

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
7 August 2003 (07.08.2003)

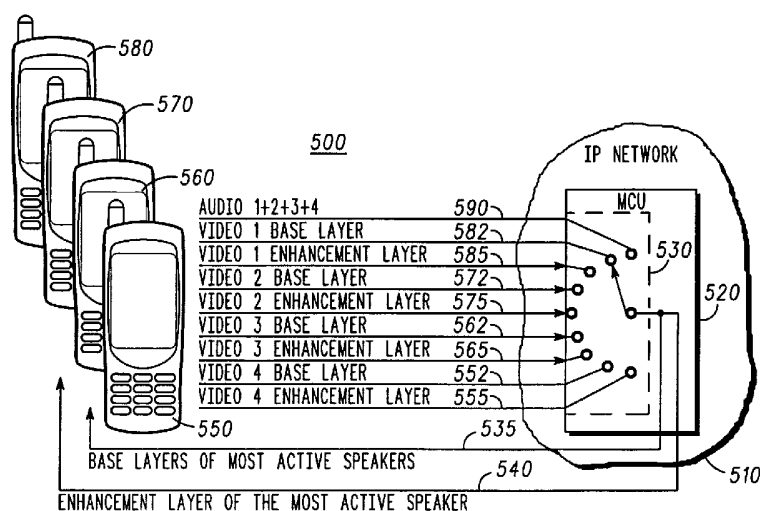
PCT

(10) International Publication Number
WO 03/065720 A1

- (51) International Patent Classification⁷: **H04N 7/14**, 7/15, 7/24
- (21) International Application Number: PCT/EP02/14337
- (22) International Filing Date:
16 December 2002 (16.12.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
0202101.2 30 January 2002 (30.01.2002) GB
- (71) Applicant (for all designated States except US): **MOTOROLA INC** [US/US]; 1303 E.Algonquin Road, Schaumburg, IL 60196 (US).
- (72) Inventor; and
- (75) Inventor/Applicant (for US only): **LALLET, Arthur** [GB/GB]; Motorola Limited, Jays Close, Viabes Industrial Estate, Basingstoke, Hampshire RG21 7PL (GB).
- (54) **Agent: TRELEVEN, Colin**; Motorola European Intellectual, Property Operations, Midpoint, Alencon Link, Basingstoke, Hampshire RG21 7PL (GB).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:
— with international search report

[Continued on next page]

(54) Title: VIDEO CONFERENCING AND METHOD OF OPERATION



(57) **Abstract:** A method of relaying video images in a multimedia videoconference between a plurality of multimedia user equipment (550, 560, 570, 580) includes the step of transmitting layered video images by a number of said plurality of user equipment wherein said layered video images include a base layer (552, 562, 572, 582) and one or more enhancement layers (555, 565, 575, 585). The transmitted layered video images are received at a multipoint control unit (520), where a number of base layer video images of a number of active speakers (535) and one or more enhancement layers (540) of a most active speaker are selected. the multipoint control unit (520) transmits the base layer video images, and one or more enhancement layers (540) of the most active speaker, to one or more of the plurality of multimedia user equipment (550, 560, 570, 580). The identification of the speakers is much improved compared to traditional videoconference systems, as the available bandwidth is shared to allow one enhancement layer and several base layers to be sent, instead of only one full quality video stream.



For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

- 1 -

Video Conferencing System And Method Of OperationField of the Invention

This invention relates to video conferencing. The
5 invention is applicable to, but not limited to, a video
switching mechanism in H.323 and/or SIP based centralised
videoconferences, using layered video coding.

Background of the Invention

10 As the pace of business accelerates, and relationships
spread around the world, the need to bridge communication
distances quickly and economically has become a major
challenge. Bringing customers and staff together
efficiently is critical to being successful in an ever
15 more competitive marketplace. Businesses are looking for
flexible solutions that support real-time information
sharing across countries and continents using various
communication methods, such as voice, video, image data
and any combination thereof.

20

In particular, multi-national organisations have an
increasing desire to eliminate costly travel and link
multiple locations in order to let groups within the
organisation communicate more efficiently and effectively.
25 A multipoint conferencing system operating over an
Internet protocol (IP) network seeks to address this need.
In the field of this invention, it is known that terminals
exchange audio and video streams in real-time in
multipoint videoconferences. The conventional method to
30 set-up multipoint conferences over an IP network is to use
a Multipoint control unit (MCU). The MCU is an endpoint
on the network that provides the capability for three or
more terminals and/or communication gateways to

- 2 -

participate in a multipoint conference. The MCU may also connect two terminals in a point-to-point conference such that they have the ability to evolve into a multipoint conference.

5

Referring first to FIG. 1, a known centralised conferencing model 100 is shown. Centralised conferences make use of an MCU-based conference bridge. All terminals (endpoints) 120, 122, 125 send and receive media
10 information 130 in the form of audio, video, and/or data signals, as well as control information streams 140, to/from the MCU 110. These transmissions are done in a point-to-point fashion. This is shown in Figure 1.

15 An MCU 110 consists of a Multipoint Controller (MC), and zero or more Multipoint Processors (MP). The MC handles the call set-up and call signalling negotiations between all terminals, to determine common capabilities for audio and video processing. The MC 110 does not deal directly
20 with any of the media streams. This is left to the MP, which mixes, switches, and processes audio, video, and/or data bits.

In this manner, MCUs provide the ability to host multi-
25 location seminars, sales meetings, group conferences and other 'face-to-face' communications. It is also known that multipoint conferences can be used in various applications, for example:

30 (i) Executives and managers at multiple locations can meet 'face-to-face', share real-time information, and make decisions more quickly without any loss of time, expense, and demands of travelling;

- 3 -

(ii) Project teams and knowledge workers can coordinate individual tasks, and view and revise shared documents, presentations, designs, and files in a real time manner; and

- 5 (iii) Students, trainees, and employees at remote locations can access shared educational/training resources across any distance or time zones.

Consequently, it is envisaged that MCU-based systems will
10 play an important role in multimedia communications over IP based networks in the future.

Such multimedia communication often employs video transmission. In such transmissions, a sequence of
15 images, often referred to as frames, is transmitted between transmitting and receiving units. Multipoint multimedia conference systems can be set-up using various methods, for example as specified by the H.323 and SIP session layer protocol standards. References for SIP can
20 be found at:

<http://www.ietf.org/rfc/rfc2543.txt>, and
<http://www.cs.columbia.edu/~hgs/sip>.

Furthermore, for example in systems using ITU H.263 video
25 compression [ITU-T Recommendation, H.263, 'Video Coding for Low Bit Rate Communication'], the first frame of a video sequence includes a comprehensive amount of image data, generally referred to as intra coded information. The intra-coded frame, as it is the first frame, provides
30 a substantial portion of the image to be displayed. This intra-coded frame is followed by inter-coded (predicted) information, which generally includes data relating to changes in the image that is being transmitted. Hence,

- 4 -

predicted inter-coded information contains much less information than intra-coded information.

In traditional multimedia conferencing system, the users
5 need to identify themselves when they speak, so that the receiving terminals know who is speaking. Clearly if the transmitting terminal fails to identify itself, the listening users will have to guess who is speaking.

10 A known technique solves this problem by analysing the audio streams and forwarding the name and video stream of the active speaker to all the participants. In a centralized conferencing system, the MCU often performs this function. The MCU can then send the name of the
15 speaker and the corresponding video and audio stream to all the participants by switching the appropriate input multimedia stream to the output ports/paths.

Video switching is a well-known technique that aims at
20 delivering to each endpoint a single video stream, equivalent to arranging multiple point-to-point sessions. The video switching can be:

- (i) Voice activated switching, where the MCU transmits the video of the active speaker.
- 25 (ii) Timed activated switching, where the video of each participant is transmitted one after another at a predetermined time interval.
- (iii) Individual video selection switching, where each endpoint can request the participant video stream that
30 he/she wishes to receive.

Referring now to FIG. 2, a functional diagram of a traditional video switching mechanism 200 is shown. In a

- 5 -

traditional centralised conferencing system, the video switching is performed as follows. The MCU 220, for example positioned within an Internet protocol (IP) based network 210, contains a switch 230. The MCU 220 receives
5 the video streams 255, 265, 275, 285 of all the participants (user equipment) 250, 260, 270, 280. The MCU may also receive, separately, a combined (multiplexed) audio stream 290 from the participants who are speaking. The MCU 220 then selects one of the video streams and
10 sends this video stream 240 to all the participants 250, 260, 270, 280.

Such traditional systems have the disadvantage that they only send the video stream of the active speaker. The
15 users might still have a problem in identifying the speaker of the video stream if several speakers are talking at the same time, or if the active speaker is constantly changing. This is particularly the case with large videoconferences.

20

Alternatively, the video of each participant can be sent to all the participants. However, this approach suffers in a wireless based conference due to the bandwidth limitation.

25

In the field of video technology, it is known that video is transmitted as a series of still images/pictures. Since the quality of a video signal can be affected during coding or compression of the video signal, it is known to
30 include additional information 'layers' based on the difference between the video signal and the encoded video bit stream. The inclusion of additional layers enables the quality of the received signal, following decoding

- 6 -

and/or decompression, to be enhanced. Hence, a hierarchy of pictures and enhancement pictures partitioned into one or more layers is used to produce a layered video bit stream.

5

In a layered (scalable) video bit stream, enhancements to the video signal may be added to the base layer either by:

- (i) Increasing the resolution of the picture (spatial scalability);
- 10 (ii) Including error information to improve the Signal to Noise Ratio of the picture (SNR scalability); or
- (iii) Including extra pictures to increase the frame rate (temporal scalability).

- 15 Such enhancements may be applied to the whole picture, or to an arbitrarily shaped object within the picture, which is termed object-based scalability. In order to preserve the disposable nature of the temporal enhancement layer, the H.263+ standard dictates that pictures included in the
- 20 temporal scalability mode should be bi-directionally predicted (B) pictures, as shown in the video stream of FIG. 3.

- FIG. 3 shows a schematic illustration of a scalable video arrangement 300 illustrating B picture prediction dependencies, as known in the field of video coding techniques. An initial intra-coded frame (I_1) 310 is followed by a bi-directionally predicted frame (B_2) 320. This, in turn, is followed by a (uni-directional)
- 25 predicted frame (P_3) 330, and again followed by a second bi-directionally predicted frame (B_4) 340. This again, in
- 30

- 7 -

turn, is followed by a (uni-directional) predicted frame (P₅) 350, and so on.

FIG. 4 is a schematic illustration of a layered video arrangement, known in the field of video coding techniques. A layered video bit stream includes a base layer 405 and one or more enhancement layers 435.

The base layer (layer 1) includes one or more intra-coded pictures (I pictures) 410 sampled, coded and/or compressed from the original video signal pictures. Furthermore, the base layer will include a plurality of predicted inter-coded pictures (P pictures) 420, 430 predicted from the intra-coded picture(s) 410.

In the enhancement layers (layers 2 or 3 or more) 435, three types of picture may be used:

- (i) Bi-directionally predicted (B) pictures (not shown);
- (ii) Enhanced intra (EI) pictures 440 based on the intra-coded picture(s) 410 of the base layer 405; and
- (iii) Enhanced predicted (EP) pictures 450, 460, based on the inter-coded predicted pictures 420, 430 of the base layer 405.

The vertical arrows from the lower layer illustrate that the picture in the enhancement layer is predicted from a reconstructed approximation of that picture in the reference (lower) layer.

In summary, scalable video coding has been used with multicast multimedia conferences, and only in the context of point-to-point or multicast video communication. However, wireless networks do not currently support

- 8 -

multicasting. Furthermore, with multicasting, each layer is sent in separate multicast sessions, with the receiver deciding itself whether to register to one or more sessions.

5

A need therefore exists for an improved video conferencing arrangement and method of operation, wherein the abovementioned disadvantages may be alleviated.

10 Statement of Invention

In accordance with the present invention there is provided a method of relaying video images in a multimedia videoconference, as claimed in claim 1, a video conferencing arrangement for relaying video images, as
15 claimed in claim 7, a wireless device for participating in a videoconference, as claimed in claim 11, a multipoint processor, as claimed in claim 12, a video communication system, as claimed in claim 16, a media resource function, as claimed in claim 18, a video communication unit, as
20 claimed in claim 19 or claim 20, a storage medium, as claimed in claim 23. Further aspects of the present invention are as claimed in the dependent claims.

In summary, the inventive concepts of the present
25 invention address the disadvantages of prior art arrangements by providing a video switching method to improve the identification of the participants and speakers in a videoconference. This invention makes use of layered video coding, in order to provide a better
30 usage of the bandwidth available for each user.

Brief Description of the Drawings

FIG. 1 shows a known centralised conferencing model.

- 9 -

FIG. 2 shows a functional diagram of a traditional video switching mechanism.

FIG. 3 is a schematic illustration of a video arrangement showing picture prediction dependencies, as known in the field of video coding techniques.

FIG. 4 is a schematic illustration of a layered video arrangement, known in the field of video coding techniques.

Exemplary embodiments of the present invention will now be described, with reference to the accompanying drawings, in which:

FIG. 5 shows a functional diagram of a video switching mechanism, in accordance with a preferred embodiment of the invention.

FIG. 6 shows a functional block diagram/flowchart of a multipoint processing unit, in accordance with a preferred embodiment of the invention.

FIG. 7 shows a video display of a wireless device participating in a videoconference using the preferred embodiment of the present invention.

FIG. 8 shows a UMTS (3GPP) communication system adapted in accordance with the preferred embodiment of the present invention.

25 Description of Preferred Embodiments

In summary, the preferred embodiment of the present invention proposes a new video switching mechanism for multimedia conferences that makes use of layered video coding. Previously, layered video coding has only been used to partition a video bit stream into more than one layer: a base layer and one or several enhancement layers, as described above with respect to FIG. 4. These known

- 10 -

techniques for scalable video communication are described in detail in standards such as H.263 and MPEG-4.

However, the inventor of the present invention has
5 recognised the benefits to be gained by adapting the concept of layered video coding and applying the adapted concepts to multimedia videoconference applications. In this manner, the present invention defines a different type of scalable video coding focused for use in
10 multimedia conferences, in contrast to point-to-point or multicast video communication.

Referring now to FIG. 5, a functional block diagram 500 of a video switching mechanism is shown, in accordance with
15 the preferred embodiment of the invention. In contrast to a traditional centralised conferencing system, the video switching is performed as follows. The MCU 520, for example positioned within an Internet protocol (IP) based network 510, contains a switch 530.

20 It is noteworthy that the MCU 520 receives 'layered' video streams including a base layer 552, 562, 572, 582 and one or more enhancement layer streams 555, 565, 575, 585 of all the participants (user equipment) 550, 560, 570, 580.
25 Only one enhancement layer video stream is per participant shown for clarity purposes only.

The MCU 520 may also receive, separately, a combined (multiplexed) audio stream 590 from the participants. The
30 MCU 520 then selects the base layer video streams of a number of active speakers 535 and the enhancement layer 540 of the most active speaker, using switch 530. The MCU

- 11 -

520 then sends these video streams 535, 540 to all the participants 550, 560, 570, 580.

The selection process to determine the most active speaker is preferably performed by the MCU 520 analysing the audio streams 590 in order to determine first whom all the active speakers are. The most active speaker is then preferably determined in the multipoint processor unit, as described with reference to FIG. 6. The one or more base layers and one enhancement layer are preferably sent to the participants according to a priority level based on the activity of each participant.

In order to effect the improved, but more complex, video switching mechanism of FIG. 5, the multipoint processing unit (MP) 600 has been adapted to facilitate the new video switching mechanism, in accordance with a preferred embodiment of the invention and as shown in FIG. 6.

The MP 600 still receives the audio stream 590 from the participants' video/multimedia communication units, through a packet-filtering module 610 and routes this audio stream to a packet routing module 630. However, the audio stream is now also routed to a speaker identification module 620 that analyses the audio streams 590 in order to determine who are the active speakers. The speaker identification module 620 allocates a priority level based on the activity of each participant and determines:

30

- (i) The most active speaker 622,
- (ii) Any other active speakers 625, and by default
- (iii) Any remaining inactive speakers.

- 12 -

The speaker identification module 620 then forwards the priority level information to the switching module 640 that has been adapted to deal with priority level of
5 speakers, in accordance with the preferred embodiment of the present invention. Furthermore, the switching module 640 has been adapted to receive layered video streams, including video base layer streams 552, 562, 572 and 582 and video enhancement layer streams 555, 565, 575 and 585
10 from the participants' video communication units through the packet filtering module 610. The switching module 640 uses this speaker information to send the video base layers of the secondary (lesser) active speakers and the most active speaker and only the video enhancement layer
15 of the most active speaker, to all the participants, via the packet routing module 630.

The one or more receiving ports of the multipoint processor have therefore been adapted to receive layered
20 video streams, including base layer video streams 552, 562, 572 and 582 and enhancement layer video streams 555, 565, 575 and 585, from a plurality of user equipment 550, 560, 570 and 580. It is within the contemplation of the invention that the switching module 640, may only select
25 one base layer video image and corresponding one or more enhancement layers if it is determined that there is only one active speaker. This speaker then automatically is designated as the most active speaker for transmitting to one or more user equipment 550, 560, 570 and 580.

30

When the most active speaker is constantly changing, as can happen in videoconferences, the enhancement layer will be constantly switching. The inventor of the present

- 13 -

invention has recognised a potential problem with such constant and rapid switching. Under such circumstances, the first frame may need to be converted into an Intra frame (EI) if it was actually a predicted frame (EP) from
5 a speaker who was previously only a secondary active speaker.

To address this potential problem, the video base layer streams 552, 562, 572 and 582 and video enhancement layer
10 streams 555, 565, 575 and 585 from the packet-filtering module 610 are preferably input to a de-packetisation function 680. The de-packetisation function 680 de-multiplexes the video streams and provides the de-multiplexed video streams to a video decoder and buffer
15 function 670.

To synchronise and co-ordinate the video decoding, the video decoder and buffer function 670 receives the indication of the most active speaker 622. After
20 extracting the video stream information for the most active speaker, the video decoder and buffer function 670 provides bi-directionally predicted (BP) 675 and/or predicted (EP) video stream data of the most active speaker 622 to an 'EP frame to EI frame Transcoding
25 Module' 660. The 'EP frame to EI frame Transcoding Module' 660 processes the input video streams to provide the primary speaker enhancement layer video stream, as an Intra-coded (EI) frame.

30 The primary speaker enhancement layer video stream is then input to a packetisation function 650, where it is packetised and input to the switching module 640. The switching module 640 then combines the primary speaker

- 14 -

enhancement layer video stream, with the video base layer streams 552, 562, 572 and 582 of the secondary active speakers and routes the combined multimedia stream to the packet routing module 630. The packet routing module then
5 routes the information to the participants in accordance with the method of FIG. 5.

In the preferred embodiment of the present invention, the video switching module 640 uses the output of the 'EP
10 frame to EI frame Transcoding module' 660 when it determines that the primary speaker has changed.

It is within the contemplation of the invention that one or more modules that are similar to module 660 could also
15 be included in the MP 600 to perform the same function for the secondary speakers, when they are deemed to have changed. Otherwise, in the embodiment that uses a single 'EP frame to EI frame Transcoding module' 660 to transcode the video stream of only the primary speaker, when say an
20 inactive speaker becomes a secondary active speaker, the speaker identification module 620 (or switching module 640) may make a request for a new Intra-frame.

Alternatively, the switching module 640 may wait for a new Intra frame of the new secondary active speaker before
25 sending the corresponding video base layer stream to all the participants.

In addition to the preferred embodiment of the present invention, where more than one enhancement layer is
30 available for use, it is within the contemplation of the invention that more classes of speakers can be used. By using more classes of speakers, a finer scalability of the multimedia messages can be attained, as the identification

- 15 -

of speakers is improved, especially for large videoconferences.

It is also within the contemplation of the invention that
5 predicted frame to Intra frame conversion could be added
for one or more of the base layers streams. In this
manner, the switching module 640 can quickly switch
between the base layers without having to wait for a new
Intra frame.

10

FIG. 7 shows the video display 710 of a wireless device
700 taking part in a videoconference using the preferred
embodiment of the present invention. By implementing the
inventive concepts hereinbefore described, improved video
15 communication is achieved. In particular, for a given
bandwidth, the participants are now able to receive better
video quality of the most active speaker 720, by lowering
the video quality of the lesser (secondary) active
speakers 730, and providing no video for the inactive
20 speakers. In order to provide such improved video
conferencing, the video communication device receives the
enhancement layer and base layer of the most active
speaker 720, the base layers of the secondary active
speakers 730 and no video from inactive speakers.

25

In such a manner, a video communication unit can provide a
constantly updated video image of the most active speaker
in a larger, higher resolution display, whilst smaller
displays can display secondary (lesser) active speakers.

30

The wireless device 700 preferably has a primary video
display 710 for displaying a higher quality video image of
the most active speaker, and one or more second distinct

- 16 -

displays for displaying respective lesser active speakers. Preferably, the manipulation of the respective video images into the respective displays is performed by a processor (not shown) that is operably coupled to the video displays. The processor receives an indication of a most active speaker 720 and lesser active speakers, and determines which video image received should be displayed in the first display and which video image(s) received from the lesser active speakers 730 should be displayed in the second display. Advantageously, the second display may be configured to provide a lower quality video image of the lesser active speakers, thereby saving cost.

It is anticipated that MCU-based systems will facilitate multimedia communications over IP based networks in the future. Therefore, the inventor of the present invention envisages that the herein described techniques could be incorporated in any H.323/SIP based multipoint multimedia conferences or systems that make use of MCU.

A preferred application of the aforementioned invention is in the Third Generation Partnership Project (3GPP) specification for wide-band code-division multiple access (WCDMA) standard. In particular, the invention can be applied to the IP Multimedia Domain (described in the 3G TS 25.xxx series of specifications), which is planning to incorporate H.323/SIP MCU into the 3GPP network. The MCU will be hosted by the Media Resource Function (MRF) 890A, see Figure 8.

FIG. 8 shows, a 3GPP (UMTS) communication system/network 800, in a hierarchical form, which is capable of being adapted in accordance with the preferred embodiment of the

- 17 -

present invention. The communication system 800 is compliant with, and contains network elements capable of operating over, a UMTS and/or a GPRS air-interface.

- 5 The network is conveniently considered as comprising:
- (i) A user equipment domain 810, made up of:
 - (a) A user SIM (USIM) domain 820 and
 - (b) A mobile equipment domain 830; and
 - (ii) An infrastructure domain 840, made up of:
 - 10 (c) An access network domain 850, and
 - (d) A core network domain 860, which is, in turn, made up of (at least):
 - (di) a serving network domain 870, and
 - (dii) a transit network domain 880 and
 - 15 (diii) an IP multimedia domain 890, with multimedia being provided by SIP (ETF RFC2543).

In the mobile equipment domain 830, UE 830A receives data from a user SIM 820A in the USIM domain 820 via the wired
20 Cu interface. The UE 830A communicates data with a Node B 850A in the network access domain 850 via the wireless Uu interface. Within the network access domain 850, the Node Bs 850A contain one or more transceiver units and communicate with the rest of the cell-based system
25 infrastructure, for example RNC 850B, via an I_{ub} interface, as defined in the UMTS specification.

The RNC 850B communicates with other RNCs (not shown) via the I_{ur} interface. The RNC 850B communicates with a SGSN
30 870A in the serving network domain 870 via the I_u interface. Within the serving network domain 870, the SGSN 870A communicates with a GGSN 870B via the Gn interface, and the SGSN 870A communicates with a VLR

- 18 -

server 870C via the Gs interface. In accordance with the preferred embodiment of the present invention, the SGSN 870A communicates with the MCU (not shown) that resides within the media resource function (890A) in the IP
5 Multimedia domain 890. The communication is performed via the Gi interface.

The GGSN 870B (and/or SSGN) is responsible for UMTS (or GPRS) interfacing with a Public Switched Data Network
10 (PSDN) 880A such as the Internet or a Public Switched Telephone Network (PSTN). The SGSN 870A performs a routing and tunnelling function for traffic within say, a UMTS core network, whilst a GGSN 870B links to external packet networks, in this case ones accessing the UMTS mode
15 of the system.

The RNC 850B is the UTRAN element responsible for the control and allocation of resources for numerous Node Bs 850A; typically 50 to 100 Node B's may be controlled by
20 one RNC 850B. The RNC 850B also provides reliable delivery of user traffic over the air interfaces. RNCs communicate with each other (via the interface Iur) to support handover and macro diversity.

25 The SGSN 870A is the UMTS Core Network element responsible for Session Control and interface to the Location Registers (HLR and VLR). The SGSN is a large centralised controller for many RNCs.

30 The GGSN 870B is the UMTS Core Network element responsible for concentrating and tunnelling user data within the core packet network to the ultimate destination (e.g., an internet service provider (ISP)). Such user data includes

- 19 -

multimedia and related signalling data to/from the IP multimedia domain 890. Within the IP multimedia domain 890, the MRF is split into a Multimedia Resource Function Controller (MRFC) 892A and a Multimedia Resource Function
5 Processor (MRFP) 891A. The MRFC 892A provides the Multipoint Controller (MC) functionalities, whereas the MRFP 891A provides the Multipoint Processor (MP) functionalities, as described previously.

10 The protocol used across the Mr reference point/interface 893A is SIP (as defined by RFC 2543). The call-state control function (CSCF) 895A acts as a call server and handles multimedia call signalling.

15 Thus, in accordance with the preferred embodiment of the invention the elements SGSN 870A, GGSN 870B and all parts within the MRF 890A are adapted to facilitate multimedia messages as herein before described. Furthermore, the UE 830A, Node B 850A and RNC 850B may also be adapted to
20 facilitate improved multimedia messages as hereinbefore described.

More generally, the adaptation may be implemented in the respective communication units in any suitable manner.
25 For example, new apparatus may be added to a conventional communication unit, or alternatively existing parts of a conventional communication unit may be adapted, for example by reprogramming one or more processors therein. As such, the required adaptation may be implemented in the
30 form of processor-implementable instructions stored on a storage medium, such as a floppy disk, hard disk, PROM, RAM or any combination of these or other storage multimedia.

- 20 -

It is also within the contemplation of the invention that such adaptation of multimedia messages may alternatively be controlled, implemented in full or implemented in part
5 by adapting any other suitable part of the communication system 800.

Although the above elements are typically provided as discrete and separate units (on their own respective
10 software/hardware platforms), divided across the mobile equipment domain 830, access network domain 850 and the serving network domain 870, it is envisaged that other configurations can be applied.

15 Further, in the case of other network infrastructures, such as a GSM network, implementation of the processing operations may be performed at any appropriate node such as any other appropriate type of base station, base station controller, mobile switching centre or operational
20 and management controller, etc. Alternatively, the aforementioned steps may be carried out by various components distributed at different locations or entities within any suitable network or system.

25 The video conferencing method using layered video coding, preferably when applied in a centralised videoconference, as described above, provides the following advantages:
(i) The identification of the speakers is much improved compared to traditional systems, because the bandwidth is
30 shared to allow one or more enhancement layers and several base layers to be sent instead of only one full quality video stream.

- 21 -

(ii) The video switching when the active speaker changes is much smoother using the inventive concepts herein described, because it defines several states active speaker, second most active speakers, inactive speakers.

5 (iii) The video quality of the most active speaker is improved.

(iv) Improved video communication units can display a variety of speakers, with each displayed image being dependent upon a priority level associated with the
10 respective video communication unit's transmission.

A method of relaying video images in a multimedia videoconference between a plurality of multimedia user equipment has been described. The method includes the
15 steps of transmitting layered video images by a number of the plurality of user equipment wherein the layered video images include a base layer and one or more enhancement layers and receiving the transmitted layered video images at a multipoint control unit. A number of base layer
20 video images of a number of active speakers are selected and one or more enhancement layers of a most active speaker. The multipoint control unit transmits the number of base layer video images of a number of active speakers and one or more enhancement layers of the most active
25 speaker to one or more of the plurality of multimedia user equipment.

In addition, a video conferencing arrangement for relaying video images between a plurality of user equipment has
30 been described. Furthermore, a wireless device for participating in a videoconference has been described, where a number of participants transmit video images.

- 22 -

Claims

1. A method of relaying video images in a multimedia videoconference between a plurality of multimedia user equipment (550, 560, 570, 580), the method comprising the
5 steps of:

transmitting layered video images by a number of said plurality of user equipment, wherein said layered video images include a base layer (552, 562, 572, 582) and one or more enhancement layers (555, 565, 575, 585);

10 receiving said transmitted layered video images at a multipoint control unit (520);

selecting a number of base layer video images of a number of active speakers (535) and one or more enhancement layers (540) of a most active speaker; and

15 transmitting, by said multipoint control unit (520) said number of base layer video images of a number of active speakers (535) and one or more enhancement layers (540) of the most active speaker to one or more of the plurality of multimedia user equipment (550, 560, 570,
20 580).

2. The method of relaying video images in a multimedia videoconference according to Claim 1, wherein the step of selecting further comprises the step of:

25 analysing a number of audio data streams (590), transmitted by said plurality of multimedia user equipment (550, 560, 570, 580), in order to determine the number of active speakers and/or said most active speaker.

30 3. The method of relaying video images in a multimedia videoconference according to Claim 1 or Claim 2, the method further characterised by the step of:

- 23 -

assigning a priority level to each layered video image and/or said audio data stream transmitted by a respective user equipment; and

5 selecting a number of base layer video images (535) and one or more enhancement layers (540) for transmitting to said one or more of said plurality of multimedia user equipment (550, 560, 570, 580), based on said assigned priority level.

10 4. The method of relaying video images in a multimedia videoconference according to any preceding Claim, the method further characterised by the step of:
transcoding (660) a first predicted frame of a video image of the most active speaker to an intra-coded frame, for
15 enhancing the video quality of the most active speaker.

5. The method of relaying video images in a multimedia videoconference according to any preceding Claim, the method further characterised by the step of:
20 receiving by said multipoint control unit (520), when more than one enhancement layer is available, an indication of a class of said one or more speakers with each layered video image transmission, in order to provide a finer scalability of said video images.

25 6. The method of relaying video images in a multimedia videoconference according to any preceding Claim, the method further characterised by the step of:
converting a predicted frame into an Intra-coded
30 frame for one or more base layer video streams.

- 24 -

7. A video conferencing arrangement for relaying video images between a plurality of user equipment (550, 560, 570, 580), the video conferencing arrangement comprising:

5 a multipoint control unit (520), adapted to receive a number of layered video images transmitted by a number of said plurality of user equipment, wherein said layered video images include a base layer (552, 562, 572, 582) and one or more enhancement layers (555, 565, 575, 585); and
10 a video switching module (530), operably coupled to said multipoint control unit (520) and adapted to select a number of base layer video images of a number of active speakers (535) and one or more enhancement layers (540) of a most active speaker; wherein

said multipoint control unit (520) being further
15 adapted to transmit said number of base layer video images of a number of active speakers (535) and one or more enhancement layers (540) of the most active speaker to one or more of the plurality of user equipment (550, 560, 570, 580).

20

8. The video conferencing arrangement according to Claim 7, further characterised by:

a predicted frame to intra-coded frame transcoding
module (660), operably coupled to said video switching
25 module (530), to provide a most active speaker enhancement layer video stream, as an Intra-coded frame, if said multipoint control unit (520) received said frame initially as a predicted frame.

30 9. The video conferencing arrangement according to Claim 7 or Claim 8, further characterised by:

- 25 -

a speaker identification module (620) that analyses a number of audio streams (590) in order to determine a number of active speakers and/or said most active speaker.

5 10. The video conferencing arrangement according to Claim 9, wherein said speaker identification module (620) allocates a priority level based on a determined activity of each participant to determine one or more of: a most active speaker (622), any other active speakers (625), and
10 any inactive speakers.

11. A wireless device (700) for participating in a videoconference where a plurality of participants transmit video images, the wireless device (700) comprising:

15 a video display (710) having a first display and one or more second distinct displays for displaying respective participants (720, 730) from the plurality of participants; and

a processor, operably coupled to said video
20 display, for receiving an indication of a most active speaker (720) and less active speakers (730), and determining that said video image received from said most active speaker (720) should be displayed in said first display offering a higher quality video image, and said
25 video images received from said number of said lesser active speakers (730) should be displayed in said one or more second display offering a lower quality video image.

12. A multipoint processor comprising:

30 one or more receiving ports adapted to receive layered video streams, including base layer video streams (552, 562, 572 582) and enhancement layer video streams

- 26 -

(555, 565, 575, 585), from a plurality of user equipment (550, 560, 570, 580); and

a switching module (640), operably coupled to said one or more receiving ports, selecting a number of base
5 layer video images of a number of active speakers (535) and one or more enhancement layers (540) of the most active speaker, for transmitting to one or more user equipment (550, 560, 570, 580).

10 13. The multipoint processor according to Claim 12, further characterised by:

a speaker identification module (620) operably coupled to said one or more receiving ports for analysing a number of audio streams (590) received from a number of
15 said plurality of user equipment, in order to determine a number of active speakers and/or said most active speaker.

14. The multipoint processor according to Claim 12 or Claim 13, wherein said speaker identification module (620)
20 allocates a priority level based on a determined activity of a number of participants to determine one or more of the following: a most active speaker (622), any other active speakers (625), and any inactive speakers.

25 15. The multipoint processor according to any of preceding Claims 12 to 14, further characterised by:

a predicted frame to intra-coded frame transcoding Module (660) operably coupled to said switching module (640), to convert the enhancement layer video stream of
30 said most active speaker to an Intra-coded frame if it has been received at a respective port as a predicted frame.

- 27 -

16. A video communication system adapted to perform the method steps of any of claims 1 to 6, or adapted to incorporate the video conferencing arrangement of any of Claims 7 to 10, or adapted to incorporate the multipoint
5 processor of any of Claims 12 to 15.

17. The video communication system according to Claim 16, wherein the video communication system is compatible with the UMTS communication standard (800) having an
10 Internet Protocol multimedia domain (890) to facilitate videoconferencing communication.

18. A media resource function (890A) adapted to perform the method steps of any of claims 1 to 6, or adapted to
15 incorporate the video conferencing arrangement of any of Claims 7 to 10, or adapted to incorporate the multipoint processor of any of Claims 12 to 15.

19. A video communication unit (700) adapted to receive
20 layered videoconference images generated in accordance with the method of claims 1 to 6.

20. A video communication unit adapted to generate layered videoconference images for use in the method of
25 claims 1 to 6, or to transmit layered videoconference images generated in accordance with the method of claims 1 to 6.

21. The video communication unit according to Claim 19,
30 wherein the video communication unit is one of:

a Node B (850A), a RNC (850B), a SGSN (870A), a GGSN (870B), a MRF (890A).

- 28 -

22. The method of relaying video images in a multimedia
videoconference of claims 1 to 6 or the video conferencing
arrangement of any of Claims 7 to 10, or multipoint
processor of any of Claims 12 to 15 or the video
communication system of Claim 16 or 17, or the media
resource function (890A) of Claim 18 or the video
communication unit of Claim 19, 20, or 21, adapted to
facilitate videoconference images based on the H.323
standard or SIP standard.

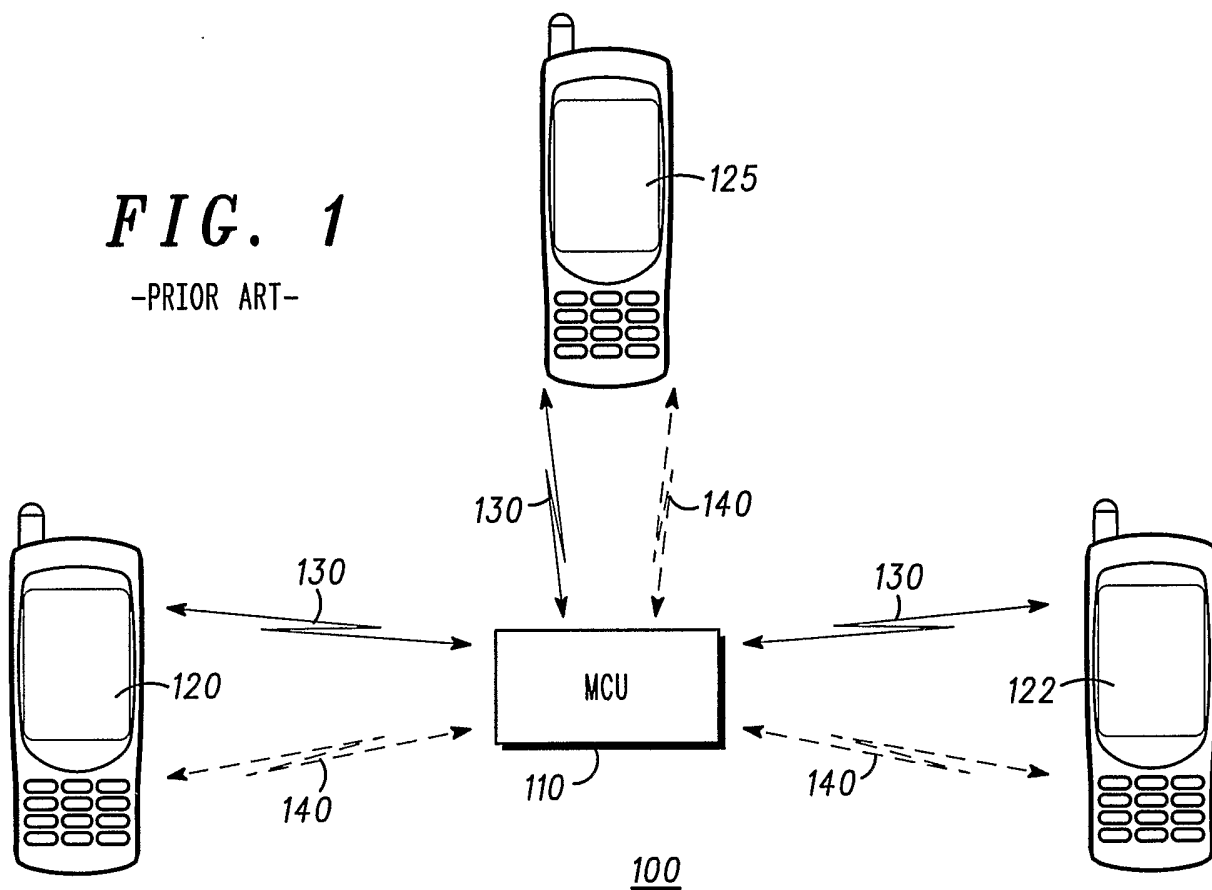
10

23. A storage medium storing processor-implementable
instructions for controlling a processor to carry out the
method of any of claims 1 to 6.

1/4

FIG. 1

-PRIOR ART-

**FIG. 2**

-PRIOR ART-

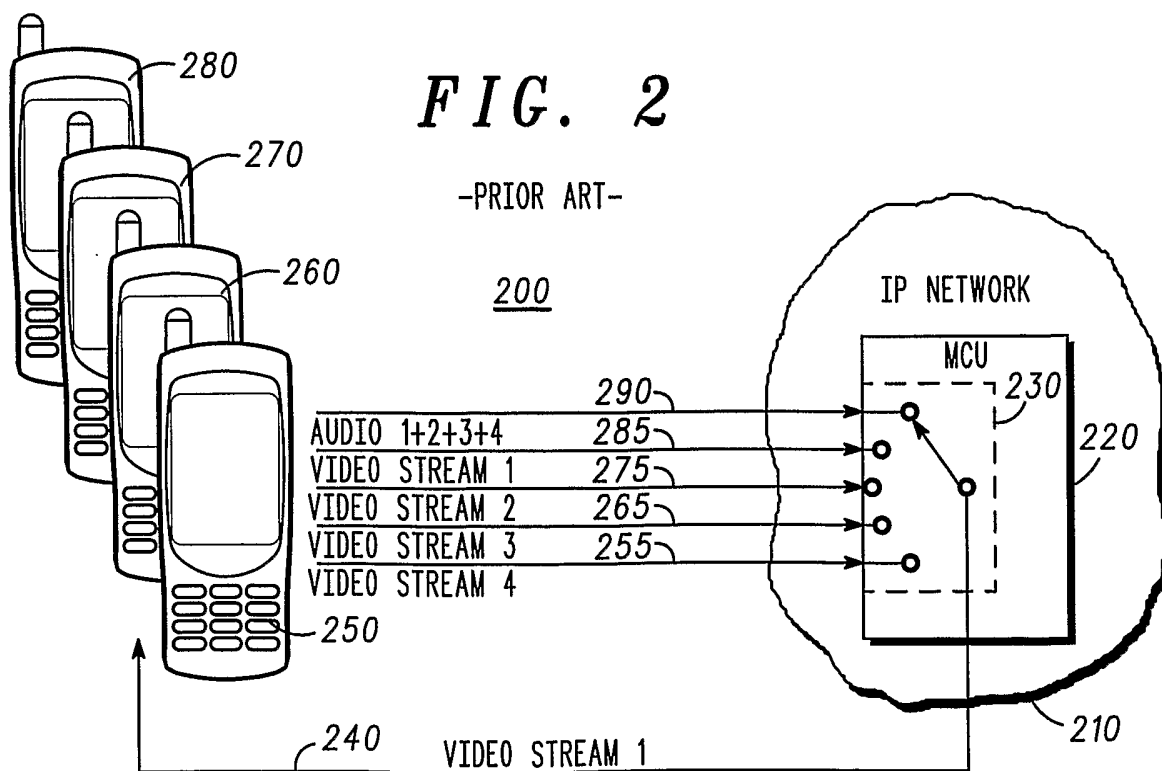


FIG. 3

-PRIOR ART-

300

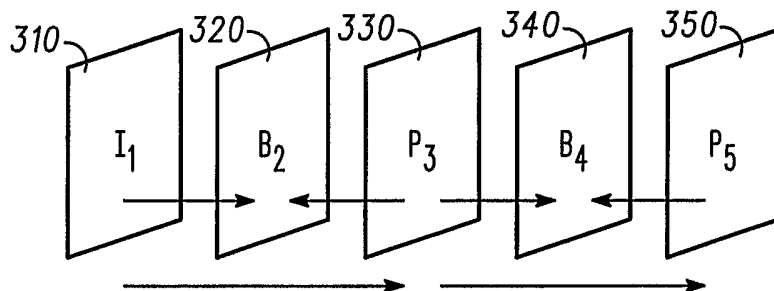


FIG. 4

-PRIOR ART-

400

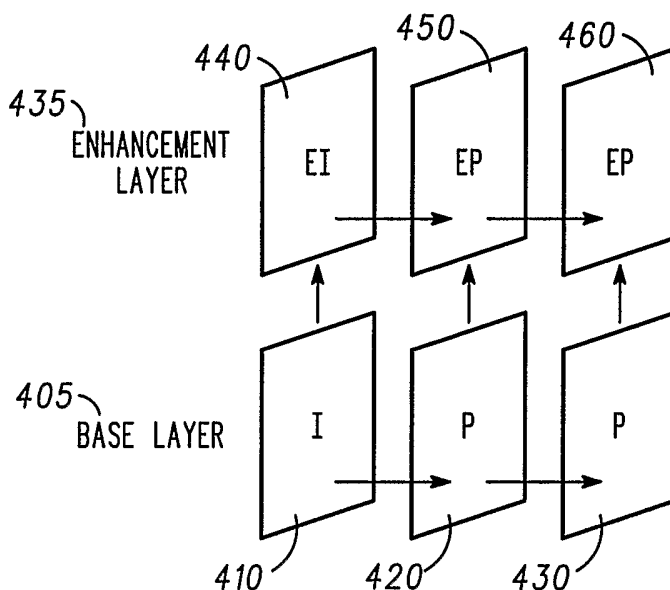
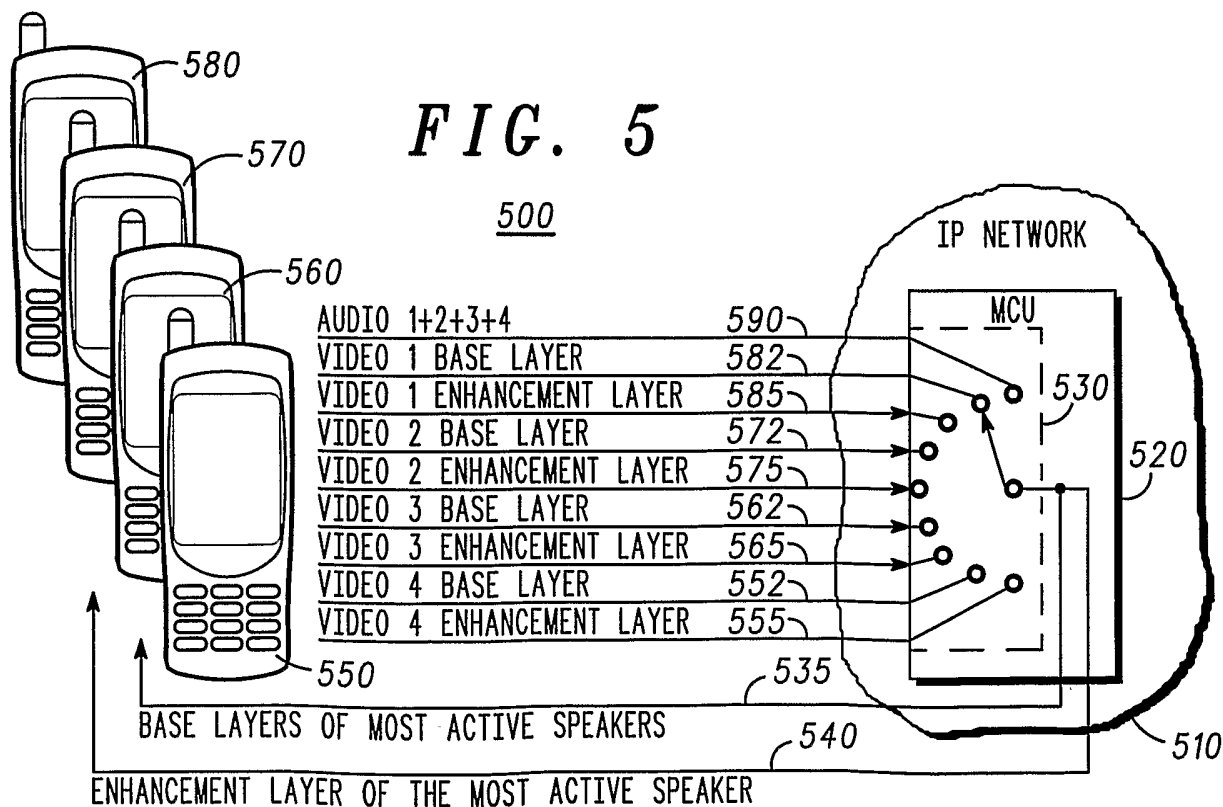
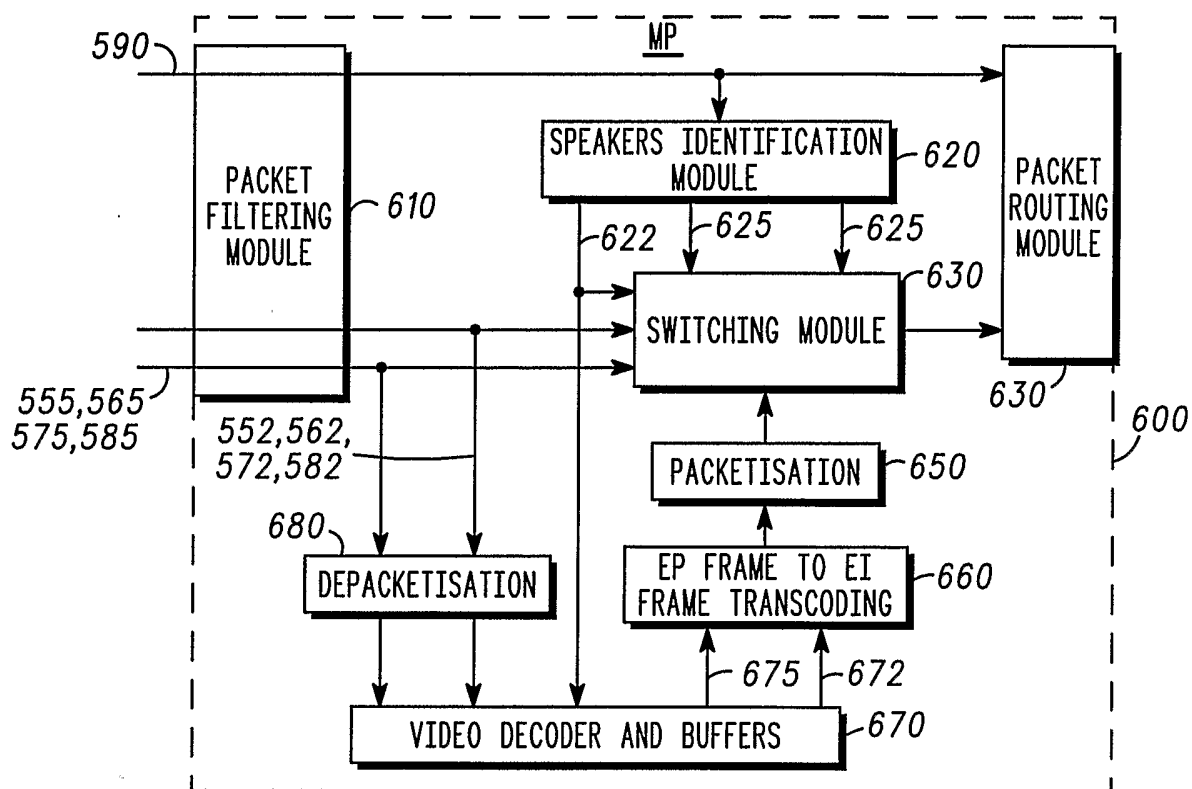
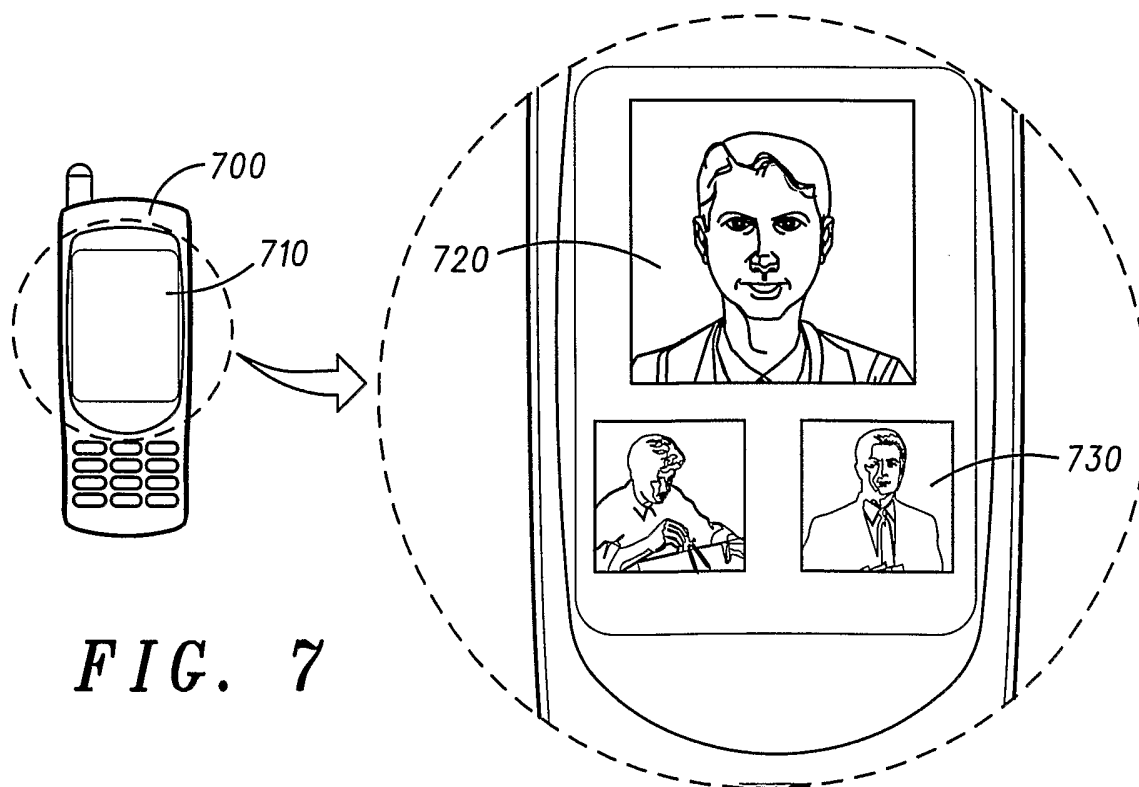
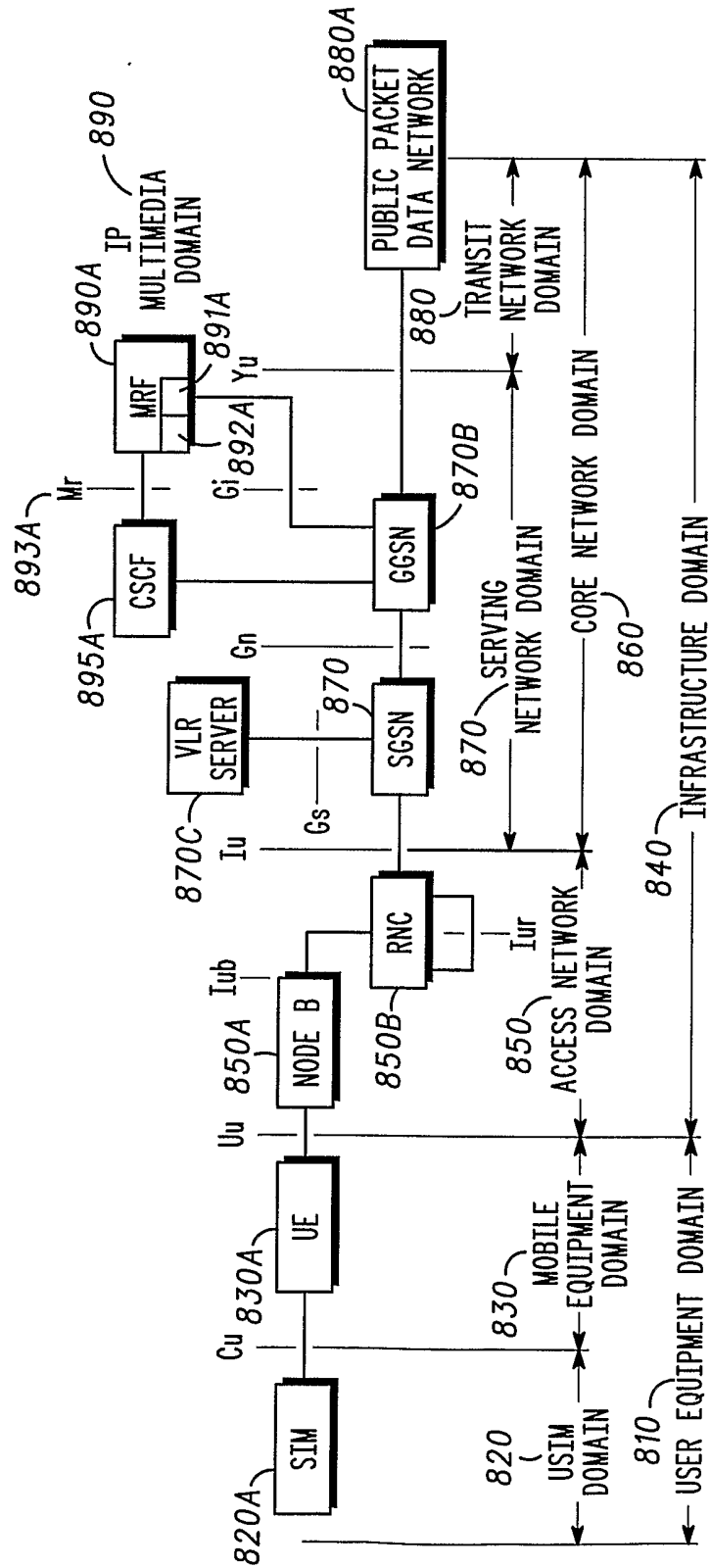


FIG. 5

500



**FIG. 6****FIG. 7**



800

FIG. 8

INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP 02/14337

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 H04N7/14 H04N7/15 H04N7/24

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 H04N

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

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP 0 905 976 A (MATSUSHITA ELECTRIC IND CO LTD) 31 March 1999 (1999-03-31) page 9, column 15, line 26 -page 11, column 19, line 27 page 19, column 35, line 16 -page 20, column 37, line 13 figures 1-4,46,47	1-23
A	EP 0 711 080 A (AT & T CORP) 8 May 1996 (1996-05-08) page 3, column 4, line 38 -page 4, column 6, line 5 page 12, column 22, line 23 -page 15, column 27, line 39 figure 2	1-23
	--- -/--	



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

* Special categories of cited documents:

- *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 when the document is taken alone
- *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.
- *Z* document member of the same patent family

Date of the actual completion of the international search

3 April 2003

Date of mailing of the international search report

10/04/2003

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Van der Zaai, R

INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP 02/14337

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>WO 98 26592 A (ERNST DETLEF ;HAFERBECK RALF (DE); SIEMENS AG (DE)) 18 June 1998 (1998-06-18) page 7, line 5 -page 14, line 18 figures 1-3 -----</p>	1-17

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/EP 02/14337

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
EP 0905976	A	31-03-1999	EP 0905976 A1	31-03-1999
			CN 1227031 T	25-08-1999
			JP 11225168 A	17-08-1999
			WO 9842132 A1	24-09-1998
			KR 2000011118 A	25-02-2000
			TW 410298 B	01-11-2000
EP 0711080	A	08-05-1996	US 5629736 A	13-05-1997
			CA 2159847 A1	02-05-1996
			EP 0711080 A2	08-05-1996
			JP 8237621 A	13-09-1996
WO 9826592	A	18-06-1998	AT 233458 T	15-03-2003
			WO 9826592 A1	18-06-1998
			DE 59709413 D1	03-04-2003
			EP 0941609 A1	15-09-1999
			JP 2000506709 T	30-05-2000
			US 6314302 B1	06-11-2001