



US009270964B1

(12) **United States Patent**
Tseytlin

(10) **Patent No.:** **US 9,270,964 B1**
(45) **Date of Patent:** **Feb. 23, 2016**

(54) **EXTRACTING AUDIO COMPONENTS OF A PORTION OF VIDEO TO FACILITATE EDITING AUDIO OF THE VIDEO**

(71) Applicant: **Google Inc.**, Mountain View, CA (US)

(72) Inventor: **Yan Tseytlin**, Wayne, NJ (US)

(73) Assignee: **Google Inc.**, Mountain View, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 12 days.

(21) Appl. No.: **13/925,434**

(22) Filed: **Jun. 24, 2013**

(51) **Int. Cl.**

- H04N 7/00** (2011.01)
- H04N 11/00** (2006.01)
- H04N 5/93** (2006.01)
- H04N 9/79** (2006.01)
- G11B 27/00** (2006.01)
- H04N 9/475** (2006.01)

(52) **U.S. Cl.**

CPC **H04N 9/79** (2013.01)

(58) **Field of Classification Search**

CPC G11B 27/00; H04N 5/93; H04N 11/00;
H04N 7/00; H04N 9/475

USPC 386/278, 280, 282, 285; 348/461, 462,
348/515

See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

7,336,890	B2	2/2008	Lu et al.	
7,424,204	B2 *	9/2008	Nakamura	386/248
7,529,464	B2	5/2009	Cheng et al.	
8,111,326	B1 *	2/2012	Talwar	348/500
8,326,623	B2	12/2012	Fujii	
2004/0143434	A1	7/2004	Divakaran et al.	
2006/0170778	A1 *	8/2006	Ely et al.	348/207.99
2010/0168881	A1	7/2010	Weber et al.	
2014/0219637	A1 *	8/2014	McIntosh et al.	386/282

* cited by examiner

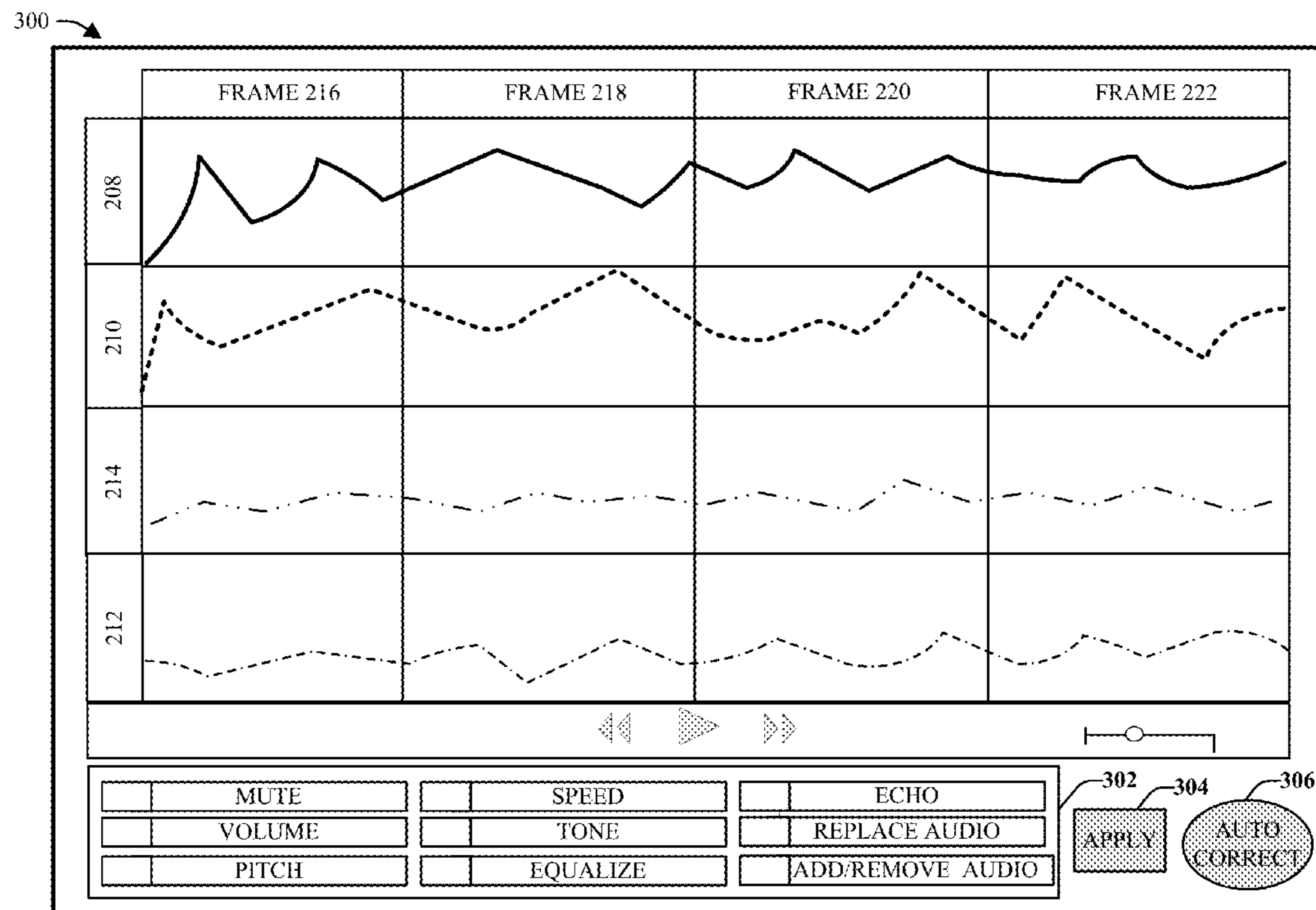
Primary Examiner — Daquan Zhao

(74) *Attorney, Agent, or Firm* — Fenwick & West LLP

(57) **ABSTRACT**

Systems and methods for extracting audio components of a portion of a video to facilitate editing the audio portion are presented. In one or more aspects, a system is provided that includes a receiving component configured to receive a video as an upload from a client device over a network and an identification component configured to identify two or more different audio components of an audio track of the video. The system further comprises an extraction component configured to extract and separate the two or more different audio components, and an editing component configured to generate an editing interface that receives input via the editing interface regarding editing the two or more different audio components separately.

24 Claims, 15 Drawing Sheets



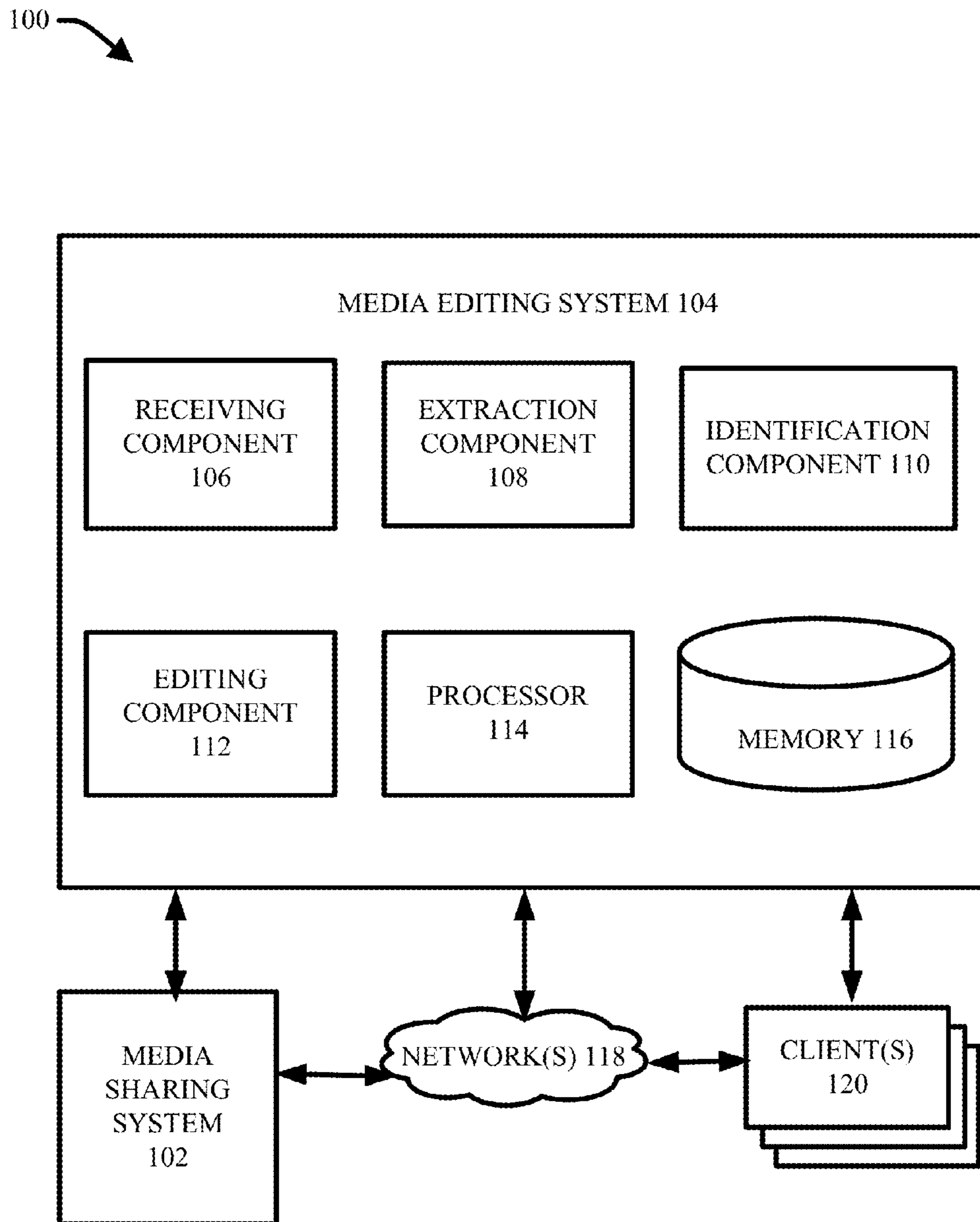


FIG. 1

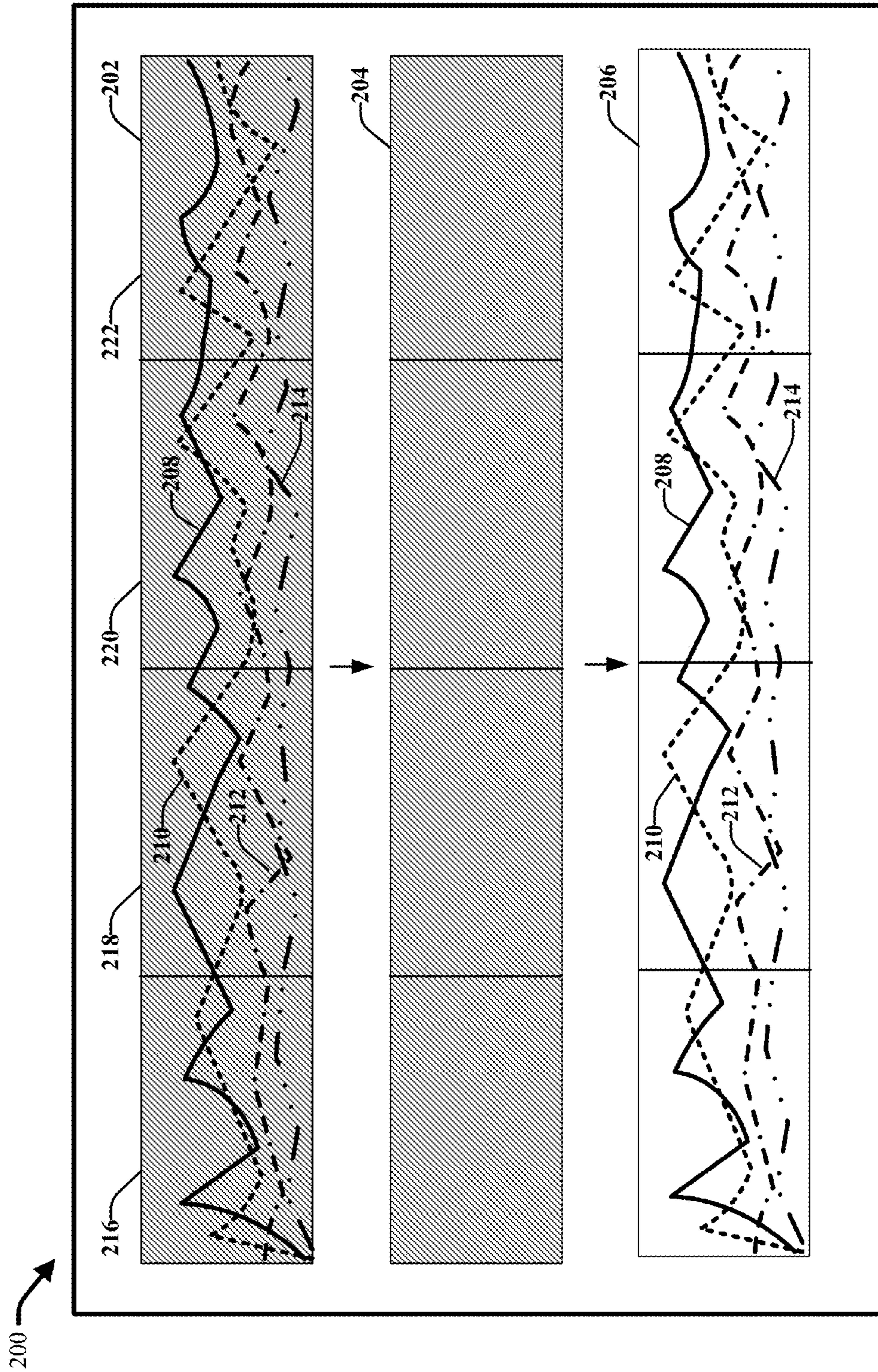


FIG. 2

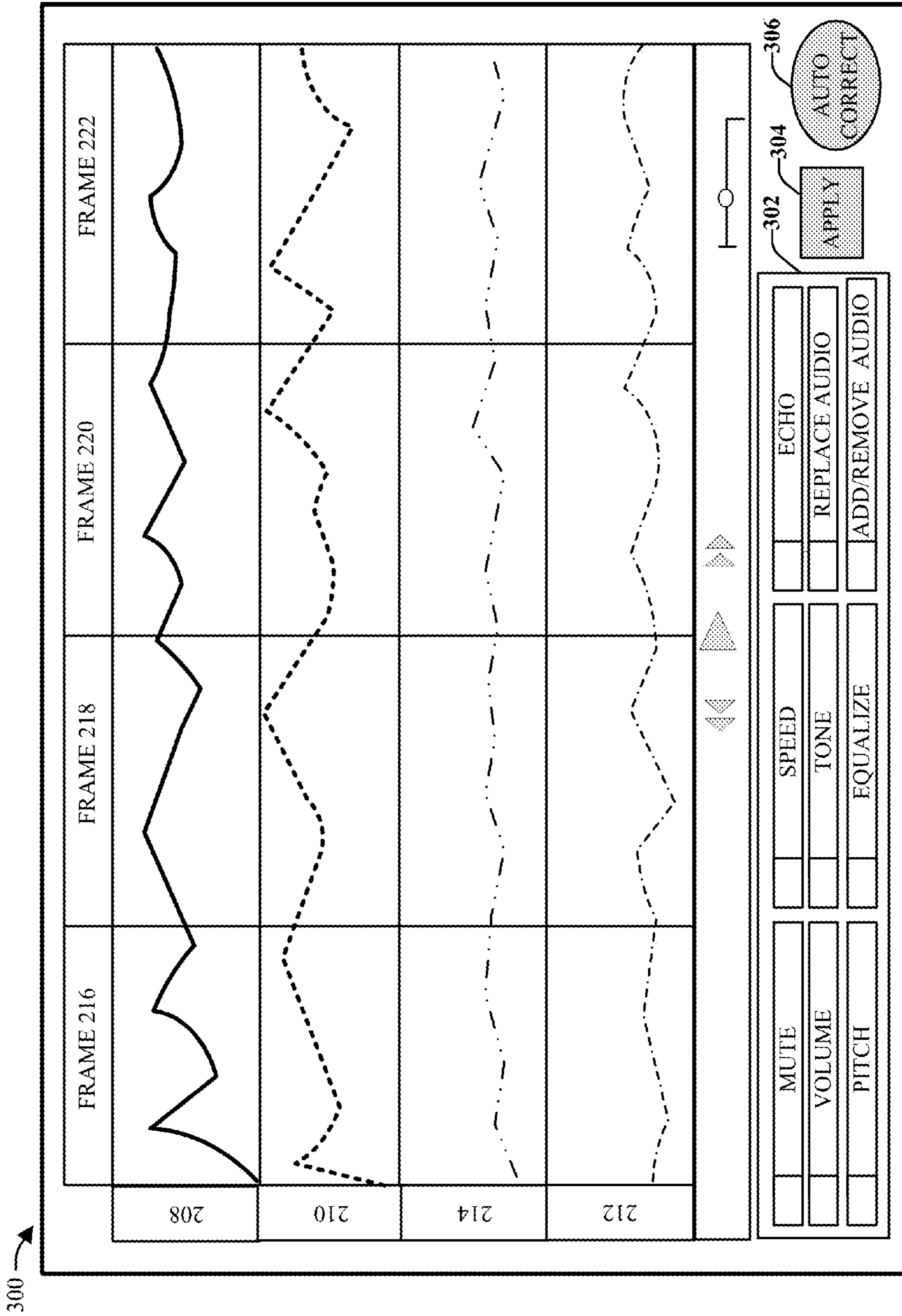


FIG. 3

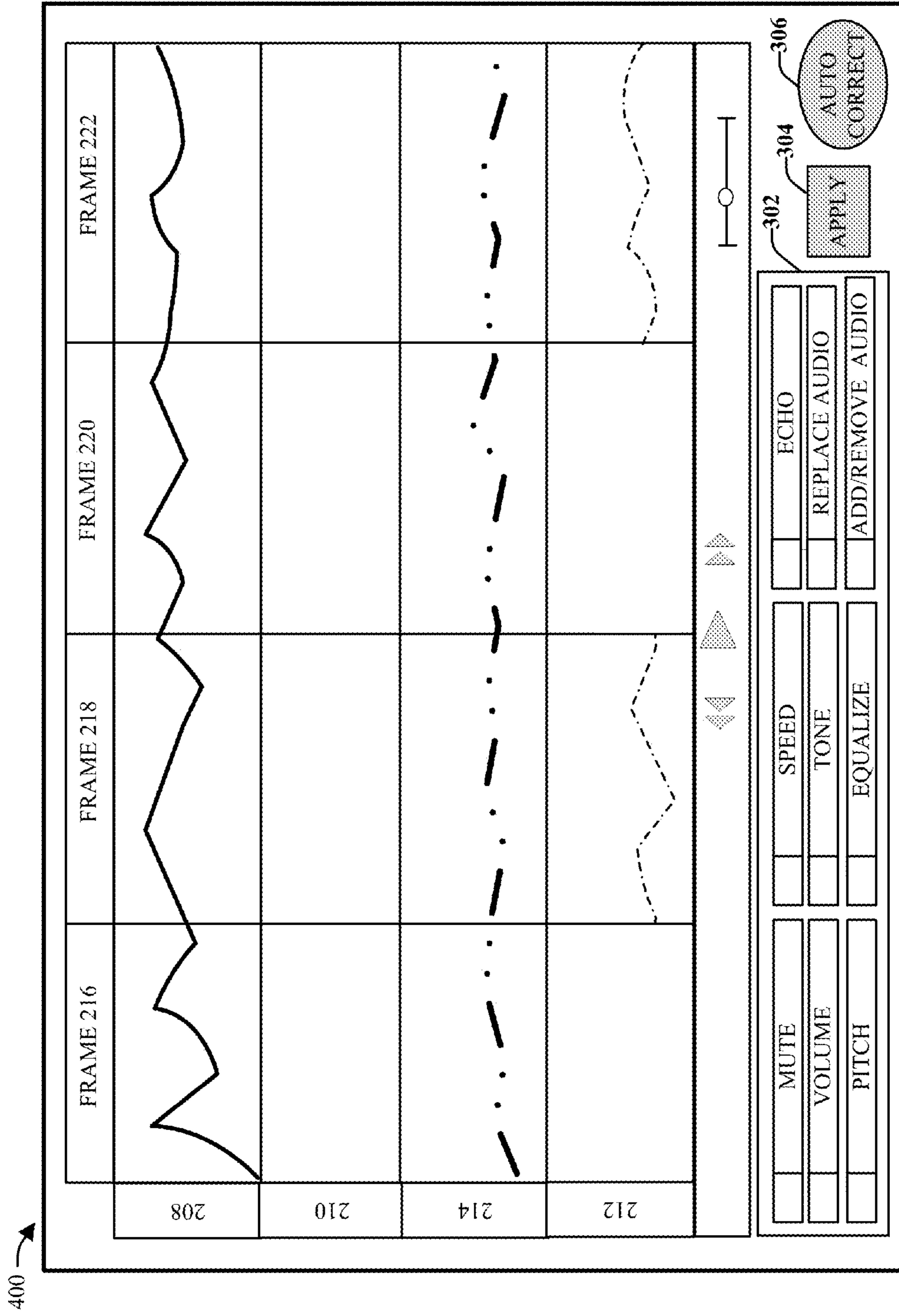


FIG. 4

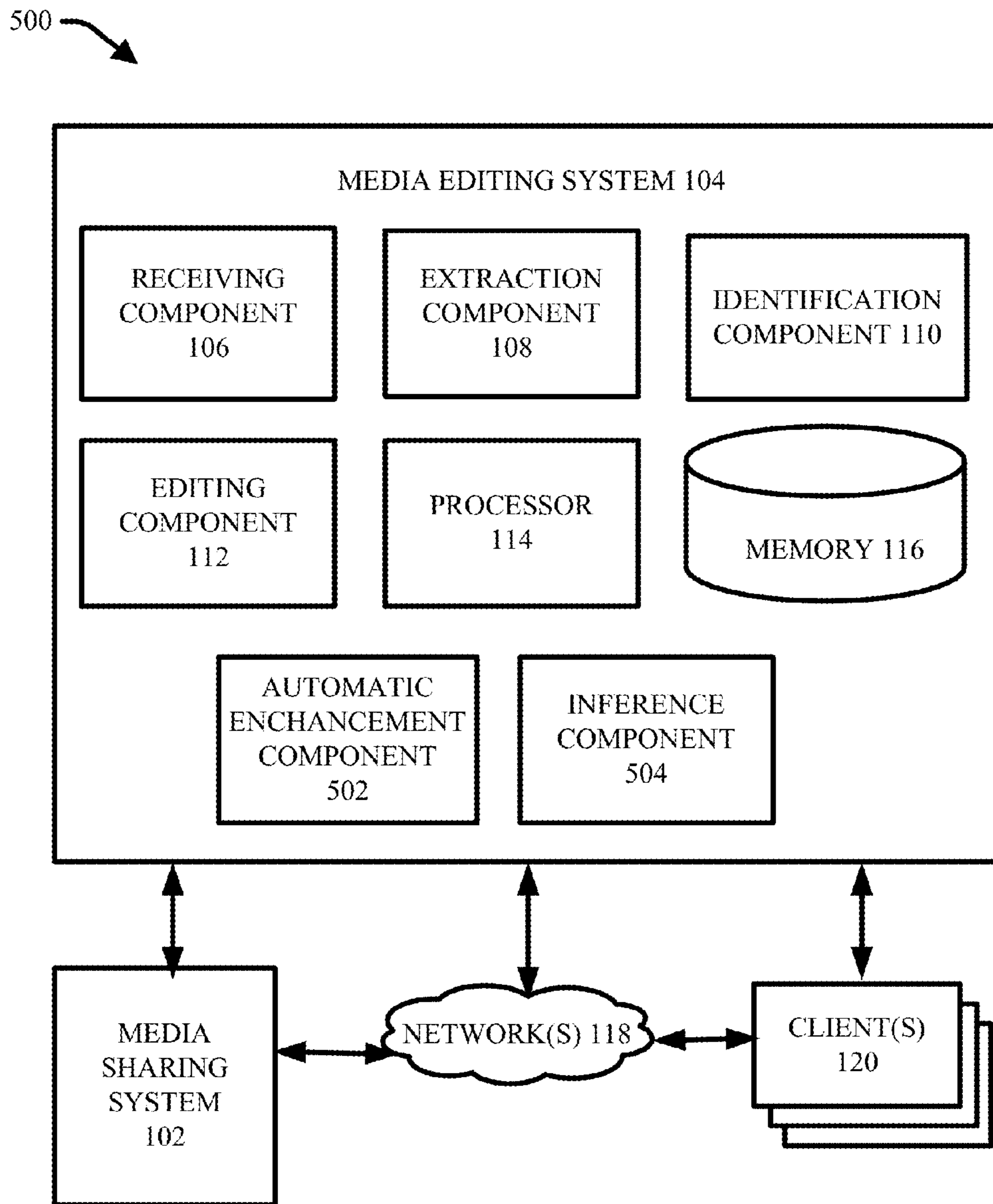


FIG. 5

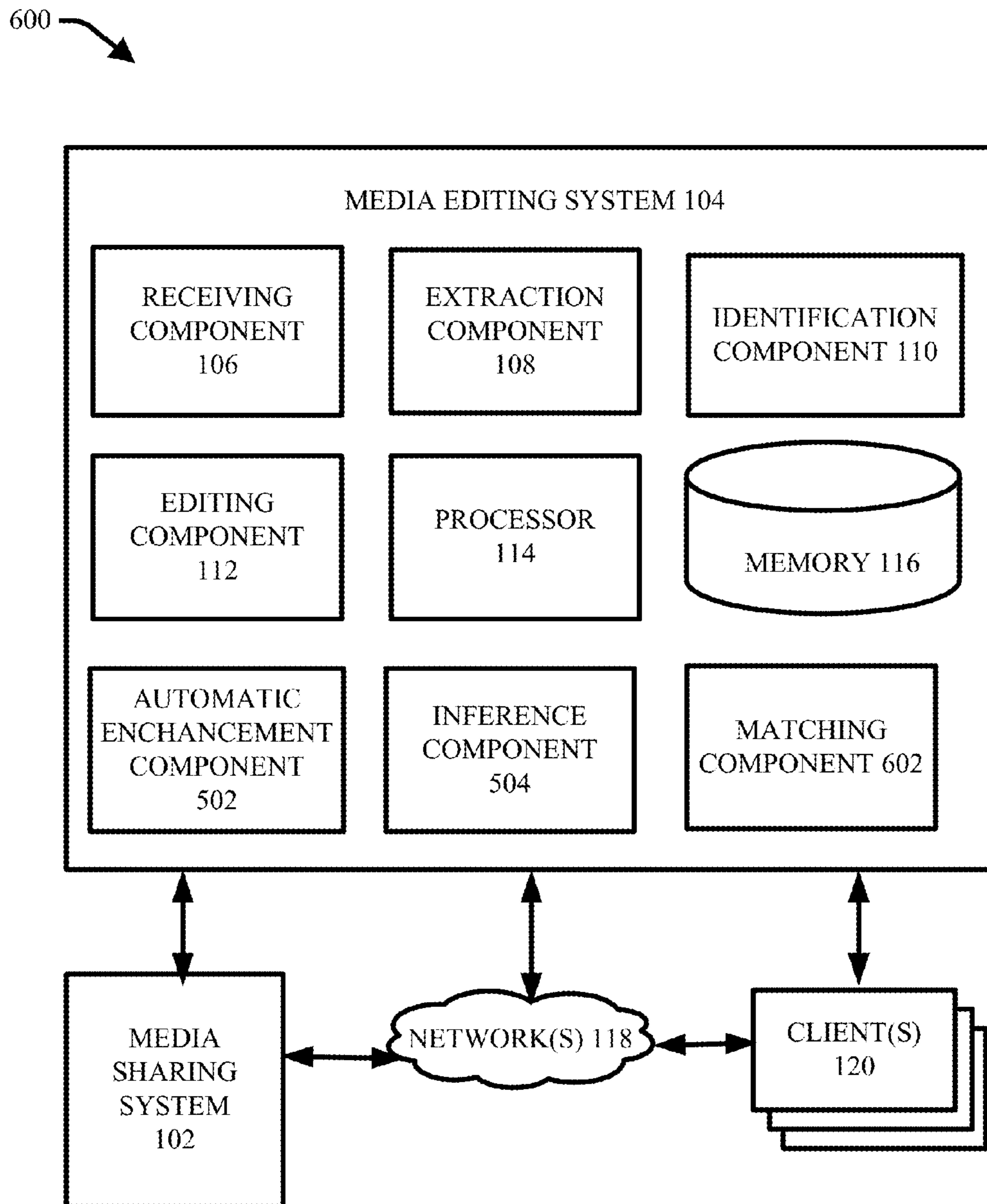


FIG. 6

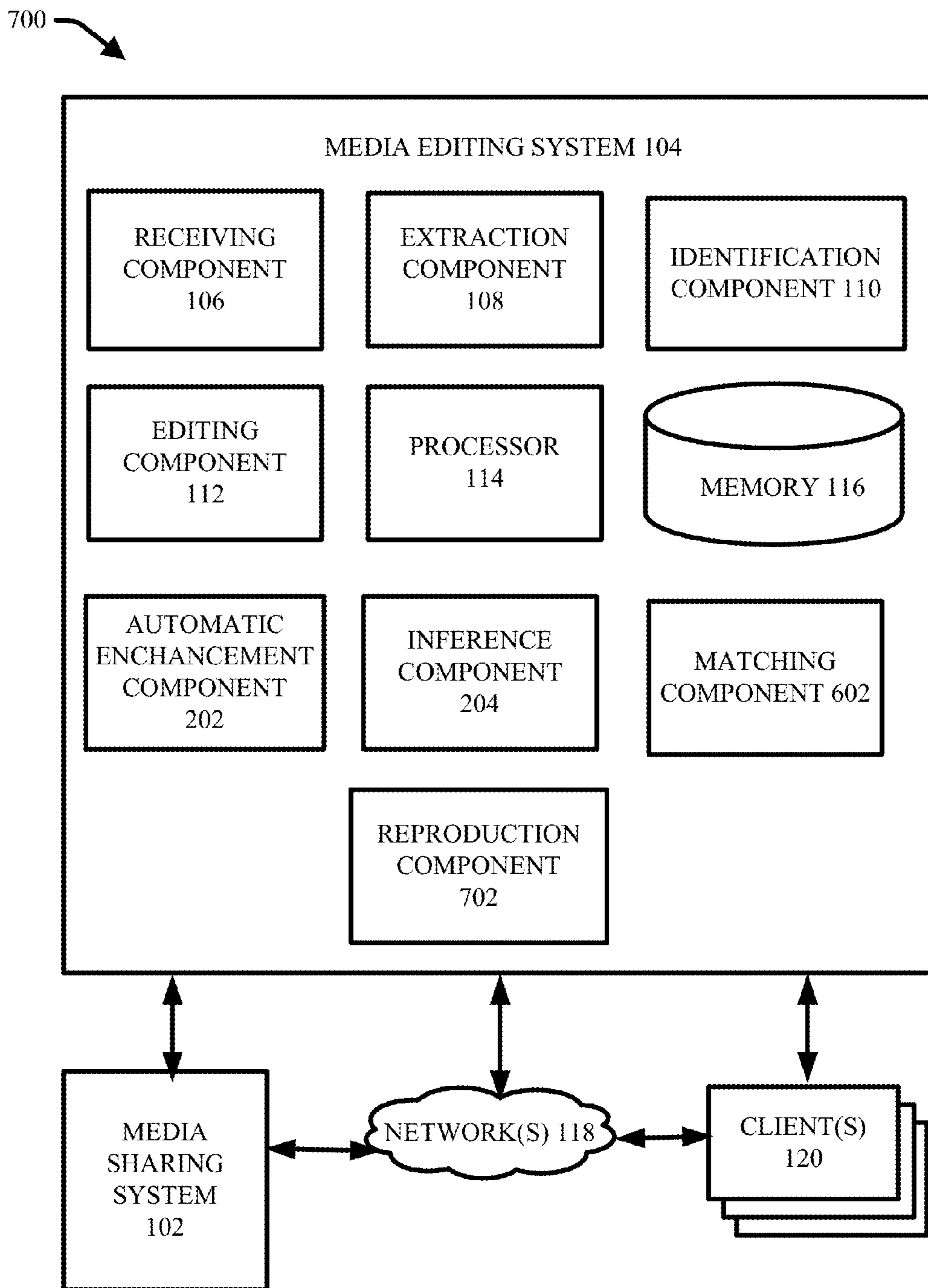


FIG. 7

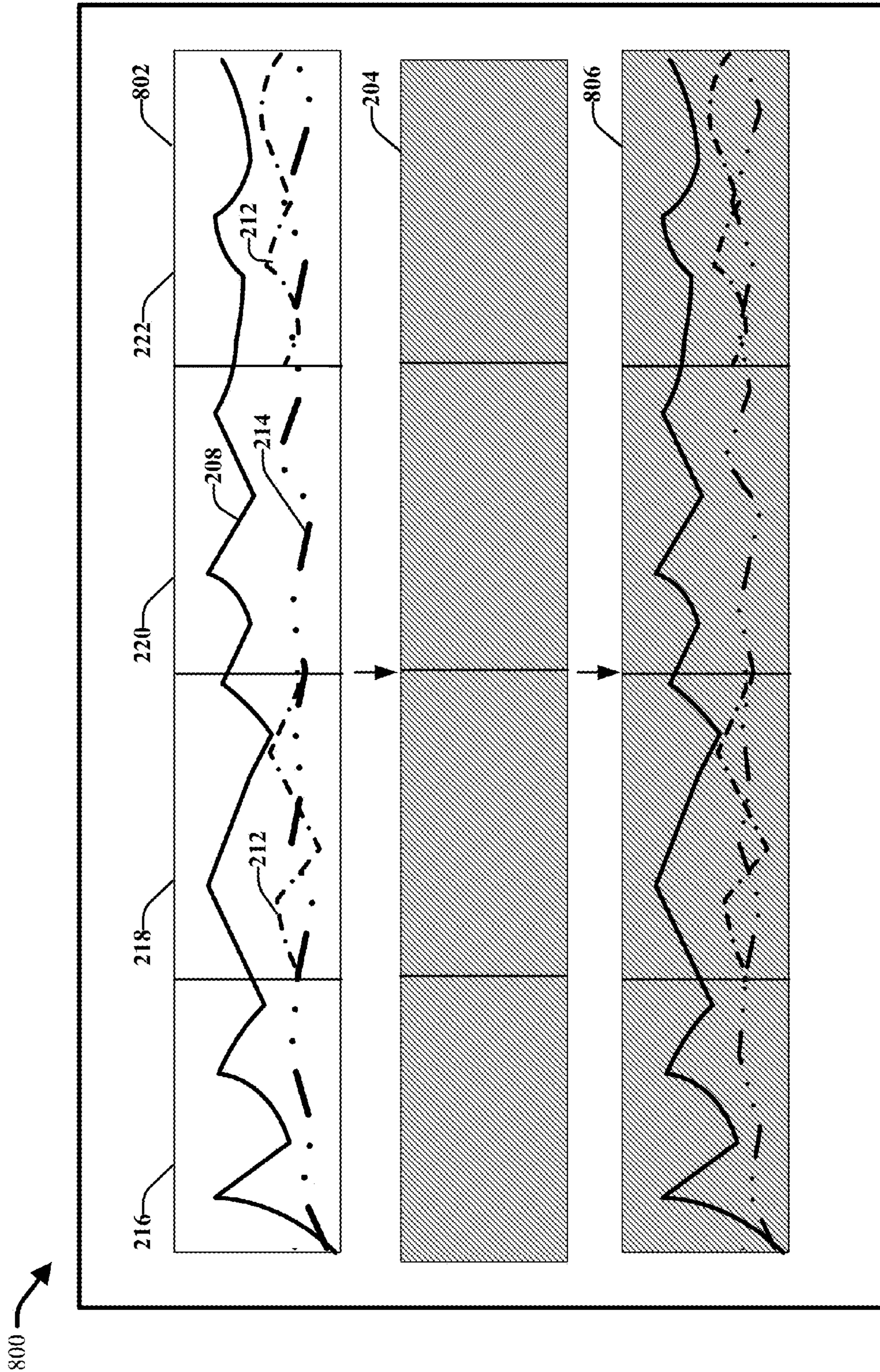


FIG. 8

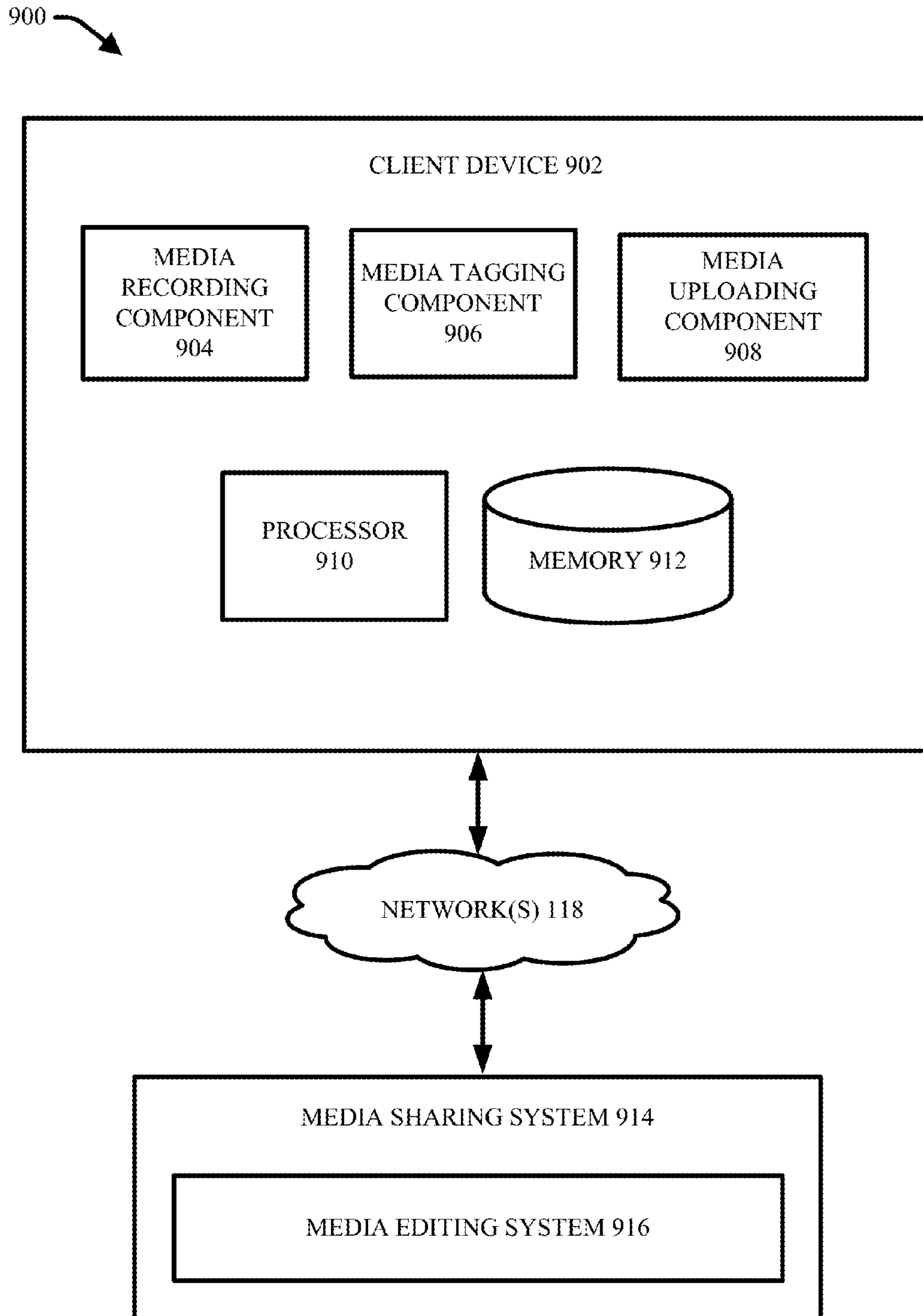


FIG. 9

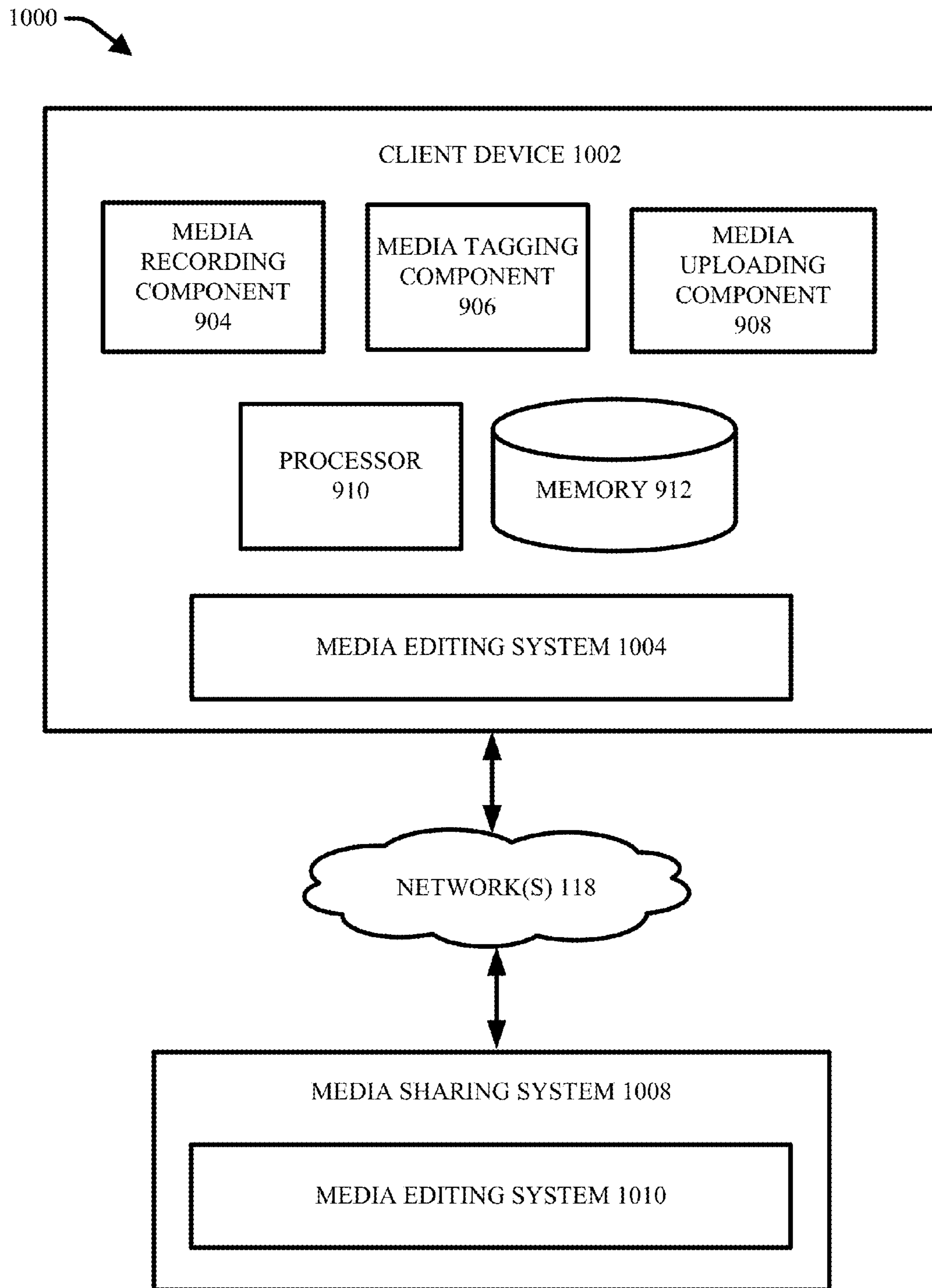


FIG. 10

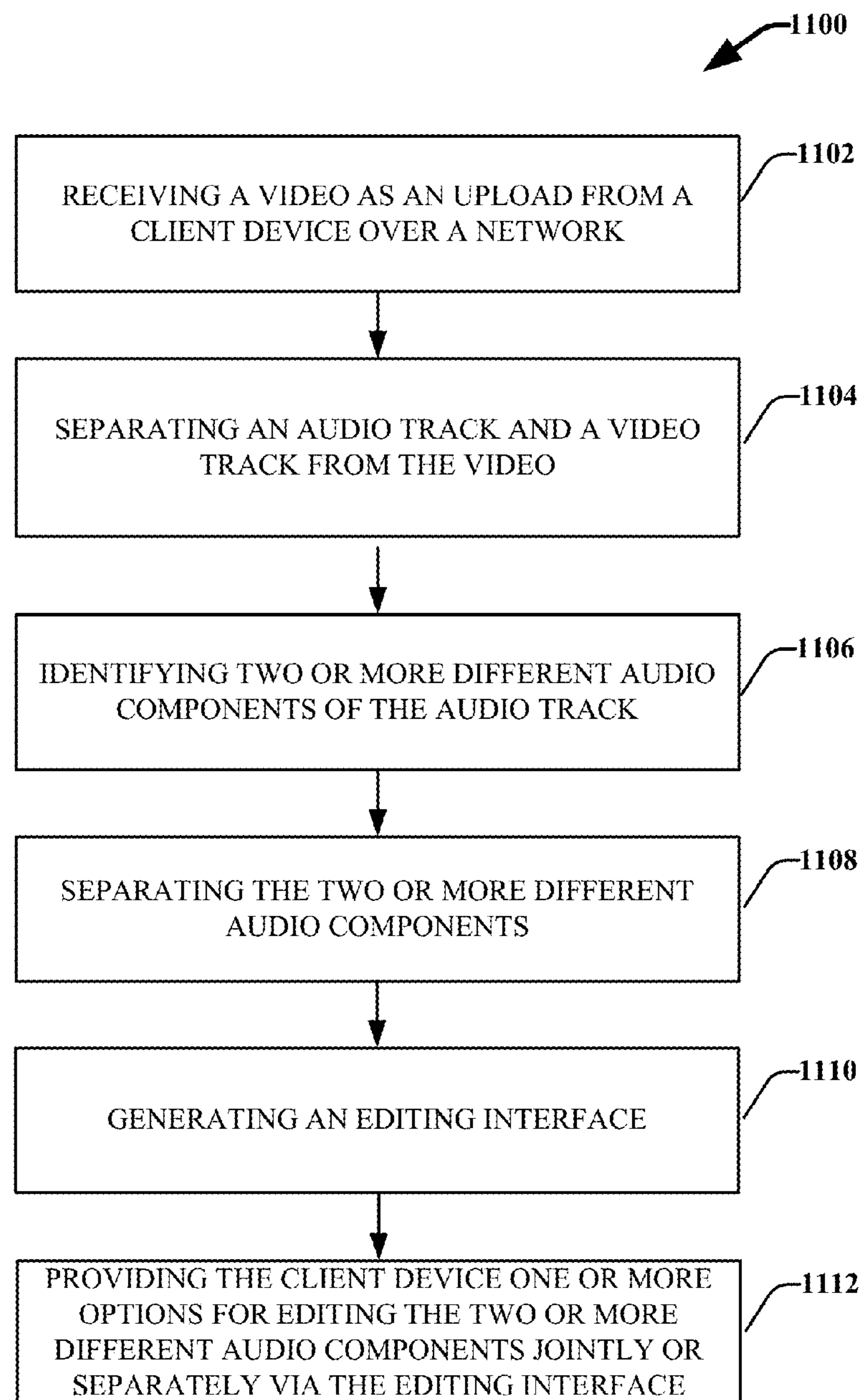


FIG. 11

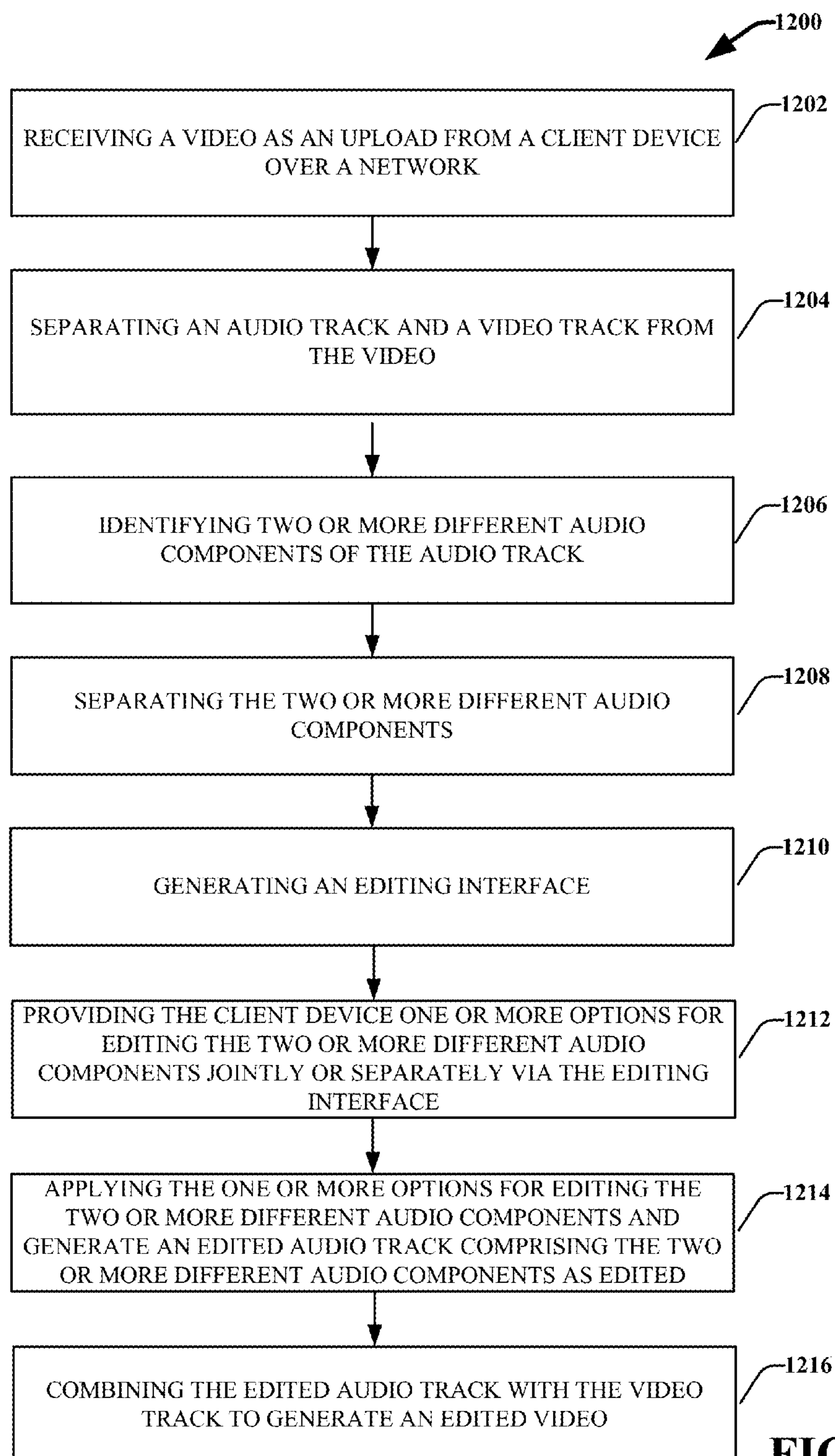


FIG. 12

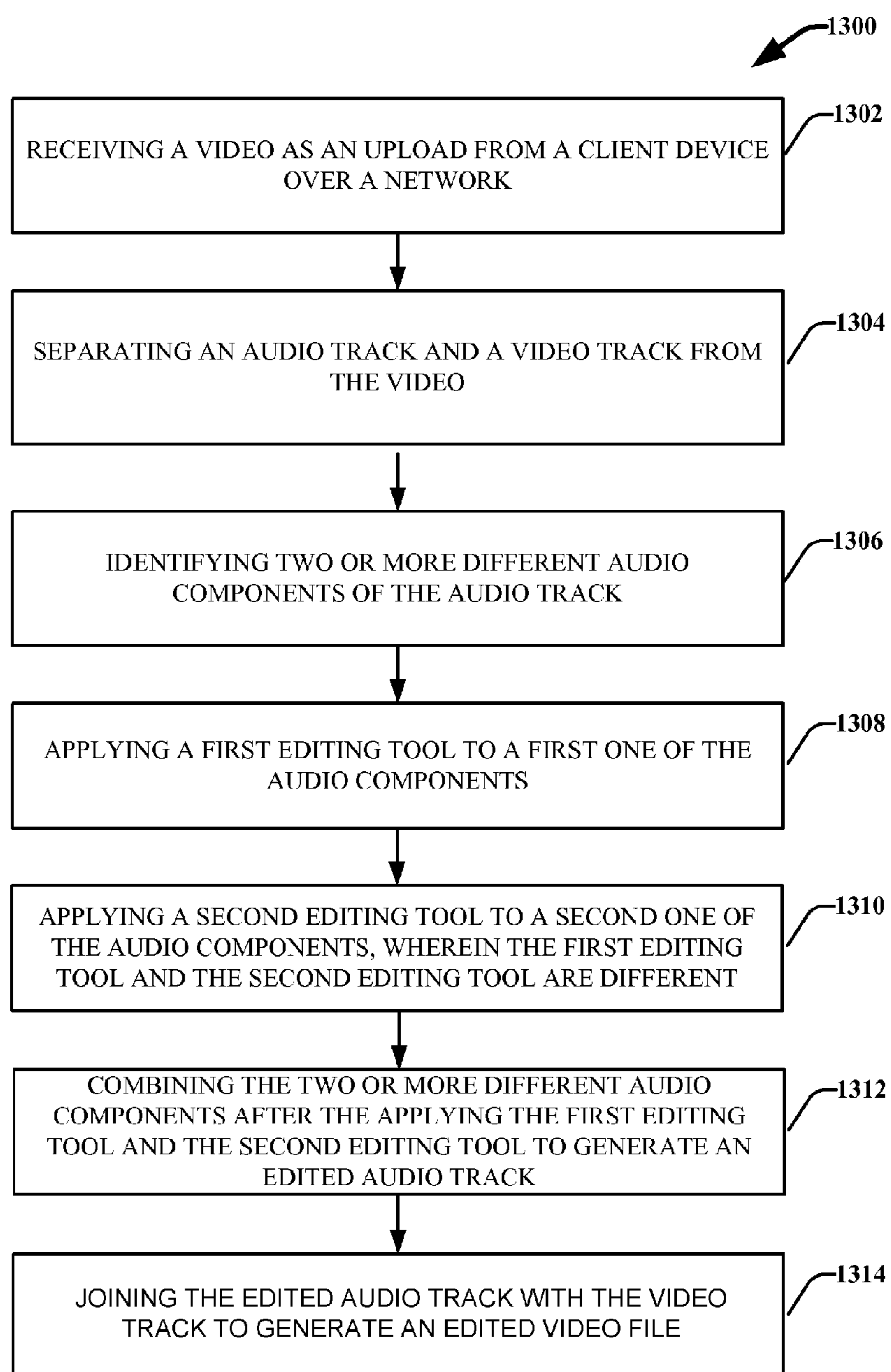


FIG. 13

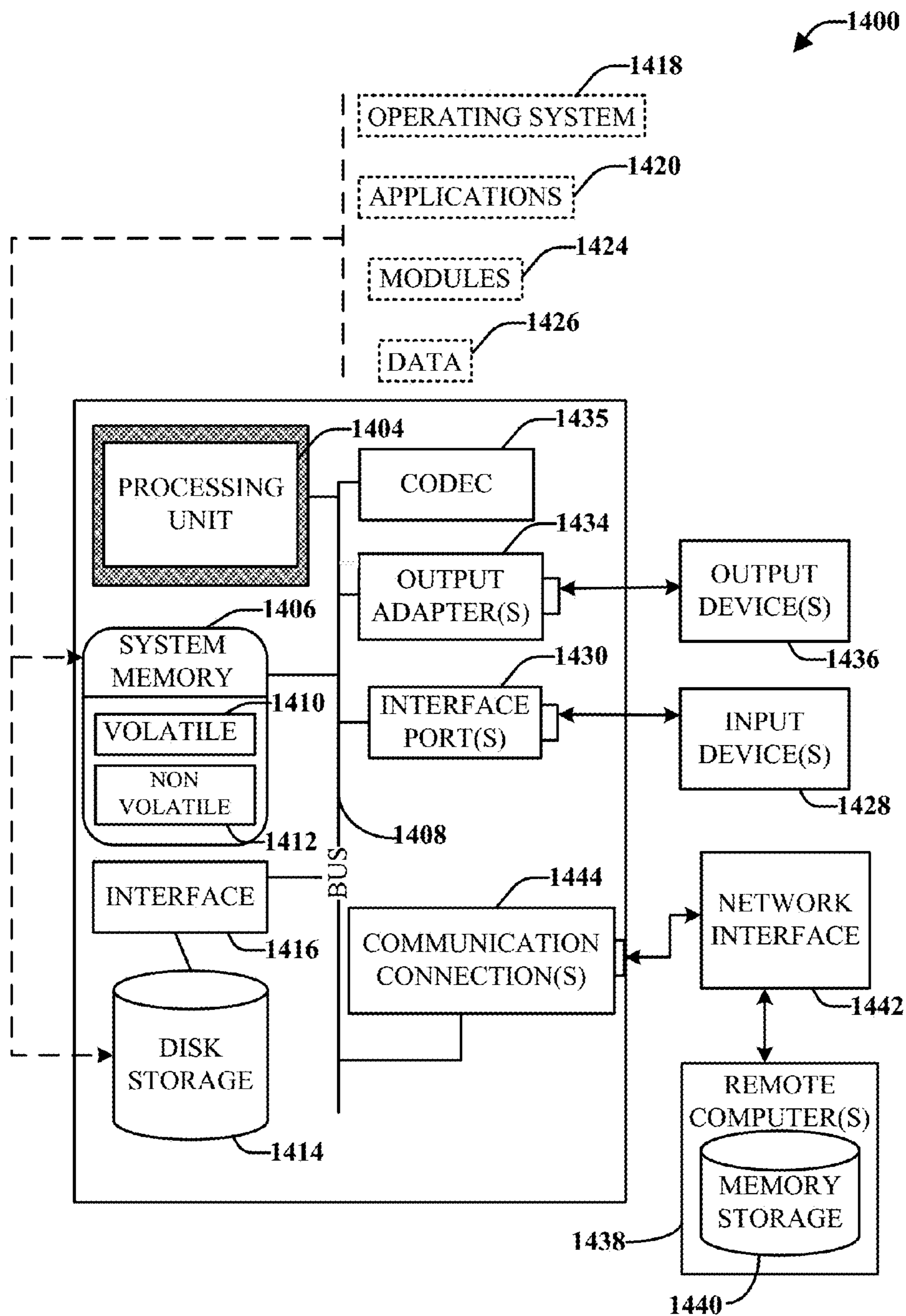


FIG. 14

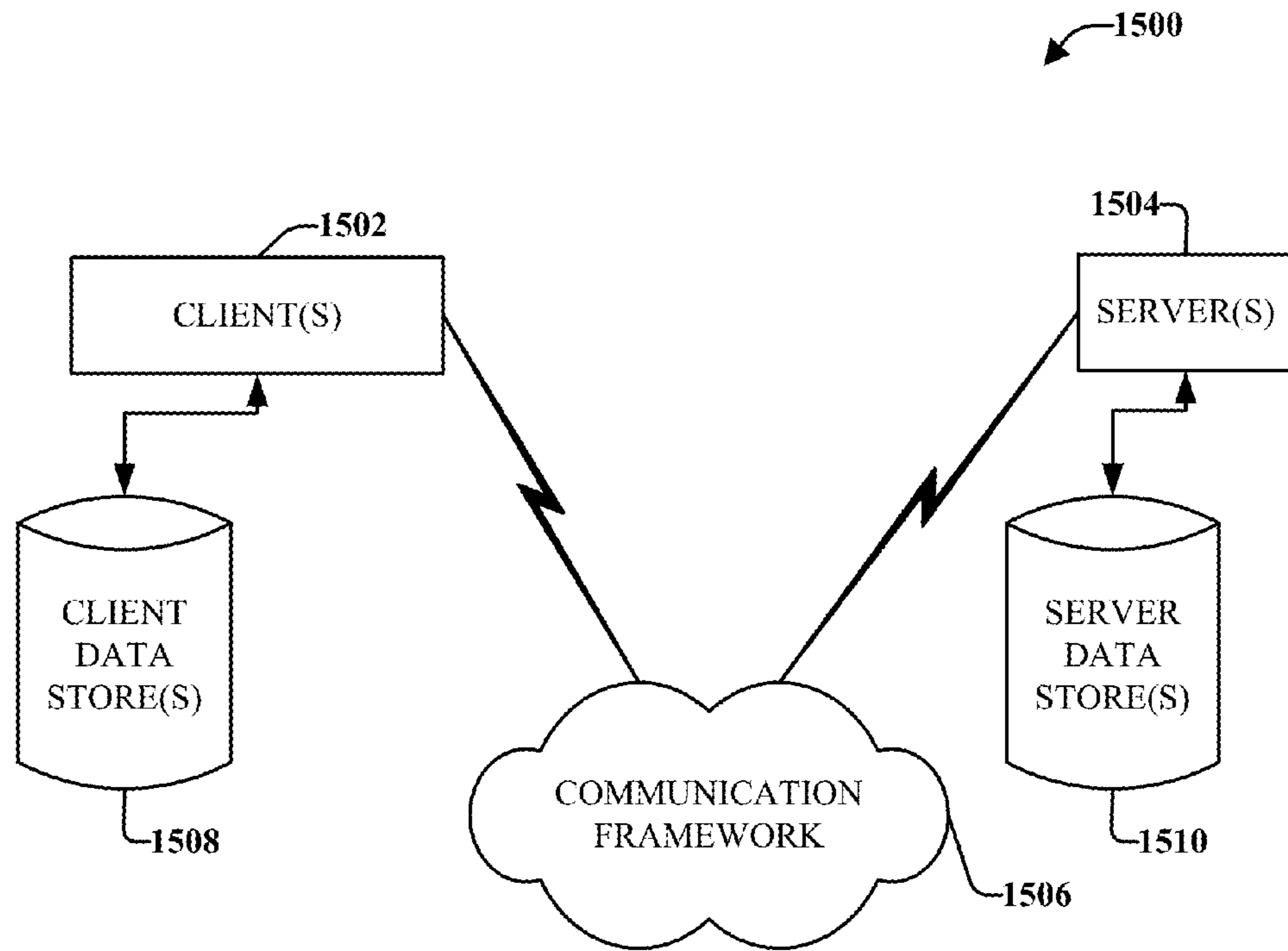


FIG. 15

1**EXTRACTING AUDIO COMPONENTS OF A
PORTION OF VIDEO TO FACILITATE
EDITING AUDIO OF THE VIDEO**

TECHNICAL FIELD

This application generally relates to systems and methods for extracting audio components of a portion of video to facilitate editing audio of the video.

BACKGROUND

Many mobile devices allow users to capture video and share captured videos with others through media publishing websites. Absent sophisticated video editing tools or videographer editing expertise, video content uploaded to media publishing websites by ordinary users is often incomplete and of amateur quality.

BRIEF DESCRIPTION OF THE DRAWINGS

Numerous aspects, embodiments, objects and advantages of the present invention will be apparent upon consideration of the following detailed description, taken in conjunction with the accompanying drawings, in which like reference characters refer to like parts throughout, and in which:

FIG. 1 illustrates an example system for editing an audio track of a video file in accordance with various aspects and embodiments described herein;

FIG. 2 presents an example process for extracting an audio track and video track from a video file in accordance with various aspects and embodiments described herein;

FIG. 3 presents an example user interface for editing audio components of an audio track jointly or separately in accordance with various aspects and embodiments described herein;

FIG. 4 presents another example user interface for editing audio components of an audio track jointly or separately in accordance with various aspects and embodiments described herein;

FIG. 5 illustrates another example system for editing an audio track of a video file in accordance with various aspects and embodiments described herein;

FIG. 6 illustrates another example system for editing an audio track of a video file in accordance with various aspects and embodiments described herein;

FIG. 7 illustrates another example system for editing an audio track of a video file in accordance with various aspects and embodiments described herein;

FIG. 8 presents an example process for combining an edited audio track with an associated video track to generate an edited video file in accordance with various aspects and embodiments described herein;

FIG. 9 illustrates another example system for editing an audio track of a video file in accordance with various aspects and embodiments described herein;

FIG. 10 illustrates another example system for editing an audio track of a video file in accordance with various aspects and embodiments described herein;

FIG. 11 is a flow diagram of an example method for editing an audio track of a video file in accordance with various aspects and embodiments described herein;

FIG. 12 is a flow diagram of another example method for editing an audio track of a video file in accordance with various aspects and embodiments described herein;

2

FIG. 13 is a flow diagram of another example method for editing an audio track of a video file in accordance with various aspects and embodiments described herein;

FIG. 14 is a schematic block diagram illustrating a suitable operating environment in accordance with various aspects and embodiments.

FIG. 15 is a schematic block diagram of a sample-computing environment in accordance with various aspects and embodiments.

DETAILED DESCRIPTION

The innovation is described with reference to the drawings, wherein like reference numerals are used to refer to like elements throughout. In the following description, for purposes of explanation, numerous specific details are set forth in order to provide a thorough understanding of this innovation. It may be evident, however, that the innovation can be practiced without these specific details. In other instances, well-known structures and components are shown in block diagram form in order to facilitate describing the innovation.

By way of introduction, the subject matter described in this disclosure relates to systems and methods for editing video using an application running at a server accessible to a client device (e.g., via the cloud). For example, a user can record a video on a client device and upload the video to a networked media sharing system accessible to the client device via a network (e.g., the Internet). Once the video is uploaded, the user can employ video editing tools provided at the networked media sharing system to edit the video. In particular, the disclosed systems and methods offer video editing tools that facilitate editing an audio track associated with a video (e.g., the audio portion of the video as opposed to the visual image portion of the video).

For example, an editing system running or employed by the networked media sharing system separates video, comprising video images and audio, into respective video and audio tracks. The editing system parses the audio track(s) and identifies two or more different audio components or layers present in the audio track(s).

For example, different audio components or layers of an audio track can include a voice of a first person, a voice of a second person, a sound of a train, a sound of a siren, music, muffled sounds of a crowd, etc. The editing system provides editing tools to apply to the different audio components jointly or separately. For example, the editing system can allow a user to mute one of the audio components and increase volume of another audio component. After editing different audio components, the editing system can re-join different audio components and generate an edited audio track that reflects edits applied to the different audio components. The system can re-join the edited audio track with the original video track to produce an edited video.

In one or more aspects, a system is provided that includes a receiving component configured to receive video as an upload from a client device over a network and an extraction component configured to separate an audio track and a video track from the video. The system further includes an identification component configured to identify two or more different audio components of the audio track, wherein the extraction component is further configured to separate the two or more different audio components; and an editing component configured to generate an editing interface and provide the client device one or more options for editing the two or more different audio components jointly or separately via the editing interface.

In another aspect, a method is disclosed that includes receiving a video as an upload from a client device over a network, separating an audio track and a video track from the video, identifying two or more different audio components of the audio track, separating the two or more different audio components, generating an editing interface, and providing the client device one or more options for editing the two or more different audio components jointly or separately via the editing interface.

Further provided is a tangible computer-readable storage medium comprising computer-readable instructions that, in response to execution, cause a computing system to perform various operations. These operations can include receiving a video as an upload from a client device over a network, separating an audio track and a video track from the video, identifying two or more different audio components of the audio track, and separating the two or more different audio components. The operations further include generating an editing interface, providing the client device one or more editing options for editing the two or more different audio components via the editing interface, receiving a request to apply an editing option of the one or more editing options to only a first one of the two or more different audio components, and applying the editing option to only the first one of the two or more different audio components in response to the request.

Referring now to the drawings, with reference initially to FIG. 1, presented is a diagram of an example system 100 that facilitates editing an audio portion of a video in accordance with various aspects and embodiments described herein. Aspects of systems, apparatuses or processes explained in this disclosure can constitute machine-executable components embodied within machine(s), e.g., embodied in one or more computer readable mediums (or media) associated with one or more machines. Such components, when executed by the one or more machines, e.g., computer(s), computing device(s), virtual machine(s), etc. can cause the machine(s) to perform the operations described.

System 100 includes media editing system 104 and one or more client devices 120 configured to connect to media editing system 104 via one or more networks 118. Media editing system 104 is configured to receive a media item from a client device 120 via a network 118 and facilitate editing the media item. For example, a client device 120 can upload a video to media editing system 104 over a network and employ the media editing system 104 to edit the video in manners discussed herein. The media editing system 104 can provide the edited video back to the client device 120 for usage thereof, send the edited video to another device via a network 118, post the edited video at a networked resource, store the edited video, etc.

In an aspect, media editing system 104 can be used in association with a media sharing system 102. According to this aspect, media sharing system 102 can include media editing system 104 or access media editing system via a network 118. Media sharing system 102 can include an entity configured to receive media content from one or more client devices 120 via a network 118 and provide the media content to one or more clients via network 118. In an aspect, media sharing system 102 employs media editing system 104 to provide tools for editing media content uploaded to media sharing system 102.

As used herein the term media content or media item can include but is not limited to streamable media (e.g., video, live video, video advertisements, music videos, audio, music, sound files and etc.) and static media (e.g., pictures, thumbnails). In an aspect, media sharing system 102 can employ one

or more server computing devices to store and deliver media content to users of client devices 120 that can be accessed using a browser. For example, media sharing system 102 can provide and present media content to a user via a website.

In an aspect, media sharing system 102 is configured to provide streamed media to users over network 118. The media can be stored in memory (not shown) associated with media sharing system 102 and/or at various servers employed by media sharing system 102 and accessed by a client device 120 using a website platform of the media sharing system 102. For example, media sharing system 102 can include a media presentation source that provides client device 120 access to a voluminous quantity (and potentially an inexhaustible number) of shared media (e.g., video and/or audio) files. The media presentation source can further stream these media files to one or more users at respective client devices 120 of the one or more users over one or more networks 118. In another aspect, media sharing system 102 is configured to receive media files from one or more client devices 120 via one or more networks 118. For example, client device 120 can upload a video to media sharing system 102 via the Internet for sharing with other users using the media sharing system 102. Videos received by media sharing system 102 can further be stored in memory (not shown) employed by the media sharing system 102.

Client device 120 can include any suitable computing device associated with a user and configured to interact with media sharing system 102 and media editing system 104. For example, client device 120 can include a desktop computer, a laptop computer, a television, a mobile phone, a smart-phone, a tablet personal computer (PC), or a personal digital assistant PDA. As used in this disclosure, the terms “content consumer” or “user” refer to a person, entity, system, or combination thereof that employs system 100 (or additional systems described in this disclosure) using a client device 120. Network(s) 118 can include wired and wireless networks, including but not limited to, a wide area network (WAN, e.g., the Internet), a cellular network, a local area network (LAN), or a personal area network (PAN). For example, client device 120 can provide and/or receive media to/from media sharing system 102 or media editing system 104 (and vice versa) using virtually any desired wired or wireless technology, including, for example, cellular, WAN, wireless fidelity (Wi-Fi), Wi-Max, WLAN, and etc. In an aspect, one or more components of system 100 are configured to interact via disparate networks.

Media editing system 104 is configured to offer editing tools for editing a media item that at least includes an audio track. In an aspect, media editing system 104 is configured to edit a video that has both an audio portion or audio track and a video image portion or video track—the terms portion and track are used herein interchangeably. According to this aspect, the media editing system 104 is configured to separate the audio track and video track of a video file and provide various tools for editing the audio track. The media editing system 104 can later join the edited audio track with the video track previously separated there from to create a version of the video with the edited audio track. In another aspect, media editing system 104 can provide editing tools for an audio track that is not associated with a video or video track.

Media editing system 104 can include receiving component 106 for receiving video and audio files, extraction component 108 for extracting an audio track from a video file and for extracting audio components from an audio track, identification component 110 for identifying different audio components of an audio track and editing component 112 for editing the different components of the audio track. Media

5

editing system **104** includes memory **116** for storing computer executable components and instructions. Media editing system **104** can further include a processor **114** to facilitate operation of the instructions (e.g., computer executable components and instructions) by media editing system **104**.

Receiving component **106** is configured to receive a video or audio track transmitted to media editing system **104** or media sharing system **102** from a client device via a network **118**. For example, a client device **120** can upload or otherwise send a video or audio track to media editing system **104** via a network **118** for editing thereof. According to this example, the video or audio track is intercepted by receiving component **106**. In another example, where media sharing system **102** includes media editing system **104**, a client device **120** can upload or otherwise send a video or audio track to media sharing system **102**. In an aspect, a user of client device **120** can further choose to edit the uploaded video or audio track. The user can then access the video at the media sharing system **102** and select the video for editing (e.g., via a user interface generated by the media sharing system **102** or media editing system **104**). Selection of the video for editing can result in sending of the video by media sharing system **102** to the media editing system **104** for editing thereof. In another aspect, the media sharing system **102** can automatically send a video or audio track uploaded thereto to media editing system **104** for editing. According to this aspect, videos or audio tracks sent to media editing system by media sharing system **102** are received by receiving component **106**.

In an aspect, extraction component **108** is configured to separate an audio track and video track from a video file received by media editing system **104**. Referring ahead to FIG. 2, presented is a diagram demonstrating an example extraction process **200** of extraction component **108**. FIG. 2 includes a first bar **202** representing a video file, a second bar **204** representing the extracted video portion of the video file and a third bar **206** representing the extracted audio portion of the video file. The video file and its extracted parts are depicted separated into four segments **216**, **218**, **220**, **222**. In an aspect, the segments are associated with frames of the video. It should be appreciated that although the video file represented by bar **202** is depicted as having four segments or frames, a video file received by media editing system **104** can include any suitable number *N* of segments or frames (*N* is an integer). Still in other aspects, a video file received and processed by media editing system **104** can be organized and displayed as a single segment or frame.

The video file represented by bar **202** includes an audio portion/audio track and a video portion/video track. The video portion of the video file is represented by the diagonal patterned lines of bar **202**. The audio portion is collectively represented by the four different lines **208**, **210**, **212**, and **214** spanning across the segments of bar **202**. The four different lines **208**, **210**, **212**, and **214** represent different audio components or layers of the audio portion of the video file. For example, line **208** can represent dialogue between actors in the video, line **210** can represent muffled background noise occurring in the video, line **212** can represent a song playing in the video, and line **214** can represent clapping and cheering of an audience during the video.

The extraction component **108** is configured to separate the video portion and audio portion of a video file. In particular, as seen in FIG. 2, the extraction component can separate the video file represented by **202** into bar **204** and bar **206**. Bar **204** represents the video portion of video file separated from the audio portion and bar **206** represents the audio portion of video file separated from the video portion.

6

Referring back to FIG. 1, in another aspect, extraction component **108** is configured to extract different audio components or layers of an audio track. (The terms audio component and audio layer are used herein interchangeably). For example, in addition to extracting an audio track from a video file, the extraction component **108** can extract different identified audio components of the audio track. In another example, when the receiving component **106** receives an audio file, the extraction component **108** can extract different identified audio components from the audio file. As discussed infra, these different audio components of an audio track or audio file can be identified by identification component **110**.

The term audio component or audio layer refers to a distinct sound present in an audio track. For example, as exemplified above with respect to FIG. 5, an audio track could include several audio components or distinct sounds such as dialogue between actors, muffled background noise, a song, and clapping and cheering of an audience. In an aspect, a sound is considered distinct as a function of the source of the sound. For example, different sources can provide different sounds (e.g., different people, different groups of people, different animals, different objects, different instruments, different inputs of sound, etc.). In another aspect, a sound present in an audio track can be considered distinct based on various features including by not limited to type, intensity, pitch, tone, and harmony. Still in other aspects, a sound can be considered distinct as a function of words spoken and language employed to create the sound.

The identification component **110** is configured to identify different audio components present in an audio track. For example, with reference to FIG. 2, bar **206** representative of the audio track is depicted with four different audio components represented by lines **208**, **210**, **212** and **214**. The identification component **110** is configured to identify these different audio components so that they can be extracted by extraction component **108**. In one aspect the identification component **110** is configured to identify different audio component merely as distinct sounds present in the audio track. In another aspect, the identification component **110** can determine or infer what the distinct sounds are. For example, with reference to FIG. 2, in one aspect, the identification component **110** can identify four different audio components present in the audio track, the components represented by lines **208**, **210**, **212** and **214**. In another aspect, the identification component **110** can further determine that the audio component represented by line **208** is dialogue between actors, the audio component represented by line **210** is muffled background noise, the audio component represented by line **212** is a song, and the audio component represented by line **214** is clapping and cheering of an audience. In one or more additional aspects, the identification component **110** can also identify and note features of the audio components such as intensity, volume, tone, pitch, harmony, etc.

The identification component **110** can employ various mechanisms to identify different audio components (and characteristics of the different audio components) present in an audio track. In an aspect, identification component **110** can analyze frequency patterns generated by the various sounds present in an audio track to identify distinguishable patterns. The identification component **110** can then classify each distinguishable pattern as a different audio component. For example, the identification component **110** can distinguish between different frequency bands based on different oscillation patterns associated with the frequency bands to identify different audio components of an audio track.

In another aspect, identification component **110** can compare known frequency patterns stored in memory **116** to fre-

quency patterns present in an audio track. The identification component **110** can further determine whether a frequency pattern present in the audio track represents a distinct audio component and/or what the distinct audio component is based in a degree of similarity between the known frequency pattern and the frequency pattern present in the audio track. For example, memory **116** can store a look-up table having various known frequency patterns respectively representative of known sounds. According to this example, a frequency pattern identified as pattern #124 could represent a frequency pattern for an ambulance siren. The identification component **110** can match a frequency pattern in an audio track to that of pattern #124 and determine that the frequency pattern in the audio track is an ambulance siren. The identification component **110** can then classify the frequency pattern present in the audio track as a distinguishable audio component and note that the audio component is an ambulance siren.

In another aspect, where an audio component includes spoken language, identification component **110** can employ voice to text recognition software to convert the spoken language into a text file to identify the audio component. The identification component **110** can analyze the text file to further identify what the audio component is and the source of the audio component (e.g., what person is speaking). In an aspect, identification component **110** can match a text file of an audio component to a known reference text file to facilitate identifying the audio component. For example, identification component **110** can access a text reference file for a speech spoken by the President to identify a voice to text interpretation of an audio component as the same speech.

In yet another aspect, identification component **110** can analyze tags or metadata associated with an audio track to facilitate identifying different audio components present in the audio track. For example, a video file can be received by media editing system **104** with annotations or tags identifying one or more audio components of the audio portion of the video file. Similarly, an audio file comprising an audio track can be received by media editing system **104** with annotations or tags identifying one or more audio components of the audio track. The identification component **110** can further employ the tags or annotations associated with an audio track to easily identify the different components of the audio track.

According to this aspect, when a video file or audio track is recorded at a client device **120**, hardware (e.g., different microphones) or software associated with the client device **120** can distinguish between different sounds being received. The client device **120** can include software that then annotates the different sounds with metadata. In some aspects, the annotations can merely identify different components or sounds in the audio track. In other aspects, the annotations can characterize the type or source of the sound (e.g., the annotations can indicate a particular sound is a person speaking or a dog barking). It should be appreciated that degree and specificity of annotations of an audio track will vary based on technical sophistication of hardware and/or software employed by the client device **120**.

For example, client device **120** may recognize dominant sounds in an audio recording and characterize those sounds while grouping extraneous sounds into a classification as background noise. According to this example, a video recording of a polo match may include a plurality of distinguishable sounds, such as the sound of a chatting crowd, the sound of the match announcer, the sound of running horses, the sound of the players grunting and calling out plays, the sound of mallets hitting the chucker, the sound of cars coming and going, the sound of the game horn, etc. As the video is being recorded (or after the video has been recorded) the client

device **120** can recognize and annotate the dominate sounds (e.g., the match announcer, the sound of running horses, the sound of the game horn) while grouping and annotating (or tagging) the non-dominate sounds as background noise.

In an aspect, identification component **110** can also identify video frames or segments of a received video file and the respective segments of the audio track for the video respectively associated with each video frame. For example, with reference to FIG. **2**, the video file represented by bar **202** includes four frames/segments **216**, **218**, **220**, and **222**. The identification component **110** can identify these frames of video and further identify different audio components associated with each frame. For example, bar **206** representative of the audio track for the video file is also broken into segments corresponding to the video frames/segments **216**, **218**, **220** and **222**, respectively. According to this aspect, the audio portion associated with different frames of a video can include different audio components. A user can further edit an audio track associated with a video on a frame by frame basis in addition to an audio component by audio component basis.

Editing component **112** is configured to provide editing tools for editing an audio track received by media editing system **104**. In particular, editing component **112** facilitates editing different components of an audio track identified by identification component **110** and extracted by extraction component **108**. For example, a user may film a video having a variety of occurring sounds or audio components. The user may desire to edit the audio portion of the video to effect different changes with respect to different sounds or audio components of the video occurring at different times of the video. Editing component **112** provides the tools that enable the user to accomplish this task.

With reference to the above example video of a polo match, the video may include the sound of a chatting crowd, the sound of the match announcer, the sound of running horses, the sound of the players grunting and calling out plays, etc. A user may want to decrease volume of the match announcer in frames identified as frames **128** and **129** and increase volume of sound of horses running at frames **128**, **129** and **160**. In another example, the user may want to remove sound of a chatting crowd altogether.

In an aspect, editing component **112** is configured to generate an editing interface that allows a user of a client device **120** to edit the audio portion of a video file. The editing interface can provide editing tools, including tools for separating an audio track from a video file, tools for applying various editing options to audio components of the audio track and tools for re-joining an edited audio track with the video track of the video file.

FIGS. **3** and **4** depict example editing interfaces generated by editing component **112**. With reference to FIG. **3**, presented is an editing interface **300** that displays components of an audio track of a video file in a layered view. FIG. **6** is an extension of FIG. **3**. In FIG. **3**, the extracted audio track for a video file is represented by bar **206**. The extracted audio track for the video file includes four different audio components represented by lines **208**, **210**, **212** and **214** identified by identification component **110**. As noted above, the extraction component **108** is configured to extract these different audio components. The editing component **112** can then present the extracted audio components via editing interface **300**.

Editing component **112** can separate each of the different audio components represented by lines **208**, **210**, **214** and **212** in different layers. Editing component **112** can also segment each of the different audio components by frame (e.g., frame **216**, frame **218**, frame **220** and frame **222**). Although audio components and frames are identified in editing interface **300**

by respective numbers, editing component **112** can apply various different titles to identify items of the interface. For example, where identification component **110** identifies an audio component for what it is, (e.g., a siren, a song, actor John Smith, etc.), editing component **112** can place a title next to the audio component indicating what it is.

Editing component **112** enables a user to edit different components of an audio track jointly or separately. In other words, editing component **112** allows a user to edit an audio track by effecting editing changes to the audio track as a whole or in a piecemeal manner. In particular, editing component **112** allows a user to select a specific audio component for editing, a specific segment of audio associated with a frame of video for editing, and/or a specific audio component associated with a specific frame for editing. For example, with reference to interface **300**, a user could select audio component represented by line **210** and apply editing tools to the entire audio component represented by line **210** (e.g., select the row for component represented by line **210** and apply editing tools to the entire row). In another example, a user can select frame **216** and apply editing tools to each of the audio segments associated with frame **216** (e.g., select the column for frame **216** and apply editing tools to the entire column). In another example, a user could select audio component represented by line **210** at frame **216** for editing individually (e.g., select the cell at component for line **210**, frame **216**). Still in other aspects, a user could select two or more of the audio components and/or two or more of the frames for editing jointly. It should be appreciated that various additional combinations of cell selection, row selection and/or column selection associated with interface **600** can be afforded by editing component **112**.

Editing component **112** can provide various editing tools or options for editing one or more components of an audio track jointly or separately. For example, as seen in interface **300**, the editing component **112** provides an option to mute audio, adjust volume, adjust pitch, adjust speed, adjust tone, equalize, adjust echo, replace audio and add/remove audio (e.g., add a sound affect, remove a sound effect, add a soundtrack, etc). It should be appreciated that the above noted editing tools are merely exemplary. Editing component **112** can be configured to provide various additional known or later developed audio editing tools. For example, additional editing tools that can be applied to one or more components of an audio track by editing component **112** can include an option to change a spoken language or an option to anonymise spoken language of a particular person. In aspect, when editing an audio track via an interface generated by editing component (e.g., using interface **300**), a user can select an one or more audio components to edit and then select one or more editing tools (e.g., one or more editing options **302** of interface **300**) to apply to the one or more audio components. The editing component **112** is further configured to apply editing changes to the audio track. For example, with respect to interface **300**, a user can select one or more audio components to edit and one or more editing options **302** to apply. The user can the select the apply button **604** to effectuated the changes.

FIG. **4** demonstrates an editing interface **400** after application of one or more editing tools to an audio track. FIG. **4** is an extension of FIG. **3**. As seen in FIG. **4**, various changes have been implemented to the audio track for the associated video file. In particular, the audio component represented by line **210** has been removed or muted. The audio component represented by line **212** has been removed or muted at frames **216** and **220**. Further, speed of the audio component represented by line **214** has been reduced, and the volume of audio com-

ponent represented by line **214** has been increased as exemplified by thickening and lengthening of line **214**.

It should be appreciated that an editing interface generated by editing component (e.g., interface **300** and **400**) can provide various other tools that facilitate editing audio in accordance with known editing audio software. For example, editing interfaces **300** and **400** can allow a user to playback an edited audio track to listen to changes, adjust changes, and/or apply additional changes to the audio track.

In some aspects, an audio track may include a large number of identified audio components. However, presenting a user with different audio components may be undesired or overwhelming. For example, where a portion of an audio track includes 15+ identified audio components, a user may not desire to listen to each component to identify specific components the user is interested in editing (e.g., where the components are not identified for what they are but merely identifies as separate/distinct sound). Accordingly, in an aspect, editing component **112** and/or identification component **110** can be configured to discriminate between audio components and select a subset of a plurality of audio components to present to a user for editing. For example, identification component **110** can be configured to select the top five most dominate/distinct sounds for editing. In another example, identification component **110** can be configured to identify extraneous sounds or background noise in an audio track for editing (e.g., for muting). In yet another example, an inference can be made as to a subset of sounds that a user would be interested in editing (e.g., based on context, preferences, historical information . . .).

Referencing FIG. **5**, presented is a diagram of another example system **500** that facilitates editing an audio portion of a video, in accordance with various aspects and embodiments described herein. System **500** includes same features and functionalities of system **100** with the addition of automatic enhancement component **502** and inference component **504**. Repetitive description of like elements employed in respective embodiments of systems and interfaces described herein are omitted for sake of brevity.

Automatic enhancement component **502** is configured to automatically apply editing tools to one or more identified components of an audio track. In particular, automatic enhancement component **502** is configured to make editing decisions on behalf of a user to automatically edit an audio track. For example, automatic enhancement component **502** can analyze an audio track based on the various identified audio components, select one or more editing tools to apply to one or more of the different audio components and apply the one or more editing tools to the one or more different audio components.

In an aspect, a user can request automatic enhancement of an audio track by automatic enhancement component **502**. For example, as seen in FIGS. **3** and **4**, an editing interface can include an auto-correct button **306**. In an aspect, selection of the auto-correct button **306** results in application of one or more editing tools selected by automatic enhancement component **502** to one or more audio components of an audio track selected by automatic enhancement component **502**. In another aspect, automatic enhancement component **502** can be configured to automatically edit an audio track in response to receipt of the audio track, or video comprising the audio track, by receiving component **106**.

In an aspect, automatic enhancement component **502** can apply various rules or algorithms stored in memory **116** that dictate manners for editing components of an audio track to automatically edit the audio track. For example, an algorithm could require audio components identified as background

noise to be muted, audio components identified as music to be adjusted to volume level 5, and audio components identified as dialogue to be adjusted to volume level 8.

In another aspect, automatic enhancement component **502** can employ inference component **504** to facilitate automatically enhancing or editing an audio track. Inference component **504** is configured to provide for or aid in various inferences or determinations associated with identifying audio components of an audio track and classifying the audio components for what the components are (e.g., identifying a component as an ambulance siren or background noise) by identification component **110**. In addition, inference component **504** can facilitate inferring what audio components of an audio track to apply editing tools to and what tools to apply to those audio components. For example, inference component **504** can infer that a particular sound identified in an audio track is a motorcycle. Inference component **504** can further infer that the motorcycle sound should be increased to volume level 8 at frame **499** of a particular video. In an aspect, all or portions of media editing system **104** and media sharing system **102** can be operatively coupled to inference component **504**.

In order to provide for or aid in the numerous inferences described herein, inference component **504** can examine the entirety or a subset of the data to which it is granted access and can provide for reasoning about or infer states of the system, environment, etc. from a set of observations as captured via events and/or data. An inference can be employed to identify a specific context or action, or can generate a probability distribution over states, for example. The inference can be probabilistic—that is, the computation of a probability distribution over states of interest based on a consideration of data and events. An inference can also refer to techniques employed for composing higher-level events from a set of events and/or data.

Such an inference can result in the construction of new events or actions from a set of observed events and/or stored event data, whether or not the events are correlated in close temporal proximity, and whether the events and data come from one or several event and data sources. Various classification (explicitly and/or implicitly trained) schemes and/or systems (e.g., support vector machines, neural networks, expert systems, Bayesian belief networks, fuzzy logic, data fusion engines, etc.) can be employed in connection with performing automatic and/or inferred action in connection with the claimed subject matter.

A classifier can map an input attribute vector, $x=(x_1, x_2, x_3, x_4, x_n)$, to a confidence that the input belongs to a class, such as by $f(x)=\text{confidence}(\text{class})$. Such classification can employ a probabilistic and/or statistical-based analysis (e.g., factoring into the analysis utilities and costs) to prognose or infer an action that a user desires to be automatically performed. A support vector machine (SVM) is an example of a classifier that can be employed. The SVM operates by finding a hyper-surface in the space of possible inputs, where the hyper-surface attempts to split the triggering criteria from the non-triggering events. Intuitively, this makes the classification correct for testing data that is near, but not identical to training data. Other directed and undirected model classification approaches include, e.g., naïve Bayes, Bayesian networks, decision trees, neural networks, fuzzy logic models, and probabilistic classification models providing different patterns of independence can be employed. Classification as used herein also is inclusive of statistical regression that is utilized to develop models of priority.

FIG. **6** presents a diagram of another example system **600** that facilitates editing an audio portion of a video, in accor-

dance with various aspects and embodiments described herein. System **600** includes same features and functionalities of system **500** with the addition of matching component **602**. Repetitive description of like elements employed in respective embodiments of systems and interfaces described herein are omitted for sake of brevity.

Matching component **602** is configured to match audio components of an audio track to same or substantially similar audio files. For example, matching component **602** can access reference media files for a plurality of known audio files stored at or associated with media sharing system **102** and/or media editing system **104**. Matching component **602** can employ various tools to identify matches between audio components and reference files including but not limited to, audio frequency pattern comparison, audio fingerprinting comparison, or voice to text file comparison. In response to an identified match, matching component **602** can suggest replacing the matched audio component with the reference file. For instance, a user may want to replace an audio component with a matched reference file where the reference is a better quality than the audio component.

For example, matching component **602** can match an audio component of an audio track of a live performance of a song to a reference file of a professional studio track recording of the song. The matching component **602** can further provide a user with the option to replace the live performance audio component with the studio version of the song. In another example, matching component **602** can match an audio component of an audio track of an audience laughing with a reference file of a laughing audience and suggest replacing the audio component with the reference file.

FIG. **7** presents a diagram of another example system **700** that facilitates editing an audio portion of a video, in accordance with various aspects and embodiments described herein. System **700** includes same features and functionalities of system **600** with the addition of reproduction component **702**. Repetitive description of like elements employed in respective embodiments of systems and interfaces described herein are omitted for sake of brevity.

Reproduction component **702** is configured to re-join or combine an edited audio track with an extracted video track of a video file. In particular, reproduction component **702** is configured to combine extracted audio components, as edited, into a single edited audio track and combine the edited audio track with the original extracted video track. With reference to FIG. **8**, presented is an example process **800** for combining an edited audio track with an extracted video track of a video file by reproduction component **702**. FIG. **8** is an extension of FIG. **4**. As seen in FIG. **8**, bar **802** represents an edited audio track of a video file. In particular, the edited audio track represented by bar **802** includes the edits applied to the audio track as shown in FIG. **4**. For example, the edited audio track represented by bar **802** includes the original audio component represented by line **208**, edited audio component represented by line **214**, no component represented by line **210** and no component represented by line **212** in frames **216** and **220**. Reproduction component **702** combines the edited audio track represented by bar **802** with the original extracted video track represented by bar **204** to generate an edited video file represented by bar **806** that includes the original video track and the edited audio track.

FIG. **9** presents a diagram of another example system **900** that facilitates editing an audio portion of a video, in accordance with various aspects and embodiments described herein. System **900** includes same or similar features and functionalities of other systems described herein. Repetitive

description of like elements employed in respective embodiments of systems and interfaces described herein are omitted for sake of brevity.

System 900 includes client device 902, one or more networks 118, media sharing system 914 and media editing system 916. With system 900, media sharing system 914 is depicted as including media editing system 916. Media sharing system 914 and media editing system 916 can include one or more features of media sharing system 102 and media editing system 104 (and vice versa). Client device 902 can include media recording component 904, media tagging component 906 and media uploading component 908. Client device 902 includes memory 912 for storing computer executable components and instructions. Client device 902 further includes a processor 910 to facilitate operation of the instructions (e.g., computer executable components and instructions) by client device 902.

System 900 emphasizes features of an example client device 902 that facilitates recording video and/or audio and annotating audio components in accordance with an embodiment. As noted with respect to the discussion of FIG. 1, in an aspect, receiving component 106 of media editing system 104 can receive a video file having an audio track that is annotated to indicate different sounds or audio components present in the audio track. The identification component 110 can employ the annotations to easily identify different audio components of the audio track for extraction by extraction component 108. Client device 902 is configured to provide such annotated video and audio files to media sharing system 914 for editing with media editing system 916.

Media recording component 904 is configured to record video and/or audio files. For example, media recording component 904 can include a video camera and one or more microphones. Media tagging component 906 is configured to identify and tag different audio components of received/recorded audio. For example, media tagging component 906 can associate metadata with an audio track indicating distinct audio components and/or identifying what the distinct audio components are (e.g., background noise, song, dialogue, etc.). In an aspect, media tagging component 906 is configured to tag different components of audio as the audio is received/recorded. Media uploading component 908 is configured to upload a tagged or annotated media file to media sharing system 914 via a network 118.

FIG. 10 presents a diagram of another example system 1000 that facilitates editing an audio portion of a video, in accordance with various aspects and embodiments described herein. System 1000 includes same or similar features and functionalities of system 900. Repetitive description of like elements employed in respective embodiments of systems and interfaces described herein are omitted for sake of brevity.

System 1000 demonstrates an example embodiment of a system that facilitates editing an audio portion of a video similar to system 900. However unlike system 900, system 1000 includes a client device 1002 that includes a media editing system 1004. Media editing system 1004 can include one or more of the features and functionalities of media editing system 104. System 1000 further includes one or more networks, media sharing system 1008 and media editing system 1010. Media sharing system 1008 and media editing system 1010 can include one or more features of media sharing system 102 and media editing system 104 (and vice versa).

In an aspect, media editing system 1004 includes a portion of the components of media editing system 104 while media editing system 1010 includes another portion of the compo-

nents of media editing system 104. For example, media editing system 1004 can include extraction component 108 and identification component 110 while media editing system 1010 includes editing component 112. In another example, media editing system 1004 can include identification component 110 while media editing system 1010 includes include extraction component 108 and editing component 112. In an aspect, (not shown) media sharing system 1008 does not include a media editing system. According to this aspect, the various features of media editing system 104 are provided at client device 1002.

In view of the example systems and/or devices described herein, example methods that can be implemented in accordance with the disclosed subject matter can be further appreciated with reference to flowcharts in FIGS. 11-13. For purposes of simplicity of explanation, example methods disclosed herein are presented and described as a series of acts; however, it is to be understood and appreciated that the disclosed subject matter is not limited by the order of acts, as some acts may occur in different orders and/or concurrently with other acts from that shown and described herein. For example, a method disclosed herein could alternatively be represented as a series of interrelated states or events, such as in a state diagram. Moreover, interaction diagram(s) may represent methods in accordance with the disclosed subject matter when disparate entities enact disparate portions of the methods. Furthermore, not all illustrated acts may be required to implement a method in accordance with the subject specification. It should be further appreciated that the methods disclosed throughout the subject specification are capable of being stored on an article of manufacture to facilitate transporting and transferring such methods to computers for execution by a processor or for storage in a memory.

FIG. 11 illustrates a flow chart of an example method 1100 for editing an audio track of a video file, in accordance with aspects described herein. At 1102, a video is received as an upload from a client device over a network (e.g., via receiving component 106). At 1104, an audio track and a video track are separated from the video (e.g., using extraction component 108). At 1106, two or more different audio components of the audio track are identified (e.g., using identification component 110). At 1108, the two or more different audio components are separated (e.g., using extraction component 108). At 1110, an editing interface is generated and at 1112, the client device is provided one or more options for editing the two or more different audio components jointly or separately via the editing interface (e.g., using editing component 112).

FIG. 12 illustrates a flow chart of another example method 1200 for editing an audio track of a video file, in accordance with aspects described herein. At 1202, a video is received as an upload from a client device over a network (e.g., via receiving component 106). At 1204, an audio track and a video track are separated from the video (e.g., using extraction component 108). At 1206, two or more different audio components of the audio track are identified (e.g., using identification component 110). At 1208, the two or more different audio components are separated (e.g., using extraction component 108). At 1210, an editing interface is generated and at 1212, the client device is provided one or more options for editing the two or more different audio components jointly or separately via the editing interface (e.g., using editing component 112). At 1214, the one or more options for editing the two or more different audio components are applied to generate an edited audio track comprising the two or more different audio components as edited (e.g., using editing com-

ponent 112). At 1216, the edited audio track is combined with the video track to generate an edited video (e.g., using reproduction component 702).

FIG. 13 illustrates a flow chart of another example method 1300 for editing an audio track of a video file, in accordance with aspects described herein. At 1302, a video is received as an upload from a client device over a network (e.g., via receiving component 106). At 1304, an audio track and a video track are separated from the video (e.g., using extraction component 108). At 1306, two or more different audio components of the audio track are identified (e.g., using identification component 110). At 1308, a first editing tool is applied to a first one of the audio components and at 3110, a second editing tool is applied to a second one of the audio components, wherein the first editing tool and the second editing tool are different (e.g., using editing component 112). At 1312, the two or more different audio components are combined after the applying the first editing tool and the second editing tool to generate an edited audio track (e.g., using reproduction component 702). At 1314, the edited audio track is joined with the video track to generate an edited video file (e.g., using reproduction component 702).

Example Operating Environments

The systems and processes described below can be embodied within hardware, such as a single integrated circuit (IC) chip, multiple ICs, an application specific integrated circuit (ASIC), or the like. Further, the order in which some or all of the process blocks appear in each process should not be deemed limiting. Rather, it should be understood that some of the process blocks can be executed in a variety of orders, not all of which may be explicitly illustrated in this disclosure.

With reference to FIG. 14, a suitable environment 1400 for implementing various aspects of the claimed subject matter includes a computer 1402. The computer 1402 includes a processing unit 1404, a system memory 1406, a codec 1405, and a system bus 1408. The system bus 1408 couples system components including, but not limited to, the system memory 1406 to the processing unit 1404. The processing unit 1404 can be any of various available processors. Dual microprocessors and other multiprocessor architectures also can be employed as the processing unit 1404.

The system bus 1408 can be any of several types of bus structure(s) including the memory bus or memory controller, a peripheral bus or external bus, and/or a local bus using any variety of available bus architectures including, but not limited to, Industrial Standard Architecture (ISA), Micro-Channel Architecture (MSA), Extended ISA (EISA), Intelligent Drive Electronics (IDE), VESA Local Bus (VLB), Peripheral Component Interconnect (PCI), Card Bus, Universal Serial Bus (USB), Advanced Graphics Port (AGP), Personal Computer Memory Card International Association bus (PCMCIA), Firewire (IEEE 13144), and Small Computer Systems Interface (SCSI).

The system memory 1406 includes volatile memory 1410 and non-volatile memory 1412. The basic input/output system (BIOS), containing the basic routines to transfer information between elements within the computer 1402, such as during start-up, is stored in non-volatile memory 1412. In addition, according to present innovations, codec 1405 may include at least one of an encoder or decoder, wherein the at least one of an encoder or decoder may consist of hardware, a combination of hardware and software, or software. Although, codec 1405 is depicted as a separate component, codec 1405 may be contained within non-volatile memory 1412. By way of illustration, and not limitation, non-volatile memory 1412 can include read only memory (ROM), programmable ROM (PROM), electrically programmable ROM

(EPROM), electrically erasable programmable ROM (EEPROM), or flash memory. Volatile memory 1410 includes random access memory (RAM), which acts as external cache memory. According to present aspects, the volatile memory may store the write operation retry logic (not shown in FIG. 14) and the like. By way of illustration and not limitation, RAM is available in many forms such as static RAM (SRAM), dynamic RAM (DRAM), synchronous DRAM (SDRAM), double data rate SDRAM (DDR SDRAM), and enhanced SDRAM (ESDRAM).

Computer 1402 may also include removable/non-removable, volatile/non-volatile computer storage medium. FIG. 14 illustrates, for example, disk storage 1414. Disk storage 1414 includes, but is not limited to, devices like a magnetic disk drive, solid state disk (SSD) floppy disk drive, tape drive, Jaz drive, Zip drive, LS-70 drive, flash memory card, or memory stick. In addition, disk storage 1414 can include storage medium separately or in combination with other storage medium including, but not limited to, an optical disk drive such as a compact disk ROM device (CD-ROM), CD recordable drive (CD-R Drive), CD rewritable drive (CD-RW Drive) or a digital versatile disk ROM drive (DVD-ROM). To facilitate connection of the disk storage devices 1414 to the system bus 1408, a removable or non-removable interface is typically used, such as interface 1416.

It is to be appreciated that FIG. 14 describes software that acts as an intermediary between users and the basic computer resources described in the suitable operating environment 1400. Such software includes an operating system 1418. Operating system 1418, which can be stored on disk storage 1414, acts to control and allocate resources of the computer system 1402. Applications 1420 take advantage of the management of resources by operating system 1418 through program modules 1424, and program data 1426, such as the boot/shutdown transaction table and the like, stored either in system memory 1406 or on disk storage 1414. It is to be appreciated that the claimed subject matter can be implemented with various operating systems or combinations of operating systems.

A user enters commands or information into the computer 1402 through input device(s) 1428. Input devices 1428 include, but are not limited to, a pointing device such as a mouse, trackball, stylus, touch pad, keyboard, microphone, joystick, game pad, satellite dish, scanner, TV tuner card, digital camera, digital video camera, web camera, and the like. These and other input devices connect to the processing unit 1404 through the system bus 1408 via interface port(s) 1430. Interface port(s) 1430 include, for example, a serial port, a parallel port, a game port, and a universal serial bus (USB). Output device(s) 1436 use some of the same type of ports as input device(s). Thus, for example, a USB port may be used to provide input to computer 1402, and to output information from computer 1402 to an output device 1436. Output adapter 1434 is provided to illustrate that there are some output devices 1436 like monitors, speakers, and printers, among other output devices 1436, which require special adapters. The output adapters 1434 include, by way of illustration and not limitation, video and sound cards that provide a means of connection between the output device 1436 and the system bus 1408. It should be noted that other devices and/or systems of devices provide both input and output capabilities such as remote computer(s) 1438.

Computer 1402 can operate in a networked environment using logical connections to one or more remote computers, such as remote computer(s) 1438. The remote computer(s) 1438 can be a personal computer, a server, a router, a network PC, a workstation, a microprocessor based appliance, a peer

device, a smart phone, a tablet, or other network node, and typically includes many of the elements described relative to computer **1402**. For purposes of brevity, only a memory storage device **1440** is illustrated with remote computer(s) **1438**. Remote computer(s) **1438** is logically connected to computer **1402** through a network interface **1442** and then connected via communication connection(s) **1444**. Network interface **1442** encompasses wire and/or wireless communication networks such as local-area networks (LAN) and wide-area networks (WAN) and cellular networks. LAN technologies include Fiber Distributed Data Interface (FDDI), Copper Distributed Data Interface (CDDI), Ethernet, Token Ring and the like. WAN technologies include, but are not limited to, point-to-point links, circuit switching networks like Integrated Services Digital Networks (ISDN) and variations thereon, packet switching networks, and Digital Subscriber Lines (DSL).

Communication connection(s) **1444** refers to the hardware/software employed to connect the network interface **1442** to the bus **1408**. While communication connection **1444** is shown for illustrative clarity inside computer **1402**, it can also be external to computer **1402**. The hardware/software necessary for connection to the network interface **1442** includes, for exemplary purposes only, internal and external technologies such as, modems including regular telephone grade modems, cable modems and DSL modems, ISDN adapters, and wired and wireless Ethernet cards, hubs, and routers.

Referring now to FIG. **15**, there is illustrated a schematic block diagram of a computing environment **1500** in accordance with this disclosure. The system **1500** includes one or more client(s) **1502** (e.g., laptops, smart phones, PDAs, media players, computers, portable electronic devices, tablets, and the like). The client(s) **1502** can be hardware and/or software (e.g., threads, processes, computing devices). The system **1500** also includes one or more server(s) **1504**. The server(s) **1504** can also be hardware or hardware in combination with software (e.g., threads, processes, computing devices). The servers **1504** can house threads to perform transformations by employing aspects of this disclosure, for example. One possible communication between a client **1502** and a server **1504** can be in the form of a data packet transmitted between two or more computer processes wherein the data packet may include video data. The data packet can include a metadata, e.g., associated contextual information, for example. The system **1500** includes a communication framework **1506** (e.g., a global communication network such as the Internet, or mobile network(s)) that can be employed to facilitate communications between the client(s) **1502** and the server(s) **1504**.

Communications can be facilitated via a wired (including optical fiber) and/or wireless technology. The client(s) **1502** include or are operatively connected to one or more client data store(s) **1508** that can be employed to store information local to the client(s) **1502** (e.g., associated contextual information). Similarly, the server(s) **1504** are operatively include or are operatively connected to one or more server data store(s) **1510** that can be employed to store information local to the servers **1504**.

In one embodiment, a client **1502** can transfer an encoded file, in accordance with the disclosed subject matter, to server **1504**. Server **1504** can store the file, decode the file, or transmit the file to another client **1502**. It is to be appreciated, that a client **1502** can also transfer uncompressed file to a server **1504** and server **1504** can compress the file in accordance with the disclosed subject matter. Likewise, server **1504** can encode video information and transmit the information via communication framework **1506** to one or more clients **1502**.

The illustrated aspects of the disclosure may also be practiced in distributed computing environments where certain tasks are performed by remote processing devices that are linked through a communications network. In a distributed computing environment, program modules can be located in both local and remote memory storage devices.

Moreover, it is to be appreciated that various components described in this description can include electrical circuit(s) that can include components and circuitry elements of suitable value in order to implement the embodiments of the subject innovation(s). Furthermore, it can be appreciated that many of the various components can be implemented on one or more integrated circuit (IC) chips. For example, in one embodiment, a set of components can be implemented in a single IC chip. In other embodiments, one or more of respective components are fabricated or implemented on separate IC chips.

What has been described above includes examples of the embodiments of the present invention. It is, of course, not possible to describe every conceivable combination of components or methodologies for purposes of describing the claimed subject matter, but it is to be appreciated that many further combinations and permutations of the subject innovation are possible. Accordingly, the claimed subject matter is intended to embrace all such alterations, modifications, and variations that fall within the spirit and scope of the appended claims. Moreover, the above description of illustrated embodiments of the subject disclosure, including what is described in the Abstract, is not intended to be exhaustive or to limit the disclosed embodiments to the precise forms disclosed. While specific embodiments and examples are described in this disclosure for illustrative purposes, various modifications are possible that are considered within the scope of such embodiments and examples, as those skilled in the relevant art can recognize.

In particular and in regard to the various functions performed by the above described components, devices, circuits, systems and the like, the terms used to describe such components are intended to correspond, unless otherwise indicated, to any component which performs the specified function of the described component (e.g., a functional equivalent), even though not structurally equivalent to the disclosed structure, which performs the function in the disclosure illustrated exemplary aspects of the claimed subject matter. In this regard, it will also be recognized that the innovation includes a system as well as a computer-readable storage medium having computer-executable instructions for performing the acts and/or events of the various methods of the claimed subject matter.

The aforementioned systems/circuits/modules have been described with respect to interaction between several components/blocks. It can be appreciated that such systems/circuits and components/blocks can include those components or specified sub-components, some of the specified components or sub-components, and/or additional components, and according to various permutations and combinations of the foregoing. Sub-components can also be implemented as components communicatively coupled to other components rather than included within parent components (hierarchical). Additionally, it should be noted that one or more components may be combined into a single component providing aggregate functionality or divided into several separate sub-components, and any one or more middle layers, such as a management layer, may be provided to communicatively couple to such sub-components in order to provide integrated functionality. Any components described in this disclosure may

also interact with one or more other components not specifically described in this disclosure but known by those of skill in the art.

In addition, while a particular feature of the subject innovation may have been disclosed with respect to only one of several implementations, such feature may be combined with one or more other features of the other implementations as may be desired and advantageous for any given or particular application. Furthermore, to the extent that the terms “includes,” “including,” “has,” “contains,” variants thereof, and other similar words are used in either the detailed description or the claims, these terms are intended to be inclusive in a manner similar to the term “comprising” as an open transition word without precluding any additional or other elements.

As used in this application, the terms “component,” “module,” “system,” or the like are generally intended to refer to a computer-related entity, either hardware (e.g., a circuit), a combination of hardware and software, software, or an entity related to an operational machine with one or more specific functionalities. For example, a component may be, but is not limited to being, a process running on a processor (e.g., digital signal processor), a processor, an object, an executable, a thread of execution, a program, and/or a computer. By way of illustration, both an application running on a controller and the controller can be a component. One or more components may reside within a process and/or thread of execution and a component may be localized on one computer and/or distributed between two or more computers. Further, a “device” can come in the form of specially designed hardware; generalized hardware made specialized by the execution of software thereon that enables the hardware to perform specific function; software stored on a computer readable storage medium; software transmitted on a computer readable transmission medium; or a combination thereof.

Moreover, the words “example” or “exemplary” are used in this disclosure to mean serving as an example, instance, or illustration. Any aspect or design described in this disclosure as “exemplary” is not necessarily to be construed as preferred or advantageous over other aspects or designs. Rather, use of the words “example” or “exemplary” is intended to present concepts in a concrete fashion. As used in this application, the term “or” is intended to mean an inclusive “or” rather than an exclusive “or”. That is, unless specified otherwise, or clear from context, “X employs A or B” is intended to mean any of the natural inclusive permutations. That is, if X employs A; X employs B; or X employs both A and B, then “X employs A or B” is satisfied under any of the foregoing instances. In addition, the articles “a” and “an” as used in this application and the appended claims should generally be construed to mean “one or more” unless specified otherwise or clear from context to be directed to a singular form.

Computing devices typically include a variety of media, which can include computer-readable storage media and/or communications media, in which these two terms are used in this description differently from one another as follows. Computer-readable storage media can be any available storage media that can be accessed by the computer, is typically of a non-transitory nature, and can include both volatile and nonvolatile media, removable and non-removable media. By way of example, and not limitation, computer-readable storage media can be implemented in connection with any method or technology for storage of information such as computer-readable instructions, program modules, structured data, or unstructured data. Computer-readable storage media can include, but are not limited to, RAM, ROM, EEPROM, flash memory or other memory technology, CD-

ROM, digital versatile disk (DVD) or other optical disk storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or other tangible and/or non-transitory media which can be used to store desired information. Computer-readable storage media can be accessed by one or more local or remote computing devices, e.g., via access requests, queries or other data retrieval protocols, for a variety of operations with respect to the information stored by the medium.

On the other hand, communications media typically embody computer-readable instructions, data structures, program modules or other structured or unstructured data in a data signal that can be transitory such as a modulated data signal, e.g., a carrier wave or other transport mechanism, and includes any information delivery or transport media. The term “modulated data signal” or signals refers to a signal that has one or more of its characteristics set or changed in such a manner as to encode information in one or more signals. By way of example, and not limitation, communication media include wired media, such as a wired network or direct-wired connection, and wireless media such as acoustic, RF, infrared and other wireless media.

In view of the exemplary systems described above, methodologies that may be implemented in accordance with the described subject matter will be better appreciated with reference to the flowcharts of the various figures. For simplicity of explanation, the methodologies are depicted and described as a series of acts. However, acts in accordance with this disclosure can occur in various orders and/or concurrently, and with other acts not presented and described in this disclosure. Furthermore, not all illustrated acts may be required to implement the methodologies in accordance with certain aspects of this disclosure. In addition, those skilled in the art will understand and appreciate that the methodologies could alternatively be represented as a series of interrelated states via a state diagram or events. Additionally, it should be appreciated that the methodologies disclosed in this disclosure are capable of being stored on an article of manufacture to facilitate transporting and transferring such methodologies to computing devices. The term article of manufacture, as used in this disclosure, is intended to encompass a computer program accessible from any computer-readable device or storage media.

What is claimed is:

1. A system, comprising:

a memory having stored thereon computer executable components;

a processor that executes at least the following computer executable components:

a receiving component configured to receive a video as an upload to a website from a client device over a network;

an identification component configured to:

analyze audio frequencies of an audio track of the video;

identify patterns in the audio frequencies;

identify two or more different and concurrent audio layers of the audio track based on the patterns; and

identify at least one of the audio layers as a dialogue audio layer based on the identified patterns including at least one pattern corresponding to one or more voices within the audio track;

an extraction component configured to extract and separate the audio layers;

an editing component configured to:

21

generate an editing interface on the website, the interface including a set of editing options and a representation of each of the audio layers;

receive, via the editing interface, input from the client device over the network regarding editing the audio layers separately, the input including a selection of at least one of the editing options and at least one of the representations of the audio layers;

edit the selected audio layers based on the selected editing options; and

generate an edited audio track comprising the audio layers as edited; and

a reproduction component configured to combine the edited audio track with an extracted video track of the video to generate an edited video to post on the website.

2. The system of claim 1, wherein the system is located at a server device accessible to one or more client devices via the network.

3. The system of claim 1, wherein the extraction component is further configured to separate the audio track from a video track of the video.

4. The system of claim 1, wherein the input regarding editing the two or more different and concurrent audio layers includes at least one of, a request to modify volume, a request to mute, a request to add a sound effect, a request to remove a sound effect, or a request to change pitch.

5. The system of claim 1, wherein the input regarding editing the two or more different and concurrent audio layers includes a request to apply a first editing option to a first one of the two or more different and concurrent audio layers and a request to apply a second editing option to a second one of the two or more different and concurrent audio layers, wherein the first one of the two or more different and concurrent audio layers includes the dialogue audio layer and the first editing option and the second editing option are different.

6. The system of claim 1, wherein the audio track comprises a plurality of sequential segments respectively associated with sequential frames of the video, wherein the identification component is configured to identify two or more different and concurrent audio layers respectively associated with respective segments of the sequential segments.

7. The system of claim 1, further comprising an inference component configured to analyze the two or more different and concurrent audio layers and determine or infer an editing option to apply to at least one of the two or more different and concurrent audio layers.

8. The system of claim 1, wherein the two or more different and concurrent audio layers span along an entirety of the audio track.

9. The system of claim 1, wherein the representations of each of the audio layers are presented within a respective frame of a set of layered frames.

10. The system of claim 1, wherein the identification component identifies a set of audio layers not including the dialogue audio layer as background noise.

11. The system of claim 1, further comprising an automatic enhancement component configured to automatically edit the audio layers by increasing a volume of the dialogue audio layer and decreasing or muting a volume of a remaining set of audio layers, wherein the extracted audio layers are automatically edited in response to the selected editing option including a selection that corresponds to a dialogue enhancement option.

12. The system of claim 1, further comprising a matching component configured to match one of the audio layers with

22

a reference file, wherein the set of editing options includes an option to replace the matched audio layer with the reference file.

13. The system of claim 12, wherein the editing component replaces the matched audio layer with the reference file in response to the selected editing option including an option to replace the matched audio layer with the reference file.

14. The system of claim 13, wherein the reference file includes a music track.

15. The system of claim 1, wherein the identification component is configured to identify patterns in the audio frequencies by referencing a look-up table storing patterns corresponding to previously identified sounds.

16. The system of claim 1, wherein the identification component is configured to identify patterns in the audio frequencies by employing a voice to text recognition to covert spoken language into a text file to identify the dialogue audio layer.

17. The system of claim 1, further comprising a media tagging component configured to associate metadata with each of the identified audio layers.

18. A method comprising:

using a processor to execute the following computer executable instructions stored in a memory to perform the following acts:

receiving a video as an upload from a client device over a network;

analyzing audio frequencies of an audio track of the video;

identifying patterns in the audio frequencies;

identifying two or more different and concurrent audio layers of the audio track based on the patterns;

identifying at least one of the audio layers as a dialogue audio layer based on the identified patterns including at least one pattern corresponding to one or more voices within the audio track;

separating the two or more different and concurrent audio layers;

generating an editing interface on a website, the interface including a set of editing options and a representation of each of the two or more different and concurrent audio layers;

receiving input from the client device over the network regarding editing the two or more different and concurrent audio layers separately via the editing interface;

editing the two or more different and concurrent audio layers based on the input; and

generating an edited audio track comprising the two or more different and concurrent audio layers as edited.

19. The method of claim 18, wherein the receiving the input comprises at receiving at least one of, a request to modify volume, a request to mute, a request to add a sound effect, a request to remove a sound effect, or a request to change pitch.

20. The method of claim 18, further comprising combining the edited audio track with an extracted video track of the video to generate an edited video.

21. The method of claim 18, wherein the receiving the input comprises receiving a request to apply a first editing option to a first one of the two or more different and concurrent audio layers and a second editing option to a second one of the two or more different and concurrent audio layers, wherein the first one of the two or more different and concurrent audio layers includes the dialogue audio layer and the first editing option and the second editing option are different.

22. The method of claim 18, further comprising analyzing the two or more different audio components and determining or inferring an editing option to apply to at least one of the two or more different and concurrent audio layers.

23

23. A non-transitory computer-readable storage storing computer-readable instructions that, in response to execution, cause a computing system to perform operations, comprising:

receiving a video as an upload from a client device over a network;

analyzing audio frequencies of an audio track of the video;

identifying patterns in the audio frequencies;

identifying two or more different and concurrent audio layers of the audio track based on the patterns;

identifying at least one of the audio layers as a dialogue audio layer based on the identified patterns including at least one pattern corresponding to one or more voices within the audio track;

separating the two or more different and concurrent audio layers;

generating an editing interface on a website;

24

receiving a request from the client device over the network to apply an editing option to a subset of the two or more different and concurrent audio layers via the editing interface;

applying the editing option to only the subset of the two or more different and concurrent audio layers in response to the request;

generating an edited audio track comprising the subset of the two or more different and concurrent audio layers in response to the applied editing option; and

combining the edited audio track with an extracted video track of the video to generate an edited video.

24. The non-transitory computer-readable storage of claim 23, wherein the editing option includes at least one of, an option to modify volume, an option to mute, an option to add a sound effect, an option to remove a sound effect, or an option to change pitch.

* * * * *