PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : H04M 11/06, H04N 7/14

A1

(11) International Publication Number: WO 93/18607

(43) International Publication Date: 16 September 1993 (16.09.93)

(21) International Application Number: PCT/GB93/00517

(22) International Filing Date: 12 March 1993 (12.03.93)

(30) Priority data:
9205580.5 13 March 1992 (13.03.92) GB

(71) Applicant *(for all designated States except US):* GEC-MARCONI LIMITED [GB/GB]; The Grove, Warren Lane, Stanmore, Middlesex HA7 4LY (GB).

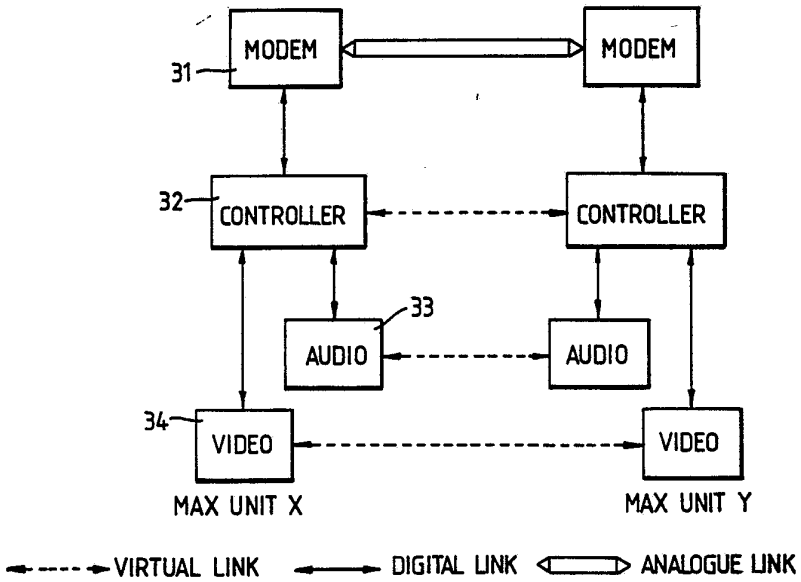
(72) Inventors; and
(75) Inventors/Applicants *(for US only) :* WILBY, W., A. [GB/GB]; 3 Sandpit Road, Welwyn Garden City, Herts (GB). HYATT, T. [GB/GB]; 5 Napier Drive, Bushey, Herts (GB). DEIGHTON, D., W. [GB/GB]; 18 Oakwood Avenue, Borehamwood, Herts WD6 1SR (GB). GROSS, S., T. [GB/GB]; 10 Cottonmill Lane, St. Albans, Herts (GB). CAESARI, A., C. [GB/GB]; 43 Mulbray Road, Brondesbury, London NW6 7QS (GB). BRETT, A. [GB/GB]; 21 Ridgeway, Kensworth LU6 3RL (GB). WHALLEY, R. [GB/GB]; 50 Grange Rise, Codicote, Hitchin, Herts SG4 8YR (GB). FIERMAN, A. [GB/GB]; 5 Malthouse Gren, Wigmore, Luton, Beds LU2 8SN (GB). RICHMAN, L. [GB/GB]; 1 Howards Close, Bushey Heath, Herts WD2 3RR (GB).

(74) Agent: WOOD, John, Irwin; GEC Patent Department, Waterhouse Lane, Chelmsford, Essex CM1 2QX (GB).

(81) Designated States: CA, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: IMPROVEMENTS RELATING TO VIDEO TELEPHONY



(57) Abstract

In a videophone a method of transmitting simultaneously video and audio data over a PTSN uses a first modem (31) which is connected to the PTSN to transmit by analogue means data by a data compression technique to a second modem. Each modem (31) has associated with it a controller (32) linked to a respective audio unit (33) and a video unit (34). Each audio unit (33) is arranged to transmit fixed sized data packages in a synchronous mode and each video unit (34) is arranged to transmit a variable bit stream of variable length in an asynchronous mode. Each controller (32) is capable of transmitting data in packs in a data frame comprising 5 packs, the audio data is contained in 4 fixed sized data packs of 160 bits of data and the video data packet is contained within a single data pack.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**FR**	France	**MR**	Mauritania
AU	Australia	**GA**	Gabon	**MW**	Malawi
BB	Barbados	**GB**	United Kingdom	**NL**	Netherlands
BE	Belgium	**GN**	Guinea	**NO**	Norway
BF	Burkina Faso	**GR**	Greece	**NZ**	New Zealand
BG	Bulgaria	**HU**	Hungary	**PL**	Poland
BJ	Benin	**IE**	Ireland	**PT**	Portugal
BR	Brazil	**IT**	Italy	**RO**	Romania
CA	Canada	**JP**	Japan	**RU**	Russian Federation
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SK**	Slovak Republic
CI	Côte d'Ivoire	**KZ**	Kazakhstan	**SN**	Senegal
CM	Cameroon	**LI**	Liechtenstein	**SU**	Soviet Union
CS	Czechoslovakia	**LK**	Sri Lanka	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**MC**	Monaco	**UA**	Ukraine
DK	Denmark	**MG**	Madagascar	**US**	United States of America
ES	Spain	**ML**	Mali	**VN**	Viet Nam
FI	Finland	**MN**	Mongolia		

IMPROVEMENTS RELATING TO VIDEO TELEPHONY

This invention relates to a method and apparatus for transmitting both video and audio data over a PTSN.

The simultaneous transmission of such data is to be used in apparatus and equipment known under a generic title of a "videophone" in which a normal telephone audio signal and a television signal are transmitted over a public telephone network and are received at a remote location so that a caller can communicate with a receiver and each can see the other and speak at the same time. In order to use the PTSN it is necessary for data to be transmitted in an analogue rather than a digital form and standards have been established for example by the International Telegraph and Telephone Consultative Committee for the transmission of such data. These standards rely on there being a low bit rate of transmission and compression of data. One of the advantages of these types of telephone is that although the signal quality may not be terribly good, it does mean that low cost telecommunication equipment can be provided to the general public and the standard is suitable for all except the most sophisticated of purposes.

In this type of communication a telephone is provided which is a two piece instrument comprising a handset having a cord to a base unit and the base unit is connected to the PTSN by a

straight line cord. Power is drawn by a separate lead from the mains via the 'in line' mains transformer. If the mains power to the telephone unit is discontinued the telephone will continue to work on the same basis as a conventional telephone. It is known to provide a base unit with such a handset and with a liquid crystal display screen which is mounted on the base unit and which incorporates a camera. A basic layout of such a system is shown in Figure 1 of the accompanying drawings to which reference is now made and here it can be seen that the base unit 1 has the power supply cable 2 at low voltage from a mains transformer 3 via a mains input cable 4 connected to a conventional mains source of electricity supply 5 via a plug.

The base unit has connected to it an LCD screen 6 which incorporates a television camera arrangement 7 and a handset 8 is separately connected by a trailing cord 9. The unit 1 can be connected to a PTSN by a conventional telephone plug 11 connected via cord 10 to the base unit 1.

The base unit incorporates the basic circuitry for the telephone and the display unit and also carries push button controls for the operations unit. The layout is preferably as shown in Figure 2 of the accompanying drawings to which reference is now made and it will be seen that it contains the conventional alpha numeric pad comprising 12 keys labelled 0 to 9, together with two function keys. The layout shows a conventional layout for a telephone pad. Above the pad is another set of push keys 13 which represent further functions in connection with the video aspects of the video phone. These incorporate as shown buttons to control contrast 14, brightness 15, freeze frame 16 and self view 17. Any provision for switching the video on and off is made via a pad 18 as well as an on/off timer 19. The pictures taken by the camera 7 of Figure 1 are slow and do not always properly control because of the slow data speed the accurate representation of what the camera sees. In order to refresh the picture the video refresh button 29 is incorporated and there is furthermore a function button 20. This function button will select additional functions through a combination of operation of other key pads or enable the video to

be set up separately. There is a third panel shown which has function pads 21, 22, 23 and 24. The purpose of all the pads is set out in Table I.

TABLE I

PAD	VIDEO KEYS	ABBREVIATION	FUNCTION
14	CONTRAST+	CU	Increase contrast of video display.
	CONTRAST-	CD	Decrease contrast of video display.
15	BRIGHTNESS+	BU	Increase brightness of display.
	BRIGHTNESS-	BD	Decrease brightness of display.
19	VIDEO REFRESH	VR	During video operation, cause the videophone at the other end of the link to restart video coding from scratch (to overcome picture errors due to data corruption)
			During freeze frame operation, obtain a new still image.
			During "on-hook" modes (self view and timer) step through diagnostic overlay information.

20	FUNCTION	FN	Selects additional functions through a combination of key presses, or to enable video set up when on-hook. Also used to force an extension receiving a transferred call into answering mode.
16	FREEZE FRAME	FF	Selects and deselects (toggles) video freeze frame.
17	SELF-VIEW	SV	Prior to making a call, provides a video loop-back (uncoded) to allow the user to adjust their position/ appearance. This can be deselected by another press of self-view, or by video call initiation (Video On). During a call a reduced size "self view" image is superimposed on the received picture ("picture in picture").
		SV/SV	During a video call, this double key press selects a self view monitor mode in which the local picture is

			the same as that transmitted to the far end of videophone.
19	TIMER ON-OFF	TIM	Toggles a display of call duration (h:mm format) overlayed on the video display. When on-hook, displays duration of last call.
18	VIDEO ON-OFF	VO	Toggles between video (digital) and audio (analogue) operation.
20/ 15		FN/BU	Increase mains phase delay (adj.during self view only)
		FN/BD	Decrease mains phase delay (adj.during self view only).

The combined FN/KEY sequences require that the FN key 20 is pressed down and released before the other key has been pressed and released.

PHONE KEYS	ABBREVIATION	FUNCTION
0 thru 9 P thru 9		Normal code indexing.
*	*	Invokes Digital Exchange Functions.
#	#	Invokes Digital Exchange Functions.

21 SECRECY MUTE		During audio operation, selects the mute facility. During video operation causes transmission of null video data (screen cleared at far end) and silent audio.
21 PAUSE	PSE	Used to add a pause in a sequence of stored digits, to cater for certain types of PABX. (The total duration of a sequence of pauses must not exceed 12 secs - this restriction shall appear in the user guide).
22 MEMORY STORE	SET	Used to set up phone numbers stored in memory.
22 MEMORY RECALL	PAGE	Used to recall a stored phone number,
23 RECALL	RCL	Invokes recall functions on business exchanges.
23 REDIAL	LNR	Redials the last number.

24 LOUDSPEAKING	LS	Selects/Deselects "hands free" operation, either in video or analogue mode.

other functions available are as follows:-

TABLE II

LED's	ABBREVIATION	FUNCTION
VIDEO ON	VONL	LED is on when the phone is in video call mode.
MUTE LED	MUTL	LED is on when SECRECY mode is selected (video or analogue mode).
HANDS FREE LED	HFL	LED is on when LOUDSPEAKING has been selected (video or analogue mode, with mains power available).

Featurephone operation includes the following features:

- Handset operation with Inductive Loop.
- Handsfree operation including on-hook dialing.
- Mute/secrecy function.
- Speaker volume control slider.
- Ringing volume control switch.
- 10 Telephone number memory.
- Last Number Redial.
- Recall.

- Loop disconnect/tonal dialing.
- Star services.
- Call timer (off-hook time in h:mm format).

These operate when mains power is available. If mains power is not available the phone is capable of operation as a basic phone.

It will be appreciated that the layouts and the functions may be varied but these are typical ones which can be found in apparatus made to the ITT CC standards.

It is an object of the present invention to provide an improved form of videophone and transmission system which incorporates all the above features and which gives a high degree of resolution in the picture and high quality audio transmission, while at the same time incorporating further features.

According to the present invention there is provided a method of transmitting simultaneously video and audio data over a PTSN in which a first modem is arranged in use to be connected to the PTSN to transmit by analogue means data by a data compression technique to a second modem, each modem has associated with it a controller linked to a respective audio and a video unit, the audio unit being arranged to transmit fixed size data packages in a synchronous mode and the video unit being arranged to transmit a variable bit stream of variable length in an asynchronous mode, each controller being capable of transmitting data in packs in a data frame comprising five packs, the audio data being contained in four fixed size data packs of 160 bits of data and the video data packet comprising a single data pack.

Preferably each audio data packet contains 77 data parts and each video packet comprises 76 data parts. The data parts are preferably comprised of 6 bits each. The check sums are also comprised of a 6 bit piece of data.

In carrying the invention into effect according to one example of construction in operation, the videophone as illustrated in Figures 1 and 2 will now be described in detail.

Commencing first with the video, the video as well as

providing a visual display indicating motion is also able to provide a laterally inverted i.e. mirror like self view image, this is to enable the caller to set up their position in front of the camera for transmitting. If used in this way the video will be uncoded although if it is used in a monitor mode during a call, it will be coded. Another feature of the video is that there is provided means to allow the user to monitor the position in front of the camera by inserting a picture within a picture.

The video screen colour image represents a picture in a 128 x 96 pixel matrix. Set in one corner of this 32 x 24 pixels can be allocated to give a black and white image of the user. Further visual information can be given to the display in the manner of an overlay along either the base or the side of the picture. This overlay is normally in textual form and represented by normal characters. The font for this text is based on 7 x 5 pixels and characters are normally represented as white on black with a surrounding border. The characters represented can be normal alphabetic characters as well as numeric ones and non-alpha numeric characters such as punctuation marks and symbols. The display may include diagnostic information such as the call time M & R and minute format up to a maximum of 10 hours and this is continually updated while the call is in operation. It can also include training message and condition message for example saying whether the line is in good repair or whether there is a reply or no reply. It can further indicate the display settings of contrast and brightness and give a far end video mute status showing full secrecy. If reference is now made to Figure 3 of the accompanying drawings, this shows the set-up of the interface from which the protocol of communication will be described for 2 modems. The layout in Figure 3 shows a maxdata interface. The two units are referred to as 'X' and 'Y' and each has a respective modem 31 which is connected to a controller 32 and the controller is fed by associated audio 33 and video 34 elements.

In operation the communications between the controller will follow the following sequence:-

a) The audio-audio link will be synchronous, i.e. fixed size data packets to be transmitted every 32 ms. in an analysis period.

b) Video-video communications consist of a bit-stream of variable length codes, and will be essentially asynchronous. However, since data errors on the video link cannot be tolerated any error detection protocol (checksum) and associated control (refresh) will be required.
An assumption is made of 10^{-5} worst case bit error rate in choosing the protocol for dealing with data errors. This corresponds to a data packet error at worst every 30 seconds of operation, in which case the image will be refreshed from scratch. If the data packet error rate is significantly higher than this in practice a re-send protocol will be required, which will

i) make the 'Video Data and Control' significantly more complicated

ii)introduce an extra lag into the video link

c) The modem link will train to 14.4 kbps, 9.6kbps, or not at all. If the modems 31 train up at 14.4kbps data symbols will comprise 6 bits each (at 2.4kHz), at 9.6kbps they will comprise 4 bits each (also at 2.4kHz symbol rate).

The controller packs data from the audio and video DSPs, along with control information and error correction data, into data packets for transmission via the modem link. These data packets are based on the audio timings in order easily to implement the synchronicity required in the audio communications. The protocol is complicated by the disparity between the audio and controller interrupt rates, leading to an average data packet size of 76.8 symbols.

The controller 32 sends a repeated sequence of:four 77 symbol data packets followed by a 76 symbol data packet. These 5 packet blocks constitute a data frame as follows:

For a 14.4kbps Link

Each data packet within a frame comprises:

Data Type	Symbols	
Video	46 at 6 bits each	) 73
Audio	27	)
Checksum A	1	
Checksum B	1	
Checksum C	1	
Slow Control	1	
Total	77	

The 'Slow Control' symbol is not present on the fifth packet of a frame.

The components are further details as follows:

a) Video

272 bits (i.e. 17 x 16bit data words) plus four spare bits. This corresponds to 8.5kbps.

b) Audio

160 bits of data, corresponding to 5.0kbps

c) Checksums A, B,C.

6 bit checksums generated by EX-ORing successive symbols within the video and audio (TBC) parts of the data packet (i.e. Checksum A looks at symbols 0,3, ..., 72; Checksum B at symbols 1,4, ..., 73; and Checksum C at symbols 2,5, ..., 71)

d) Slow Control

A 6 bit control symbol, repeated four times per data frame to ensure valid transmission.

(i) Bits 4, 5 both to zero (normal data transmission mode)

Bits 0,1: 00 Null
01 Refresh
10 Freezeframe
11 Continue

Bit 2 1 Acknowledge

Bit 3 1 Link Terminated

(ii) Bits 4, 5 not both zero

These values are used when keypress data are being sent to an STTE unit, with the following formats for transmission:

Bit 5 = 1

Bit 4 = 0

Bits 3 to 0 = 0x0 Function key 10 pressed (Fig. 2)
0x1 Key 19 pressed
0x2 Key 16 pressed
0x3 Key 15 pressed
0x4 Key 15 pressed
0x5 Key 14 pressed
0x6 Key 14 pressed
0x7 Mute asserted (following a change in mute status)
0x8 Mute removed (following a change in mute status)

When receiving control data with Bits 4, 5 not set to 00, the controller 31 ignores the data.

In order to establish a precedence hierarchy it is decided that:

When connected to STTE, the normal video control signals shall take priority, i.e. reporting of key presses to STTE will wait for the next slot ('data frame') for which the control data would have been 000000 (binary).

For a 9.6kbps link

Each data packet within a frame shall comprise:

Data Type	Symbols
Video	32
Audio	40
Checksum A	1
Checksum B	1
Checksum C	1
Checksum D	1
Slow Control	1
Total	77

The 'Slow Control' symbol is not present on the fifth packet of a frame.

The components are further detailed:

a) Video

128 bits (i.e. 8 x 16 bit data words) plus four spare bits. This corresponds to 4.0kbps.

b) Audio

160 bits of data, corresponding to 5.0kbps

c) Checksums A, B, C, D

4 bit checksums generated by EX-ORing successive symbols within the video and audio (TBC) parts of the data packet (i.e. Checksum A looks at symbols 0,4,..., 72; Checksum B at symbols 1,5,...,69; Checksum C at symbols 2,6,..., 70; and Checksum D at 3,7,..., 71)

d) Slow Control

A 4 bit control symbol is repeated four times per data frame to ensure valid transmission.

Bits 0,1:	00	Null
	01	Refresh
	10	Freezeframe
	11	Continue
Bit 2	1	Acknowledge
Bit 3	1	Link Terminated
Bits 4,5	XX	Not Used

Synchronisation of the Controller 31 is tied to audio analysis frames. In order to achieve this the following procedure is adopted.

1. Following modem training the SPM controller transmits zeroes up to and including the next audio-synchronised interrupt. At this point the Audio DSP receive a 'Start' command.
2. On the next 2.4kHz controller interrupt cycle, the controller starts transmitting its initialisation sequence, consisting of 228 symbols.
3. After the synchronisation sequence, data transmission, as described above commences immediately.

The synchronisation sequence consists of four zero symbols followed by repeated (but not identical) blocks of 16 symbols as follows:

In One Block

(Block delimiter	1 symbol F Hex
(Counter 1 symbol	E Hex. counting down to 1 Hex.
(Unit Identifier	14 symbols (TBD)

giving a total of (4 + 14 x (1+1+14)) = 228 symbols.

For the purposes of commonality, the structure for 9.6kbps and 14.4kbps have been kept identical. The Unit Identifier format is defined as:

Unit Version	:8 bits as defined at Production Test (Set to 0xFF for STTE equipment)
Unit Operational Mode	:8 bits determined locally (=0x00 for initial units)
Serial number	:24 bits
Spare	:16 bits

Total = 56 bits

The identifier only serves to indicate that equipment is an STTE unit or otherwise, for the initial system, i.e. by checking whether the incoming Unit Version =0xFF, the local Controller can determine whether or not it is attached to STTE (this will only occur for a 14.4kbps link).

At the far end unit the (block delimiter and counter) pair effectively point ahead to the start of real data transmission (i.e. it will start 16 symbols x counter value max 14 from the beginning of that block). After four agreeing pointers, the unit identifier shall be extracted by taking the value of the next matching pair of unit identifiers. Following the agreeing pointers and an identity match, synchronisation is deemed to have been achieved.

If synchronisation is not achieved before a nominal 3 second time-out the unit defaults back to analogue operation.

If attached to STTE equipment then a note can be made, and for the rest of the call, key presses shall be reported as noted above. In addition, the start of normal data transmission shall be delayed by three data frames (data packets) as described above. This allows time for status data to be transferred to the STTE/ATE and for the STTE/ATE to download set-up parameters as necessary.

Control data is protected by repetition. A two-out-of-four match criterion is applied to the four control symbols received in a data frame as follows:

(i) Three out of four match or better.
Accept the control value.

(ii) Two-out-of-four match, other two values not consistent with each other.
Accept matched value.

(iii) Two inconsistent matched pairs.
Revert to an analogue phone.

(iv) No matches.
Revert to an analogue phone.

The chances of (iii) or (iv) occuring are minimal under normal conditions, even at a 10^{-4} bit error rate. If a major modem failure has occurred then the controller will have been alerted by the modem itself.

As additional protection the 'accepted' control word shall be checked for the following invalid values which shall, if they occur, cause the system to revert to an analogue phone

(i) Unexpected 'Acknowledge'

(ii) Top two bits non-zero when not connected to STTE.

(iii) 'Freezeframe' if already frozen.

(iv) 'Refresh' if currently in a 'Refresh' cycle but an 'Acknowledge' has not yet been sent to the far end of the link.

(v) 'Continue' if already in motion video mode.

A detailed description of the video DSP/ASIC interaction is given below and with reference to Figures 4, 5, and 6 of the accompanying drawings in which Figure 4 is a block diagram showing the Decoder/Display process, Figure 5 is a block diagram showing the Encode/Camera Operation Camera ready before DSP and Figure 6 is similar to Figure 5 except this shows the situation of DSP ready before Camera.

The main considerations in the operation of the Video hardware are:

a) Software. The Decode and Encode tasks are distinct and asynchronous (except in the special cases of Monitor and Self View modes). Timing must be controlled by polling as the chosen DSP does not support interrupts well.

b) Camera. The total time to gather a video frame will be in the region 70-1000 ms, the uncertainty being up to 10 ms. waiting for mains synchronisation and up to 20 ms shutter time (this will be increased further under very poor lighting). The data are to be gathered into a separate Live RAM without affecting the DSP's access to its own Current and Previous RAMs. When a frame has been gathered and the DSP is ready to process it, the Live RAM data are transferred, overwriting the Previous RAM. The physical Live and Previous RAMs are then bank switched so that the new data can be accessed by the DSP at the logical addresses of Current RAM. On completion, the act of transferring data from live RAM is therefore used to trigger the gathering of a new image by the camera in order to keep the DSP waiting time to a minimum. A second consideration is that of AGC and white balance. The data for updating these are collected during the gathering of a camera frame. To avoid step changes and the resulting settling characteristics at the start/restart of image processing (e.g. after a spell in Freezeframe) the DSP must therefore continue to request new frames both in the 'Idle' state and when its encoder has completed a Freezeframe sequence.

c) Display. The LCD is continually updated at a rate equivalent to 625-line TV. The Live area of display RAM must therefore be physically separate from that used by the DSP to avoid contention. Also, the LCD update process is free running, i.e. asynchronous to any DSP processes, and has a short 'Vertical Blank' period when the image data are not needed. The transfer of a new image to the display must therefore wait for a 'Vertical Blank', and during this waiting time ($<$20ms) the DSP will not be allowed

to access its Decode memory space. The initialisation sequences for camera and display are as follows: Camera: Once the SPM Controller has loaded the AGC Metering FIFO and Camera Control FIFO the first frame will be gathered. The DSP will be aware of this via the 'Live Encode Memory Prepared' flag. Display: Once the Display Control FIFO has been filled by the SPM Controller and the DSP has transferred an initial blank/blue screen into the Display Live RAM, the display will start up.

Referring now to Figures 4, 5 and 6 these illustrate normal (i.e. not initialisation) operation, using the status flags and updates described earlier. P = Previous frame data C = Current frame data L = Live data During data transfers (which last > 1ms) the Decode (for Display update) or Encode (for new camera frame) memory space will not be accessible to the DSP (except via Scratchpad mode). Note:

1) that the flags 'Decode Memory Prepared' and 'Encode Memory Prepared' are actually inverse logic in the interface protocol (but are shown the logical way round here).

2) that the designation Previous/Current is somewhat arbitrary. In a scheme where the image is post-processed (e.g. Laplace, the reference frame will be in 'C' space at the start of processing, an updated reference will then be generated in 'P' space and, finally, a post-processed image generated in 'C' space for transfer to Live RAM. At the start of the next frame's processing the new reference will have been bank switched into 'C' space, and so on.

3) The DSP can select/deselect the 'Live Frame Hold' feature for freezeframe operation. When selected the Camera ASIC will pass data across as normal when 'Encode Processing Complete' is flagged by the DSP. In this case, however, the data being passed across will remain fixed (NB. the ASIC will still take the normal frame-gathering time even when 'Live Frame Hold' is selected). Following selection/deselection the new ASIC state will have been instated after the next time 'Encode Processing Complete' is asserted by the DSP.

The moderm produces "symbols" at a rate of 2.4kHz. For a

bit rate of 9.6kbps, the modem symbols are 4 bits in size. At 14.4 kbps they are 7 bits in size (6 bits + 1 redundant bit). It follows that the modem DSP must be sychronised to a clock which is harmonically related to 2.4kHz.

The audio DSP takes/reproduces speech samples taken at a rate of 8kHz, which is standard for narrow bandwidth speech. It follows that the audio DSP must be synchronised to a clock which is harmonically related to 8kHZ.

The need for synchronisation arises when the audio DSP and the modem DSP need to communicate with the controller. Rather than have a completely asynchronous protocol, the communications protocol is designed to use a synchronous frame structure. To achieve this a master clock produces audio DSP interrupts at a rate of 8kHz, and interrupts at 9.6kHz to the controller and modem DSP. Synchronisation is achieved by transferring data between audio and controller after every 10 x 8kHz interrupts, during which time there will have been exactly 12 x 9.6kHz interrupts.

The other issue is synchronisation between the modems at opposite ends of the line; it is possible to simplify the system by phase locking one of the modems to the other.

In accordance with CCITT standards a buffer control algorithm is developed for video decoding.

This comprises a hybrid transform data compression system for full motion video images. Hybrid denotes a technique which involves more than one redundancy reduction and in this case it is achieved by interframe methods and orthogonal transformations.

Interframe methods consist of coding the error between the previous and current frames. It is assisted by the use of motion estimation which attempts to minimise the error signal. Implicit in this method is the necessity for the encoder to hold a local version of the reconstructed image, in order that the correct error signal is obtained for each frame. The encode/decode process is therefore:

$$e(t)=f(t-1)-f(t)$$
$$f(t)=f(t-1)-e(t)$$

where e = error frame
f = image data
t = time

The orthogonal transform is designed to map highly correlated special data into a highly uncorrelated transform domain. In this case the Discrete Cosine Transform is used. Further compression is achieved by thresholding out all insignificant transform coefficients and quantising the remaining prior to transmission.

The decode process is the inverse of the encode process.

The CCITT standard defines a picture format (Common Intermediate Format or CIF) of 352 cols by 288 rows. They also define a Quarter CIF format of 176 x 144. Our format is a further reduced version of the 1/4 CIF.

The thresholding and quantisation process is linked to the number of bits required to code a frame: higher threshold/coarser quantisation mean fewer bits of information per frame, but poorer image quality, and vice versa.

Considering first the thresholds, 64 frequency components resulting from the DCT are subjected to a variable thresholder designed to remove as many insignificant coefficients as possible. In order to prevent bias in any direction the 2 dimensional array of DCT coefficients shall first be sujected to a mapping into 1 dimension as shown in Figure 7 of the accompanying drawings.

In Figure 7 the numbers given in the grid represent the index into the original 2 dimensional array.

The thresholding algorithm is thus as follows:

1. If [x[i]]<T then 2, else 3.
2. x[i]=0, If T<Tmax then 4, else 5.
3. T=g.
4. T=T+1.
5. T=Tmax.

where T=Threshold level relevant to that coefficient

g= current quantiser value.

Tmax=1.5*g.

i=array index.

and for i=0, T=g.

Considering now the quantisation, this is carried out after thresholding. The decision and representation levels are as follows:

$$Qdec(0)=0$$

$$Qdec(n)=n/|n|\{T+(|n|-1)g\}, \quad |n| =1,2,3...$$

$$Qrep(0)=0$$

$$Qrep(n)= 1/2 \{ Qdec(n)+Qdec(n+n/|n|)\}, \quad |n| =1,2..$$

where Qdec=decision level

Qrep=representation level

T=Threshold

g=quantiser value

Both luminance and chrominance coefficients are subjected to the same quantiser value. The quantiser value is modified at the start of each frame according to the estimator polynomial given under Buffer Control.

Legal quantiser values shall range from 4 to 64 inclusive, in steps of 2.

Frame 0 is coded with a quantiser value of 8.

All INTRA DC coefficients are quantised and reconstructed linearly with a value of 8 and no dead zone.

In order to achieve Buffer Control it is important that the number of bits per frame averages to the available bit rate over the PSTN line, since this gives the most efficient use of the bandwidth, and therefore the best quality image. It is also important that the number of bits per frame does not fluctuate

widely, since this varies the time for transmission and the quality of each frame.

The number of bits per frame is controlled by the thresholding an quantisation processes, which in turn must depend upon the number of bits used to code previous frames (i.e. transmit buffer level).

It has been observed that certain buffer control algorithms used in H261 coders exhibit instabilities. (Clearly this is a non-linear feedback process, and therefore instability and chaotic behaviour are possible).

Accordingly the invention uses estimator polynomials which produce a more uniform control of the number of bits per video frame, and hence more consistent video coding delays and picture quality.

The form of the estimator polynomial is as follows:

$B=A0+A1(M/S)^2$

Where B= number of bits

M=S (total number of macroblocks) MAE

S=quantiser

and S(s)=summation over all a.

A0=-480

A1=22

(The above values of A0 and A1 are typical values, but may be changed to optimise for different situations).

For each frame this equation is rearranged into the following format:

$S=\mathrm{sqrt}(A1/(B-A0))M$

Recognising that the polynomial is only a rough approximation to the actual number of bits generated, the value for B in this equation is calculated as follows:

Define T=bit rate/frame rate.

Frame 0: B=T

All other frames: B=T+ (T-X)

where X = actual number of bits generated for frame f(t-1)

The actual value of S obtained from the equation above is converted to a legal value of S, by taking the nearest which is not less than the actual.

Consideration will now be given to the Modem. The basic requirement of the modem section is the ability to transmit in full-duplex over the PSTN at data rate of around 12 kb/s. The proposed realisation is to be based on the CCITT V32 recommendation and high data rate extension of this. Will will allow the following specific data rates: 9.6kb/s. 12kb/s and 14.4kb/s. The transmission method involved is based on trellis coding, and each increased bit rate step is 3dBs (about 3 orders of magnitude in error rate) worse in signal-to-noise terms.

An overview of the functional elements of a modem transmission system is now given on the basis that it will be realised round a single powerful DSP processor, but with a necessary amount of external analogue hardware for the DSP and line interface.

Each functional block is described, together with how each must be realised - hardware or DSP software - and an indication of their complexity.

The nominal interfaces around a modem will be referred to the terminal interface, which will be internal in the use of the videophone, and the line interface where the connection is made to the PSTN.

The two relevant interfaces of a modem are shown in Figure 8 of the accompanying drawings. The terminal interface may be used separately and is a well-defined demarcation point between the modem and the rest of the videophone. The line interface is effectively where the "modem" plugs into the wall socket.

The relevant standards and necessary approvals are (a) the CCITT V-series which concerns the interface signals and (b) the BABT approvals and the corresponding British Standards which concern the user-PSTN and environment interaction. Their application points are shown by dotted lines in Figure 8.

V24 this covers the terminal equipment/modem equipment connection and specifies the data and control protocol lines. It is effectively RS232.

V32 the actual modem specification V32 refers to the signal requirements such as signal spectrum, modulation and coding, transmit clock accuracy, carrier frequency accuracy.

BABT: the relevant BP specifications cover:

Electrical safety	- user from PSTN
	- PSTN from user
Interaction with PSTN	- spurious line signals
	- auto-calling (loop-disconnect, MFetc)
	- auto-answering (ringing detection etc)

A functional block diagram is shown in Figure 9 of the accompanying drawings which show the primary elements in a modem. The transmitter, receiver, echo canceller, control and clock generation elements will be performed in the DSP. However, the control could be performed in an external microprocessor as is done in the V32 modem manufactured and sold by GPT Limited. Also, some elements of clock generation may have to be done in some sort of glue logic gate array if suitable timers and hardware pins are not available on the DSP.

Of the remaining elements, the D/A,A/D and line interfaces are all hardware, amd the V24 interface circuitry is likely to be nominal unless the modem in the videophone is required to function with external terminal equipment.

Consider now the V24 Interface. In a standard modem application, this block covers the standard RS232-type serial

interface by which terminals such as PCs and printers can be connected. The signals which exist are such as transmit and receive data and element timing and the communication links such as request and clear-to-send. The signal protocols and their electrical characteristics are all defined in the CCITT V24 specification.

The V24 interface can also be used to request auto-dialling by the modem via the line interface components. The V24 lines are used to transfer the dialling code, the relevant CCITT specification being V25bis.

In the videophone, the V24 signals are purely an internal interface. However, the signalling protocol is maintained for two reasons:

a) the modem element can be controlled entirely separately from the videophone and by the videophone via the V24 control lines;

b) the modem element can also be used separately as a modem by providing a V24 socket. Note however that the lines have to be driven by the drivers that are subject to BABT testing.

Functions: Protocol handling, signal generation, (line driving)
Implementation: Hardware and software.
Complexity: 2 (on a scale of 10)

Transmission to the transmitter Tx of Figure 9. The purpose of this is to take the input data at the data rate and convert to a sampled analogue signal which fits into the telephone signal bandwidth, 300 - 3.4kHz. This function is entirely defined in the CCITT V32 specification (and also in V33 for the 12 and 14.4kb/s rates).

Functions: Test pattern scramblers, rate-sequence generation, main scrambler, differential and trellis coding, I/Q mapping, pulse shaping, modulation, tone generation, output power scaling.
Implementation: DSP software.
Complexity :3

The Echo canceller EC of Figure 9 is provided so as to

remove the component of the transmitted signal present in the received signal. This component arises from mismatches in the 4-wire to 2-wire converters present in the signal path between the near-end and far-end modems.

There are two distinct sources of transmit echo:

(a) At the near-end modem 4w-to-2w hybrid. The "echo" here is instantaneous and has an effective time span of a few ms.

(b) At the far-end local exchange 4w-to-2w hybrid. The echo again has a few ms duration, but is delayed by the round trip period of the trunk exchange connection. In the national network this is a few ms, but in an international call this may be up to 1.5 secs. Also the echo may be phase rolled when crossing administration boundaries because of lack of network synchronisation.

Functions: National - Single complex adaptive FIR filter.

International - Dual adaptive filter, Bulk delay, Phase roll extractor.

Implementation:	National - DSP software.	
	International DSP Software + RAM (4K X 16)	
Complexity:	National	2
	International	4

The receiver RX of Figure 9 takes the sampled data from the echo canceller EC and processes this to recover the original data from the far-end modem. It also performs the function of achieving synchronisation to the far-end clock to generate the timing required in conjunction.

Functions: Timing interpolation, clock extraction, Complex adaptive equalisation, Carrier extraction, Demodulation, Trellis decoding, Line decoding, Descrambling, Test sequence detection, Signal quality detection.

Implementation: DSP software.

Complexity: 10 (of which the trellis decoder is 5)

The D/A, A/D block in Figure 9 provides the analogue to digital interface for the DSP. It sets the quantisation level which should have linearities of 11 bits and 12 bits for the D/A

and A/D respectively to cope with the dynamic range of the received signal level (approx. 40dB).
Functions: D/A low-pass filter (tx side). Anti-alias low pass and mains rejection high pass filters, A/D.
Implementation: Hardware, either discrete analogue or custom LSI.
Complexity: Discrete 3, Custom 12.

The function of the line interface is to connect the modem to the PSTN 2w subscriber line as prescribed by the relevant BS standards. Constraints include minimising non-linear distortion in any transmit echo to less than -80dB relative to the transmit signal level.
Functions: Isolating hybrid transformer, Active line hold, Line relays (used also for loop-disconnect dialling), Dialling detection.
Implementation: Discrete analogue, pcb design.
Complexity: 4.

The main function of the control is to sequence the modem through the handshake and to provide monitoring during date transmission via signal diagnostics. Other functions include activating the connect-to-line circuitry, performing automatic calling, and controlling the V24 interface line.
Functions: Auto-calling, Connect-to-Line, Handshake sequencing, Data transfer monitoring.
Implementation: DSP software (assuming no separate control processor, the modem being controlled via the V24 interface).
Complexity: 5.

The purpose of the clock generator is to provide clocks for the transfer of data cross the V24 interface and also for the A/D, D/A conversion. The complication is that the near-end and far-end transmitters are not phase-linked but have plesiochronous clocks. Thus the transmit and receiving sampling and data clocks are independent. The simplest arrangement is to use the transmit clock for the transmitter, the A/D, D/A conversion, and the echo

canceller. The echo canceller output is interpolated to the receiver timing in the receiver.
Functions: tx and rx data and sampling clocks.
Implementation: DSP software timer, with separate timer for tx clocks (hardware?).
Complexity: 3 but dependent on hardware.

The whole outline diagrammatic connection of the various parts of the system are shown in Figure 10, to which reference is now made and this shows all the features previously discussed and how they are inter-connected. A camera can be used which has a low cost image capture system with frame rates and interface compatible with the digital coding systems and suitable for a wide range of lighting conditions. The display from the video LCD will be flicker free and compatible with coded video transmission at variable and/or low frame rates. Assuming that standard video components are used, i.e. CCD, CCD driver in a non-standard manner, the following additions may be made to the camera:

a) Pipelined digital processing end-to-end including data conversion of colour parameters. This minimises the component count and the number of adjustments required at the production stage (both of which reduce production costs), and also leads to a more flexible white balance facility.

b) Electronic exposure control thus allows a greater than 10 ft. dynamic range in lighting levels. Normally this is not possible because the quantisation in exposure time is limited to multiples of horizontal line timings, in order not to interfere with the read-out process. We operate the CCD as a still-frame capture system and are therefore never reading out simultaneously with controlling exposure.

c) Voltage supply requirements to be reduced in complexity by using non-standard CCD drive waveforms which are then level-shifted. From only OV and 15V supply rails, a negative CCD supply and a super-voltage CCD supply are generated and, additionally, the need for adjustment at manufacture of these two derived supplies is eliminated.

The result of this is that the component costs in terms of supply generation and manufacturing cost in terms of set-up time, test equipment and yield as well as errors in set-up, will be reduced.

A further advantage of the invention is that three-way memory buffering is used on both display and camera. This allows the DSP to use its working areas independently of the display with the result that there is no flicker and camera, so that the DSP does not normally have to wait for a new image, except for roughly 1½ms transfer priods. In addition to this the DSP's camera and display memory spaces are organised so as to minimise the overhead incurred due to DSP techniques which code the difference between video images.

CLAIMS

1. A method of transmitting simultaneously video and audio data over a PTSN in which a first modem is arranged in use to be connected to the PTSN to transmit by analogue means data by a data compression technique to a second modem, each modem having associated with it a controller linked to a respective audio and video unit characterised in that the audio unit is arranged to transmit fixed size data packages in a synchronous mode and the video unit is arranged to transmit a variable bit stream of variable length in an asynchronous mode, and in that each controller is capable of transmitting data in packs in a data frame comprising five packs, the audio data being contained in four fixed size data packs of 160 bits of data and the video data comprising a single data pack.

2. A method according to Claim 1 characterised in that each audio data packet contains 77 data parts and each video data packet comprises 76 data parts.

3. A method according to Claim 2 characterised in that the data parts each comprise 6 bits.

4. A method according to Claim 2 or Claim 3 characterised by including at least one check sum part, said part comprising 6 bits of data.

5. A method as claimed in any preceding claim characterised in that the audio link is synchronous and fixed size data packets are transmitted at regular intervals in an analysis period.

6. A method as claimed in any preceding claim characterised in that the video link is asynchronous and comprises a bit stream of variable length codes.

7. A method as claimed in Claim 6 and characterised by including a refresh facility for the video link in which a 10^{-5} worst case bit error rate is established as the protocol for errors.

8. A method as claimed in Claims 2, 3, or 4 characterised in that the data packets are arranged to operate at a 14.4 bps link.

9. A method as claimed in any preceding claim characterised in that a precedence hierarchy is established in which normal video control signals are given priority.

10. A method as claimed in any preceding claim characterised in that control data is protected by repetition wherein a two out of four or better match of four control signals received in the data frame is accepted.

11. A method as claimed in any preceding claim characterised in that an image gathered by a camera in the video unit is arranged to gather the image in a 70 to 1000ms time period.

12. A method as claimed in any preceding claim characterised in that the video decoding is effected by a buffer control algorithm including a video hybrid transformer data compression technique embodying interframe methods and orthogonal transformations.

13. A method according to Claim 12 characterised by coding in the interframe method the error between previous and current frames.

14. A method according to Claim 13 characterised by an encoder being arranged to hold a local version of a reconstructed image to enable a correct error signal for that frame to be obtained.

15. A method as claimed in any one of Claims 12, 13 or 14 characterised in that further compression of the video signal is obtained by thresholding out all insignificant transform coefficients in a discrete cosine transform and quantising those remaining prior to transmission.

16. A method as claimed in any one of Claims 12 to 15 characterised in that estimator polynomials are used to provide an improved picture quality.

17. A method as claimed in any preceding claim characterised by removing any echo component of the transmitted signal in the received signal.

18. A method of transmitting simultaneously video and audio data substantially as hereinbefore described with reference to the accompanying drawings.

19. Apparatus for use with the method as claimed in any preceding claim and substantially as hereinbefore described with reference to the accompanying drawings.

1/7

Fig.1.

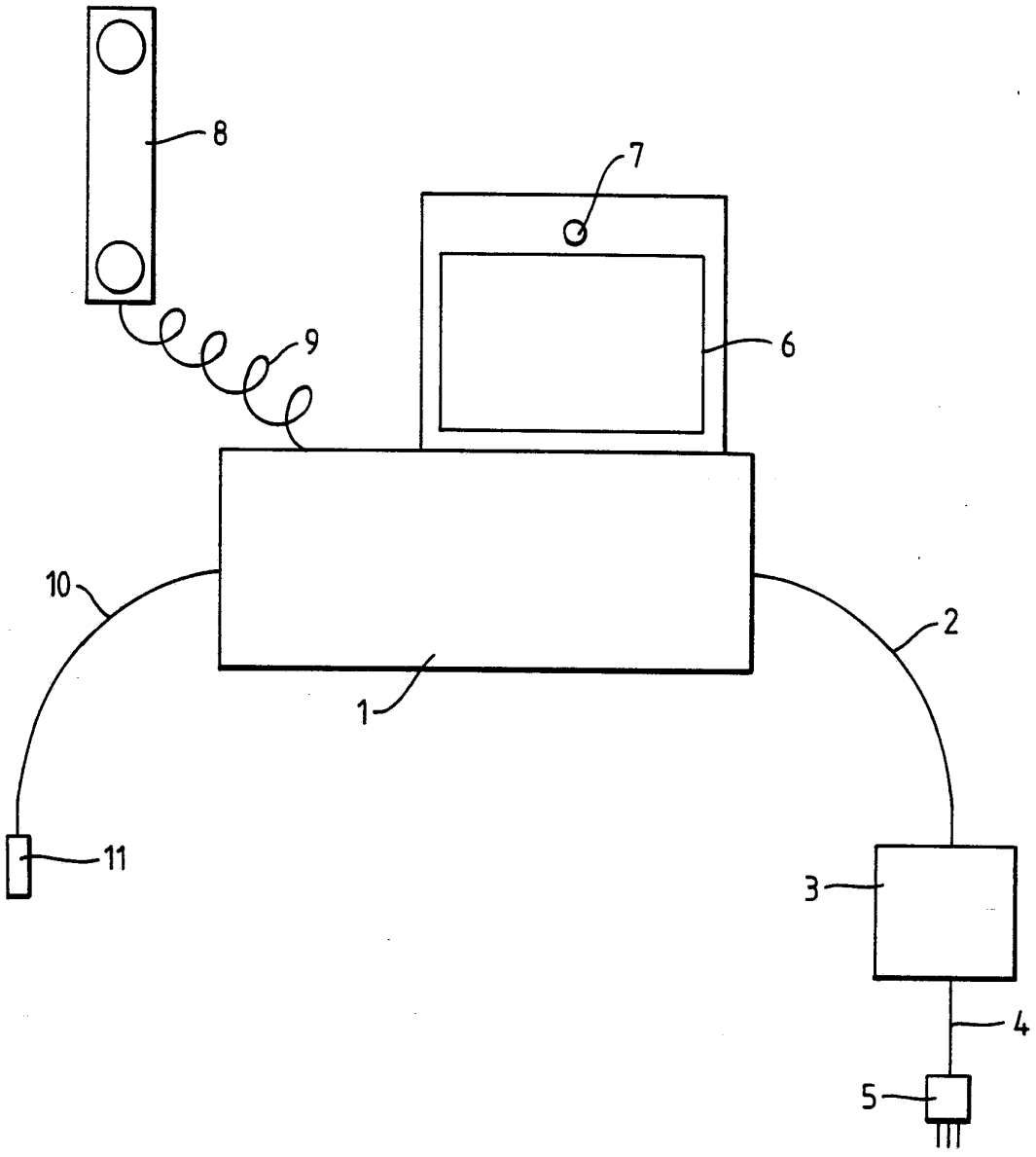


Fig.2.

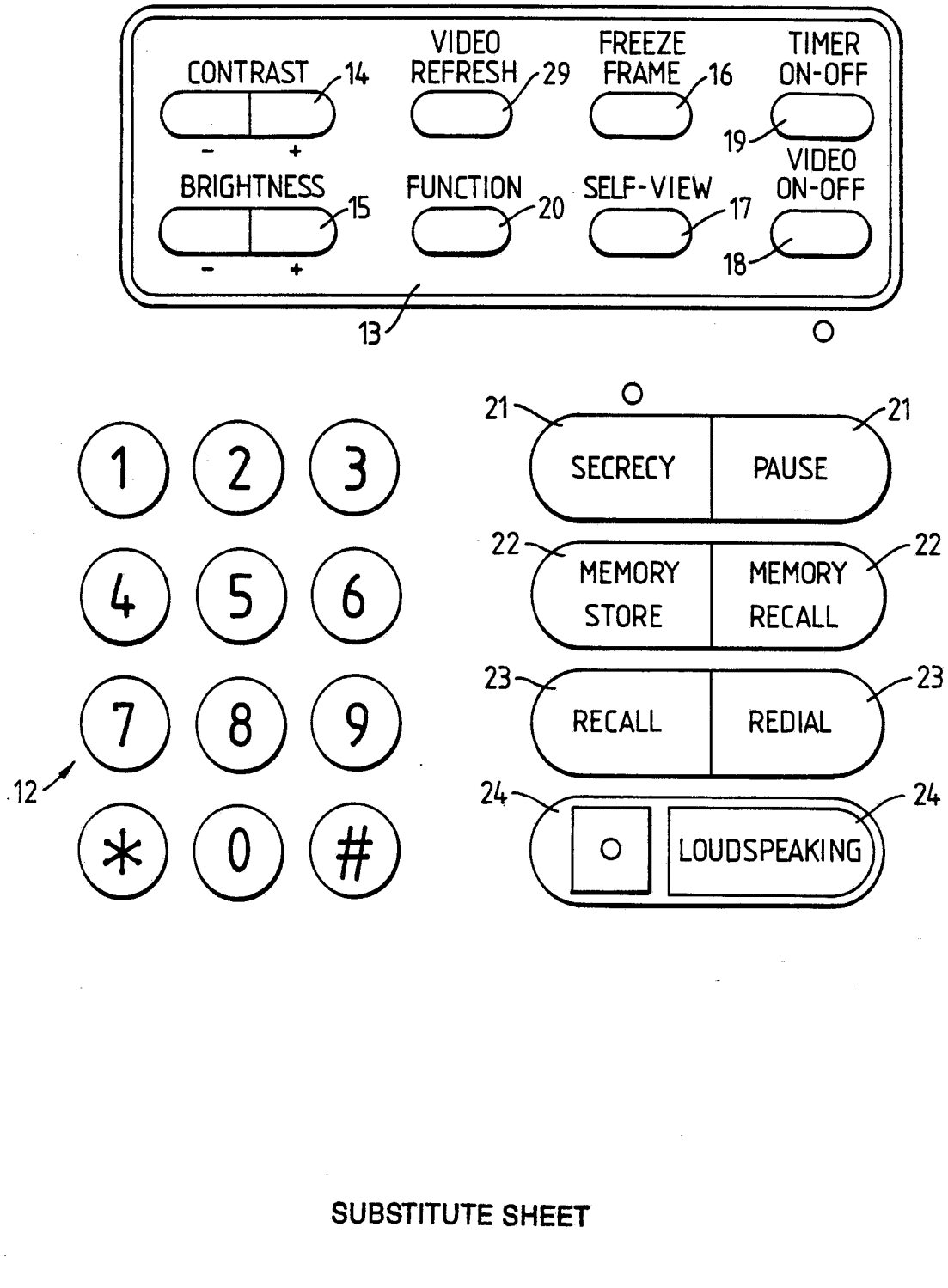

CONTRAST
14
VIDEO REFRESH
29
FREEZE FRAME
16
TIMER ON-OFF
19
BRIGHTNESS
15
FUNCTION
20
SELF-VIEW
17
VIDEO ON-OFF
18
13
1
2
3
4
5
6
7
8
9
12
*
0
#
21
SECRECY
PAUSE
22
MEMORY STORE
MEMORY RECALL
23
RECALL
REDIAL
24
LOUDSPEAKING

Fig.3.

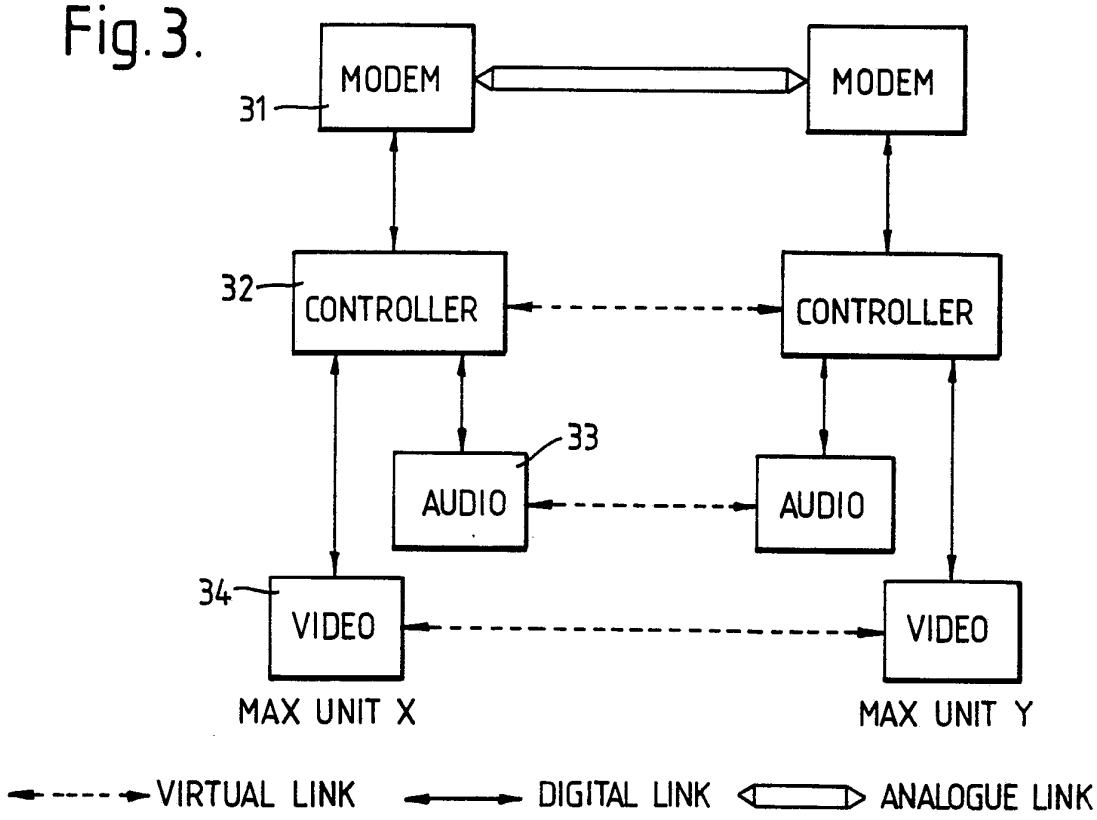

MODEM
MODEM
31
32
CONTROLLER
CONTROLLER
33
AUDIO
AUDIO
34
VIDEO
VIDEO
MAX UNIT X
MAX UNIT Y
VIRTUAL LINK
DIGITAL LINK
ANALOGUE LINK

Fig.4.

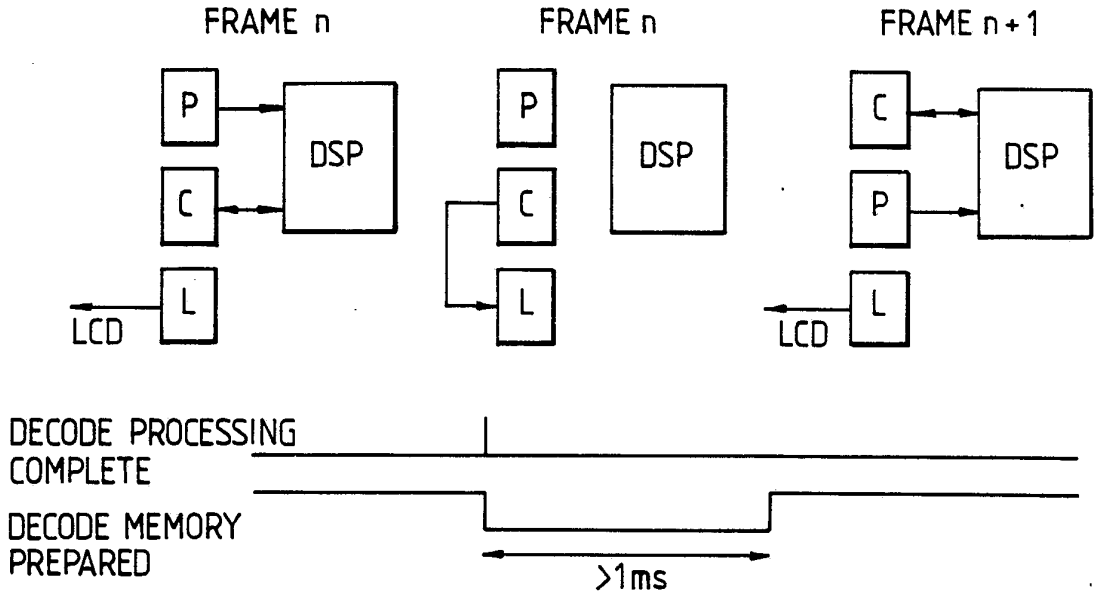

FRAME n
FRAME n
FRAME n+1
P
C
DSP
L
LCD
P
C
DSP
L
C
P
DSP
L
LCD
DECODE PROCESSING COMPLETE
DECODE MEMORY PREPARED
>1ms

Fig. 5.

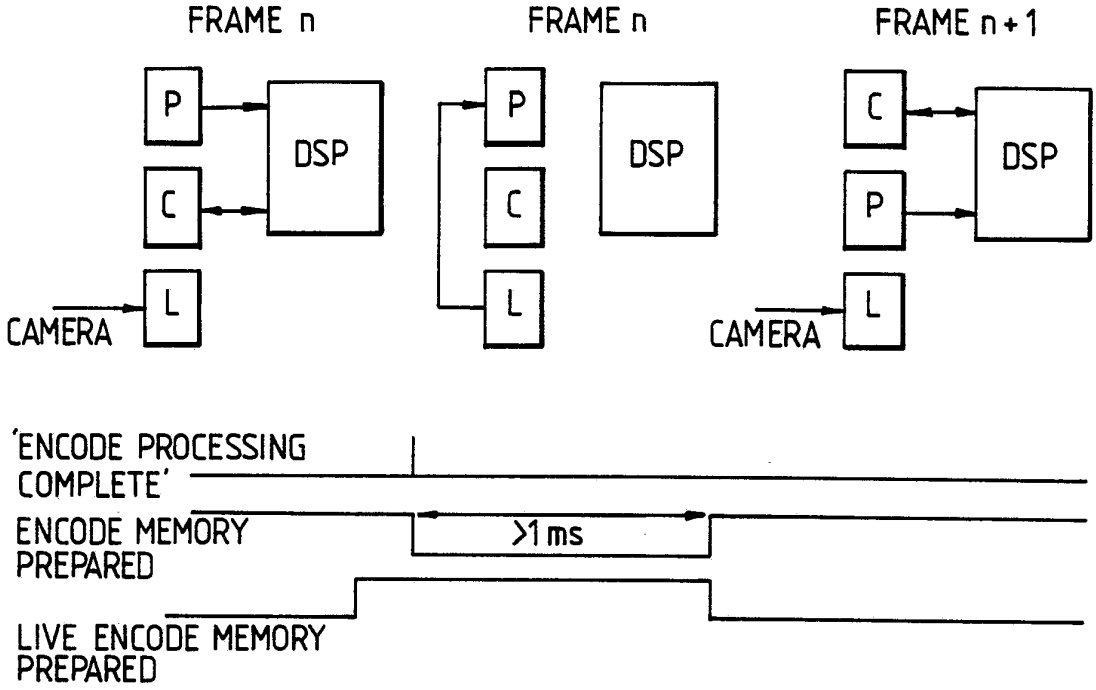

FRAME n
FRAME n
FRAME n+1
P
C
L
DSP
CAMERA
'ENCODE PROCESSING COMPLETE'
ENCODE MEMORY PREPARED
>1 ms
LIVE ENCODE MEMORY PREPARED

Fig. 6.

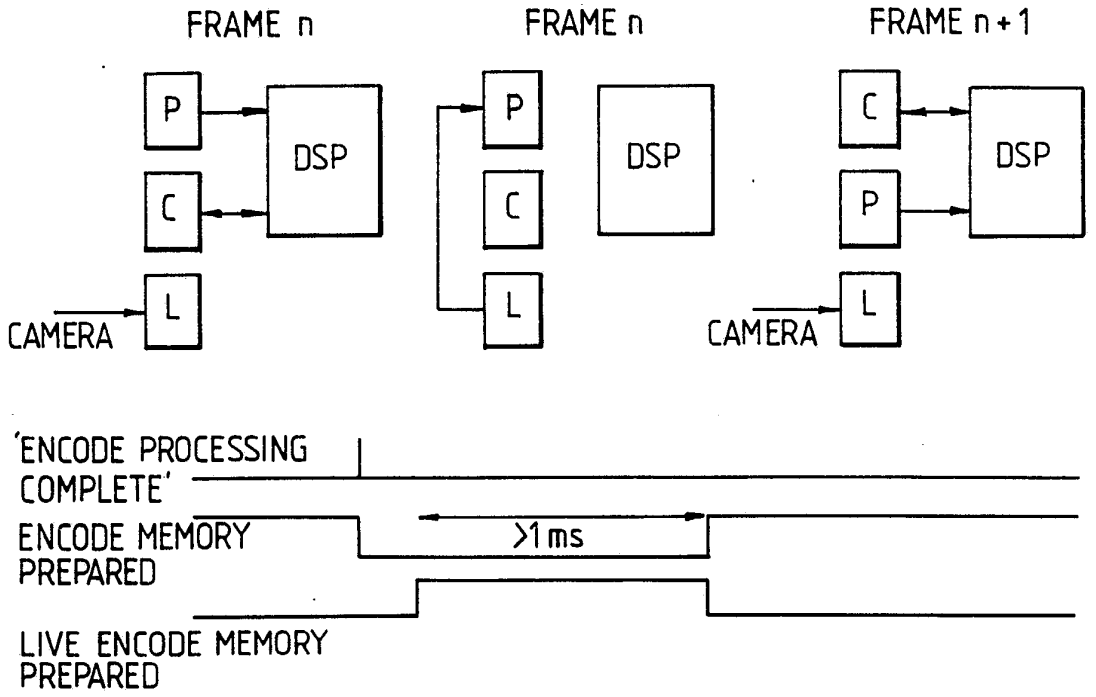

FRAME n
FRAME n
FRAME n+1
P
C
L
DSP
CAMERA
'ENCODE PROCESSING COMPLETE'
ENCODE MEMORY PREPARED
>1 ms
LIVE ENCODE MEMORY PREPARED

Fig. 7.

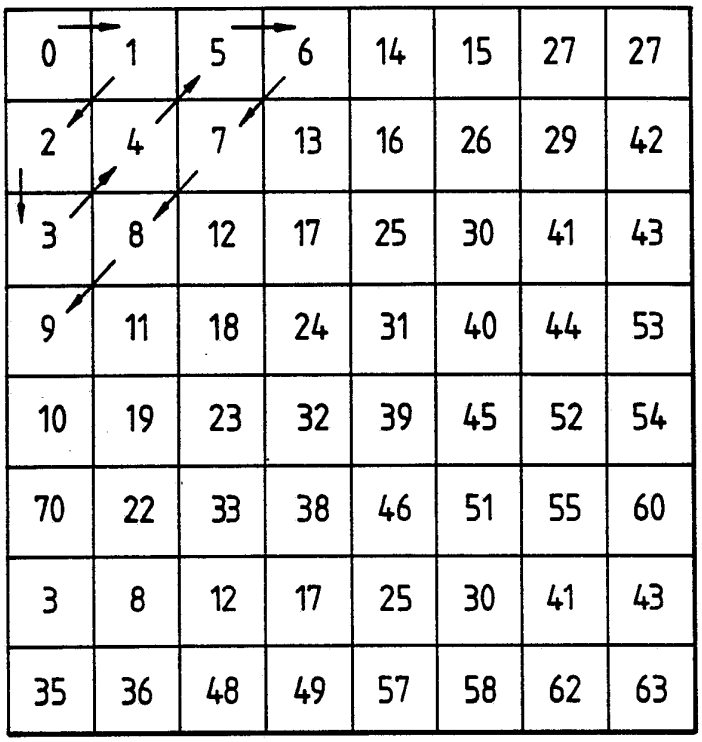

0 1 5 6 14 15 27 27
2 4 7 13 16 26 29 42
3 8 12 17 25 30 41 43
9 11 18 24 31 40 44 53
10 19 23 32 39 45 52 54
70 22 33 38 46 51 55 60
3 8 12 17 25 30 41 43
35 36 48 49 57 58 62 63

Fig.8.

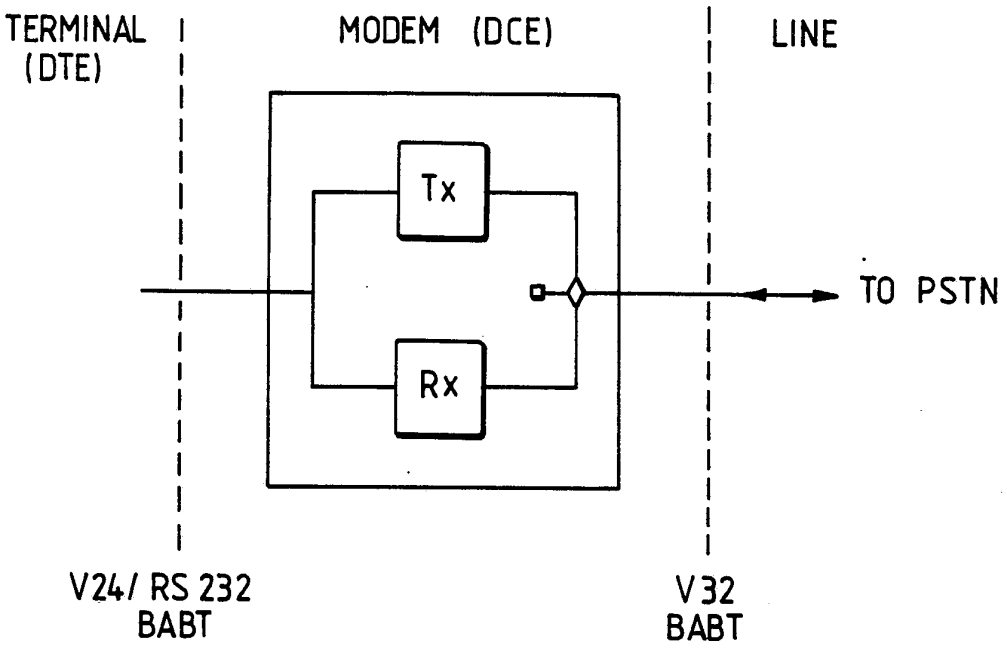
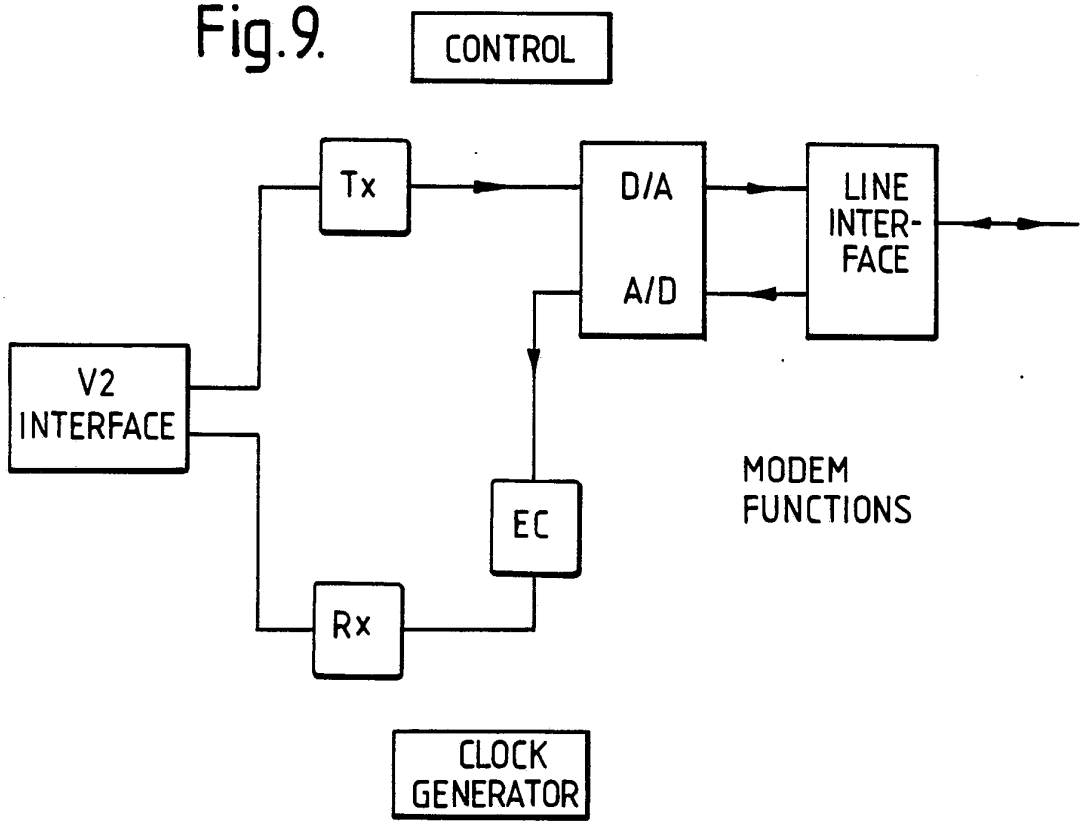

TERMINAL
(DTE)
MODEM (DCE)
LINE
Tx
Rx
TO PSTN
V24/ RS 232
BABT
V32
BABT

Fig.9.


CONTROL
Tx
D/A
A/D
LINE
INTER-
FACE
V2
INTERFACE
EC
Rx
MODEM
FUNCTIONS
CLOCK
GENERATOR

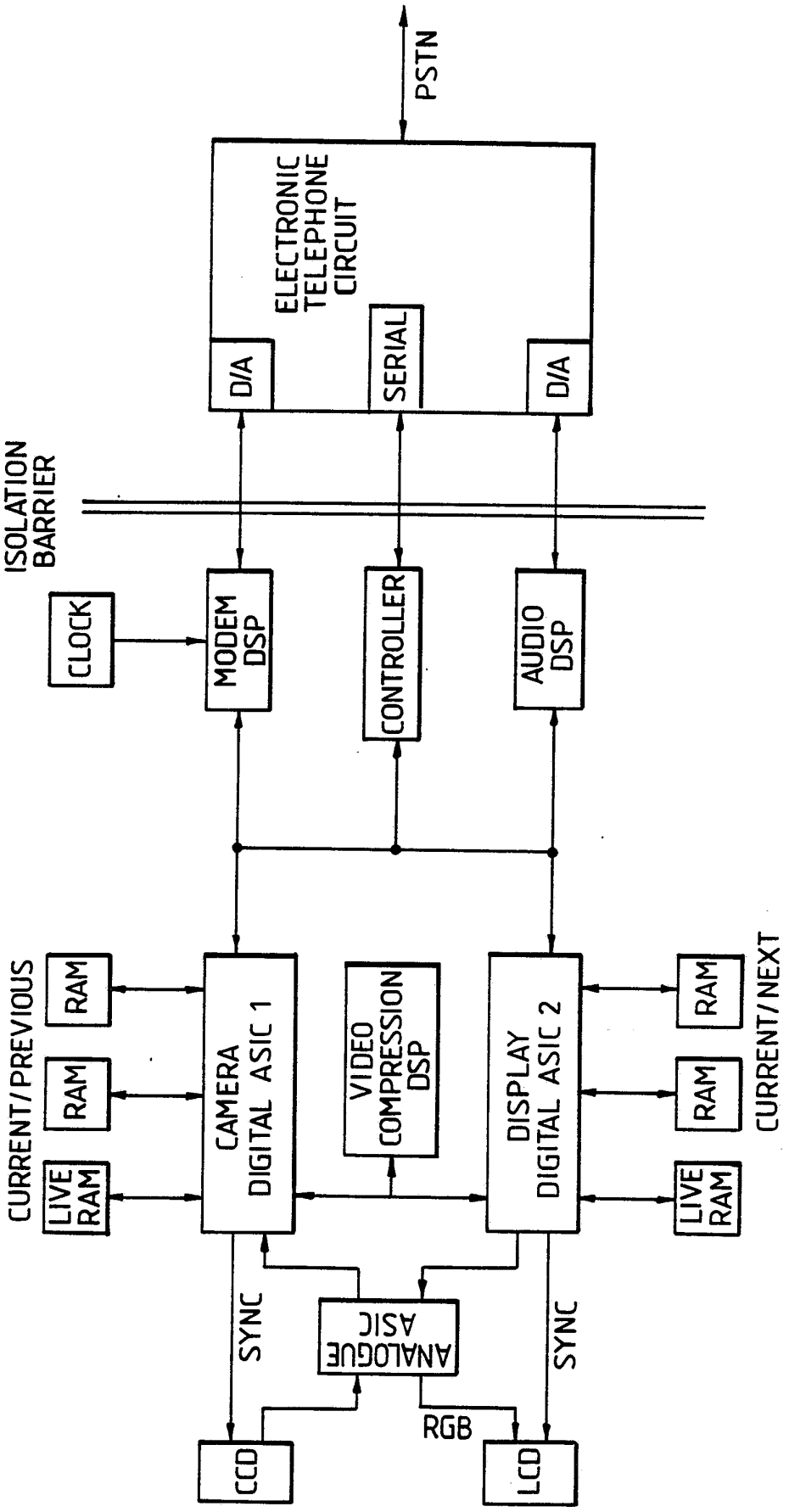

PSTN
ELECTRONIC TELEPHONE CIRCUIT
D/A
SERIAL
D/A
ISOLATION BARRIER
CLOCK
MODEM DSP
CONTROLLER
AUDIO DSP
CURRENT/PREVIOUS
RAM
RAM
LIVE RAM
CAMERA DIGITAL ASIC 1
VIDEO COMPRESSION DSP
DISPLAY DIGITAL ASIC 2
RAM
RAM
LIVE RAM
CURRENT/NEXT
SYNC
ANALOGUE ASIC
SYNC
RGB
CCD
LCD

Fig.10.

INTERNATIONAL SEARCH REPORT

International Application No PCT/GB 93/00517

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all)[6]

According to International Patent Classification (IPC) or to both National Classification and IPC

Int.Cl. 5 H04M11/06; H04N7/14

II. FIELDS SEARCHED

Minimum Documentation Searched[7]

Classification System	Classification Symbols
Int.Cl. 5	H04M ; H04N

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched[8]

III. DOCUMENTS CONSIDERED TO BE RELEVANT[9]

Category °	Citation of Document,[11] with indication, where appropriate, of the relevant passages[12]	Relevant to Claim No.[13]
A	BRITISH TELECOM TECHNOLOGY JOURNAL vol. 8, no. 3, July 1990, UK pages 43 - 54 M.W. WHYBRAY ET AL. 'Videophony' see paragraph 4.2 see paragraph 5	1,12,15, 16
A	US,A,5 033 062 (S.E. MORROW ET AL) 16 July 1991 see column 2, line 8 - line 26 see column 4, line 58 - column 5, line 10	1,5,6
A	EP,A,0 312 434 (J. GUICHARD ET AL.) 19 April 1989 see column 5, line 11 - line 20 see claim 1 see figure 6	1-3

° Special categories of cited documents :[10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
30 JUNE 1993	13.07.93
International Searching Authority	**Signature of Authorized Officer**
EUROPEAN PATENT OFFICE	PEREZ PEREZ J.C.

ANNEX TO THE INTERNATIONAL SEARCH REPORT ON INTERNATIONAL PATENT APPLICATION NO. GB 9300517 SA 71226

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report. The members are as contained in the European Patent Office EDP file on 30/06/93
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US-A-5033062	16-07-91	None	
EP-A-0312434	19-04-89	FR-A- 2621767	14-04-89
		DE-A- 3874478	15-10-92
		JP-A- 1243792	28-09-89
		US-A- 4887158	12-12-89