



(51) International Patent Classification:

H04N 17/00 (2006.01) *H04N 21/234* (201 1.01)
H04N 21/8547 (201 1.01) *H04N 21/2343* (201 1.01)
H04N 21/233 (201 1.01) *H04N 21/2362* (201 1.01)

(21) International Application Number:

PCT/US20 14/053029

(22) International Filing Date:

28 August 2014 (28.08.2014)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

14/015,278 30 August 2013 (30.08.2013) US

(71) Applicant: MICROSOFT CORPORATION [US/US];
One Microsoft Way, Redmond, Washington 98052-6399
(US).

(72) Inventors: DALAL, Firoz; c/o Microsoft Corporation,
LCA - International Patents (8/1 172), One Microsoft Way,
Redmond, Washington 98052-6399 (US). WU, Yongjun;
c/o Microsoft Corporation, LCA - International Patents

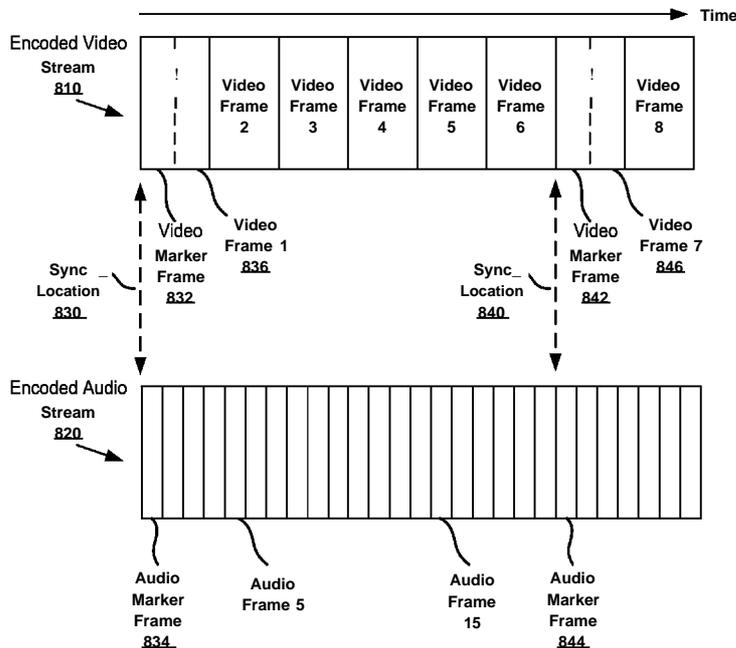
(8/1 172), One Microsoft Way, Redmond, Washington
98052-6399 (US). SADHWANI, Shyam; c/o Microsoft
Corporation, LCA - International Patents (8/1 172), One
Microsoft Way, Redmond, Washington 98052-6399 (US).
BONAPARTE, Jarred; c/o Microsoft Corporation, LCA -
International Patents (8/1 172), One Microsoft Way, Red-
mond, Washington 98052-6399 (US).

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

(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,

[Continued on nextpage]

(54) Title: AUDIO VIDEO PLAYBACK SYNCHRONIZATION FOR ENCODED MEDIA



(57) Abstract: Techniques are described for inserting
encoded markers into encoded audio-video content.
For example, encoded audio-video content can be re-
ceived and corresponding encoded audio and video
markers can be inserted. The encoded audio and
video markers can be inserted without changing the
overall duration of the encoded audio and video
streams and without changing most or all of the prop-
erties of the encoded audio and video streams. Cor-
responding encoded audio and video markers can be
inserted at multiple locations (e.g., sync locations) in
the encoded audio and video streams. Audio-video
synchronization testing can be performed using en-
coded audio-video content with inserted encoded au-
dio-video markers.

FIG. 8

WO 2015/031548 A1

GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(H))*
- *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(in))*

Published:

- *with international search report (Art. 21(3))*

AUDIO VIDEO PLAYBACK SYNCHRONIZATION FOR ENCODED MEDIA**BACKGROUND**

[001] People are increasingly using different types of devices and software applications
5 to play multimedia content. For example, people use computing devices, such as desktop
computers and mobile devices, to view movies and video clips, to download on-demand or
streaming multimedia content, to record and capture audio-video content (e.g., a video
chat or on-line conference), and to perform other recording and playback tasks using
multimedia content.

10 [002] In order for the user to have a positive experience when producing or consuming
multimedia content, it is important that audio and video information in the multimedia
content be synchronized. For example, if a user is watching a movie, the video content
should be synchronized to the audio content (e.g., so that an actor's mouth is moving in
time with the words that the actor is speaking).

15 [003] However, with the increasing number of different types of software and
hardware used to consume and produce multimedia content, testing for audio-video
synchronization can be problematic and time consuming.

[004] Some solutions have been developed to help test audio-video synchronization
using uncompressed audio and video content. However, such solutions may only be useful
20 in detecting synchronization problems at the encoding or authoring stage, and may not be
able to detect or isolate problems at the playback stage.

[005] Therefore, there exists ample opportunity for improvement in technologies
related to testing audio-video synchronization.

SUMMARY

25 [006] This Summary is provided to introduce a selection of concepts in a simplified
form that are further described below in the Detailed Description. This Summary is not
intended to identify key features or essential features of the claimed subject matter, nor is
it intended to be used to limit the scope of the claimed subject matter.

[007] Techniques and tools are described for inserting encoded markers into encoded
30 audio-video content. For example, encoded video markers can be inserted into an encoded
video stream without increasing the overall duration of the encoded video stream.

Furthermore, the original video stream can remain substantially unchanged, retaining all
original (or nearly all original) video properties. Encoded audio markers can be inserted
into an encoded audio stream (e.g., at sync locations corresponding to the inserted video

markers) without increasing the overall duration of the encoded audio stream.

Furthermore, the original audio stream can remain substantially unchanged, retaining all original (or nearly all original) audio properties. Audio-video synchronization testing can be performed using encoded audio-video content with inserted encoded audio-video

5 markers.

[008] For example, a method can be provided for inserting encoded markers into encoded audio-video content. The method comprises receiving encoded audio-video content comprising an encoded video stream and an encoded audio stream, inserting an encoded video marker into the encoded video stream at a video sync location, inserting an
10 encoded audio marker into the encoded audio stream at an audio sync location corresponding to the video sync location, and outputting the encoded video stream with the inserted encoded video marker and the encoded audio stream with the inserted encoded audio marker. The encoded video marker can be inserted without decoding or encoding (or re-encoding) the encoded video stream, and the encoded audio marker can be inserted
15 without decoding or encoding (or re-encoding) the encoded audio stream.

[009] As another example, a method can be provided for inserting encoded markers into encoded audio-video content. The method comprises receiving encoded audio-video content comprising an encoded video stream and an encoded audio stream, analyzing the encoded video stream to determine video encoding parameters, encoding a video marker
20 using, at least in part, the determined video encoding parameters to create an encoded video marker compatible with the encoded video stream, inserting the encoded video marker into the encoded video stream at a video sync location, analyzing the encoded audio stream to determine audio encoding parameters, encoding an audio marker using, at least in part, the determined audio encoding parameters to create an encoded audio marker
25 compatible with the encoded audio stream, inserting the encoded audio marker into the encoded audio stream at an audio sync location corresponding to the video sync location, and outputting the encoded video stream with the inserted encoded video marker and the encoded audio stream with the inserted encoded audio marker. The encoded video marker can be inserted without decoding or encoding (or re-encoding) the encoded video stream,
30 and the overall duration of the encoded video stream can remain the same after the encoded video marker is inserted. The encoded audio marker can be inserted without decoding or encoding (or re-encoding) the encoded audio stream, and the overall duration of the encoded audio stream can remain the same after the encoded audio marker is inserted.

[010] As another example, a method can be provided for testing synchronization of encoded audio-video content. The method comprises receiving encoded audio-video content comprising an encoded video stream and an encoded audio stream, the encoded video stream comprising one or more video markers, and the encoded audio stream
5 comprising one or more corresponding audio markers, initiating playback of the encoded audio-video content, during playback of the encoded audio-video content: capturing decoded video content (e.g., where the captured video content is captured at a reduced resolution), and capturing decoded audio content (e.g., where the captured audio content is captured with a reduced number of audio channels and/or reduced quality), from the
10 captured video content and the captured audio content, matching the one or more video markers and the one or more corresponding audio markers, and based on the matching, outputting audio-video synchronization information.

[011] As another example, computing devices comprising processing units and memory can be provided for performing operations described herein. For example, a
15 computing device can receive encoded audio-video content, insert encoded audio-video markers, and output encoded audio-video content with the inserted markers (e.g., output as test audio-video content). A computing device can test audio-video synchronization by receiving encoded audio-video content with inserted markers, capturing audio-video content during playback, and matching audio-video markers to determine synchronization
20 results.

[012] As described herein, a variety of other features and advantages can be incorporated into the technologies as desired.

BRIEF DESCRIPTION OF THE DRAWINGS

[013] FIG. 1 is an example diagram depicting insertion of encoded audio-video
25 markers into encoded audio-video streams.

[014] FIG. 2 is an example diagram depicting insertion of encoded audio-video markers into encoded audio-video streams, including de-multiplexing and multiplexing audio content.

[015] FIG. 3 is a flowchart of an example method for inserting encoded markers into
30 encoded audio-video content.

[016] FIG. 4 is a flowchart of an example method for inserting encoded video markers while maintaining the same overall duration.

[017] FIG. 5 is a flowchart of an example method for creating encoded audio-video markers based on audio-video encoding parameters.

[018] FIG. 6 is a prior art diagram of an example video stream and video timestamp table.

[019] FIG. 7 is a diagram of an example video stream and video timestamp table showing an inserted encoded video marker frame.

5 [020] FIG. 8 is a diagram showing example video and audio streams with inserted video and audio markers at sync locations.

[021] FIG. 9 is a flowchart of an example method for testing synchronization of encoded audio-video content with inserted encoded audio-video markers.

10 [022] FIG. 10 is a diagram of an exemplary computing system in which some described embodiments can be implemented.

[023] FIG. 11 is an exemplary mobile device that can be used in conjunction with the technologies described herein.

[024] FIG. 12 is an exemplary cloud-support environment that can be used in conjunction with the technologies described herein.

15 **DETAILED DESCRIPTION**

Example 1 - Overview

[025] As described herein, various techniques and solutions can be applied for testing synchronization between encoded audio and encoded video streams. For example, encoded video markers can be inserted into one or more encoded video streams and
20 encoded audio markers can be inserted into one or more encoded audio streams. Encoded video markers and encoded audio markers can be inserted at corresponding locations in audio and video streams (e.g. sync locations or sync points), such as locations with corresponding timestamps (e.g., with the same or nearly the same timestamp).

[026] There are a number of existing solutions for testing audio-video synchronization
25 that add detectable content to uncompressed audio and video, encode the audio and video, and then detect synchronization errors. However, such existing solutions suffer from a number of limitations. For example, access to the raw uncompressed audio and video is required, or encoded audio and video may need to be decoded and re-encoded with the inserted content (e.g., which can lose some or all of the original properties of the encoded
30 audio and video). In addition, such existing solutions may be unable to isolate potential causes of synchronization errors (e.g., between encoding operations and playback operations).

[027] In the techniques and solutions described herein, encoded audio-video markers can be inserted into existing encoded audio-video streams. Synchronization analysis and

testing can be performed using the encoded audio-video streams with the encoded markers. For example, synchronization testing can be performed with various software and/or hardware playback systems. In this way, the end-to-end playback pipeline can be tested for synchronization issues.

5 [028] In some implementations, encoded audio and video markers are inserted into encoded audio and video streams without changing the original duration or length of the encoded audio and video streams. For example, for video streams, existing video frames in encoded video streams can be reduced in duration and the marker frames can be inserted. For audio, existing audio frames can be replaced with marker frames.

10 [029] In some implementations, encoded audio and video streams are analyzed to determine encoding properties. The encoding properties can be used when encoding the audio and video markers for insertion into the encoded streams to maintain compatibility and proper playback of the encoded streams.

Example 2 - Encoded Audio-Video Content

15 [030] In the technologies described herein, markers can be inserted into encoded audio-video content to test audio-video synchronization during playback. Encoded audio-video content comprises one or more video streams encoded according to one or more video codecs (e.g., according to one or more video coding standards), and one or more audio streams encoded according to one or more audio codecs (e.g., according to one or more
20 audio coding standards). For example, an encoded video stream can be encoded according to MPEG-1/MPEG-2 coding standard, the SMPTE VC-1 coding standard, the H.264/AVC coding standard, the emerging H.265/HEVC coding standard, or according to another video coding standard. An encoded audio stream can be encoded according to the AAC coding standard, the MP3, MPEG-1, and MPEG-2 coding standard, or according to
25 another audio coding standard.

[031] Encoded audio-video content can be received from a variety of sources. For example, encoded audio-video content can be obtained from a file, such as from a file storing encoded audio and video streams in a digital container format. Encoded audio-video content can be received from a network streaming source, from a capture device
30 (e.g., video and audio from a camera and microphone, which is then encoded), or from another source.

[032] Encoded audio-video content can be received in a digital container format. The digital container format can group one or more encoded video streams and one or more encoded audio streams. The digital container format can also comprise meta-data (e.g.,

describing the different video and audio streams). Examples of digital container formats include MP4 (defined by the MPEG-4 standard), AVI (defined by Microsoft®), MKV (the open standard Matroska Multimedia Container format), MPEG-2 Transport Stream/Program Stream, and ASF (advanced streaming file format).

5 **Example 3 - Audio-Video Markers**

[033] In the technologies described herein, encoded video markers can be inserted into encoded video streams and encoded audio markers can be inserted into encoded audio streams. The encoded audio and video markers can be used to test audio-video synchronization when the encoded streams are played back (e.g., tested during playback on a variety of computing devices using a variety of playback software and/or hardware).

10

[034] A video marker can be any type of marker that can later be recognized during playback (e.g., that contains video content that can later be recognized). For example, a video marker can include specific picture content (e.g., all black content, all white content, a particular pattern, etc.). A video marker can also include content that represents information, such as a frame number, a synchronization location number, a timestamp, and/or other types of information. The content of a video marker can be different from the content of a video stream into which the marker will be inserted.

15

[035] A video marker can comprise one or more pictures (e.g., frames and/or fields) of video content. In a specific implementation, a single frame with black content is used as a video marker.

20

[036] An audio marker can be any type of marker that can later be recognized during playback (e.g., that contains audio content that can later be recognized). For example, an audio marker can include an audible tone or chirp that can be detected later during playback. An audio marker can also include content that conveys information, such as a series of tones or frequencies, each indicating a different marker frame identifier (e.g., so that multiple audio markers in an audio frame can be distinguished from one another). The content of an audio marker can be selected so that it can be distinguished from (e.g., different than) other audio content of the audio stream.

25

[037] An audio marker can comprise one or more frames of audio content. In a specific implementation, a sequence of two audio frames with an audible tone or chirp is used as an audio marker.

30

Example 4 - Encoding Audio-Video Markers

[038] In the technologies described herein, audio and video markers can be encoded and inserted into encoded audio and video streams. The encoded audio and video streams

with the inserted encoded markers can be used to test audio-video synchronization when the encoded streams are played back (e.g., tested during playback on a variety of computing devices using a variety of playback software and/or hardware).

5 [039] In some implementations, audio and/or video markers are encoded according to encoding parameters of encoded audio and/or video streams. For example, an encoded video stream can be analyzed to determine video encoding parameters. Video encoding parameters can include information indicating the video codec and corresponding video standard (e.g., VC-1, H.264, H.265, H.263, MPEG-1, MPEG-2, etc.) used to encode the video stream and/or other parameters used in the encoding process (e.g., bitrate, 10 resolution, progressive or interlaced options, frame rate, aspect ratio, etc.).

[040] Once video encoding parameters have been determined, a video marker (e.g., a single black frame) can be encoded using some or all of the determined video encoding parameters. Encoding the video marker based on the determined video encoding parameters can be used to create an encoded video marker that is compatible with the 15 encoded video stream (e.g., that can be inserted into the encoded video stream without causing decoding or playback errors).

[041] Similarly, an encoded audio stream can be analyzed to determine audio encoding parameters. Audio encoding parameters can include information indicating the audio codec and corresponding audio standard (e.g., AC3, E-AC3, AAC, MP3, WMA, etc.) used 20 to encode the audio stream and/or other parameters used in the audio encoding process (e.g., bitrate, channel information, sample rate, etc.).

[042] Once audio encoding parameters have been determined, an audio marker (e.g., a sequence of audio frames) can be encoded using some or all of the determined audio encoding parameters. Encoding the audio marker based on the determined audio encoding 25 parameters can be used to create an encoded audio marker that is compatible with the encoded audio stream (e.g., that can be inserted into the encoded audio stream without causing decoding or playback errors).

[043] In some implementations, audio and/or video markers are encoded based on encoding parameters determined from encoded audio and/or video streams. For example, 30 encoded audio and/or video streams can be analyzed to determine audio and/or video encoding parameters, and based on the determined audio and/or video encoding parameters, audio and/or video markers can be encoded and inserted. In other implementations, pre-encoded audio and/or video markers are selected according to determined audio and/or video encoding parameters from analyzed encoded audio and/or

video streams. For example, a collection of pre-encoded audio and/or video markers can be maintained for use with encoded audio and/or video streams using common encoding parameters.

Example 5 - Inserting Encoded Audio-Video Markers

5 [044] In the technologies described herein, encoded audio and video markers can be inserted into encoded audio and video streams. The encoded audio and video streams with the inserted encoded markers can be used to test audio-video synchronization when the encoded streams are played back. For example, the end-to-end playback path can be tested.

10 [045] Encoded video markers can be inserted into encoded video streams. For example, encoded video markers can be inserted as new pictures (e.g., frames, fields, and/or slices), or the encoded video markers can be inserted by replacing existing pictures.

[046] Encoded video markers can be inserted without having to decode or encode the encoded video stream (e.g., without having to re-encode the video stream with the inserted
15 encoded video markers). For example, encoded video markers can be inserted as key frames (e.g., as I-frame or intra-coded frame). The encoded video markers can be inserted at particular locations in the encoded video stream, such as immediately before existing key frames (e.g., I-frames) in the encoded video stream. Such particular locations can be identified in the encoded video stream as sync locations (e.g., by only scanning the
20 compressed bitstream, or by parsing the index information present in some container formats). For example, a sequence of sync locations can be identified in a video stream by identifying existing key frames occurring approximately every few seconds in the video stream.

[047] Encoded video markers can be inserted as one or more key frames (e.g., I-
25 frames) that do not have any dependent frames. Encoded video markers can be inserted with their own sequence parameter header and picture parameter header (e.g., as part of the meta-data of the encoded video frame).

[048] Encoded audio markers can be inserted into encoded audio streams. For example, encoded audio markers can be inserted as new audio frames, or the encoded audio markers
30 can be inserted by replacing existing audio frames.

[049] Encoded audio markers can be inserted without having to decode or encode the encoded audio stream (e.g., without having to re-encode the audio stream with the inserted encoded audio markers). For example, encoded audio markers can be inserted as new audio frames between existing audio frames or by replacing existing audio frames.

[050] By not having to decode or encode (e.g., re-encode or transcode) the encoded video and audio streams, encoded markers can be efficiently inserted. For example, encoded audio and video streams can be received from a variety of sources (e.g., files, network streams, live capture and encoding, etc.) and encoded marker frames can be
5 inserted.

[051] In addition, inserting encoded markers into encoded audio and video streams can provide for testing playback systems (e.g., various computing devices, various types of software and/or hardware, etc.). For example, inserting encoded markers into encoded streams can allow isolated testing of the playback path (e.g., the end-to-end playback path)
10 without being affected by encoding processes (e.g., compared to inserting uncompressed markers into uncompressed audio-video content and then encoding the audio-video content with the inserted markers). In addition, inserting encoded markers into encoded audio and video streams can be used for testing audio-video synchronization when access to the original encoder(s) that were used to encode the audio-video content is not available
15 (e.g., which could make it difficult to decode the encoded audio and video streams to insert uncompressed markers).

[052] Encoded audio and video markers can be inserted at particular locations (e.g., sync locations) in the encoded audio and video streams. A sync location can be determined to be a location in an encoded audio stream and an encoded video stream with the same
20 timestamp (e.g., at the same time position as indicated by a timestamp or time code) or nearly the same timestamp (e.g., the closest time position, such as within a few milliseconds). In some implementations, a sync location is determined by locating a key frame in a video stream (e.g., a key frame at a specific timestamp or time code). A corresponding location in an encoded audio stream is then determined (e.g., an audio
25 frame at the same timestamp, or the closest timestamp, as the timestamp of the key frame in the video stream).

[053] Encoded audio and video markers can be inserted at a number of sync locations. For example, a number of sync locations can be selected according to an interval (e.g., a user-selected or system defined interval, such as a number of seconds or minutes). For
30 example, corresponding audio and video markers can be inserted approximately every 10 seconds into the encoded audio and video streams.

[054] Fig. 1 is an example block diagram 100 depicting insertion of encoded audio-video markers into encoded audio-video streams. In the example diagram 100, encoded audio-video content is input 110. For example, the encoded audio-video content can be

input via a file, a network connection, a live encoded stream, or via another source of encoded audio-video content.

[055] Encoded video markers 125 are inserted into one or more encoded video streams 120. For example, the encoded video markers 125 can be inserted at one or more sync
5 locations (e.g., on a periodic basis, such as every minute).

[056] Encoded audio markers 135 are inserted into one or more encoded audio streams 130. For example, the encoded audio markers 135 can be inserted at one or more sync locations corresponding to the inserted encoded video markers 125.

[057] The encoded video markers 125 and encoded audio markers 135 can be inserted
10 at the same time (e.g., at the same, or nearly the same, timestamp location in the encoded video streams 120 and the encoded audio streams 130).

[058] Once the encoded video markers and encoded audio markers have been inserted, the encoded video streams 120 and encoded audio streams 130 are output 140 (e.g., as test audio and video streams). For example, the encoded video streams 120 with the inserted
15 encoded video markers 125 and the encoded audio streams 130 with the inserted encoded audio markers 135 can be output as one or more files (e.g., in a digital container format), as streaming audio-video content, directly to a decoder for audio-video playback and testing, to a remote device such as a television for playback and testing, etc. The output encoded audio-video content with the inserted markers can be played back to test audio-
20 video synchronization (e.g., end-to-end synchronization testing of the playback path).

[059] Fig. 2 is an example block diagram 200 depicting insertion of encoded audio-video markers into encoded audio-video streams, including de-multiplexing and multiplexing audio-video content. In the example diagram 200, encoded audio-video content is received at 210. For example, the encoded audio-video content can be received
25 via a file, a network connection, a live encoded stream, or via another source of encoded audio-video content.

[060] The received encoded audio-video content is de-multiplexed 220 to separate the one or more encoded video streams 230 and the one or more encoded audio streams 240. In some implementations, encoded audio-video content can comprise multiple video
30 and/or multiple audio streams. In some implementations, only one encoded video stream and only one encoded audio stream are needed for inserting the encoded markers (e.g., only the encoded video stream and the encoded audio stream that will be used for playback and testing) even if multiple audio and/or video streams are present.

[061] In some implementations, the received encoded audio-video content 210 may not need to be de-multiplexed 220. For example, the encoded audio-video content may be received as separate streams and therefore not require de-multiplexing in order to separate the audio from the video streams.

5 [062] Encoded video markers 235 are then inserted into one or more encoded video streams 230. For example, the encoded video markers 235 can be inserted at one or more sync locations (e.g., on a periodic basis, such as every 10 seconds or every minute). Encoded audio markers 245 are inserted into one or more encoded audio streams 240. For example, the encoded audio markers 245 can be inserted at one or more sync locations
10 corresponding to the inserted encoded video markers 235 .

[063] Once the encoded video markers and encoded audio markers have been inserted, the encoded video stream 230 and the encoded audio streams 240 are multiplexed 250 (re-multiplexed) to create encoded audio-video content with the inserted markers. For example, the encoded streams can be multiplexed 250 to create audio-video content in a
15 digital container format, such as an AVI format file or stream.

[064] The multiplexed audio-video content is then output 260. For example, the multiplexed audio-video content with the inserted markers can be saved to a file, streamed via a network connection, provided for playback (e.g., via local or remote audio and video components).

20 **Example 6 - Methods for Inserting Encoded Markers**

[065] In any of the examples herein, methods can be provided for inserting encoded audio and video markers into encoded audio and video streams. The markers can be inserted without changing the overall duration (length) of the encoded audio and video streams. The markers can be inserted without having to decode or encode (or re-encode)
25 the encoded audio and video streams.

[066] Fig. 3 is a flowchart of an example method 300 for inserting encoded markers into encoded audio-video content. The example method 300 can be performed, at least in part, by a computing device.

[067] At 310, encoded audio-video content comprising an encoded video stream and an
30 encoded audio stream is received. For example, the encoded audio-video content can be received from a file, from a network connection (e.g., as streaming encoded audio-video content), or from another source. The encoded audio-video content can be de-multiplexed to separate the encoded audio stream and the encoded video stream. Alternatively, the encoded audio-video content can be received as separate encoded streams.

[068] At 320, an encoded video marker is inserted into the encoded video stream. The encoded video marker can be inserted at a video sync location (e.g., at an existing video key frame located at a particular video timestamp). The encoded video marker can be inserted without decoding or encoding the encoded video stream. The encoded video marker can be inserted without affecting the overall duration of the encoded video stream (e.g., while maintaining the same duration or length of the encoded video stream before and after the insertion).

[069] At 330, an encoded audio marker is inserted into the encoded audio stream. The encoded audio marker can be inserted at an audio sync location (e.g., at an existing audio frame, or set of frames, located at a particular audio timestamp) corresponding to the video sync location (e.g., at the same timestamp location, or the closest audio frame to the timestamp location). The encoded audio marker can be inserted without decoding or encoding the encoded audio stream. The encoded audio marker can be inserted without affecting the overall duration of the encoded audio stream (e.g., while maintaining the same duration of the encoded audio stream before and after the insertion).

[070] At 340, the encoded video stream and the encoded audio stream with the inserted markers is output. For example, the encoded streams can be output to a file (e.g., in a digital container format), to a network connection, for playback on audio-video components (e.g., a display and speakers), etc.

[071] The example method 300 can be used to insert encoded video markers and encoded audio markers into multiple encoded audio streams and/or multiple encoded video streams. In addition, the example method 300 can be used to insert corresponding encoded video and audio markers at multiple sync locations (e.g., on a periodic basis, such as every 10 seconds or every minute within the encoded audio and video streams).

[072] Fig. 4 is a flowchart of an example method 400 for inserting encoded video markers while maintaining the same overall duration. The example method 400 can be performed, at least in part, by a computing device.

[073] At 410, a key video frame (e.g., an I-frame or intra-coded frame) is selected in an encoded video stream. The key frame can be selected based on various criteria. For example, the key frame can be selected based on a frequency with which markers are to be inserted (e.g., every 10 seconds, every 5 minutes, etc.). The key frame can also be selected based on a comparison of timing information between the encoded video stream and an associated encoded audio stream. For example, a key video frame can be selected that has a timestamp corresponding to a timestamp of an audio frame in the encoded audio stream

(e.g., where the key video frame and the audio frame have the same timestamp, or nearly the same timestamp, such as within a few milliseconds of each other).

[074] At 420, the duration of the key video frame is reduced resulting in an unused duration. For example, for 30 FPS (frames-per-second) video content, each frame is
5 displayed for 1730th of a second (approximately 33 milliseconds). If the key video frame selected at 410 is encoded within 30 FPS video content, then the duration of the key video frame can be reduced in duration by half, to 1/60th of a second (approximately 16
10 milliseconds). After the reduction, there will be an unused duration that was previously used by the key video frame. In this example, the unused duration will be 1/60th of a second (approximately 16 milliseconds).

[075] In some implementations, the duration of the existing key video frame is reduced at 420 by half (e.g., from 1/30th of a second to 1/60th of a second). Alternatively, the duration of the existing key video frame can be reduced by more or less than one-half.

[076] At 430, an encoded video marker frame is inserted into the encoded video stream
15 using the unused duration. For example, if the key video frame is reduced from 1730th of a second in duration to 1/60th of a second in duration, then the inserted encoded video marker frame can use the unused 1760th of a second.

[077] In some implementations, the example method 400 is performed, in part, by updating meta-data associated with the encoded video stream, such as a meta-data table
20 indicating timing information (e.g., a timestamp table or index table). Such meta-data can specify the timing of the video pictures (e.g., video frames and/or video fields). By modifying the meta-data, the existing key video frame duration can be set to the reduced duration (e.g., at 420) and the inserted encoded video marker can be set to use the unused duration (e.g., at 430).

[078] In some implementations, the reduction in duration of the existing key video
25 frame at 420, and thus the unused duration, is determined so that the inserted encoded marker frame will be displayed when the encoded stream is played back. For example, if a display (e.g., a built-in mobile device display, external computer display, or another type of display) displays video content at 60Hz, then a video frame may need to be at least
30 1760th of a second in duration in order to be displayed. In this situation, the existing key video frame can be reduced in duration to leave at least 1760th of a second in unused duration for the inserted encoded video frame. Depending on the display rate (e.g., 30Hz, 60Hz, 120Hz, etc.) of the display on which the encoded stream will be played back, the duration of the inserted encoded video frame may need to be adjusted, and in some

situations (e.g., where the display rate is less than the video frame rate) multiple encoded video frames may need to be inserted in order to ensure that the marker is displayed.

[079] Fig. 5 is a flowchart of an example method 500 for creating encoded audio-video markers based on audio-video encoding parameters. The example method 500 can be
5 performed, at least in part, by a computing device.

[080] At 510, encoded audio-video content comprising an encoded video stream and an encoded audio stream is received. For example, the encoded audio-video content can be received from a file, from a network connection (e.g., as streaming encoded audio-video content), or from another source. The encoded audio-video content can be de-multiplexed
10 to separate the encoded audio stream and the encoded video stream. Alternatively, the encoded audio-video content can be received as separate encoded streams.

[081] At 520, the encoded video stream received with the encoded audio-video content 510 is analyzed to determine video encoding parameters. The video encoding parameters indicate how the video stream was encoded (e.g., the codec and associated video coding
15 standard used, resolution, frame rate, and/or other codec-specific encoding parameters and options).

[082] At 530, a video marker (e.g., comprising one or more video frames and/or fields) is encoded based, at least in part, on the determined video encoding parameters 520 to create an encoded video marker. For example, all, or most, of the determined video
20 encoding parameters 520 can be used to create the encoded video marker. Using the determined video encoding parameters 520, the video marker can be encoded in a manner that is compatible with the encoded video stream (e.g., that will display properly when the encoded video stream is played back).

[083] At 540, the encoded video marker that was created at 530 is inserted into the
25 encoded video stream. The encoded video marker can be inserted at a video sync location (e.g., a particular video timestamp). The encoded video marker can be inserted without decoding or encoding the encoded video stream. The encoded video marker can be inserted without affecting the overall duration of the encoded video stream (e.g., while maintaining the same duration of the encoded video stream before and after the insertion).

[084] At 550, the encoded audio stream received with the encoded audio-video content 510 is analyzed to determine audio encoding parameters. The audio encoding parameters indicate how the audio stream was encoded (e.g., the codec and associated audio coding
30 standard used, bit rate, sampling rate, channel information, and/or other codec-specific encoding parameters and options).

[085] At 560, an audio marker (e.g., comprising one or more audio frames) is encoded based, at least in part, on the determined audio encoding parameters 550 to create an encoded audio marker. For example, all, or most, of the determined audio encoding parameters 550 can be used to create the encoded audio marker. Using the determined audio encoding parameters 550, the audio marker can be encoded in a manner that is compatible with the encoded audio stream (e.g., that plays properly when the encoded audio stream is played back).

[086] At 570, the encoded audio marker that was created at 560 is inserted into the encoded audio stream. The encoded audio marker can be inserted at an audio sync location (e.g., a particular audio timestamp) corresponding to the video sync location (e.g., at the same, or nearly the same, timestamp location in both the encoded video stream and the encoded audio stream). The encoded audio marker can be inserted without decoding or encoding the encoded audio stream. The encoded audio marker can be inserted without affecting the overall duration of the encoded audio stream (e.g., while maintaining the same duration of the encoded audio stream before and after the insertion).

[087] At 580, the encoded video stream and the encoded audio stream with the inserted markers is output. For example, the encoded streams can be output to a file (e.g., in a digital container format), to a network connection, for playback on audio-video components (e.g., a display and speakers), etc.

Example 7 - Example Implementations for Inserting Encoded Markers

[088] Fig. 6 depicts a prior art diagram of an example video stream 610 and corresponding video timestamp table 620. The example video stream 610 is encoded at 30 FPS. Therefore, each frame has a duration of $1/30^{\text{th}}$ of a second (approximately 33 milliseconds).

[089] The video stream 610 depicts a number of video frames. Specifically, eight video frames are depicted. For example, Frame 1 could be a key frame (e.g., an I-frame), frames 2-7 could be predicted frames (e.g., P-frames) that are predicted from Frame 1, and Frame 8 could be another key frame.

[090] The video timestamp table 620 indicates the time at which each frame is displayed, which also indicates the duration of display for each frame. As depicted in the video timestamp table 620, Frame 1 is displayed at 0 milliseconds (ms), Frame 2 is displayed at 33 ms, Frame 3 is displayed at 66 ms, and so on. Each of the frames depicted in the video stream 610 is displayed for a duration of approximately 33 ms.

[091] Fig. 7 is a diagram depicting an example encoded video stream 710 and corresponding video timestamp table showing an inserted encoded video marker frame. Fig. 7 illustrates an example implementation where a marker video frame is inserted into the encoded video stream 710 without changing the overall duration.

5 [092] The encoded video stream depicted at 710 comprises eight video frames. In order to insert an encoded video marker into the encoded video stream 710 while maintaining the same overall duration of the encoded video stream 710, the duration of Frame 1 (730) has been reduced. Specifically, in this example, the duration of Frame 1 (730) has been reduced by half, from $1/30^{\text{th}}$ of a second to $1/60^{\text{th}}$ of a second (from approximately 33 ms to approximately 16 ms). The reduction in duration is depicted graphically in Fig. 7 as
10 Frame 1 (730) now occupies the right-hand portion (to the right of the dashed line) of the original first frame duration.

[093] Using the unused duration, an encoded video marker frame 720 is inserted into the encoded video stream 710. In this example, the duration of the encoded video marker
15 frame 720 is inserted using the remaining $1/60^{\text{th}}$ of a second left unused by the reduction in duration of existing Frame 1 (730). The inserted encoded video marker frame 720 is depicted graphically in Fig. 7 as occupying the left-hand portion (to the left of the dashed line) of the original first frame duration.

[094] In some implementations, inserting an encoded video marker involves modifying
20 a timestamp table (or other meta-data associated with the encoded video stream) in order to specify display times, durations, and/or other timing information. In Fig. 7, an original timestamp table 740 is depicted. As indicated by the original timestamp table 740, the encoded video stream 710 is encoded at a rate of 30 FPS, so each frame is $1/30^{\text{th}}$ of a second in duration (approximately 33 ms). When Frame 1 (730) is reduced in duration and
25 the encoded video marker frame 720 is inserted, the original timestamp table 740 is modified as depicted in the modified timestamp table 750. As indicated by the modified timestamp table 750, the encoded video marker frame 720 is displayed first (at 0 ms in this example) for a duration of $1/60^{\text{th}}$ of a second (approximately 16 ms), followed by Frame 1 (730), which is displayed next at 16 ms for a duration of $1/60^{\text{th}}$ of a second, followed by
30 Frame 2, which is displayed next at 33 ms for a duration of $1/30^{\text{th}}$ of a second, and so on. In this manner, the encoded video marker frame 720 and Frame 1 (730) take up the duration previously occupied by Frame 1 (730), and the overall duration of the encoded video stream 710 remains the same. Furthermore, as depicted in the modified timestamp table 750, only the timing information for the marker frame and the frame being reduced

in duration need to be modified; the timing information for remaining frames in the video stream remain unchanged.

[095] Fig. 8 is a diagram showing example video and audio streams with inserted video and audio markers at sync locations. Fig. 8 illustrates an example implementation where
5 marker frames are inserted at sync locations into encoded video and audio streams without changing the overall duration of the streams.

[096] In Fig. 8, an example encoded video stream 810 is depicted. The example encoded video stream 810 is encoded at a rate of 30 FPS, and each video frame has an original duration of $1/30^{\text{th}}$ of a second. An example encoded audio stream 820 is also
10 depicted. The example encoded audio stream 820 is encoded using audio frames with a duration of 10 ms each.

[097] In order to insert audio-video markers into the encoded video stream 810 and the encoded audio stream 820, sync locations are determined at corresponding locations in the two streams. Specifically, in this example, two sync locations have been determined, sync
15 location 830 and sync location 840. For example, if both the encoded video stream 810 and the encoded audio stream 820 start at time 0 ms, then the first sync location 830 will be at the same timestamp location (0 ms) in both streams, and the second sync location 840 will also be at the same timestamp location (200 ms) in both streams.

[098] After the sync locations have been determined, encoded markers are inserted.
20 Specifically, in this example, an encoded video marker frame 832 and an encoded audio marker frame 834 have been inserted at the first sync location 830 in both the encoded video stream 810 and the encoded audio stream 820. The encoded video marker frame 832 has been inserted using unused duration resulting from reduction in duration of existing video frame 1 (836). The encoded audio marker frame 834 has been inserted by replacing
25 the original audio frame at the sync location 830. Another encoded video marker frame 842 and corresponding audio marker frame 844 have been inserted at the second sync location 840 in both the encoded video stream 810 and the encoded audio stream 820. Once again, the encoded video marker frame 842 has been inserted using unused duration resulting from reduction in duration of existing video frame 7 (846). The encoded audio
30 marker frame 844 has been inserted by replacing the original audio frame at the sync location 840.

[099] In some implementations, the encoded audio marker uses a duration at least as long as the encoded video marker. For example, using the example depicted in Fig. 8, if the video marker frame 832 is $1/30^{\text{th}}$ of a second in duration (approximately 16 ms), then

the corresponding encoded audio marker can take up two audio frames (20 ms total), instead of the single audio frame as depicted at 834.

Example 8 - Testing Audio-Video Synchronization

[0100] In the examples herein, techniques are provided for testing audio-video
5 synchronization using encoded audio and video markers inserted into encoded audio and
video streams. Audio-video synchronization testing can be performed by obtaining
encoded audio-video content and inserting encoded markers without having to decode or
encode the encoded audio-video content. The encoded audio-video content with the
inserted markers can be played back and the decoded audio-video content can be captured.
10 For example, the decoded audio-video content can be captured with reduced quality (e.g.,
reduced resolution for video and reduced channels/quality for audio), which can reduce
capture overhead and optimize capture performance (e.g., reduce capture latency).
Corresponding markers can be detected and matched in the captured decoded audio-video
content and audio-video synchronization information can be output.

15 [0101] Fig. 9 is a flowchart of an example method 900 for testing synchronization of
encoded audio-video content with inserted encoded audio-video markers. The example
method 900 can be performed, at least in part, by a computing device.

[0102] At 910, encoded audio-video content is received with inserted encoded audio-
video markers. The encoded audio-video content comprises an encoded video stream with
20 one or more inserted encoded video markers and an encoded audio stream with one or
more corresponding inserted encoded audio markers.

[0103] At 920, playback of the encoded audio-video content is initiated. For example,
playback can be initiated by providing the encoded audio-video content to operating
system components (e.g., media player components), or other software and/or hardware,
25 of a computing device.

[0104] At 930, decoded video content is captured during playback. The decoded video
content can be captured as it would be displayed on a display (e.g., computer monitor or
integrated mobile device display). For example, the decoded video content can be captured
via a software application programming interface (API) that provides access to decoded
30 video content prior to, or contemporaneous with, display (e.g., that captures decoded video
at the video endpoint). For example, a software-based solution can capture decoded video
content using screen scraping. In a specific implementation, the decoded video content is
captured as uncompressed YUV raw video. The decoded video content can also be
captured from the display by a separate device, such as via an external video camera or an

HDMI capture device. The decoded video content can be captured with associated video playback timing information (e.g., high precision timing information) indicating the time at which pictures are displayed (e.g., just for video marker content or for additional video content as well.).

5 [0105] In some implementations, the decoded video content that is captured during playback 930 is captured at a reduced resolution. Capturing the decoded video content at a reduced resolution can be more efficient (e.g., in terms of latency and computing resources) than capturing the decoded video content at the original resolution. Encoded video markers can be recognized even if captured at a reduced resolution. For example, a
10 video marker that is a black frame (e.g., in a video stream that does not contain a black frame) can be recognized at full resolution or at a reduced resolution.

[0106] At 940, decoded audio content is captured during playback. The decoded audio content can be captured as it would be played on a speaker (e.g., a built-in speaker, external speakers, headphones, etc.). For example, the decoded audio content can be
15 captured via a software application programming interface (API) that provides access to decoded audio content prior to, or contemporaneous with, playback (e.g., that captures decoded audio at the audio endpoint). For example, decoded audio can be captured using a software-based solution that uses a loop back capture feature available from the operating system. In a specific implementation, the decoded audio content is captured as
20 uncompressed PCM raw audio. The decoded audio content can also be captured by a separate device, such as via an external microphone. The decoded audio content can be captured with associated audio playback timing information (e.g., high precision timing information) indicating the time at which audio content is played (e.g., just for audio marker content or for additional audio content as well).

25 [0107] In some implementations, the decoded audio content that is captured during playback 940 is captured with a reduced number of audio channels and/or with reduced quality (e.g., with a reduced bit depth and/or sampling rate). For example, decoded audio content with 2-channel stereo audio can be captured as a single channel (e.g., by selecting one of the two channels for capture) and/or at a reduced bit depth (e.g., 8-bit audio) or
30 sampling rate (e.g., 22 kHz). Capturing the decoded audio content with reduced channels (e.g., a single channel) and/or with reduced quality can be more efficient (e.g., in terms of latency and computing resources) than capturing the decoded audio content with higher quality (e.g., corresponding to the quality of the encoded audio stream). Encoded audio markers can be recognized even if captured with reduced channels and/or reduced quality.

For example, a specific audio tone marker that is present in all channels (e.g., in an audio stream that does not otherwise contain such a tone) can be recognized even if only one channel is captured and/or if capture quality is reduced.

5 [0108] In a specific implementation, playback, decoding, and capture (e.g., at 920, 930, and 940) is performed using, at least in part, Microsoft® Media Foundation APIs.

[0109] At 950, matching audio and video markers are detected in the captured decoded video content (from 930) and the captured decoded audio content (from 940). Matching corresponding pairs of audio and video markers can be performed, for example, by matching the first pair of audio and video markers, the second pair of audio and video markers, and so on. Corresponding audio and video markers can also be matched by examining the content of the marker. For example, video markers can comprise an identifier (e.g., a sequence number or timestamp) which can be matched to a corresponding audio marker that comprises specific audio content (e.g., a specific tone, sequence of tones, or other recognizable audio pattern) that has been pre-determined to correspond to the video marker identifier.

15 [0110] In some implementations, the matching 950 is performed in real-time or near real-time as the decoded video and audio is captured (at 930 and 940). In other implementations, the decoded video and audio is captured (at 930 and 940) and saved for later analysis (e.g., for matching, as described with regard to 950).

20 [0111] At 960, audio-video synchronization information is output based on the matching performed at 950. For example, the synchronization information can comprise indications of differences between corresponding audio and video markers detected in the decoded audio-video content (e.g., based on associated audio-video playback timing information). For example, if an encoded video marker and corresponding encoded audio marker were inserted at the same timestamp (e.g., at 5 minutes, 10 seconds, 100 ms timestamp in both the encoded video stream and the encoded audio stream), then the difference detected between playback of the markers can be output (e.g., if the video marker is played back beginning at time t_1 and the audio marker is played back beginning at time $t_1 + 115$ ms, then information indicating that audio synchronization is off by 115 ms at the location of the video and audio markers can be output). Audio-video synchronization information can also indicate whether synchronization difference between corresponding audio and video markers is within a threshold value (e.g., within a pre-set or user-configured threshold, such as within 20 ms) or not.

[0112] In a specific implementation, corresponding audio and video markers present in captured decoded audio and video streams (e.g., as described at 930 and 940) are matched and their presentation times (display times) are recorded. Audio-video synchronization issues can then be reported. For example, synchronization issues can be reported if the presentation times for the corresponding markers are not within a threshold value (e.g., a pre-determined or user-configured threshold that indicates an allowable gap or offset in milliseconds).

[0113] In the specific implementation, the auto-correlation of an audio marker is computed using the following equation (Equation 1):

10 [0114]
$$CorrM = \sum_{k=0}^{N-1} (M[k] \times M[k])$$

(Equation 1)

In Equation 1, N is the length of the audio marker, M[k]. Then cross-correlation between the audio marker M[k] and the captured audio stream T[i + k] at position i in the captured sequence is computed using the following equation (Equation 2):

15 [0115]
$$CorrTi = \sum_{k=0}^{N-1} (T[i + k] \times M[k])$$

(Equation 2)

[0116] If $CorrM \times \alpha > CorrTi > CorrM \times \beta$, where α and β can be chosen as 1.1 and 0.9, respectively, in the specific implementation, the audio marker is detected at sample position i in the captured audio stream.

20 [0117] In the specific implementation, the video marker is detected using the following equation (Equation 3):

[0118]
$$\gamma_{c1} < P_c(x,y) < \gamma_{c2}$$

(Equation 3)

[0119] Using Equation 3, the video marker is detected for any position (x, y) in the captured frame, where $P_c(x, y)$ is the pixel value at the spatial position of (x, y) inside the luma (Y) component or the chroma (U or V) component of the frame. In the specific implementation γ_{c1} and γ_{c2} are set to 0 and 16, respectively, for Y, and 120 and 136, respectively, for U and V.

[0120] In other implementations, other detection techniques can be employed to detect audio and video markers in captured decoded audio and video content.

Example 9 - Computing Systems

[0121] FIG. 10 depicts a generalized example of a suitable computing system 1000 in which the described innovations may be implemented. The computing system 1000 is not

intended to suggest any limitation as to scope of use or functionality, as the innovations may be implemented in diverse general-purpose or special-purpose computing systems.

[0122] With reference to FIG. 10, the computing system 1000 includes one or more processing units 1010, 1015 and memory 1020, 1025. In FIG. 10, this basic configuration 5 1030 is included within a dashed line. The processing units 1010, 1015 execute computer-executable instructions. A processing unit can be a general-purpose central processing unit (CPU), processor in an application-specific integrated circuit (ASIC), or any other type of processor. In a multi-processing system, multiple processing units execute computer-executable instructions to increase processing power. For example, FIG. 10 shows a 10 central processing unit 1010 as well as a graphics processing unit or co-processing unit 1015. The tangible memory 1020, 1025 may be volatile memory (e.g., registers, cache, RAM), non-volatile memory (e.g., ROM, EEPROM, flash memory, etc.), or some combination of the two, accessible by the processing unit(s). The memory 1020, 1025 stores software 1080 implementing one or more innovations described herein, in the form 15 of computer-executable instructions suitable for execution by the processing unit(s).

[0123] A computing system may have additional features. For example, the computing system 1000 includes storage 1040, one or more input devices 1050, one or more output devices 1060, and one or more communication connections 1070. An interconnection 20 mechanism (not shown) such as a bus, controller, or network interconnects the components of the computing system 1000. Typically, operating system software (not shown) provides an operating environment for other software executing in the computing system 1000, and coordinates activities of the components of the computing system 1000.

[0124] The tangible storage 1040 may be removable or non-removable, and includes magnetic disks, magnetic tapes or cassettes, CD-ROMs, DVDs, or any other medium 25 which can be used to store information and which can be accessed within the computing system 1000. The storage 1040 stores instructions for the software 1080 implementing one or more innovations described herein.

[0125] The input device(s) 1050 may be a touch input device such as a keyboard, mouse, pen, or trackball, a voice input device, a scanning device, or another device that 30 provides input to the computing system 1000. For video encoding, the input device(s) 1050 may be a camera, video card, TV tuner card, or similar device that accepts video input in analog or digital form, or a CD-ROM or CD-RW that reads video samples into the computing system 1000. The output device(s) 1060 may be a display, printer, speaker, CD-writer, or another device that provides output from the computing system 1000.

[0126] The communication connection(s) 1070 enable communication over a communication medium to another computing entity. The communication medium conveys information such as computer-executable instructions, audio or video input or output, or other data in a modulated data signal. A modulated data signal is a signal that
5 has one or more of its characteristics set or changed in such a manner as to encode information in the signal. By way of example, and not limitation, communication media can use an electrical, optical, RF, or other carrier.

[0127] The innovations can be described in the general context of computer-executable instructions, such as those included in program modules, being executed in a computing
10 system on a target real or virtual processor. Generally, program modules include routines, programs, libraries, objects, classes, components, data structures, etc. that perform particular tasks or implement particular abstract data types. The functionality of the program modules may be combined or split between program modules as desired in various embodiments. Computer-executable instructions for program modules may be
15 executed within a local or distributed computing system.

[0128] The terms "system" and "device" are used interchangeably herein. Unless the context clearly indicates otherwise, neither term implies any limitation on a type of computing system or computing device. In general, a computing system or computing device can be local or distributed, and can include any combination of special-purpose
20 hardware and/or general-purpose hardware with software implementing the functionality described herein.

[0129] For the sake of presentation, the detailed description uses terms like "determine" and "use" to describe computer operations in a computing system. These terms are high-level abstractions for operations performed by a computer, and should not be confused
25 with acts performed by a human being. The actual computer operations corresponding to these terms vary depending on implementation.

Example 10 - Mobile Device

[0130] FIG. 11 is a system diagram depicting an exemplary mobile device 1100 including a variety of optional hardware and software components, shown generally at
30 1102. Any components 1102 in the mobile device can communicate with any other component, although not all connections are shown, for ease of illustration. The mobile device can be any of a variety of computing devices (e.g., cell phone, smartphone, handheld computer, Personal Digital Assistant (PDA), etc.) and can allow wireless two-

way communications with one or more mobile communications networks 1104, such as a cellular, satellite, or other network.

[0131] The illustrated mobile device 1100 can include a controller or processor 1110 (e.g., signal processor, microprocessor, ASIC, or other control and processing logic
5 circuitry) for performing such tasks as signal coding, data processing, input/output processing, power control, and/or other functions. An operating system 1112 can control the allocation and usage of the components 1102 and support for one or more application programs 1114. The application programs can include common mobile computing applications (e.g., email applications, calendars, contact managers, web browsers,
10 messaging applications), or any other computing application. Functionality 1113 for accessing an application store can also be used for acquiring and updating application programs 1114.

[0132] The illustrated mobile device 1100 can include memory 1120. Memory 1120 can include non-removable memory 1122 and/or removable memory 1124. The non-
15 removable memory 1122 can include RAM, ROM, flash memory, a hard disk, or other well-known memory storage technologies. The removable memory 1124 can include flash memory or a Subscriber Identity Module (SIM) card, which is well known in GSM communication systems, or other well-known memory storage technologies, such as "smart cards." The memory 1120 can be used for storing data and/or code for running the
20 operating system 1112 and the applications 1114. Example data can include web pages, text, images, sound files, video data, or other data sets to be sent to and/or received from one or more network servers or other devices via one or more wired or wireless networks. The memory 1120 can be used to store a subscriber identifier, such as an International Mobile Subscriber Identity (IMSI), and an equipment identifier, such as an International
25 Mobile Equipment Identifier (IMEI). Such identifiers can be transmitted to a network server to identify users and equipment.

[0133] The mobile device 1100 can support one or more input devices 1130, such as a touchscreen 1132, microphone 1134, camera 1136, physical keyboard 1138 and/or
30 trackball 1140 and one or more output devices 1150, such as a speaker 1152 and a display 1154. Other possible output devices (not shown) can include piezoelectric or other haptic output devices. Some devices can serve more than one input/output function. For example, touchscreen 1132 and display 1154 can be combined in a single input/output device.

[0134] The input devices 1130 can include a Natural User Interface (NUI). An NUI is any interface technology that enables a user to interact with a device in a "natural"

manner, free from artificial constraints imposed by input devices such as mice, keyboards, remote controls, and the like. Examples of NUI methods include those relying on speech recognition, touch and stylus recognition, gesture recognition both on screen and adjacent to the screen, air gestures, head and eye tracking, voice and speech, vision, touch, gestures, and machine intelligence. Other examples of a NUI include motion gesture detection using accelerometers/gyroscopes, facial recognition, 3D displays, head, eye, and gaze tracking, immersive augmented reality and virtual reality systems, all of which provide a more natural interface, as well as technologies for sensing brain activity using electric field sensing electrodes (EEG and related methods). Thus, in one specific example, the operating system 1112 or applications 1114 can comprise speech-recognition software as part of a voice user interface that allows a user to operate the device 1100 via voice commands. Further, the device 1100 can comprise input devices and software that allows for user interaction via a user's spatial gestures, such as detecting and interpreting gestures to provide input to a gaming application.

[0135] A wireless modem 1160 can be coupled to an antenna (not shown) and can support two-way communications between the processor 1110 and external devices, as is well understood in the art. The modem 1160 is shown generically and can include a cellular modem for communicating with the mobile communication network 1104 and/or other radio-based modems (e.g., Bluetooth 1164 or Wi-Fi 1162). The wireless modem 1160 is typically configured for communication with one or more cellular networks, such as a GSM network for data and voice communications within a single cellular network, between cellular networks, or between the mobile device and a public switched telephone network (PSTN).

[0136] The mobile device can further include at least one input/output port 1180, a power supply 1182, a satellite navigation system receiver 1184, such as a Global Positioning System (GPS) receiver, an accelerometer 1186, and/or a physical connector 1190, which can be a USB port, IEEE 1394 (FireWire) port, and/or RS-232 port. The illustrated components 1102 are not required or all-inclusive, as any components can be deleted and other components can be added.

Example 11 - Cloud-Supported Environment

[0137] Fig. 12 illustrates a generalized example of a suitable implementation environment 1200 in which described embodiments, techniques, and technologies may be implemented. In the example environment 1200, various types of services (e.g., computing services) are provided by a cloud 1210. For example, the cloud 1210 can comprise a

collection of computing devices, which may be located centrally or distributed, that provide cloud-based services to various types of users and devices connected via a network such as the Internet. The implementation environment 1200 can be used in different ways to accomplish computing tasks. For example, some tasks (e.g., processing user input and presenting a user interface) can be performed on local computing devices (e.g., connected devices 1230, 1240, 1250) while other tasks (e.g., storage of data to be used in subsequent processing) can be performed in the cloud 1210.

[0138] In example environment 1200, the cloud 1210 provides services for connected devices 1230, 1240, 1250 with a variety of screen capabilities. Connected device 1230 represents a device with a computer screen 1235 (e.g., a mid-size screen). For example, connected device 1230 could be a personal computer such as desktop computer, laptop, notebook, netbook, or the like. Connected device 1240 represents a device with a mobile device screen 1245 (e.g., a small size screen). For example, connected device 1240 could be a mobile phone, smart phone, personal digital assistant, tablet computer, and the like. Connected device 1250 represents a device with a large screen 1255. For example, connected device 1250 could be a television screen (e.g., a smart television) or another device connected to a television (e.g., a set-top box or gaming console) or the like. One or more of the connected devices 1230, 1240, 1250 can include touchscreen capabilities. Touchscreens can accept input in different ways. For example, capacitive touchscreens detect touch input when an object (e.g., a fingertip or stylus) distorts or interrupts an electrical current running across the surface. As another example, touchscreens can use optical sensors to detect touch input when beams from the optical sensors are interrupted. Physical contact with the surface of the screen is not necessary for input to be detected by some touchscreens. Devices without screen capabilities also can be used in example environment 1200. For example, the cloud 1210 can provide services for one or more computers (e.g., server computers) without displays.

[0139] Services can be provided by the cloud 1210 through service providers 1220, or through other providers of online services (not depicted). For example, cloud services can be customized to the screen size, display capability, and/or touchscreen capability of a particular connected device (e.g., connected devices 1230, 1240, 1250).

[0140] In example environment 1200, the cloud 1210 provides the technologies and solutions described herein to the various connected devices 1230, 1240, 1250 using, at least in part, the service providers 1220. For example, the service providers 1220 can provide a centralized solution for various cloud-based services. The service providers

1220 can manage service subscriptions for users and/or devices (e.g., for the connected devices 1230, 1240, 1250 and/or their respective users).

Example 12 - Implementations

5 [0141] Although the operations of some of the disclosed methods are described in a particular, sequential order for convenient presentation, it should be understood that this manner of description encompasses rearrangement, unless a particular ordering is required by specific language set forth below. For example, operations described sequentially may in some cases be rearranged or performed concurrently. Moreover, for the sake of simplicity, the attached figures may not show the various ways in which the disclosed
10 methods can be used in conjunction with other methods.

[0142] Any of the disclosed methods can be implemented as computer-executable instructions or a computer program product stored on one or more computer-readable storage media and executed on a computing device (e.g., any available computing device, including smart phones or other mobile devices that include computing hardware).
15 Computer-readable storage media are any available tangible media that can be accessed within a computing environment (e.g., one or more optical media discs such as DVD or CD, volatile memory components (such as DRAM or SRAM), or nonvolatile memory components (such as flash memory or hard drives)). By way of example and with reference to Fig. 10, computer-readable storage media include memory 1020 and 1025,
20 and storage 1040. By way of example and with reference to Fig. 11, computer-readable storage media include memory and storage 1120, 1122, and 1124. The term computer-readable storage media does not include signals and carrier waves. In addition, the term computer-readable storage media does not include communication connections (e.g., 1070, 1160, 1162, and 1164).

25 [0143] Any of the computer-executable instructions for implementing the disclosed techniques as well as any data created and used during implementation of the disclosed embodiments can be stored on one or more computer-readable storage media. The computer-executable instructions can be part of, for example, a dedicated software application or a software application that is accessed or downloaded via a web browser or
30 other software application (such as a remote computing application). Such software can be executed, for example, on a single local computer (e.g., any suitable commercially available computer) or in a network environment (e.g., via the Internet, a wide-area network, a local-area network, a client-server network (such as a cloud computing network), or other such network) using one or more network computers.

[0144] For clarity, only certain selected aspects of the software-based implementations are described. Other details that are well known in the art are omitted. For example, it should be understood that the disclosed technology is not limited to any specific computer language or program. For instance, the disclosed technology can be implemented by
5 software written in C++, Java, Perl, JavaScript, Adobe Flash, or any other suitable programming language. Likewise, the disclosed technology is not limited to any particular computer or type of hardware. Certain details of suitable computers and hardware are well known and need not be set forth in detail in this disclosure.

[0145] Furthermore, any of the software-based embodiments (comprising, for example, computer-executable instructions for causing a computer to perform any of the disclosed
10 methods) can be uploaded, downloaded, or remotely accessed through a suitable communication means. Such suitable communication means include, for example, the Internet, the World Wide Web, an intranet, software applications, cable (including fiber optic cable), magnetic communications, electromagnetic communications (including RF,
15 microwave, and infrared communications), electronic communications, or other such communication means.

[0146] The disclosed methods, apparatus, and systems should not be construed as limiting in any way. Instead, the present disclosure is directed toward all novel and nonobvious features and aspects of the various disclosed embodiments, alone and in
20 various combinations and sub combinations with one another. The disclosed methods, apparatus, and systems are not limited to any specific aspect or feature or combination thereof, nor do the disclosed embodiments require that any one or more specific advantages be present or problems be solved.

[0147] The technologies from any example can be combined with the technologies described in any one or more of the other examples. In view of the many possible
25 embodiments to which the principles of the disclosed technology may be applied, it should be recognized that the illustrated embodiments are examples of the disclosed technology and should not be taken as a limitation on the scope of the disclosed technology. Rather, the scope of the disclosed technology includes what is covered by the following claims.

30 We therefore claim as our invention all that comes within the scope and spirit of the claims.

CLAIMS

1. A method, implemented at least in part by a computing device, for inserting encoded markers into encoded audio-video content, the method comprising:

receiving, by the computing device, encoded audio-video content comprising an encoded video stream and an encoded audio stream;

inserting, by the computing device, an encoded video marker into the encoded video stream at a video sync location, wherein the encoded video marker is inserted without decoding or re-encoding the encoded video stream;

inserting, by the computing device, an encoded audio marker into the encoded audio stream at an audio sync location corresponding to the video sync location, wherein the encoded audio marker is inserted without decoding or re-encoding the encoded audio stream; and

outputting, by the computing device, the encoded video stream with the inserted encoded video marker and the encoded audio stream with the inserted encoded audio marker.

2. The method of claim 1 wherein:

the receiving comprises de-multiplexing the encoded audio-video content to produce the encoded video stream and the encoded audio stream; and

the outputting comprises re-multiplexing the encoded video stream with the inserted encoded video marker and the encoded audio stream with the inserted encoded audio marker.

3. The method of claim 1 further comprising:

analyzing the encoded video stream to determine video encoding parameters; and encoding a video marker using, at least in part, the determined video encoding parameters to create the encoded video marker.

4. The method of claim 1 wherein overall duration of the encoded video stream remains the same after the encoded video marker is inserted, and wherein substantially all original properties of the encoded audio stream and the encoded video stream remain the same after the encoded audio and video streams are output.

5. The method of claim 4 wherein the encoded video marker is an encoded video marker frame, wherein inserting the encoded video marker frame comprises:

selecting an existing key video frame, wherein the existing key video frame is located at the video sync location;

reducing a duration of the existing key video frame, creating an unused duration; and

inserting the encoded video marker frame using the unused duration.

6. The method of claim 5 wherein the duration of the existing key video frame is reduced by half, and wherein the encoded video marker frame is inserted as a key video frame immediately before the existing key video frame using the unused duration.

7. The method of claim 4 further comprising:

modifying a meta-data table associated with the encoded video stream according to the reduced duration of the existing key video frame and the unused duration for the inserted encoded video marker frame.

8. The method of claim 1 further comprising:

analyzing the encoded audio stream to determine audio encoding parameters; and encoding an audio marker using, at least in part, the determined audio encoding parameters to create the encoded audio marker;

wherein inserting the encoded audio marker comprises replacing an existing audio frame in the encoded audio stream with an encoded audio marker frame; and

wherein the audio sync location is a closest timestamp location to the video sync location.

9. A computing device comprising:

a processing unit; and

memory;

the computing device configured to perform operations for inserting encoded markers into encoded audio-video content, the operations comprising:

receiving encoded audio-video content comprising an encoded video stream and an encoded audio stream;

analyzing the encoded video stream to determine video encoding parameters;

encoding a video marker using, at least in part, the determined video encoding parameters to create an encoded video marker compatible with the encoded video stream;

inserting the encoded video marker into the encoded video stream at a video sync location, wherein the encoded video marker is inserted without decoding or re-encoding the encoded video stream, and wherein overall duration of the encoded video stream remains the same after the encoded video marker is inserted;

analyzing the encoded audio stream to determine audio encoding parameters;

encoding an audio marker using, at least in part, the determined audio encoding parameters to create an encoded audio marker compatible with the encoded audio stream;

inserting the encoded audio marker into the encoded audio stream at an audio sync location corresponding to the video sync location, wherein the encoded audio marker is inserted without decoding or re-encoding the encoded audio stream; and

outputting the encoded video stream with the inserted encoded video marker and the encoded audio stream with the inserted encoded audio marker.

10. A computer-readable storage medium storing computer-executable instructions for testing synchronization of encoded audio-video content, the method comprising:

receiving encoded audio-video content comprising an encoded video stream and an encoded audio stream, the encoded video stream comprising one or more video markers, and the encoded audio stream comprising one or more corresponding audio markers;

initiating playback of the encoded audio-video content;

during playback of the encoded audio-video content:

capturing decoded video content, wherein the captured video content is captured at a reduced resolution; and

capturing decoded audio content, wherein the captured audio content is captured with a reduced number of audio channels;

from the captured video content and the captured audio content, matching the one or more video markers and the one or more corresponding audio markers; and

based on the matching, outputting audio-video synchronization information comprising outputting an indication of playback timing differences between the matched one or more video markers and the one or more corresponding audio markers.

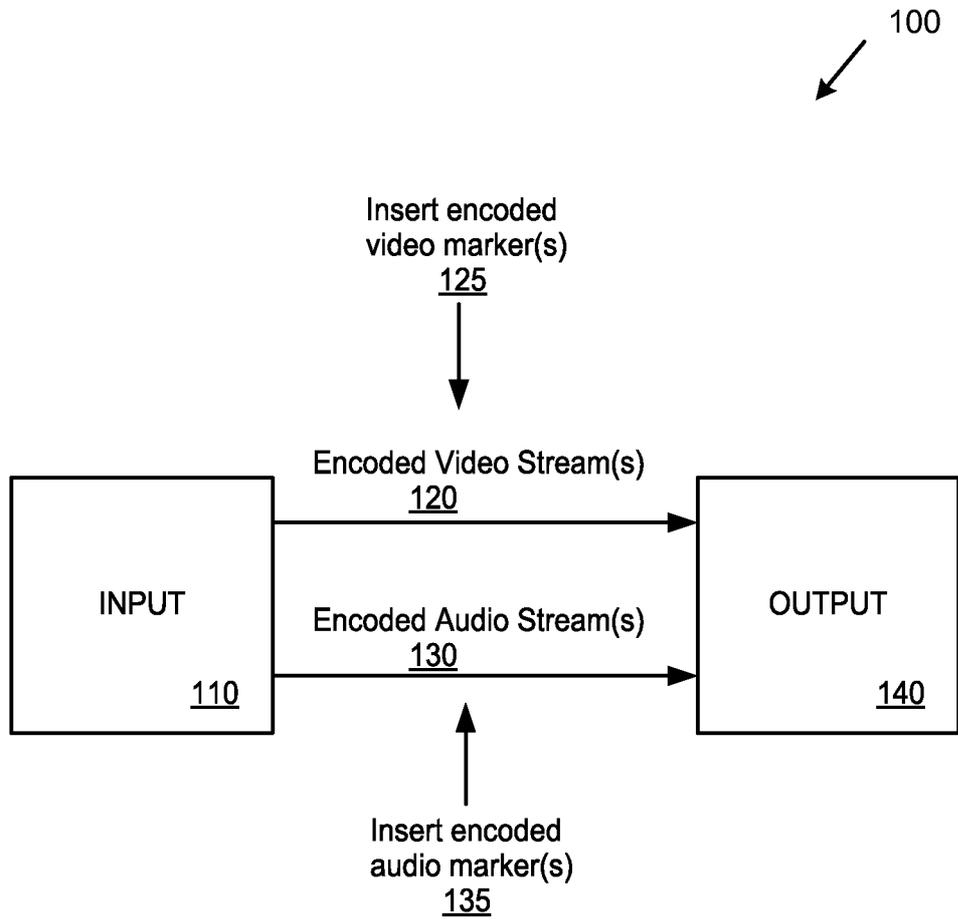


FIG. 1

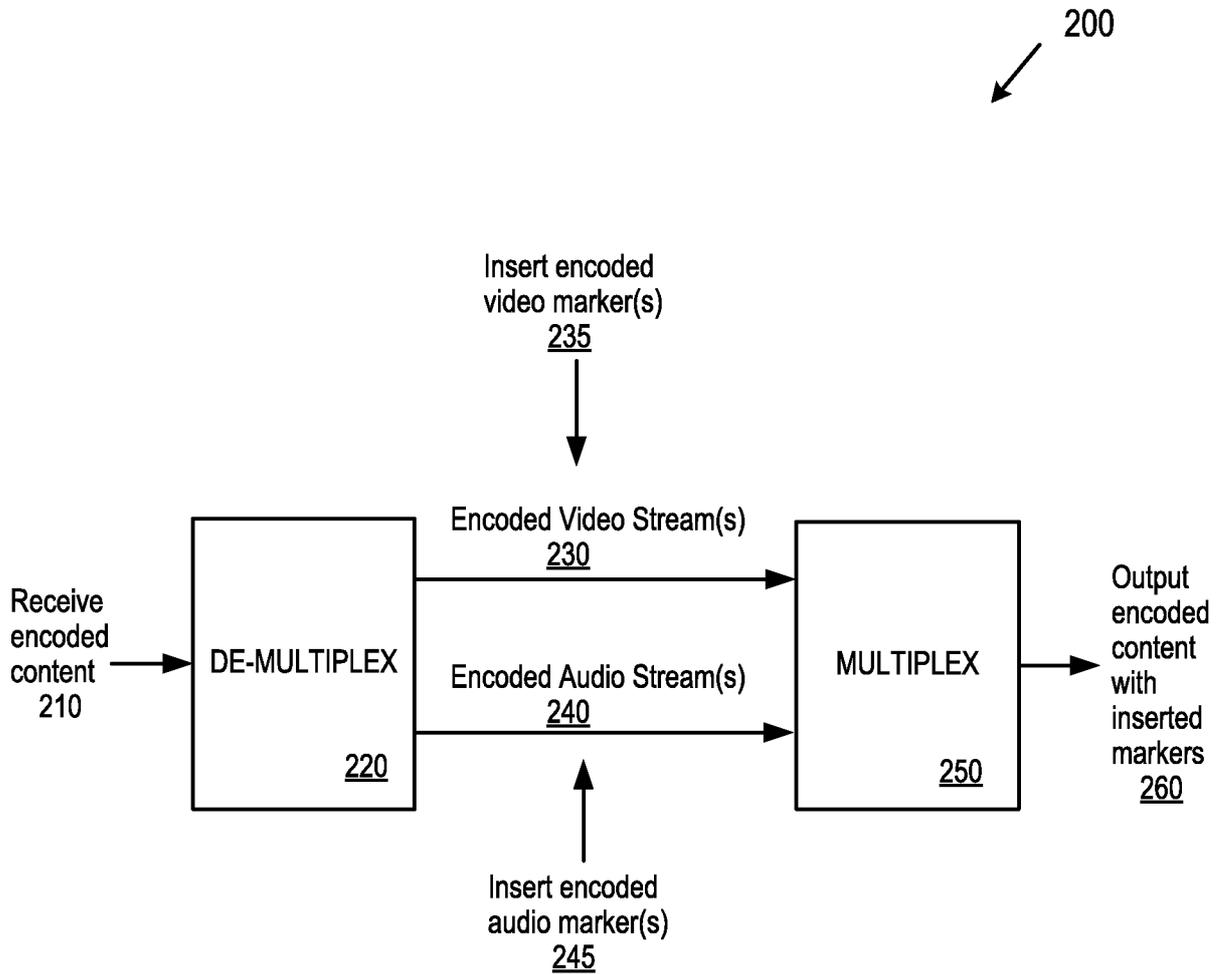


FIG. 2

300 ↙

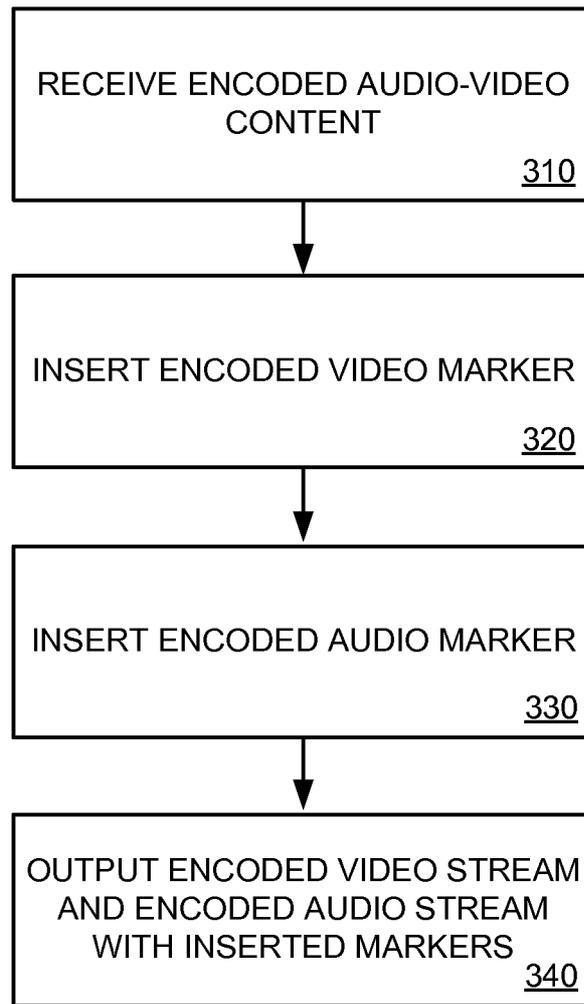


FIG. 3

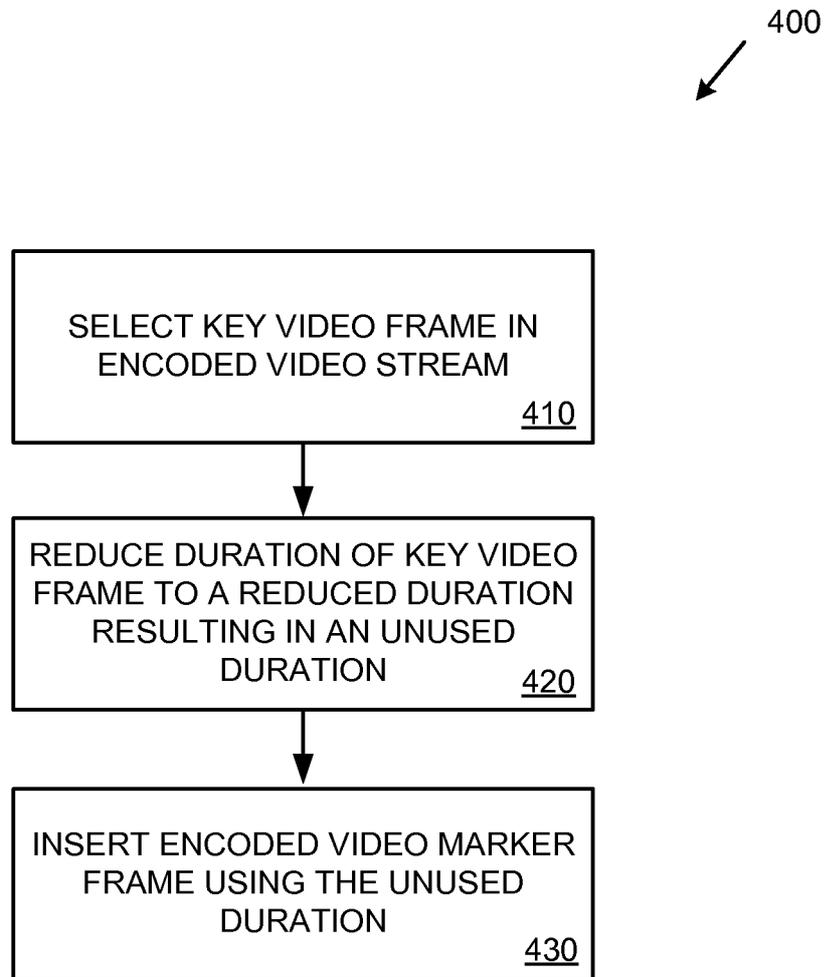


FIG. 4

5/12

500

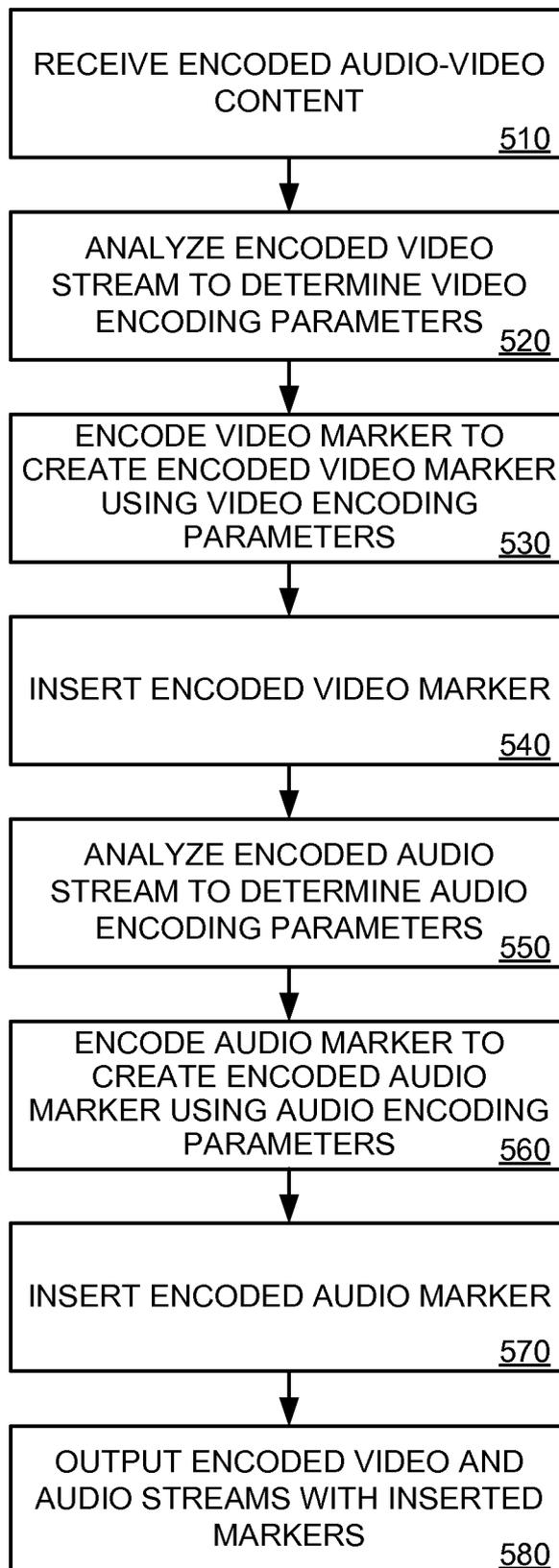


FIG. 5

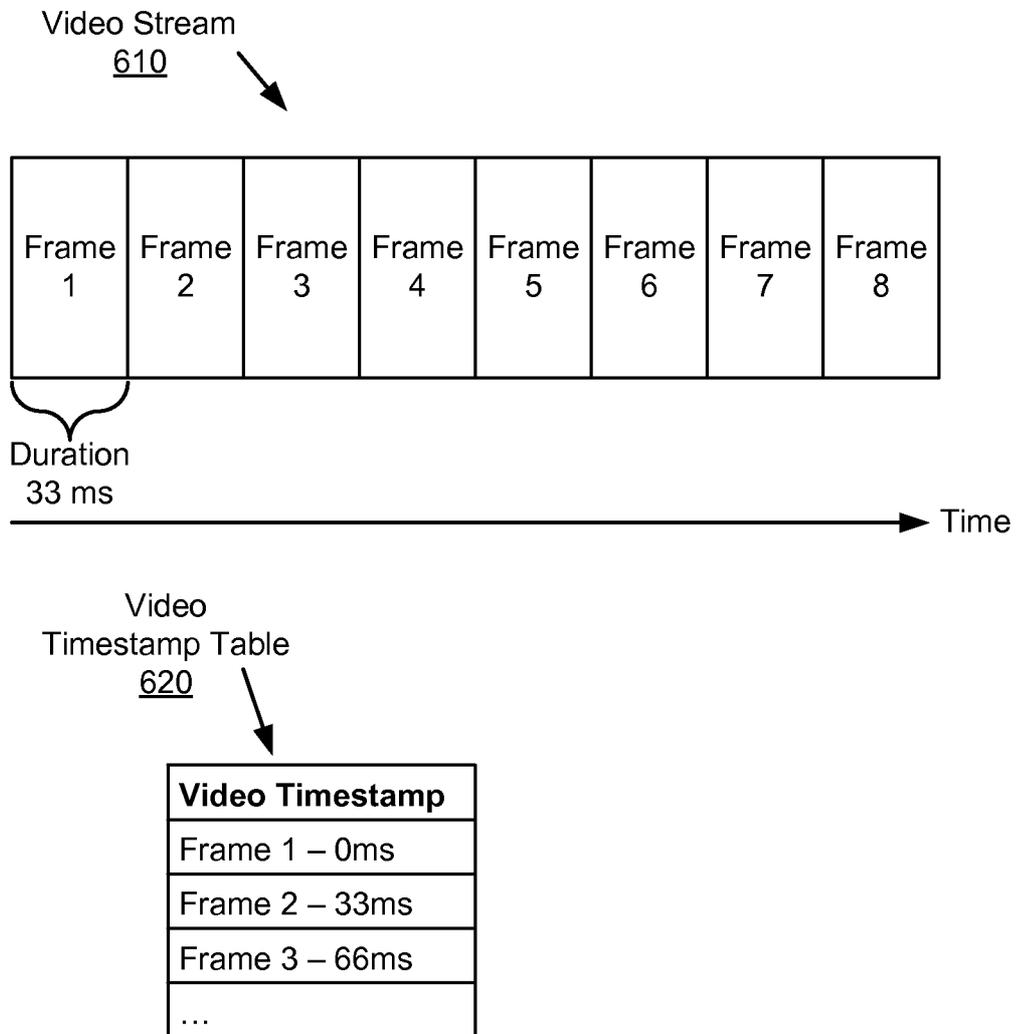


FIG. 6

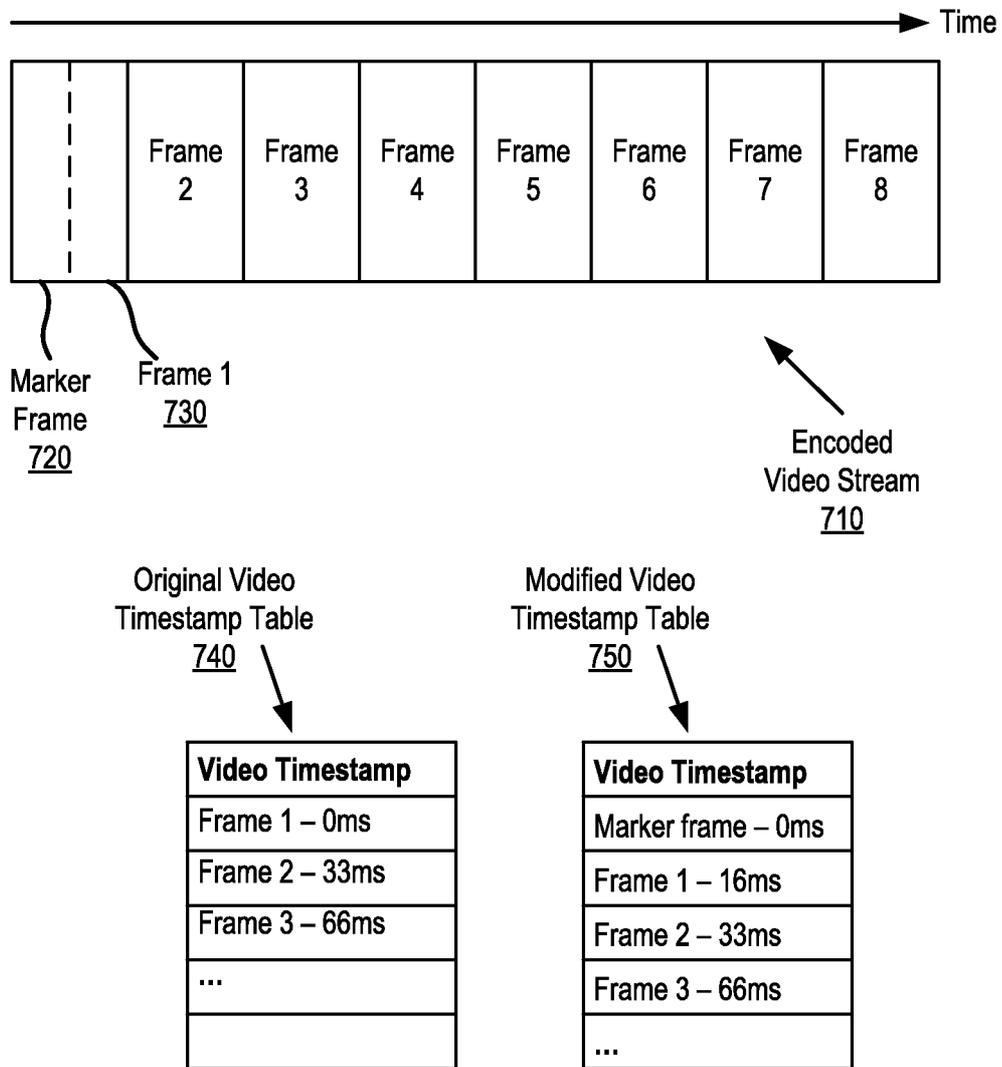


FIG. 7

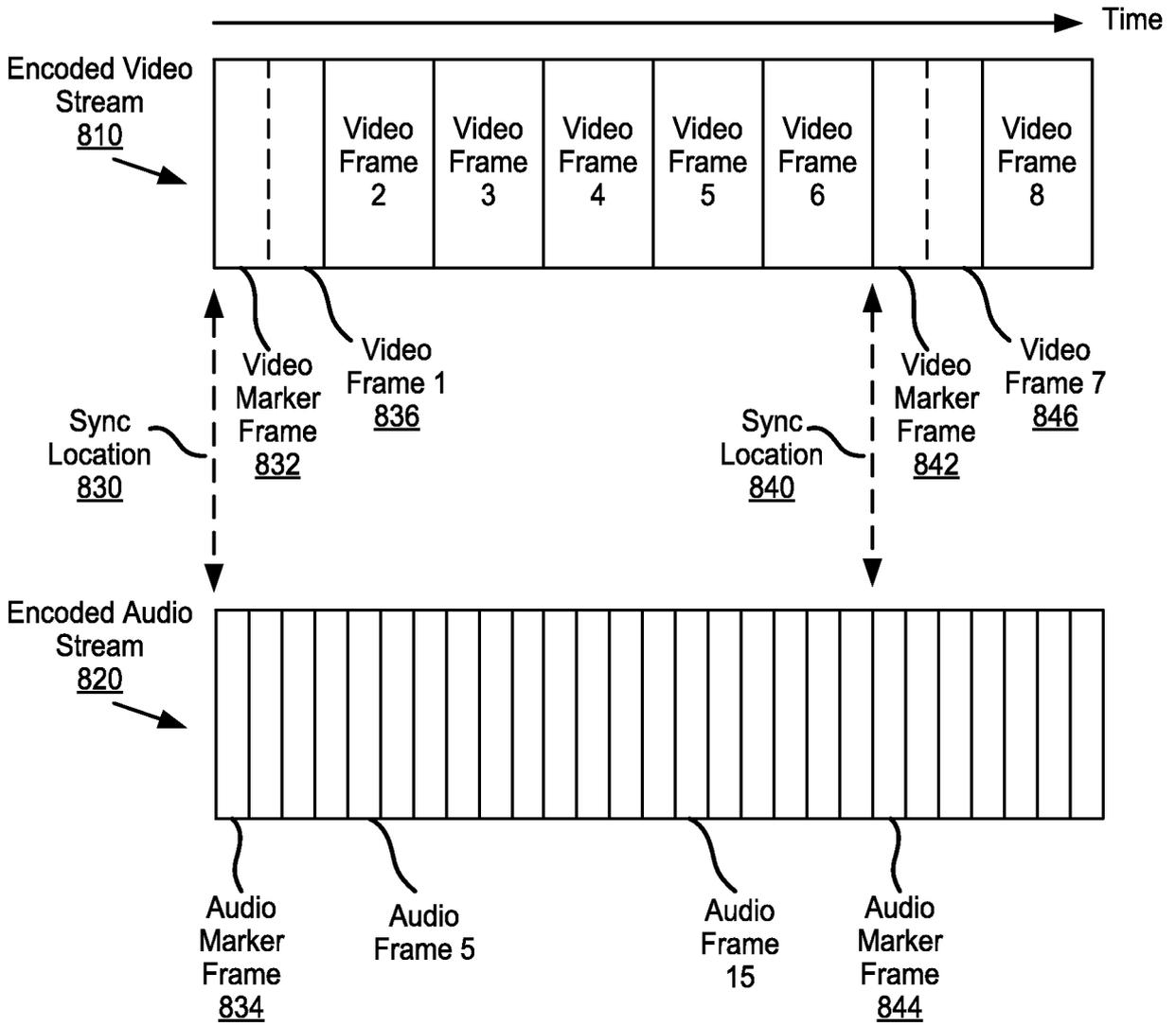
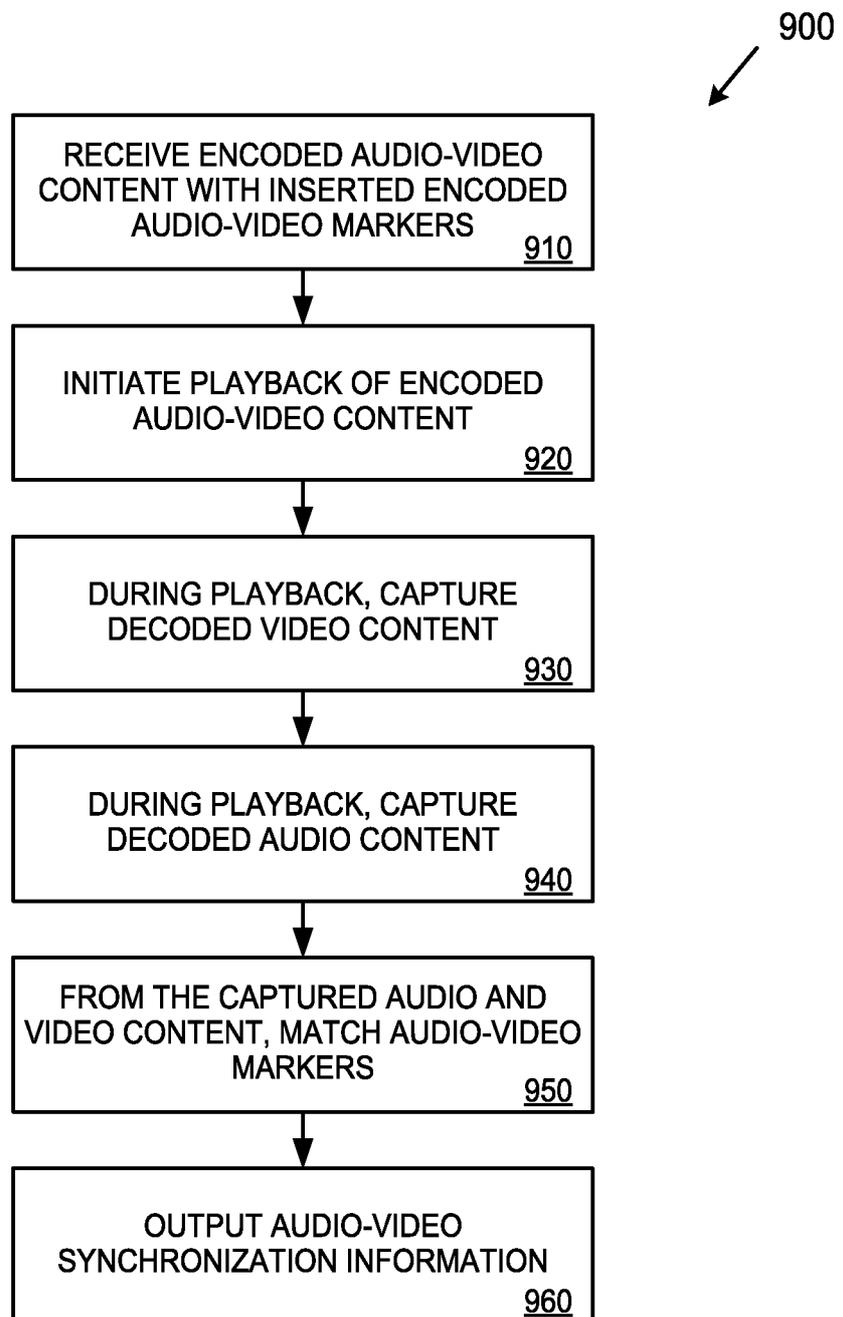


FIG. 8

9/12

**FIG. 9**

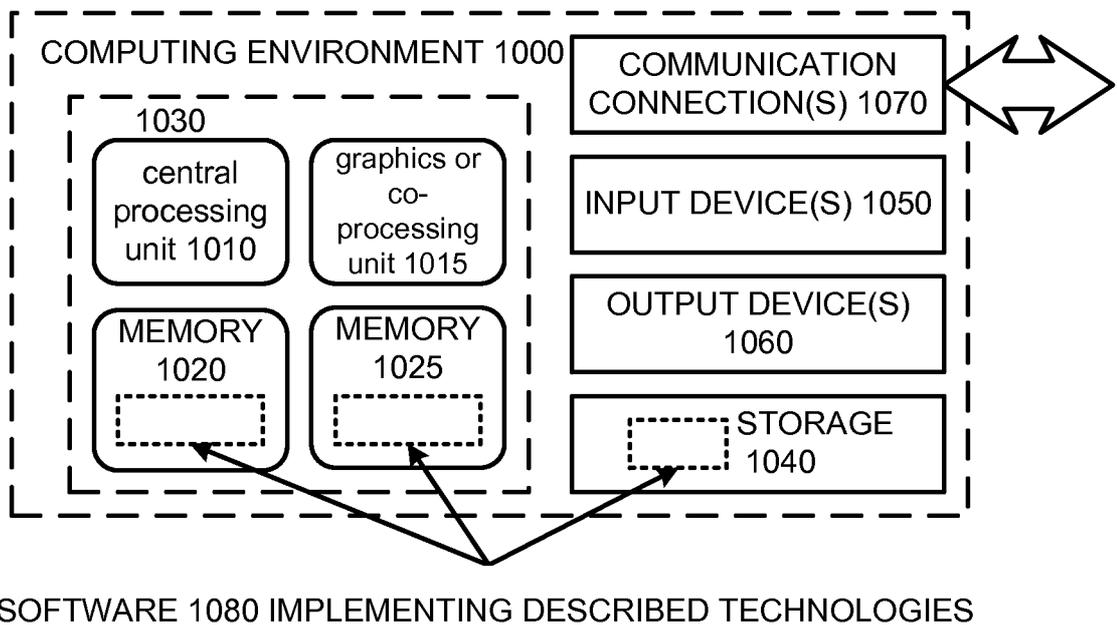


FIG. 10

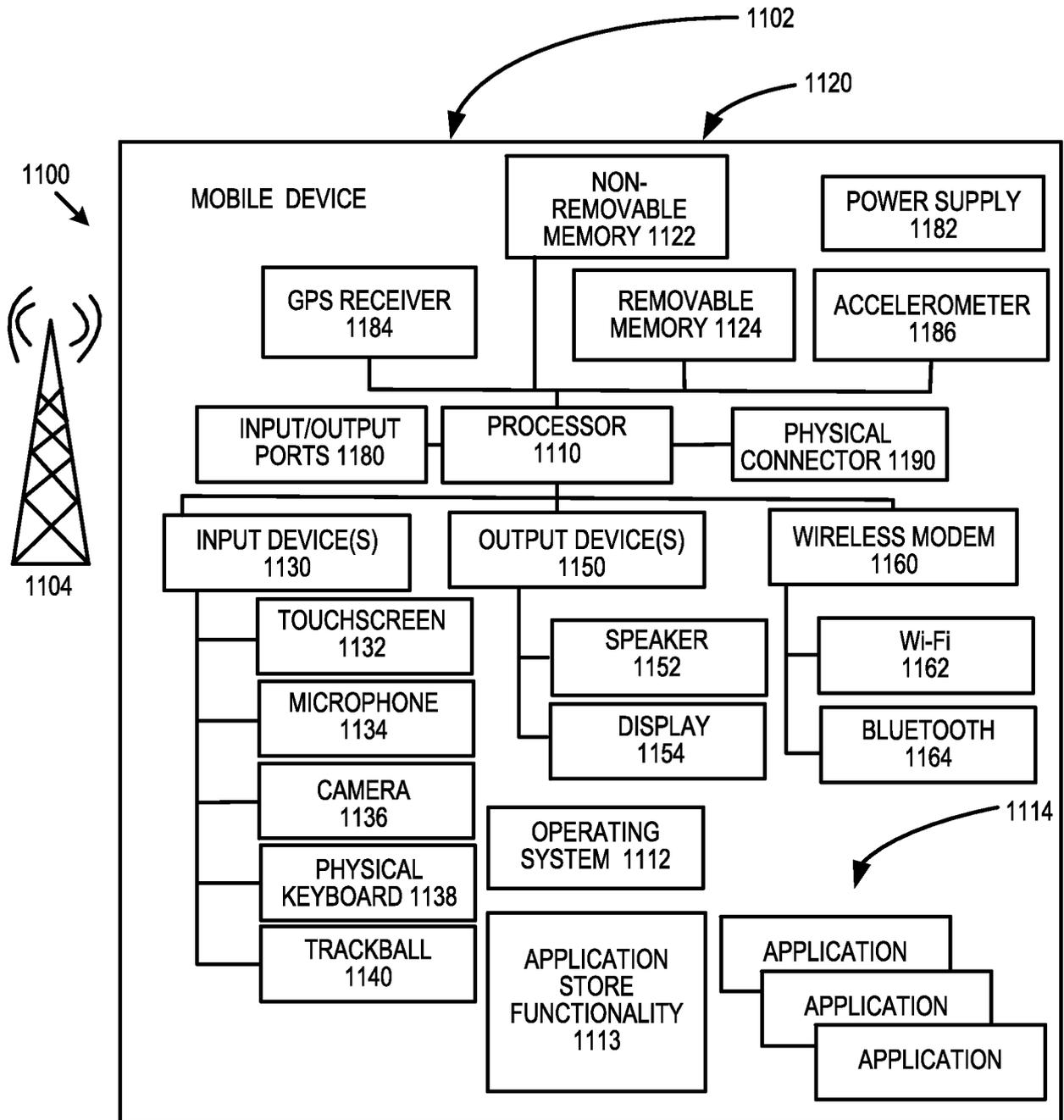


FIG. 11

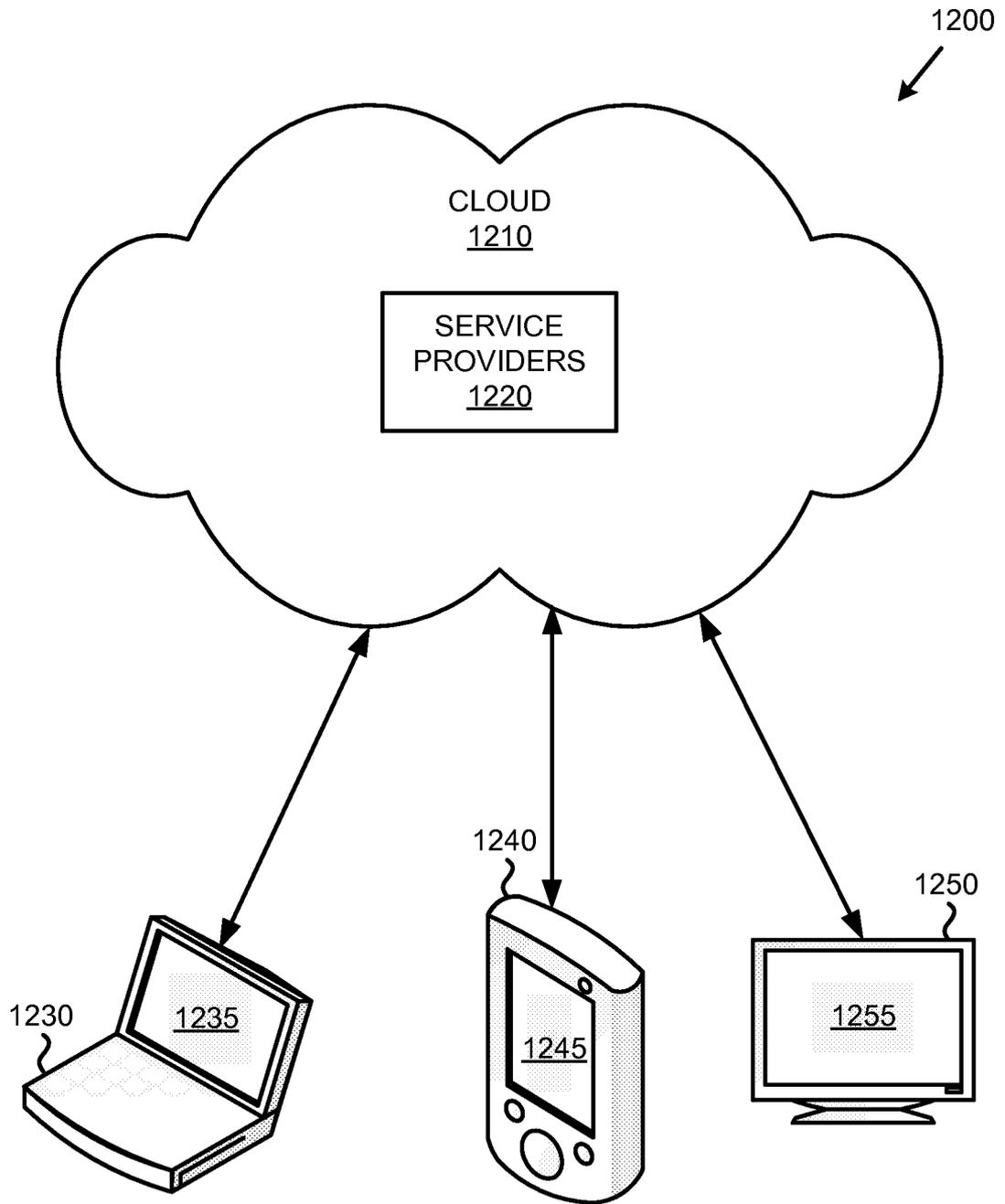


FIG. 12

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2014/053029

A. CLASSIFICATION OF SUBJECT MATTER
INV. H04N17/00 H04N21/8547 H04N21/233 H04N21/234 H04N21/2343
H04N21/2362
 ADD.
 According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED
 Minimum documentation searched (classification system followed by classification symbols)
H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal , WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CA 2 541 560 A1 (LEITCH TECHNOLOGY INTERNAT INC [CA]) 30 September 2007 (2007-09-30)	1-4,8-10
A	figures la, lb, 2 page 1, line 3 - line 5 page 2, line 7 - line 16 page 3, line 4 - line 11 page 4, line 11 -- line 26 page 5, line 13 -- line 24 page 6, line 15 -- line 18 page 7, line 4 - line 18 page 8, line 3 - page 9, line 9 page 11, line 8 -- page 12, line 3 ----- -/- .	5-7

Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"E" earlier application or patent but published on or after the international filing date	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O" document referring to an oral disclosure, use, exhibition or other means	"&" document member of the same patent family
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 31 October 2014	Date of mailing of the international search report 06/11/2014
---	---

Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Authorized officer Boubal , o i v i e r
--	---

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2014/053029

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 1 773 072 AI (AVAYA TECH LLC [US]) 11 April 2007 (2007-04-11)	1,8, 10
A	paragraph [0001] paragraph [0007] column 5, line 37 - line 42 paragraph [0023] - paragraph [0026] -----	2-7 ,9
X	US 2008/049846 AI (NAGAFUJI NOBUYUKI [JP]) 28 February 2008 (2008-02-28)	1,8, 10
A	figures 1,3 paragraph [0003] paragraph [0006] paragraph [0010] paragraph [0046] - paragraph [0047] paragraph [0055] - paragraph [0056] -----	2-7 ,9
A	US 6 480 902 BI (YUANG MARIA C [TW] ET AL) 12 November 2002 (2002-11-12) figures 1,2 column 1, line 9 - line 12 column 2, line 8 - line 10 column 2, line 58 - line 63 column 3, line 5 - line 19 -----	1-10
A	GB 2 328 099 A (BRITISH BROADCASTING CORP [GB]) 10 February 1999 (1999-02-10) figure 2 page 2, line 2 - line 10 page 3, line 10 - line 14 page 5, line 7 - line 8 page 8, line 27 - line 30 page 11, line 14 - line 15 page 12, line 9 - line 11 page 15, line 2 - page 16, line 16 -----	1-10
A	US 2007/276670 AI (PEARLSTEIN LARRY [US]) 29 November 2007 (2007-11-29) paragraph [0059] - paragraph [0077] paragraph [0084] paragraph [0106] paragraph [0110] paragraph [0156] -----	1-10

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No PCT/US2014/053029
--

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
CA 2541560	AI	30-09-2007	CA 2541560 AI 30-09-2007
			CN 101796812 A 04-08-2010
			EP 2008442 AI 31-12-2008
			US 2007245222 AI 18-10-2007
			Wo 2007112552 AI 11-10-2007

EP 1773072	AI	11-04 -2007	CA 2552769 AI 28-03-2007
			CN 1941916 A 04-04-2007
			EP 1773072 AI 11-04-2007
			JP 2007097185 A 12-04-2007
			US 2007071037 AI 29-03-2007

US 2008049846	AI	28-02 -2008	JP 4655006 B2 23-03-2011
			JP 2008053872 A 06-03-2008
			US 2008049846 AI 28-02-2008

US 6480902	BI	12-11 -2002	NONE

GB 2328099	A	10-02-1999	AU 754879 B2 28--11 -2002
			AU 8738198 A 01--03 -1999
			CA 2300389 AI 18--02 -1999
			DE 69814642 DI 18--06 -2003
			DE 69814642 T2 18--03 -2004
			EP 1002424 A2 24--05 -2000
			ES 2194338 T3 16--11 -2003
			GB 2328099 A 10--02 -1999
			JP 2001513606 A 04--09 -2001
			US 6741290 BI 25--05 -2004
			Wo 9908441 A2 18--02 -1999

US 2007276670	AI	29-11-2007	NONE
