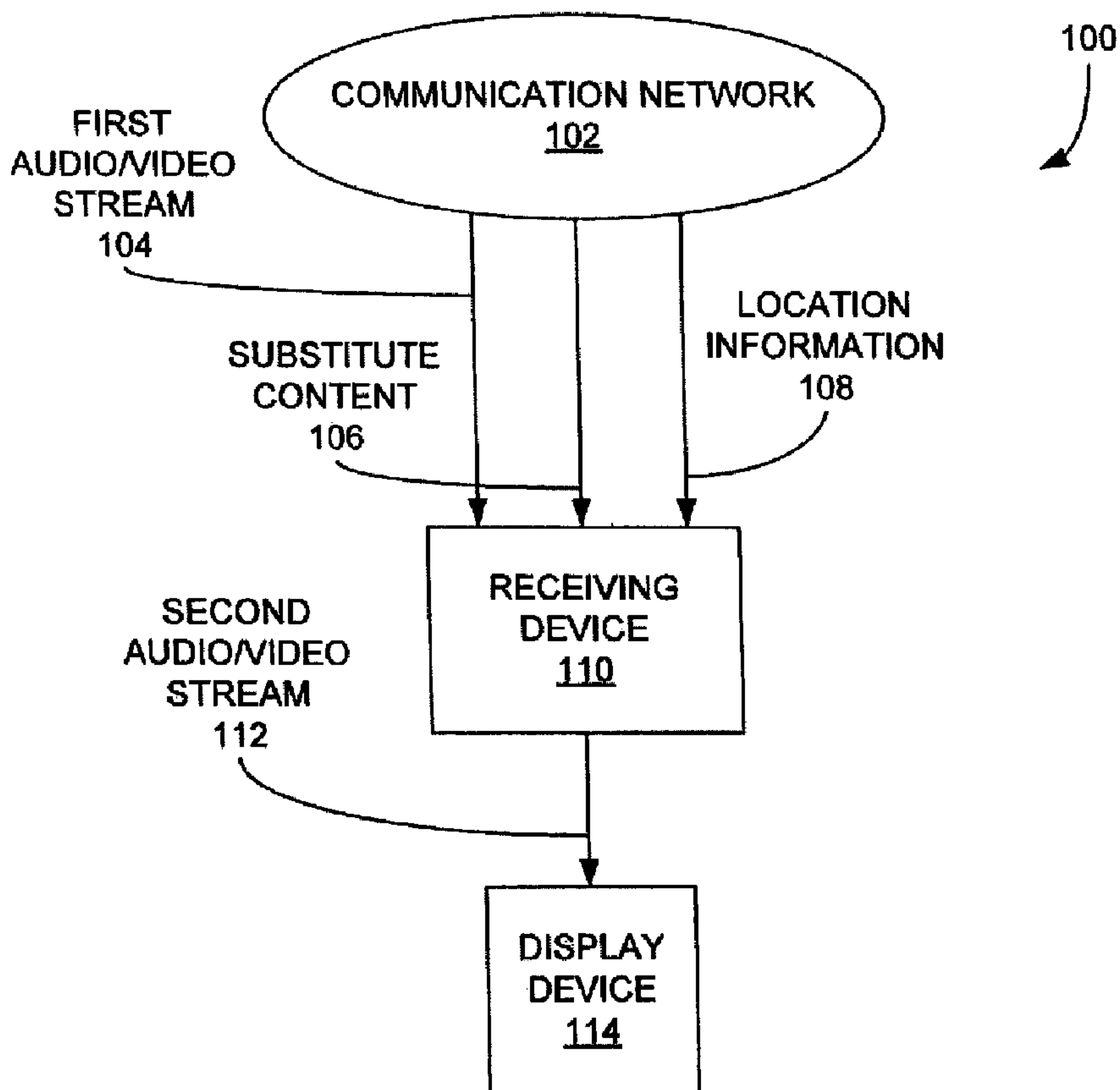




(22) Date de dépôt/Filing Date: 2009/05/12  
 (41) Mise à la disp. pub./Open to Public Insp.: 2009/11/30  
 (30) Priorité/Priority: 2008/05/30 (US12/130,792)

(51) Cl.Int./Int.Cl. *H04N 5/445* (2006.01),  
*H04N 5/76* (2006.01), *H04N 7/08* (2006.01)  
 (71) Demandeur/Applicant:  
ECHOSTAR TECHNOLOGIES L.L.C., US  
 (72) Inventeurs/Inventors:  
CASAGRANDE, STEVEN M., US;  
KUMMER, DAVID A., US  
 (74) Agent: SIM & MCBURNEY

(54) Titre : METHODES ET DISPOSITIF DE PRESENTATION DE CONTENU DE REMPLACEMENT DANS UN FLUX  
AUDIO-VIDEO DE DONNEES TEXTUELLES  
 (54) Title: METHODS AND APPARATUS FOR PRESENTING SUBSTITUTE CONTENT IN AN AUDIO/VIDEO STREAM  
USING TEXT DATA



(57) Abrégé/Abstract:  
 Various embodiments of apparatus and/or methods are described for skipping, filtering and/or replacing content from an audio/video stream using text data associated with the audio/video stream. The text data is processed using location information

(57) **Abrégé(suite)/Abstract(continued):**

that references a segment of the text data of the first audio/video stream to identify a location within the first audio/video stream. The location within the first audio/video stream is utilized to identify portions of the audio/video stream that are to be skipped during presentation. The portions of the audio/video stream that are to be skipped are filtered from the audio/video stream, and some of the skipped portions of the audio/video stream are replaced with substitute content. The filtered video stream, including the substitute content, is outputted for presentation to a user.

**METHODS AND APPARATUS FOR PRESENTING SUBSTITUTE CONTENT  
IN AN AUDIO/VIDEO STREAM USING TEXT DATA**

**ABSTRACT**

Various embodiments of apparatus and/or methods are described for skipping, filtering and/or replacing content from an audio/video stream using text data associated with the audio/video stream. The text data is processed using location information that references a segment of the text data of the first audio/video stream to identify a location within the first audio/video stream. The location within the first audio/video stream is utilized to identify portions of the audio/video stream that are to be skipped during presentation. The portions of the audio/video stream that are to be skipped are filtered from the audio/video stream, and some of the skipped portions of the audio/video stream are replaced with substitute content. The filtered video stream, including the substitute content, is outputted for presentation to a user.

## **METHODS AND APPARATUS FOR PRESENTING SUBSTITUTE CONTENT IN AN AUDIO/VIDEO STREAM USING TEXT DATA**

### **Background**

[0001] Digital video recorders (DVRs) and personal video recorders (PVRs) allow viewers to record video in a digital format to a disk drive or other type of storage medium for later playback. DVRs are often incorporated into set-top boxes for satellite and cable television services. A television program stored on a set-top box allows a viewer to perform time shifting functions, (e.g., watch a television program at a different time than it was originally broadcast). However, commercials within the recording may be time sensitive, and may no longer be relevant to the user when they finally get around to watching the program. Thus, the user is essentially presented with commercials and other advertisements which are of little use to both the advertiser and the viewer.

### **Brief Description of the Drawings**

[0002] The same number represents the same element or same type of element in all drawings.

[0003] FIG. 1 illustrates an embodiment of a system for presenting content to a user.

[0004] FIG. 2 illustrates an embodiment of a graphical representation of a first audio/video stream received by the receiving device, and a second audio/video stream outputted by the receiving device.

[0005] FIG. 3 illustrates an embodiment of a second audio/video stream in which the substitute content is presented before the segments of a show.

[0006] FIG. 4 illustrates an embodiment of a second audio/video stream in which the substitute content is presented after the segments of a show.

[0007] FIG. 5 illustrates an embodiment in which the boundaries of a segment of an audio/video stream are identified based on a text string included with the text data associated with the audio/video stream.

[0008] FIG. 6 illustrates an embodiment of a receiving device for presenting a recorded audio/video stream.



**[0009]** FIG. 7 illustrates an embodiment of a system in which multiple receiving devices are communicatively coupled to a communication network.

**[0010]** FIG. 8 illustrates an embodiment of a process for presenting a recorded audio/video stream.

### **Detailed Description of the Drawings**

**[0011]** The various embodiments described herein generally provide apparatus, systems and methods which facilitate the reception, processing, and outputting of audio/video content. More particularly, the various embodiments described herein provide for the identification of portions of an audio/video stream that are to be skipped during presentation of the audio/video stream. The various embodiments further provide for the insertion of substitute content into locations of the audio/video stream during presentation. In short, various embodiments described herein provide apparatus, systems and/or methods for replacing content in an audio/video stream based on data included in or associated with the audio/video stream.

**[0012]** In at least one embodiment, the audio/video stream to be received, processed, outputted and/or communicated may come in any form of an audio/video stream. Exemplary audio/video stream formats include Motion Picture Experts Group (MPEG) standards, Flash, Windows Media and the like. It is to be appreciated that the audio/video stream may be supplied by any source, such as an over-the-air broadcast, a satellite or cable television distribution system, a digital video disk (DVD) or other optical disk, the internet or other communication networks, and the like. In at least one embodiment, the audio/video data may be associated with supplemental data that includes text data, such as closed captioning data or subtitles. Particular portions of the closed captioning data may be associated with specified portions of the audio/video data.

**[0013]** In various embodiments described herein, the text data associated with an audio/video stream is processed to identify portions of the audio/video stream. More particularly, the text data may be processed to identify boundaries of portions of the audio/video stream. The portions of the audio/video stream between identified boundaries may then be designated for presentation to a user, or may be designated for skipping during presentation of the audio/video stream. Thus, in at least one

embodiment, portions of an audio/video stream that a user desires to view may be presented to the user, and portions of the audio/video stream that a user desires not to view may be skipped during presentation of the audio/video stream. Further, substitute content may be identified for presentation in association with portions of the original audio/video stream. The substitute content may be inserted within any identified location of the audio/video stream. For example, the original commercials included in a recorded audio/video stream may be replaced with updated commercials during subsequent presentation of the recorded audio/video stream.

[0014] Generally, an audio/video stream is a contiguous block of associated audio and video data that may be transmitted to, and received by, an electronic device, such as a terrestrial (“over-the-air”) television receiver, a cable television receiver, a satellite television receiver, an internet connected television or television receiver, a computer, a portable electronic device, or the like. In at least one embodiment, an audio/video stream may include a recording of a contiguous block of programming from a television channel (e.g., an episode of a television show). For example, a digital video recorder may record a single channel between 7:00 and 8:00, which may correspond with a single episode of a television program. Generally, an hour long recording includes approximately 42 minutes of video frames of the television program, and approximately 18 minutes of video frames of commercials and other content that is not part of the television program.

[0015] The television program may be comprised of multiple segments of video frames, which are interspersed with interstitials (e.g., commercials). As used herein, an interstitial is the video frames of a recording that do not belong to a selected show (e.g., commercials, promotions, alerts, and other shows). A segment of video includes contiguous video frames of the program that are between one or more interstitials.

[0016] Further, an audio/video stream may be delivered by any transmission method, such as broadcast, multicast, simulcast, closed circuit, pay-per-view, point-to-point (by “streaming,” file transfer, or other means), or other methods. Additionally, the audio/video stream may be transmitted by way of any communication technology, such as by satellite, wire or optical cable, wireless, or other means. The audio/video stream may also be transferred over any type of communication network, such as the internet or other wide area network, a local area network, a private network, a mobile



communication system, a terrestrial television network, a cable television network, and a satellite television network.

[0017] FIG. 1 illustrates an embodiment of a system 100 for presenting content to a user. The system of FIG. 1 is operable for replacing audio/video content within a contiguous block of audio/video data with substitute content for presentation to a user. The system 100 includes a communication network 102, a receiving device 110 and a display device 114. Each of these components is discussed in greater detail below.

[0018] The communication network 102 may be any communication network capable of transmitting an audio/video stream. Exemplary communication networks include television distribution networks (e.g., over-the-air, satellite and cable television networks), wireless communication networks, public switched telephone networks (PSTN), and local area networks (LAN) or wide area networks (WAN) providing data communication services. The communication network 102 may utilize any desired combination of wired (e.g., cable and fiber) and/or wireless (e.g., cellular, satellite, microwave, and radio frequency) communication mediums and any desired network topology (or topologies when multiple mediums are utilized).

[0019] The receiving device 110 of FIG. 1 may be any device capable of receiving an audio/video stream from the communication network 102. For example, in the case of the communication network 102 being a cable or satellite television network, the receiving device 110 may be a set-top box configured to communicate with the communication network 102. The receiving device 110 may be a digital video recorder in some embodiments. In another example, the receiving device 110 may be computer, a personal digital assistant (PDA), or similar device configured to communicate with the internet or comparable communication network 102. While the receiving device 110 is illustrated as receiving content via the communication network 102, in other embodiments, the receiving device may receive, capture and record video streams from non-broadcast services, such as video recorders, DVD players, personal computers or the internet.

[0020] The display device 114 may be any device configured to receive an audio/video stream from the receiving device 110 and present the audio/video stream to a user. Examples of the display device 114 include a television, a video monitor, or similar

device capable of presenting audio and video information to a user. The receiving device 110 may be communicatively coupled to the display device 114 through any type of wired or wireless connection. Exemplary wired connections include coax, fiber, composite video and high-definition multimedia interface (HDMI). Exemplary wireless connections include WiFi, ultra-wide band (UWB) and Bluetooth. In some implementations, the display device 114 may be integrated within the receiving device 110. For example, each of a computer, a PDA, and a mobile communication device may serve as both the receiving device 110 and the display device 114 by providing the capability of receiving audio/video streams from the communication network 102 and presenting the received audio/video streams to a user. In another implementation, a cable-ready television may include a converter device for receiving audio/video streams from the communication network 102 and displaying the audio/video streams to a user.

[0021] In the system 100, the communication network 102 transmits each of a first audio/video stream 104, substitute content 106 and location information 108 to the receiving device 110. The first audio/video stream 104 includes audio data and video data. In one embodiment, the video data includes a series of digital frames, or single images to be presented in a serial fashion to a user. Similarly, the audio data may be composed of a series of audio samples to be presented simultaneously with the video data to the user. In one example, the audio data and the video data may be formatted according to one of the MPEG encoding standards, such as MPEG-2 or MPEG-4, as may be used in DBS systems, terrestrial Advanced Television Systems Committee (ATSC) systems or cable systems. However, different audio and video data formats may be utilized in other implementations.

[0022] Also associated with the first audio/video stream 104 is supplemental data providing information relevant to the audio data and/or the video data of the first audio/video stream 104. In one implementation, the supplemental data includes text data, such as closed captioning data, available for visual presentation to a user during the presentation of the associated audio and video data of the audio/video data stream 104. In some embodiments, the text data may be embedded within the audio/video stream during transmission across the communication network 102 to the receiving device 110. In one example, the text data may conform to any text data or closed captioning standard,



such as the Electronic Industries Alliance 708 (EIA-708) standard employed in ATSC transmissions or the EIA-608 standard. When the text data is available to the display device 114, the user may configure the display device 114 to present the text data to the user in conjunction with the video data.

[0023] Each of a number of portions of the text data may be associated with a corresponding portion of the audio data or video data also included in the audio/video stream 104. For example, one or more frames of the video data of the audio/video stream 104 may be specifically identified with a segment of the text data included in the first audio/video stream 104. A segment of text data (e.g., a string of bytes) may include displayable text strings as well as non-displayable data strings (e.g., codes utilized for positioning the text data). As a result, multiple temporal locations within the audio/video stream 104 may be identified by way of an associated portion of the text data. For example, a particular text string or phrase within the text data may be associated with one or more specific frames of the video data within the first audio/video stream 104 so that the text string is presented to the user simultaneously with its associated video data frames. Therefore, the particular text string or phrase may provide an indication of a location of these video frames, as well as the portion of the audio data synchronized or associated with the frames.

[0024] The communication network 102 also transmits substitute content 106 and location information 108 to the receiving device 110. The substitute content 106 and/or the location information 108 may be transmitted to the receiving device 110 together or separately. Further, the substitute content 106 and/or the location information 108 may be transmitted to the receiving device 110 together or separately from the first audio/video stream 104. Generally, the substitute content 106 is provided to replace or ~~supplant a portion of the first audio/video stream 104.~~ The location information 108 specifies locations within the first audio/video stream 104 that are to be skipped and/or presented during presentation of the audio/video data of the first audio/video stream 104 by the receiving device 110. For example, if the first audio/video stream 104 includes one or more segments of a television show interspersed with one or more interstitials, then the location information 108 may identify the locations of the segments, which are to be presented, and/or identify the locations of the interstitial, which are to be skipped.

**[0025]** The location information 108 may identify the boundaries of either the segments or the interstitials. More particularly, the location information 108 may reference the text data to identify a video location within the first audio/video stream 104. The video location may then be utilized to determine the boundaries of either the segments or the interstitials. Generally, the beginning boundary of a segment corresponds with the ending boundary of an interstitial. Similarly, the ending boundary of a segment corresponds with the beginning boundary of an interstitial. Thus, the receiving device 110 may utilize the boundaries of segments to identify the boundaries of the interstitials, and vice versa. In some embodiments, the first audio/video stream 104 may not include both segments and interstitials, but nonetheless may include portions of audio/video data that a user desires to skip during presentation of the audio/video content of the first audio/video stream 104. Thus, the location information 108 may identify which portions of the audio/video content of the first audio/video stream are to be presented and/or skipped during presentation to a user.

**[0026]** In at least one embodiment, the insertion location of the substitute content 106 may be designated by the location information 108. For example, the substitute content 106 may be designated to replace an interstitial of the first audio/video stream 104. However, other locations for the substitute content 106 may also be identified by either the location information 108 or by the receiving device 110. For example, the substitute content 106 may be presented before the beginning of audio/video data of the first audio/video stream 104.

**[0027]** The receiving device 110 is operable for processing the text data to identify the portions of the audio/video stream which are to be presented to a user. More particularly, the receiving device 110 operates to identify the segments of the audio/video stream 104 which are to be presented to a user. The receiving device 110 further identifies substitute content 106 to present in association with the identified segments of the first audio/video stream 104. The receiving device 110 outputs a second audio/video stream 112, including the segments of the first audio/video stream 104 and the substitute content 106, for presentation on the display device 114. Thus, in some embodiments, the receiving device 110 operates to filter the interstitials from the first audio/video stream



104 and replaces the interstitials with the substitute content when outputting the second audio/video stream 112.

[0028] FIG. 2 illustrates an embodiment of a graphical representation of the first audio/video stream 104 received by the receiving device 110, and a second audio/video stream 112 outputted by the receiving device 110. More particularly, FIG. 2 illustrates an embodiment in which an interstitial of the first audio/video stream 104 is replaced by the substitute content 106 during presentation of the second audio/video stream 112. FIG. 2 will be discussed in reference to the system 100 of FIG. 1.

[0029] The first audio/video stream 104 includes a first audio/video segment 202 of a show, an interstitial 204 and a second audio/video segment 206 of the show. Also indicated are beginning and ending boundaries 208 and 210 of the interstitial 204, which are indicated to the receiving device 110 (see FIG. 1) by way of the location information 108. It is to be recognized that the boundaries 208 and 210 of the interstitial 204 are also boundaries of the segments 202 and 206. The supplemental data of the audio/video stream 104 is not shown in FIG. 2 to simplify the diagram.

[0030] In the specific example of FIG. 2 the boundary 208 (e.g., the ending boundary of segment 202) is the starting point at which the substitute content 106 is to replace a portion of the first audio/video stream 104. Likewise, the boundary 210 (e.g., the beginning boundary of segment 206) is the ending point at which the substitute content 106 is to replace a portion of the first audio/video stream 104. In FIG. 2, the portion of the first audio/video stream 104 to be replaced is the interstitial 204, located between the segments 202 and 206. As a result of this replacement, a second audio/video stream 112 is produced, in which the substitute content 106 is presented in place of the interstitial 204 during presentation of the second audio/video stream 112.

~~[0031] While FIG. 2 illustrates the substitute content 106 replacing the interstitial 204, it is to be appreciated that other locations of the substitute content 106 may also be utilized. FIG. 3 illustrates an embodiment of a second audio/video stream 112B in which the substitute content 106 is presented before the segments 202 and 206. Thus, the second audio/video stream 112B includes the substitute content 106 followed by the segment 202 and the segment 206. The interstitial 204 (see FIG. 2) is thus skipped during presentation of the second audio/video stream 112B.~~

[0032] FIG. 4 illustrates an embodiment of a second audio/video stream 112C in which the substitute content 106 is presented after the segments 202 and 206. The second audio/video stream 112C includes the segment 202 followed by the segment 206 which is followed by the substitute content 106. Again, the interstitial 204 (see FIG. 2) is skipped during presentation of the second audio/video stream 112C. The substitute content 106 may be inserted at any logical location within the second audio/video stream 112B.

[0033] Returning to FIGS. 1 and 2, while the substitute content 106 is illustrated as having the same length as the interstitial 204, it is to be appreciated that the substitute content 106 may have a duration that is the same as, or different than the original content it replaces (e.g., interstitial 204). For example, the length of substitute commercials utilized during playback of the recording may be selected to maintain the original length of the recording. In another embodiment, the length of the substitute content 106 utilized may be significantly shorter or longer than the commercials or other content it replaces. For example, an interstitial may originally include four commercials totaling two minutes in length, and these four commercials may be replaced with a single commercial that is thirty seconds in length. In at least one embodiment, the receiving device 110 may restrict the user from utilizing trick mode functions (e.g., fast forwarding) in order to skip over the substitute content.

[0034] The substitute content 106 may be shown to the user to offset the costs associated with removing the original interstitials 204. Thus, by watching a substitute commercial, the user is able to avoid watching an additional 1.5 minutes of commercials that were originally in the show. In at least one embodiment, the substitute content 106 may also be selected to replace a commercial with a timelier commercial from the same advertiser. For example, a department store may have originally advertised a sale during the original broadcast of the show, but that particular sale may have since ended. Thus, the substitute content 106 may replace that particular commercial with another commercial advertising a current sale at the store.

[0035] In at least one embodiment, the substitute content may be selected based on characteristics or demographics of the user. For example, if the user is a small child, then a commercial for a toy may be selected, whereas if the viewer is an adult male, then a commercial for a sports car may be shown. In some embodiments, the characteristics



utilized may be viewing characteristics of the user. Thus, the receiving device 110 may track what the user watches, and the substitute content 106 may be selected based on the collected data. For example, if the user watches many detective shows, then the substitute content may be a preview for a new detective show on Friday nights, whereas, if the user watches many reality shows, then the substitute content may be a preview for the new season of a reality show on Thursday nights.

[0036] As described above, the receiving device 110 (see FIG. 1) may identify the boundaries 208 and 210 (see FIG. 2) of the first audio/video stream 104 by processing the text data associated with the first audio/video stream 104. The boundaries 208 and 210 are identified based on the location of one or more video locations within the first audio/video stream 104. More particularly, the beginning and ending boundaries of a segment of the first audio/video stream 104 may be specified by a single video location within the segment. Thus, each segment may be identified by a unique video location within the first audio/video stream 104.

[0037] To specify a video location within the first audio/video stream 104, the location information 108 references a portion of the text data associated with the first audio/video stream 104. A video location within the first audio/video stream 104 may be identified by a substantially unique text string within the text data that may be unambiguously detected by the receiving device 110. The text data may consist of a single character, several characters, an entire word, multiple consecutive words, or the like. Thus, the receiving device 110 may review the text data to identify the location of the unique text string. Because the text string in the text data is associated with a particular location within the first audio/video stream 104, the location of the text string may be referenced to locate the video location within the first audio/video location.

[0038] In some embodiments, multiple video locations may be utilized to specify the beginning and ending boundaries of a segment. In at least one embodiment, a single video location is utilized to identify the beginning and ending boundaries of a segment. The video location may be located at any point within the segment, and offsets may be utilized to specify the beginning and ending boundaries of the segment relative to the video location. In one implementation, a human operator, of a content provider of the first audio/video stream 104, bears responsibility for selecting the text string, the video



location and/or the offsets. In other examples, the text string, video location and offset selection occurs automatically under computer control, or by way of human-computer interaction. A node within the communication network 102 may then transmit the selected text string to the receiving device 110 as the location information 108, along with the forward and backward offset data.

[0039] FIG. 5 illustrates an embodiment in which the boundaries of a segment of an audio/video stream 500 are identified based on a text string included with the text data associated with the audio/video stream 500. FIG. 5 will be discussed in reference to system 100 of FIG. 1. The audio/video stream 500 includes a segment 502, an interstitial 504 and text data 506. The segment 502 is defined by a boundary 508 and a boundary 510. The location information 108 received by the receiving device 110 identifies the segment 502 using a selected string 518 and offsets 512 and 514. Each of these components is discussed in greater detail below.

[0040] The receiving device 110 reviews the text data 506 to locate the selected string 518. As illustrated in FIG. 5, the selected string 518 is located at the video location 516. More particularly, in at least one embodiment, the beginning of the selected string 518 corresponds with the frame located at the video location 516. After locating the video location 516, the receiving device 110 utilizes the negative offset 512 to identify the beginning boundary 508. Likewise, the receiving device 110 utilizes the positive offset 514 to identify the ending boundaries 510. The offsets 512 and 514 are specified relative to the video location 516 to provide independence from the absolute presentation times of the video frames associated with the boundaries 508 and 510 within the audio/video stream 500. For example, two users may begin recording a particular program from two different affiliates (e.g., one channel in New York City and another channel in Atlanta). Thus, the absolute presentation time of the boundaries 508 and 510 will vary within the recordings. The technique described herein locates the same video frames associated with the boundaries 508 and 510 regardless of their absolute presentation times within a recording.

[0041] In at least one embodiment, the receiving device 110 filters the content of the audio/video stream 500 by outputting the video content of segment 502, while omitting from the presentation the interstitial 504 located outside of the boundaries 508



and 510. The receiving device 110 may additionally present the substitute content 106 adjacent to either of the boundaries 508 and 510. In some embodiments, the receiving device 110 may output the video content within the boundaries 508 and 510 and may also present video content within another set of similar boundaries 508 and 510, thus omitting presentation of the interstitial 504.

**[0042]** In at least one embodiment, a receiving device 110 identifies a set of boundaries 508 and 510 for a portion of the audio/video stream 500, and omits presentation of the content within the boundaries while presenting the other video content that is outside of the boundaries 508 and 510. For example, a user may watch the commercials within a football game, while skipping over the actual video content of the football game.

**[0043]** Depending on the resiliency and other characteristics of the text data, the node of the communication network 102 generating and transmitting the location information 108 may issue more than one instance of the location information 108 to the receiving device 110. For example, text data, such as closed captioning data, is often error-prone due to transmission errors and the like. As a result, the receiving device 110 may not be able to detect some of the text data, including the text data selected for specifying the video location 516. To address this issue, multiple unique text strings may be selected from the text data 506 of the audio/video stream 500 to indicate multiple video locations (e.g., multiple video locations 516), each having a different location in the audio/video stream 500. Each string has differing offsets relative to the associated video location that point to the same boundaries 508 and 510. The use of multiple text strings (each accompanied with its own offset(s)) may thus result in multiple sets of location information 108 transmitted over the communication network 102 to the receiving device 110, each of which is associated with the segment 502. Each set of location information 108 may be issued separately, or may be transmitted in one more other sets.

**[0044]** The location information 108 and the substitute content 106 may be logically associated with one another to prevent incorrect association of the location information 108 with other substitute content 106 being received at the receiving device 110. To this end, the substitute content 106 may include an identifier or other indication

associating the substitute content 106 with its appropriate location information 108. Conversely, the location information 108 may include such an identifier, or both the substitute content 106 and the location information 108 may do so. Use of an identifier may be appropriate if the substitute content 106 and the location information 108 are transmitted separately, such as in separate data files. In another embodiment, the substitute content 106 and the location information 108 may be packaged within the same transmission to the receiving device 110 so that the receiving device 110 may identify the location information 108 with the substitute content 106 on that basis.

[0045] Further, both the substitute content 106 and the location information 108 may be associated with the first audio/video stream 104 to prevent any incorrect association of the data with another audio/video stream. Thus, an identifier, such as that discussed above, may be included with the first audio/video stream 104 to relate the audio/video stream 104 to its substitute content 106 and location information 108. In one particular example, the identifier may be a unique program identifier (UPID). Each show may be identified by a UPID. A recording (e.g., one file recorded by a receiving device between 7:00 and 8:00) may include multiple UPIDs. For example, if a television program doesn't start exactly at the hour, then the digital video recorder may capture a portion of a program having a different UPID. The UPID allows a digital video recorder to associate a particular show with its corresponding location information 108 and/or substitute content 106.

[0046] Use of an identifier in this context addresses situations in which the substitute content 106 and the location information 108 are transmitted after the first audio/video stream 104 has been transmitted over the communication network 102 to the receiving device 110. In another scenario, the substitute content 106 and the location information 108 may be available for transmission before the time the first audio/video stream 104 is transmitted. In this case, the communication network 102 may transmit the substitute content 106 and the location information 108 before the first audio/video stream 104.

[0047] A more explicit view of a receiving device 610 according to one embodiment is illustrated in FIG. 6. The receiving device 610 includes a communication interface 602, a storage unit 616, an audio/video interface 618 and control logic 620. In



some implementations, a user interface 622 may also be employed in the receiving device 610. Other components possibly included in the receiving device 610, such as demodulation circuitry, decoding logic, and the like, are not shown explicitly in FIG. 6 to facilitate brevity of the discussion.

[0048] The communication interface 602 may include circuitry to receive a first audio/video stream 604, substitute content 606 and location information 608. For example, if the receiving device 610 is a satellite set-top box, the communication interface 602 may be configured to receive satellite programming, such as the first audio/video stream 604, via an antenna from a satellite transponder. If, instead, the receiving device 610 is a cable set-top box, the communication interface 602 may be operable to receive cable television signals and the like over a coaxial cable. In either case, the communication interface 602 may receive the substitute content 606 and the location information 608 by employing the same technology used to receive the first audio/video stream 604. In another implementation, the communication interface 602 may receive the substitute content 606 and the location information 608 by way of another communication technology, such as the internet, a standard telephone network, or other means. Thus, the communication interface 602 may employ one or more different communication technologies, including wired and wireless communication technologies, to communicate with a communication network, such as the communication network 102 of FIG. 1.

[0049] Coupled to the communication interface 602 is a storage unit 616, which is configured to store both the first audio/video stream 604 and the substitute content 606. The storage unit 616 may include any storage component configured to store one or more such audio/video streams. Examples include, but are not limited to, a hard disk drive, an optical disk drive, and flash semiconductor memory. Further, the storage unit 616 may include either or both volatile and nonvolatile memory.

[0050] Communicatively coupled with the storage unit 616 is an audio/video interface 618, which is configured to output audio/video streams from the receiving device 610 to a display device 614 for presentation to a user. The audio/video interface 618 may incorporate circuitry to output the audio/video streams in any format recognizable by the display device 614, including composite video, component video, the

Digital Visual Interface (DVI), the High-Definition Multimedia Interface (HDMI), Digital Living Network Alliance (DLNA), Ethernet, Multimedia over Coax Alliance (MOCA), WiFi and IEEE 1394. Data may be compressed and/or transcoded for output to the display device 614. The audio/video interface 618 may also incorporate circuitry to support multiple types of these or other audio/video formats. In one example, the display device 614, such as a television monitor or similar display component, may be incorporated within the receiving device 610, as indicated earlier.

[0051] In communication with the communication interface 602, the storage unit 616, and the audio/video interface 618 is control logic 620 configured to control the operation of each of these three components 602, 616, 618. In one implementation, the control logic 620 includes a processor, such as a microprocessor, microcontroller, digital signal processor (DSP), or the like for execution of software configured to perform the various control functions described herein. In another embodiment, the control logic 620 may include hardware logic circuitry in lieu of, or in addition to, a processor and related software to allow the control logic 620 to control the other components of the receiving device 610.

[0052] Optionally, the control logic 620 may communicate with a user interface 622 configured to receive user input 623 directing the operation of the receiving device 610. The user input 623 may be generated by way of a remote control device 624, which may transmit the user input 623 to the user interface 622 by the use of, for example, infrared (IR) or radio frequency (RF) signals. In another embodiment, the user input 623 may be received more directly by the user interface 622 by way of a touchpad or other manual interface incorporated into the receiving device 610.

[0053] The receiving device 610, by way of the control logic 620, is configured to receive the first audio/video stream 604 by way of the communication interface 602, and store the audio/video stream 604 in the storage unit 616. The receiving device 610 is also configured to receive the substitute content 606 over the communication interface 602, possibly storing the substitute content 606 in the storage unit 616 as well. The location information 608 is also received at the communication interface 602, which may pass the location information 608 to the control logic 620 for processing. In another embodiment,



the location information 608 may be stored in the storage unit 616 for subsequent retrieval and processing by the control logic 620.

[0054] At some point after the location information 608 is processed, the control logic 620 generates and transmits a second audio/video stream 612 over the audio/video interface 618 to the display device 614. In one embodiment, the control logic 620 generates and transmits the second audio/video stream 612 in response to the user input 623. For example, the user input 623 may command the receiving device 610 to output the first audio/video stream 604 to the display device 614 for presentation. In response, the control logic 620 instead generates and outputs the second audio/video stream 612. As described above in reference to FIG. 1, the second audio/video stream 612 includes portions of the audio/ video data of the first audio/video stream 604, with the substitute content 606 also being presented in association with the portions of the first audio/video stream 604. In some embodiments, the substitute content 606 may replace portions of the original audio/video content of the first audio/video stream 604 at a location specified in the location information 608, as described in detail above with respect to the first audio/video stream 104 of FIG. 1. For example, the first audio/video stream 604 may include portions of a movie that are not appropriate for viewing by children. The substitute content 606 may be utilized to replace these portions of the first audio/video stream 604 with more appropriate portions of video content for output in the second audio/video stream 612. In other embodiments, the substitute content 606 may be utilized to augment portions of the first audio/video stream 604 which are presented as part of the second audio/video stream 612.

[0055] Depending on the implementation, the second audio/video stream 612 may or may not be stored as a separate data structure in the storage unit 616. In one example, ~~the control logic 620 generates and stores the entire second audio/video stream 612 in the~~ storage unit 616. The control logic 620 may further overwrite the first audio/video stream 604 with the second audio/video stream 612 to save storage space within the storage unit 616. Otherwise, both the first audio/video stream 604 and the second audio/video stream 612 may reside within the storage unit 616.

[0056] In another implementation, the second audio/video stream 612 may not be stored separately within the storage unit 616. For example, the control logic 620 may

instead generate the second audio/video stream 612 “on the fly” by transferring selected portions of the audio data and the video data of the first audio/video stream 604 in presentation order from the storage unit 616 to the audio/video interface 618. At the point at which the substitute content 606 indicated by the location information 608 is to be outputted, the control logic 620 may then cause the substitute content 606 to be transmitted from the storage unit 616 to the audio/video interface 618 for output to the display device 614. Once the last of the substitute content 606 has been transferred from the storage unit 616, the control logic 620 may cause remaining portions of the first audio/video stream 604 which are to be presented to a user to be outputted to the audio/video interface 618 for presentation to the display device 614.

**[0057]** In one implementation, a user may select by way of the user input 623 whether the first audio/video stream 604 or the second audio/video stream 612 is outputted to the display device 614 by way of the audio/video interface 618. In another embodiment, a content provider of the first audio/video stream 604 may prevent the user from maintaining such control by way of additional information delivered to the receiving device 610.

**[0058]** If more than one portion of substitute content 606 is available in the storage unit 616 to replace a specified portion of the audio/video of the first audio/video stream 604 or augment the first audio/video stream 604, then the user may select via the user input 623 which of the substitute content 606 are to replace the corresponding portion of the audio data of the first audio/video stream 604 upon transmission to the display device 614. Such a selection may be made in a menu system incorporated in the user interface 622 and presented to the user via the display device 614. In other embodiments, the control logic 620 may select the substitute content 606 based on various criteria, such as information specified in the location information 608, user characteristics such a demographic information or user viewing characteristics.

**[0059]** In a broadcast environment, such as that depicted in the system 700 of FIG. 7, multiple receiving devices 710A-E may be coupled to a communication network 702 to receive audio/video streams, any of which may be recorded, in whole or in part, by any of the receiving devices 710A-E. In conjunction with any number of these audio/video streams, substitute content serving to replace content in an audio/video



stream or to augment content in an audio/video stream, as well as the location information for portions of the audio/video stream which are to be skipped and/or presented to a user, may be transferred to the multiple receiving devices 710A-E. In response to receiving the audio/video streams, each of the receiving devices 710A-E may record any number of the audio/video streams received. For any substitute content and associated location information that are transmitted over the communication network 702, each receiving device 710A-E may then review whether the received audio/video data segments and location information are associated with an audio/video stream currently stored in the device 710A-E. If the associated stream is not stored therein, the receiving device 710A-E may delete or ignore the related audio data segment and location information received.

[0060] In another embodiment, instead of broadcasting each possible substitute content and related location information, the transfer of an audio/video stream stored within the receiving device 710A-E to an associated display device 714A-E may cause the receiving device 710A-E to query the communication network 702 for any outstanding substitute content that apply to the stream to be presented. For example, the communication network 702 may comprise an internet connection. As a result, the broadcasting of each portion of substitute content and related location information would not be required, thus potentially reducing the amount of consumed bandwidth over the communication network 702.

[0061] FIG. 8 illustrates an embodiment of a process for presenting a recorded audio/video stream. The operation of FIG 8 is discussed in reference to filtering a broadcast television program. However, it is to be appreciated that the operation of the process of FIG. 8 may be applied to filter other types of video stream content. The operations of the process of FIG. 8 are not all-inclusive, and may comprise other operations not illustrated for the sake of brevity.

[0062] The process includes recording a first audio/video stream including at least one segment of a show and at least one interstitial of the show (operation 802). The process further includes recording supplemental data associated with the first audio/video stream (operation 804). The supplemental data includes closed captioning data associated with the first audio/video stream. Closed captioning data is typically



transmitted in two or four byte intervals associated with particular video frames. Because video frames don't always arrive in their presentation order, the closed captioning data may be sorted according to the presentation order (e.g., by a presentation time stamp) of the closed captioning data. In at least one embodiment, the sorted closed captioning data may then be stored in a data file separate from the first audio/video stream.

[0063] The process further includes receiving location information associated with the first audio/video stream (operation 806). The location information references the closed captioning data to identify a video location within the first audio/video stream. The location information may be utilized to filter portions of an audio/video stream, and may be further utilized to insert substitute content to locations within the audio/video stream. Operations 802 and 806 may be performed in parallel, sequentially or in either order. For example, the location information may be received prior to recording the audio/video stream, subsequently to recording the audio/video stream, or at the same time as the audio/video stream. In at least one embodiment, the location information is received separately from the first audio/video stream.

[0064] As described above, closed captioning data may be sorted into a presentation order and stored in a separate data file. In at least one embodiment, the sorting process is performed responsive to receiving the location information in step 806. Thus, a digital video recorder may not perform the sorting process on the closed captioning data unless the location information used to filter the audio/video stream is available for processing. In other embodiments, the closed captioning data may be sorted and stored before the location information arrives at the digital video recorder. For example, the sorting process may be performed in real-time during recording.

[0065] The process further includes processing the closed captioning data to identify boundaries of a segment of the first audio/video stream based on the video location (operation 808). More particularly, a text string included within the closed captioning data may be utilized to identify a specific location within the audio/video stream (e.g., a video location). The text string may be a printable portion of the text data or may comprise formatting or display options, such as text placement information, text coloring information and the like. The audio/video contained within the boundaries may then either be designated for presentation or may be skipped when the digital video



recorder outputs portions of the first audio/video stream to a display device. It is to be appreciated that operation 808 may identify either the boundaries of the segments of the interstitials or the segments of the show to filter the interstitials (or other portions of the first audio/video stream) from the audio/video stream.

**[0066]** Operation 808 may be performed to identify and skip portions of an audio/video stream for a variety of reasons. For example, a user may desire to skip commercials, portions of a television program or other content which is of no interest to the user, or portions of the audio/video stream which are offensive or should otherwise not be shown to certain users. The video location identified by a text string may be located within a portion of the audio/video stream that is designated for presentation (e.g., part of a television program), or may be within a portion of the audio/video stream that is designated for skipping (e.g., in a portion of the program that a user does not desire to view).

**[0067]** The process further includes identifying substitute content to present during presentation of the audio/video stream in association with the segments of the show (operation 810). The process further includes outputting a second audio/video stream for presentation on a presentation device (operation 812). The second audio/video stream includes at least one segment of the show and the substitute content. Thus, a user does not see the original interstitials of the show, but rather, may see the original segments of the show interspersed with substitute content. The substitute content may be presented during playback in any logical location of the audio/video stream.

**[0068]** For example, the substitute content may include a lead-in ad presented before the first segment of the show. In at least one embodiment, the segments of the show may then be presented back-to-back with no additional substitute content or interstitials presented there between. Thus, for the option of automatically filtering interstitials from within the show, the user may be presented with one or more lead-in ads, which may be specifically targeted to the user. This is advantageous to a user, because they receive automatic filtering of interstitials within the show. Likewise, advertisers and/or broadcasters benefit, because this ensures that a user will see at least some form of advertisement during playback of the recording. Otherwise, a viewer could

manually fast forward through all advertising, and the broadcaster and/or advertiser lose all benefit to the advertising slots within the program.

[0069] In some embodiments, the substitute content is presented at the original interstitial locations within the first audio/video stream. For example, a digital video recorder may present video frames between beginning and ending boundaries of a segment of the show. The substitute content may then be presented after a video frame of the segment that is associated with the ending boundary. In at least one embodiment, only some of the original interstitials are replaced with substitute content. Thus, other interstitials may be filtered from the original recording during playback, or even presented to the user during playback.

[0070] Thus, through the process illustrated in FIG. 8, broadcasters, advertisers and content providers (e.g., satellite television providers and cable providers) may offer various combinations of advertisement viewing during playback of recorded content. Advertisers can offer timelier and more relevant advertising to users that the users are more likely to view. Additionally, broadcasters and service providers may offer services which allow users to skip over some commercials within a recording, as long as the users are willing to watch some replacement commercials as well. This offers a compromise between the interests of broadcasters to reap the economic benefits of their television programs, while allowing users the advantages offered by time shifting devices.

[0071] Under another scenario, some programs may contain content that some users deem offensive or objectionable. To render the program palatable to a wider range of viewers, the content provider may make alternative content segments of the program available to viewers. A user who has recorded the program may then select a milder form of the audio/video content portion for viewing.

[0072] In each of these examples, the replacement audio/video content may be made available to the receiving device after the audio/video stream has been recorded at the device, thus providing a significant level of flexibility as to when the replacement audio data is provided.

[0073] Although specific embodiments were described herein, the scope of the invention is not limited to those specific embodiments. The scope of the invention is defined by the following claims and any equivalents therein.



## Claims

What is claimed is:

1. A method for presenting a recorded audio/video stream, the method comprising:
  - recording a first audio/video stream including at least one segment of a show and at least one interstitial of the show;
  - recording supplemental data associated with the first audio/video stream, the supplemental data including closed captioning data associated with the first audio/video stream;
  - receiving autonomous location information referencing the closed captioning data to identify a video location within the first audio/video stream;
  - processing the closed captioning data to identify boundaries of the at least one segment of the show based on the video location;
  - identifying substitute content to present in association with the at least one segment of the show; and
  - outputting a second audio/video stream for presentation on a display device, the second audio/video stream including the at least one segment of the show and the substitute content.
2. The method of claim 1, wherein receiving the autonomous location information further comprises:
  - receiving the autonomous location information separately from the first audio/video stream.
3. The method of claim 2, further comprising:
  - ~~sorting the closed captioning data according to a presentation order of the closed captioning data; and~~
  - storing the sorted closed captioning data in a data file separate from the first audio/video stream.
4. The method of claim 2, wherein outputting a second audio/video stream further comprises:

replacing the at least one interstitial with the substitute content.

5. The method of claim 2, wherein outputting a second audio/video stream further comprises:

outputting the substitute content before the at least one segment of the show in the second audio/video stream.

6. The method of claim 2, wherein receiving the location information further comprises:

receiving a string of bytes contained in the closed captioning data that is associated with the video location;

receiving a beginning offset, associated with the string of bytes, that is relative to the video location, the beginning offset identifying a beginning location of the at least one segment; and

receiving an ending offset, associated with the string of bytes, that is relative to the video location, the ending offset identifying an ending location of the at least one segment.

7. The method of claim 6, wherein outputting the second audio/video stream further comprises:

outputting the at least one segment of the first audio/video stream between the beginning location and the ending location; and

presenting the substitute content after presenting a video frame associated with the ending location.

8. The method of claim 6, wherein the string of bytes comprises a text string of displayable closed captioning data.

9. The method of claim 8, wherein the text string is unique within the at least one segment of the show.



10. The method of claim 6, wherein the string of bytes comprises a data string of non-displayable closed captioning data.
11. A receiving device comprising:  
a communication interface that receives a first audio/video stream including at least one segment of a show and at least one interstitial of the show, and that further receives supplemental data associated with the first audio/video stream, the supplemental data including closed captioning data associated with the first audio/video stream;  
a storage unit that stores the first audio/video stream and the supplemental data;  
control logic that:  
receives autonomous location information that references the closed captioning data to identify a video location within the first audio/video stream;  
processes the closed captioning data to identify boundaries of the at least one segment of the show based on the video location;  
identifies substitute content to present in association with the at least one segment of the show; and  
an audio/video interface that outputs a second audio/video stream for presentation on a display device, the second audio/video stream including the at least one segment of the show and the substitute content.
12. The receiving device of claim 11, wherein the processor receives the location information separately from the first audio/video stream.
13. The receiving device of claim 12, wherein the control logic sorts the closed captioning data according to a presentation order of the closed captioning data and stores the sorted closed captioning data in a data file separate from the first audio/video stream.
14. The receiving device of claim 12, wherein the audio/video interface replaces the at least one interstitial with the substitute content when outputting the second audio/video stream.

15. The receiving device of claim 12, wherein the audio/video interface outputs the substitute content before the at least one segment of the show in the second audio/video stream.
16. The receiving device of claim 12, wherein the location information received by the control logic includes:
  - a string of bytes contained in the closed captioning data that is associated with the video location;
  - a beginning offset, associated with the string of bytes, that is relative to the video location, the beginning offset identifying a beginning location of the at least one segment; and
  - an ending offset, associated with the string of bytes, that is relative to the video location, the ending offset identifying an ending location of the at least one segment.
17. The receiving device of claim 16, wherein the audio/video interface outputs the second audio/video stream including the at least one segment of the first audio/video stream between the beginning location and the ending location and the substitute content after a video frame that is associated with the ending location.
18. The receiving device of claim 16, wherein the string of bytes comprises a text string of displayable closed captioning data.
19. The receiving device of claim 18, wherein the text string is unique within the at least one segment of the show.
20. The receiving device of claim 16, wherein the text string is unique within the at least one segment of the show.
21. A method for presenting a recorded audio/video stream, the method comprising:
  - recording a first audio/video stream including at least one segment of a show and at least one interstitial of the show;



recording closed captioning data associated with the first audio/video stream;  
 receiving location information separately from the first audio/video stream, the  
 location information including at least one segment that is contained in the closed  
 captioning data, the at least one segment associated with a video location within the first  
 audio/video stream, beginning and ending offsets, associated with the at least one text  
 string, that are relative to the video location, the beginning and ending offsets identifying  
 beginning and ending locations of the at least one segment;  
 sorting the closed captioning data according to a presentation order;  
 processing the sorted closed captioning data to identify the video location within  
 the first audio/video stream based on the at least one text string;  
 identifying the beginning location and the ending location of the at least one  
 segment in the first audio/video stream based on the video location, the beginning offset  
 and the ending offset;  
 identifying substitute content;  
 replacing the at least one interstitial of the first audio/video stream with the  
 substitute content to generate a second audio/video stream; and  
 outputting the second audio/video stream for presentation on a display device.

22. The method of claim 21, wherein identifying the substitute content further comprises identifying the substitute content based on demographics of the user.
23. The method of claim 21, wherein identifying the substitute content further comprises identifying the substitute content based on viewing characteristics of the user.
24. A receiving device comprising:
- a communication interface that receives a first audio/video stream including at least one segment of a show and at least one interstitial of the show, and that further receives supplemental data, the supplemental data including closed captioning data associated with the first audio/video stream;
  - a storage unit that stores the first audio/video stream and the supplemental data;
  - control logic that:

sorts the closed captioning data according to a presentation order;  
receives location information separately from the first audio/video stream, the location information including at least one text string that is contained in the closed captioning data, the at least one text string associated with a video location within the first audio/video stream, beginning and ending offsets, associated with the at least one text string, that are relative to the video location, the beginning and ending offsets identifying beginning and ending locations of the at least one segment;

processes the sorted closed captioning data to identify the video location within the first audio/video stream based on the at least one text string;

identifies the beginning location and the ending location of the at least one segment within the first audio/video stream based on the video location, the beginning offset and the ending offset;

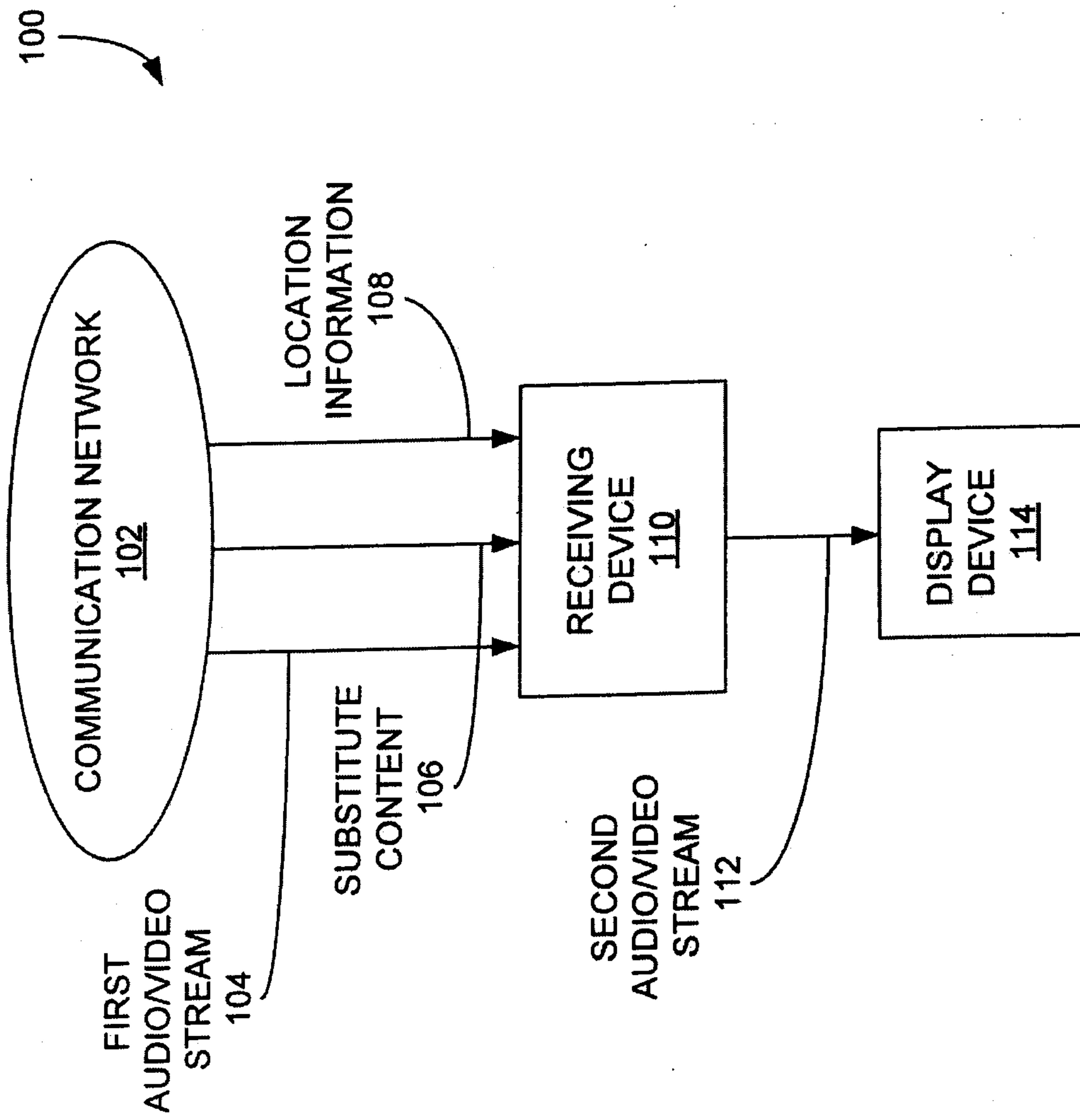
identifies substitute content; and

replaces the at least one interstitial of the first audio/video stream with the substitute content to generate a second audio/video stream; and  
an audio/video interface that outputs the second audio/video stream for presentation on a display device.

25. The receiving device of claim 24, wherein the control logic identifies the substitute content based on demographics of the user.

26. The receiving device of claim 24, wherein the control logic identifies the substitute content based on viewing characteristics of the user.





**FIG. 1**

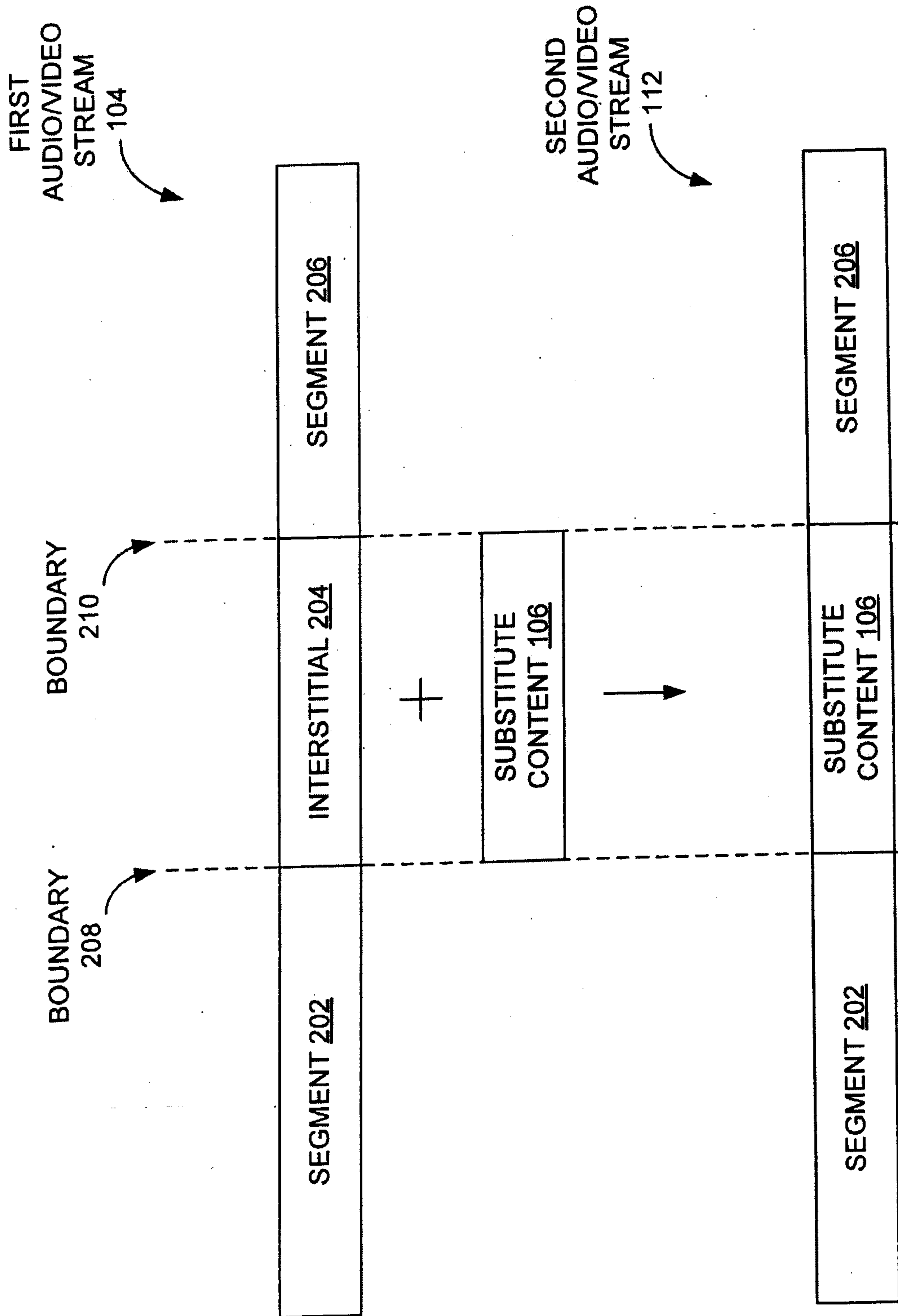


FIG. 2

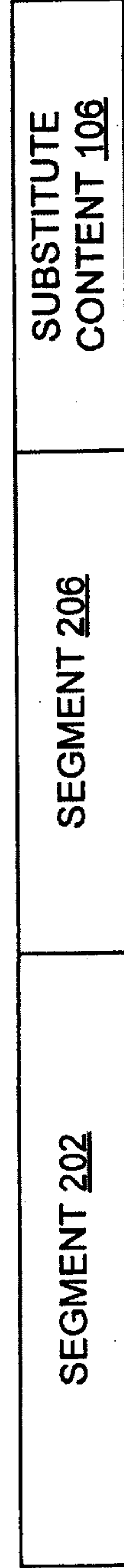


SECOND  
AUDIO/VIDEO  
STREAM  
112B



**FIG. 3**

SECOND  
AUDIO/VIDEO  
STREAM  
112C



**FIG. 4**

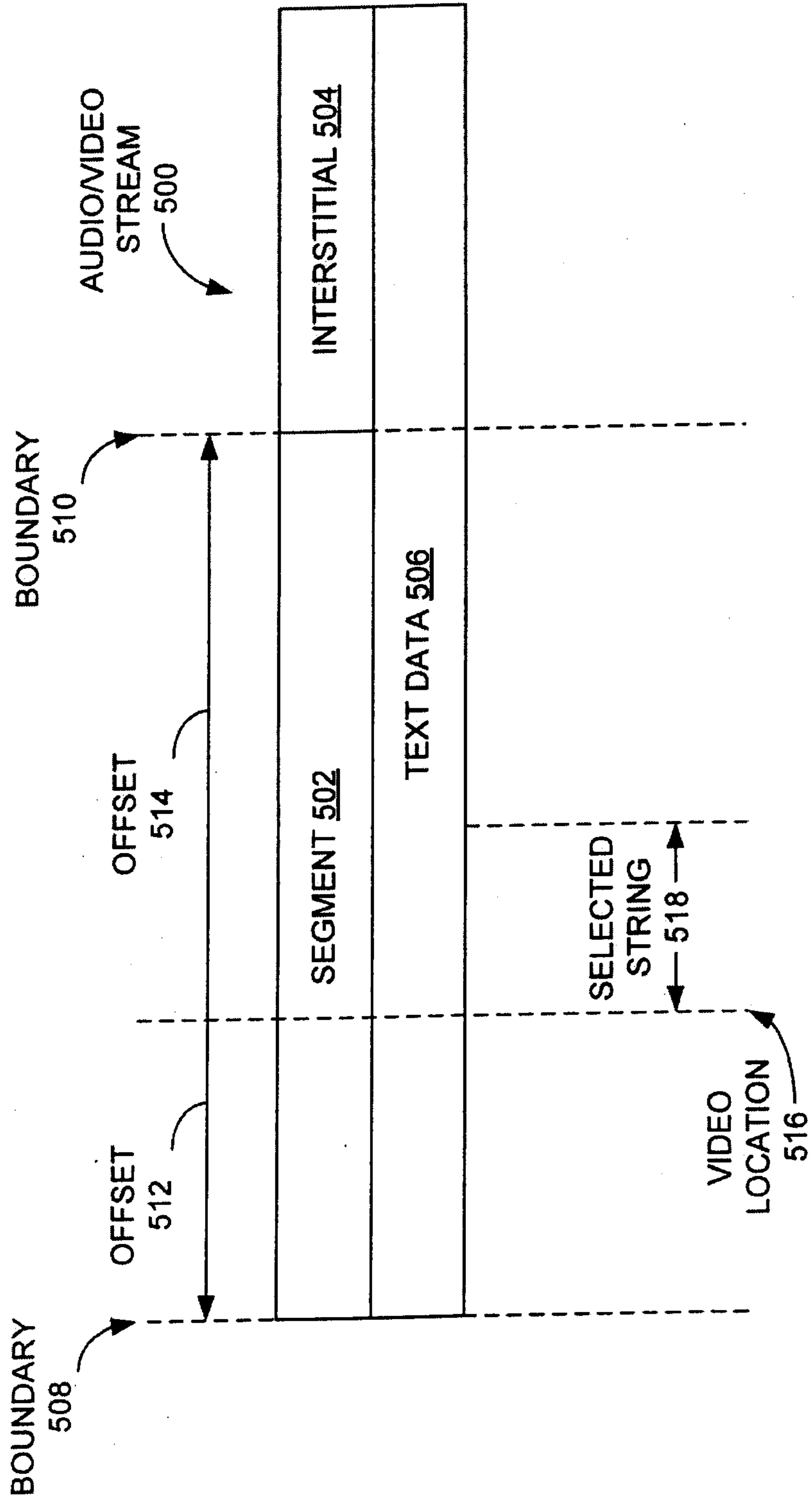


FIG.5



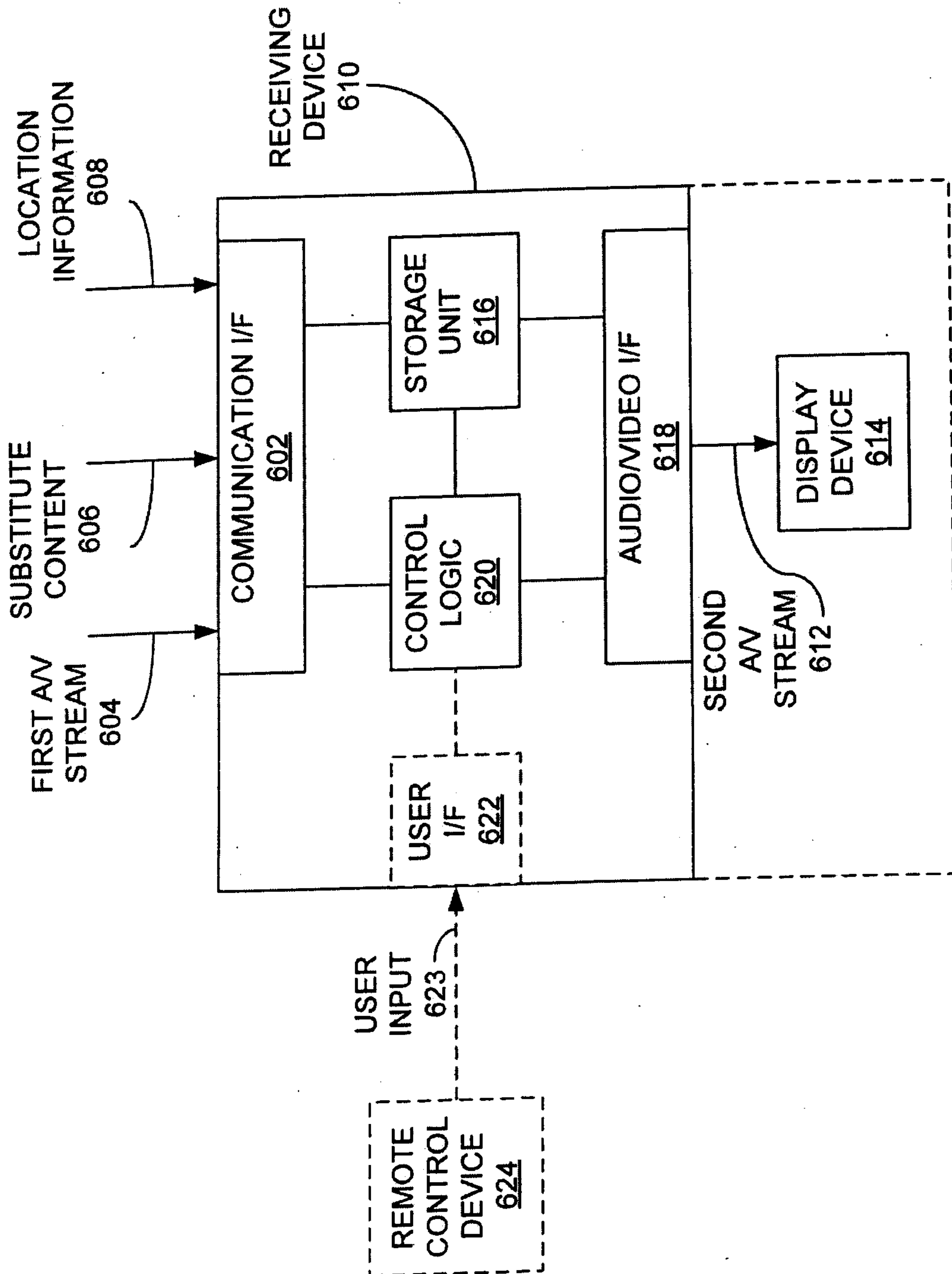
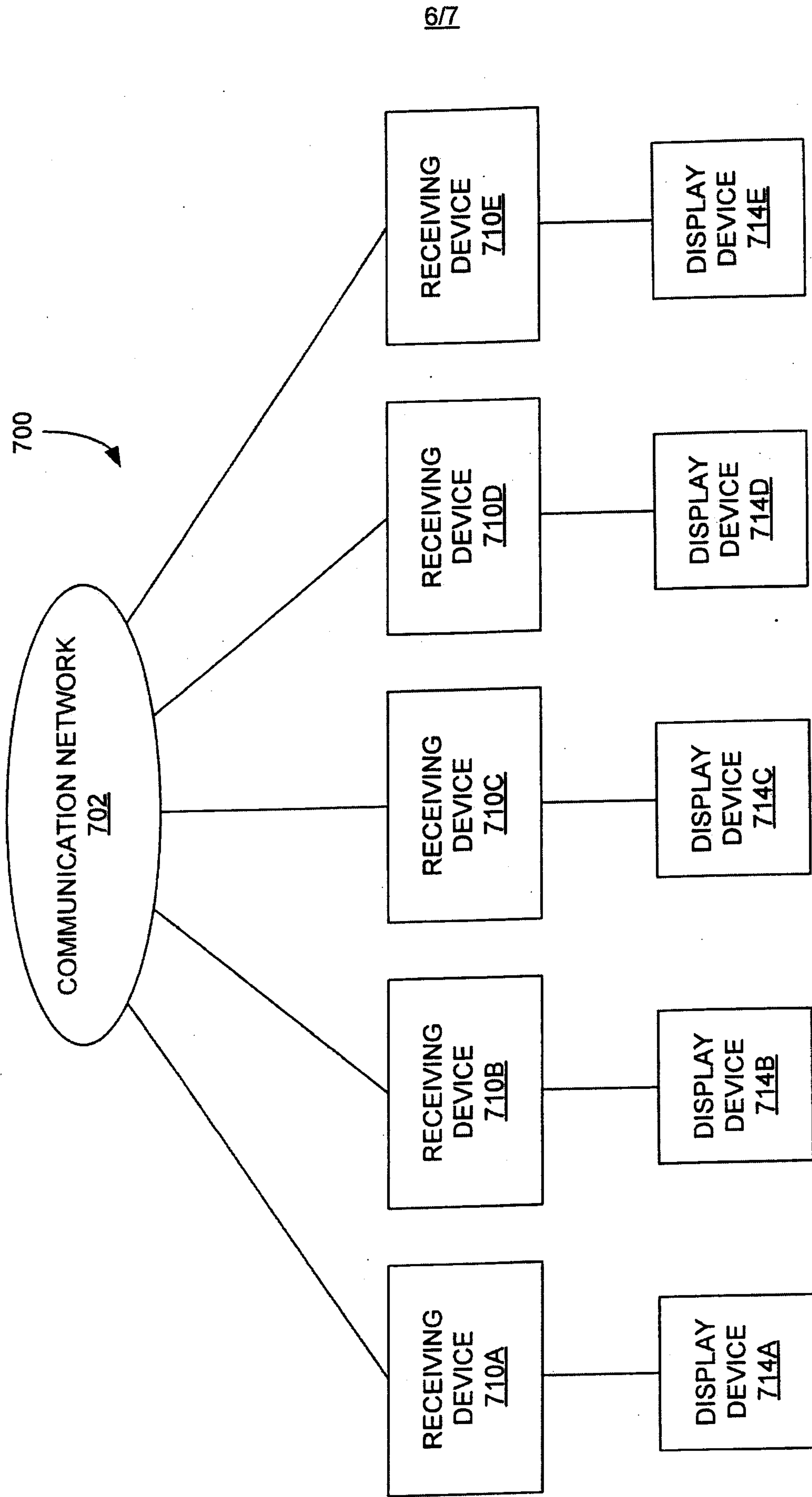


FIG. 6



6/7

FIG. 7



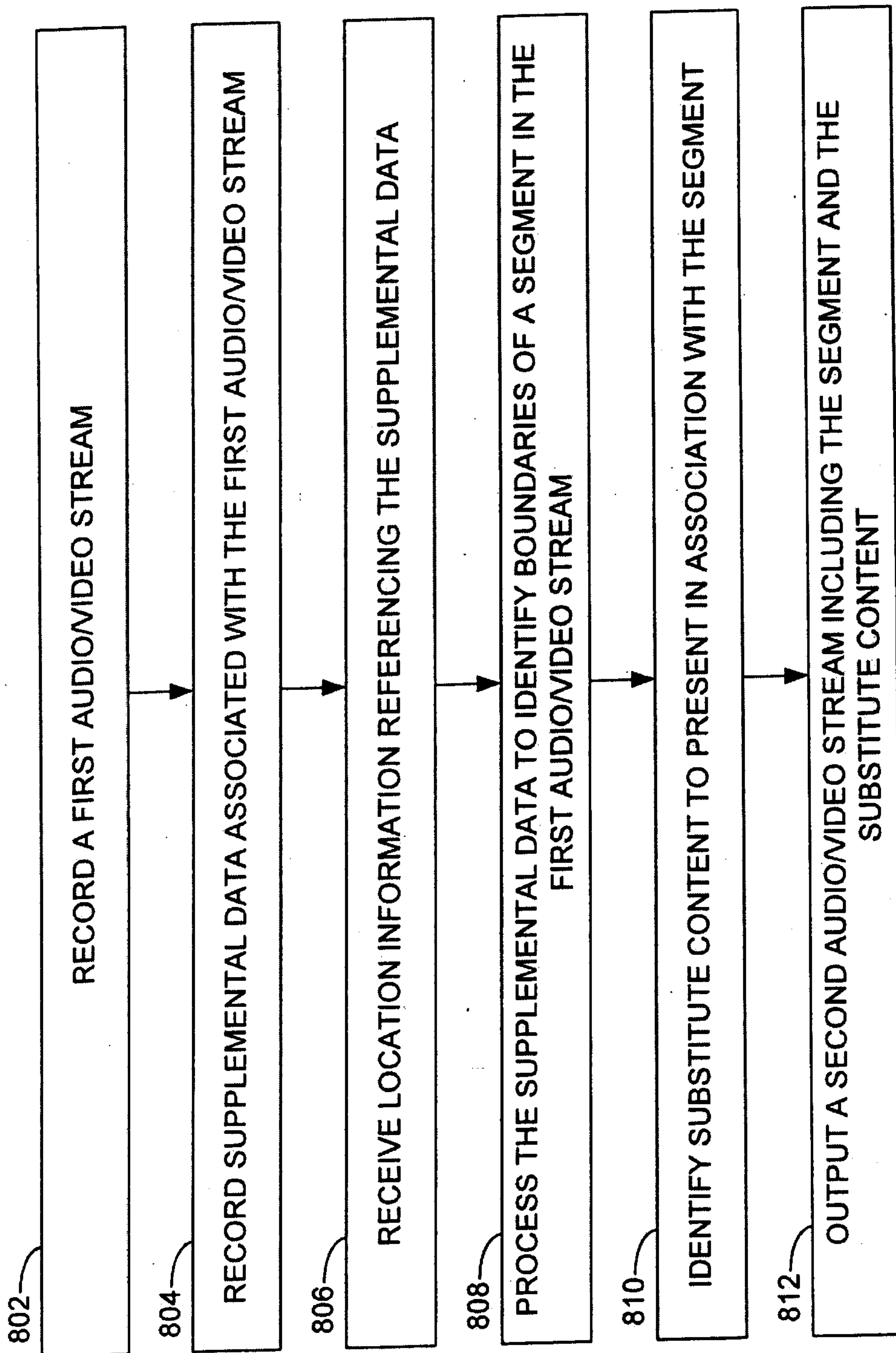


FIG. 8

