

AUDIO/VIDEO SYNCHRONIZATION IN A DIGITAL TRANSMISSION SYSTEM

This invention relates to a method and apparatus for synchronizing the audio and video component signals at an audio/video signal decompression apparatus

Audio and video (A/V) transmission and reception requires that the audio and video components be properly synchronized. EIA standards RS-250-B limit the time differential between associated audio and video signals to 25 ms. lead or 40 ms. lag time. Film standards limit the time differential of associated audio and video to $\pm 1/2$ frame, which corresponds to 20.8 ms. An acceptable goal for source to viewer A/V synchronization is therefore a time differential of ± 20 ms.

Digital communication systems typically time-multiplex associated signal components over a single channel. Such multiplexing is common among A/V transmission systems proposed and implemented for cable, fiber, terrestrial and satellite applications. The time multiplexing of the signal components may destroy their natural time relationships between the transmission and display of the information. Therefore, time critical components of the transmitted component signals may be associated with a time reference before being multiplexed. This is referred to as "stamping" the information, and timing samples are referred to as time stamps. The receiver may then output the respective components in time relative to their respective time stamps. However to accomplish this the receiver must maintain a very precise local time reference, which is synchronous to the encoder time reference.

One reason that the receiver must be tightly coupled to the time base of the transmitter is to insure that the output of real time data matches that of the input to the receiver. If the receiver provides (displays) the data too rapidly, buffers in the receiver may underflow resulting in an interruption of the output signal. If the receiver outputs the data too slowly, the buffers may overflow (assuming finite rate buffers) resulting is a loss of data.

In one proposed system the receiver is synchronized to the transmitter by supplemental time stamps (system clock references, SCR) associated with predetermined packets of



transmitted information. The timing of the capture of the time stamps, SCR, bear no relation to presentation time stamps (PTS) which are related to the video data, other than by virtue of being derived from the same counter. The SCR codes are generated by
 5 sampling a modulo 2^N counter ($N \geq 32$) which counts a substantially constant frequency crystal clock at the transmitter. The receiver incorporates a phase locked loop which has a free running frequency substantially equal to the frequency of the clock in the transmitter. The receiver clock (local clock) is also
 10 counted modulo 2^N , and each time a SCR arrives at the receiver the local counter is sampled to provide a local clock reference or LCR. No attempt is made to force the LCR to equal the SCR. Rather the local clock is adjusted based upon processing changes in the difference between the LCR and SCR time stamps. An error signal
 15 is generated according to the relation

$$ERR = |SCR_n - SCR_{n-1}| - |LCR_n - LCR_{n-1}|$$

The signal ERR is utilized to control the local clock frequency. Via this process the LCR can be made arbitrarily close to the transmitter clock frequency. Note that since both the system and
 20 local clocks are counting modulo N, they periodically wrap around. On these occurrences the respective terms $SCR_n - SCR_{n-1}$ and $LCR_n - LCR_{n-1}$ will be negative and erroneous. The system monitors the polarity of the respective differences and when one of the differences is negative the difference is ignored.

25 Video signal coded according to the MPEG standard includes presentation time stamps, PTS_{vid} , which are synchronized to the input video frames. The respective PTS_{vid} indicate the relative times that the respective frames are to be displayed at the receiver, nominally 30 Hz, for NTSC source material.
 30 Associated audio is also encoded with presentation time stamps PTS_{aud} based on the same time base as the system time and which time stamps are placed in an MPEG system packet layer encompassing the encoded audio data. An audio system packet layer may contain several "frames" of audio data and respective
 35 frames equal, in this example, 24 ms. of original audio data. Audio frames are approximately six times the duration of a (127 byte) transport packet. (Information to be transmitted, audio, video, data, etc. are segmented into respective transport packets



of predetermined size, with a variety of control words appended, to provide an extra layer of error correction/detection and synchronization.) In addition, according to the MPEG protocol, the number of audio frames per MPEG system layer is a variable.

5 Hence there may be little or no correlation between the video PTS_{vid} and audio PTS_{aud} presentation time stamps for associated audio and video source material. Thus synchronizing the audio and video components is difficult if one attempts to do so by comparing the PTS_{vid} with the PTS_{aud} . It is an object of the

10 present invention to simplify the process of synchronizing associated audio and video components.

A receiver for decoding associated video and audio information components transmitted in mutually exclusive "frames" of data with respective time stamps PTS_{vid} and PTS_{aud}

15 respectively, includes a controller which is responsive to the respective received time stamps to provide coarse synchronization by delaying or skipping respective frames of one or the other of the components to approximately time align the two components. Fine synchronization is provided by adjusting the processing or

20 clock frequency of the audio signal processor independent of the video processor. The control for the frequency adjustment is related to the difference between audio and video time stamps.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGURE 1 is a block diagram of audio/video

25 compression apparatus.

FIGURE 2 is a block diagram of a audio/video decompression apparatus embodying the present invention.

FIGURE 3 is a block diagram of apparatus for providing the receiver system clock signal having substantially the same

30 rate as the system clock of the compression apparatus.

FIGURE 4 is a flow chart of the operation of the FIGURE 2 apparatus.

FIGURES 5 and 6 are block diagrams of alternative audio processing clock signal generation which may be

35 implemented for element 215 in FIGURE 2.

FIGURE 1 illustrates a typical system in which the invention may be practiced, which system is a compressed digital video signal transmission arrangement. In this system, video



signal from a source 10 is applied to a video signal compression element 11 which may include a motion compensated predictive encoder utilizing discrete cosine transforms. Compressed video signal from the element 11 is coupled to a formatter 12. The
5 formatter arranges the compressed video signal and other ancillary data according to some signal protocol such as MPEG, a standard developed by the International Organization for Standardization (Organisation Internationale De Normalisation). The standardized signal is applied to a transport processor 13,
10 which divides the signal into packets of data and adds certain overhead to provide some noise immunity for the transmission purposes. The transport packets, which normally occur at a non-uniform rate are applied to a rate buffer 14 which provides output data at a relatively constant rate conducive to efficient use
15 of a relatively narrow bandwidth transmission channel. The buffered data is coupled to a modem 15 which performs the signal transmission.

A system clock 22 provides clocking signal to operate much of the apparatus, at least including the transport processor.
20 This clock will operate at a fixed frequency such as 27 MHz for example. As shown herein, however, it is used to generate timing information. The system clock is coupled to the clock input of a counter 23 which may be arranged to count modulo 2^{30} , for example. The count values output by the counter are applied to
25 two latches 24 and 25. The latch 24 is conditioned by the video source to latch count values on the occurrence of respective frame intervals. These count values are denoted presentation time stamps, PTS's, and are included in the compressed video signal stream by the formatter 12, and are used by the receiver to
30 provide lip-synchronization of associated audio and video information. The latch 25 is conditioned by the transport processor 13 (or the system controller 21) to latch count values according to a predetermined schedule. These count values are denoted system clock references, SCR's, and are embedded as
35 auxiliary data within respective auxiliary transport packets.

Audio signal associated with the video signal from source 10 is applied to an audio signal compressor 18. The compressor 18 provides frame sampling pulses (independent of



video frames) to control a latch 19. Responsive to the sampling pulses, the latch 19 captures count values provided by the counter 23. These latched values correspond to audio presentation time stamps PTS_{aud} . The PTS_{aud} are incorporated in the compressed audio signal provided by the compressor 18. The compressed audio signal is coupled to a transport processor 17 which divides the signal into packets of data and adds certain overhead to provide some noise immunity for transmission purposes. The audio transport packets provided by the processor 17 are coupled to a multiplexor 16 which time division multiplexes the audio and video transport packets. In the figure, separate transport processors are shown in the audio and video signal processing channels. For systems wherein the data rate is moderate, the functions of the two transport processors and the multiplexor 16 may be subsumed in a single transport processor.

The system controller 21 is a variable state machine programmed to coordinate the various processing elements. Note that the controller 21, the compressors 11 and 18, the transport processors 13 and 17, and the rate buffer 14 may or may not operate synchronously via a common clocking arrangement as long as proper handshaking is provided between processing elements. However, the two compressors both derive PTS values from the same reference counter 23, thus a precise timing relationship between the two compressed signals is provided in the compressed output signal.

FIGURE 2 illustrates an exemplary receiver apparatus embodying the invention wherein the modem 200 performs the inverse function of the modem 15 and the rate buffers 204 and 206 perform, in effect, the inverse function of the rate buffer 14. In FIGURE 2 a single inverse transport processor 202 is indicated, which processor divides respective transport packets by service and allocates the respective data to the appropriate processing channel. In so doing the respective transport packet signal payloads are separated from the auxiliary data, with the respective payloads being applied to the appropriate processing channel and the auxiliary data applied to the system controller 210. In an alternative arrangement a separate transport processor may be included in each processing channel and



arranged to recognize and process only data associated with the respective channels.

Compressed video data from the inverse transport processor 202 is applied to the rate buffer 204, which provides
5 compressed video signal according to the system protocol to the decompressor 214. The rate buffer 204 accepts data at a bursty or inconsistent rate, and provides data on demand to the decompressor 214. The decompressor, responsive to the compressed video signal, generates non-compressed video signal
10 for display or for storage etc., in appropriate display or storage devices (not shown).

Compressed audio data from the inverse transport processor 202 is applied to the rate buffer 206, which provides compressed audio signal according to the system protocol to the
15 audio decompressor 212. The decompressor, 212, responsive to the compressed audio signal generates non-compressed audio signal for reproduction or for storage etc., in appropriate speakers or storage devices (not shown).

The inverse processor 202 also provides SCR's from
20 the auxiliary transport data, and control signals to a system clock generator 208. The clock generator responsive to these signals generates a system clock signal synchronous with at least the transport processor operation. This system clock signal is applied to the receiver system controller 210 to control the timing of
25 appropriate processing elements.

Figure 3 illustrates details of an exemplary clock regenerator 208. Data from the receiver modem 200 is coupled to the inverse transport processor 202', including an auxiliary packet detector 31. The inverse transport processor 202'
30 separates transport header data from the respective transport packet payloads. Responsive to the transport header data, the processor 202' applies audio and video signal payloads to, for example, respective decompression apparatus (not shown), and auxiliary data (designated as AUX) to the appropriate auxiliary
35 data processing elements (not shown). The SCR's residing within the auxiliary data are routed and stored in a memory element, 34.

The auxiliary packet detector 31, which may be a matched filter arranged to recognize the codewords designating an



auxiliary transport packet containing a SCR, produces a control pulse on the occurrence of transport packets containing such data. The control pulse is utilized to capture and store within a latch 35, the count value currently exhibited by the local counter 36 at a
 5 time precisely related to the time of detection. The local counter 36 is arranged to count pulses provided by a voltage controlled oscillator 37. The counter 36 is arranged to count modulo M which may be, but is not necessarily, the same number as its counterpart counter in the encoder (counter 23). If M differs
 10 from N the difference may be accommodated in the error equation.

The voltage controlled oscillator 37 is controlled by a low pass filtered 38 error signal provided by a clock controller 39. The error signal is generated in the following fashion. Let the SCR
 15 arriving at time n be designated SCR_n and the local count value concurrently captured in the latch 35 be designated LCR_n . The clock controller reads the successive values of SCR's and LCR's and forms an error signal E proportional to the differences

$$E \Rightarrow |SCR_n - SCR_{n-1}| - |LCR_n - LCR_{n-1}|$$

20 The error signal E, is utilized to condition the voltage controlled oscillator 37 to a frequency tending to equalize the differences. As indicated previously, negative differences exhibited due to modulo counter wrap around, may be ignored. The error signal produced by the clock controller 39 may be in the form of a pulse
 25 width modulated signal, which may be rendered into an analog error signal by implementing the low pass filter 38 in analog components.

The constraints on this system are that the counters at the two ends of the system count the same frequency or even
 30 multiples thereof. This requires that the nominal frequency of the voltage controlled oscillator be fairly close to the frequency of the system clock at the encoder.

The foregoing approach provides rather rapid synchronization but may introduce a long term error. Long term
 35 error LTE is proportional to the difference

$$LTE \Rightarrow |LCR_n - LCR_0| - |SCR_n - SCR_0|$$

where SCR_0 and LCR_0 are for example the first occurring SCR and the corresponding latched value of the receiver counter.



Nominally the error signals E and LTE will vary in discrete steps. As such, once the system is "synchronized" the error signal will dither one unit about the null point. The preferred method of synchronization is to initiate control of the voltage controlled oscillator using the error signal E until a one unit dither occurs in the error signal E, and then to switch to the use of the long term error signal LTE to control the voltage controlled oscillator.

The system clock signal provided by the VCXO 37 may be utilized to operate at least the transport processor and rate buffers. Since it is synchronized at least in frequency with the encoder system clock, the possibility of rate buffer overflow or underflow due to clock timing errors is substantially non-existent.

Refer again to FIGURE 2 for an explanation of audio/video synchronization. Recall that a presentation time stamp PTS_{vid} , is included in the compressed video signal associated with predetermined video data. The PTS_{vid} is indicative of the relative time that the associated video is to be displayed. Similarly the compressed audio signal includes presentation time stamps PTS_{aud} associated with audio to be reproduced at times related to the respective PTS_{aud} . At the receiver the PTS_{aud} and PTS_{vid} may not be compared directly to provide A/V synchronization because the respective samples were determined at different instants. The respective PTS values are compared to a continuous time base which is the receiver clock provided by the VCXO 37. This is done by sampling local time stamps of the count values LCR.

When data associated with an associated PTS is presented, The LCR is sampled. For example, The audio decompressor 212 issues a PTS_{aud} when a respective audio frame is output for reproduction. At these times a control signal conditions the latch 220 to sample the LCR, the values of which will be designated LAS, for local audio stamp. Similarly when the video decompressor provides a video frame for display, it provides a PTS_{vid} and a control pulse to condition a latch 222 to store the current value of the LCR. These LCR values are designated LVS for local video stamps.

The LAS and the corresponding PTS_{aud} are coupled to respective input terminals of a subtracter 218 which develops the signal $\Delta A-PTS$ according to the relation;



$$\Delta A\text{-PTS} = \text{PTS}_{\text{aud}} - \text{LAS}$$

The LVS and the corresponding PTS_{vid} are coupled to respective input terminals of the subtractor 217 which develops the signal $\Delta V\text{-PTS}$ according to the relation;

5
$$\Delta V\text{-PTS} = \text{PTS}_{\text{vid}} - \text{LVS}.$$

The signals $\Delta V\text{-PTS}$ and $\Delta A\text{-PTS}$ are coupled to respective input terminals of a further subtractor, 219, which develops an A/V synchronization error signal ERR_{PTS} according to the relation;

$$\text{ERR}_{\text{PTS}} = \Delta V\text{-PTS} - \Delta A\text{-PTS}$$

10

Synchronization of the audio and video requires that the A/V synchronization error be driven to zero. This indicates that when the difference in the values of the corresponding audio and video PTS's equals the time, in units of the local reference, between the
15 occurrence of the corresponding PTS's, the audio and video signal will be in synchronization.

Two mechanisms may be used to adjust the A/V synchronization based upon the error signal ERR_{PTS} ; skips and repeats of data sections and conversion clock deviation. Skipping
20 fixed intervals or "frames" of audio advances the audio data stream by a fixed interval relative to the video signal. Repeating (or muting without consuming data) delays the audio data stream by fixed intervals relative to the video signal. Skipping and repeating audio frames is audible under many conditions, and
25 therefore is only utilized for coarse adjustment of synchronization. Even so, brief skipping or repeating may be preferable to discernible audio/video synchronization errors. If the audio frames are less than 40 msec, coarse adjustment by skipping/repeating may result in synchronization errors within
30 ± 20 msec., which is within industry standards for A/V synchronization. However, this synchronization will degrade if the audio conversion time base does not match that of the source. Once synchronization is coarsely adjusted, variations of the audio conversion clock frequency are arranged to further refine the A/V
35 synchronization.

The error signal ERR_{PTS} is applied to a filter and processing element 216. A filter function therein smoothes the signal ERR_{PTS} to minimize aberrant effects that might otherwise



be generated by signal noise. The processing function in the element 216 then examines the smoothed error signal and determines whether a skip/repeat of audio should be utilized to effect coarse synchronization of the audio and video signals
5 and/or whether an adjustment to the audio processing frequency should be utilized to effect fine synchronization. If a coarse synchronization adjustment is determined to be necessary, the processor 216 provides a control signal to the audio decompressor 212 to condition the decompressor to skip or repeat the current
10 decompressed audio frame. Alternatively, or in addition to the coarse adjustment, if a fine adjustment is determined to be necessary, the processor 216 provides a control signal to the audio time base 215 to adjust the frequency of the audio processing clock signal.

15 The processing algorithm is detailed in the flow chart of FIGURE 4. After initialization 400 of the system, which is designated START, the system monitors 401 the audio decompressor for the occurrence of a PTS_{aud} and if a PTS_{aud} is detected it is read 403 and a local clock reference LAS is captured
20 and stored. If a PTS_{aud} has not occurred, the system monitors the video compressor for a PTS_{vid} 402. If a PTS_{vid} has occurred, the PTS_{vid} is read and a local clock reference LVS is captured and stored 404. When both a PTS_{aud} and a PTS_{vid} have been read the ERR_{PTS} is calculated 405 according to the equation;

25
$$ERR_{PTS} = \Delta V - PTS - \Delta A - PTS.$$

The magnitude of the error signal is examined 406 to determine if it is greater than one half an audio frame interval. If it is greater than one half an audio frame interval, the error signal is checked for polarity 407. If the polarity is positive, the current audio
30 frame is repeated 409. If it is negative the current audio frame is skipped 408. After skipping or repeating a frame the system iterates back to the start position to wait for the next occurrence of PTS's.

At step 406, if the magnitude of the error signal is less than
35 one half an audio frame interval, the error is examined to determine if it is greater than zero 410. If the error is greater than zero, the error is checked 412 to determine if it is less than the previous error signal. If it is less than the previous error



signal, this is an indication that the system is converging toward synchronization, and the synchronization control parameters are not changed. The system returns to the start position to wait for the next PTS's. Conversely, if the error has increased over the
5 previous error signal, the audio system processing clock is adjusted to lessen its frequency 414.

At step 410, if the error is less than zero (negative), it is checked 411 to determine whether it is greater than the previous error signal. If it is greater than the previous error signal, this is
10 also an indication that the system is converging toward synchronization, and the synchronization control parameters are not changed. Alternatively if the current error signal is less than the previous error signal the system is moving further out of synchronization and the audio processing clock frequency is
15 increased 413. After processing steps 412 and 413 the system returns to wait for the next occurrence of PTS's. It will be noted in this example, that the system performs only coarse adjustments via skipping or repeating audio frames until the A/V synchronization error is reduced to less than one half an audio
20 frame interval.

In an alternative embodiment, the filtered error signal is compared against a predetermined threshold related to the size of respective audio frames. If the error signal is less than the threshold, indicative that the audio-video timing error is less than
25 an audio frame, the error signal is coupled to the audio time base circuit 215, wherein it is utilized to adjust the frequency of the audio signal processing (decompression) clock. Alternatively, if the error signal is greater than the threshold, the error signal may be divided by the audio frame interval to determine the number
30 of audio frames that the audio and video signal are misaligned. The integer portion of the quotient is applied to the audio decompressor to condition the audio decompressor to skip or repeat that number of audio frames. The polarity of the error signal will determine whether audio frames should be skipped or
35 repeated. Nominally the decompressed data is arranged in a buffer memory prior to being output, hence skipping or repeating audio frames is a simple matter of controllably enabling the memory read/write commands.



The fractional portion of the quotient is coupled to the audio time base circuit 215, where it is utilized to adjust the audio processing clock to fine tune the A/V synchronization.

The rate of production of audio PTS's is proportional to the processing speed of the audio decompressor. The processing speed of the audio decompressor is directly proportional to the frequency of the clock signal used to operate the audio decompressor. If the clock frequency of the audio decompressor is independent of the clock used to operate the video decompressor, and is finely adjustable then the relative rate of occurrence of the audio and video PTS's can be adjusted and the A/V finely synchronized.

FIGURE 5 is a first embodiment of circuitry for generating an adjustable audio processing clock signal. In FIGURE 5 a voltage controlled oscillator, 503, is connected in a phase locked loop including elements 500 and 501. The output of the oscillator is coupled to one input of a phase detector 500. The system clock is coupled to a second input of the phase detector 500 via a binary rate multiplier 505. A phase error signal developed by the phase detector is low pass filtered in filter 501. This smoothed phase error signal is coupled to the control input terminal of the oscillator to condition the oscillator to oscillate at the same frequency and phase as the output of the binary rate multiplier. In this example the system clock frequency is approximately 27 MHz and the desired audio processing clock frequency is approximately $1/380$ of the 27 MHz signal. A control signal from the process element 216 is coupled to a control input port of the binary rate multiplier to control its output frequency. This control signal is selected to represent a nominal division ratio of $1/380$, but modulates about this value so as to condition the output of the binary rate multiplier to produce an output frequency commensurate with the instantaneous desired audio processing rate.

FIGURE 6 illustrates a second example of adjustable audio processing clock generation. In this example, a constant clock frequency, from the system clock generator 208 or some other stable oscillator, is applied to one input of a binary rate multiplier 600. The binary rate multiplier is controlled by a

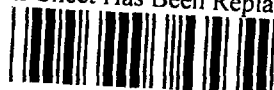


control signal from the process element 216 to generate a nominal desired audio processing clock frequency. Responsive to the A/V error signal the process element 216 alters this control signal to condition the binary rate multiplier to either increase or
5 decrease the nominal audio processing clock frequency.

A further alternative audio signal processing rate changing apparatus (not shown) may include an oscillator for providing a nominal maximum audio processing clock frequency and a gating circuit. The oscillator is coupled to the audio processing circuit via
10 the gating circuit. The gating circuit is controlled by the process element 216 to excise ones of the oscillator output pulses to provide, on average, a desired audio processing clock frequency.

An alternative embodiment of the synchronizing apparatus, indicated by the dashed arrow in FIGURE 2, may be
15 arranged to skip or repeat video frames to effect synchronization. Alternatively video frames may be skipped for the audio leading (lagging) video and audio frames may be skipped for the audio frames lagging (leading) video. However, in the preferred embodiment, the audio frames are skipped/repeated for audio
20 lagging and leading video respectively.

This Sheet Has Been Replaced



CLAIMS:

1. Apparatus for providing synchronized, reproduced audio and video signals from recovered compressed audio and video signal, which audio and video signals include respective time stamps PTSaud and PTSvid determined at predetermined intervals and related to an encoder system clock, said apparatus characterized by:

a source (200) of said recovered compressed audio and video signals, and a source (208, 37) of reference frequency;

a counter (208, 36) responsive to said reference frequency for providing a sequence of values modulo N where N is an integer;

decompression apparatus (214) responsive to said recovered compressed video signal for providing decompressed video signal and said time stamp PTSvid;

decompression apparatus (212) responsive to said recovered compressed audio signal for providing decompressed audio signal and said time stamp PTSaud;

means (222, 217) for capturing values (LVS) of said sequence at predetermined times related to the provision of said time stamps PTSvid and for generating the difference ($\Delta_V\text{-PTS}$) of said values LVS and corresponding time stamps PTSvid; and (220, 218) for capturing values (LAS) of said sequence at predetermined times related to the provision of said time stamps PTSaud and generating the difference ($\Delta_A\text{-PTS}$) of said values LAS and corresponding time stamps PTSaud; and

means (219, 216, 215) responsive to said differences $\Delta_V\text{-PTS}$ and $\Delta_A\text{-PTS}$ for synchronizing the decompressed audio and video signals.

This Sheet Has Been Replaced





2. The apparatus set forth in claim 1 characterized in that said means for synchronizing the decompressed audio and video signals includes means (216, 212) for skipping/repeating predetermined intervals of decompressed audio signal responsive to the difference the between values $\Delta V-PTS$ and $\Delta A-PTS$ exceeding a predetermined value.

3. The apparatus set forth in claim 1 characterized in that said means for synchronizing the decompressed audio and video signals includes means (216,214) for skipping/repeating predetermined intervals of decompressed video signal responsive to the difference the between values $\Delta V-PTS$ and $\Delta A-PTS$ exceeding a predetermined value.

4. The apparatus set forth in claim 1 characterized in that said decompression apparatus (212) responsive to said recovered compressed audio signal provides decompressed audio signal at a rate determined by a processing (215) clock signal and said means for synchronizing the decompressed audio and video signals further includes means (216) responsive to the differences $(\Delta V-PTS)-(\Delta A-PTS)$, for variably adjusting the rate of said processing clock signal.

This Sheet Has Been Replaced





5. Apparatus for providing synchronized, reproduced audio and video signals from recovered compressed audio and video signal, which audio and video signals include respective time stamps PTSaud and PTSvid determined at predetermined intervals and related to an encoder system clock, said apparatus characterized by:

a source (200) of said recovered compressed audio and video signals, and a source (208) of a local clock signal;

decompression apparatus (214) responsive to said recovered compressed video signal for providing decompressed video signal and said time stamp PTSvid;

decompression apparatus (212) responsive to said recovered compressed audio signal for providing decompressed audio signal and said time stamp PTSaud;

means (217, 218, 220, 222) for determining, in cycles of said local clock signal, the time T between occurrences of corresponding time stamps PTSaud and PTSvid;

means (219) for calculating the difference between values of corresponding time stamps PTSaud and PTSvid and comparing such difference with said time T to generate an A/V synchronization error signal; and

means (215, 216) responsive to said A/V synchronization error signal for synchronizing the decompressed audio and video signals.

This Sheet Has Been Replaced





6. The apparatus set forth in claim 5 characterized in that said means for determining the time T comprises means including a counter for counting cycles of said local clock signal and providing a value representing said time T which is equal to the number of cycles of said clock signal occurring between the occurrences of said corresponding time stamps PTSaud and PTSvid.

7. The apparatus set forth in claim 6 characterized in that said means including a counter comprises:

a counter responsive to said local clock signal for providing a sequence of count values modulo N where N is an integer; and

means responsive to an occurrence of said time stamps PTSaud for capturing a first count value and responsive to the occurrence of a corresponding time stamp PTSvid for capturing a second count value;

a subtractor for determining the difference between said first and second count values.

8. The apparatus set forth in claim 5 characterized in that said recovered compressed audio and video signal includes system clock references (SCR's), and said apparatus further includes means responsive to said system clock references for generating said local clock signal synchronous with said encoder system clock.

9. The apparatus set forth in claim 5 characterized in that said respective decompression apparatus provide decompressed video signals in predetermined intervals and decompressed audio signal in further predetermined intervals, and said means for synchronizing includes apparatus responsive to said A/V synchronization error signal for skipping or repeating ones of said predetermined intervals of said decompressed audio signals.

This Sheet Has Been Replaced





10. The apparatus set forth in claim 9 characterized in that said decompression apparatus for providing decompressed audio signal includes a source of a processing clock signal for operating said decompression apparatus and said means for synchronizing includes further apparatus responsive to said A/V synchronization error signal for varying the frequency of said processing clock signal.

11. The apparatus set forth in claim 10 characterized in that said apparatus for varying the frequency of said processing clock signal includes a binary rate multiplier.

12. The apparatus set forth in claim 7 characterized in that said respective decompression apparatus provide decompressed video signals in predetermined intervals and decompressed audio signal in further predetermined intervals, and said means for synchronizing includes apparatus responsive to said A/V synchronization error signal for skipping or repeating ones of said predetermined intervals of said decompressed audio signals.

13. The apparatus set forth in claim 12 characterized in that said decompression apparatus for providing decompressed audio signal includes a source of a processing clock signal for operating said decompression apparatus and said means for synchronizing includes further apparatus responsive to said A/V synchronization error signal for varying the frequency of said processing clock signal.

This Sheet Has Been Replaced





14. Apparatus for providing synchronized, reproduced audio and video signals from recovered compressed audio and video signal, which audio and video signals include respective time stamps PTSaud and PTSvid determined at predetermined intervals and related to an encoder system clock, said apparatus comprising respective decompression apparatus for decompressing said compressed audio and video signals and producing said time stamps PTSaud and PTSvid; and means responsive to the occurrence of said time stamps PTSaud and PTSvid for measuring the time between an occurrence of a PTSaud and the next occurring PTSvid and adjusting the decompression rate of at least one of said decompression apparatus such that the time between an occurrence of a PTSaud and the next occurring PTSvid substantially equals the difference in values of said PTSaud and the next occurring PTSvid.

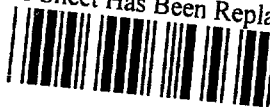
This Sheet Has Been Replaced



ABSTRACT OF THE DISCLOSURE

A receiver (FIGURE 2) for decoding associated compressed video and audio information components transmitted in mutually exclusive "frames" of data with respective time stamps PTS_{vid} and PTS_{aud} respectively, includes a controller (216) which is responsive to the respective received time stamps to provide coarse synchronization by delaying or skipping respective frames of one or the other of the components to approximately time align the two components. Fine synchronization is provided by adjusting the processing or clock frequency (215) of the audio signal processor (212) independent of the video processor (214). The control for the frequency adjustment is related to the difference between audio and video time stamps.

This Sheet Has Been Replaced



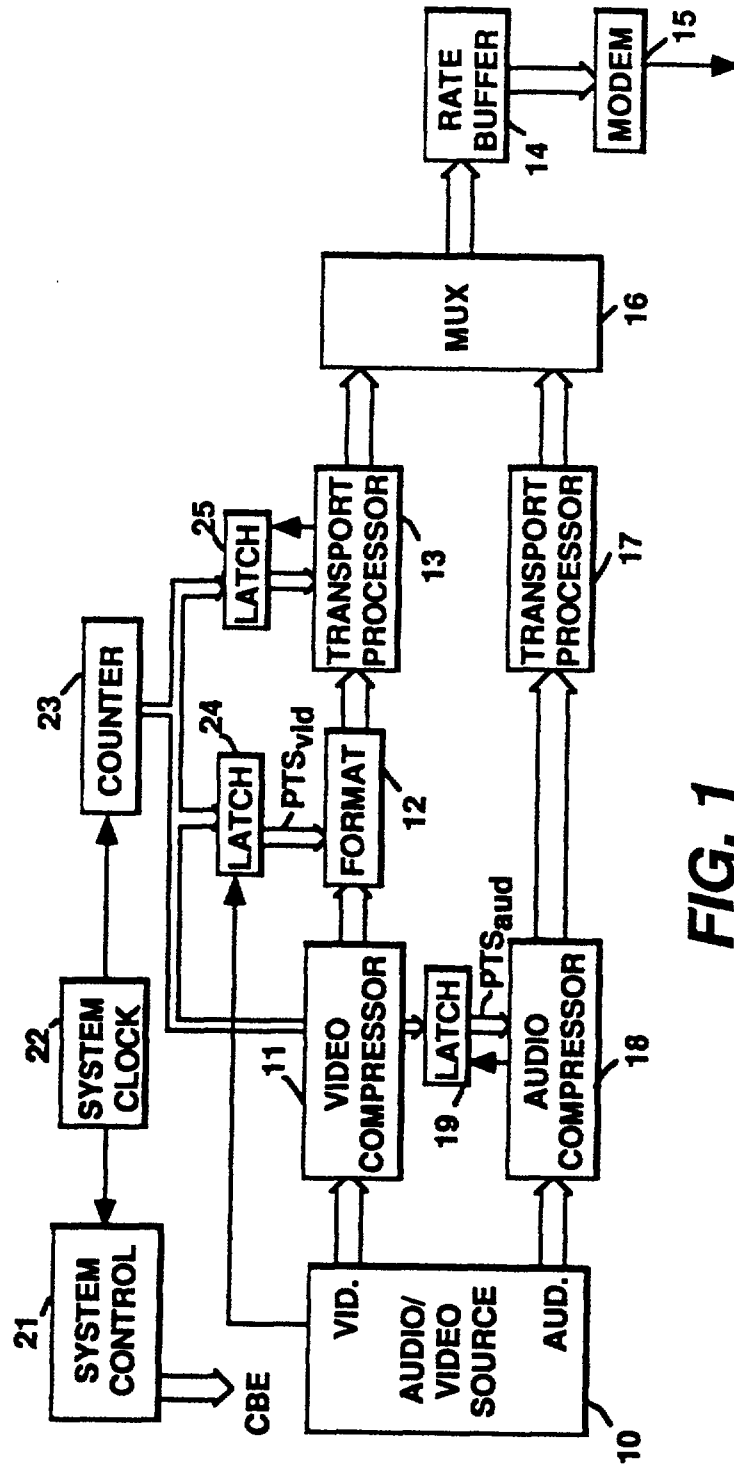
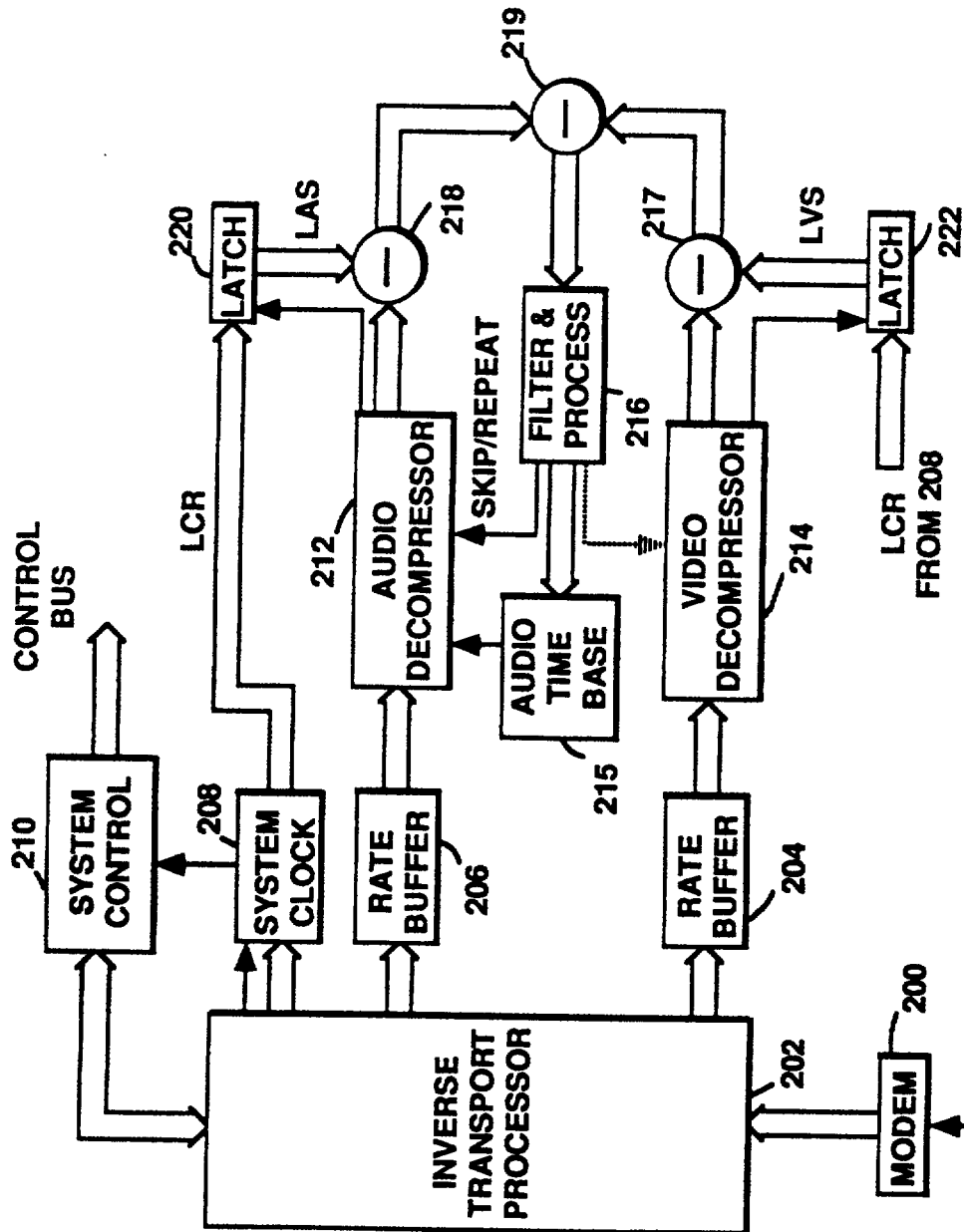
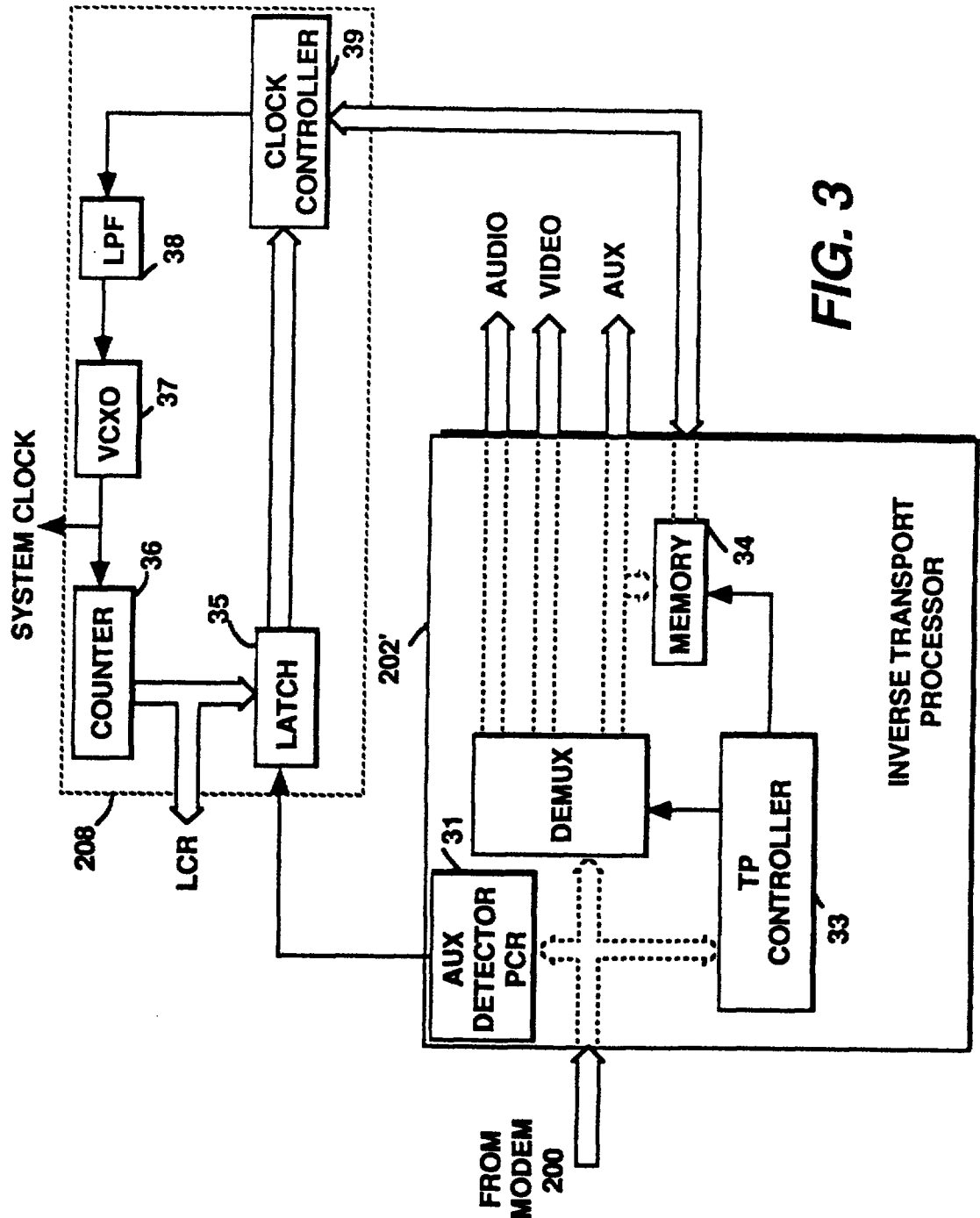
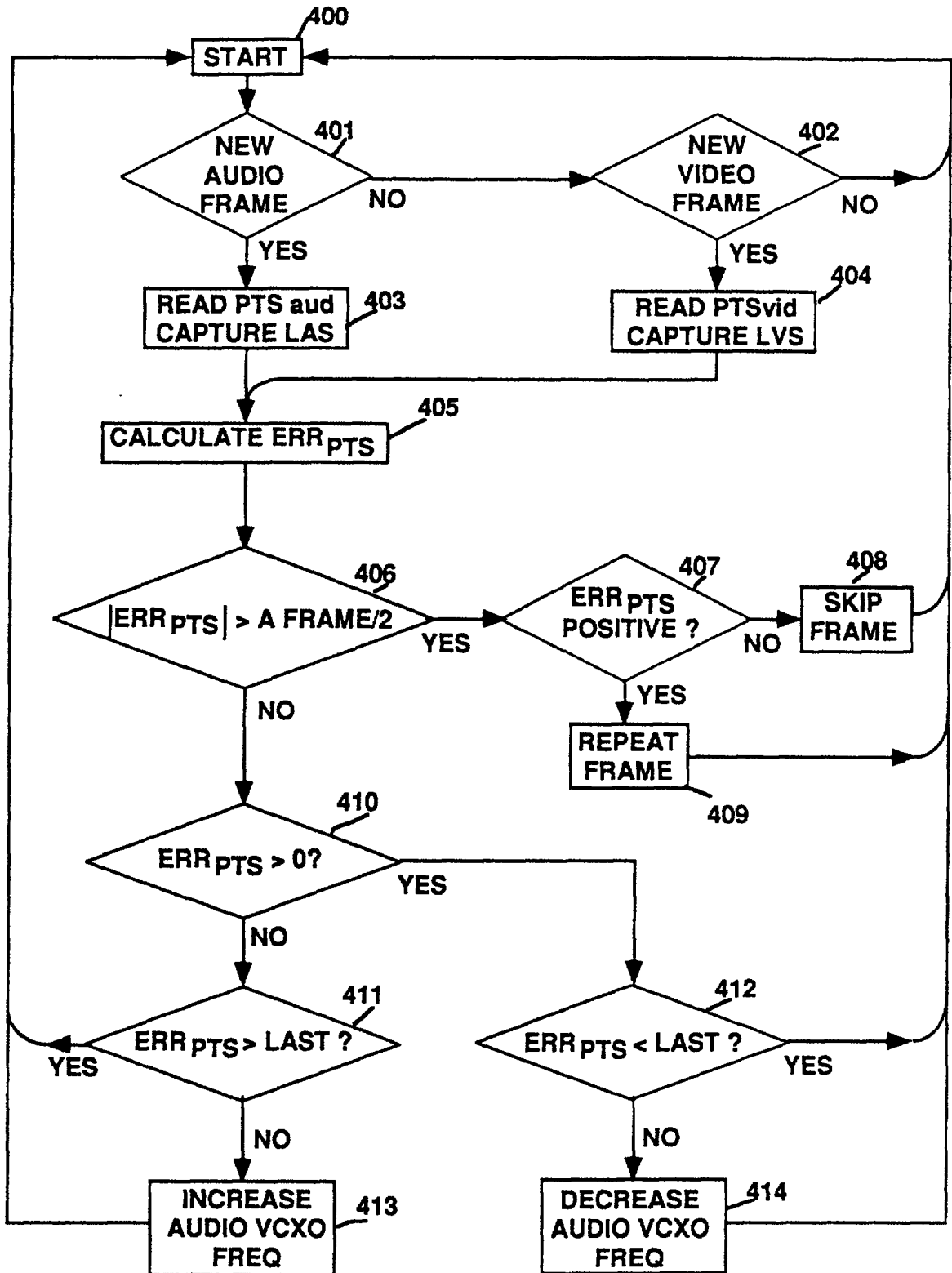


FIG. 1

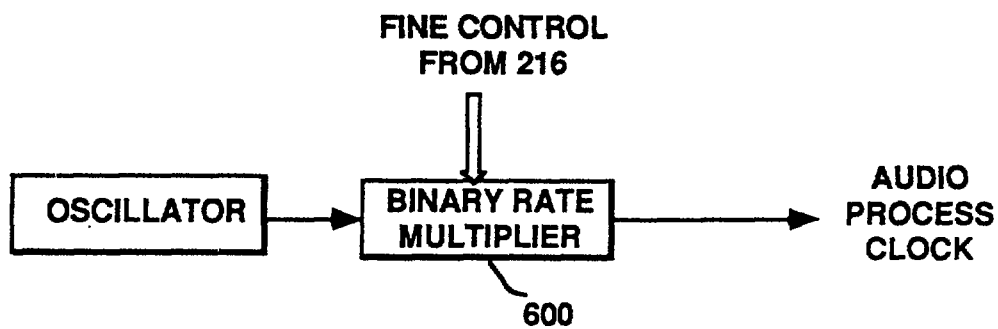
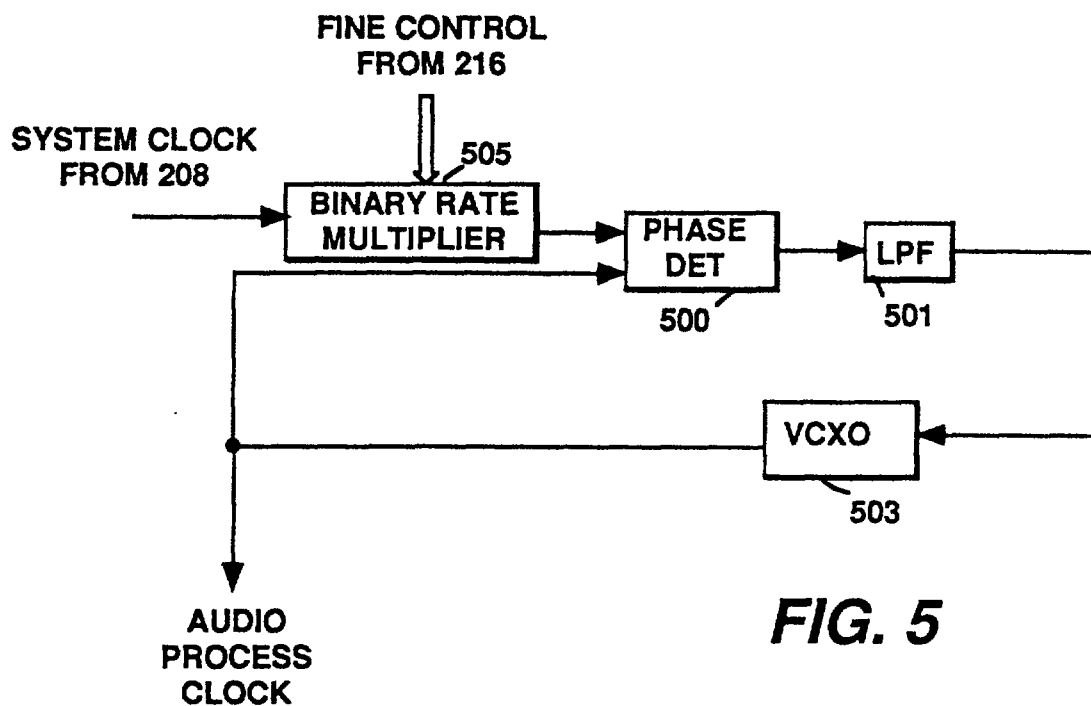




**FIG. 4**

This Sheet Has Been Replaced





This Sheet Has Been Replaced



REPLACEMENT SHEET IN
✓ Description : pages 1. to page 13
✓ (Claims) : pages 14. to page 17
✓ Drawing(s) : pages 18. to page 22.
✓ Abstract : pages 23. to page 24.

- 1 -

REPLACEMENT SHEET

30 JUL 1998

DATE OF RECEIPT

30/7/98

**AUDIO/VIDEO SYNCHRONIZATION IN
A DIGITAL TRANSMISSION SYSTEM**

This invention relates to a method and apparatus for
synchronizing the audio and video component signals at an
5 audio/video signal decompression apparatus

Audio and video (A/V) transmission and reception
requires that the audio and video components be properly
synchronized. EIA standards RS-250-B limit the time differential
between associated audio and video signals to 25 ms. lead or 40
10 ms. lag time. Film standards limit the time differential of
associated audio and video to $\pm 1/2$ frame, which corresponds to
20.8 ms. An acceptable goal for source to viewer A/V
synchronization is therefore a time differential of ± 20 ms.

Digital communication systems typically
15 time-multiplex associated signal components over a single
channel. Such multiplexing is common among A/V transmission
systems proposed and implemented for cable, fiber, terrestrial
and satellite applications. The time multiplexing of the signal
components may destroy their natural time relationships between
20 the transmission and display of the information. Therefore, time
critical components of the transmitted component signals may be
associated with a time reference before being multiplexed. This is
referred to as "stamping" the information, and timing samples are
referred to as time stamps. The receiver may then output the
25 respective components in time relative to their respective time
stamps. However to accomplish this the receiver must maintain a
very precise local time reference, which is synchronous to the
encoder time reference.

One reason that the receiver must be tightly coupled
30 to the time base of the transmitter is to insure that the output of
real time data matches that of the input to the receiver. If the
receiver provides (displays) the data too rapidly, buffers in the
receiver may underflow resulting in an interruption of the output
signal. If the receiver outputs the data too slowly, the buffers
35 may overflow (assuming finite rate buffers) resulting in a loss of
data.

In one proposed system the receiver is synchronized
to the transmitter by supplemental time stamps (system clock
references, SCR) associated with predetermined packets of

transmitted information. The timing of the capture of the time stamps, SCR, bear no relation to presentation time stamps (PTS) which are related to the video data, other than by virtue of being derived from the same counter. The SCR codes are generated by
5 sampling a modulo 2^N counter ($N \geq 32$) which counts a substantially constant frequency crystal clock at the transmitter. The receiver incorporates a phase locked loop which has a free running frequency substantially equal to the frequency of the clock in the transmitter. The receiver clock (local clock) is also
10 counted modulo 2^N , and each time a SCR arrives at the receiver the local counter is sampled to provide a local clock reference or LCR. No attempt is made to force the LCR to equal the SCR. Rather the local clock is adjusted based upon processing changes in the difference between the LCR and SCR time stamps. An error signal
15 is generated according to the relation

$$ERR = |SCR_n - SCR_{n-1}| - |LCR_n - LCR_{n-1}|$$

The signal ERR is utilized to control the local clock frequency. Via this process the LCR can be made arbitrarily close to the transmitter clock frequency. Note that since both the system and
20 local clocks are counting modulo N, they periodically wrap around. On these occurrences the respective terms $SCR_n - SCR_{n-1}$ and $LCR_n - LCR_{n-1}$ will be negative and erroneous. The system monitors the polarity of the respective differences and when one of the differences is negative the difference is ignored.

25 Video signal coded according to the MPEG standard includes presentation time stamps, PTS_{vid} , which are synchronized to the input video frames. The respective PTS_{vid} indicate the relative times that the respective frames are to be displayed at the receiver, nominally 30 Hz, for NTSC source material.
30 Associated audio is also encoded with presentation time stamps PTS_{aud} based on the same time base as the system time and which time stamps are placed in an MPEG system packet layer encompassing the encoded audio data. An audio system packet layer may contain several "frames" of audio data and respective
35 frames equal, in this example, 24 ms. of original audio data. Audio frames are approximately six times the duration of a (127 byte) transport packet. (Information to be transmitted, audio, video, data, etc. are segmented into respective transport packets

of predetermined size, with a variety of control words appended, to provide an extra layer of error correction/detection and synchronization.) In addition, according to the MPEG protocol, the number of audio frames per MPEG system layer is a variable.

5 Hence there may be little or no correlation between the video PTS_{vid} and audio PTS_{aud} presentation time stamps for associated audio and video source material. Thus synchronizing the audio and video components is difficult if one attempts to do so by comparing the PTS_{vid} with the PTS_{aud} . It is an object of the
10 present invention to simplify the process of synchronizing associated audio and video components.

A receiver for decoding associated video and audio information components transmitted in mutually exclusive "frames" of data with respective time stamps PTS_{vid} and PTS_{aud}
15 respectively, includes a controller which is responsive to the respective received time stamps to provide coarse synchronization by delaying or skipping respective frames of one or the other of the components to approximately time align the two components. Fine synchronization is provided by adjusting the processing or
20 clock frequency of the audio signal processor independent of the video processor. The control for the frequency adjustment is related to the difference between audio and video time stamps.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGURE 1 is a block diagram of audio/video
25 compression apparatus.

FIGURE 2 is a block diagram of a audio/video decompression apparatus embodying the present invention.

FIGURE 3 is a block diagram of apparatus for providing the receiver system clock signal having substantially the same
30 rate as the system clock of the compression apparatus.

FIGURE 4 is a flow chart of the operation of the FIGURE 2 apparatus.

FIGURES 5 and 6 are block diagrams of alternative audio processing clock signal generation which may be
35 implemented for element 215 in FIGURE 2.

FIGURE 1 illustrates a typical system in which the invention may be practiced, which system is a compressed digital video signal transmission arrangement. In this system, video

30 JUL 1996

- 4 -

signal from a source 10 is applied to a video signal compression element 11 which may include a motion compensated predictive encoder utilizing discrete cosine transforms. Compressed video signal from the element 11 is coupled to a formatter 12. The
5 formatter arranges the compressed video signal and other ancillary data according to some signal protocol such as MPEG, a standard developed by the International Organization for Standardization (Organisation Internationale De Normalisation). The standardized signal is applied to a transport processor 13,
10 which divides the signal into packets of data and adds certain overhead to provide some noise immunity for the transmission purposes. The transport packets, which normally occur at a non-uniform rate are applied to a rate buffer 14 which provides output data at a relatively constant rate conducive to efficient use
15 of a relatively narrow bandwidth transmission channel. The buffered data is coupled to a modem 15 which performs the signal transmission.

A system clock 22 provides clocking signal to operate much of the apparatus, at least including the transport processor.
20 This clock will operate at a fixed frequency such as 27 MHz for example. As shown herein, however, it is used to generate timing information. The system clock is coupled to the clock input of a counter 23 which may be arranged to count modulo 2^{30} , for example. The count values output by the counter are applied to
25 two latches 24 and 25. The latch 24 is conditioned by the video source to latch count values on the occurrence of respective frame intervals. These count values are denoted presentation time stamps, PTS's, and are included in the compressed video signal stream by the formatter 12, and are used by the receiver to
30 provide lip-synchronization of associated audio and video information. The latch 25 is conditioned by the transport processor 13 (or the system controller 21) to latch count values according to a predetermined schedule. These count values are denoted system clock references, SCR's, and are embedded as
35 auxiliary data within respective auxiliary transport packets.

Audio signal associated with the video signal from source 10 is applied to an audio signal compressor 18. The compressor 18 provides frame sampling pulses (independent of

video frames) to control a latch 19. Responsive to the sampling pulses, the latch 19 captures count values provided by the counter 23. These latched values correspond to audio presentation time stamps PTS_{aud} . The PTS_{aud} are incorporated in the compressed audio signal provided by the compressor 18. The compressed audio signal is coupled to a transport processor 17 which divides the signal into packets of data and adds certain overhead to provide some noise immunity for transmission purposes. The audio transport packets provided by the processor 17 are coupled to a multiplexor 16 which time division multiplexes the audio and video transport packets. In the figure, separate transport processors are shown in the audio and video signal processing channels. For systems wherein the data rate is moderate, the functions of the two transport processors and the multiplexor 16 may be subsumed in a single transport processor.

The system controller 21 is a variable state machine programmed to coordinate the various processing elements. Note that the controller 21, the compressors 11 and 18, the transport processors 13 and 17, and the rate buffer 14 may or may not operate synchronously via a common clocking arrangement as long as proper handshaking is provided between processing elements. However, the two compressors both derive PTS values from the same reference counter 23, thus a precise timing relationship between the two compressed signals is provided in the compressed output signal.

FIGURE 2 illustrates an exemplary receiver apparatus embodying the invention wherein the modem 200 performs the inverse function of the modem 15 and the rate buffers 204 and 206 perform, in effect, the inverse function of the rate buffer 14. In FIGURE 2 a single inverse transport processor 202 is indicated, which processor divides respective transport packets by service and allocates the respective data to the appropriate processing channel. In so doing the respective transport packet signal payloads are separated from the auxiliary data, with the respective payloads being applied to the appropriate processing channel and the auxiliary data applied to the system controller 210. In an alternative arrangement a separate transport processor may be included in each processing channel and

20 JUL 1996

- 6 -

arranged to recognize and process only data associated with the respective channels.

Compressed video data from the inverse transport processor 202 is applied to the rate buffer 204, which provides
5 compressed video signal according to the system protocol to the decompressor 214. The rate buffer 204 accepts data at a bursty or inconsistent rate, and provides data on demand to the decompressor 214. The decompressor, responsive to the compressed video signal, generates non-compressed video signal
10 for display or for storage etc., in appropriate display or storage devices (not shown).

Compressed audio data from the inverse transport processor 202 is applied to the rate buffer 206, which provides compressed audio signal according to the system protocol to the
15 audio decompressor 212. The decompressor, 212, responsive to the compressed audio signal generates non-compressed audio signal for reproduction or for storage etc., in appropriate speakers or storage devices (not shown).

The inverse processor 202 also provides SCR's from
20 the auxiliary transport data, and control signals to a system clock generator 208. The clock generator responsive to these signals generates a system clock signal synchronous with at least the transport processor operation. This system clock signal is applied to the receiver system controller 210 to control the timing of
25 appropriate processing elements.

Figure 3 illustrates details of an exemplary clock regenerator 208. Data from the receiver modem 200 is coupled to the inverse transport processor 202', including an auxiliary packet detector 31. The inverse transport processor 202'
30 separates transport header data from the respective transport packet payloads. Responsive to the transport header data, the processor 202' applies audio and video signal payloads to, for example, respective decompression apparatus (not shown), and auxiliary data (designated as AUX) to the appropriate auxiliary
35 data processing elements (not shown). The SCR's residing within the auxiliary data are routed and stored in a memory element, 34.

The auxiliary packet detector 31, which may be a matched filter arranged to recognize the codewords designating an

auxiliary transport packet containing a SCR, produces a control pulse on the occurrence of transport packets containing such data. The control pulse is utilized to capture and store within a latch 35, the count value currently exhibited by the local counter 36 at a time precisely related to the time of detection. The local counter 36 is arranged to count pulses provided by a voltage controlled oscillator 37. The counter 36 is arranged to count modulo M which may be, but is not necessarily, the same number as its counterpart counter in the encoder (counter 23). If M differs from N the difference may be accommodated in the error equation.

The voltage controlled oscillator 37 is controlled by a low pass filtered 38 error signal provided by a clock controller 39. The error signal is generated in the following fashion. Let the SCR arriving at time n be designated SCR_n and the local count value concurrently captured in the latch 35 be designated LCR_n . The clock controller reads the successive values of SCR's and LCR's and forms an error signal E proportional to the differences

$$E \Rightarrow |SCR_n - SCR_{n-1}| - |LCR_n - LCR_{n-1}|$$

The error signal E, is utilized to condition the voltage controlled oscillator 37 to a frequency tending to equalize the differences. As indicated previously, negative differences exhibited due to modulo counter wrap around, may be ignored. The error signal produced by the clock controller 39 may be in the form of a pulse width modulated signal, which may be rendered into an analog error signal by implementing the low pass filter 38 in analog components.

The constraints on this system are that the counters at the two ends of the system count the same frequency or even multiples thereof. This requires that the nominal frequency of the voltage controlled oscillator be fairly close to the frequency of the system clock at the encoder.

The foregoing approach provides rather rapid synchronization but may introduce a long term error. Long term error LTE is proportional to the difference

$$LTE \Rightarrow |LCR_n - LCR_0| - |SCR_n - SCR_0|$$

where SCR_0 and LCR_0 are for example the first occurring SCR and the corresponding latched value of the receiver counter.

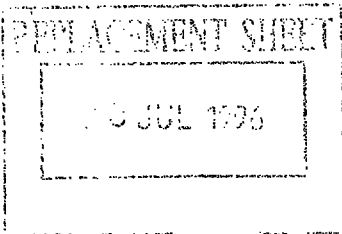
Nominally the error signals E and LTE will vary in discrete steps. As such, once the system is "synchronized" the error signal will dither one unit about the null point. The preferred method of synchronization is to initiate control of the voltage controlled oscillator using the error signal E until a one unit dither occurs in the error signal E, and then to switch to the use of the long term error signal LTE to control the voltage controlled oscillator.

The system clock signal provided by the VCXO 37 may be utilized to operate at least the transport processor and rate buffers. Since it is synchronized at least in frequency with the encoder system clock, the possibility of rate buffer overflow or underflow due to clock timing errors is substantially non-existent.

Refer again to FIGURE 2 for an explanation of audio/video synchronization. Recall that a presentation time stamp PTS_{vid} , is included in the compressed video signal associated with predetermined video data. The PTS_{vid} is indicative of the relative time that the associated video is to be displayed. Similarly the compressed audio signal includes presentation time stamps PTS_{aud} associated with audio to be reproduced at times related to the respective PTS_{aud} . At the receiver the PTS_{aud} and PTS_{vid} may not be compared directly to provide A/V synchronization because the respective samples were determined at different instants. The respective PTS values are compared to a continuous time base which is the receiver clock provided by the VCXO 37. This is done by sampling local time stamps of the count values LCR.

When data associated with an associated PTS is presented, The LCR is sampled. For example, The audio decompressor 212 issues a PTS_{aud} when a respective audio frame is output for reproduction. At these times a control signal conditions the latch 220 to sample the LCR, the values of which will be designated LAS, for local audio stamp. Similarly when the video decompressor provides a video frame for display, it provides a PTS_{vid} and a control pulse to condition a latch 222 to store the current value of the LCR. These LCR values are designated LVS for local video stamps.

The LAS and the corresponding PTS_{aud} are coupled to respective input terminals of a subtractor 218 which develops the signal $\Delta A-PTS$ according to the relation;



$$\Delta A-PTS = PTS_{aud} - LAS$$

The LVS and the corresponding PTS_{vid} are coupled to respective input terminals of the subtractor 217 which develops the signal $\Delta V-PTS$ according to the relation;

5
$$\Delta V-PTS = PTS_{vid} - LVS.$$

The signals $\Delta V-PTS$ and $\Delta A-PTS$ are coupled to respective input terminals of a further subtractor, 219, which develops an A/V synchronization error signal ERR_{PTS} according to the relation;

$$ERR_{PTS} = \Delta V-PTS - \Delta A-PTS$$

10

Synchronization of the audio and video requires that the A/V synchronization error be driven to zero. This indicates that when the difference in the values of the corresponding audio and video PTS's equals the time, in units of the local reference, between the
15 occurrence of the corresponding PTS's, the audio and video signal will be in synchronization.

Two mechanisms may be used to adjust the A/V synchronization based upon the error signal ERR_{PTS} ; skips and repeats of data sections and conversion clock deviation. Skipping
20 fixed intervals or "frames" of audio advances the audio data stream by a fixed interval relative to the video signal. Repeating (or muting without consuming data) delays the audio data stream by fixed intervals relative to the video signal. Skipping and repeating audio frames is audible under many conditions, and
25 therefore is only utilized for coarse adjustment of synchronization. Even so, brief skipping or repeating may be preferable to discernible audio/video synchronization errors. If the audio frames are less than 40 msec, coarse adjustment by skipping/repeating may result in synchronization errors within
30 ± 20 msec., which is within industry standards for A/V synchronization. However, this synchronization will degrade if the audio conversion time base does not match that of the source. Once synchronization is coarsely adjusted, variations of the audio conversion clock frequency are arranged to further refine the A/V
35 synchronization.

The error signal ERR_{PTS} is applied to a filter and processing element 216. A filter function therein smoothes the signal ERR_{PTS} to minimize aberrant effects that might otherwise

be generated by signal noise. The processing function in the element 216 then examines the smoothed error signal and determines whether a skip/repeat of audio should be utilized to effect coarse synchronization of the audio and video signals and/or whether an adjustment to the audio processing frequency should be utilized to effect fine synchronization. If a coarse synchronization adjustment is determined to be necessary, the processor 216 provides a control signal to the audio decompressor 212 to condition the decompressor to skip or repeat the current decompressed audio frame. Alternatively, or in addition to the coarse adjustment, if a fine adjustment is determined to be necessary, the processor 216 provides a control signal to the audio time base 215 to adjust the frequency of the audio processing clock signal.

The processing algorithm is detailed in the flow chart of FIGURE 4. After initialization 400 of the system, which is designated START, the system monitors 401 the audio decompressor for the occurrence of a PTS_{aud} and if a PTS_{aud} is detected it is read 403 and a local clock reference LAS is captured and stored. If a PTS_{aud} has not occurred, the system monitors the video compressor for a PTS_{vid} 402. If a PTS_{vid} has occurred, the PTS_{vid} is read and a local clock reference LVS is captured and stored 404. When both a PTS_{aud} and a PTS_{vid} have been read the ERR_{PTS} is calculated 405 according to the equation;

$$ERR_{PTS} = \Delta V - PTS - \Delta A - PTS.$$

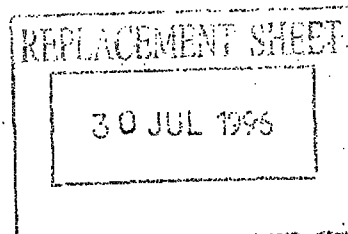
The magnitude of the error signal is examined 406 to determine if it is greater than one half an audio frame interval. If it is greater than one half an audio frame interval, the error signal is checked for polarity 407. If the polarity is positive, the current audio frame is repeated 409. If it is negative the current audio frame is skipped 408. After skipping or repeating a frame the system iterates back to the start position to wait for the next occurrence of PTS's.

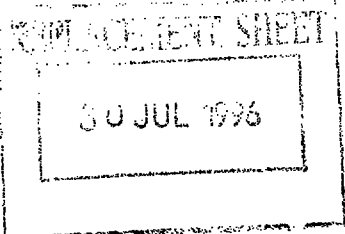
At step 406, if the magnitude of the error signal is less than one half an audio frame interval, the error is examined to determine if it is greater than zero 410. If the error is greater than zero, the error is checked 412 to determine if it is less than the previous error signal. If it is less than the previous error

signal, this is an indication that the system is converging toward synchronization, and the synchronization control parameters are not changed. The system returns to the start position to wait for the next PTS's. Conversely, if the error has increased over the
5 previous error signal, the audio system processing clock is adjusted to lessen its frequency 414.

At step 410, if the error is less than zero (negative), it is checked 411 to determine whether it is greater than the previous error signal. If it is greater than the previous error signal, this is
10 also an indication that the system is converging toward synchronization, and the synchronization control parameters are not changed. Alternatively if the current error signal is less than the previous error signal the system is moving further out of synchronization and the audio processing clock frequency is
15 increased 413. After processing steps 412 and 413 the system returns to wait for the next occurrence of PTS's. It will be noted in this example, that the system performs only coarse adjustments via skipping or repeating audio frames until the A/V synchronization error is reduced to less than one half an audio
20 frame interval.

In an alternative embodiment, the filtered error signal is compared against a predetermined threshold related to the size of respective audio frames. If the error signal is less than the threshold, indicative that the audio-video timing error is less than
25 an audio frame, the error signal is coupled to the audio time base circuit 215, wherein it is utilized to adjust the frequency of the audio signal processing (decompression) clock. Alternatively, if the error signal is greater than the threshold, the error signal may be divided by the audio frame interval to determine the number
30 of audio frames that the audio and video signal are misaligned. The integer portion of the quotient is applied to the audio decompressor to condition the audio decompressor to skip or repeat that number of audio frames. The polarity of the error signal will determine whether audio frames should be skipped or
35 repeated. Nominally the decompressed data is arranged in a buffer memory prior to being output, hence skipping or repeating audio frames is a simple matter of controllably enabling the memory read/write commands.





The fractional portion of the quotient is coupled to the audio time base circuit 215, where it is utilized to adjust the audio processing clock to fine tune the A/V synchronization.

5 The rate of production of audio PTS's is proportional to the processing speed of the audio decompressor. The processing speed of the audio decompressor is directly proportional to the frequency of the clock signal used to operate the audio decompressor. If the clock frequency of the audio decompressor is independent of the clock used to operate the video
10 decompressor, and is finely adjustable then the relative rate of occurrence of the audio and video PTS's can be adjusted and the A/V finely synchronized.

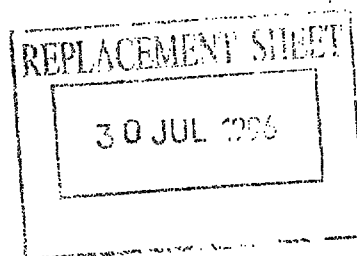
FIGURE 5 is a first embodiment of circuitry for generating an adjustable audio processing clock signal. In FIGURE 5 a voltage
15 controlled oscillator, 503, is connected in a phase locked loop including elements 500 and 501. The output of the oscillator is coupled to one input of a phase detector 500. The system clock is coupled to a second input of the phase detector 500 via a binary rate multiplier 505. A phase error signal developed by the phase
20 detector is low pass filtered in filter 501. This smoothed phase error signal is coupled to the control input terminal of the oscillator to condition the oscillator to oscillate at the same frequency and phase as the output of the binary rate multiplier. In this example the system clock frequency is approximately 27
25 MHz and the desired audio processing clock frequency is approximately 1/380 of the 27 MHz signal. A control signal from the process element 216 is coupled to a control input port of the binary rate multiplier to control its output frequency. This control signal is selected to represent a nominal division ratio of 1/380,
30 but modulates about this value so as to condition the output of the binary rate multiplier to produce an output frequency commensurate with the instantaneous desired audio processing rate.

FIGURE 6 illustrates a second example of adjustable
35 audio processing clock generation. In this example, a constant clock frequency, from the system clock generator 208 or some other stable oscillator, is applied to one input of a binary rate multiplier 600. The binary rate multiplier is controlled by a

control signal from the process element 216 to generate a nominal desired audio processing clock frequency. Responsive to the A/V error signal the process element 216 alters this control signal to condition the binary rate multiplier to either increase or
5 decrease the nominal audio processing clock frequency.

A further alternative audio signal processing rate changing apparatus (not shown) may include an oscillator for providing a nominal maximum audio processing clock frequency and a gating circuit. The oscillator is coupled to the audio processing circuit via
10 the gating circuit. The gating circuit is controlled by the process element 216 to excise ones of the oscillator output pulses to provide, on average, a desired audio processing clock frequency.

An alternative embodiment of the synchronizing apparatus, indicated by the dashed arrow in FIGURE 2, may be
15 arranged to skip or repeat video frames to effect synchronization. Alternatively video frames may be skipped for the audio leading (lagging) video and audio frames may be skipped for the audio frames lagging (leading) video. However, in the preferred embodiment, the audio frames are skipped/repeated for audio
20 lagging and leading video respectively.





CLAIMS:

1. Apparatus for providing synchronized, reproduced audio and video signals from recovered compressed audio and video signal, which audio and video signals include respective time stamps PTSaud and PTSvid determined at predetermined intervals and related to an encoder system clock, said apparatus characterized by:

a source (200) of said recovered compressed audio and video signals, and a source (208, 37) of reference frequency;

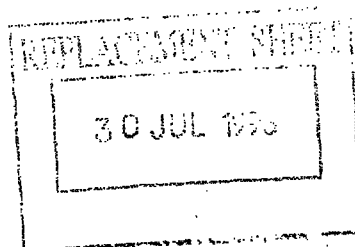
a counter (208, 36) responsive to said reference frequency for providing a sequence of values modulo N where N is an integer;

decompression apparatus (214) responsive to said recovered compressed video signal for providing decompressed video signal and said time stamp PTSvid;

decompression apparatus (212) responsive to said recovered compressed audio signal for providing decompressed audio signal and said time stamp PTSaud;

means (222, 217) for capturing values (LVS) of said sequence at predetermined times related to the provision of said time stamps PTSvid and for generating the difference ($\Delta_V\text{-PTS}$) of said values LVS and corresponding time stamps PTSvid; and (220, 218) for capturing values (LAS) of said sequence at predetermined times related to the provision of said time stamps PTSaud and generating the difference ($\Delta_A\text{-PTS}$) of said values LAS and corresponding time stamps PTSaud; and

means (219, 216, 215) responsive to said differences $\Delta_V\text{-PTS}$ and $\Delta_A\text{-PTS}$ for synchronizing the decompressed audio and video signals.

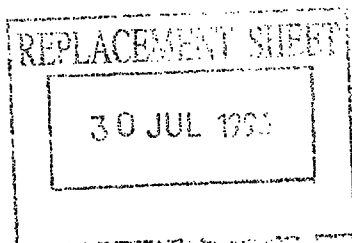




2. The apparatus set forth in claim 1 characterized in that said means for synchronizing the decompressed audio and video signals includes means (216, 212) for skipping/repeating predetermined intervals of decompressed audio signal responsive to the difference the between values $\Delta V\text{-PTS}$ and $\Delta A\text{-PTS}$ exceeding a predetermined value.

3. The apparatus set forth in claim 1 characterized in that said means for synchronizing the decompressed audio and video signals includes means (216,214) for skipping/repeating predetermined intervals of decompressed video signal responsive to the difference the between values $\Delta V\text{-PTS}$ and $\Delta A\text{-PTS}$ exceeding a predetermined value.

4. The apparatus set forth in claim 1 characterized in that said decompression apparatus (212) responsive to said recovered compressed audio signal provides decompressed audio signal at a rate determined by a processing (215) clock signal and said means for synchronizing the decompressed audio and video signals further includes means (216) responsive to the differences $(\Delta V\text{-PTS})-(\Delta A\text{-PTS})$, for variably adjusting the rate of said processing clock signal.





5. Apparatus for providing synchronized, reproduced audio and video signals from recovered compressed audio and video signal, which audio and video signals include respective time stamps PTSaud and PTSvid determined at predetermined intervals and related to an encoder system clock, said apparatus characterized by:

a source (200) of said recovered compressed audio and video signals, and a source (208) of a local clock signal;

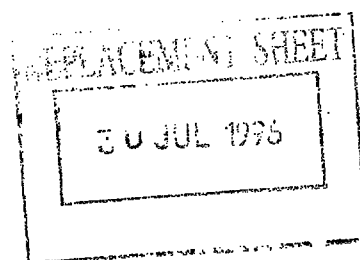
decompression apparatus (214) responsive to said recovered compressed video signal for providing decompressed video signal and said time stamp PTSvid;

decompression apparatus (212) responsive to said recovered compressed audio signal for providing decompressed audio signal and said time stamp PTSaud;

means (217, 218, 220, 222) for determining, in cycles of said local clock signal, the time T between occurrences of corresponding time stamps PTSaud and PTSvid;

means (219) for calculating the difference between values of corresponding time stamps PTSaud and PTSvid and comparing such difference with said time T to generate an A/V synchronization error signal; and

means (215, 216) responsive to said A/V synchronization error signal for synchronizing the decompressed audio and video signals.





6. The apparatus set forth in claim 5 characterized in that said means for determining the time T comprises means including a counter for counting cycles of said local clock signal and providing a value representing said time T which is equal to the number of cycles of said clock signal occurring between the occurrences of said corresponding time stamps PTSaud and PTSvid.

7. The apparatus set forth in claim 6 characterized in that said means including a counter comprises:

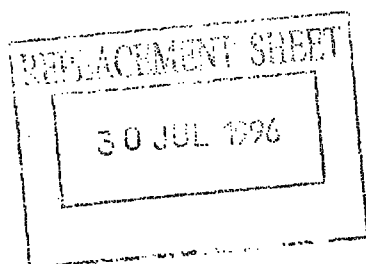
a counter responsive to said local clock signal for providing a sequence of count values modulo N where N is an integer; and

means responsive to an occurrence of said time stamps PTSaud for capturing a first count value and responsive to the occurrence of a corresponding time stamp PTSvid for capturing a second count value;

a subtractor for determining the difference between said first and second count values.

8. The apparatus set forth in claim 5 characterized in that said recovered compressed audio and video signal includes system clock references (SCR's), and said apparatus further includes means responsive to said system clock references for generating said local clock signal synchronous with said encoder system clock.

9. The apparatus set forth in claim 5 characterized in that said respective decompression apparatus provide decompressed video signals in predetermined intervals and decompressed audio signal in further predetermined intervals, and said means for synchronizing includes apparatus responsive to said A/V synchronization error signal for skipping or repeating ones of said predetermined intervals of said decompressed audio signals.



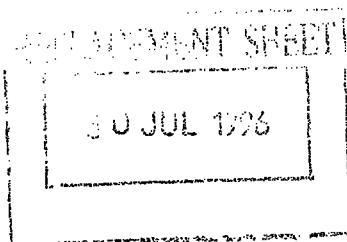


10. The apparatus set forth in claim 9 characterized in that said decompression apparatus for providing decompressed audio signal includes a source of a processing clock signal for operating said decompression apparatus and said means for synchronizing includes further apparatus responsive to said A/V synchronization error signal for varying the frequency of said processing clock signal.

11. The apparatus set forth in claim 10 characterized in that said apparatus for varying the frequency of said processing clock signal includes a binary rate multiplier.

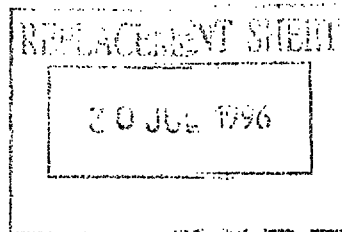
12. The apparatus set forth in claim 7 characterized in that said respective decompression apparatus provide decompressed video signals in predetermined intervals and decompressed audio signal in further predetermined intervals, and said means for synchronizing includes apparatus responsive to said A/V synchronization error signal for skipping or repeating ones of said predetermined intervals of said decompressed audio signals.

13. The apparatus set forth in claim 12 characterized in that said decompression apparatus for providing decompressed audio signal includes a source of a processing clock signal for operating said decompression apparatus and said means for synchronizing includes further apparatus responsive to said A/V synchronization error signal for varying the frequency of said processing clock signal.





14. Apparatus for providing synchronized, reproduced audio and video signals from recovered compressed audio and video signal, which audio and video signals include respective time stamps PTSaud and PTSvid determined at predetermined intervals and related to an encoder system clock, said apparatus comprising respective decompression apparatus for decompressing said compressed audio and video signals and producing said time stamps PTSaud and PTSvid; and means responsive to the occurrence of said time stamps PTSaud and PTSvid for measuring the time between an occurrence of a PTSaud and the next occurring PTSvid and adjusting the decompression rate of at least one of said decompression apparatus such that the time between an occurrence of a PTSaud and the next occurring PTSvid substantially equals the difference in values of said PTSaud and the next occurring PTSvid.



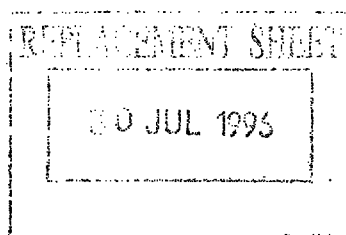
THIS SHEET HAS BEEN
REPLACED

AUDIO/VIDEO SYNCHRONIZATION IN A DIGITAL TRANSMISSION SYSTEM

ABSTRACT OF THE DISCLOSURE

A receiver (FIGURE 2) for decoding associated compressed video and audio information components transmitted in mutually exclusive "frames" of data with respective time stamps PTS_{vid} and PTS_{aud} respectively, includes a controller (216) which is responsive to the respective received time stamps to provide coarse synchronization by delaying or skipping respective frames of one or the other of the components to approximately time align the two components. Fine synchronization is provided by adjusting the processing or clock frequency (215) of the audio signal processor (212) independent of the video processor (214). The control for the frequency adjustment is related to the difference between audio and video time stamps.

Figure 2



This Sheet Has Been Replaced



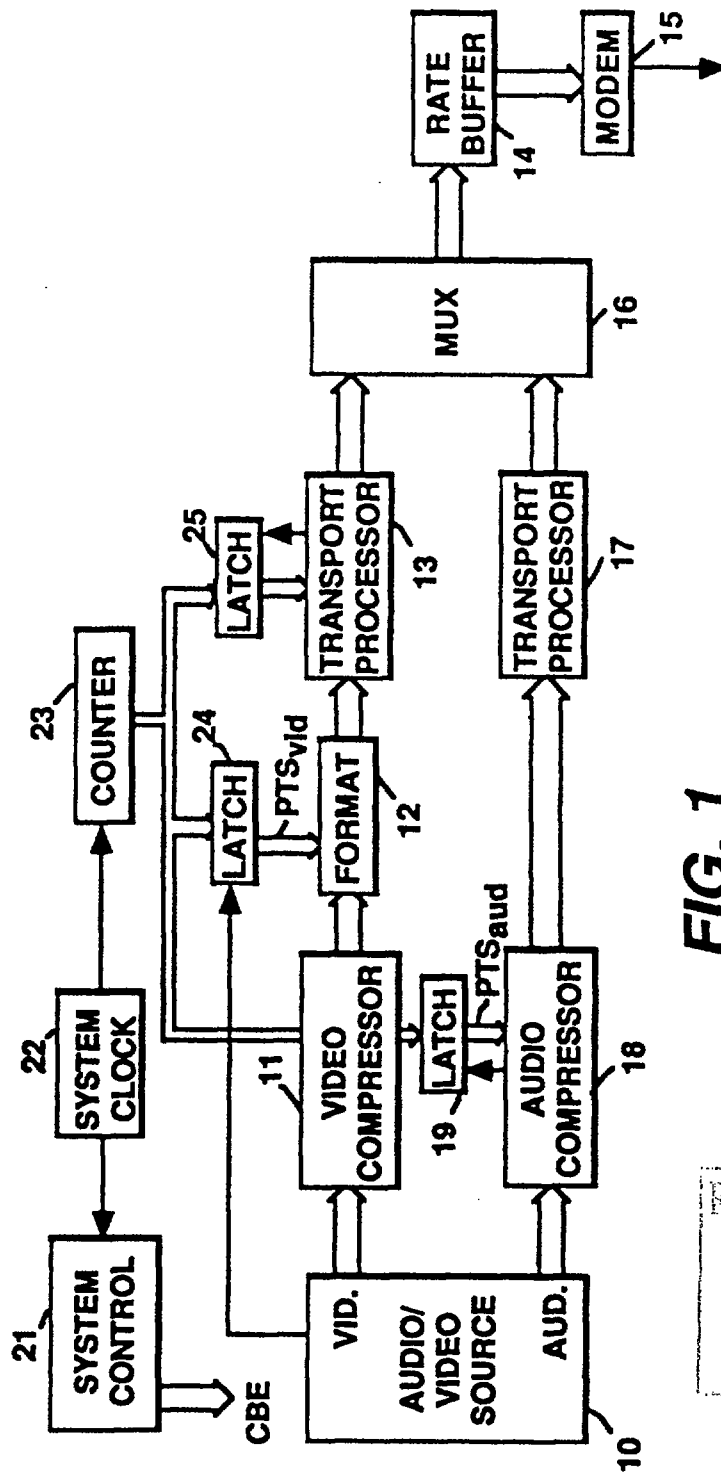
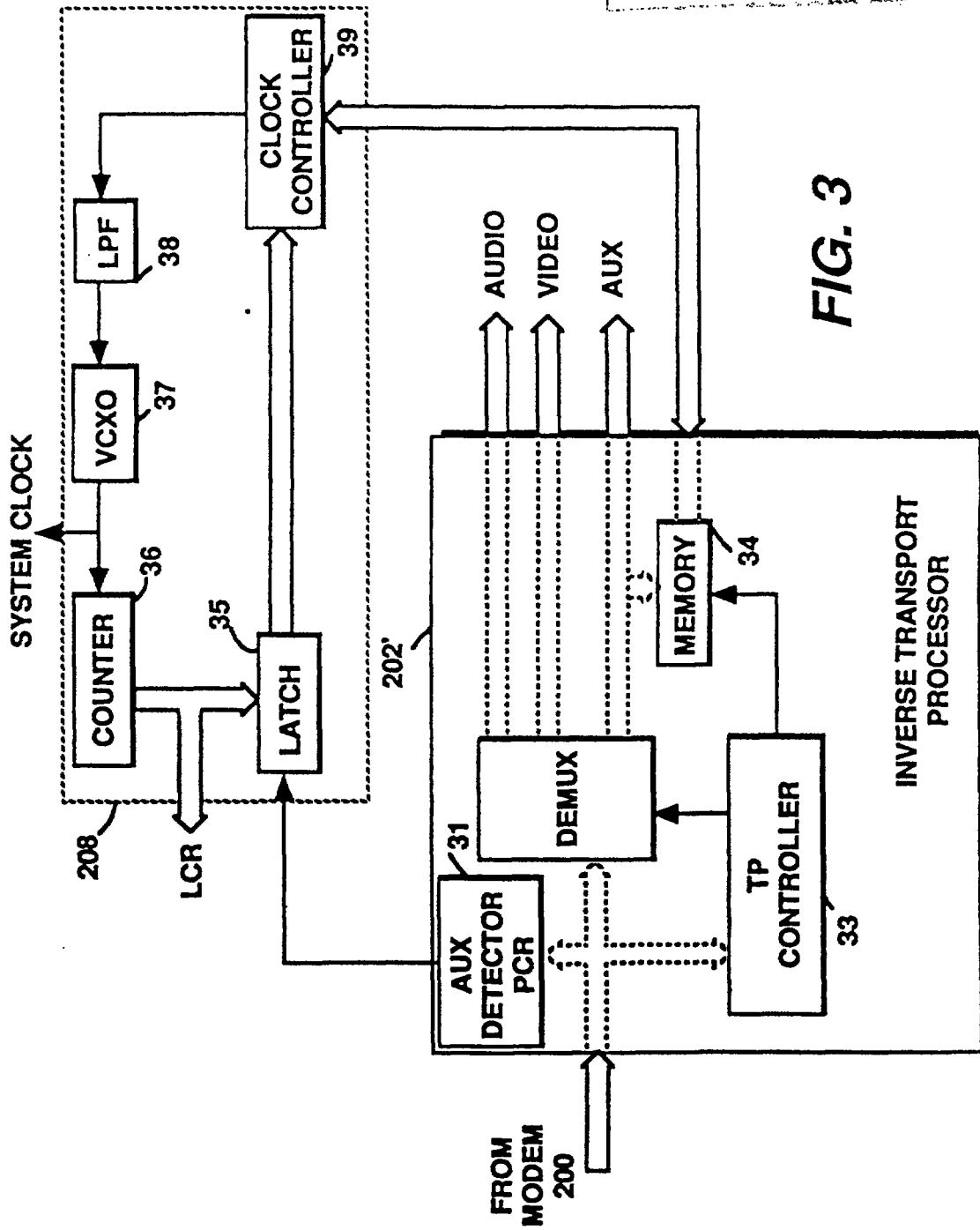
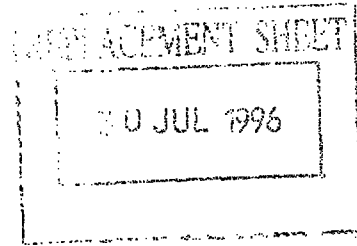


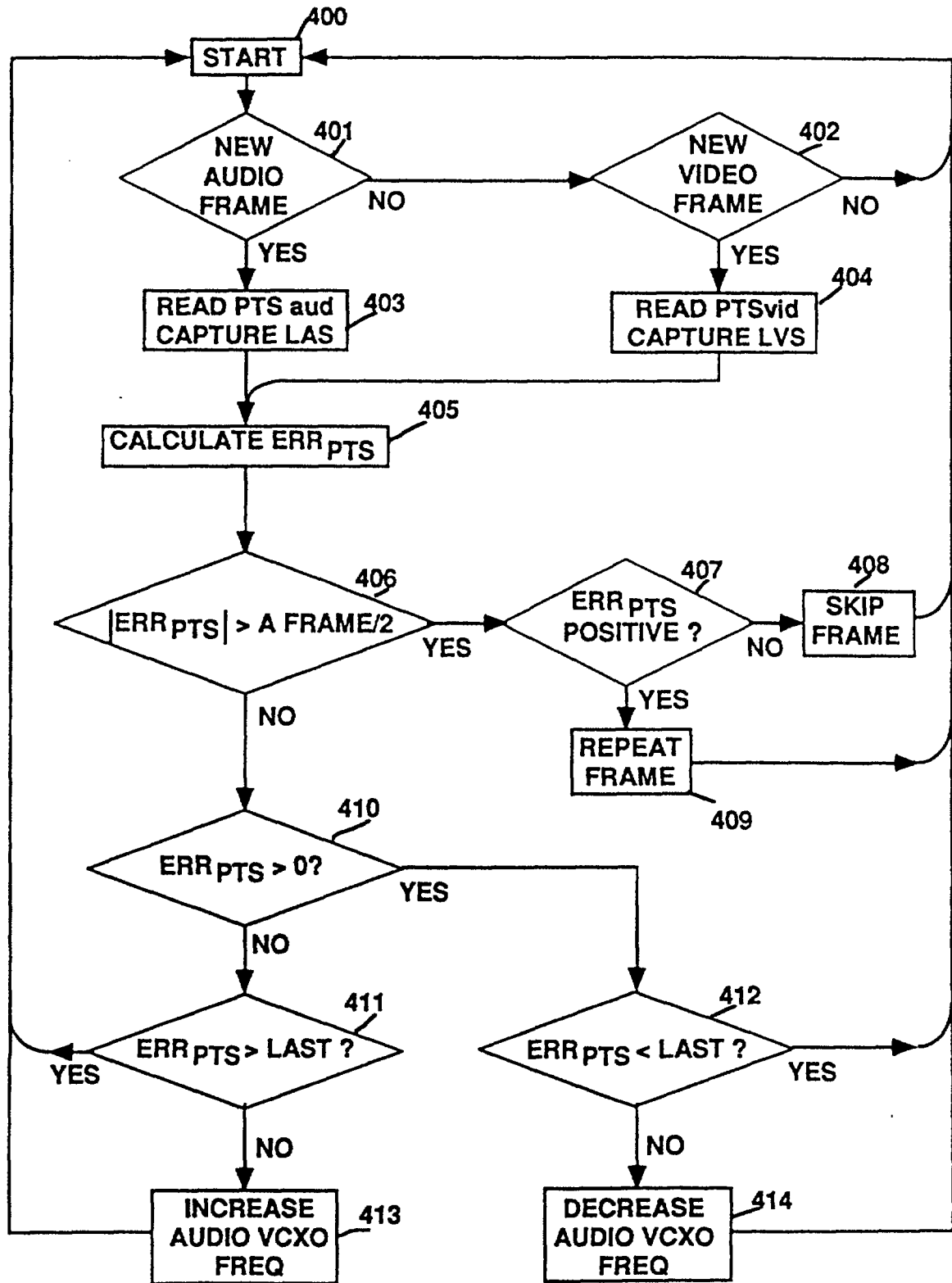
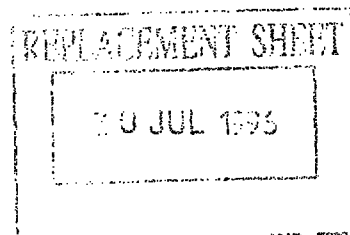
FIG. 1

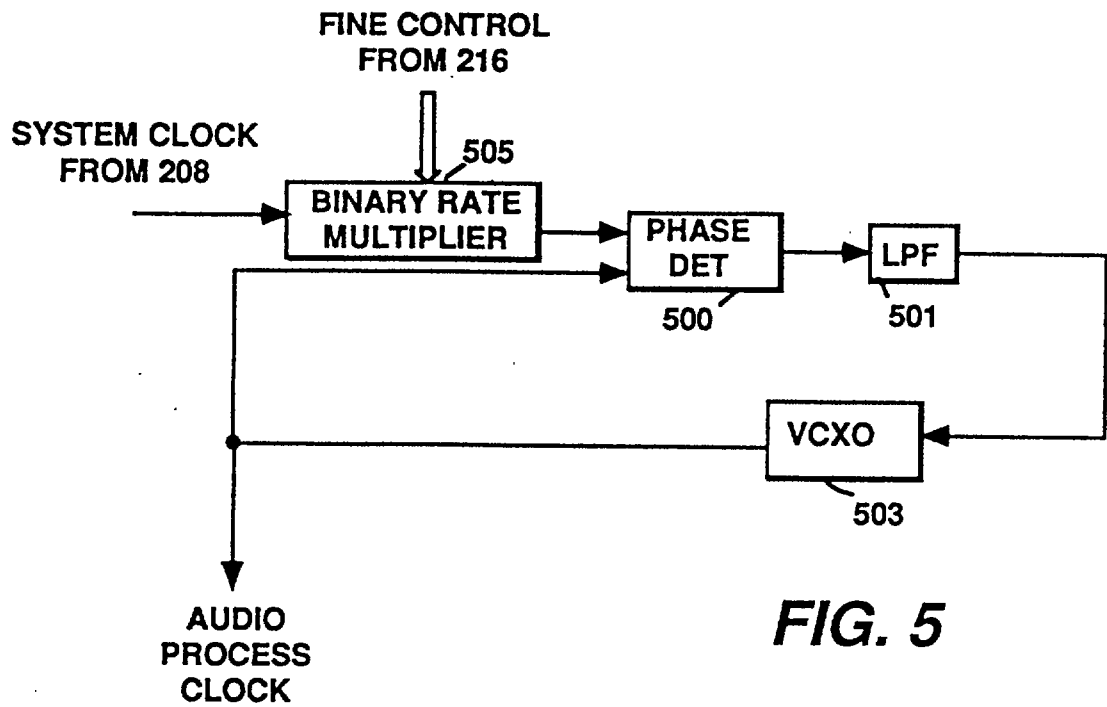
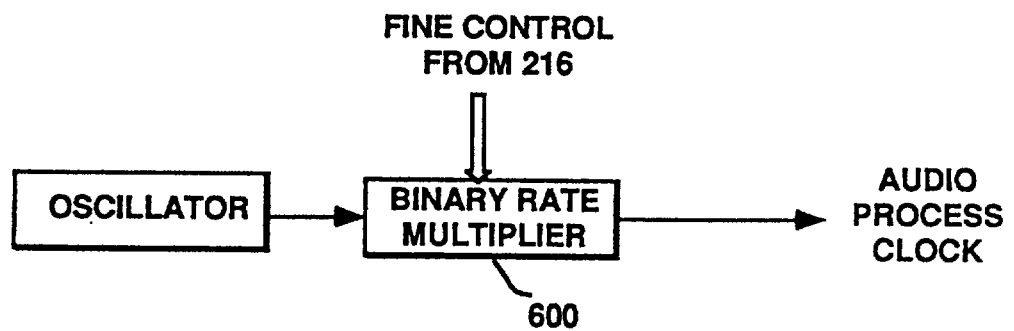
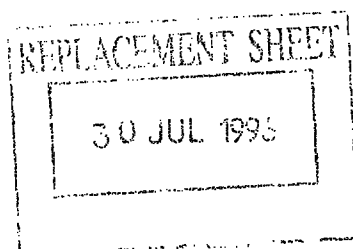
REPLACEMENT SHEET

30 JUL 1996





**FIG. 4**

**FIG. 5****FIG. 6**



EXT OF THE ABSTRACT

AUDIO/VIDEO SYNCHRONIZATION IN A DIGITAL
TRANSMISSION SYSTEM

A receiver for decoding associated compressed video and audio information components transmitted in mutually exclusive "frames" of data with respective time stamps PTS_{vid} and PTS_{aud} respectively, includes a controller (216) which is responsive to the respective received time stamps to provide coarse synchronization by delaying or skipping respective frames of one or the other of the components to approximately time align the two components. Fine synchronization is provided by adjusting the processing or clock frequency (215) of the audio signal processor (212) independent of the video processor (214). The control for the frequency adjustment is related to the difference between audio and video time stamps.

Fig. 2

