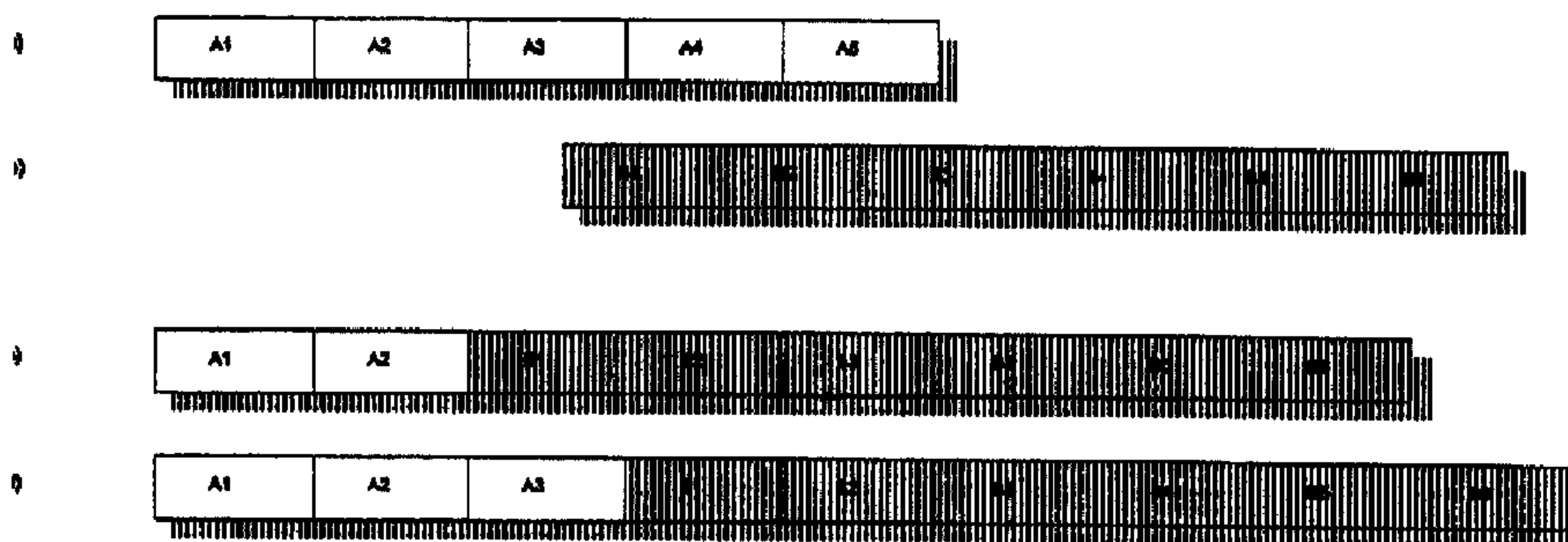




(86) Date de dépôt PCT/PCT Filing Date: 1998/05/29  
 (87) Date publication PCT/PCT Publication Date: 1998/12/03  
 (45) Date de délivrance/Issue Date: 2007/03/13  
 (85) Entrée phase nationale/National Entry: 1999/11/29  
 (86) N° demande PCT/PCT Application No.: GB 1998/001574  
 (87) N° publication PCT/PCT Publication No.: 1998/054909  
 (30) Priorité/Priority: 1997/05/30 (GB9711313.8)

(51) Cl.Int./Int.Cl. *H04N 7/52* (2006.01),  
*G11B 27/031* (2006.01), *G11B 27/10* (2006.01),  
*G11B 27/30* (2006.01)  
 (72) Inventeur/Inventor:  
 WELLS, NICHOLAS DOMINIC, GB  
 (73) Propriétaire/Owner:  
 BRITISH BROADCASTING CORPORATION, GB  
 (74) Agent: FETHERSTONHAUGH & CO.

(54) Titre : TRAITEMENT DE SIGNAUX VIDEO ET AUDIO  
 (54) Title: VIDEO AND AUDIO SIGNAL PROCESSING



Examples of switching audio at frame boundaries (Audio frame  
 period = 24 ms)

- a) Bitstream A (co-timed with video A)
- b) Bitstream B (co-timed with video B)
- c) Switched bitstream with Audio B advanced
- b) Switched bitstream with Audio B delayed

(57) Abrégé/Abstract:

To maintain lip-sync when a video signal has been edited or switched at a video frame boundary, the associated audio signal is advanced or retarded to the closest audio frame boundary. Any error introduced by this constraint is accumulated and carried forward to direct a subsequent advance or retard decision. In this way, the cumulative error can be kept within an acceptable tolerance.

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> <b>H04N 7/52, G11B 27/031</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 98/54909</b> <b>(43) International Publication Date:</b> 3 December 1998 (03.12.98)
<b>(21) International Application Number:</b> PCT/GB98/01574 <b>(22) International Filing Date:</b> 29 May 1998 (29.05.98)  <b>(30) Priority Data:</b> 9711313.8              30 May 1997 (30.05.97)              GB  <b>(71) Applicant (for all designated States except US):</b> BRITISH BROADCASTING CORPORATION [GB/GB]; Broadcasting House, London W1A 1AA (GB).  <b>(72) Inventor; and</b> <b>(75) Inventor/Applicant (for US only):</b> WELLS, Nicholas, Dominic [GB/GB]; 17 Wellington Road, Brighton, East Sussex BN2 3AB (GB).  <b>(74) Agents:</b> GARRATT, Peter, Douglas et al.; Mathys & Squire, 100 Gray's Inn Road, London WC1X 8AL (GB).		<b>(81) Designated States:</b> AU, CA, JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Published</b> <i>With international search report.</i>
<b>(54) Title:</b> VIDEO AND AUDIO SIGNAL PROCESSING		
<b>(57) Abstract</b>  <p>To maintain lip-sync when a video signal has been edited or switched at a video frame boundary, the associated audio signal is advanced or retarded to the closest audio frame boundary. Any error introduced by this constraint is accumulated and carried forward to direct a subsequent advance or retard decision. In this way, the cumulative error can be kept within an acceptable tolerance.</p>		

**VIDEO AND AUDIO SIGNAL PROCESSING**

This invention relates to the editing and switching of digital television signals consisting of video and associated sound components. It is particularly relevant to systems which handle the audio component in compressed form.

5 The component parts of the TV signal necessarily carry with them information which enables their relative replay/display timing to be maintained. However, if for any reason, the replayed/displayed audio/video timings get out of step then this is most noticeable when people are speaking. Consequently, an error in audio/video replay/display timings is often referred to as "lip-sync error".

10 A variety of techniques have been proposed for maintaining lip-sync in circumstances where the video and audio components have undergone separate processing with the introduction of different delays. Reference is directed, for example, to United Kingdom patent number 2 273 215 filed on November 24, 1993 which discloses a technique by which a delay signal is associated with either the  
15 video or audio signal, that delay signal representing the relative delay between video and audio. Signal processing elements which introduce delay are provided with synchronizing units which read the delay signal at the input, add or subtract the appropriate delay interval and insert the modified delay signal at the output. At the end of a signal processing chain, the delay signal is read and the appropriate delay  
20 interval is inserted. Since there is usually more signal processing conducted on the video component, the greater delay will usually be suffered by the video signal. The compensating delay at the end of the signal processing chain is typically an audio delay only. If a video delay is required, the above-referenced United Kingdom patent (number 2 273 215) explains that this can only be adjusted in coarse steps in order to  
25 maintain synchronization between its input and output. An additional audio delay will then be used to restore precise synchronization.

The present invention recognizes that in a number of key applications, the audio signal itself will be arranged in discrete frames having a defined time interval. It will no longer be possible in such



applications to achieve precise synchronization – to any desired level of accuracy – through coarse frame-by-frame control of video delay and fine, continuous control of audio delay. The prior art proposals, when audio processing is necessarily organised into discrete audio frame intervals, will enable synchronization to be achieved only to  
5 within the resolution of one audio frame. The time intervals of audio frames – as will later be discussed in more detail – are considerably shorter than video frames and a lip-sync error of a fraction of an audio frame time interval may not be objectionable. The difficulty is that the effect of such sub-frame synchronization errors is cumulative and conventional equipment may still produce objectionable lip-sync errors even if  
10 every delay within the chain has been defined within the accuracy of one audio frame.

It is an object of the present invention to provide improved methods for use in editing, switching and like signal processing of temporally associated frame-based signals, which address this problem.

15 It is a further object of this invention to provide improved methods for use in editing, switching and like signal processing where sub-frame synchronization errors are controlled or managed.

In accordance with a broad aspect, the invention provides a method for use in the editing and switching of video and associated audio signals in which editing and  
20 switching of a video signal is constrained to video frame boundaries and editing and switching of an audio signal is constrained to audio frame boundaries, wherein accumulated sub-frame synchronization error information is carried forward.

Preferably, accumulated sub-frame synchronization error information is employed to direct a frame advance/retard decision in a frame accurate  
25 synchronization process.

Suitably, said advance/retard decision operates on audio frames.

Advantageously, accumulated sub-frame synchronization error information is used to keep the total sub-frame synchronization error introduced by the cascading of processing stages within certain tolerances.

30 In accordance with another broad aspect, the present invention provides a method for use in editing, switching and like signal processing of temporally associated frame-based signals, in which steps are taken to ensure frame accurate

synchronization, characterized in that sub-frame synchronization error feedback is provided to minimize cumulative sub-frame synchronization error.

In accordance with yet another aspect, the present invention provides a method for use in editing, switching and like signal processing of temporally associated frame-based signals which are provided in digital form with information prescribing the time of presentation frames, characterized by the maintenance through said processes of signal synchronization information, independent of said information prescribing the time of presentation of frames.

In accordance with another broad aspect, the invention provides a method of carrying synchronisation information for use in the editing and switching of video and associated audio signals in which editing and switching of a video signal is constrained to video frame boundaries and editing and switching of an audio signal is constrained to audio frame boundaries. The method comprises conveying the audio and video signals to be edited or switched and carrying cumulative sub-frame synchronisation error information in a lip-sync-error field associated with the signals.

In accordance with yet another broad aspect, the invention provides a method of editing or switching of video and associated audio signals in which editing and switching of a video signal is constrained to video frame boundaries and editing and switching of an audio signal is constrained to audio frame boundaries. The method comprises editing or switching a video signal at a video frame boundary, editing or switching an audio signal at an audio frame boundary, reading a lip-sync error field to obtain accumulated sub-frame synchronization error information and using the information to minimize cumulative sub-frame synchronization error resulting from said editing or switching.

In accordance with yet another broad aspect, the invention provides an apparatus for editing or switching of video and associated audio signals in which editing and switching of a video signal is constrained to video frame boundaries and editing and switching of an audio signal is constrained to audio frame boundaries. The apparatus comprises means for determining the time at which an audio frame should be replayed in order to maintain synchronism with the associated video. The apparatus also comprises means for reading a lip-sync-error field carrying cumulative sub-frame synchronization error information and means for making a



-3a-

frame advance/retard decision based on the cumulative sub-frame synchronization error information.

The invention will now be described by way of example with reference to the accompanying drawings, in which:-

5           Figure 1 is a block diagram illustrating the synchronization of frame based audio signals; and

          Figure 2 is a block diagram illustrating apparatus performing a method according to one embodiment of this invention.

10           When digital audio is compressed, the compressed audio data is usually organized into "audio frames". For example, in MPEG2 Layer II audio coding, the audio frames are 24ms long.

          When switching/editing such audio data it may be important to switch between source audio clips at times coincident with audio frame boundaries (for example in order to avoid or minimize the build up of re-quantisation noise). If this is the case  
15           then, when switching/editing video and associated audio signals it is unlikely that it will be possible to maintain continuity and regularity of video frames and simultaneously maintain continuity and regularity of audio frames without introducing some lip-sync error after the switch/edit point. This is because the video frame period and the audio frame period are not equal or related in a very simple fashion.

20           For the MPEG Layer II example the maximum lip-sync error introduced at a switch/edit point could be up to +/- 12 ms as shown in

- 4 -

Figure 1.

Along the process of programme production, continuity switching and distribution, the signal may be switched and edited several times, and successive errors could accumulate to an unacceptable value. Therefore, some mechanism is required to prevent the build up of lip-sync errors through successive generations of editing/switching. This mechanism is in addition to other mechanisms which may be employed to ensure the overall audio/video synchronization at points along the combined signal chain.

Coded digital audio frames usually contain bytes for signalling framing information together with information as to how to decode that audio frame. Also, there are usually slots in the audio frame which can be used to carry user-specified data.

The proposed solution to the lip-sync error problem is to use some of these user-specified data bytes to signal the lip-sync error introduced by an edit switch process. At each edit switch process in the production and distribution chain, any equipment performing a switch can read the existing value in the lip-sync field and decide whether to delay or advance the switched audio (by including or removing one extra audio frame) such that the overall lip-sync error is minimized after the new switch edit point.

Following the switch point the equipment should re-insert the new value for the total lip-sync error in the lip-sync error field.

For the example of MPEG Layer II switching and editing, this technique would enable the total lip-sync error to be kept within +/- 12 ms. Some systems may be defined which switch on sub-frame boundaries (e.g. at 8ms intervals), but this technique would still be important to prevent unacceptable accumulation of lip-sync errors.

It should be noted that compressed bitstreams such as those defined by the MPEG standard contain information, called *Presentation Time Stamps (PTS)*, which instruct a decoder at what time a given audio or video frame should be replayed/displayed. It is intended that this *PTS* information is the means by which the audio and video should be replayed/displayed in synchronism at the output of a decoder. However, in situations such as



- 5 -

that described by Figure 1, the decoder is being instructed by the audio *PTS* values to decode the audio with a possible lip-sync error. It is not possible to simply change the audio *PTS* values for the audio frames following a switch point because this would lead to a discontinuity in the decoded audio signal.

For audio (and video) editing systems, the audio/video signal is replayed from disk according to an *edit list* or a *playout list*. In this case, the audio/video replay/display timing of any particular audio/video frame is specified through the *edit list* rather than via *PTS* values. In switching between separate audio source clips, the requirement for continuity of audio frames will inevitably lead to lip-sync errors as shown in Figure 1.

In summary therefore, editing/switching equipment should perform the following operations around each edit/switch point:

- i) Note the time at which an audio frame should be replayed in order to maintain synchronism with the associated video. (This information might come from synchronization information carried in the bitstream or from an edit list).
- ii) Read the lip-sync-error field.
- iii) In order to maintain continuity of audio frames (or some sub-multiple of the audio frame period), decide whether to advance or delay the following sequence of audio frames in a manner which minimizes the accumulated lip-sync error.
- iv) Re-insert in the each frame of the audio a new value corresponding the new accumulated lip-sync error.

Referring now to Figure 2, an audio switch 20 receives audio signal A through fixed delay 22 and audio signal B through variable delay 24. A switch controller 26 receives an Edit Decision List (EDL) prescribing in this case the desired point of switching from A to B. The switch controller additionally receives audio signal A, from which it determines the frame timings, and audio signal B, from which it determines the current accumulated sub-frame synchronization error. Utilising this information, the switch controller decides whether to advance or retard signal B at the



switch point, as illustrated in Figure 1. To enable signal B to be effectively advanced or retarded, a fixed delay is inserted in the path of signal A and the switch controller selects the appropriate value for the variable delay in the path of signal B. A residual sub-frame synchronization error is calculated and provided as an output for optional embedding in the switched output through embedder 28.

It will be recognised that for simplicity, apparatus has been shown which is capable of switching from signal A to signal B only; a symmetrical version of the described apparatus would also permit switching from signal B to signal A.

In one project, a TV switching and editing system is being developed which uses compressed MPEG2 coded video and MPEG Layer II coded audio. The audio is stored in the form of Packetised Elementary Streams (PES) with one audio frame per PES packet. The PES packet consists of a header of 40 bytes followed by the bytes of the compressed audio frame data. The header contains 16 bytes of user data. Two bytes of this user data field have been reserved to signal the lip-sync error in units of 90kHz clock periods.

In some applications switching may be done using audio signals which have been decoded from previously compressed data. The switched signal may then be re-compressed and/or passed on to equipment which can correct for any lip-sync error. In these cases, it is beneficial to be able to carry lip-sync error information within the decoded audio signal.

Several methods for carrying such data within a decoded audio signal which may be in digital or analogue form have been described in a previous patent application [patent application in the United Kingdom serial no. 9701616.6 filed on January 27, 1997].

For decoded audio signals, the concept *audio frame* can still be applied. In the decoded domain, an audio frame has a one-to-one correspondence with the relevant audio frame in the compressed domain.

The ideas described in this document can also be applied to the switching and editing of any program related data (PRD) where relative timing of the replay of the PRD signal needs to be maintained to within a

- 7 -

given tolerance in relation to other signals in the digital multiplex.



**CLAIMS**

1. A method of carrying synchronisation information for use in the editing and switching of video and associated audio signals in which editing and switching of a video signal is constrained to video frame boundaries and editing and switching of an audio signal is constrained to audio frame boundaries, the method comprising:
- conveying the audio and video signals to be edited or switched;
- carrying cumulative sub-frame synchronisation error information in a lip-sync-error field associated with the signals.
2. A method of editing or switching of video and associated audio signals in which editing and switching of a video signal is constrained to video frame boundaries and editing and switching of an audio signal is constrained to audio frame boundaries the method comprising:
- editing or switching a video signal at a video frame boundary;
- editing or switching an audio signal at an audio frame boundary;
- reading a lip-sync error field to obtain accumulated sub-frame synchronization error information;
- using the information to minimize cumulative sub-frame synchronization error resulting from said editing or switching.
3. A method according to claim 2, wherein using the information comprises directing a frame advance/retard decision in a frame accurate synchronization process based on the information.
4. A method according to claim 3, in which said advance/retard decision operates on audio frames.
5. A method according to any of claims 2 to 4 further comprising inserting a new value into the lip-sync-error field following a switch point.
6. A method according to any of claims 1 to 5, in which accumulated sub-frame synchronization error information is carried as data within compressed audio data.

7. A method according to claim 6 wherein the accumulated sub-frame synchronization information is carried in the digital header of a compressed audio frame.

5

8. A method according to any of claims 1 to 5, in which accumulated sub-frame synchronization error information is carried as data within a decoded/uncompressed digital or audio signal.

10

9. A method according to any of claims 1 to 8, in which accumulated sub-frame synchronization error information is used to keep the total sub-frame synchronization error introduced by the cascading of processing stages within certain tolerances.

15

10. A method according to claim 9 wherein for an MPEG Layer II signal the error is kept within +/- 12 ms.

20

11. A method according to any of claims 1 to 10 wherein the sub-frame synchronization error information is maintained independent of information prescribing the time of presentation of frames.

25

12. Apparatus for editing or switching of video and associated audio signals in which editing and switching of a video signal is constrained to video frame boundaries and editing and switching of an audio signal is constrained to audio frame boundaries, the apparatus comprising:

means for determining the time at which an audio frame should be replayed in order to maintain synchronism with the associated video, including:

a) means for reading a lip-sync-error field carrying cumulative sub-frame synchronization error information;

b) means for making a frame advance/retard decision based on the cumulative sub-frame synchronization error information.

30

13. Apparatus according to claim 12 further comprising means for inserting a new value in the lip-sync-error field following switching.



1/2

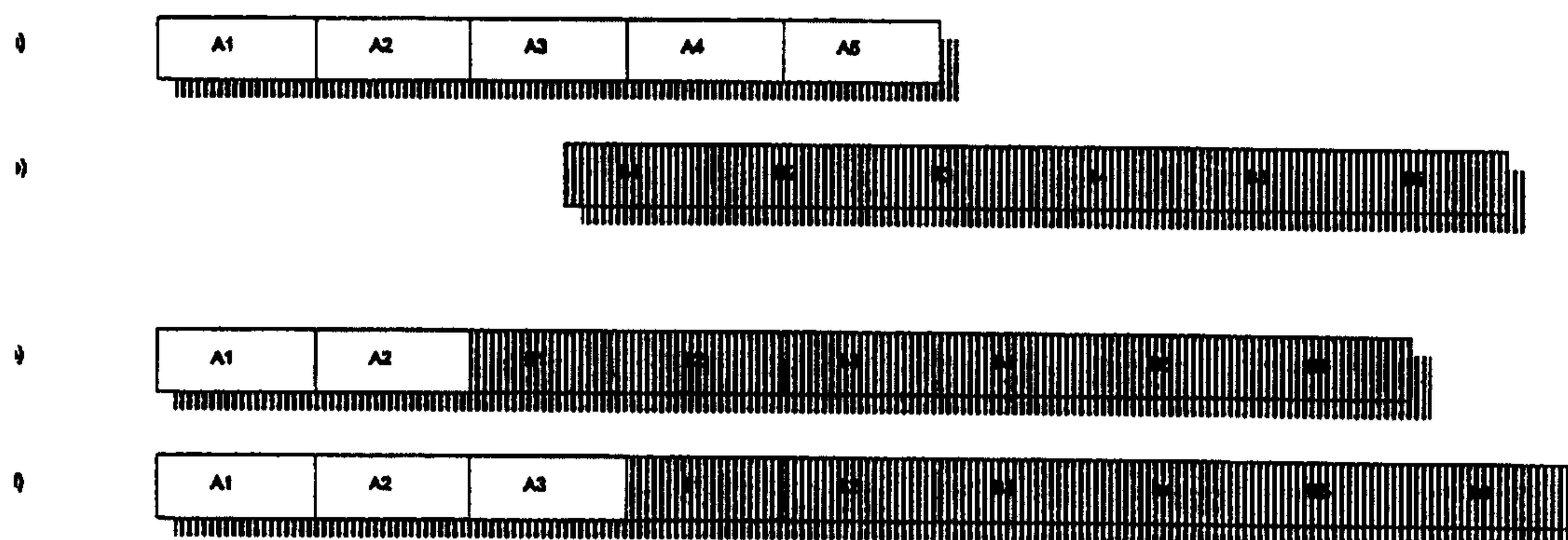
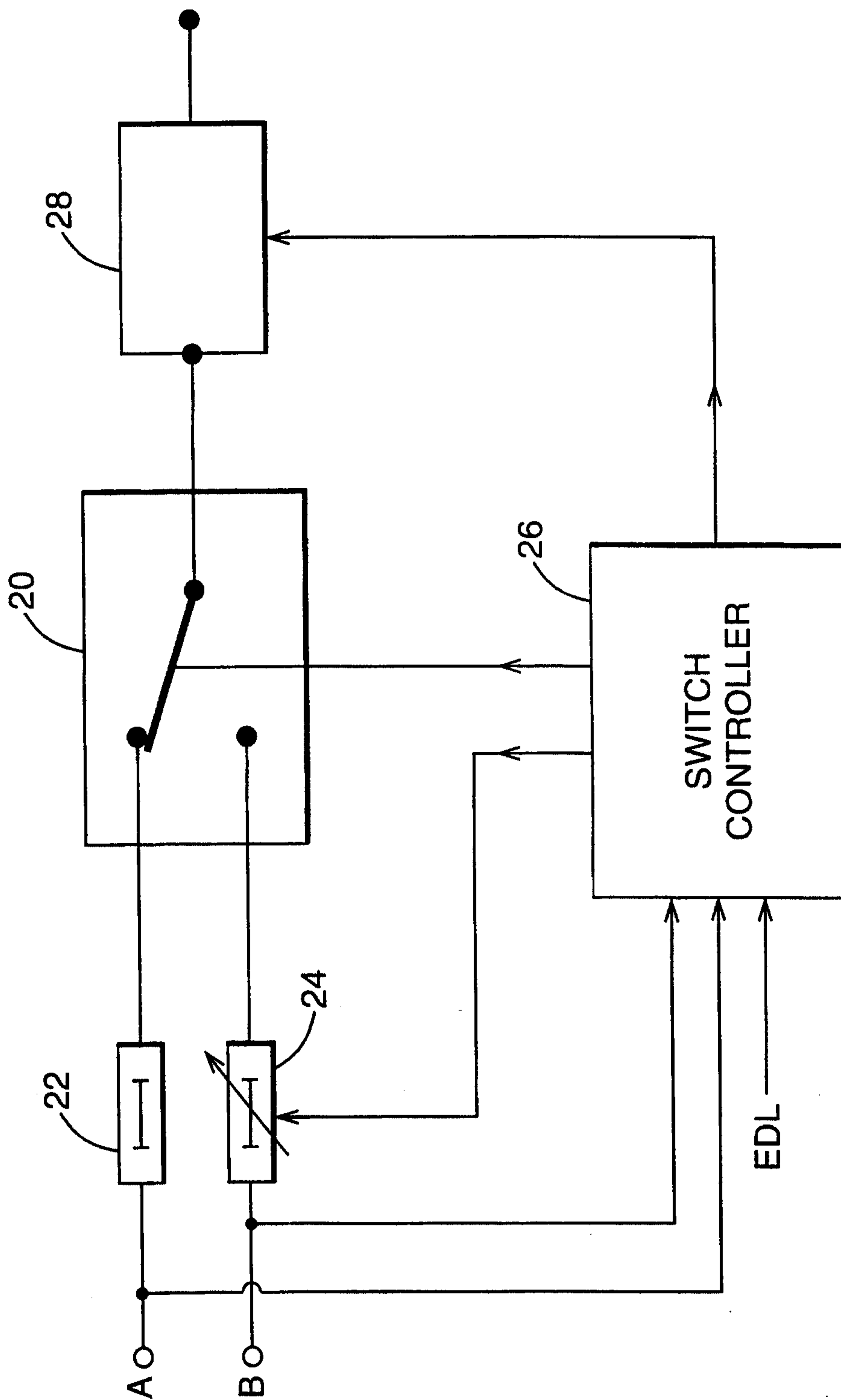


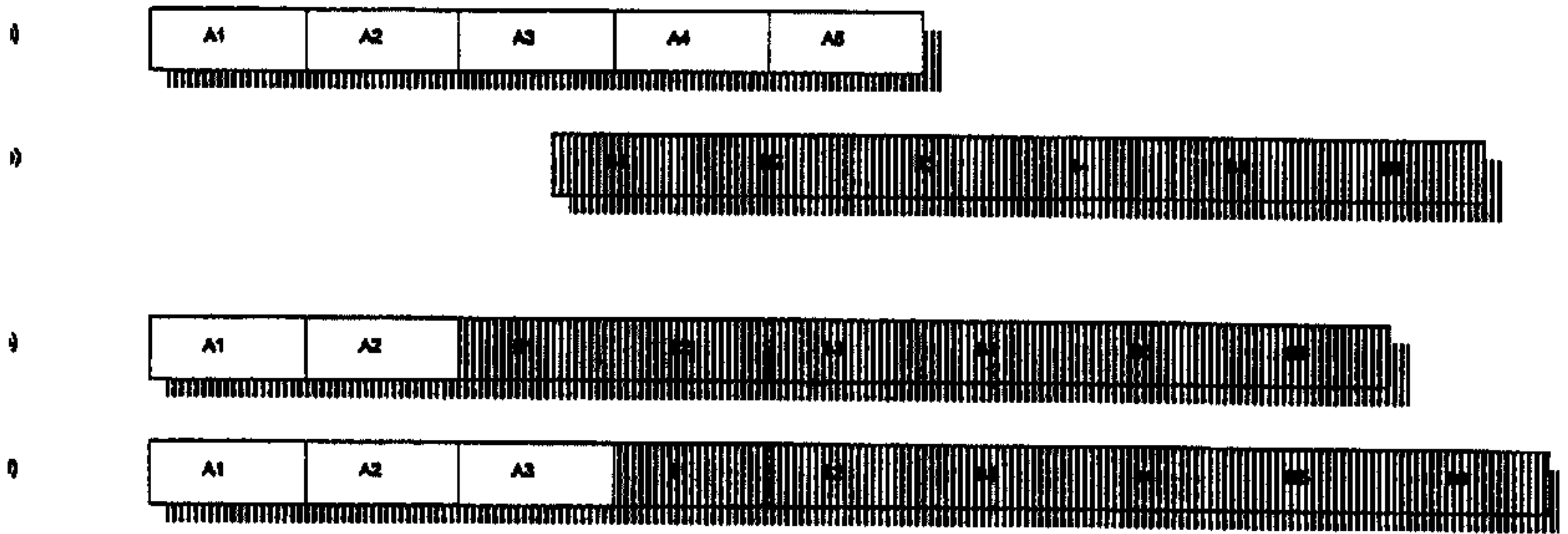
Figure 1: Examples of switching audio at frame boundaries (Audio frame period = 24 ms)

- a) Bitstream A (co-timed with video A)
- b) Bitstream B (co-timed with video B)
- c) Switched bitstream with Audio B advanced
- b) Switched bitstream with Audio B delayed

FIG. 2







Examples of switching audio at frame boundaries (Audio frame period = 24 ms)

- a) Bitstream A (co-timed with video A)
- b) Bitstream B (co-timed with video B)
- c) Switched bitstream with Audio B advanced
- b) Switched bitstream with Audio B delayed