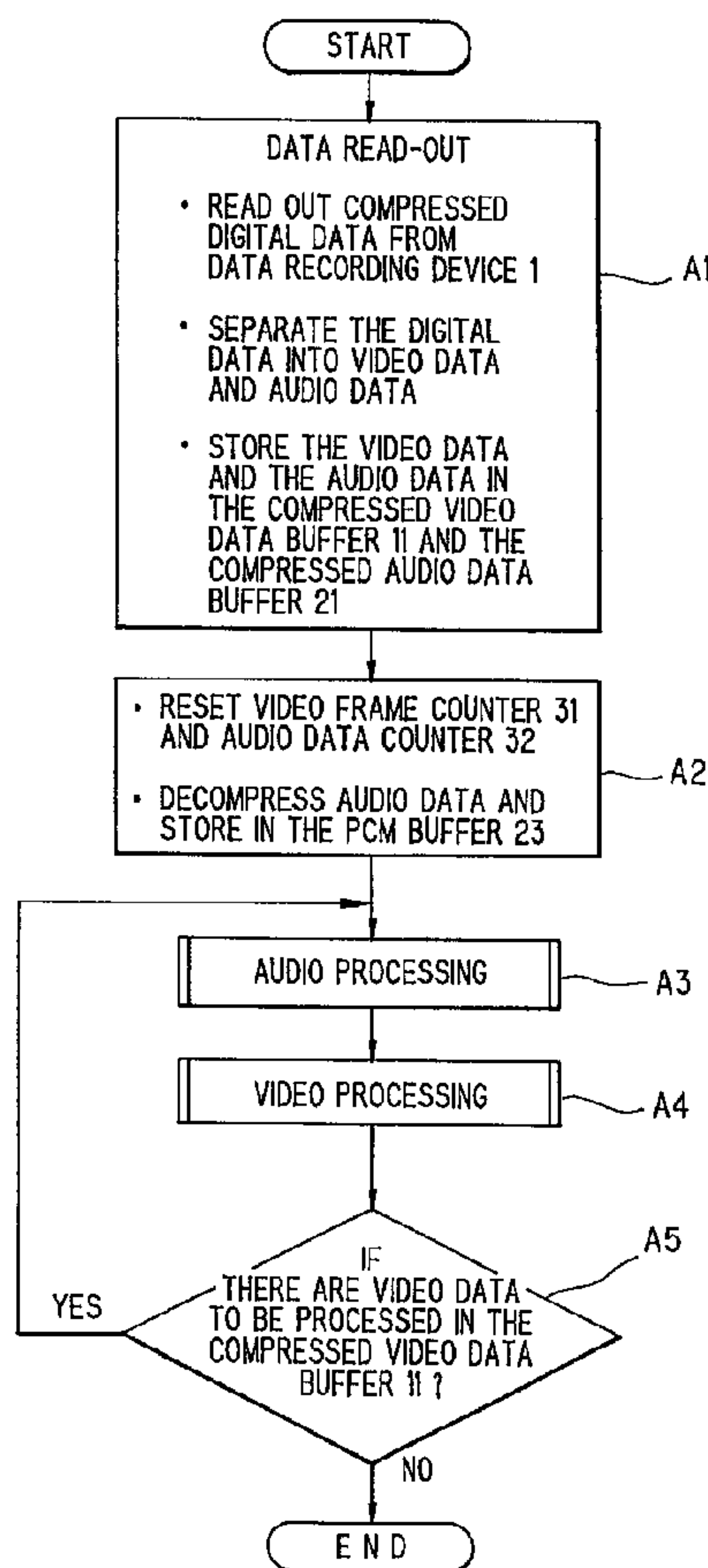




(11) (21) (C) **2,219,154**
(22) 1997/10/24
(43) 1998/04/29
(45) 2000/10/17

(72) TANAKA, Mitsumasa, JP
(72) SAMESHIMA, Takashi, JP
(72) TERASHIMA, Masaru, JP
(72) SAWADA, Hideki, JP
(73) NEC CORPORATION, JP
(51) Int.Cl.⁶ H04N 5/04, H04N 7/04
(30) 1996/10/29 (8-287075) JP

(54) **APPAREIL DE LECTURE SYNCHRONISEE DE SIGNAUX
AUDIO ET VIDEO**
(54) **AN APPARATUS FOR A SYNCHRONIZED PLAYBACK OF
AUDIO-VIDEO SIGNALS**



(57) An apparatus for a synchronized playback of compressed digital data of audio-video signals with a simple configuration to be applied to a system having comparatively low data processing capability, independent of the resolution of a software timer of the system, comprises means (32 and 33) for obtaining a reference time indicating reproduction time passage of the audio signal component from reproduced data amount of the audio signal component, and means (31 and 34) for detecting progress of the video signal component in reference to the reference time.



ABSTRACT

An apparatus for a synchronized playback of compressed digital data of audio-video signals with a simple configuration to be applied to a system having comparatively low data processing capability, independent of the resolution of a software timer of the system, comprises means (32 and 33) for obtaining a reference time indicating reproduction time passage of the audio signal component from reproduced data amount of the audio signal component, and means (31 and 34) for detecting progress of the video signal component in reference to the reference time.

AN APPARATUS FOR A SYNCHRONIZED PLAYBACK OF AUDIO-VIDEO SIGNALS

The present invention relates to an apparatus for synchronized playback of audio-video (hereafter abbreviated as AV) signals, wherein compressed digital data of AV signals are decompressed into a video signal component and an audio signal component to be represented in synchronization with each other.

In the synchronized playback of AV signals, compressed digital data of AV signals read out from recording media such as a CD-ROM (Compact Disk Read Only Memory) or a hard disk device are decompressed before being represented through output devices such as a CRT monitor and a speaker. When the apparatus for the synchronized playback is provided with data processing capability sufficiently powerful for the data decompression of the AV signals, or when the size of video signal of the compressed digital data is comparatively small, the data decompression of both of the video signal component and the audio signal component can be performed within a short time, and so, audio data and video frames can be reproduced without interruption.

However, when the data processing capability of the apparatus is not sufficiently powerful compared to the size of the video signal of the compressed digital data, the apparatus cannot reproduce both the audio signal component and the video signal component in real time with which they are recorded originally as taking too much time is taken for the data decompression processes. In consequence, the video signal component may be represented as a slow-motion film and the audio signal component is reproduced intermittently.

The intermittent reproduction of the audio signal component gives a sense of intense incongruity to the human hearing. Therefore, the audio signal component is usually processed to be reproduced continuously with a priority over the video signal component. The data decompression and displaying process of the video signal component is performed when the audio signal component is not processed. Here, adequate thin down of the picture frames (so-called frame omission) becomes necessary for synchronizing the reproduction timings of the audio signal component and the video signal component so that the video signal component is not delayed relative to the audio signal component.

As for the digital data compression, there is known an MPEG (Motion Picture Experts Group) standard, wherein the audio signal component and the video signal component of the AV signals are multiplexed and compressed into digital data each added with timing information indicating the reproduction time of each. By referring to the timing

information, the audio signal component and the video signal component are reproduced in synchronization in the decompression procedure.

5 As a prior art technique of the synchronized playback of the audio (voice) signal component and the video (picture) signal component, there is, for example, a Japanese patent application entitled "An apparatus for compression and reproduction of AV signals" and laid open as Provisional Publication No. 107514/95 (hereafter called the first prior art apparatus).

10 In the first prior art apparatus, color difference information is thinned down in a video data compression procedure considering the processing speed of the reproduction apparatus. Compensating the thinned down color difference information by simple interpolation, the decompression procedure of the video data is performed at a high speed, enabling synchronized reproduction of the video data and the audio data.

15 In another Japanese patent application entitled "Variable speed synchronized playback of recorded digital audio and video data" and laid open as Provisional Publication No. 303240/95 (hereafter called the second prior art apparatus), here is proposed another method of synchronized playback with a variable speed of the audio signal component and the video signal component.

20 In this method, a time required for decompression and playback of a frame of the video signal component, and a time required for decompression and reproduction of a block of the audio signal component are calculated from respective attribute information of each. One of the calculated times is used as a master time clock for decompression of the audio signal component and the video signal component. The audio signal component and the video signal component are decompressed more speedily or more slowly according to a scaling factor of the master time clock indicated by a user.

25 A problem of the first prior art apparatus is that feedback control means for the synchronization are not provided. In the first prior art apparatus, the thinning down of the color difference information is performed a priori in the data compression procedure according to the amount of information in the video signal component to be processed, and no means are provided for detecting lead or lag of a displayed video frame to the
30 corresponding audio data. Therefore, unnecessary thinning down of color difference information is performed even when the playback apparatus is provided with sufficient data processing capability, and restoration of the lead or lag of the video signal component to the audio signal component cannot be performed flexibly, in the first prior art apparatus.

A problem of the second prior art apparatus is that a TDHS (Time Domain Harmonic Scaling) system is used for synchronizing reproduction of the audio signal component with the scaled master time clock. The TDHS system needs a somewhat complicated processing inconvenient to be applied to the playback system with low data processing capability, without saying of the distortion of the audio signal component.

Furthermore, there is another problem with the prior art apparatus. When the playback apparatus is implemented with software operating on an operating system with a low resolution software timer, there is a problem that correct synchronization between the audio signal component and the video signal component cannot be obtained because of the low resolution of the software timer, even if the feedback control means are provided. That is, when the resolution of the software timer is 10ms and decompression process of the audio signal component corresponding to one frame of the video signal component having 30 frames per second takes 13ms, for example, the decompression process of the audio signal component is measured to be accomplished in 10ms, and 23ms is allotted for decompression process of the video signal component of one frame. Therefore, a delay of 3ms per frame is derived in the reproduction of the video signal component.

Therefore, a primary object of the present invention is to provide an apparatus for synchronized playback of audio-video signals with a simple configuration to be applied to a system having comparatively low data processing capability such as a personal computer, wherein the video signal component can be decompressed and reproduced in synchronization with the audio signal component adequately from the compressed digital data according to the data processing capability of the system, independent of resolution of the software timer of the system.

In order to achieve the object, an apparatus for synchronized playback of audio-video signals of the invention, for representing an audio signal component and a video signal component in synchronization with each other by decompressing compressed digital data of the audio signal component and the video signal component, comprises:

means for obtaining a reference time indicating reproduction time passage of the audio signal component from reproduced data amount of the audio signal component, a video frame counter for counting and maintaining a real frame number of the video signal component which indicates a sum of a frame number actually decompressed and a frame number omitted to be decompressed,

a delay detector for calculating an ideal frame number of the video signal component to be displayed originally from the reference time, and detecting progress of the video signal component by comparing the real frame number to the ideal frame number, and

5 means for restoring a difference of display timing of the video signal component from the reference time by performing frame omission and decompression waiting of the video signal component according to the progress of the video signal component.

Therefore, the video signal component can be decompressed and reproduced in synchronization with the audio signal component adequately from the compressed digital data with a simple configuration according to the data processing capability of a system wherein the apparatus of the invention is applied, independent of resolution of the software timer of the system, without needing any special hardware timer.

According to another aspect of the present invention, there is provided an apparatus for synchronized playback of audio-video signals for representing an audio signal component and a video signal component in synchronization with each other by decompressing compressed digital data of the audio signal component and the video signal component, said apparatus comprising:

means for obtaining a reference time indicating reproduction time passage of the audio signal component from reproduced data amount of the audio signal component; and

20 means for detecting progress of the video signal component in reference to said reference time.

The foregoing, further objects, features, and advantages of this invention will become apparent from a consideration of the following description, the appended claims, and the accompanying drawings wherein the same numerals indicate the same or the corresponding parts.

In the drawings:

Figure 1 is a block diagram illustrating a fundamental configuration of an apparatus for synchronized playback of audio-video signals according to an embodiment of the invention;

Figure 2 is a schematic diagram illustrating an example of data stream of the video signal component subjected to frame omission;

Figure 3 is a schematic diagram illustrating the PCM buffer 23 of Figure 1 for storing the data of the audio signal component to be reproduced;

Figure 4 is a flowchart illustrating main processes of the embodiment of Figure 1;

Figure 5 is a flowchart illustrating details of the audio processing performed by an interruption at step A3 of Figure 4; and

5 Figure 6 is a flowchart illustrating details of the video processing performed at step A4 of Figure 4.

Now, embodiments of the present invention will be described in connection with the drawings.

10 Figure 1 is a block diagram illustrating a fundamental configuration of an apparatus for synchronized playback of AV signals according to an embodiment of the invention.

Referring to Figure 1, the apparatus comprises:

a data recording device 1 for storing compressed digital data of AV signals,
an AV separation unit 2 for extracting an audio signal component and a video
15 signal component multiplexed in the compressed digital data read out of the data recording device 1,

a video signal processor 3 for performing decompression of the video signal component extracted by the AV separation unit 2,

20 an audio signal processor 4 for performing decompression of the audio signal component extracted by the AV separation unit 2,

an AV synchronization controller 5 for performing synchronization control between the audio signal component and the video signal component,

a CRT monitor 6 for displaying the video signal component, and

a speaker 7 for reproducing the audio signal component.

25 The video signal processor 3 includes:

a compressed video data buffer 11 for storing compressed data of the video signal component extracted by the AV separation unit 2,

30 a video decoder 12 for decompressing the compressed data of the video signal component buffered in the compressed video data buffer 11 by analyzing attribute information (video header information) thereof,

a frame buffer 13 for buffering frame data of the video signal component decompressed by the video decoder 12, and

a video display controller 14 for generating display signals to be supplied to the CRT monitor 6 by reading out the frame data prepared in the frame buffer 13 in synchronization with synchronous signals for the CRT monitor 6.

The audio signal processor 4 includes:

5 a compressed audio data buffer 21 for storing compressed data of the audio signal component extracted by the AV separation unit 2,

an audio decoder 22 for decompressing the compressed data of the audio signal component buffered in the compressed audio data buffer 21 by analyzing attribute information (audio header information) thereof,

10 a PCM buffer 23 for storing data of the audio signal component decompressed by the audio decoder 22, and

an audio reproduction controller 24 for generating sound signals to be supplied to the speaker 7 according to the data of the audio signal component supplied from the PCM buffer 23.

15 The AV synchronization controller 5 includes:

a video frame counter 31 for counting and maintaining a real frame number of the video signal component which indicates a sum of a frame number actually decompressed and a frame number omitted to be decompressed by the video decoder 12, from the beginning of reproduction,

20 an audio data counter 32 for counting and maintaining the amount of data of the audio signal component actually decompressed and reproduced by the audio reproduction controller 24 from the beginning of reproduction according to reproduction information supplied from the audio reproduction controller 24,

25 a clock generator 33 for calculating a reference time indicating reproduction time passage of the audio signal component from the data amount counted by the audio data counter 32 referring to the audio header information analyzed by the audio decoder 22,

30 a delay detector 34 for calculating an ideal frame number of the video signal component to be displayed originally from the reference time calculated by the clock generator 33 referring to the video header information analyzed by the video decoder 12, and detecting lead or lag of the video signal component to the audio signal component by comparing the ideal frame number and the real frame number counted by the video frame counter 31, and

a frame omission controller 35 for performing the frame omission when the video signal component is delayed from the audio signal component according to the lead or lag detected by the delay detector 34 referring to priority of frames to be decompressed.

In other words, the video frame counter 31 of the AV synchronization controller
5 5 operates as means for counting the real frame number of the video signal component actually processed (decompressed or omitted), and the audio data counter 32 and the clock generator 33 operate as means for obtaining the reference time indicating reproduction time passage of the audio signal component, as will be described afterwards, independent of a software timer of the system which may not be available for correct
10 synchronization because of its insufficient resolution.

Further, the delay detector 34 operates as means for calculating the ideal frame number of the video signal component to be decompressed and displayed originally in an interval indicated by the reference time, as well as operating as means for detecting progress of the video signal component with reference to the reference time, that is,
15 detecting the lead or lag of the video signal component to the audio signal component by comparing the ideal frame number and the real frame number. The frame omission controller 35 operates as means for restoring the lead or lag of display timing of the video signal component to reproduction timing of the audio signal component by indicating the frame omission or decompression waiting of the video signal component to the video
20 decoder 12, according to the progress of the video signal component detected by the delay detector 34.

Here, the frame omission is performed according to the order of priority of frames as will be described in the following paragraphs, and the decompression waiting is performed when a lead of the video signal component is detected by the delay detector
25 34 until the audio signal component overtakes the video signal component.

Now, the order of priority of frames is described.

Figure 2 is a schematic diagram illustrating an example of data stream of the video signal component subjected to the frame omission performed by the frame omission controller 35. In the example, frame data of the video signal component are ranged according to an inter-frame prediction coding system such as the MPEG standard, for
30 example.

Referring to Figure 2, the data of the video signal component has a frame composition comprising a plurality of GOP (Group of Pictures) data, each including data of I (Intra), P (Predictive) and B (Bidirectionally predictive) pictures. The I-picture, coded

from intra-frame data, can be decompressed independently from itself without any key frame, and is used as a key frame to be referred to when following frames are decompressed. The P-picture, coded according to forward prediction, is decompressed referring to a preceding key frame. The P-picture is also used as a key frame to be referred to when following and preceding frames are decompressed. The B-picture, coded according to bi-directional prediction, is decompressed referring to a preceding key frame and a following key frame, being referred to by no other frame.

The key frame is a picture to be referred to when other pictures are decompressed, and so, when a key frame is omitted to be decompressed, following P and B-pictures until a next I-picture appears should be also omitted, unable to be decompressed.

Therefore, the frame omission priority is attributed to the B-picture, the P-picture and the I-picture, in that order.

Now, the PCM buffer 23 for storing the data of the audio signal component to be reproduced is described referring to a schematic diagram of Figure 3.

The PCM buffer 23 operates as a ring buffer comprising a plurality of memory blocks (1) to (n), wherein the data of the audio signal component decompressed by the audio decoder 22 are stored to be read out and supplied to the audio reproduction controller 24 in turns cyclically to be reproduced.

When reproduction by the audio reproduction controller 24 of data read out and transferred from a memory block of the PCM buffer 23 is accomplished, an interruption occurs and data in the next memory block is transferred to the audio reproduction controller 24 to be reproduced. The audio decoder 22 decompresses the audio signal component to be stored in the next memory block which is read out, interrupting processing of the video signal component even when the video signal component is under processing. For example, when reproduction of data of the memory block (1) of Figure 3 is accomplished, data of the memory block (2) is read out, transferred to the audio reproduction controller 24 and begins to be reproduced. In the memory block (1), data of the audio signal component newly decompressed by the audio decoder 22 is stored to be read out after reproduction of data of the memory block (n) is accomplished.

Thus, the audio signal component is decompressed with the interruption processing prior to other procedures such as the decompression process of the video signal component, in the embodiment. Therefore, the audio signal component can be reproduced without interruption, and accordingly, a precise reproduction time can be

obtained from reproduction progress of the audio signal component at the audio reproduction controller 24.

Now, operation of the synchronized playback of AV signals in the embodiment of Figure 1 is described referring to flowcharts of Figures 4 to 6.

5 Figure 4 is a flowchart illustrating main processes of the embodiment.

As an initializing procedure, data read-out (at step A1) and audio data decompression (at step A2) are performed at the beginning of reproduction of AV signals.

10 At step A1, a block of multiplexed digital data of compressed AV signals is read out from the data recording device 1, and separated into compressed data of the video signal component and those of the audio signal component, which are stored in the compressed video data buffer 11 and the compressed audio data buffer 21, respectively. Then, at step A2, the video frame counter 31 and the audio data counter 32 are reset, and audio data decompression is executed, wherein compressed data of the audio signal component to be prepared in all the memory blocks of the PCM buffer 23 are read out from
15 the compressed audio data buffer 21 and decompressed referring to the audio header information therein.

After the initialization procedure, audio processing (at step A3) and the video processing (at step A4) are repeated until it is detected (at step A5) that there is no compressed data left to be processed in the compressed video data buffer 11. Here, in
20 Figure 4, the audio processing of step A3 and the video processing at step A4 are depicted to be performed sequentially. However, in the embodiment, the audio processing at step A3 is performed whenever an interruption occurs for the audio data processing, as beforehand described, interrupting the video processing at step A4 even when it is under execution at the time.

25 Figure 5 is a flowchart illustrating details of the audio processing performed by an interruption at step A3 of Figure 4.

When there is no data in the audio reproduction controller 24, an interruption occurs and data of a first/next memory block in the PCM buffer 23 are read out and transferred to the audio reproduction controller 24 (at step B1), which are converted into
30 analog signals to be supplied to the speaker 7 according to a sampling frequency indicated by the audio header information. Then (at step B5), the compressed data of the audio signal component are read out from the compressed audio data buffer 21, decompressed by the audio decoder 22 and stored in the first/next memory block which is read out, on condition that there are compressed data left to be processed in the compressed audio

data buffer 21. Then, the audio processing returns and waits for a next interruption. When it is detected (at step B2) that there is no compressed data to be processed in the compressed audio data buffer 21, the data read-out at step B4 is performed in the same way with step A1 described in connection with the initialization procedure of Figure 4, on condition it is detected (at step B3) that there are compressed digital data of AV signals left to be reproduced in the data recording device 1. When there is detected no compressed digital data of AV signals left to be reproduced in the data recording device 1, only the data read-out and transfer at step B1 is performed for each interruption until data of all the memory blocks in the PCM buffer 23 are transferred to the audio reproduction controller 24.

Thus, the audio signal component is reproduced without interruption, in the embodiment.

Now, details of the video processing at step A4 of Figure 4 is described referring to the flowchart of Figure 6, wherein the synchronization control and decompression of the video signal component are performed.

First (at step C1), the data amount of the audio signal component actually decompressed and reproduced by the audio reproduction controller 24 from the beginning of reproduction is acquired from the audio data counter 32. Then (at step C2), the reference time indicating the reproduction time passage of the audio signal component is calculated by the clock generator 33 from the data amount referring to the audio header information obtained at step A2 of Figure 4, as follows.

When the audio header information indicates audio data of Layer I, bit rate of 192,000 bits/sec, sampling frequency of 44.1KHz and 2-channel stereo signal, for example, 2 channels of analog signals are generated by the audio reproduction controller 24 from 2 x 16bits x 441,000/sec of the decompressed data supplied from the PCM buffer 23. Therefore, when the data amount of D bytes is acquired from the audio counter 32, the reference time T is calculated as $T(\text{sec}) = D \times 8 / (44,100 \times 2 \times 16)$.

Thus, the reference time T is calculated according to the data amount D counted by the audio data counter 32 from the reproduction information supplied from the audio reproduction controller 24, independent of the software timer of the system without needing any special hardware timer, in the embodiment.

By multiplying the reference time T by a frame rate of 30 frames/sec, for example, indicated by the video header information, the ideal frame number $FI = 30T$ is

obtained, which is compared to the real frame number FR (at step C3) by calculating progress $PR = FR - FI$ of the video signal component.

5 When the progress $PR = 0$, the video signal component is synchronized to the audio signal component, whereas it is leading when the progress PR is positive and lagging when the progress PR is negative. However, when the decompression waiting or the frame omission is performed directly according to positive/negative of the progress PR, the control becomes unstable. Therefore, a threshold range is preferably provided for the synchronization discrimination. In the embodiment, when $-4 < \text{the progress PR} < 1$, the video signal component is discriminated to be synchronized to the audio signal component
10 (at step C4), considering the fact that the lead of the video signal component gives a greater sense of incongruity than the lag thereof.

Returning to Figure 6, when the ideal frame number FI is 10 and the real frame number FR is 11, for example, the video signal component is discriminated to be leading the audio signal component, the progress PR being one, and control of the video
15 processing returns directly from step C4. The video processing of step A4 of Figure 4, that is, steps C1 to C4 of Figure 6 are repeated (including the audio processing of step A3, when there is an interruption) until the progress PR becomes less than one, the ideal frame number becoming more than 11, in the example. Thus, the decompression waiting is performed in the embodiment.

20 When the ideal frame number FI is 10 and the real frame number is 8, for example, the video signal component is discriminated to be synchronized with the audio signal component, the progress PR being -2. In this case, the control goes to step C5 from step C4. Compressed data of a frame are read out from the compressed video data buffer 11, decompressed by the video decoder 12 and stored in the frame buffer 13 (at step C6)
25 to be displayed on the CRT monitor 6 by the video display controller 14. Then (at step C7), the real frame number FR is increased by one, and the control returns.

When the ideal frame number FI is 10 and the real frame number is 5, for example, the video signal component is discriminated to be lagging from the audio signal component, the progress PR being -5. In this case, the control goes to step C8, where the
30 omission priority of the concerning frame is checked. In the embodiment, the frame omission is set not to be performed for the I and the P-picture, the control going to step C5 in the case for performing the decompression of the concerning frame. If it is the B-picture, the control goes to step C9 for rejecting data of the concerning frame from the

compressed data buffer 11, and the real frame number is increased by one at step C10 in the same way with step C7.

Thus, the synchronized playback of the AV signals is performed in the embodiment.

5 Here, the frame omission is described to be performed only for B-pictures. However, the frame omission may be set to be performed for the P or I-frames too, by providing a second and a third threshold value according to the omission priority, for example, when the progress PR is lower than the second and the third threshold value. In this case, data of the compressed video data buffer 11 are to be rejected until data of
10 next I-frame at step C9 and the real frame number FR is to be added at step C10 with a number of frames rejected at step C9.

As heretofore described, the synchronization control of the AV signals is performed according to the data amount D counted by the audio data counter 32 from the reproduction information supplied from the audio reproduction controller 24 in the
15 embodiment, and so, the video signal component can be decompressed and reproduced in synchronization with the audio signal component adequately from the compressed digital data with a simple configuration, independent of resolution of the software timer of a system wherein the apparatus of the embodiment is applied, without needing any special hardware timer.

20 The audio signal component is processed with a priority to be reproduced without any interruption, and even when the frame omission is necessary, it is performed according to priority of frames, enabling a smooth presentation of the video signal component, and eliminating unnecessary frame omission according to the data processing ability of the system.

25

What is claimed is:

1. An apparatus for synchronized playback of audio-video signals for representing an audio signal component and a video signal component in synchronization with each other by decompressing compressed digital data of the audio signal component and the video signal component, said apparatus comprising:

means for obtaining a reference time indicating reproduction time passage of the audio signal component from reproduced data amount of the audio signal component; and

means for detecting progress of the video signal component in reference to said reference time.

2. The apparatus for synchronized playback of audio-video signals recited in claim 1, said means for detecting progress of the video signal component having:

a video frame counter for counting and maintaining a real frame number of the video signal component which indicates a sum of a frame number actually decompressed and a frame number omitted to be decompressed; and

a delay detector for calculating an ideal frame number of the video signal component to be displayed originally from said reference time, and detecting said progress of the video signal component by comparing said real frame number to said ideal frame number.

3. The apparatus for synchronized playback of audio-video signals recited in claim 1, further comprising means for restoring difference of display timing of the video signal component from said reference time by performing frame omission and decompression waiting of the

5 video signal component according to said progress of the video signal component.

4. The apparatus for synchronized playback of audio-video signals recited in claim 3; wherein said frame omission is performed according to an order of priority attributed to frames of the video signal component subject to said frame omission in connection with kinds of
5 compressed data of said frames.

5. The apparatus for synchronized playback of audio-video signals recited in claim 3; wherein said decompression waiting is performed by controlling decompression of compressed data of a frame of the video signal component to wait until said difference of display timing of said
5 frame from said reference time becomes within a threshold value.

FIG. 1

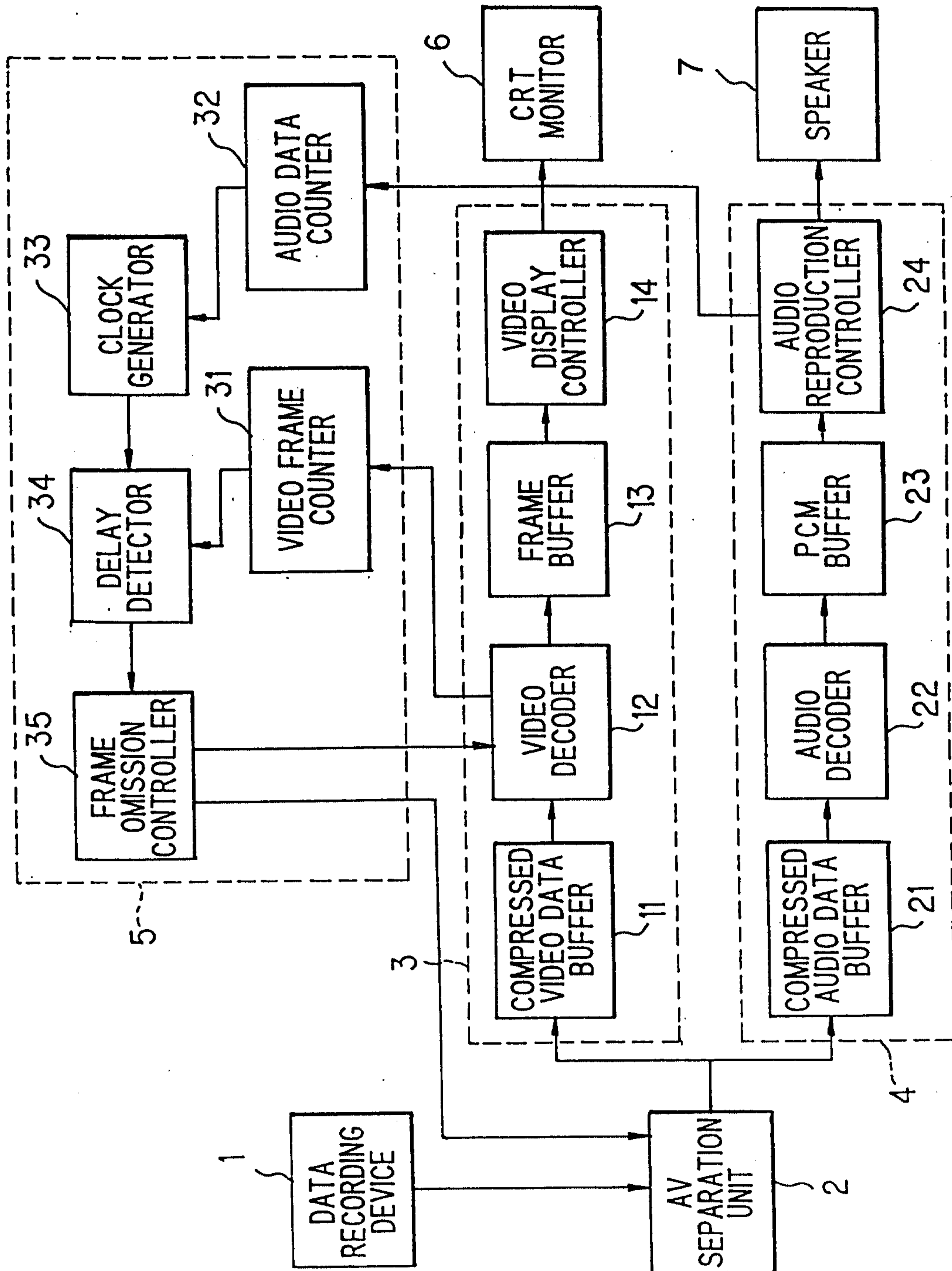


FIG. 2

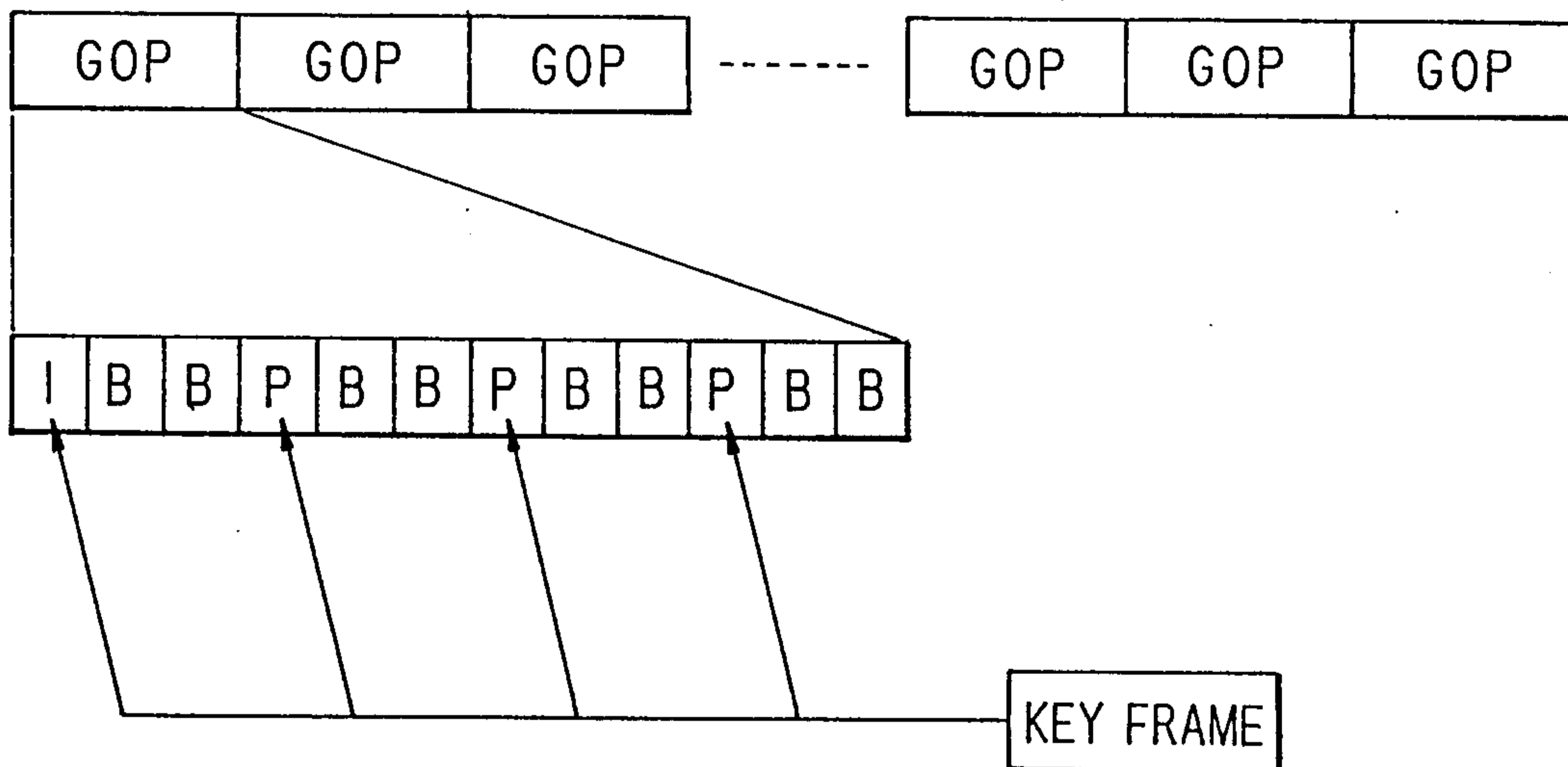
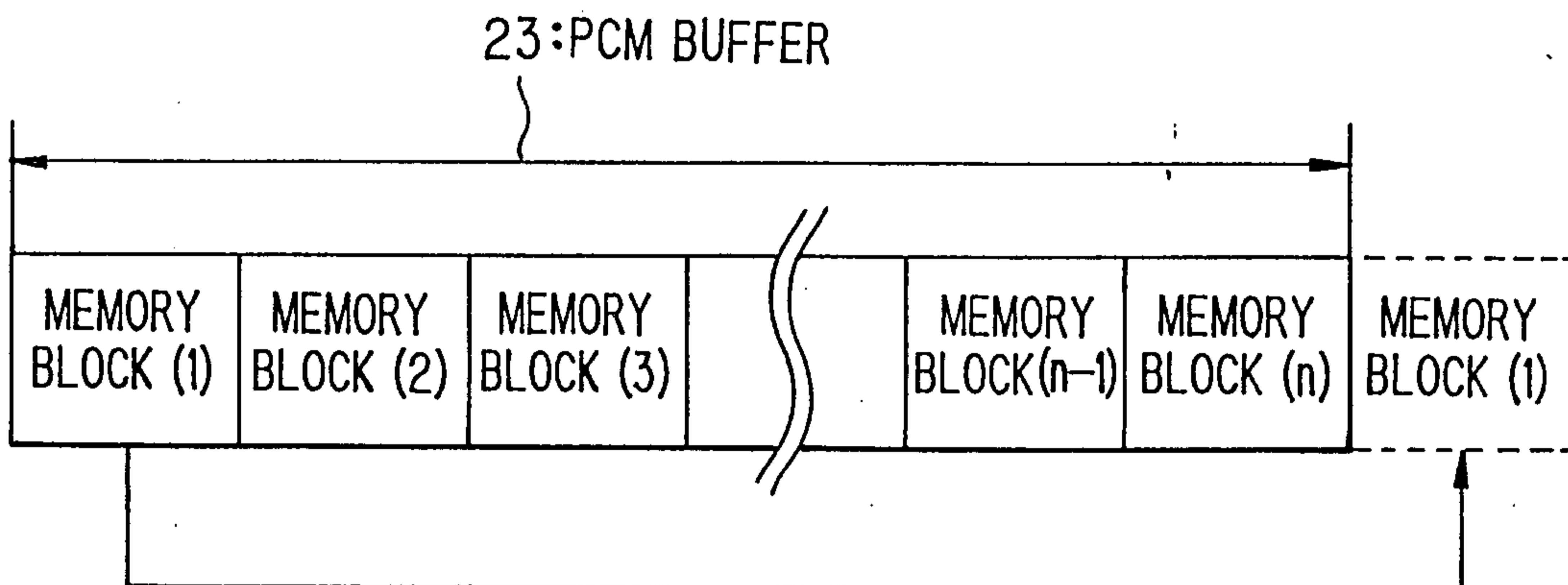
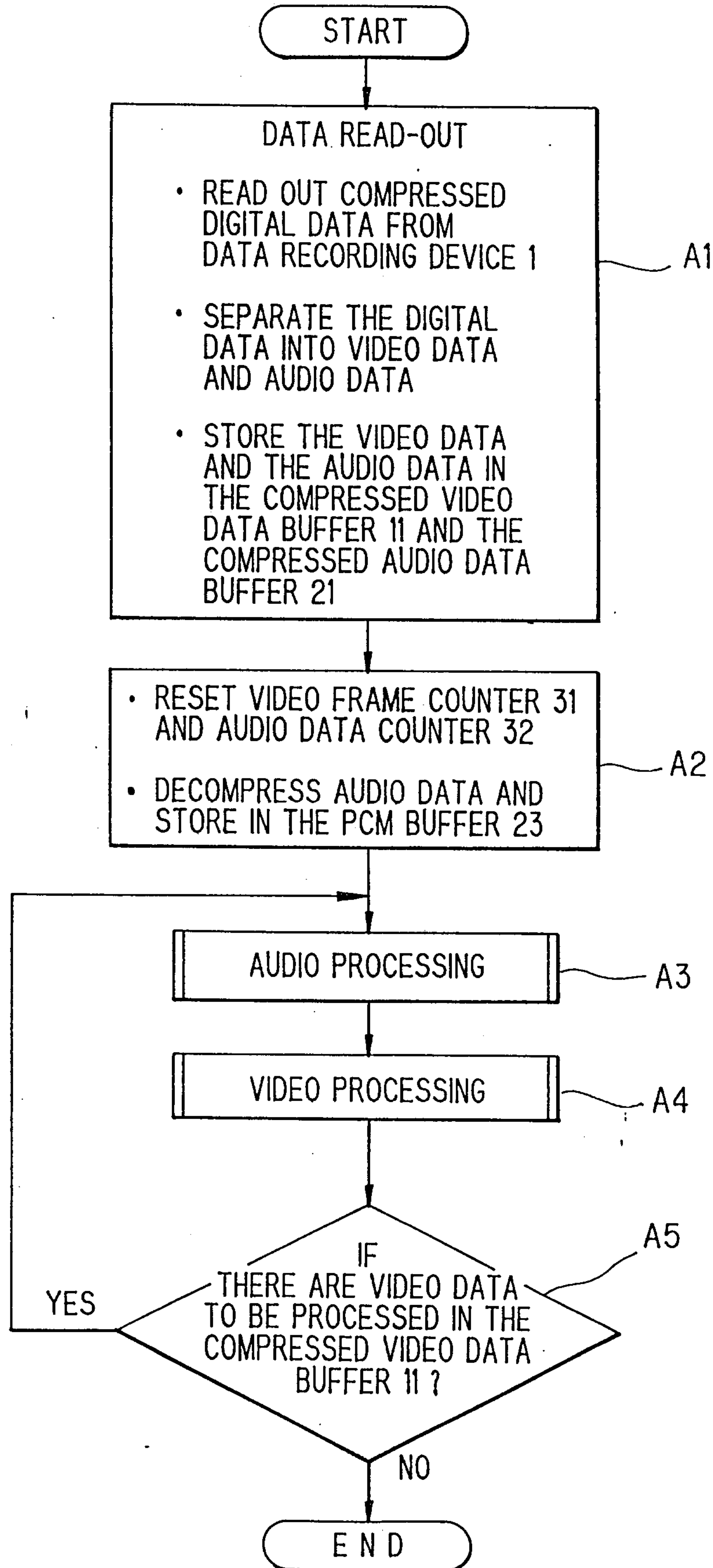


FIG. 3



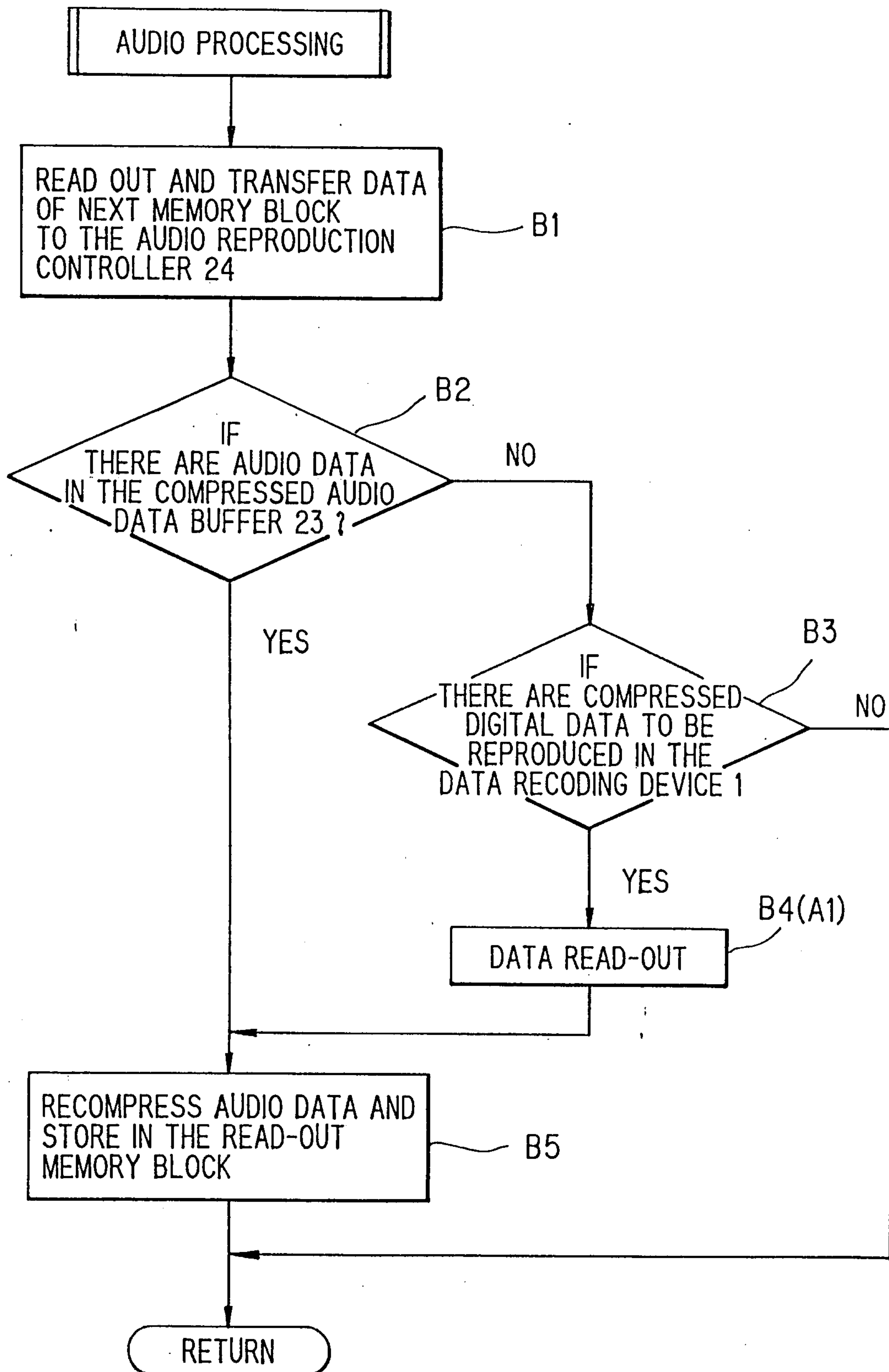
3/5

FIG. 4



4/5

FIG. 5



5/5

FIG. 6

