US 20070291108A1

(19) United States

(12) Patent Application Publication

(10) Pub. No.: US 2007/0291108 A1

(43) Pub. Date: Dec. 20, 2007

Huber et al.

(54) **CONFERENCE LAYOUT CONTROL AND CONTROL PROTOCOL**

(75) Inventors: **Richard E. Huber**, Harmony, PA (US); **Arun Punj**, Cranberry Township, PA (US)

Correspondence Address:
Ansel M. Schwartz
Suite 304, 201 N. Craig Street
Pittsburgh, PA 15213

(73) Assignee: **Ericsson, Inc.**

(21) Appl. No.: **11/800,838**

(22) Filed: **May 8, 2007**

Related U.S. Application Data

(60) Provisional application No. 60/814,476, filed on Jun. 16, 2006, provisional application No. 60/814,477, filed on Jun. 16, 2006, provisional application No. 60/814,491, filed on Jun. 16, 2006.

Publication Classification

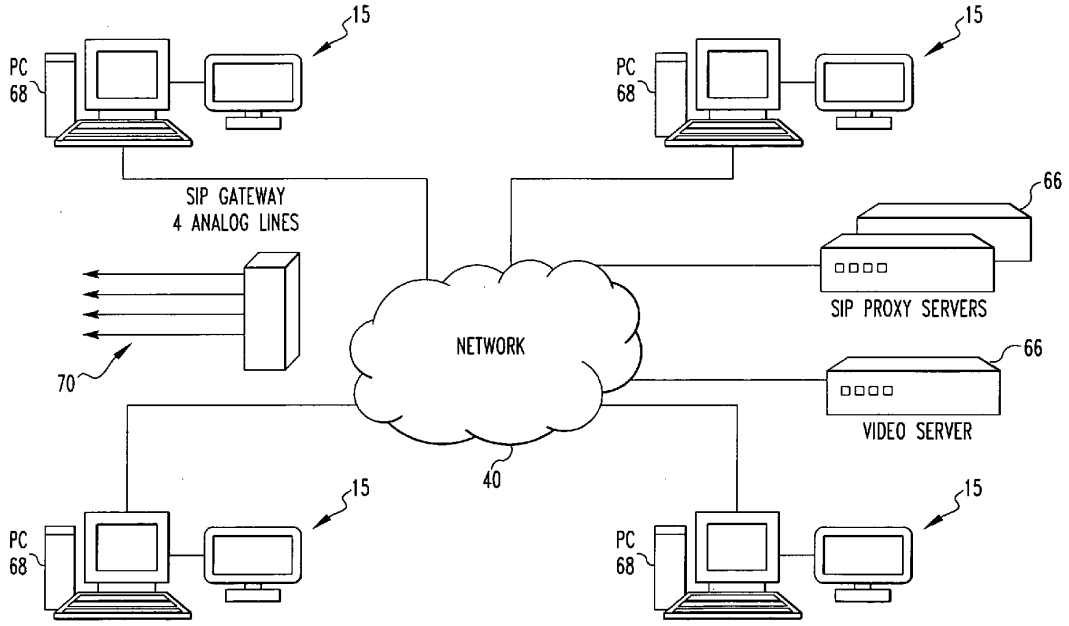
(51) **Int. Cl.**

H04N 7/14	(2006.01)
H04L 12/16	(2006.01)
H04Q 11/00	(2006.01)

(52) **U.S. Cl.** **348/14.02**; 370/260

(57) **ABSTRACT**

A teleconferencing system includes a network. The system includes a plurality of nodes in communication with each other to form a conference. Each node having a video display with a display layout, at least one of the nodes individually controlling at least partially the display layout of each node in the conference with a specific format able to be unique for each node. A method for providing a teleconference including the steps of forming a conference with a plurality of nodes in communication with each other through a network. Each node having a video display with a display layout. There is the step of controlling individually with at least one of the nodes at least partially the display layout of each node in the conference with a specific format able to be unique for each node. A teleconferencing node. A teleconferencing system includes a network. The system includes a plurality of nodes in communication with each other through the network to form a conference. Each node communicating only a change to the other nodes when the change occurs. A method for conducting a telecommunications conference between at least three parties includes the steps of establishing a conference between the party. There is the step of making a change to the conference. There is the step of communicating only the change to the parties.



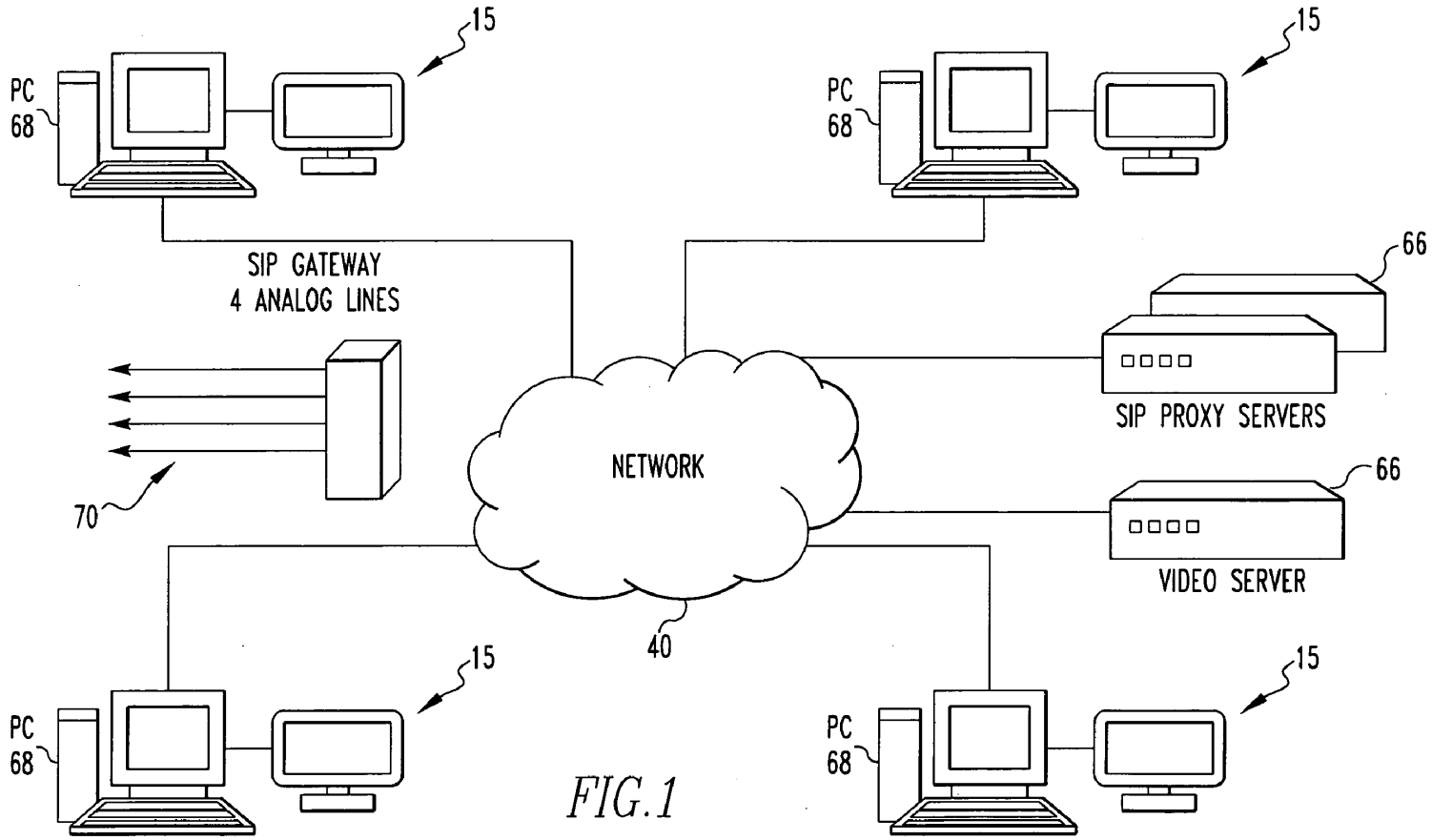

15
PC
68
SIP GATEWAY
4 ANALOG LINES
70
NETWORK
40
66
SIP PROXY SERVERS
VIDEO SERVER

FIG.1

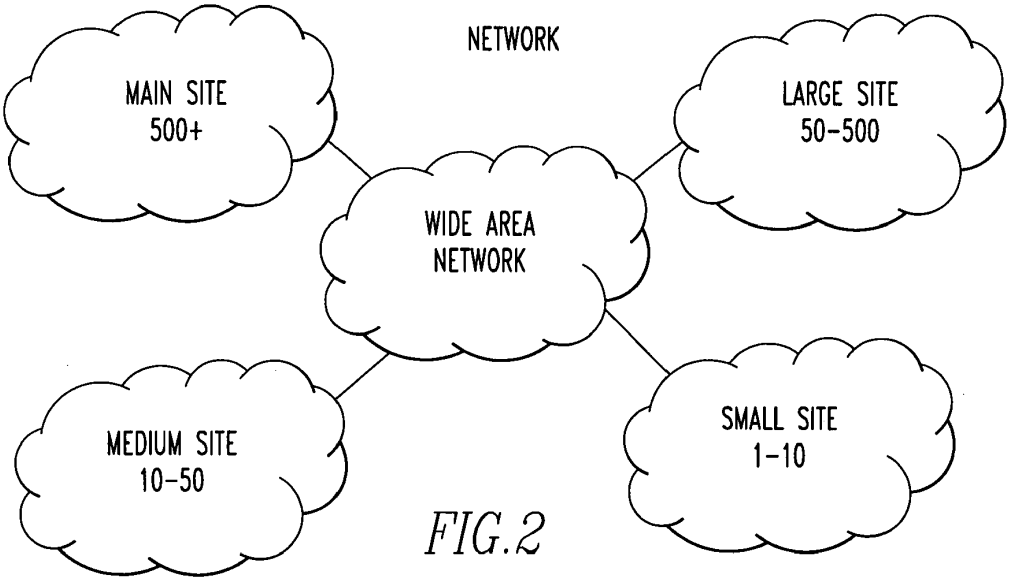

NETWORK
MAIN SITE
500+
LARGE SITE
50-500
WIDE AREA
NETWORK
MEDIUM SITE
10-50
SMALL SITE
1-10

FIG.2

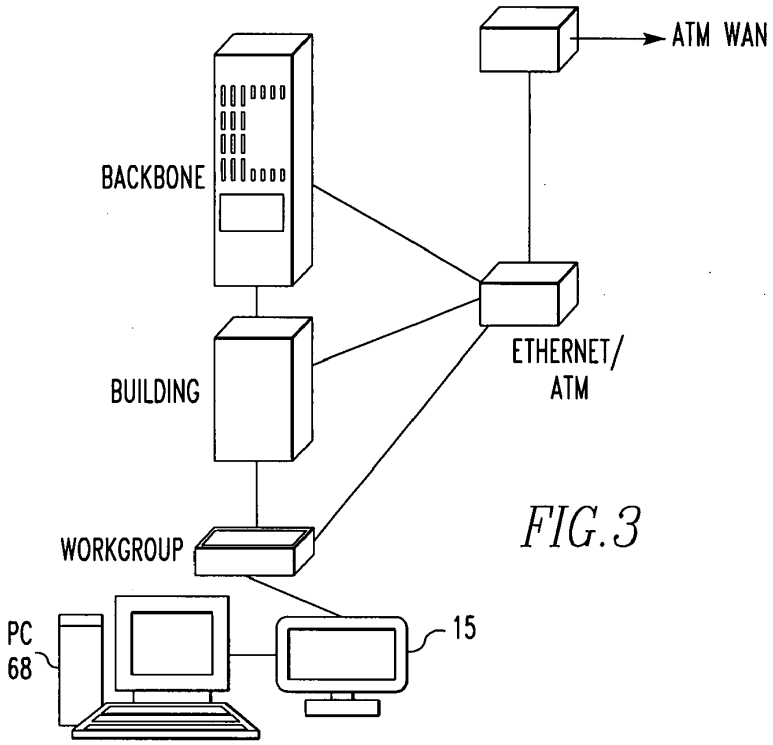

ATM WAN
BACKBONE
ETHERNET/
ATM
BUILDING
WORKGROUP
PC
68
15

FIG.3

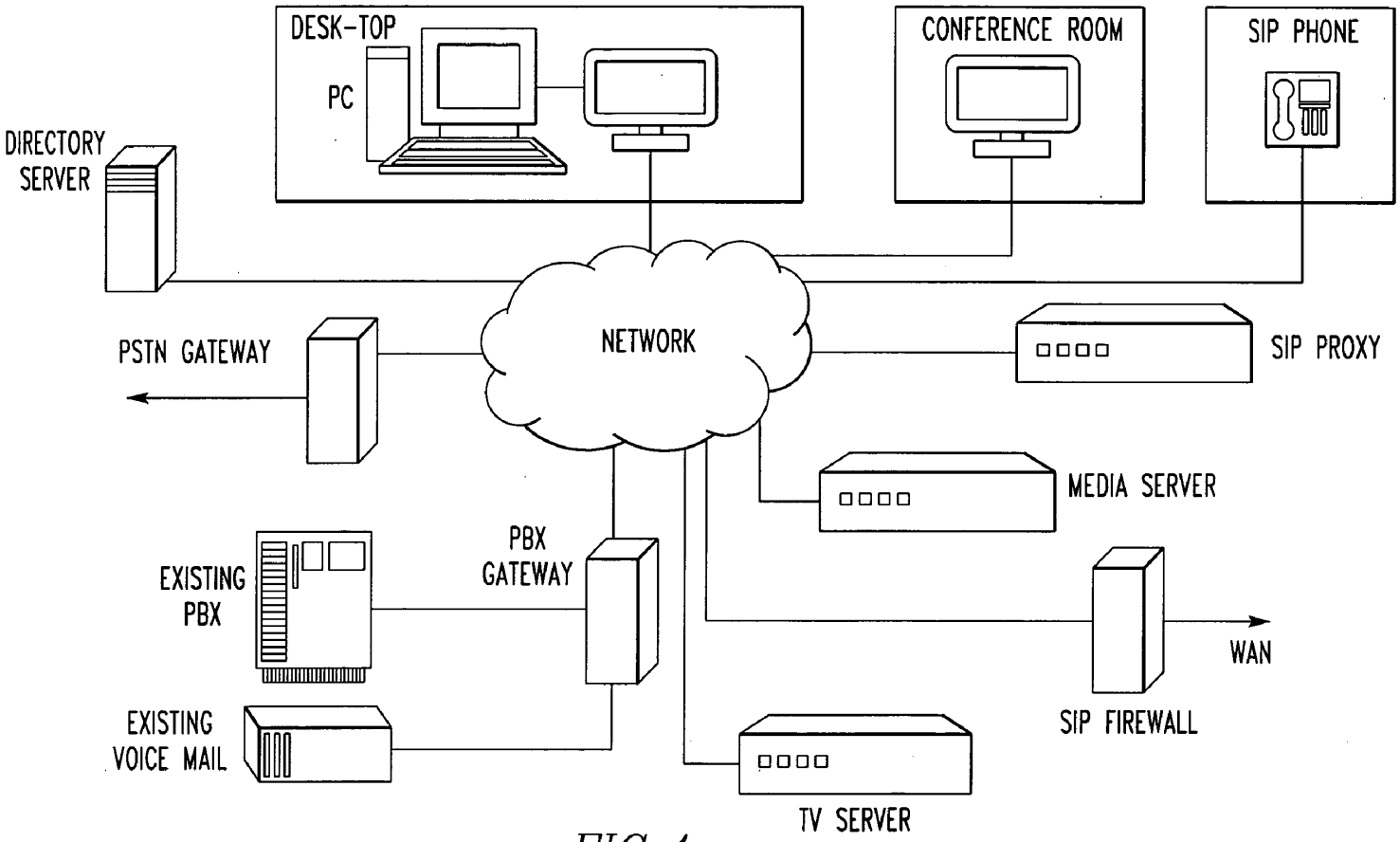

DESK-TOP
PC
CONFERENCE ROOM
SIP PHONE
DIRECTORY SERVER
NETWORK
PSTN GATEWAY
SIP PROXY
MEDIA SERVER
PBX GATEWAY
EXISTING PBX
WAN
SIP FIREWALL
EXISTING VOICE MAIL
TV SERVER

FIG. 4

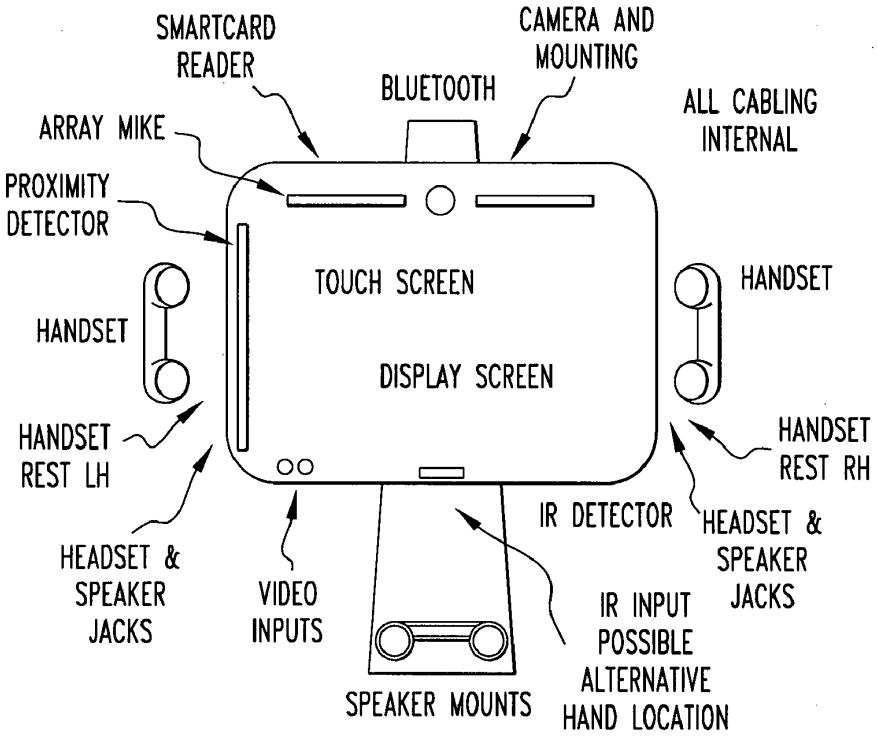

SMARTCARD READER
CAMERA AND MOUNTING
BLUETOOTH
ALL CABLING INTERNAL
ARRAY MIKE
PROXIMITY DETECTOR
TOUCH SCREEN
HANDSET
HANDSET
DISPLAY SCREEN
HANDSET REST LH
HANDSET REST RH
IR DETECTOR
HEADSET & SPEAKER JACKS
HEADSET & SPEAKER JACKS
VIDEO INPUTS
IR INPUT POSSIBLE ALTERNATIVE HAND LOCATION
SPEAKER MOUNTS

FIG.5a

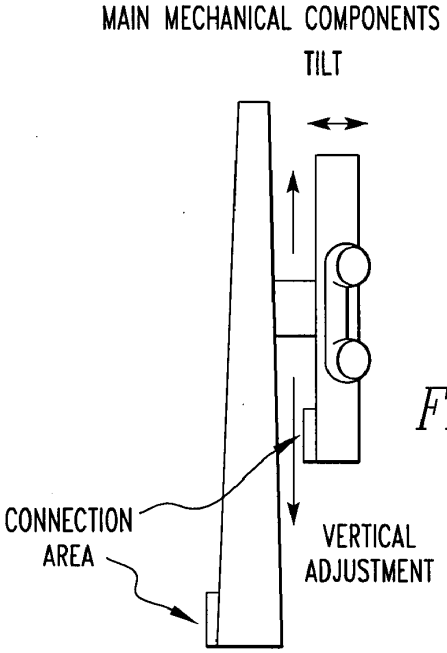

MAIN MECHANICAL COMPONENTS
TILT
CONNECTION AREA
VERTICAL ADJUSTMENT

FIG.5b

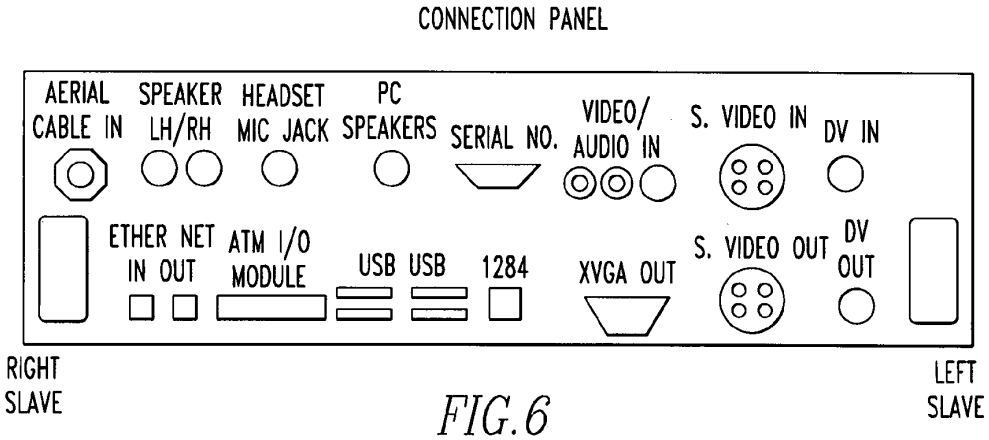

CONNECTION PANEL
AERIAL CABLE IN
SPEAKER LH/RH
HEADSET MIC JACK
PC SPEAKERS
SERIAL NO.
VIDEO/ AUDIO IN
S. VIDEO IN
DV IN
ETHER NET IN OUT
ATM I/O MODULE
USB USB
1284
XVGA OUT
S. VIDEO OUT
DV OUT
RIGHT SLAVE
LEFT SLAVE

FIG. 6

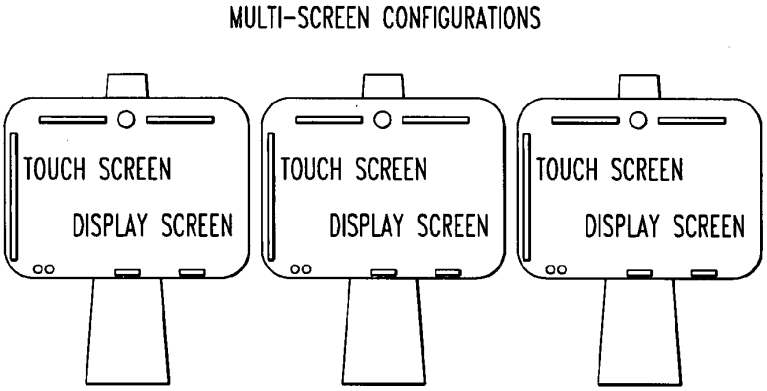

MULTI-SCREEN CONFIGURATIONS
TOUCH SCREEN
DISPLAY SCREEN
TOUCH SCREEN
DISPLAY SCREEN
TOUCH SCREEN
DISPLAY SCREEN

FIG. 7

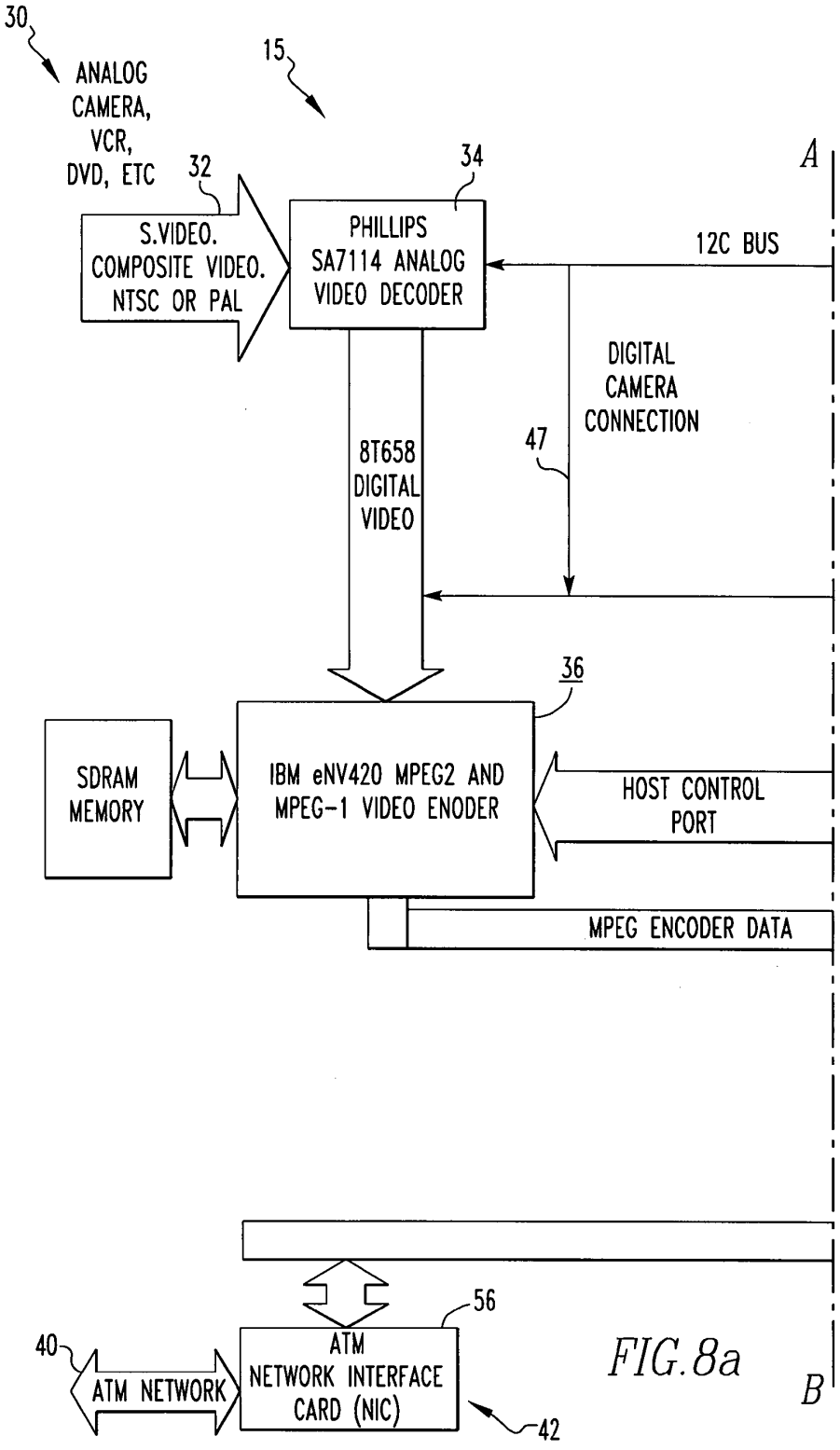

30
ANALOG CAMERA, VCR, DVD, ETC
15
32
S.VIDEO. COMPOSITE VIDEO. NTSC OR PAL
34
PHILLIPS SA7114 ANALOG VIDEO DECODER
A
12C BUS
DIGITAL CAMERA CONNECTION
47
8T658 DIGITAL VIDEO
36
SDRAM MEMORY
IBM eNV420 MPEG2 AND MPEG-1 VIDEO ENODER
HOST CONTROL PORT
MPEG ENCODER DATA
56
40
ATM NETWORK
ATM NETWORK INTERFACE CARD (NIC)
42
B

FIG.8a

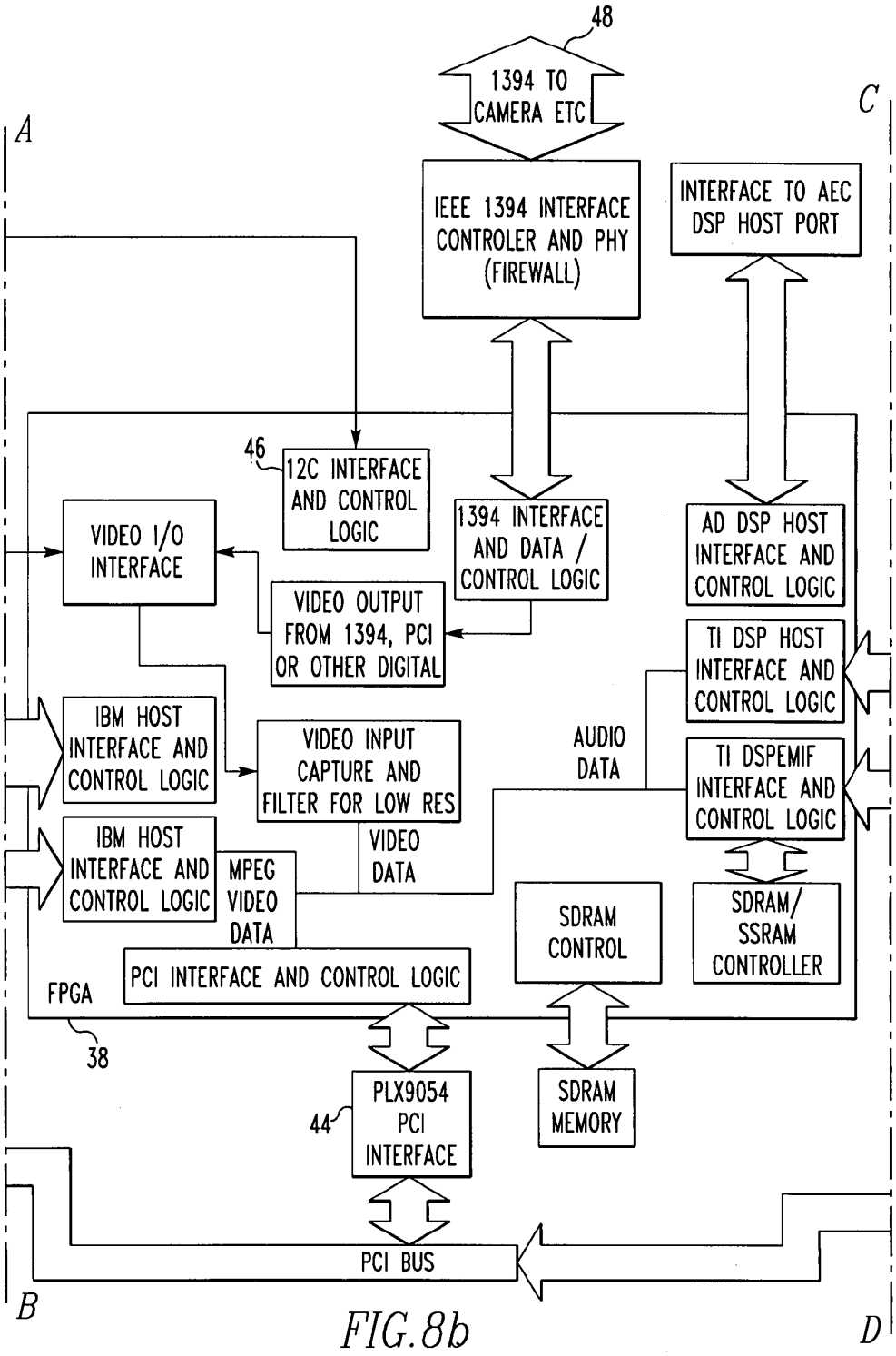

48
1394 TO CAMERA ETC
A
C
IEEE 1394 INTERFACE CONTROLER AND PHY (FIREWALL)
INTERFACE TO AEC DSP HOST PORT
46
12C INTERFACE AND CONTROL LOGIC
VIDEO I/O INTERFACE
1394 INTERFACE AND DATA / CONTROL LOGIC
AD DSP HOST INTERFACE AND CONTROL LOGIC
VIDEO OUTPUT FROM 1394, PCI OR OTHER DIGITAL
TI DSP HOST INTERFACE AND CONTROL LOGIC
IBM HOST INTERFACE AND CONTROL LOGIC
VIDEO INPUT CAPTURE AND FILTER FOR LOW RES
AUDIO DATA
TI DSPEMIF INTERFACE AND CONTROL LOGIC
IBM HOST INTERFACE AND CONTROL LOGIC
MPEG VIDEO DATA
VIDEO DATA
SDRAM CONTROL
SDRAM/ SSRAM CONTROLLER
PCI INTERFACE AND CONTROL LOGIC
FPGA
38
PLX9054 PCI INTERFACE
44
SDRAM MEMORY
PCI BUS
B
D

FIG.8b

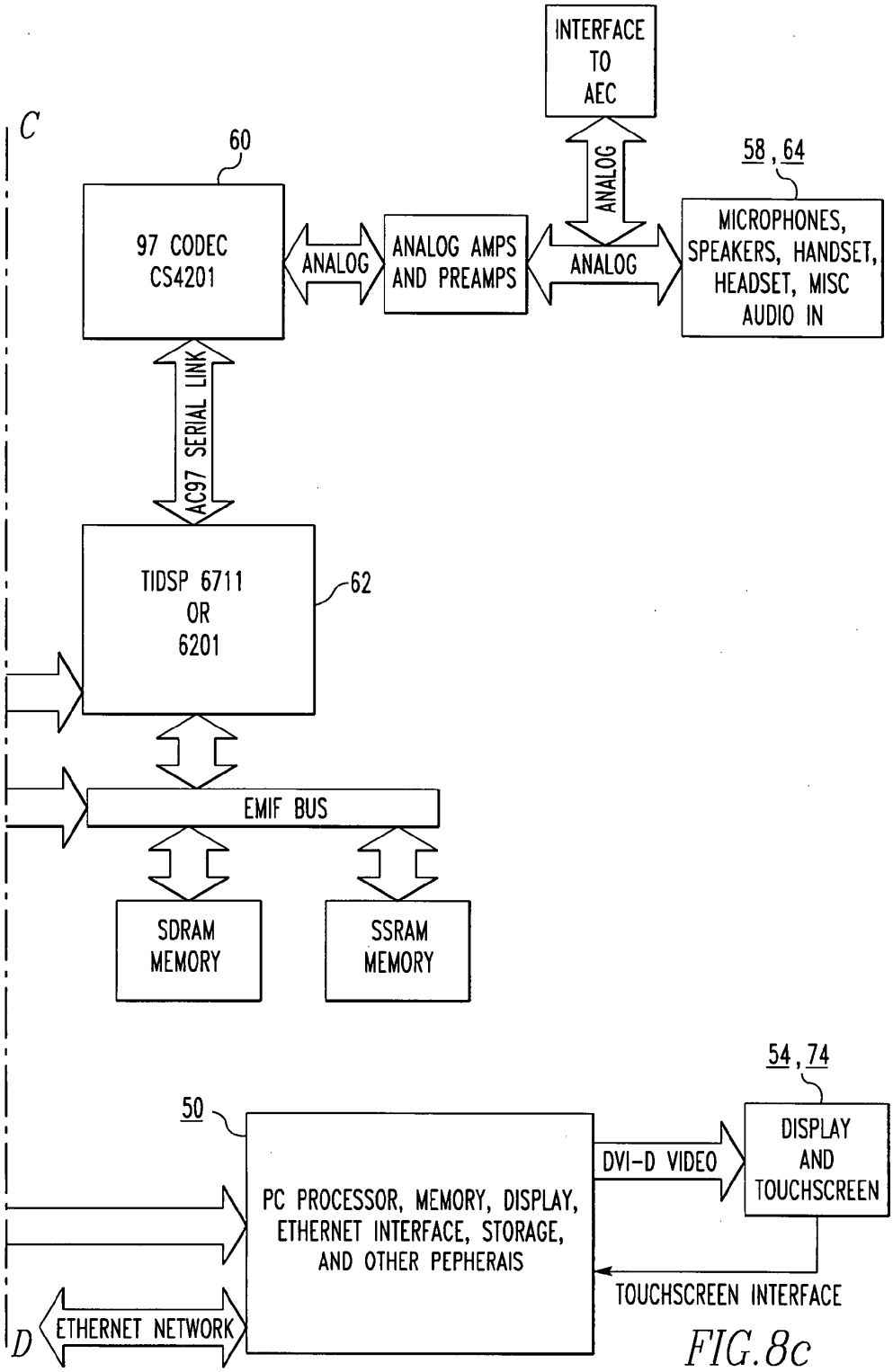

C
60
97 CODEC
CS4201
ANALOG
ANALOG AMPS
AND PREAMPS
ANALOG
INTERFACE
TO
AEC
ANALOG
58, 64
MICROPHONES,
SPEAKERS, HANDSET,
HEADSET, MISC
AUDIO IN
AC97 SERIAL LINK
TIDSP 6711
OR
6201
62
EMIF BUS
SDRAM
MEMORY
SSRAM
MEMORY
54, 74
50
PC PROCESSOR, MEMORY, DISPLAY,
ETHERNET INTERFACE, STORAGE,
AND OTHER PEPHERALS
DVI-D VIDEO
DISPLAY
AND
TOUCHSCREEN
TOUCHSCREEN INTERFACE
D
ETHERNET NETWORK

FIG.8c

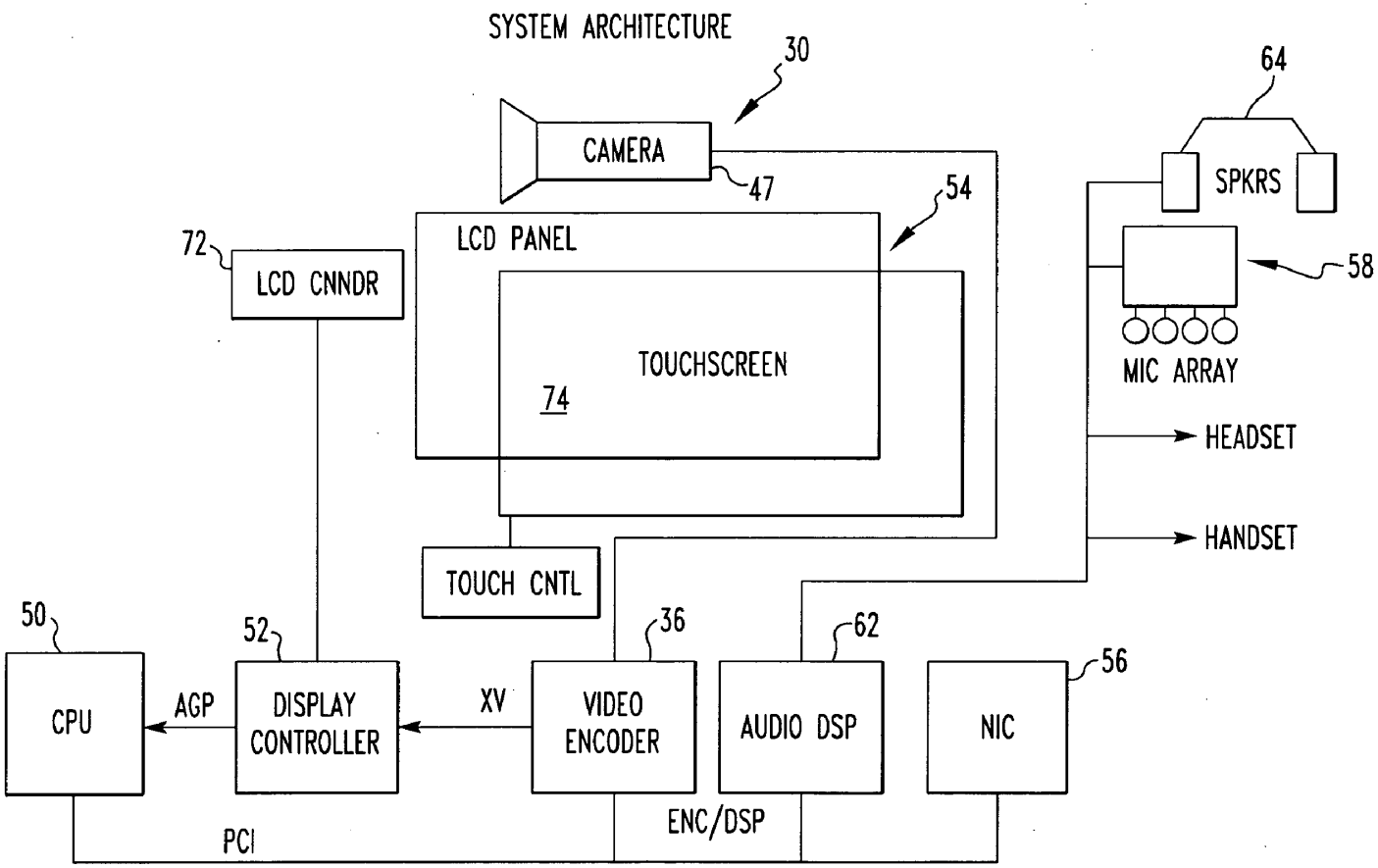

SYSTEM ARCHITECTURE
30
CAMERA
47
54
LCD PANEL
TOUCHSCREEN
74
72
LCD CNNDR
TOUCH CNTL
64
SPKRS
58
MIC ARRAY
HEADSET
HANDSET
50
CPU
52
DISPLAY CONTROLLER
AGP
XV
36
VIDEO ENCODER
62
AUDIO DSP
56
NIC
PCI
ENC/DSP

FIG. 9

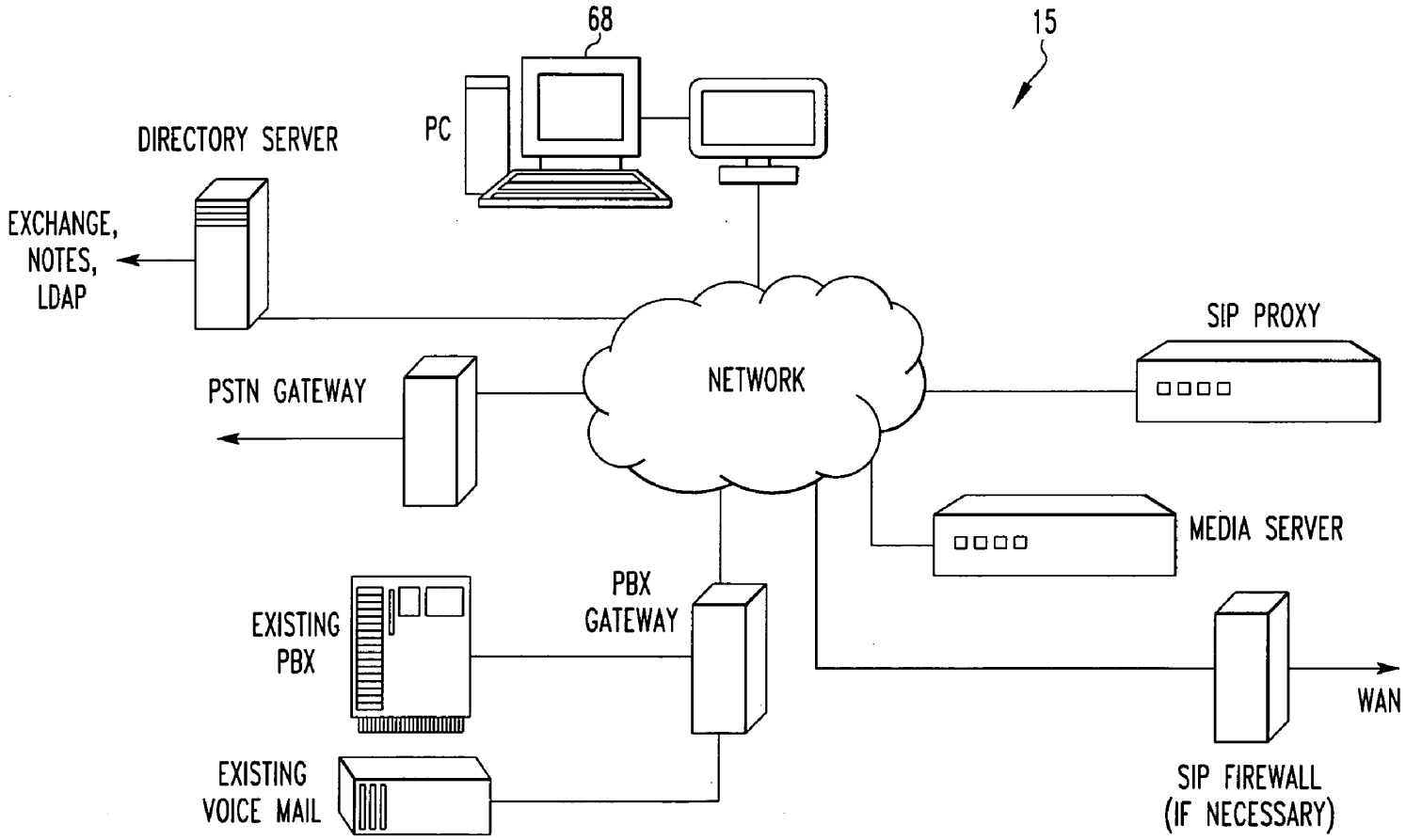

68
15
PC
DIRECTORY SERVER
EXCHANGE,
NOTES,
LDAP
NETWORK
SIP PROXY
PSTN GATEWAY
MEDIA SERVER
PBX
GATEWAY
EXISTING
PBX
WAN
EXISTING
VOICE MAIL
SIP FIREWALL
(IF NECESSARY)

FIG.10

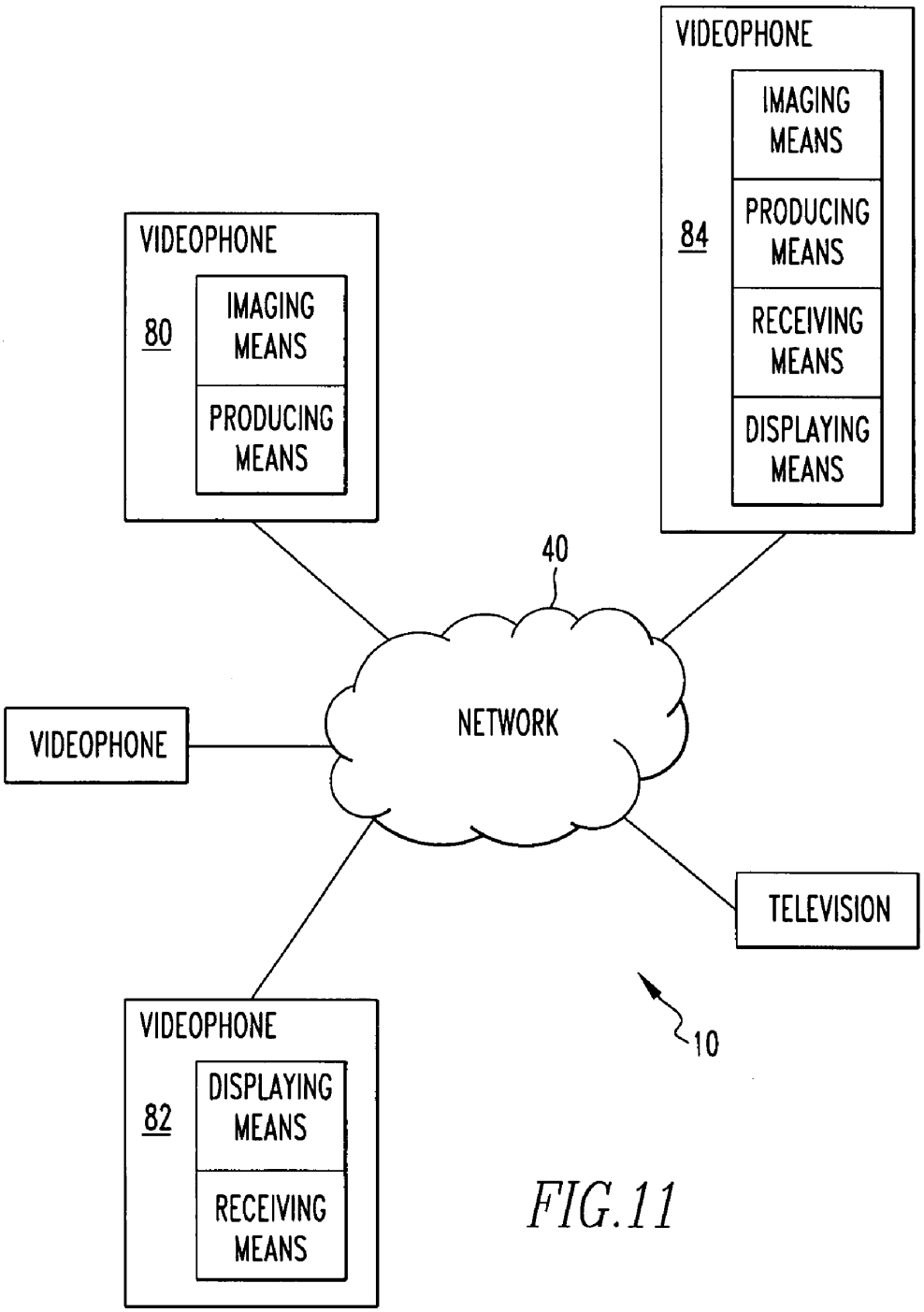

VIDEOPHONE
IMAGING MEANS
PRODUCING MEANS
RECEIVING MEANS
DISPLAYING MEANS
84
VIDEOPHONE
80
IMAGING MEANS
PRODUCING MEANS
40
NETWORK
VIDEOPHONE
TELEVISION
10
VIDEOPHONE
82
DISPLAYING MEANS
RECEIVING MEANS

FIG.11

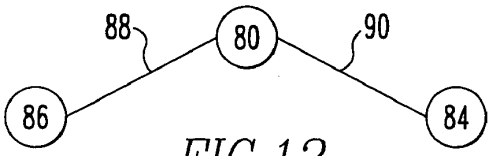


FIG.12

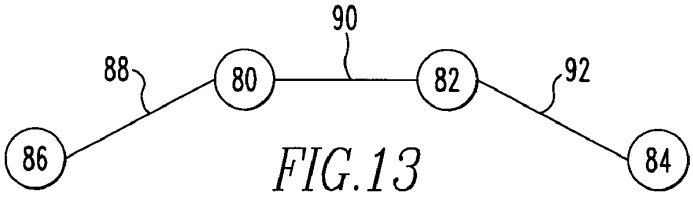


FIG.13

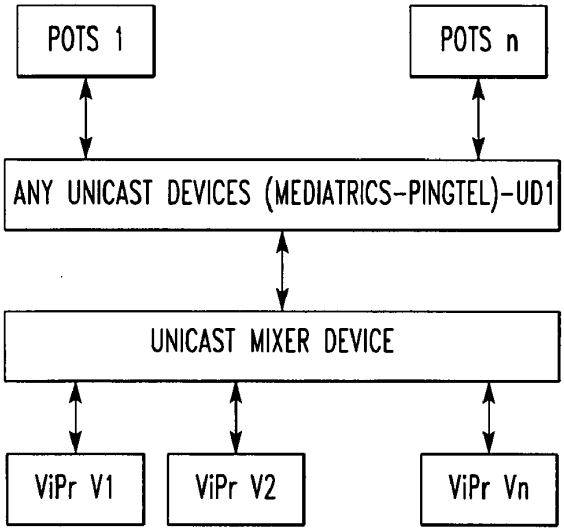


FIG.14

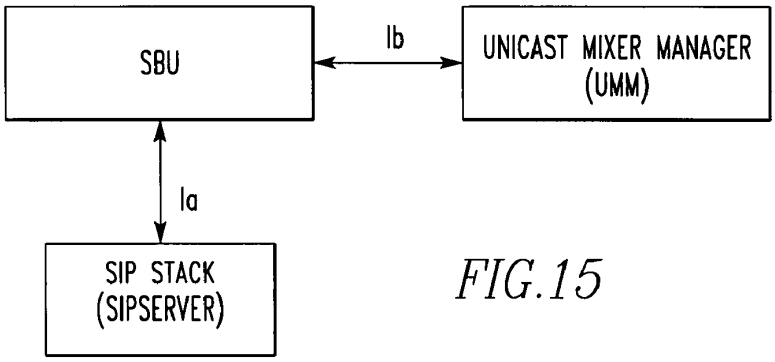


FIG.15

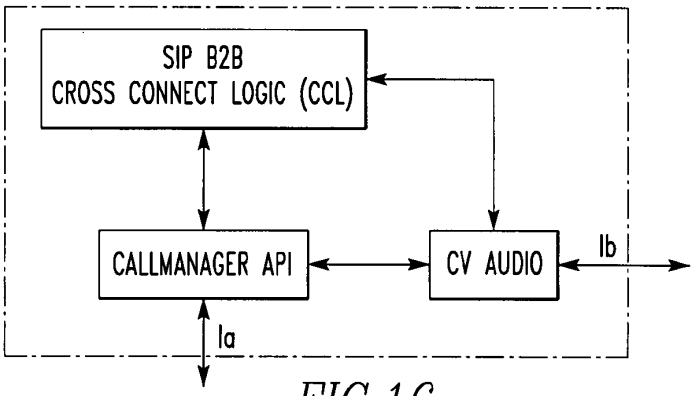

SIP B2B
CROSS CONNECT LOGIC (CCL)
CALLMANAGER API
CV AUDIO
Ib
Ia

FIG.16

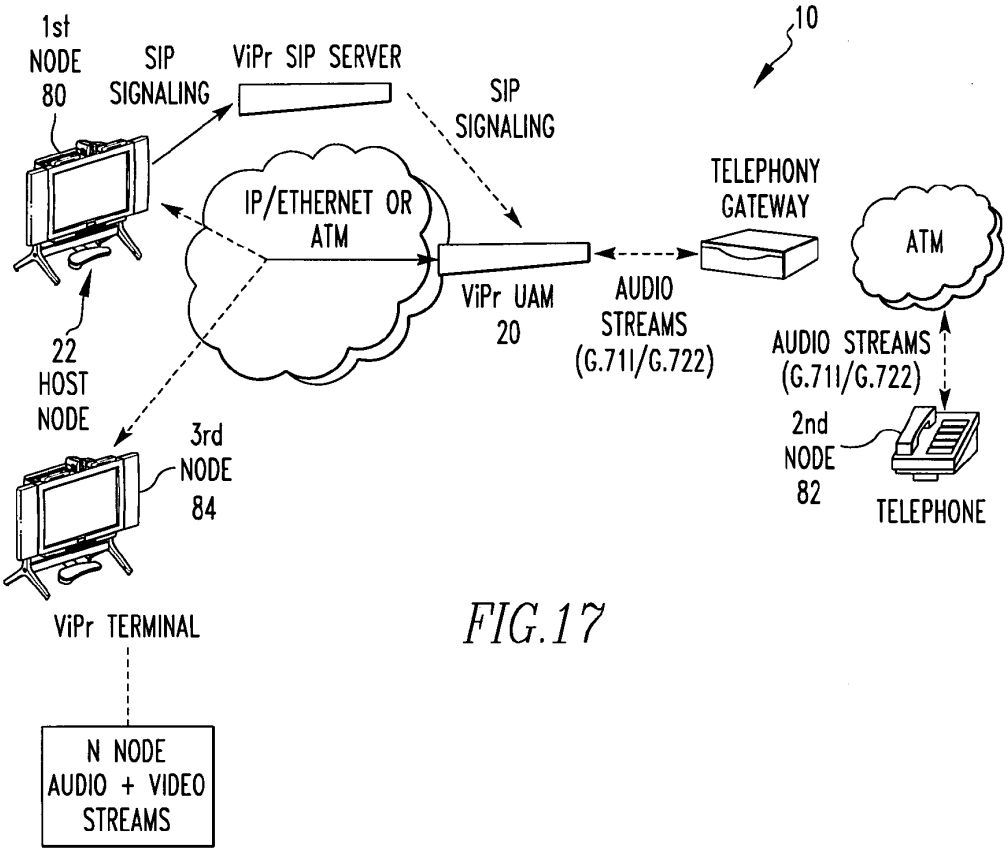

1st
NODE
80
SIP
SIGNALING
ViPr SIP SERVER
SIP
SIGNALING
10
IP/ETHERNET OR
ATM
TELEPHONY
GATEWAY
ATM
ViPr UAM
20
AUDIO
STREAMS
(G.711/G.722)
AUDIO STREAMS
(G.711/G.722)
22
HOST
NODE
3rd
NODE
84
2nd
NODE
82
TELEPHONE
ViPr TERMINAL
N NODE
AUDIO + VIDEO
STREAMS

FIG.17

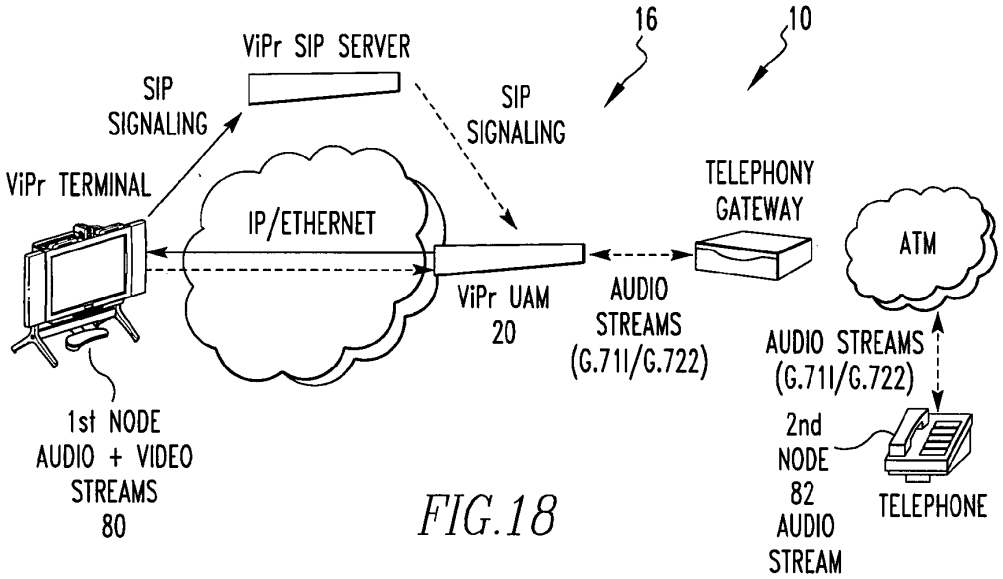

16
10
ViPr SIP SERVER
SIP SIGNALING
SIP SIGNALING
ViPr TERMINAL
IP/ETHERNET
TELEPHONY GATEWAY
ATM
ViPr UAM 20
AUDIO STREAMS (G.711/G.722)
AUDIO STREAMS (G.711/G.722)
1st NODE AUDIO + VIDEO STREAMS 80
2nd NODE 82 AUDIO STREAM
TELEPHONE

FIG.18

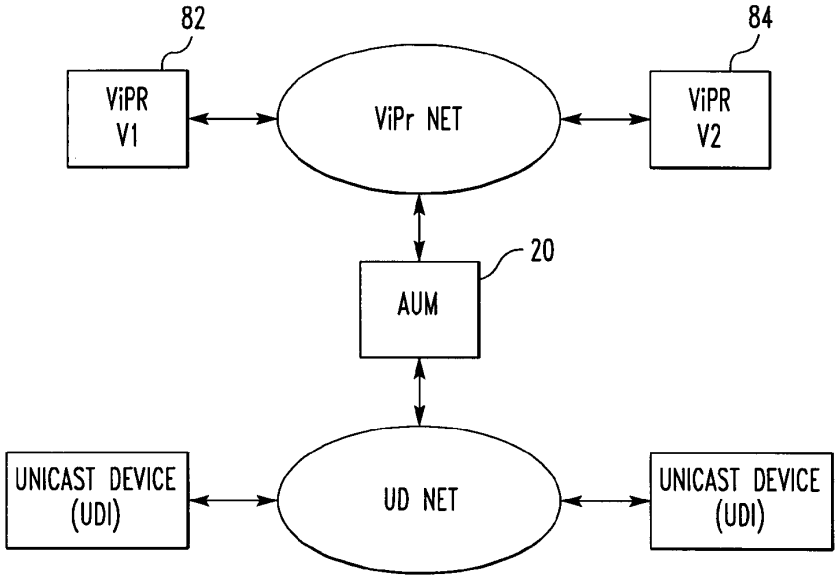

82
84
ViPR V1
ViPr NET
ViPR V2
20
AUM
UNICAST DEVICE (UDI)
UD NET
UNICAST DEVICE (UDI)

FIG.19

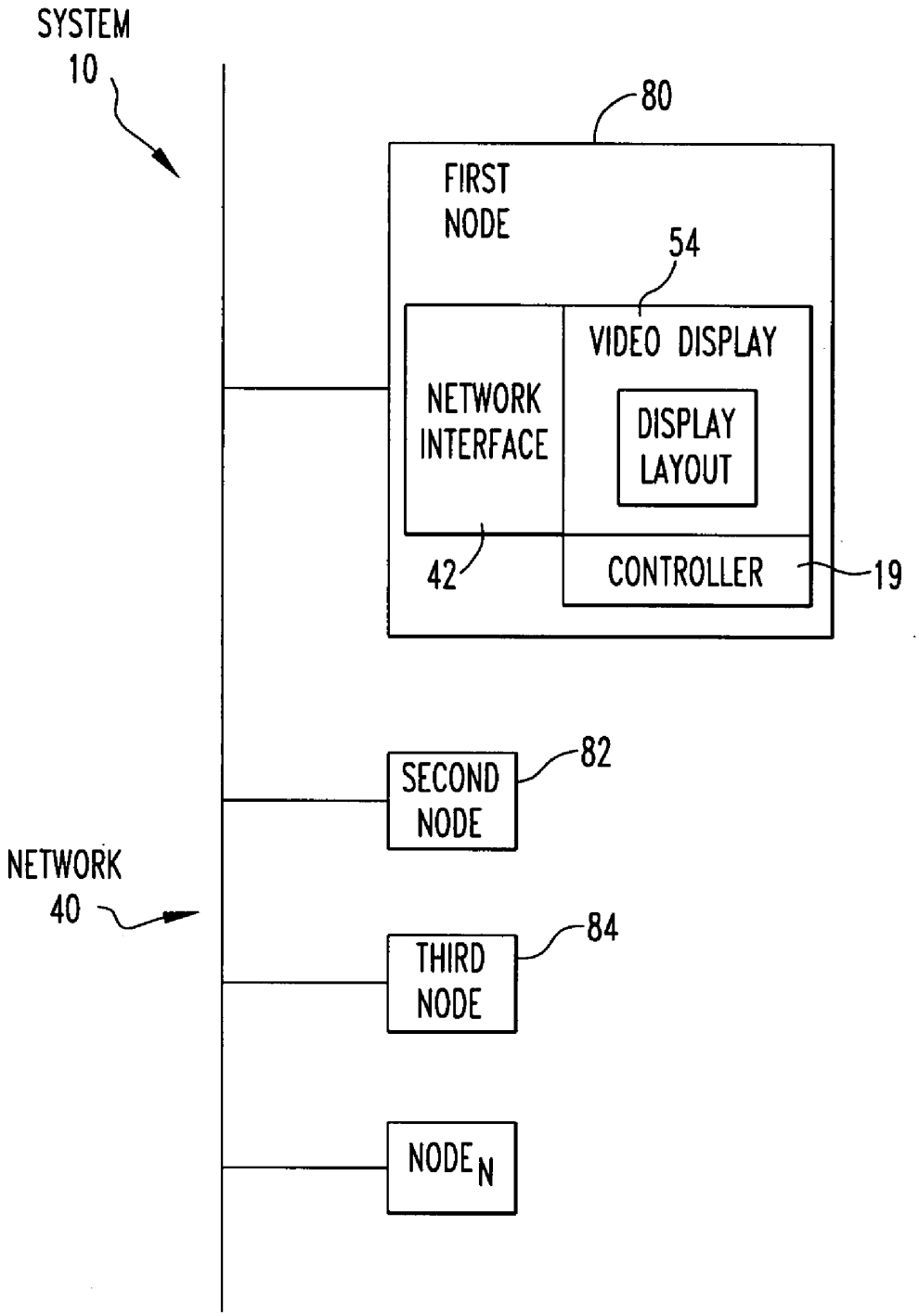

SYSTEM
10
80
FIRST
NODE
54
VIDEO DISPLAY
NETWORK
INTERFACE
DISPLAY
LAYOUT
42
CONTROLLER
19
SECOND
NODE
82
NETWORK
40
THIRD
NODE
84
NODE N

FIG.20

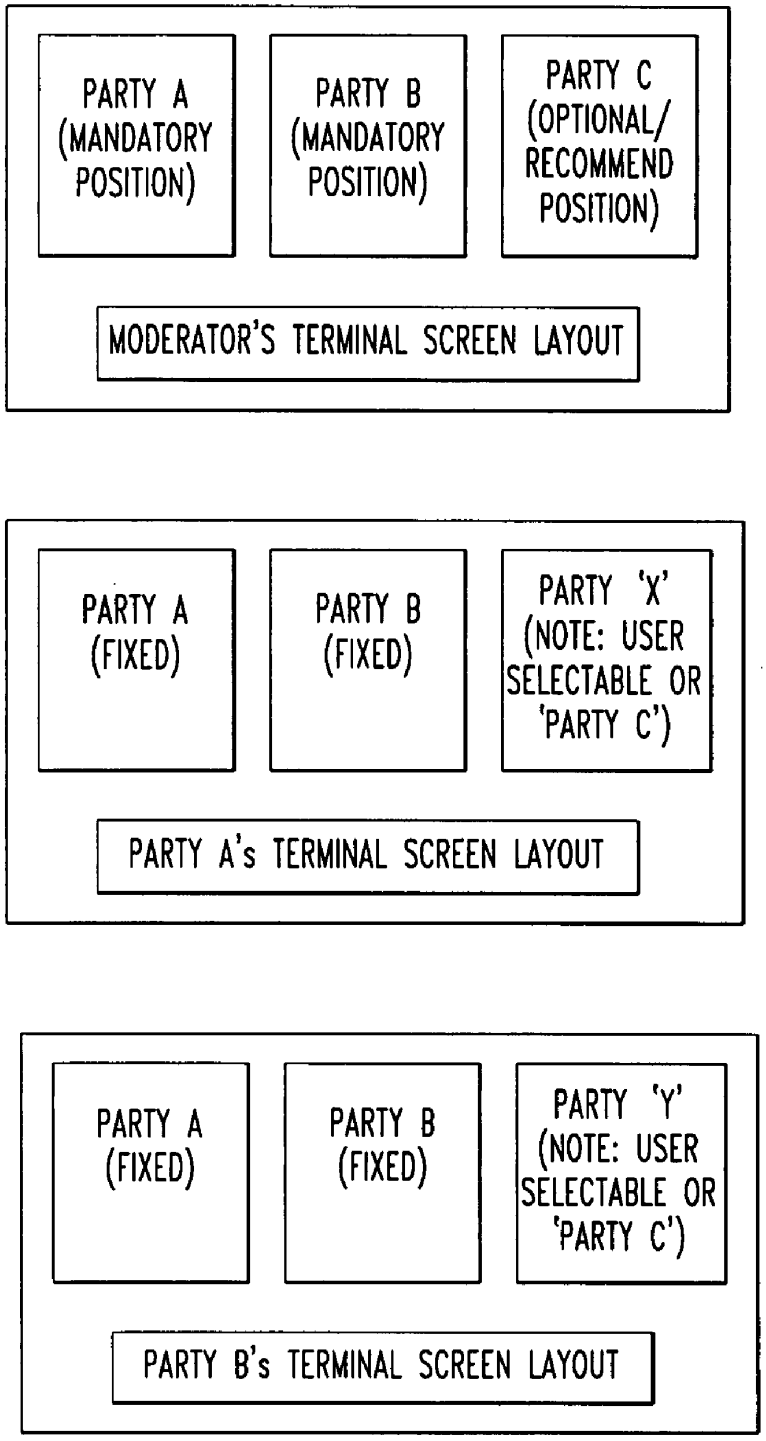

PARTY A (MANDATORY POSITION)
PARTY B (MANDATORY POSITION)
PARTY C (OPTIONAL/ RECOMMEND POSITION)
MODERATOR'S TERMINAL SCREEN LAYOUT
PARTY A (FIXED)
PARTY B (FIXED)
PARTY 'X' (NOTE: USER SELECTABLE OR 'PARTY C')
PARTY A's TERMINAL SCREEN LAYOUT
PARTY A (FIXED)
PARTY B (FIXED)
PARTY 'Y' (NOTE: USER SELECTABLE OR 'PARTY C')
PARTY B's TERMINAL SCREEN LAYOUT

FIG.21

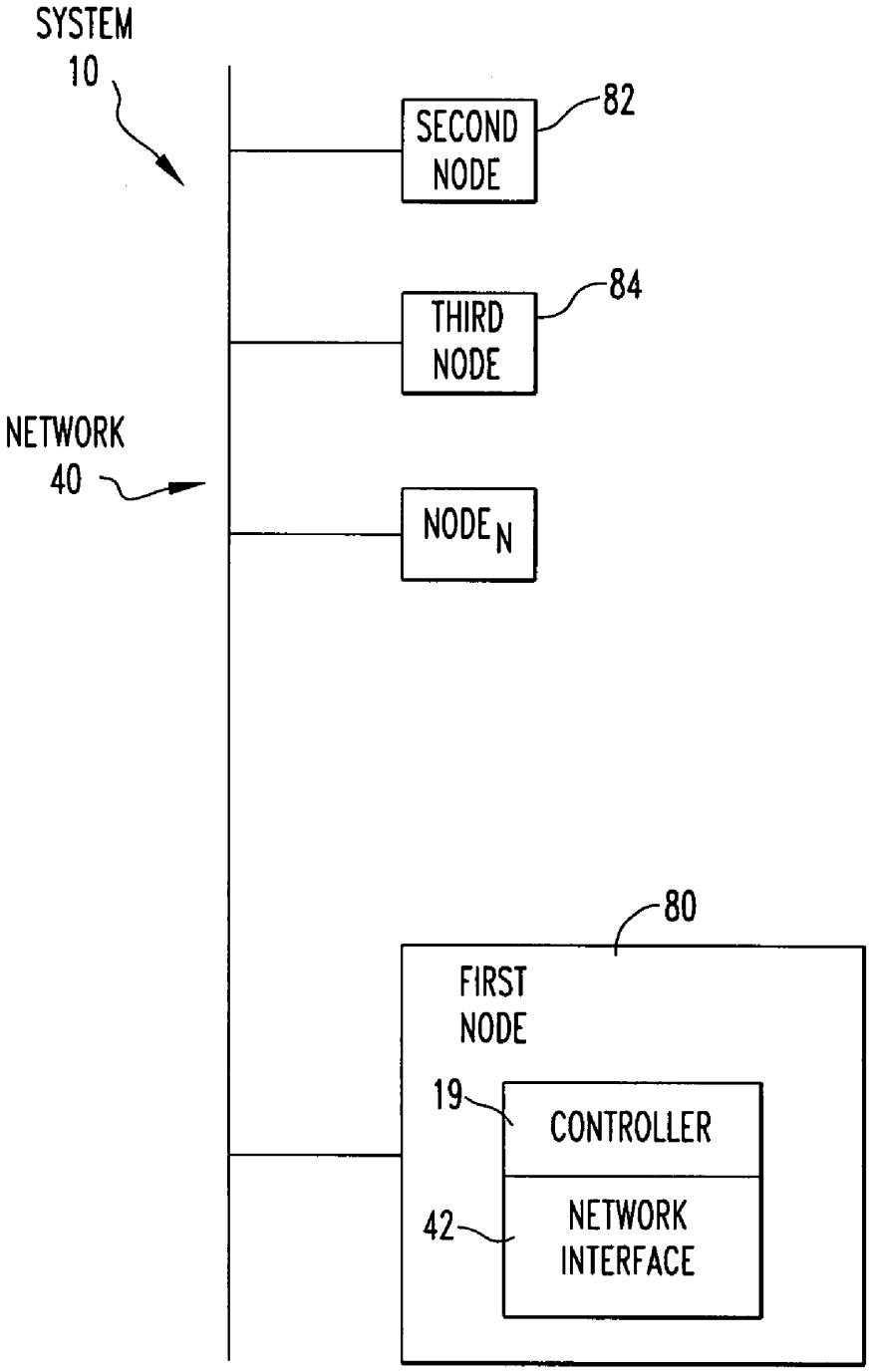

SYSTEM
10
SECOND NODE
82
THIRD NODE
84
NETWORK
40
NODE N
80
FIRST NODE
19
CONTROLLER
42
NETWORK INTERFACE

FIG.22

CONFERENCE LAYOUT CONTROL AND CONTROL PROTOCOL

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is related to contemporaneously filed U.S. provisional patent application: Ser. No. 60/814,477, titled "Intelligent Audio Limit Method", by Richard E. Huber, Arun Punj and Peter D. Hill, having attorney docket number FORE-119; Ser. No. 60/814,491, titled "Associating Independent Multimedia Sources Into a Conference Call", by Arun Punj, Richard E. Huber and Gregory H. Smith, having attorney docket number FORE-121, both of which are incorporated by reference herein.

FIELD OF THE INVENTION

[0002] The present invention is related to control of a video display of a teleconference. More specifically, the present invention is related to control of a video display of a teleconference where at least one of the nodes of the teleconference individually controls at least partially the display layout of each node in the conference with a specific format able to be unique for each node.

[0003] The present invention is related to a teleconference between nodes where each node communicates only a change to the conference to the other nodes when the change occurs. More specifically, the present invention is related to a teleconference between nodes where each node communicates only a change to the conference only to the nodes affected by the change when the change occurs.

BACKGROUND OF THE INVENTION

[0004] In regard to display layout, in a regular MCU based conference call, the MCU controls the layout of the video streams on each participant. In fact, the MCU sends the same image to all the participants. For example, in a conference call with **10** participants, the MCU would pick any 4 lets say B, C, D, E and form a composite image with B, C, D and E (probably as Hollywood squares) and send it to all the participants.

[0005] In ViPr, this model has been extended to where each participant could individually choose the layout independently. Thus, A could view 2 as large videos (say B and C) and view the other 7 as small videos. B could pick 1 large video, 3 small videos and a TV channel as its display.

[0006] In regard to protocol, consider a conference call with 10 participants. In the traditional signaling protocol, when there was a change in the state of the conference, for example, if P**1** disables its video, a message used to be sent with information for all the parties P**1** through P**10**; this causes serious scalability issues. The present invention provides a technique to control a very large conference call (with 100s of participants) in an efficient manner. The technique is whereby only the difference needs to be sent out, for instance, in the case mentioned above, a small NOTIFY event is sent with information that P**1** has turned off its transmitter.

BRIEF SUMMARY OF THE INVENTION

[0007] The present invention is related to a teleconferencing system. The system comprises a network. The system comprises a plurality of nodes in communication with each other to form a conference, preferably of live scenes at each node. Each node having a video display with a display layout, at least one of the nodes individually controlling at least partially the display layout of each node in the conference with a specific format able to be unique for each node.

[0008] The present invention is related to a method for providing a teleconference. The method comprises the steps of forming a conference with a plurality of nodes in communication with each other through a network, preferably of live scenes at each node. Each node having a video display with a display layout. There is the step of controlling individually with at least one of the nodes at least partially the display layout of each node in the conference with a specific format able to be unique for each node.

[0009] The present invention pertains to a teleconferencing node for a network with other nodes. The node comprises a network interface which communicates with the other nodes to form a conference, preferably of live scenes at each node. The node comprises a controller individually controlling at least partially the display layout of each node in the conference with a specific format able to be unique for each node.

[0010] The present invention pertains to a teleconferencing system. The system comprises a network. The system comprises a plurality of nodes in communication with each other through the network to form a conference, preferably of live scenes at each node. Each node communicating only a change to the other nodes when the change occurs.

[0011] The present invention pertains to a method for conducting a telecommunications conference between at least three nodes, for instance, parties. The method comprises the steps of establishing a conference between the nodes, preferably of live scenes at each node. There is the step of making a change to the conference. There is the step of communicating only the change to the nodes, preferably of live scenes at each node.

[0012] The present invention pertains to a teleconferencing node for a network with other nodes. The node comprises a network interface which communicates with the other nodes to form a conference, preferably of live scenes at each node. The node comprises a controller which communicates only a change to the other nodes when the change occurs.

[0013] The ability to efficiently control a large number of conference participants is very desirable. This can be especially true for low bandwidth links. Additionally, this is also a reduction of intermediate node processing since much smaller messages need to be exchanged.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWING

[0014] In the accompanying drawings, the preferred embodiment of the invention and preferred methods of practicing the invention are illustrated in which:

[0015] FIG. **1** is a schematic representation of a system for the present invention.

[0016] FIG. **2** is a schematic representation of a network for the present invention.

[0017] FIG. **3** is a schematic representation of a videophone connected to a PC and a network.

[0018] FIG. **4** is a schematic representation of the system for the present invention.

[0019] FIGS. **5***a* and **5***b* are schematic representations of front and side views of the videophone.

[0020] FIG. **6** is a schematic representation of a connection panel of the videophone.

[0021] FIG. **7** is a schematic representation of a multi-screen configuration for the videophone.

[0022] FIG. **8** is a block diagram of the videophone.

[0023] FIG. **9** is a block diagram of the videophone architecture.

[0024] FIG. **10** is a schematic representation of the system.

[0025] FIG. **11** is a schematic representation of the system.

[0026] FIG. **12** is a schematic representation of a system of the present invention.

[0027] FIG. **13** is a schematic representation of another system of the present invention.

[0028] FIG. **14** is a schematic representation of an audio mixer of the present invention.

[0029] FIG. **15** is a block diagram of the architecture for the mixer.

[0030] FIG. **16** is a block diagram of an SBU.

[0031] FIG. **17** is a schematic representation of a videophone UAM in a video phone conference.

[0032] FIG. **18** is a schematic representation of a videophone UAM in a two-way telephone call.

[0033] FIG. **19** is a schematic representation of a network for a mixer.

[0034] FIG. **20** is a block diagram of the present invention.

[0035] FIG. **21** is a block diagram of the present invention showing several nodes.

[0036] FIG. **22** is a block diagram of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

[0037] Referring now to the drawings wherein like reference numerals refer to similar or identical parts throughout the several views, and more specifically to FIGS. **20** and **21** thereof, there is shown a teleconferencing system **10**. The system **10** comprises a network **40**. The system **10** comprises a plurality of nodes in communication with each other to form a conference, preferably of live scenes at each node. Each node having a video display **54** with a display layout, at least one of the nodes individually controlling at least partially the display layout of each node in the conference with a specific format able to be unique for each node.

[0038] Preferably, each node is forced to display video in the specific format. Each node preferably is locked into the specific format. Preferably, each node is forced to display certain video streams from the other nodes of the conference at specific locations on the display. Each node preferably controls what is displayed on any portion of the screen not controlled by the one of the nodes. Preferably, the one of the nodes completely controls the display layout of each node.

[0039] The present invention is related to a method for providing a teleconference. The method comprises the steps of forming a conference with a plurality of nodes in communication with each other through a network **40**, preferably of live scenes at each node. Each node having a video display **54** with a display layout. There is the step of controlling individually with at least one of the nodes at least partially the display layout of each node in the conference with a specific format able to be unique for each node.

[0040] The present invention pertains to a teleconferencing node for a network **40** with other nodes. The node comprises a network interface **42** which communicates with the other nodes to form a live conference, preferably of live scenes at each node. The node comprises a controller **19** individually controlling at least partially the display layout of each node in the conference with a specific format able to be unique for each node.

[0041] In the operation of the invention, this invention provides a technique to control layout of an individual conference participant screen from one of the conference participants. For example, if there are participants P**1** through P**10** in a conference call, one of the participants could become moderator and force P**2** to view (P**1** and P**5**) in large video and rest in small video the respective live scenes of each participant. Each party's display could be individually controlled in this manner.

[0042] This layout control can be enforced on an individual window rather than the entire screen to offer individualized control of the non-managed windows on the screen.

[0043] A remote party can control the screen layout of each conference participant. Typically, a moderator would force all parties to use the same layout as the moderator. However, there could be cases where fine grain control could grant a sub-moderator control over a subsection of the conference participants.

[0044] The layout control mechanism generates a layout message that contains the desired screen layout for conference participants. The layout message also contains the list of participants, which should receive this message. This layout message is then sent via a SIP NOTIFY event to the conference focus or host. The conference focus will then add this message to the outgoing message queue of each party contained in this list. The focus will then send this message as it processes all of the queued events for each party. When the message is sent to and received by a particular party, the party will modify its screen layout to match the request contained in the message. If the screen layout change requires that the party connect or disconnect to a new media stream, then the party will issue the appropriate events to make the requested changes.

[0045] In the present invention, functionality has been added which allows a user or a set of users (moderator/moderators) to individually control the display layout on each ViPr video phone in the conference. This control could be partial or complete. In complete control display layout format, each participant in a conference is forced to display video in a specific format. It is somewhat like an MCU, however it is different in regard that each participant could be locked into a different format. But once a participant is locked into a given format it has no control over display layout of the conference call. For example, A could be locked to display 3 big videos of live scenes from C, D and E and 6 small videos from F, G, H, I, J, K.

[0046] In partial control display format, each participant is instructed to display certain streams at specific locations. But it has control over what to display on the rest of screen. For example, A could be instructed to display B in its left big video. However, it could choose whether to display 1, 2 or 3 big videos. Similarly, scheme would be available for audio/small video.

[0047] The Layout Control messages are sent as SIP/NOTIFY messages although they 'could' be sent via other SIP or HTTP means.

[0048] In a 'typical' ViPr conference call, each terminal is provided with all of the available audio and video streams from each other participant in the call. Each user at each terminal usually would let the local terminal automatically

place each video in sequential order on the screen. Then the user CAN manually choose to select which of the parties video get shown as large video or small video windows on the screen.

[0049] If a call participant wants to moderate the call then they can use 'Layout Control' feature to place restrictions on some or all of the other terminals. These restrictions can include which participants to display as large video windows. They can also control which of the small video windows are locked to a particular participant. The restrictions can be specified as mandatory or optional which in turn specifies whether a participant can override the moderators choice or not. The layout control can also be used to control the placement of secondary images on the screen. The layout control can also control the size of any secondary images on the screen. The layout control can also control the audio mute of remote parties and the ability for remote parties to request the floor when they wish to be recognized by the floor and unmuted.

[0050] A SIP/NOTIFY message containing the layout control options is sent to the conference host and then the conference host distributes this message to all of the other parties.

[0051] The call moderator uses the standard SIP/SDP "a=Rx-List: A B C" mechanism to specify which parties the call moderator is displaying, where 'A B C' represent the Party-Identifiers of the remote parties which are being viewed. When operating in call moderated mode these are all treated as 'optional' layout positions unless an 'm' is appended to the party, such as 'Am Bm Cm'. The 'm' is the mandatory flag and tells the remote User Interface which parties 'must' be locked into position on the screen. The optional parties can still be changed independently by each remote terminal if the user interface wishes to enable this. The user interface CAN enforce an 'all mandatory' mode on the conference, in which case all parties are treated as mandatory when operating in a moderated call.

[0052] The event named "moderator-layout" is used to identify a layout control message for controlling other aspects of the screen. The option string keywords used to control the size of secondary videos and images are "chan_size" and "col_size" respectively. The events named "floor-request" and "floor-withdrawn" are used to tell the moderator that a party wants the floor or wishes to withdraw their request. The event named "floor-granted" is sent by the moderator to a party to inform them that they have been unmuted and can now speak. The terminal User Interface will respect each of these events and control the screen as directed by the call moderator.

[0053] The following applications are all incorporated by reference herein:

[0054] U.S. patent application Ser. No. 10/114,402 titled VIDEOPHONE AND METHOD FOR A VIDEO CALL

[0055] U.S. patent application Ser. No. 10/871,852 titled AUDIO MIXER AND METHOD

[0056] U.S. patent application Ser. No. 11/078,193 titled METHOD AND APPARATUS FOR CONFERENCING WITH STREAM

[0057] A node can include a member, party, terminal, or participant of a conference. A conference typically comprises at least 3 nodes, and could have 10 or 20 or even 50 or 100 or 150 or greater nodes.

[0058] Referring now to the drawings wherein like reference numerals refer to similar or identical parts throughout the several views, and more specifically to FIG. **22** thereof, there is shown a teleconferencing system **10**. The system **10** comprises a network **40**. The system **10** comprises a plurality of nodes in communication with each other through the network **40** to form a conference, preferably of live scenes at each node. Each node communicating only a change to the other nodes when the change occurs.

[0059] Preferably, each node communicates only the change only to the parties affected by the change.

[0060] The present invention pertains to a method for conducting a telecommunications conference between at least three nodes, for instance, parties. The method comprises the steps of establishing a live conference between the nodes conference, preferably of live scenes at each node. There is the step of making a change to the conference. There is the step of communicating only the change to the nodes.

[0061] Preferably, the communicating step includes the step of communicating only the change only to the nodes affected by the change. The making step preferably includes the step of making the change to one of the nodes' status. Preferably, the making step includes the step of making the change to the conference's state. There is preferably the step of sending a directed message from one of the nodes to only certain of but less than all of the nodes. Preferably, the establishing step includes the step of establishing the conference based on SIP NOTIFY/OK techniques.

[0062] The change to the conference preferably includes a change to one of the nodes' status. Preferably, the change to the conference includes a change to the conference's state. One of the nodes sends a directed message to only certain of but less than all of the nodes. Preferably, the plurality of nodes establishes the conference based on SIP NOTIFY/OK techniques. Each node preferably has a controller **19** and a Network Interface **42** in communication with the controller **19** and the network **40**. The controller **19** effecting any change to its node and sending the change through the network interface **42** to through the network **40** to the other nodes of the conference.

[0063] The present invention pertains to a teleconferencing node for a network **40** with other nodes. The node comprises a network interface **42** which communicates with the other nodes to form a live conference, preferably of live scenes at each node. The node comprises a controller **19** which communicates only a change to the other nodes when the change occurs.

[0064] In the operation of the preferred embodiment, the large conference control mechanism uses a conference control message to manage all of the parties on the call. A participant generates a conference control message that contains both the transmitted streams properties and the desired lists of receive streams from other conference participants. This conference control message is then sent via a SIP NOTIFY event to the conference focus or host. The conference focus will then add this message to the outgoing message queue of each party affected by this message. The focus will then send these messages as it processes all of the queued events for each party. When the message is sent to and received by a particular party, the party will add the requesting party to its list of outgoing streams. A conference

control message can also be sent to indicate the desire to place video streams on or off "hold" for control of voice-only operation.

Large Conference Signaling Invention

[0065] ViPr conference previously was based on an offer/answer model. In this model, the complete state of the conference was carried in each message exchanged between conference participants. For example, consider a conference between 5 participants P**1** through P**5**. In this case, these five parties would be connected into a conference via a central point called host. The Host would form a table which contains the complete state of the conference. Like If parties P**1** through P**3** are transmitting/Receiving Video/Audio and parties P**2** and P**3** are only transmitting/receiving audio only this table would indicate that complete state. When ever any changes happened in the conference the Host would recomputed the table and send that information to everyone. For example, if P**3** stopped transmitting video, it would change the table 1 to table 2 and send that complete table to everyone.

TABLE 1

P1	Video (tx = On, rx = On), Audio (tx = on, rx = on)
P2	Video (tx = On, rx = On), Audio (tx = on, rx = on)
P3	Video (tx = On, rx = On), Audio (tx = on, rx = on)
P4	Video (tx = Off, rx = on), Audio (tx = on, rx = on)
P5	Video (tx = off, rx = on), Audio (tx = on, rx = on)

TABLE 2

P1	Video (tx = On, rx = On), Audio (tx = on, rx = on) Video (tx = On, rx = On), Audio (tx = on, rx = on) Video (tx = On, rx = Off), Audio (tx = on, rx = on)
P4	Video (tx = Off, rx = on), Audio (tx = on, rx = on) Video (tx = off, rx = on), Audio (tx = on, rx = on)

[0066] This scheme worked well because it was consistent with the existing SIP standards and allowed us to add extensions to the basic SIP protocol to enable ViPr style conferencing.

[0067] This scheme works well if the number of participants in the conference is less than 15 or so, beyond that the table, which contains all the parties, becomes too big to be passed around efficiently. Additionally, not every party is affected by every change in the state of the conference—and flooding it with messages which it does not need to handle is unnecessary. Another problem with this model is that it does not allow for the ability to send application level non media related messages between two SIP peers.

[0068] To remedy the above issues we invented this new scheme. In this new scheme following critical changes have been made:

[0069] The state of the conference can change in two respects. Either one of the participants had requested a change in its status (party local) or one of party has requested a change in the state of the whole conference (global change). When ever, any such change is made only the information specific to the party being changed or the global conference state is communicated between conference participants. To use example from table 2, only the information corresponding to P**3** would be sent out, rather than the complete table. [Thus only row referring to p**3** is re-sent to other participants].

[0070] Examples of Global Events:

[0071] → Conference Name Change

[0072] → Conference Moderator State Change

[0073] → Floor Request Status Change

[0074] → Conference Type Change

[0075] → Conference Status Messages (A party being rejected from joining the conference etc)

[0076] Examples of Party Events

[0077] → Party Toggling camera

[0078] → Party Toggling Hold status

[0079] → Party Delete

[0080] → Party Add

[0081] → Party Status Change (becomes moderator/relinquish moderator)

[0082] → Party requesting a change in Receive Media Stream

[0083] Only the parties, which are affected by the change, receive the changed information.

[0084] It is possible to send a directed message from one party P**1** to any number of parties. For example, P**1** could request the host to relay a message to P**2**, p**3** and p**4**. But not to P**5**. This last functionality is important to enable group based signaling in conferencing.

[0085] The new design is based on SIP NOTIFY/OK methods and defines a new event package. Standards RFC 3261 and RFC 3264 define the base specification for SIP. RFC 3265 defines a framework for using events, both of which are incorporated by reference herein.

As an example, a host is always required for Vipr conference. In fact, parties in a conference are not ever directly signaling with each other. For example, if there is a conference call between A, B and C. There are three SIP Call Legs

[0086] Host to A

[0087] Host to B

[0088] Host to C

Media flows directly between A, B and C.

[0089] In the present invention, B is currently receiving video from A and now wishes to also receive video from C. This change could be communicated in either of two ways:

[0090] B Sends a NOTIFY to Host, with a field in the NOTIFY [called dest-party-list:C] indicating that this message may only be sent to C.

[0091] Alternatively, Host can detect that this message only affects C and may send that message only to C although B has not been explicitly put in any such field.

Videophone

[0092] Referring to FIGS. **8**, **9**, **10** and **11**, an imaging device **30**, such as a conventional analog camera **32** provided by Sony with S video, converts the images of a scene from the imaging device **30** to electrical signals which are sent along a wire to a video decoder **34**, such as a Philips SAA7114 NTSC/PAL/decoder. The video decoder **34** con-

verts the electrical signals to digital signals and sends them out as a stream of pixels of the scene, such as under BT **656** format. The stream of pixels are sent out from the video decoder **34** and split into a first stream and a second stream identical with the first stream. An encoder **36**, preferably an IBM eNV 420 encoder, receives the first stream of pixels, operates on the first stream and produces a data stream in MPEG-2 format. The data stream produced by the video encoder **36** is compressed by about 1/50 the size as compared to the data as it was produced at the camera. The MPEG-2 stream is an encoded digital stream and is not subject to frame buffering before it is subsequently packetized so as to minimize any delay. The encoded MPEG-2 digital stream is packetized using RTP by a Field Programmable Gate Array (FPGA) **38** and software to which the MPEG-2 stream is provided, and transmitted onto a network **40**, such as an Ethernet 802.p or ATM at 155 megabits per second, using a network interface **42** through a PLX 9054 PCI interface **44**. If desired, a video stream associated with a VCR or a television show, such as CNN or a movie, can be received by the decoder **34** and provided directly to the display controller **52** for display. A decoder controller **46** located in the FPGA **38** and connected to the decoder **34**, controls the operation of the decoder **34**.

[0093] Alternatively, if a digital camera **47** is used, the resulting stream that is produced by the camera is already in a digital format and does not need to be provided to a decoder **34**. The digital stream from the digital camera **47**, which is in a BT 656 format, is split into the first and second streams directly from the camera, without passing through any video decoder **34**.

[0094] In another alternative, a fire wire camera **48**, such as a 1394 interface fire wire camera **48**, can be used to provide a digital signal directly to the FPGA **38**. The fire wire camera **48** provides the advantage that if the production of the data stream is to be at any more than a very short distance from the FPGA **38**, then the digital signals can be supported over this longer distance by, for instance, cabling, from the fire wire camera **48**. The FPGA **38** provides the digital signal from the fire wire camera **48** to the encoder **36** for processing as described above, and also creates a low fame rate stream, as described below.

[0095] The second stream is provided to the FPGA **38** where the FPGA **38** and software produce a low frame rate stream, such as a motion JPEG stream, which requires low bandwidth as compared to the first stream. The FPGA **38** and a main controller **50** with software perform encoding, compression and packetization on this low frame rate stream and provide it to the PCI interface **44**, which in turn transfers it to the network interface **42** through a network interface card **56** for transmission onto the network **40**. The encoded MPEG-2 digital stream and the low frame rate stream are two essentially identical but independent data streams, except the low frame rate data stream is scaled down compared to the MPEG-2 data stream to provide a smaller view of the same scene relative to the MPEG-2 stream and require less resources of the network **40**.

[0096] On the network **40**, each digital stream is carried to a desired receiver videophone **15**, or receiver videophones **15** if a conference of more than two parties is involved. The data is routed using SIP. The network interface card **56** of the receive videophone **15** receives the packets associated with first and second data streams and provides the data from the packets and the video stream (first or second) chosen by the main controller to a receive memory. A main controller **50** of the receive videophone **15** with software decodes and expands the chosen received data stream and transfers it to a display controller **52**. The display controller **52** displays the recreated images on a VGA digital flat panel display using standard scaling hardware. The user at the receive videophone **15** can choose which stream of the two data streams to view with a touch screen **74**, or if desired, chooses both so both large and small images of the scene are displayed, although the display of both streams from the transmitting videophone **15** would normally not happen. A discussion of the protocols for display is discussed below. By having the option to choose either the larger view of the scene or the smaller view of the scene, the user has the ability to allocate the resources of the system **10** so the individuals at the moment who are more important for the viewer to see in a larger, clearer picture, can be chosen; while those which the user still would like to see, but are not as important at that moment, can still be seen.

[0097] The display controller **52** causes each distinct video stream, if there is more than one (if a conference call is occurring) to appear side by side on the display **54**. The images that are formed side by side on the display **54** are clipped and not scaled down so the dimensions themselves of the objects in the scene are not changed, just the outer ranges on each side of the scene associated with each data stream are removed. If desired, the images from streams associated with smaller images of scenes can be displayed side by side in the lower right corner of the display **54** screen. The display controller **52** provides standard digital video to the LCD controller **72**, as shown in FIG. **9**. The display controller **52** produced by ATI or Nvidia, is a standard VGA controller. The LCD controller **72** takes the standardized digital video from the display controller **52** and makes the image proper for the particular panel used, such as a Philips for Fujistu panel.

[0098] To further enhance the clipping of the image, instead of simply removing portions of the image starting from the outside edge and moving toward the center, the portion of the image which shows no relevant information is clipped. If the person who is talking appears in the left or right side of the image, then it is desired to clip from the left side in if the person is on the right side of the image, or right side in if the person is on the left side of the image, instead of just clipping from each outside edge in, which could cause a portion of the person to be lost. The use of video tracking looks at the image that is formed and analyzes where changes are occurring in the image to identify where a person is in the image. It is assumed that the person will be moving more relative to the other areas of the image, and by identifying the relative movement, the location of the person in the image can be determined. From this video tracking, the clipping can be caused to occur at the edge or edges where there is the least amount of change. Alternatively, or in combination with video tracking, audio tracking can also be used to guide the clipping of the image which occurs. Since the videophone **15** has microphone arrays, standard triangulation techniques based on the different times it takes for a given sound to reach the different elements of the microphone array are used to determine where the person is located relative to the microphone array, and since the location of a microphone array is known relative to the scene that is being imaged, the location of the person in the image is thus known.

[0099] The functionalities of the videophone **15** are controlled with a touch screen **74** on the monitor. The touch screen **74**, which is a standard glass touchscreen, provides raw signals to the touch screen controller **76**. The raw signals are sensed by the ultrasonic waves that are created on the glass when the user touches the glass at a given location, as is well known in the art. The touch screen controller **76** then takes the raw signals and converts them into meaningful information in regard to an X and Y position on the display and passes this information to the main controller **50**.

[0100] If a television or VCR connection is available, the feed for the television or movie is provided to the decoder **34** where the feed is controlled as any other video signal received by the videophone **15**. The television or movie can appear aside a scene from the video connection with another videophone **15** on the display **54**.

[0101] The audio stream of the scene essentially follows a parallel and similar path with the audio video stream, except the audio stream is provided from an audio receiver **58**, such as a microphone, sound card, headset or hand set to a CS crystal 4201 audio interface **60** or such as a Codec which performs analog to digital and digital analog conversion of the signals, as well as controls volume and mixing, which digitizes the audio signal and provides it to a TCI 320C6711 or 6205 DSP **62**. The DSP **62** then packetizes the digitized audio stream and transfers the digitized audio stream to the FPGA **38**. The FPGA **38** in turn provides it to the PCI interface **44**, where it is then passed on to the network interface card **56** for transmission on the network **40**. The audio stream that is received by the receive videophone **15**, is passed to the FPGA **38** and on to the DSP **62** and then to the audio interface **60** which converts the digital signal to an analog signal for playback on speakers **64**.

[0102] The network interface card **56** time stamps each audio packet and video packet that is transmitted to the network **40**. The speed at which the audio and video that is received by the videophone **15** is processed is quick enough that the human eye and ear, upon listening to it, cannot discern any misalignment of the audio with the associated in time video of the scene. The constraint of less than 20-30 milliseconds is placed on the processing of the audio and video information of the scene to maintain this association of the video and audio of the scene. To insure that the audio and video of the scene is in synchronization when it is received at a receive videophone **15**, the time stamp of each packet is reviewed, and corresponding audio based packets and video based packets are aligned by the receiving videophone **15** and correspondingly played at essentially the same time so there is no misalignment that is discernible to the user at the receiver videophone **15** of the video and audio of the scene.

[0103] An ENC-DSP board contains the IBM eNV 420 MPEG-2 encoder and support circuitry, the DSP **62** for audio encoding and decoding, and the PCI interface **44**. It contains the hardware that is necessary for full videophone **15** terminal functionality given a high performance PC **68** platform and display **54** system **10**. It is a full size PCI 2.2 compliant design. The camera, microphone(s), and speakers **64** interface to this board. The DSP **62** will perform audio encode, decode, mixing, stereo placement, level control, gap filling, packetization, and other audio functions, such as stereo AEC, beam steering, noise cancellation, keyboard click cancellation, or de-reverberation. The FPGA **38** is developed using the Celoxia (Handel-C) tools, and is fully reconfigurable. Layout supports parts in the 1-3 million gate range.

[0104] This board includes a digital camera **47** chip interface, hardware or "video DSP" based multi-channel video decoder **34** interface, video overlay using the DVI in and out connectors, up to full dumb frame buffer capability with video overlay.

[0105] Using an NTSC or PAL video signal, the encoder **36** should produce a 640×480, and preferably a 720×480 or better resolution, high-quality video stream. Bitrate should be controlled such that the maximum bits per frame is limited in order to prevent transmission delay over the network **40**. The decoder **34** must start decoding a slice upon receiving the first macroblock of data. Some buffering may be required to accommodate minor jitter and thus improve picture.

[0106] MPEG-2 is widely used and deployed, being the basis for DVD and VCD encoding, digital VCR's and time shift devices such as TiVo, as well as DSS and other digital TV distribution. It is normally considered to be the choice for 4 to 50 Mbit/sec video transmission. Because of its wide use, relatively low cost, highly integrated solutions for decoding, and more recently, encoding, are commercially available now.

[0107] MPEG-2 should be thought of as a syntax for encoded video rather than a standard method of compression. While the specification defines the syntax and encoding methods, there is very wide latitude in the use of the methods as long as the defined syntax is followed. For this reason, generalizations about MPEG-2 are frequently misleading or inaccurate. It is necessary to get to lower levels of detail about specific encoding methods and intended application in order to evaluate the performance of MPEG-2 for a specific application.

[0108] Of interest to the videophone **15** project are the issues of low delay encode and decode, as well as network **40** related issues. There are three primary issues in the MPEG-2 algorithm that need to be understood to achieve low delay high quality video over a network **40**:

- **[0109]** The GOP (Group Of Pictures) structure and its effect on delay
- **[0110]** The effect of bit rate, encoded frame size variation, and the VBV buffer on delay and network **40** requirements
- **[0111]** The GOP structure's effect on quality with packet loss

The GOP Structure and Delay:

[0112] MPEG-2 defines 3 kinds of encoded frames: I, P, and B. The most common GOP structure in use is 16 frames long: IPBBPBBPBBPBBPBB. The problem with this structure is that each consecutive B frame, since a B frame is motion estimated from the previous and following frame, requires that the following frames are captured before encoding of the B frame can begin. As each frame is 33 msec, this adds a minimum of 66 msec additional delay for this GOP structure over one with no B frames. This leads to

a low delay GOP structure that contains only I and/or P frames, defined in the MPEG-2 spec as SP@ML (Simple Profile) encoding.

Bit Rate, Encoded Frame Size, and the VBV

[0113] Once B frames are eliminated to minimize encoding delay, the GOP is made up of I frames and P frames that are relative to the I frames. Because an I frame is completely intraframe coded, it takes a lot of bits to do this, and fewer bits for the following P frames.

[0114] Note that an I frame may be 8 times as large as a P frame, and 5 times the nominal bit rate. This has direct impact on network **40** requirements and delay: if there is a bandwidth limit, the I frame will be buffered at the network **40** restriction, resulting in added delay of multiple frame times to transfer over the restricted segment. This buffer must be matched at the receiver because the play-out rate is set by the video, not the network **40** bandwidth. The sample used for the above data was a low motion office scene; in high motion content with scene changes, frames will be allocated more or less bits depending on content, with some large P frames occurring at scene changes.

[0115] To control this behavior, MPEG-2 implements the VBV buffer (Video Buffering Verifier), which allows a degree of control over the ratio between the maximum encoded frame size and the nominal bit rate. By tightly constraining the VBV so that the I frames are limited to less than 2× the size indicated by the nominal bit rate, the added buffering delay can be limited to 1 additional frame time. The cost of constraining the VBV size is picture quality: the reason for large I frames is to provide a good basis for the following P frames, and quality is seriously degraded at lower bit rates (<4 Mbit) when the size of the I frames is constrained. Consider that at 2 Mbit, the average frame size is 8 Kbytes, and even twice this size is not enough to encode a 320×240 JPEG image with good quality, which is DCT compressed similar to an I frame.

[0116] Going to I frame only encoding allows a more consistent encoded frame size, but with the further degradation of quality. Low bit rate I frame only encoding does not take advantage of the bulk of the compression capability of the MPEG-2 algorithm.

[0117] The MPEG-2 specification defines CBR (Constant Bit Rate) and VBR (Variable Bit Rate) modes, and allows for variable GOP structure within a stream. CBR mode is defined to generate a consistent number of bits for each GOP, using padding as necessary. VBR is intended to allow consistent quality, by allowing variation in encoding bandwidth, permitting the stream to allocate more bits to difficult to encode areas as long as this is compensated for by lower bit rates in simpler sections. VBR can be implemented with two pass or single pass techniques. Variable GOP structure allows, for example, the placement of I frames at scene transition boundaries to eliminate visible compression artifacts. Due to the low delay requirement and the need to look ahead a little bit in order to implement VBR or variable GOP, these modes are of little interest for the videophone **15** application.

[0118] Because P and B frames in a typical GOP structure are dependant on the I frame and the preceding P and B frames, data loss affects all of the frames following the error until the next I frame. This also affects startup latency, such as when flipping channels on a DSS system **10**, where the decoder **34** waits for an I frame before it can start displaying an image. For this reason, GOP length, structure, and bit rate need to be tuned to the application and delivery system **10**. In the case of real time collaboration using IP, an unreliable transport protocol such as RTP or UDP is used because a late packet must be treated as lost, since you can't afford the delay required to deal with reliable protocol handshaking and retransmission. Various analysis has been done on the effect of packet loss on video quality, with results showing that for typical IPB GOP structures, a 1% packet loss results in 30% frame loss. Shorter GOP structures, and ultimately I frame only streams (with loss of quality), help this some, and FEC (Forward Error Correction) techniques can help a little when loss occurs, but certainly one of the problems with MPEG-2 is that it is not very tolerant of data loss.

[0119] A GOP structure called Continuous P frame encoding addresses all of the aforementioned issues and provides excellent video quality at relatively low bit rates for the videophone **15**. Continuous P encoding makes use of the ability to intra-frame encode macro-blocks of a frame within a P frame. By encoding a pseudo-random set of 16×16 pixel macro-blocks in each frame, and motion-coding the others, the equivalent of I-frame bits are distributed in each frame. By implementing the pseudo-random macro-block selection to ensure that all blocks are updated on a frequent time scale, startup and scene change are handled in a reasonable manner.

[0120] IBM has implemented this algorithm for the S420 encoder, setting the full frame DCT update rate to 8 frames (3.75 times per second). The results for typical office and conference content is quite impressive. The encoding delay, encoded frame size variation, and packet loss behavior is nearly ideal for the videophone **15**. Review of the encoded samples shows that for scene changes and highly dynamic content that encoder **36** artifacts are apparent, but for the typical talking heads content of collaboration, the quality is very good.

[0121] High-quality audio is essential prerequisite for effective communications. High-quality is defined as full-duplex, a 7 kHz bandwidth, (telephone is 3.2 kHz), >30 dB signal-to-noise ratio, no perceivable echo, clipping or distortion. Installation will be very simple involving as few cables as possible. On board diagnostics will indicate the problem and how to fix it. Sound from the speakers **64** will be free of loud pops and booms and sound levels either too high or too low.

[0122] An audio signal from missing or late packets can be "filled" in based on the preceding audio signal. The audio buffer should be about 50 ms as a balance between network **40** jitter and adding delay to the audio. The current packet size of 320 samples or 20 ms could be decreased to decrease the encode and decode latency. However, 20 ms is a standard data length for RTP packets.

[0123] Some of the processes described below are available in commercial products. However, for cost and integration reasons, they will be implemented on a DSP **62**. In another embodiment, a second DSP **62** can perform acoustic echo cancellation instead of just one DSP **62** performing this function also.

[0124] The audio system **10** has a transmit and a receive section. The transmit section is comprised of the following:

Microphones

[0125] One of the principal complaints of the speaker phone is the hollow sound that is heard at the remote end.

This hollow sound is due to the room reverberation and is best thought of as the ratio of the reflected (reverberant) sound power over the direct sound power. Presently, the best method to improve pickup is to locate microphones close to the talker and thus increase the direct sound power. In an office environment, microphones could be located at the PC **68** monitor, on the videophone **15** terminal and at a white board.

Automatic Gain Control

[0126] The gain for the preamplifier for each microphone is adjusted automatically such that the ADC range is fully used. The preamp gain will have to be sent to other audio processes such as AEC and noise reduction.

CODEC

[0127] In its simplest form, this is an ADC device. However, several companies such as Texas Instruments and Analog Devices Inc have CODECS with analog amplifiers and analog multiplexers. Also, resident on the chip is a DAC with similar controls. The automatic gain control described in the previous section is implemented in the CODEC and controlled by the DSP **62**.

Noise Reduction

[0128] Two methods of noise reduction can be used to improve the SNR. The first method is commonly called noise gating that turns on and off the channel depending on the level of signal present. The second method is adaptive noise cancellation (ANC) and subtracts out unwanted noise from the microphone signal. In office environment, it would be possible use ANC to remove PA announcements, fan noise and in some cases, even keyboard clicks.

[0129] Noise reduction or gating algorithms are available in commercial audio editing packages such as Cool Edit and Goldwave that can apply special effects, remove scratch and pop noise from records and also remove hiss from tape recordings.

Acoustic Echo Cancellation

[0130] Echo is heard when the talker's voice returns to the talker after more than 50 ms. The echo is very distracting and thus needs to be removed. The two sources of echo are line echo and acoustic echo. The line echo is due to characteristics of a two-line telephone system **10**. The PSTN removes this echo using a line echo canceller (LEC). When using a speaker phone system **10**, acoustic echo occurs between the telephone speaker and the microphone. The sound from the remote speaker is picked by the remote microphone and returned to talker. Acoustic echo cancellation (AEC) is more difficult than LEC since the room acoustics are more complicated to model and can change suddenly with movement of people. There are many AEC products ranging from the stand-alone devices such as ASPI EF1210 to Signal Works object modules optimized to run on DSP **62** platforms.

Automixing

[0131] Automixing is selecting which microphone signals to mix together and send the monaural output of the mixer to the encoder **36**. The selection criteria is based on using the microphone near the loudest source or using microphones that are receiving sound that is above a threshold level. Automixers are commercially available from various vendors and are used in teleconferencing and tele-education systems.

Encoding

[0132] To reduce data transmission bandwidth, the audio signal is compressed to a lower bit rate by taking advantage of the typical signal characteristics and our perception of speech. Presently, the G.722 codec offers the best audio quality (7 kHz bandwidth @14 bits) at a reasonable bit rate of 64 kbits/sec.

RTP Transmission

[0133] The encoded audio data is segmented into **20** msec segments and sent as RealTime Protocol (RTP) packets. RTP was specifically designed for realtime data exchange required for VoIP and teleconference applications.

[0134] The receive section is:

RTP Reception

[0135] RTP packets containing audio streams from one or more remote locations are placed in their respective buffers. Missing or late packets are detected and that information is passed to the Gap Handler. Out of order packets are a special case of late packets and like late packets are likely to be discarded. The alternative is to have a buffer to delay playing out the audio signal for at least one packet length. The size of the buffer will have to be constrained such that the end-to-end delay is no longer than 100 ms.

Decoding

[0136] The G.722 audio stream is decoded to PCM samples for the CODEC.

Gap Handling

[0137] Over any network, RTP packets will be lost or corrupted. Therefore, the Gap Handler will "fill in" the missing data based on the spectrum and statistics of the previous packets. As a minimum, zeros should be padded in the data stream to make up data but a spectral interpolation or extrapolation algorithm to fill in the data can be used.

Buffering

[0138] Network jitter will require buffering to allow a continuous audio playback. This buffer will likely adjust its size (and hence latency) based on a compromise between the short-term jitter statistics and the effect of latency.

Rate Control

[0139] The nominal sample rate for a videophone **15** terminal is 16 kHz. However, slight differences will exist and need to be handled. For example, suppose that videophone **15** North samples at precisely 16,001 Hz while videophone **15** South samples at 15,999 Hz. Thus, the South terminal will accumulate 1 more samples per second than it outputs to the speaker and the North terminal will run a deficit of equal amount. Long-term statistics on the receiving buffer will be able to determine what the sample rate

differential is and the appropriate interpolation (for videophone **15** North) or decimation (for videophone **15** South) factor can be computed.

Volume Control

[0140] Adjusting the volume coming from the speakers **64** is typically done by the remote listeners. A better way might be to automatically adjust the sound from the speakers **64** based on how loud it sounds to the microphones in the room. Other factors such as the background noise and the listener's own preference can be taken into account.

Stereo Placement

[0141] Remote talkers from different locations can be placed in the auditory field. Thus, a person from location A would consistently come from the left, the person from location B from the middle and the person from location C from the right. This placement makes it easier to keep track of who is talking.

Speakers

[0142] The quality of the sound to some extent is determined by the quality of the speakers **64** and the enclosure. In any case, self-amplified speakers **64** are used for the videophone **15** terminal.

Differentiation

[0143] Present conferencing systems such as the PolyCom Soundstation offer satisfactory but bandlimited full-duplex audio quality. However, the bandwidth is limited to 3500 Hz and the resulting sound quality strains the ear and especially in distinguishing fricative sounds.

[0144] Videophone **15** extends the bandwidth to 7 kHz and automixes multiple microphones to minimize room reverberation. When three or more people are talking, each of the remote participants will be placed in a unique location in the stereo sound field. Combined with the high-quality audio pick-up and increased bandwidth, a conference over the network **40** will quickly approach that of being there in person.

[0145] The audio system **10** uses multiple microphones for better sound pick-up and a wideband encoder (G.722) for better fidelity than is currently offered by tollgrade systems. Additionally, for multiple party conferences, stereo placement of remote talkers will be implemented and an acoustic echo cancellation system **10** to allow hands free operation. Adjustment of volume in the room will be controlled automatically with a single control for the end user to adjust the overall sound level.

[0146] In the videophone **15** network **40**, a gateway **70** connects something non-SIP to the SIP environment. Often there are electrical as well as protocol differences. Most of the gateways **70** connect other telephone or video conference devices to the videophone **15** system **10**.

[0147] Gateways **70** are distinguished by interfaces; one side is a network **40**, for videophone **15** this is Ethernet or ATM. The external side may be an analog telephone line or RS-232 port. The type, number and characteristics of the ports distinguishes one gateway **70** from another. On the network **40** side, there are transport protocols such as RTP or AAL2, and signaling protocols such as SIP, Megaco or MGCP.

[0148] On the external side, there may be a wide variety of protocols depending on the interfaces provided. Some examples would be ISDN (Q.931) or POTS signaling. PSTN gateways **70** connect PSTN lines into the videophone **15** system **10** on site. PBX gateways **70** allow a videophone **15** system **10** to emulate a proprietary telephone to provide compatibility to existing on-site PBX. POTS gateways **70** connect dumb analog phones to a videophone **15** system **10**. H.323 gateways **70** connect an H.323 system **10** to the SIP based videophone **15** system **10**. This is a signaling-only gateway **70**—the media server **66** does the H.261 to MPEG conversion.

[0149] Three enabling technologies for the videophone **15** are the Session Initiation Protocol (SIP), the Session Description Protocol (SDP) and the Real-time Transport Protocol (RTP), all of which are incorporated by reference herein.

[0150] SIP is a signaling protocol for initiating, managing and termination voice and video sessions across packet networks.

[0151] SDP is intended for describing multimedia sessions for the purposes of session announcement, session invitation, and other forms of multimedia session initiation. SIP uses SDP to describe media sessions.

[0152] RTP provides end-to-end network **40** transport functions suitable for applications transmitting real-time data, such as audio, video or simulation data, over multicast or unicast network **40** services. SIP uses RTP for media session transport.

[0153] The videophone **15** can perform conferences with three or more parties without the use of any conferencing bridge or MCU. This is accomplished by using ATM point to multipoint streams as established by SIP. More specifically, when the MPEG-2 stream and the low frame rate stream is packetized for transmission onto the network **40**, the header information for each of the packets identifies the addresses of all the receive videophones **15** of the conference, as is well known in the art. From this information, when the packets are transmitted to the network **40**, SIP establishes the necessary connectivity for the different packets to reach their desired videophone **15** destinations.

[0154] As an example of a conference that does not use any conferencing bridge, let there be **10** videophones **15** at discreet locations who are parties to a conference. Each videophone **15** produces an audio based stream, and an MPEG-2 based stream and a low frame rate based stream. However, each videophone **15** will not send any of these streams back to itself, so effectively, in a 10 party conference of videophones **15**, each communicate with the nine other videophones **15**. While it could be the case that the videophone **15** communicates with itself, to maximize the bandwidth utilization, the video produced by any videophone **15** and, if desired, the audio produced by a videophone **15** can be shown or heard as it essentially appears to the other videophones **15**, but through an internal channel, which will be described below, that does not require any bandwidth utilization of the network **40**.

[0155] In the conference, each videophone **15** receives nine audio based streams of data. Three MPEG-2 based streams of data and six low frame rate based streams of data. If desired, the receiver could choose up to nine streams of low frame rate based streams so the display **54** only shows the smaller images of each videophone **15**, or up to four of the MPEG-2 based streams of data where the display **54** is

filled with four images from four of the videophones **15** of the conference with no low frame rate based streams having their image shown, since there is no room on the display **54** for them if four MPEG-2 based streams are displayed. By having three MPEG-2 based streams shown, this allows for six of the low frame rate based streams to be shown. Each of the streams are formed as explained above, and received as explained above at the various videophones **15**.

[0156] If more than four large images are desired to be shown of a conference, then the way that this is accomplished is additional videophones **15** are connected together so that the displays of the different videophones **15** are lined up side by side, as shown in FIG. **7**. One videophone **15** can be the master, and as each additional videophone is added, it becomes a slave to the master videophone **15**, which controls the display **54** of the large and small images across the different videophones **15**.

[0157] In terms of the protocols to determine who is shown as a large image and who is shown as a small image on the displays of the videophones **15** of the conference, one preferred protocol is that the three most recent talkers are displayed as large, and the other parties are shown as small. That is, the party who is currently talking and the two previous talkers are shown as large. Since each videophone **15** of the conference receives all the audio based streams of the conference, each videophone **15** with its main controller **50** can determine where the talking is occurring at a given moment and cause the network interface card **56** to accept the MPEG-2 stream associated with the videophone **15** from which talking is occurring, and not accept the associated low frame rate stream. In another protocol, one videophone **15** is established as the lead or moderator videophone **15**, and the lead videophone **15** picks what every other videophone **15** sees in terms of the large and small images. In yet another protocol, the choice of images as to who is large and who is small is fixed and remains the same throughout the conference. The protocol can be that each videophone **15** can pick how they want the images they receive displayed. Both the MPEG-2 based stream and the low frame rate stream are transmitted onto the network **40** to the receive videophones of the conference. Accordingly, both video based streams are available to each receive videophone **15** to be shown depending on the protocol for display **54** that is chosen.

[0158] In regard to the audio based streams that are transmitted by each videophone **15**, to further effectively use the bandwidth, and to assist in the processing of the audio by decreasing the demands of processing placed on any transmit videophone **15** or receive videophone **15**, an audio based stream can only the transmitted by a videophone **15** when there is audio above a predetermined decibel threshold at the transmit videophone **15**. By only transmitting audio based streams that have a loud enough sound, with the assumption that the threshold would be calibrated to be met or exceeded when talking is occurring, this not only eliminates extraneous background noise from having to be sent and received, which essentially contributes nothing but uses bandwidth, but assists in choosing the MPEG-2 stream associated with the talking since only the audio streams that have talking are being received.

[0159] As mentioned above, if a given videophone **15** desires to see its own image that is being sent out to the other videophones **15**, then the low frame rate stream that is formed by the FPGA **38** is sent to a local memory in the videophone **15**, but without any compression, as would be the case for the low frame rate stream that is to be packetized and sent onto the network **40** from the videophone **15**. From this local memory, the main processor with software will operate on it and cause it to be displayed as a small image on the display **54**.

[0160] Furthermore, the videophone **15** provides for the control of which audio or video streams that it receives from the network **40** are to be heard or seen. In situations where the conference has more parties than a user of the videophone **15** wishes to see or hear, the user of the videophone **15** can choose to see only or hear only a subset of the video or audio streams that comprise the total conference. For instance, in a 100 party conference, the user chooses to see three of the video streams as large pictures on the screen, and 20 of the video streams as a small images on the screen, for a total of 23 pictures out of the possible 100 pictures that could be shown. The user of the videophone **15** chooses to have the three loudest talkers appear as the large pictures, and then chooses through the touch screen **74** of the parties in the conference, which are listed on a page of the touch screen, to also be displayed as the small pictures. Other protocols can be chosen, such as the 20 pictures that are shown as small pictures can be the last 20 talkers in the conference starting from the time the conference began and each party made his introductions. By controlling the number of video streams shown, organization is applied to the conference and utilization of the resources of the videophone **15** are better allocated.

[0161] In regard to the different pictures that are shown on the screen, a choice can be associated with each picture. For example, one picture can be selected by a moderator of the conference call, two of the pictures can be based on the last/loudest talkers at a current time of the conference, and the other picture can be associated with a person the user selects from all the other participants of the conference. In this way, every participant or user of the conference could potentially see a different selection of pictures from the total number of participants in the conference. The maximum bandwidth that is then needed is for one video stream being sent to the network, and four video streams being received from the network, regardless of the number of participants of the conference.

[0162] In regard to the audio streams, the limitation can be placed on the videophone **15** that only the audio streams associated with the three loudest talkers are chosen to be heard, while their respective picture is shown on the screen. The DSP **62** can analyze the audio streams that are received, and allow only the three audio streams associated with the loudest speakers to be played, and at the same time, directing the network interface **42** to only receive the first video streams of the large pictures associated with the three audio streams having the loudest talkers. Generally speaking, the more people that are talking at the same time, the more confusion and less understanding occurs. Thus, controls by the user are exercised over the audio streams to place some level of organization to them.

[0163] As part of the controls in regard to the audio streams, as mentioned above, each videophone **15** will only send out an audio stream if noise about the videophone **15** is above a threshold. Preferably, the threshold is dynamic and is based on the noise level of the three loudest audio streams associated with the three loudest talkers at a given time. This follows, since for the audio stream to be considered as one of the audio streams with the three loudest

talkers, the noise level of other audio streams must be monitored and identified in regard to their noise level. The DSP **62** upon receiving the audio streams from the network interface **42** through the network **40**, reviews the audio stream and identifies the three streams having the loudest noise, and also compares the noise level of the three received audio streams which have been identified with the three loudest talkers with the noise level of the scene about the videophone **15**. If the noise level from the scene about the videophone **15** is greater than any one of the audio streams received, then the videophone **15** sends its audio stream to the network **40**. This type of independent analysis by the DSP **62** occurs at each of the videophones in the conference, and is thus a distributive analysis throughout the conference. Each videophone, independent of all the other videophones, makes its own analysis in regard to the audio streams it receives, which by definition have only been sent out by the respective videophone **15** after the respective videophone **15** has determined that the noise about its scene is loud enough to warrant that at a given time it is one of the three loudest. Each videophone **15** than takes this received audio stream information and uses it as a basis for comparison of its own noise level. Each videophone **15** is thus making its own determination of threshold.

[0164] An alternative way of performing this distributed analysis is that each videophone, after determining what it believes the threshold should be with its DSP **62**, can send this threshold to all the other videophones of the conference, so all of the videophones can review what all the other videophones consider the threshold to be, and can, for instance, average the thresholds, to identify a threshold that it will apply to its scene.

[0165] By using the technique of choosing the video streams of the three loudest talkers, there may be moments when parties start talking loudly all at once, and creating confusion and inability for understanding, but by doing so it raises the noise in the threshold level, resulting in very shortly the elimination of the audio streams that are not producing as much noise as others, so that only the audio streams of the three largest talkers will once again be chosen and heard, with the others not being chosen, and thus removing some of the noise that the other audio streams might be contributing. This implies that there may be times when more than three audio streams are received by the videophone **15** since more than three videophones may have a noise level above the threshold at a given moment, allowing each of such videophones to produce an audio stream at that time and to send it to the network **40**. However, as just explained, once the threshold is changed, the situation will stop. This distributed analysis in regard to audio streams, is not limited to the videophone **15** described here but is also applicable to any type of an audio conference, whether there is also present video streams or not.

[0166] Consistent with the emphasis on conserving the use of bandwidth, and to send only what is necessary to conserve the bandwidth, clipping of an image occurs at the encoder **36** rather than at the receive videophone **15**. In the instances where the transmit videophone **15** is aware of how its image will appear at the receive videophones **15**, the encoder **36** clips the large image of the scene before it is transmitted, so there is that much less of the image to transmit and utilize bandwidth. If clipping is to occur at the receiver videophone **15**, then the main processor with software will operate on the received image before it is provided to the display controller **52**.

[0167] A second camera can be connected to the videophone **15** to provide an alternative view of the scene. For instance, in a room, the first camera, or primary camera, can be disposed to focus on the face of the viewer or talker. However, there may be additional individuals in the room which the person controlling the videophone **15** in the room wishes to show to the other viewers at the receive videophones **15**. The second camera, for instance, can be disposed in an upper corner of the room so that the second camera can view essentially a much larger portion of the room than the primary camera. The second camera feed can be provided to the decoder **34**. The decoder **34** has several ports to receive video feeds. Alternatively, if the stream from the second camera is already digitized, it can be provided to the processing elements of the videophone **15** through similar channels as the primary camera. Preferably, each videophone **15** controls whatever is sent out of it, so the choice of which camera feed is to be transmitted is decided by the viewer controlling the videophone **15**. Alternatively, it is possible to provide a remote receive videophone **15** the ability to control and choose which stream from which camera at a given videophone **15** is to be transmitted. The control signals from the control videophone **15** would be transmitted over the network **40** and received by the respective videophone **15** which will then provide the chosen stream for transmission. Besides a second camera, any other type of video feed can also be provided through the videophone **15**, such as the video feed from a DVD, VCR or whiteboard camera.

[0168] In a preferred embodiment, the videophone **15** operates in a peak mode. In the peak mode, the videophone **15** camera takes a still image of the scene before it and transmits this image to other videophones **15** that have been previously identified to receive it, such as on a list of those videophones **15** on its speed dial menu. Alternatively, in the peak mode, the still image that is taken is maintained at the videophone **15** and is provided upon request to anyone who is looking to call that videophone **15**. Ideally, as is consistent with the preferred usage of the videophone **15**, each videophone **15** user controls whatever is sent out of the videophone **15**, and can simply choose to turn off the peak mode, or control what image is sent out. When an active call occurs, the peak mode is turned off so there is no conflict between the peak mode and the active call in which a continuous image stream is taken by the camera. The peak mode can have the still image of the scene be taken at predetermined time intervals, say at one-minute increments, five-minute increments, 30-minute increments, etc. In the peak mode, at a predetermined time before the still image is taken, such as five or ten seconds before the image is taken, an audible queue can be presented to alert anyone before the camera that a picture is about to be taken and that they should look presentable. The audible queue can be a beep, a ping or other recorded noise or message. In this way, when the peak mode is used, a peak into the scene before the camera of the videophone **15** is made available to other videophones **15** and provides an indication of presence of people in regard to the camera to the other videophones **15**.

[0169] As another example of a presence sensor, the location of the automatic lens of the camera in regard to the field before it can act as a presence sensor. When no one is

before the camera, then the automatic lens of the camera will focus on an object or wall that is in its field. When a person is before the camera, the automatic lens will focus on that person, which will cause the lens to be in a different position than when the person is not before the lens. A signal from the camera indicative of the focus of the lens can be sent from the camera to the FPGA **38** which then causes the focus information to be sent to a predetermined list of videophone **15** receivers, such as those on the speed dial list of the transmit videophone **15**, to inform the receive videophones **15** whether the viewer is before the videophone **15** to indicate that someone is present.

[0170] The videophone **15** also provides for video mail. In the event a video call is attempted from one videophone **15** to another videophone **15**, and the receive videophone **15** does not answer the video call after a predetermined time, for instance 4 rings, then a video server **66** associated with the receive videophone **15** will respond to the video call. The video server **66** will answer the video call from the transmit videophone **15** and send to the transmit videophone **15** a recorded audio message, or an audio message with a recorded video image from the receive videophone **15** that did not answer, which had been previously recorded. The video server **66** will play the message and provide an audio or an audio and video queue to the caller to leave their message after a predetermined indication, such as a beep. When the predetermine indication occurs, the caller will then leave a message that will include an audio statement as well as a video image of the caller. The video and audio message will be stored in memory at the video server **66**. The message can be as long as desired, or be limited to a predetermined period of time for the message to be defined. After the predetermined period of time has passed, or the caller has finished and terminated the call, the video server **66** saves the video message, and sends a signal to the receive videophone **15** which did not answer the original call, that there is a video message waiting for the viewer of the receive videophone **15**. This message can be text or a video image that appears on the display **54** of the receive videophone **15**, or is simply a message light that is activated to alert the receive videophone **15** viewer that there is video mail for the viewer.

[0171] When the viewer wishes to view the video mail, the viewer can just choose on the touch screen **74** the area to activate the video mail. The user is presented with a range of mail handling options, including reading video mail, which sends a signal to the video server **66** to play the video mail for the viewer on the videophone **15** display **54**. The image stream that is sent from the video server **66** follows the path explained above for video based streams to and through the receive videophone **15** to be displayed. For the videophone **15** viewer to record a message on the video server **66** to respond to video calls when the viewer does not answer the video calls, the viewer touches an area on the touch screen **74** which activates the video server **66** to prompt the viewer to record a message either audio or audio and video, at a predetermined time, which the viewer than does, to create the message.

[0172] The videophone **15** provides for operation of the speakers **64** at a predetermined level without any volume control by the user. The speakers **64** of the videophone **15** can be calibrated with the microphone so that if the microphone is picking up noise that is too loud, then the main controller **50** and the DSP **62** lowers the level of audio output of the speakers **64** to decrease the noise level. By setting a predetermined and desirable level, the videophone **15** automatically controls the loudness of the volume without the viewer having to do anything.

[0173] The videophone **15** can be programmed to recognize an inquiry to speak to a specific person, and then use the predetermined speech pattern that is used for the recognition as the tone or signal at the receive videophone **15** to inform the viewer at the receive videophone **15** a call is being requested with the receive videophone **15**. For instance, the term "Hey Craig" can be used for the videophone **15** to recognize that a call is to be initiated to Craig with the transmit videophone **15**. The viewer by saying "Hey Craig" causes the transmit videophone to automatically initiate a call to Craig which then sends the term "Hey Craig" to the receive videophone **15** of Craig. Instead of the receive videophone **15** of Craig ringing to indicate a call is being requested with Craig, the term "Hey Craig" is announced at the videophone **15** of Craig intermittently in place of the ringing that normally would occur to obtain Craig's attention. The functionality to perform this operation would be performed by the main controller **50** and the DSP **62**. The statement "Hey Craig" would be announced by the viewer and transmitted, as explained above, to the server **66**. The server **66**, upon analyzing the statements, would recognize the term as a command to initiate a call to the named party of the command. The server **66** would then utilize the address information of the videophone **15** of Craig to initiate the call with the videophone **15** of Craig, and cause the signal or tone to be produced at the videophone **15** of Craig to be "Hey Craig".

[0174] As is well known in the art, the encoder **36** is able to identify the beginning and the end of each frame. As the encoder **36** receives the data, it encodes the data for a frame and stores the data until the frame is complete. Due to the algorithm that the encoder **36** utilizes, the stored frame is used as a basis to form the next frame. The stored frame acts as a reference frame for the next frame to be encoded. Essentially this is because the changes to the frame from one frame to the next are the focus for the encoding, and not the entire frame from the beginning. The encoded frame is then sent directly for packetization, as explained above, with out any buffering, except for packetization purposes, so as to minimize any delay. Alternatively, as the encoder **36** encodes the data for the frame, to even further speed the transmission of the data, the encoded data is ordered on for packetization purposes without waiting for the entire frame to be encoded. The data that is encoded is also stored for purposes of forming the frame, for reasons explained above, so that a reference frame is available to the encoder **36**. However, separately, the data as it is encoded is sent on for packetization purposes and forms into a frame as it is also being prepared for packetization, although if the packet is ready for transmission and it so happens only a portion of the frame has been made part of the packet, the remaining portion of the frame will be transmitted with a separate packet, and the frame will not be formed until both packets with the frame information are received at the receive videophone **15**.

[0175] Referring to FIG. **1**, videophones **15** are connected to the network **40**. Videophones **15** support 10/100 ethernet connections and optionally ATM 155 Mbps connections, on either copper or Multimode Fiber. Each videophone **15** terminal is usually associated with a users PC **68**. The role

of the videophone **15** is to provide the audio and Video aspects of a (conference) call. The PC **68** is used for any other functions. Establishing a call via the videophone **15** can automatically establish a Microsoft Netmeeting session between associated PCs **68** so that users can collaborate in Windows-based programs, for example, a Power Point presentation, or a spread sheet, exchange graphics on an electronic whiteboard, transfer files, or use a text-based chat program, etc. The PC **68** can be connected to Ethernet irrespective of how the videophone **15** terminal is connected. It can, of course, also be connected to an ATM LAN. The PC **68** and the associated transmit videophone **15** communicate with each other through the network **40**. The PC **68** and the associated transmit videophone **15** communicate with each other so the PC **68** knows to whom the transmit videophone **15** is talking. The PC **68** can then communicate with the PC **68** of the receive videophone **15** to whom the transmit videophone **15** is talking. The PC **68** can also place a call for the videophone **15**.

[0176] Most of the system **10** functionality is server based, and is software running of the videophone **15** Proxy Server, which is preferably an SIP Proxy Server. One server **66** is needed to deliver basic functionality, a second is required for resilient operation, i.e. the preservation of services in the event that one server **66** fails. Software in the servers and in the videophone **15** terminal will automatically swap to the back up server **66** in this event. With this configuration, videophone **15** terminals can make or receive calls to any other videophone **15** terminal on the network **40** and to any phones, which are preferably SIP phones, registered on the network.

[0177] Media Servers provide a set of services to users on a set of media streams. The media server **66** is controlled by a feature server **66** (preferably an feature server **66**). It is employed to provide sources and sinks for media streams as part of various user-invocable functions. The services provided on the media server **66** are:

[0178] Conference Bridging

[0179] Record and Playback

[0180] Transcoding

[0181] Tones and announcements

[0182] The media server **66** is a box sitting on the LAN or WAN. In general, it has no other connections to it. It is preferably an SIP device. The feature servers are in the signaling path from the videophone **15** terminals. The media path, however, would go direct from the media server **66** to the appliance.

[0183] In operation, the user may ask for a function, such as videomail. The feature server **66** would provide the user interface and the signaling function, the media server **66** would provide the mechanisms for multimedia prompts (if used) and the record and playback of messages.

[0184] To enable a videophone **15** terminal to make or accept calls to any non protocol or standard (such as SIP) (video) phones, a Gateway **70**, such as an SIP gateway, is added. A four analogue line gateway **70** can be connected either directly to the PSTN, or to analogue lines of the local PBX. The normal rules for provisioning outgoing lines apply. Typically one trunk line is provisioned for every six users, i.e. it assumes any one user uses his phone to dial an external connection 10 minutes out of any hour. If the videophone **15** terminal is to act as an extension on a current PBX as far as incoming calls are concerned then one analogue line is needed for every videophone **15**.

[0185] TV sources, such as CNN, are available to the videophone **15** user. The videophone **15** Video Server **66** enables this service. The Server **66** supports the connection of a single Video channel that is then accessible by any videophone **15** user on the network **40**. The Video channel is the equivalent of two normal conference sessions. A tuner can set the channel that is available. A new videophone **15** Video Server **66** should be added to the configuration for each different channel the customer wishes to have available simultaneously.

[0186] The videophone **15** server **66** (preferably SIP) also contains a database for user data, including a local cache of the users contact information. This database can be synchronized with the users main contact database. Synchronization can be used, for instance, with Outlook/Exchange users and for Lotus Notes users. A separate program that will run on any NT based server **66** platform does synchronization. Only one server **66** is required irrespective of the number of sites served.

[0187] As shown in FIG. **2**, usually videophone **15** terminals will be distributed across several sites, joined by a Wide Area network **40**. One server **66** is sufficient to serve up to 100+ videophones **15** on a single campus. As the total number of videophones **15** on a site increases, at some stage more servers need to be installed.

[0188] With videophones **15** distributed across several sites, it is possible for them to operate based on central servers, but this is not a recommended configuration, because of the WAN bandwidth used and the dependence on the WAN. Preferably, each site has at least one server **66**, which is preferably an SIP server **66** when SIP is used. For the more cautious, the simplest and easiest configuration is if each site has duplicated servers, preferably each being SIP servers. However using a central server **66** as the alternate to remote site servers will work too.

[0189] Videophones **15** anywhere in the network **40** can make PSTN or PBX based outgoing calls from a single central gateway **70**. However, if there is the need for the videophone **15** to also be an extension on a local PBX to accept incoming calls then a PSTN gateway **70** needs to be provided at each location. There needs to be a port on the gateway **70** for every videophone **15** on that site.

[0190] A central CNN server **66** can distribute TV channel to any videophone **15** on the network **40**. Nevertheless, it may be preferable to include site specific servers than take that bandwidth over the WAN.

[0191] A videophone **15** is available to connect to either a 10/100 Ethernet network **40** or an ATM network **40** at 155 Mbits/sec (with both Fiber and Copper options). An ATM connected videophone **15** uses an IP control plane to establish the ATM addresses of the end-points for a call, and then uses ATM signaling to establish the bearer channel between those end points. The bearer channel is established a Switched Virtual Circuit (SVC), with the full QoS requirements specified.

[0192] Each video stream is between 2 Mbps and 6 Mbps duplex as determined by settings and bandwidth negotiation. As the display means can show more than a single video stream, the overall required connection bandwidth to each videophone increases with the number of parties in the call. Transmit end clipping ensures that the maximum required bandwidth is approximately 2.5 times the single video stream bandwidth in use. If there are several videophones **15** on a site, the normal telephone ratio between users and

trunks will apply to videophone **15** sessions. In other words, a videophone **15** user is expected to talk on average to two other people in each call, i.e. two streams and will use the videophone **15** on average 10 minutes in the hour. For the average encoding rate of 3 Mbps, this gives a WAN bandwidth need of **6**Mbps which can be expected to support up to 6 users.

[0193] As shown in FIG. **3**, the videophone **15** operates on a 'p' enabled Ethernet network **40**, when there is a low density of videophone **15** terminals. The videophone **15** system **10** will establish an SVC across the ATM portion of the network **40** linking the two videophones **15** together, and make use of the 'p' enabled Ethernet to ensure sufficient Quality of Service is delivered over the Ethernet part of the connection.

[0194] The essential elements of the videophone **15** system **10** are shown in FIG. **4**. Together they create multimedia collaboration tools greatly enhancing the ability of geographically dispersed teams to interact. Such teams are increasingly common in almost every large enterprise, yet the tools to help them work effectively and efficiently are little changed from a decade ago and are in many respects unsatisfactory. Videophone **15** addresses the many issues of existing systems in a comprehensive way to create a discontinuous improvement in remote collaboration. It is enabled by newly available technology, differentiated by Quality of Service and the right mix of functions, made useable by the development of an excellent user interface, and designed to be extensible by using a standards based architecture.

[0195] The audio and video streams, as explained above, are transmitted from the originating videophone **15** to terminating videophones **15** on the network using, for example, well known SIP techniques. SIP messages may be routed across heterogeneous networks using IP routing techniques. It is desirable for media streams in heterogeneous networks to have a more direct path. Preferably, in instances where the originating videophone **15** of a conference is connected to an Ethernet, and a terminating videophone **15** of the conference is connected to an ATM network, as shown in FIG. **15**, the following addressing of the packets that cross the network between the originating and terminating videophones occurs. The originating videophone **15** sends a packet onto the Ethernet to which it is an communication with the originating videophone's IP address. The packet reaches an originating gateway **80** which links the Ethernet with the ATM network. At the originating gateway **80**, the IP address of the originating videophone **15** is saved from the packet, and the originating gateway **80** adds to the packet the ATM address of the originating gateway **80** and sends the packeton to the terminating videophone **15**. When the terminating videophone **15** receives the packet, it stores the ATM address of the originating gateway **80** from the packet, and sends back to the originating gateway **80** a return packet indicating that it has received the packet, with the ATM address of the terminating videophone **15**. The originating gateway **80**, when it receives the return packet saves the ATM address of the terminating videophone **15** and adds the IP address of the originating gateway **80** to the return packet. The return packet is then sent from the originating gateway **80** back to the originating videophone **15**.

[0196] In this way, the specific addresses of each critical node of the overall path between and with the originating videophone **15** and the terminating videophone **15** is known to each critical node of the path. At minimum, each node on the path knows the address of the next node of the path, and if desired, additional addresses can be maintained with the respective packets as they move along the path so each node of the path knows more in regard to addresses of the critical nodes then the next node that the packet goes to. This is because as the packet moves from node to node, and specifically in the example, from the originating videophone **15** to the originating gateway **80** to the terminating videophone **15** and then back to the originating gateway **80** and then to the originating videophone **15**, each node saves the critical address of the previous node from which to the respective packet was received, and introduces its own address relative to the type of network the next node is part of. Consequently, all the critical addresses that each node needs to send the packet onto the next node are distributed throughout the path.

[0197] This example of transferring a packet from an originating videophone **15** on an Ethernet to a terminating videophone **15** on an ATM network also is applicable for the reverse, where the originating terminal or videophone **15** is in communication with an ATM network and the terminating videophone **15** is in communication with an Ethernet.

[0198] Similarly, the path can involve an originating videophone **15** in communication with an Ethernet and a terminating videophone **15** in communication with an Ethernet where there is an ATM network traversed by the packet in between, as shown in FIG. **16**. In such a case, there would be two gateways at each edge where there is an interface between the Ethernet and the ATM network. As explained above, the process would simply add an additional node to the path, where the originating gateway **80** introduces its own ATM address to the packet and sends it to the terminating gateway **82** which saves the originating gateway's ATM address and adds the terminating gateway's IP address to the packet, which it then sends onto the terminating videophone **15** on the Ethernet. With the return packet, the same thing happens in reverse, and each gateway saves the respective address information from the previous gateway or terminating videophone **15**, and adds its own address to the return packet that it sends on ultimately to the originating videophone **15**, with the originating gateway **80** and the originating videophone **15** saving the ATM address of the terminating gateway **82** or the originating gateway **80**, respectively, so the respective addresses in each link of the overall path is stored to more efficiently and quickly send on subsequent packets of a connection.

[0199] For instance, the main controller **50** and the network interface **42** of the videophone **15** can add the address of the videophone **15** to each packet that it sends to the network **40** using the same techniques that are well known to one skilled in the art of placing SIP routing information (or whatever standard routing information is used) with the packet. The network interface **42** also stores the address information it receives from a packet from a node on the network in a local memory. Similarly, for a gateway on the network **40**, the same can be applied. As is well known, the gateway has controlling means and a data processing means for moving a packet on to its ultimate destination. A network interface **42** and a main controller **50** of the controlling mechanism of the gateway, operating with well known techniques in regard to SIP routing information, stores address information received from a packet and places its own address information relative to a network **40** in which

it is going to send the packet, with the packet. For example, the address information of the gateway, or the videophone **15**, can be placed in a field that is in the header portion associated with the packet. It should be noted, that while the example speaks to the use of videophones **15** as terminating and originating sources, any type of device which produces and receives packets can be used as a node in this overall scheme.

[0200] The Virtual Presence Video-Phone (videophone) **15** is a desk top network **40** appliance that is a personal communications terminal. It replaces the phone on the users desk, providing all the features of a modern PBX terminal with the simplicity of user interface and ease of use afforded by videophones' **15** large touch screen **74**.

[0201] Videophone **15** adds the video dimension to all interpersonal communications, changing the experience to that of virtual presence. In the past the quality of video on video conference systems has not been high enough for the technology to be transparent videophone **15** is the first personal videophone to deliver high enough video quality to create the right experience. For effective real time video communication not only has the picture quality to be close to broadcast TV quality, but the latency must be kept very low. Lip Sync is also important if a natural conversation is to flow. All these issues have been addressed in the design of the videophone **15** video subsystem videophone **15** uses the latest encoder **36** and decoder **34** technology configured specifically for this application. In other words, videophone **15** gets as close as possible to 'being there'.

[0202] Videophone **15** also greatly improves on conventional speaker phone performance through the use of a high fidelity, near CD quality audio channel that delivers crystal clear voice. Stereo audio channels provide for spatial differentiation of each participants audio. Advanced stereo echo cancellation cancels not only all the sound from the units speakers **64** but enables the talker to carry on a conversation at normal conversational levels, even when in a noisy room.

[0203] Videophone **15** directly supports the establishment of up to 4 remote party (i.e. 5 way) video conference calls and or up to 10 party audio conference calls. Each user has visibility on the availability of all other members of his/her work group. The videophone **15** preferably uses Session Initiation Protocol (SIP) as a means of establishing, modifying and clearing multi-stream multi-media sessions. Videophone **15** can establish an audio call to any other SIP phone or to any other phone via a gateway **70**.

[0204] Videophone **15** places high demands on the network **40** to which it is attached. Videophone **15** video calls demand a network **40** that can supply continuous high bandwidth, with guarantees on bandwidth, latency and jitter. Marconi plc specializes in providing networks that support high Quality of Service applications. A conference room version of videophone **15** is also available.

[0205] The videophone **15** is a communications terminal (platform) that has the capability of fully integrating with a user's PC **68**, the computing platform. A videophone **15** application for the PC **68** provides a number of integration services between the PC **68** and the associated videophone **15** terminal. This will include the automatic establishment of NetMeeting sessions between the parties in a videophone **15** conference call, if so enabled, for the purpose of sharing applications such as whiteboard, or presentations, etc. other capabilities including "drag and drop" dialing by videophone **15** of a number on the PC **68**.

[0206] A set of servers, preferably each being SIP servers, provide call control and feature implementation to the network **40** appliances. These are software servers running on standard computing platforms, capable of redundancy. These servers also run a local copy of the users contact information database and users preference database. Applications available on these servers provide access to corporate or other LDAP accessible directories.

[0207] A synchronization server **66** maintains synchronization between the users main contact database and the local copy on the server **66** (preferably SIP). Outlook Exchange or Lotus Notes synchronization is supported. A set of Media Gateways **70** are used to the analogue or digital PSTN network **40**. A set of Media Gateways **70** interfaces to the most common PABX equipment, including the voice mail systems associated with those PABX's.

[0208] The Media server **66** provides a number of services to the videophone **15** terminal. It acts as a Bridging-Conference server **66** for video conference over 4 parties, if desired. It can also provide transcoding between the videophone **15** standards and other common audio or video formats, such as H320/H323. It can provide record and playback facilities, enabling sessions to be recorded and playback. It can provide the source of tones and announcements.

[0209] A Firewall according to the standard being used, such as an SIP Firewall, is required to securely pass the dynamically created RTP streams under the control of standard proxy software (such as SIP proxy software). A TV server **66** acts as a source of TV distribution, allowing videophone **15** users to select any channel supported, for example CNN.

[0210] Videophone **15** is for Ethernet and ATM desktops. The videophone **15** terminal will support end to end ATM SVC's and use them to establish connections with the requisite level of Quality of Service. Videophone **15** will also support IP connectivity via LANE services. For this to guarantee the required QoS, LANE **2** is required. The videophone **15** provides ATM passthrough to an ATM attached desk-top PC **68**, or an ATM to Ethernet pass through to attach the PC **68** via Ethernet.

[0211] The videophone **15** requires the support of end to end QoS. For an Ethernet attached videophone **15** the user connection needs to support 802.1p, DiffServ and/or IntServ or better. If the destination is reachable via an ATM network **40**, an Ethernet to ATM gateway **70** will be provided. The SIP proxy server **66** and SIP signaling will establish the ATM end-point nearest to the target videophone **15** terminal, i.e. its ATM address if it is ATM attached, or the ATM Ethernet gateway **70** that is closest. Signaling will establish an SVC across the ATM portion of the network **40** with the appropriate QoS. This SVC will be linked to the specific Ethernet flow generating the appropriate priority indication at the remote end.

[0212] The videophone **15** product line consists of several end terminals (appliances), a set of servers which provide features not built into the appliances, and a set of gateways **70** that connect the products to existing facilities and outside PSTN services. The basic functionality provided by the system **10** is:

[0213] Telephony Services, with video available on all "on-net" calls, very high quality audio and video

[0214] Multiparty Conference Services, audio and video, ad hoc or prescheduled, completely self-serve, fully integrated into the telephony services

[0215] Presence Services—with a variety of tools to determine availability for collaboration

[0216] Shared Surface Services—electronic whiteboard, application sharing, document sharing, presentation broadcast

[0217] Other value added services such as broadcast video (Mikes message to the troops) TV distribution. Online interactive training, etc. Session recording services is also available, if desired.

[0218] Videophone **15** is a telephone with dramatic new functionality, not a computer trying to do what a telephone does. This allows full concurrent use of a computer for the things that it is good at, while providing a flexible but application specific appliance for communication. The user interface and physical design can be tuned for this application, providing an instant on, highly reliable communications device like current phones, something that the PC **68** will never be. This approach also provides control over the operating environment of the device, eliminating the support problems related to PC **68** hardware and software configuration issues.

[0219] Human factor studies have demonstrated time after time that audio quality is the single most important factor for effective, transparent communication. While a handset is necessary, excellent quality hands free audio including Acoustic Echo Cancellation (AEC), Automatic Gain Control (AGC), wide band audio capability (G.722 8 kHz bandwidth or better), stereo output, and integration with the PC **68** sound output provides new levels of effective remote collaboration. A high quality microphone array, designed and processed to limit tin-can effects is also present.

[0220] A simple, clean, intuitive, fully flexible platform for visual output and button/selection input is used. In the first videophone model, this is a high quality TFT full color screen, 17" diagonal 16 by 9 screen with 1260×768 resolution or better, overlaid with a medium resolution high life touch-panel. A bright (>200 nit), extended viewing angle (>+−60°) active matrix panel is used to display full motion video for comfortable viewing in an office environment. Larger, brighter, faster, higher contrast, and higher viewing angle screens can be used.

[0221] The videophone **15** uses a TFT color LCD, having PC **68** like architecture with a VGA type display **54** interface based on an Intel Celeron/440 MMX and a Lynx VGA controller.

[0222] A high quality digital **480** line progressive scan camera is used to provide 30 frames per second of at least 640×480 video. Videophone **15** uses MPEG2 encoding taking advantage of the video encoder **36** technology for set top boxes. A variety of different bit rates can be generated, allowing the video quality to adapt to the available resources for one-to-one calls, and to the highest quality participant for one or many-to-many calls. An integrated high quality camera module is positioned close to the screen, with an external video input (Firewire) provided to allow the use of additional cameras, VCRs, or other video sources.

[0223] An existing 10/100 BaseT Ethernet connection to the desktop is the only connection necessary for communication to the LAN, WAN, PC **68** desktop, and various servers, proxies, and gateways **70**. Time critical RTP streams for audio and video are marked with priority using 802.1p, supplying the mechanism within the Ethernet domain of the LAN for QoS. DiffServ is also supported, with RSVP as an option. In order to eliminate the need for additional building wiring to the desktop, the videophone **15** will include a small 10/100 Ethernet switch, allowing the existing desktop port to be used for both the phone and the PC **68**.

[0224] Videophone **15** also supports an ATM interface. The interface is based on using the HE155 Mbits/sec card with either a fiber or copper interface. The videophone **15** provides an ATM pass-through port to connect to an ATM connected desktop or to connect an Ethernet connected PC **68** to the ATM connected videophone **15**.

[0225] The cost and performance tradeoffs for the conference room environment are obviously different than those for the desktop. Video projection, multiple cameras with remote Pan/Tilt/Zoom, multiple microphones, multiple video channels, rear projection white boards, and other products appropriate for the conference room environment are integrated into a conference room videophone **15**. The interworking of the conference room environment and the desktop is seamless and transparent. This environment will make heavy use of OEM equipment that is interfaced to the same infrastructure and standards in place for the desktop. The hardware design is essentially the same, with additional audio support for multiple microphones, and additional video support for multiple cameras and displays. Alternatively, a PC **68** application, either mouse or touch screen **74** driven, if the PC **68** has as touch screen **74**, that links to a low cost SIP phone can be used. For those desktops and other places that do not require the collaboration capabilities described above, a standard phone can be used that works with the system **10** without requiring additional wiring or a PBX.

[0226] Using the SIP (Session Initiation Protocol) standard, the terminal devices are supported by one or more servers that provide registration, location, user profile, presence, and various proxy services. These servers are inexpensive Linux or BSD machines connected to the LAN.

[0227] The videophone **15** is the phone, so a key set of PBX functions must be provided, including transfer, forward, 3 (and 4, 5, . . .) party conferencing, caller ID+, call history, etc. Some of these features may be built on top of a SIP extension mechanism called "CPL", which is actually a language to provide call handling in a secure, extensible manner.

[0228] The videophone **15** provides for active presence and instant messaging. Perhaps the most revolutionary tool for improving day to day distributed group collaborative work, presence allows people to know who's in and what they're doing. It provides the basis for very low overhead calling, eliminating telephone tag and traditional number dialing, encouraging groups to communicate as a group rather than through the disjoint one-to-one phone conversations that are common now. Integration with Instant Messaging (real time email) provides a no delay way of exchanging short text messages, probably making use of the PC **68** keyboard for input.

[0229] The videophone **15** provides for distributed/redundant architecture. This is the phone system **10** and it must be reliable. It should also be able to be centrally managed with local extensions, with distributed servers providing "instant" response to all users. Each of the different SIP proxy functions, for instance, if SIP is used, will be deployed such

that they can be arbitrarily combined into a set of physical servers, with redundant versions located in the network **40**.

[0230] Microsoft NetMeeting is used for shared surface and shared application functionality. Computer/Telephony Interface (CTI) for the PC **68** and PDA, with features such as integrated contact lists, auto-dialing of selected phone numbers or names, calendar logging of call history, automatic entry of contacts, etc. can be used.

[0231] SIP presents challenges to firewalls because the RTP flows use dynamically allocated UDP ports, and the address/port information is carried in SIP messages. This means the Firewall has to track the SIP messages, and open "pin holes" in the firewall for the appropriate address/port combinations. Further, if NAT is employed, the messages must be altered to have the appropriate translated address/ports. There are two ways to accomplish such a task. One is to build the capability into the firewall. The top 3 firewall vendors (Checkpoint, Network Associates and Axxent) provide this. An alternative is to have a special purpose firewall that just deals with SIP in parallel with the main firewall. There are commercial versions of such a firewall, for example, that of MicroAppliances. It should be noted that SIP or NetMeeting are preferred embodiments that are available to carry out their necessary respective functionality. Alternatives to them can be used, if the necessary functionality is provided.

[0232] FIG. **5** shows the main physical components of the videophone **15** terminal. The stand provides a means of easily adjusting the height of the main display **54** panel and of securing the panel at that height. The range of height adjustment is to be at least 6 inches of travel to accommodate different user heights. It is assumed that the stand will sit on a desk and that desktop heights are standardized. The link between the stand and the main unit must provide for a limited degree of tilt out of the vertical to suit user preference and be easily locked at that angle. The amount of tilt needed −0+15° from the vertical. The main unit can directly wall mount without the need of the stand assembly as an option.

[0233] The main unit case provides the housing for all the other elements in the videophone **15** design including all those shown in FIG. **5** and all the internal electronics. The case provides for either left-hand or right-hand mounting of the handset. Right-handed people tend to pick up the handset with the left hand (because they will drive the touch screen **74** and write with the right) and left handed people the reverse. Though the left hand location will be the normal one, it must be possible to position the handset on the right. A Speaker jack is provided on the case to allow the speakers **64** to be mounted remote from the videophone **15**. Inputs are provided to handle the speaker outputs from the associated PC **68**, so that videophone **15** can control the PC **68** and videophone **15** audio. Implementation of a wireless connection to speakers **64** (via Bluetooth, or SONY standards) can be used.

[0234] A handset is provided with the unit and should connect using a standard RJ9 coiled cable and connector jack. When parked the handset should be easy to pick-up and yet be unobtrusive. A handset option provides an on handset standard keypad. A wireless handset to improve mobility of the terminal user can be used.

[0235] A jack is provided for the connection of a stereo headset+microphone. Use of headsets for normal phone conversations is increasing. The user shall be able to choose to use a headset+boom mounted microphone, or a headset only, employing the microphone array as the input device. There is an option for a wireless headset to improve mobility of the terminal user.

[0236] An IR port is provided to interface to PDA's and other IR devices, in a position on the main case to allow easy access. For the moment IR interfaces on phones and PDA's are the most common and therefore for the same reasons as a bluetooth interface is required so too is an IR interface.

[0237] An array microphone is embedded in the casing. The array must not generate extraneous noise as a consequence of the normal operation of the terminal. Specifically, it should not be possible to detect user action on the touch-panel. The array microphone allows a user to talk at normal conversational levels within an arc (say 6 feet) round the front of the units and 110° in the horizontal plane and in the presence of predefined dbs of background noise. The unit must provide unambiguous indication that the microphone is active/not active, i.e. the equivalent of 'on-hook' or 'off-hook'. A videophone **15** user will want re-assurance that he is not being listened into without his knowledge. This is the audio equivalent of the mechanical camera shutter.

[0238] The main videophone **15** unit may have a smart card reader option to provide secure access to the terminal for personal features. Access to videophone **15** will need an array of access control features, from a simple password logon on screen, to security fob's. A smart card reader provides one of these access methods.

[0239] There is clearly an advantage if the tilt and pan is controllable from the screen, and preferably, if Pan and Tilt are electronic only and need no mechanical mechanisms. The camera mount should be mounted as close to the top of the main screen as possible to improve eye contact.

[0240] The camera should be a digital camera **47** capable of generating 480 p outputs. The camera output feeds an MPEG-2 encoder **36**. It should be possible to dynamically configure the camera so that the camera output is optimized for feeding the encoder **36** at the chosen encoder **36** output data-rate. Faces form the majority of input the camera will receive, and therefore the accurate capture under a wide range of lighting conditions of skin tone is an essential characteristic.

[0241] The camera should be operated in a wide range of lighting conditions down to a value of 3 lux. The camera should provide automatic white balance. White balance changes must be slow, so that transients on the captured image do not cause undue picture perturbation. Only changes that last over 5 seconds should change the white balance. The camera should be in focus from 18 inches inches to 10 feet, i.e. have a large depth of field and desirably be in focus to 20 feet. Both the user and the information if any on his white board both need to be in focus. Auto-focus, where the camera continually hunts for the best focus as the user moves, produces a disturbing image at the receiver end and must be avoided.

[0242] The camera should allow a limited zoom capability, from the setting where one user is directly in front of the camera, to another setting where a few users are simultaneously on one videophone **15**. As an alternative, different lenses may be provided. This can be specified in terms of lens field of view, from say a 30° field to view to a 75° field of view.

[0243] The camera should be able to input a larger picture than needed for transmission, for example a 1280×960

image. This would allow for limited zoom and horizontal and vertical pan electronically, removing the need for electromechanical controls associated with the camera. The camera should be physically small, so that an 'on-screen' mounting is not eliminated simply by the size of the camera.

[0244] A medium resolution long life touch panel forms the primary method of communicating with the videophone **15** and forms the front of the main display **54**. The panel will get a lot of finger contact and therefore must withstand frequent cleaning to remove smears and other finger prints that would otherwise affect the display **54** quality. It should be easy to calibrate the touch panel, i.e. ensure that the alignment between the area touched on the touch panel and the display **54** underneath will result in meeting the 'false touch' requirement.

[0245] The touch screen **74** surface must minimize surface reflections so that the display **54** is clear even when facing a window. The requirement is that 'false touches' are rare events. The resolution requirement on the touch panel is therefore heavily dependent on the smallest display **54** area touch is trying to distinguish. The resolution and the parallax error combined should be such that the chance of a 'false touch' due to these factors by the average trained user is less than 5%. (One false touch in 20 selections). It is desirable that this false touch ratio is less than 2%, i.e. one false touch in 50 selections.

[0246] Where appropriate, audible and or visible feedback of a successful touch must be given to the user. These tones may vary depending on what is on the touch screen **74** display **54** at the time. For example when using a keyboard, keyboard like sounds are appropriate, when using a dial-pad different sounds are likely to be relevant and so on. Audible feedback may not be needed in all circumstances, though usually some audible or visible indication of a successful touch is helpful to the user. It should be possible for the user to be able to turn tones on and off and set the tones, tone duration and volume level associated with the touch on some settings screen. Default values should be provided. The touch screen **74** can also be used with a stylus as well as the finger.

[0247] The display **54** panel should be at least 17" diagonal flat panel (or better) full color display **54** technology, with a 16×9 aspect ratio preferred but a 16×10 aspect ratio being acceptable.

[0248] The screen resolution should be at least 1280×768. The viewable angle should be at least 6° off axis in both horizontal and vertical planes. The screen contrast ratio should be better than 300:1 typical. The color resolution should be at least 6 bits per color, i.e. able to display 262K colors 6 bits per color is acceptable for the prototype units. 8 bits per color is preferred, other things being equal, for the production units. The display **54** panel should have a high enough brightness to be viewed comfortably even in a well lit or naturally lit room. The brightness should be at least 300 cd/m^2. The display **54** and the decode electronics should be able to display 720 P high resolution images from appropriate network **40** sources of such images.

[0249] The back light shall have a minimum life to 50% of minimum brightness of at least 25,000 hours. If the backlight is turned off due to inactivity on the videophone **15** terminal, then it should automatically turn on if there is an incoming call and when the user touches anywhere on the touchscreen. The inactivity period after which the touchscreen is turned off should be settable by the user, up to "do not turn off".

[0250] The connections required in the connection area of the videophone **15** are as shown in FIG. **6**. Each connector requirement will be briefly described in paragraphs below.

[0251] Two RJ 45 10/100 Ethernet connectors are for connection to the network **40** and from the associated PC **68**.

[0252] An optional plug in ATM personality module shall be provided that enables the videophone **15** to easily support 155 Mbits/sec interfaces for both Optical and copper interfaces.

[0253] A USB port shall be provided to allow various optional peripherals to be easily connected, for example a keyboard, a mouse, a low cost camera, etc.

[0254] A 1394 (Firewire) interface should be provided to permit connection to external (firewire) cameras or other video sources. The interface should permit full inband camera control over the firewire interface. Where necessary external converters should be used to convert from say S-Video to the firewire input. It should be possible to use this source in place of the main camera source in the videophone **15** output to the conference. It should also be possible to specify normal or "CNN" mode i.e. clippable or not clippable on this video source. An XVGA video output should be provided to enable the videophone **15** to drive external projectors with an image that reflects that displayed on the main display **54**.

[0255] An audio input shall be provided for PCAudio output. To ensure integration of the PC **68** audio and videophone **15** audio, only one set of speakers **64** will be deployed. The PC **68** sound will pass through the audio channel of the videophone **15**. A jack or pair of jacks shall be provided to connect to a head-set and attached boom microphone. Headset only operation, using the built in microphone array must also be possible. If the headset jack is relatively inaccessible, it should be possible to leave the headset plugged in, and select via a user control whether audio is on the headset or not. Connections are provided to external left and right hand speakers **64**. It is possible to use one, two or three videophone **15** units as though they were a single functional unit, as illustrated in FIG. **7**.

[0256] In configurations of more than one videophone **15**, only one unit acts as the main control panel, the other unit(s) display video and those controls directly associated with the video being displayed. Only one set of speakers **64** will be needed for any of these configurations.

[0257] A number of options shall be provided as far as microphone inputs and audio streams are concerned, from using a single common microphone input, to transmitting the audio from each microphone array to the sources of the video on that videophone **15**.

[0258] A number of options shall be provided for Video inputs. The default shall be to transmit the view of the 'control panel' videophone **15**. If more bandwidth is available then each user can get the Video from the screen on which the user is displayed, yielding a more natural experience. All co-ordination of the multiple videophone **15** terminals can be achieved over the LAN connection, i.e. not need any special inter-unit cabling.

[0259] The videophone **15** videophone provides its user with a number of main functions:

[0260] It is the office phone

[0261] It is the users Phone

[0262] It is a videophone
[0263] It is a conference phone
[0264] It is a video conference phone
[0265] It provides easy access to and management of contact details
[0266] It provides access and management of voice/video mail

[0267] The units functionality falls into two categories, user functions and systems functions.

[0268] User functions are any functions to which the user will have access.

[0269] System **10** functions are those required by I.T. to set up monitor and maintain the videophone **15** terminal and which are invisible to the normal user. Indeed, an important objective of the overall design is to make sure the user is presented with a very simple interface where he can use videophone **15** with virtually no training.

[0270] The following defines the basic feature set that is the minimum set of features that must be available.

[0271] The videophone **15** videophone acts as a conventional telephone when no user is logged onto the terminal. Its functionality must not depend at all on there being an associated PC **68**.

[0272] The following describes the functionality of videophone **15** as a conventional phone in an office.

[0273] The terminal is able to have a conventional extension number on the PABX serving the site.

[0274] The terminal is able to accept an incoming call from any phone, whether on the PABX, on the videophone **15** network **40** or any external phone without discrimination.

[0275] The videophone **15** is able to accept calls from other compatible SIP phones.

[0276] An incoming call will generate a ring tone as configured (see set up screen requirements below). Specifically, the ring tone for videophone **15** calls that include Video will have an option for a distinguishing ring from audio only calls, whether from videophone **15** terminals or not.

[0277] An incoming call will generate an incoming call indication in the status area on the display **54**. This display **54** must give as much Caller ID information as provided by the incoming call, or indicate that none is available.

[0278] It is possible to accept the incoming call:
[0279] a) By pressing the call accept button on the incoming call status display **54**.
[0280] b) By picking up the handset—which will always accept all the offered options i.e. video and audio.

[0281] It is possible for the user to switch between handset and hands free (speaker phone) operation easily within a call. Picking up the handset within a call should automatically switch to handset mode from speaker phone mode. Replacing the handset without reselecting speaker phone mode will disconnect the call.

[0282] An on screen indication should be given of the mode, i.e. handset or hands-free.

[0283] The call status bar can display the call duration.

[0284] It is possible to adjust the volume of the incoming call by readily available controls on the main display **54**. Headset and speaker volumes should be independently adjustable.

[0285] When in speaker phone mode, it is possible to return the handset to the handset stand without disconnecting the call.

[0286] A call is terminated:
[0287] If the user presses the clear call button on the call status display **54**.
[0288] If the user replaces the handset when in handset mode and hands free is not selected.
[0289] If the remote party hangs up the call provided it is reliably indicated to the videophone **15**.

[0290] HOLD—It should be possible to place a call on Hold and to take the call off Hold again. Hold status should be displayed on the status display **54**, with a button to allow that held call to be picked up.

[0291] CALL WAITING—Additional incoming calls must generate an incoming call indication in the status area of the display **54**. It must not generate a call tone, unless enabled in the settings menu.

[0292] It is possible to accept a new incoming call in the current operating mode, i.e. handset or hands free, from the call accept button on the status display **54**.

[0293] Accepting another incoming call will automatically place current calls on HOLD.

[0294] Pressing the "take off hold" button on any call must automatically transfer any other calls to Hold.

[0295] The number of simultaneous incoming calls that can be handled is set by the availability of status display **54** space. It must not be less than two calls.

[0296] When the number of current calls exceeds the number that can be handled, any other incoming calls:
[0297] a) Get a busy tone or
[0298] b) Are immediately forwarded to voice mail
[0299] c) Are immediately forwarded to the configured forwarding number
[0300] d) Are sent a recorded message.

[0301] As determined by the users "call forward busy" settings.

[0302] If incoming calls that are within the acceptable limit are not answered within a (configurable) interval, the calls are:
[0303] a) forwarded to voice mail
[0304] b) forwarded to the pre-configured forwarding number
[0305] c) sent a recorded message.

[0306] As determined by the user's "call forward no answer" settings.

[0307] CALL TRANSFER—It is possible for the user to easily transfer any call to any other number. The transfer function will put the call on hold and allow a new number to be dialed. Once ringing tone is heard, the user will have the option of completing the transfer. Alternatively, the user will be able to talk to the new number and then either initiate the transfer or first join all (three) parties in a conference call. If the latter, a function will be provided for the user to exit that conference call. In the event that there is no reply or just voice mail from the called terminal, the user will have the option of returning to the original call.

[0308] CALL FORWARD—It must be possible to set the phone up to automatically forward incoming calls to a pre-configured number. Call forwarding can be:
[0309] a) unconditional
[0310] b) forward on busy
[0311] c) forward on No Answer

[0312] CONFERENCE CALLS—It is possible to conference calls into an audio only conference, irrespective of the origin of the voice call. It is possible to conference at least 3 calls, i.e. a four-way conversation. It is required only to

support a single conference at any one time, but still be able to accept one other incoming call as described in call waiting above. It is acceptable that the prototype be only able to accept one incoming call to a particular conference, i.e. an external bridge will be needed for non-videophone calls.

[0313] Options associated with the incoming call status display **54** will allow the user to add or remove a call from a conference connection.

[0314] It is possible to add calls to a conference irrespective of whether they are incoming or outgoing calls.

[0315] If remote conference user hangs up, that call leg must be cleared automatically.

[0316] Calls can be made hands free or whilst using the handset. Picking up the handset should bring up the dial pad if not in a call and connect the audio to the handset. An on-screen tone dial pad (i.e. numbers 1 through 0 plus '*' and '#') is required. In addition, there should be a pause button to insert a pause into a dialed string (for getting through PABXs unless the gateway(s) **70** can be programmed to remove this requirement) Consideration should be given to adding a **+** key and arranging that the **+** sign is automatically translated into the international access string for that location.

[0317] A key to correct entry errors (eg [BACK] key and a clear key to clear the entry are also required. A short press of the [BACK] key should remove the last entered number, a longer press continue to remove numbers, a press over should clear the number register.

[0318] The number display **54** should be automatically formatted to the local number format. [This may require a user setting to select country of operation as each country has a different style or if an international code is entered that code should be used as the basis of formatting the remaining part of the number.]

[0319] When connected to services that make use of the tone number pad to select features, the correct tones must be generated in the direction of that service, when the on screen key pad or the handset key pad is used. The dial-pad must be able to provide this function irrespective of how the call is initiated.

[0320] REDIAL—It is possible to redial the last dialed number through a single touch on an appropriately identified function.

[0321] AUTO REDIAL—It is possible to trigger an auto-redial mechanism, for example by holding the [REDIAL] button. Auto redial will automatically repeat the call if the previous attempts return a busy signal a number of tries.

[0322] CAMP ON BUSY—When making a call to a device that permits its support, a "Camp on Busy" function is available. Camp on Busy calls the user back once the called party is available. A message shall be generated to say 'this service is not available' if the called number cannot support Camp on Busy.

[0323] There can be an appropriate log on screen displayed when no user is logged onto the videophone **15**.

[0324] A log of incoming, outgoing frequent and missed calls should be displayed on an appropriate view of the integrated dial screens. One or two touch access to 'last number re-dial' facility should always be available on the dial screens. Further definitions of these logs are given below.

[0325] To access the full set of features available on the videophone **15** terminal, a user must log into the terminal. A login screen is provided in which the user can enter his name and password. This can be the same as his normal network **40** access name and password The videophone **15** terminal will therefore make use of the sites user authentication services. Any screens needed to enable IT personnel to configure the videophone **15** to use these authentication services must be provided. Alternative methods of identifying the user are available, for example, the use of a smart card or ID fob. There is no requirement for the user to already be logged on to a PC **68** prior to logging in to a videophone **15** terminal.

[0326] Multiple users can be logged onto a single videophone **15** and distinct incoming ring tones for each user can be provided. The incoming call indication should also identify the called parties name and well as the calling parties name. If multiple users are logged onto a single videophone **15**, all the call forwarding functions are specific to the user to whom the call is addressed.

[0327] If the user is already logged in at his PC **68**, the action of logging onto the videophone **15** shall create an association between the PC **68** where the User was logged on and the videophone **15** terminal provided this is confirmed from the PC **68**. It is possible for a user to be logged on to multiple videophone **15** terminals simultaneously. The active videophone **15** is the one on which any call for that user is answered first.

[0328] The home page screen contains a status area that is visible on all screens (except in full screen mode). Status includes the name of the logged on user—or "no user logged on". The User's "Presence" status, Icons for video and audio transmission, Voice mail "Message" indication and the date and time.

[0329] A "message" indication is lit and flashing if there is unheard voice mail on the user voicemail system **10**. Pressing the indicator brings up the Voicemail handling screen.

[0330] Touching the Date time area gives access to the Calendar functions.

[0331] The home page has a control bar area that is visible across all screens (except in full screen mode).

[0332] The control bar gives direct access to the most frequently used call control features and access to all other functions. Icons should be used on the buttons, but text may also be used to emphasize functional purpose.

[0333] The control panel also has global controls for the microphone, Camera and Speakers **64**. The controls should clearly indicate their operational state, e.g. ON or OFF and where possible Icons should be used.

[0334] A self-image is available that indicates both the picture being taken by camera and that portion that is visible to the remote end of the active call. It is possible to turn self-image on and off and to determine whether it is always on or only once an active call has been established.

[0335] It is possible to display the camera image in the main video area of the screen at any time, i.e. in a call, not in a call, etc. The image should be that for a single Video call and should overlay any other video present. It should be possible to request a full screen version of that video. This can be thought of as a digital mirror and allows the user to make sure he/she is happy with what the camera will/is show(ing).

[0336] It is desirable for diagnostic purposes that the user can also see the image after encoding and decoding, so that he is aware of the quality of the image that will be seen at the far end. If this mode is supported then both the camera

direct and the encoded decoded image side by side. The user can capture his self image, for use as the image associated with his contact information.

[0337] The major part of the Home screen is allocated to an Integrated Dial functions. There are four main sub-functions, a speed dial display **54**, a directories access display **54**, a dial-pad and access to call logs. The dial-pad and access to call logs are to occupy the minimum screen area compatible with ease of use, maximizing the area available to the Speed Dial/Contacts pages. The speed dial area is detailed first, any common requirements across all the main sub-functions are only detailed under speed dial and are implied for the other three functions. The function of the Dial area is to select a user to whom a call is to be made.

[0338] The speed dial area is as large as possible, consistent with the other requirements for the dial screen. >20 speed dial locations is adequate. Each location should be large enough to make the identification of the persons detailed stored at that location very easily readable at the normal operational distance from the screen say 3 feet.

[0339] The user's information stored in a speed dial location includes the persons name, 'presence status' if known, the number that will be called if that speed dial is selected and an icon to indicate whether the user supports video calls. The detailed information also stores what kind of video, e.g. videophone **15**, compatible MPEG2, H261 etc.

[0340] The area provides a clear area to be touched to initiate a call. A thumbnail view of the person is included if available. A method of handling long names (i.e. names that do not fit in the space allocated on the Speed Dial button) is provided.

[0341] Conventional telephone numbers in standard international format i.e. "+ country code area code number" are automatically translated to the external access plus the international access codes needed to make a call to this number.

[0342] The full contact details associated with a person on the Speed dial page is available. The contact details provide all the numbers at which the user can be reached and a means of selecting one of the numbers as the default number that is used on the Speed Dial page. It is possible to select and dial an alternative number for that user via this link to the contacts page.

[0343] The User information includes most recent call history for that person, for example the last 10 calls either incoming missed or outgoing. Just providing the 'last call' information would be an acceptable minimum functionality.

[0344] It is possible to edit the contact details associated with the Speed dial entry and or create a new contact entry for the Speed dial page. It is possible to copy an entry from the contacts, directories or call log screens onto the Speed Dial page. It is possible copy an entry from the Speed Dial page to the contacts or Directory screens. It is possible to delete a Speed dial entry, or to move that entry to another contacts page. (i.e. copy and then delete original).

[0345] It is possible to control the placing of users on the Speed Dial page. It should also be possible in some manner (color coding) to distinguish between different classes of Speed Dial users, i.e. business, family, colleagues, vendors, customers. The speed dial page may well contain names from multiple other categories in the contacts information. Some form of automatic organization is available, for example, last name first name company or by class followed by last name first name company etc.

[0346] It is possible to define a group of users as a single speed dial entry. It is acceptable if the group size is limited to the size of the maximum conference call. It is possible to select the Directories view from the Speed Dial page. The Directories view will occupy the same screen area as the Speed Dial page. It is possible select from the range of on-line directories to which videophone **15** has access. The default will be the Outlook and or Lotus Notes directory that contains the users main contact details. The name of the selected directory should be displayed.

[0347] The categories established by the user in his Outlook or Notes contacts list is available as selections. If the number of categories do not fit in the display **54** area, buttons are provided to scroll either up or down the list. The list should be organized alphabetically.

[0348] The Speed Dial category is the category used to populate the Speed Dial page. There is some indication on when the Speed dial page is full and it no longer possible to add further names to this contacts category, unless they replace an existing entry. The ability to order Speed dial entries in order of most recent call, i.e. the least used Speed Dial entry would be at the bottom. This would be used to see which entry was best candidate for deletion to allow a more used number to be entered.

[0349] It is possible to easily find and select an entry from the selected category, with the minimum of user input. The entry selection mechanisms must work for relatively short lists and for very long lists (10,000's of names). The mechanisms must include the ability to enter a text string on which to search. It is possible to select the sort order for the presented data, by last name, first name or organization. There is a method of correcting entry errors, and quickly re-starting the whole search.

[0350] It is desirable if each order of the search keys was significant and could be changed by the user. In other words for example pressing and holding the left most search key enables the user to select to search on Last Name, First Name or Company (or an extended list of attributes. This is useful for example for finding someone in a particular department, or at a particular location—"who is in Korea"). The second key then qualifies the first key search and so on. Thus, the keys are set Company, Last Name First Name; say Marconi, then do an alphabetic user search within last names at Marconi. Clearly when each sort category is selected there is some implied sub-ordering of entries with the same value in that category field. So for last name selected, the implied sub-order is first name then company, for company the implied sort order is last name first name, and for first name, say last name company.

[0351] The call log screen displays the most recent entries of three categories of calls, outgoing, incoming, and missed calls, with a clear indication of which category is selected. In addition there should be a "frequent" category, that lists numbers by the frequency of use, over the last (<200) calls of any type. There should be access to the Dial Pad from the call log screen. The analysis of the value of providing a far greater degree of handling call log data is deferred.

[0352] At minimum, when the "message" is touched a connection is made to the users voice mail system **10**, the voice mail for this user is entered and the dial-pad is displayed to control the voice mail using the conventional phone key presses. The larger part of the "voice-mail" screen should bring up buttons to access each feature of the mail system **10**, for example Next Message, Previous Message,

Play Message, Forward Message, Reply to Message, call sender, etc. with all the equivalents of key presses within each function e.g. start recording stop recording review recording delete recording etc. All the functions need to be on buttons, converted to the respective DMF tones.

[0353] It is desirable that the "Forward to" number or any voice mail command that requires a list of users numbers to be entered can be selected from the Speed Dial or Directory views and that selection automatically inserts just the appropriate part of the users number. This could be particularly useful in forwarding a voice message to a group. It is possible for the user to set the time and date on the videophone **15**. It is desirable that the time and date can be set automatically by appropriate network **40** services.

[0354] It is desirable that Calendar functionality is available that is integrated with the users Outlook/Palm/Notes Schedule/Calendar application. The minimum requirement would be simply to view the appointments at any date, by day, week or month (as per Outlook or Palm screens) with changes and new entries only possible via the Outlook or Palm database.

[0355] It is likely that quite a few of the users will not maintain their own calendars and indeed may NOT have PCs **68** on their desk, but do need to view the information. Touching the User Status area of the status part of the screen allows a user to set his status. The user will have a range of Status options to choose from, including:

[0356] i) Available

[0357] ii) Busy—on a call where another call will not be accepted

[0358] iii) Do not disturb—not on a call but not interruptable

[0359] iv) Back in five minutes

[0360] v) Out of the office

[0361] vi) On Holiday

[0362] A single call instance on the videophone **15** terminal supports from one incoming stream to the maximum number of streams in a conference. For Video conferencing, the Terminal will support at least four connections to other parties as part of a single conference call. It is possible to accept at least two independent audio only calls, even when a maximum size video conference call is present, so that an audio call can be consultation hold transferred. The videophone **15** is able to support at least three simultaneous "call instances", i.e. up to three independent calls. Only one call can be active, i.e. the call controls can be applied only to one call at a time. More than one call can be accepted, i.e. users audio and video are being transmitted on each accepted call, whether active or not. Calls in progress may also be placed on HOLD, when the users audio and video is not transmitted to the user on HOLD and the audio and video from that user is also suppressed.

[0363] Incoming calls status is shown in Control display **54** area. Calls themselves and in-call controls are shown in the main section of the display **54**.

[0364] Call states are:

[0365] i) Incoming call

[0366] ii) Accepted and active—the user's audio (and video if a video call) are, subject to the various mute controls, connected to this call. Call controls apply to this call.

[0367] iii) Accepted and not active—as above, but the call controls do not apply to this call.

[0368] iv) Accepted and on hold—users audio (and video if a video call) are not being transmitted to this call.

[0369] v) Accepted and being transferred

[0370] Call states are indicated on each call. Only one accepted call can be active. An accepted call is made active by touching in the call display **54** area associated with that call, or the call status in the control panel. Any previous active call is set not active. A second touch will turn off the active state. An incoming call indication indicates if the call is offering a video connection. No indication implies an audio only call. The incoming call indication will show the name(s) of the parties associated with that incoming call. This shows immediately if the user is being called one on one, or being invited to join a conference.

[0371] The user has the following options to handle an incoming call:

[0372] i) Accept the call as a voice only call

[0373] ii) Accept the call as a video call (voice is implied)

[0374] iii) Send to voice mail

[0375] A setting is available to set the videophone **15** terminal to auto-answers incoming calls, up to the maximum number of supported calls. Auto-answer creates an audio and video connection if one is offered. Once a call is in progress, the Users status should be automatically changed to "In a call". The Users status will revert back to its previous state (typically "Available") once no calls are active.

[0376] The user is able to configure if call user data is also distributed. If the user already has one or more calls accepted and if all calls are on HOLD or not active, this call will create a new call instance if accepted. All the accepted but not active calls will continue to see and hear the user as he deals with this new call. If one of the accepted calls is accepted and active, the new call will be joined to that call and all parties to the call will be conferenced to the new caller, if the call is accepted.

[0377] If the user does not pick up after (>10) seconds, the call will automatically be forwarded as determined by the "Forward on No Answer" settings. As above the forwarding is specific to the user to whom the call is addressed. If the users status is marked "Do not disturb" or "Busy" or the "Busy" state has been set by there being the maximum number of calls being handled, the call is forwarded "immediately" as determined by the "Forward on Busy" and "Forward on Do not disturb" settings, as modified by the "show forwarded calls" setting if implemented.

[0378] Depending on the "show forwarded calls" settings, the user can chose to see the incoming call indication for (>5 seconds) before it is forwarded. (This means the user needs to take no action unless he wishes to pick up the call, rather than the positive action required on a call above.) This does not function if the Busy state is due to the videophone **15** already handling the maximum number of calls.

[0379] The ability to generate a (very short) text message that is sent with the call is a useful way of conveying more information about the importance of the call and how long it will take. The requirements associated with generating and adding a message to an outgoing call are dealt with below. If present, the incoming call text message should be displayed associated with the incoming call. The display **54** copes with the display of text messages on multiple incom-

ing calls simultaneously. The text message is also stored in the incoming or missed call log.

[0380] Call parameter negotiation is limited to that needed to establish the call within the network **40** policy parameters and the current network **40** usage. Settings are provided to allow the user to specify his preference for calls to other videophone **15** terminals, for example always offer Video, never offer video, ask each call if I want to offer video or not.

[0381] Camp on Available is supported for calls to other videophone **15** users. This will initiate a call to the user once his status changes to "available". If the user to be called is a group, the calls will only be initiated once all members of the group are 'Available'.

[0382] A conference call is when one location in the Speed Dial or Directories list represents a group of people, each of which are to be participants in a call. The suggested process of implementing this feature is to make each call in turn and once active request confirmation that the call should be added to the conference. This gives an escape route if the call goes through to voice mail. Once the actions on the first caller are completed, i.e. in the call or rejected the next number is processed.

[0383] It is possible to create an outgoing call that is half-duplex, in other words that requests audio and or video from the called party, but does not transmit either on this type of call. This is pull mode. Equally, it is possible to create a push mode, where the outgoing call does send audio and or video, but does not require any audio or video back. This mode may be used to selectively broadcast content to unattended terminals, or terminals with users playing only a passive role in the conference.

[0384] The overall volume of the speakers **64**, the handset and the headset are independently adjusting. The speaker can be turned ON and OFF. Turning the speaker off will also turn off the microphone. Status indicators show the status of the Speaker and Microphone.

[0385] The microphone can be turned off and turned back on. Status indicators show the status of the microphone mute.

[0386] The camera can be turned off and turned back on. Status indicators show the status of the camera mute.

[0387] In call controls work only on the active call. An accepted call is made active if it is not active, either by touching the call in progress status indicator in the control panel, or anywhere in the call display **54** area except for the specific in-call control function areas. Any other currently active call is turned in-active. The active call can be turned in-active by a subsequent press in the same area. A control is provided that hangs up the active call. In a conference call it clears all elements of the call instance.

[0388] A call must be accepted and active for the Conference control to function. Touching the Conference control will join the currently active call instance to the next call made active. Conference control will indicate it is active either until it is pressed again, making it inactive, or another call instance is made active. After all the calls in the now active call are joined to the Conferenced call instance the call becomes a single conferenced call and the Conference control active indication goes out. Just to re-state, conference selects the call to which other calls will be joined and then selects the call to join to that call.

[0389] The method of terminating one party to a conference call is for that party to hang-up. For a variety of reasons, the user may wish to have independent control on each part of a call instance. This can be achieved by a de-conference capability. For example, by touching the call instance for longer than three seconds, a sub-menu appears that allows the individual members of the call instance be identified and selected for de-conferencing. This call is then removed from the conference and established as a separate call instance, where all the normal controls apply, specifically it can be cleared.

[0390] The transfer function transfers the active call. When the transfer control is touched, the integrated dialing screen is displayed and the active call is placed on hold, but indicating that it is engaged in an in-call operation. The Transfer control indicates it is active, until it is pressed a second time, canceling the Transfer, or until the user selects and presses dial on the number to which he wishes the call to be transferred.

[0391] Once the outgoing call has been initiated, the Transfer control indicates a change of state, so that touching the control cause a 'blind' transfer and the call instance is removed from the screen. Alternatively, the user can wait until the called number answers, at which point a new call instance is created, allowing the user to talk to the called party, and the Transfer function changes state again, to indicate that pressing it again will complete the transfer and terminate both calls. Otherwise, the requirement is to go back to talking to the caller being transferred and re-start the transfer process or terminate the call. Transfer is the main mechanism whereby an 'admin' sets up a call and then transfer it to the 'boss'. In this case, it is essential that it is not possible for the admin to continue to 'listen into' the transferred call. This will be especially true in a secure environment.

[0392] The active call can be placed on HOLD by touching the HOLD control. In HOLD, the outgoing video and audio streams are suspended and an indication given to the remote end it is on HOLD The incoming audio and video streams are no longer displayed. The HOLD state is indicated on the call status display **54** on the control bar. The Hold control indicates hold is active if any call is on hold. Pressing HOLD again when the active call is in HOLD removes the HOLD and returns the call to the displayed state.

[0393] There is a control on the main control panel that brings up the home screen and gives access to all the other non-call functions. There is an indication that Main has been selected. Pressing Main a second time re-establishes the current call displays and de-selects Main. Separate controls are provided for each accepted and displayed party within a call, and for each call displayed. Adjusting the volume of the audio from each particular user is required. It is possible to individually mute audio and or video of each user displayed on the screen. There is a status indicator to indicate if audio or video mute is ON.

[0394] If more than one call instance can be displayed at any one time, for example, a conference call with two others, plus a new call to one other user, then it is possible to mute audio and or video for a complete call instance, for example mute the two party conference for audio, whilst speaking to the second call.

[0395] Requesting video on an audio only connection that could support video is provided. Accepting or rejecting a video request is provided. A video connection is established if the connection is agreed. A settings page item enables the user to always accept or always reject video requests.

[0396] It is possible to display the bearer channel parameters for each connection, i.e. the incoming and outgoing encoding rates for video if present and audio. In a call, controls work only on the active call. An accepted call is made active if it is not active.

[0397] It is possible to enable a 'bearer channel quality monitor' for any user. This monitor, a bit like a signal strength meter on a mobile, would show, for example, 100% Green bar when there were no errors or lost packets on the audio and video channels, a yellow bar once loss rate or the latency exceeds a predetermined rate and a red bar once it exceeds a higher rate. The time integral should be short, say 50 milliseconds as errors in this timeframe will affect the users video. So, for example, if the receiver sees video artifacts, but at the same time sees the monitor bar move yellow or red, he knows it is network **40** congestion induced.

[0398] Requesting a change in video encoding parameters, i.e. increase or decrease encoding rate, within the call is provided. Accepting or rejecting this request and a method of changing the outgoing video rate is provided. The videophone **15** generates a single outgoing encoding rate to all participants. It is possible for it to accept different incoming rates on all of the incoming streams.

[0399] A request for a side-bar with the ability to accept or reject the request is provided. If accepted, sidebar turns off the audio stream from both participants to everyone else, so they can have a private conversation, whilst continuing to hear all the discussion and continue to see and be seen by all the participants. The ability to send short messages both ways with the video and sidebar requests is provided.

[0400] Irrespective of whether the call is an incoming or outgoing call, the screen transition to the video view should be smooth. The audio may anticipate the video. The video should not be displayed until this transition can be made. (i.e. there should be no jumpy pictures, half formed frames etc in the transition to the video.) The transition to the user display **54** video screen should only start after the call is "in progress" and not at the time of initiating the call. The display of the video from the user should make maximum use of the area of the display **54** allocated to user display **54**. An in display **54** control is able to convert this single call instance single user display **54** to a full screen display **54**. Touching anywhere inside the "full screen" display **54** will revert to the standard display **54**. In addition to the in call controls already mentioned, the users name should be displayed. The display **54** and the call instance on the control panel must indicate if the call is active or not, i.e. if the in call general controls will operate or not. With one call instance up, active inactive is by pressing on the call instance or anywhere on the main display **54** apart from the in call specific control areas.

[0401] The transition from a one call instance two party call should be smooth and should be initiated once the second call is "in progress". The display **54** should make maximum use of the display **54** area allocated to user display **54**. If necessary, the videos can be clipped at each edge, rather than scaled, to fit the available area. There is no requirement for a full screen display **54** for two or more up. In addition to the in call controls already mentioned, the user name should be displayed for each party. There must be an indication that both parties are part of a single call instance. The display **54** and the call instance on the control panel must indicate if the call is active or not. The incoming video can be progressively clipped to fit the available display **54** area as more parties are added to the video call.

[0402] In two call instances both single party calls, there are two separate calls to single users, both of which are displayed. The on-screen display **54** and the call control indication clearly indicate these are two separate and independent calls and also indicate which if any is active. If either call is placed on HOLD, that call is no longer displayed and the display **54** reverts to a single call instance single call display **54**.

[0403] The user area should be capable of displaying any of the following combinations in addition to those described above.

[0404] Four call instances each single party calls;

[0405] Three call instances where one call can be two party and the others are single party calls;

[0406] Two call instances where one can be up to three party or two can be two party call.

[0407] The requirements of a "CNN" style display **54** are those of the single call instance single call above, including the ability to have a full screen display **54**. It is also possible to display "CNN" style call in half the screen and use the other half for one or two user display areas, the latter as two independent call instances or a single two party call instance.

[0408] The ability to provide various levels of encryption for the voice and data streams is provided. Access to diagnostic, test, measurement and management facilities shall make use of SMF (simple management framework), in other words access will be possible to all facilities in three ways, via SNMP, via the web and via a craft interface. The videophone **15** terminal must be remotely manageable, requiring no on site IT expertise for every day operation, or for software upgrades that do bug fixes. Fault diagnosis is also possible remotely and be able to determine if the problem is with the unit hardware, the units configuration, the units software, the network **40** or the network **40** services. Management can assume IP connectivity, but must assume a relatively low bandwidth connection to the videophone **15**.

[0409] Under normal operation, the videophone **15** should perform a shortened version of hardware system **10** test as it powers up. If this fails, the videophone **15** should display a boot failure message on the main screen. The terminal can be forced into an extended hardware diagnostic mode. This could be by attaching a keyboard to a USP port, or by pressing in the top right hand corner of the touch screen **74** as the unit powers up. This mode would give access to the underlying operating system **10** and more powerful diagnostics, to determine if the there is a hardware failure or not.

[0410] A series of simple tests can be included that the user can run in the event that the videophone **15** passes the boot-up test but is not providing the correct functionality for the user. The terminal provides a technical interface, in association with a local keyboard (and mouse) to assist in diagnosing unit or system **10** problems. This would give access to the various diagnostics for audio and video, etc.

[0411] It is possible to download safely new versions of the videophone **15** terminal software under remote control. By safely, it means being able to revert to the previous version if faults occur in the downloaded version, without local intervention (i.e. someone having to install a CD). It is possible to read the software version number of the software on a particular videophone **15** terminal, and the units hardware serial number, assembly revision number and the serial

number and assembly revision number of key sub-assemblies via the management interfaces. In the event of a system **10** crash, the videophone **15** should store or have stored information to assist in the diagnosis of the cause of that crash. This information must be retrievable on line from a remote site for analysis once the videophone **15** has rebooted.

[0412] The videophone **15** keeps a running log of all actions, events and status changes since power up, within the limits of the storage that can be allocated to this feature. It should enable at least one month's worth of activity to be stored. This data may need to be in a number of categories, for example a secure category that contains the users data, such as the numbers he called would only be releasable by the user. Generic data, such as number of calls, call state (i.e. number of call instances and endpoints per instance, encoder **36** and decoder **34** characteristics, bearer channel error reports and so on are not so sensitive information. It may be useful to be able to record every key press as a way of helping diagnose a system **10** level issue and re-create the chain of events.

[0413] It is possible for the videophone **15** to copy the exchanges at the control plane level at both the IP level and the SIP level, to a remote diagnostic terminal (the equivalent of having a line monitor remotely connected to the videophone **15** terminal). Terminal management will monitor a number of parameters, for example, network **40** quality. It must be possible to set thresholds and generate alarms when those thresholds are exceeded. Both the ATM interface and the Ethernet interface have standard measurements (rmon like, for example) that should be available for the videophone **15**. The videophone **15** should be able to send those alarms to one or more Network Management Systems.

Audio Mixer

[0414] In regard to the audio mixer, a first node **80** which can produce an audio stream and a video stream, and which is part of an ATM network having quality of service capability, wishes to form a point to point call with a second node **82**. The second node **82** only has audio capability and is, for instance, a PSTN phone. The second node **82** is not a part of the ATM network.

[0415] The first node **80** begins the formation of the call to the second node **82** by sending signaling information to an SIP server, also part of the ATM network, which identifies to the server that the second node **82** is the destination of the call that the first node **80** is initiating. The server, which already has address information concerning the second node **82**, adds the address information to the signaling information received from the first node **80**, and transmits the signaling information with the address information of the second node **82** to an audio mixer **20** that is also part of the ATM network.

[0416] When the mixer **20** receives the signaling information that has originated from the first node **80**, it determines from this information that it is the second node **82** with which the first node **80** wishes to form a connection. The mixer **20** then sends an invitation to the second node **82** through which it is somehow in communication, such as by a T1 line or ethernet but not by way of the ATM network, to identify itself in regard to its features and the form that the data needs to be provided to it so it can understand the data. In response, the second node **82** identifies to the mixer **20** the specific form the data needs to be in so that the second node **82** can understand the data, and also indicates to the mixer **20** it is OK to send data to it so the connection can be formed.

[0417] The mixer **20** then sends a signal to the first node **80** that it is ready to form the connection. To the first node **80**, the mixer **20**, which is part of the ATM network, represents the second node **82** and gives the impression to the first node **80** that the second node **82** is part of the ATM network and is similar to the first node **80**. To the second node **82**, the mixer **20**, which is also part of the network or connectivity that the second node **82** belongs, represents the first node **80** and gives the impression to the second node **82** that the first node **80** is part of the same network or connectivity to which the second node **82** belongs and is similar to the second node **82**.

[0418] The first node **80** then initiates streaming of the data, which includes audio data, and unicast packets of the data to the mixer **20**, as is well known in the art. When the mixer **20** receives the packets, it buffers the data in the packets, as is well known in the art, effectively terminating the connection in regard to the packets from the first node **80** that are destined for the second node **82**. The mixer **20**, having been informed earlier through the invitation that was sent to the second node **82**, of the form the data needs to be in so that the second node **82** can understand it, places the buffered data into the necessary format, and then subject to proper time constraints, sends the properly reformatted data effectively in a new and separate connection from the mixer **20** to the first node **80**. In this way, a point to point call is formed, although it really comprises two distinct connections, and neither the first node **80** nor the second node **82** realize that two connections are utilized to create the desired point to point call between the first node **80** in the second node **82**. Similarly, when data is sent from the second node **82** back to the first node **80**, the process is repeated, although in reverse so that after the data from the second node **82** is received by the mixer **20**, the mixer **20** reformats the data into a form that the first node **80** can understand and unicasts the data from the second node **82**, that has been buffered in the mixer **20**, to the first node **80**. If IP instead of ATM is used, then the mixer **20** sends unicast IP packets to the first node **80**, as is well known in the art.

[0419] A scenario involving conferencing, otherwise known as a point to multi point connection, will now be described using the present invention. Continuing the discussion involving a point to point connection from above, the first node **80** desires to join in the connection to form a conference, a third node **84** that is part of the ATM network and has essentially the same characteristics as the first node **80**. The first node **80** sends a signaling invitation to a host node **22** that will host the conference. The host node **22** can be the first node **80** or it can be a distinct node. The first node **80** communicates with the host node **22** through the server to form a conference and join the third node **84** into the conference. The host node **22** invites and then forms a connection for signaling purposes with the mixer **20** and causes the original signaling connection between the first node **80** and the mixer **20** to be terminated. The host node **22** also invites and forms a connection with the third node **84** in response to the request from the first node **80** for the third node **84** to be joined into the connection. In each case that a node which is part of the ATM network is to be joined into the connection, signaling goes through the server and is properly routed, as is well known in the art. The host node

22 acts as a typical host node for a conferencing connection in the ATM network. The mixer **20** represents any nodes that are not part of the ATM network, but that are to be part of the overall conferencing connection.

[0420] In regard to any of the nodes on the ATM network, the mixer **20** makes any nodes that are part of the connection but not part of the ATM network appear as though they are just like the other nodes on the ATM network. Through the signaling connections, that are formed between the host and the mixer **20**, and the mixer **20** and the second node **82** (as represented by the mixer **20**), the required information from all the nodes of the connection is provided to each of the nodes so that they can understand and communicate with all the other nodes of the connection. In fact, the host node **22** informs all the other nodes, not only the information of the characteristics of the other nodes, but also returns the information to the nodes that they had originally provided to the host node **22** so that essentially each node gets its own information back. Once this information is distributed, the streaming information is carried out as would normally be the case in any typical conferencing situation. In an ATM network scenario, the first node **80** and the third node **84** would ATM multicast using PMP tree the information in packets to each other and to the mixer **20**. In an IP environment, the first node **80** and the third node **84** would IP multicast packets to all nodes (the mixer **20** being a node for this purpose) in the network, and only those nodes which are part of the connection would understand and utilize the specific packet information that was part of the connection.

[0421] The mixer **20** receives the packets from the first node **80** and the third node **84** and buffers them, as described above. The packets from the different nodes that are received by the mixer **20** are reformatted as they are received and mixed or added together according to standard algorithms well known to one skilled in the art. At a predetermined time, as is well known in the art, the reformatted data by the mixer **20** is then transmitted to the second node **82**. In the same way, but only in reverse, the data from the second node **82** is received by the mixer **20** and buffered. It is then multicast out in a reformatted form to the first node **80** and the third node **84**.

[0422] When a fourth node, that only has audio capability, like the second node **82**, and which is not part of the ATM network, is joined into the conference, the host node **22** forms a second signaling connection with the mixer **20**. The mixer **20** in turn forms a distinct connection with the fourth node separate from the connection the mixer **20** has formed with the second node **82**. The mixer **20** maintains a list of sessions that it is supporting. In the session involving the subject conference, it identifies two cross connects through the mixer **20**. The first cross connect is through the signaling connection from the host node **22** to the second node **82**, and the second cross connect is from the host node **22** to the fourth node. In this way, the first and third nodes **80**, **84**, as well as the host node **22**, believes that there are two separate nodes, representing the second node **82** and the fourth node, to which they are communicating. In fact, the mixer **20** represents both the second node **82** and the fourth node and separately multicasts data from each of them to maintain this illusion, as well as the illusion the second node **82** and the fourth node are like the first node **80** and the third node **84**, to the first node **80** and the third node **84**.

[0423] The ViPr system is a highly advanced videoconferencing system providing 'Virtual Presence' conferencing quality that far exceeds the capabilities of any legacy videoconferencing systems on the market today. The ViPr system relies on point-to-multipoint SVCs (PMP-SVC) and IP multicast to establish point-to-multipoint audio/video media streams among conference participants. While users participating in a ViPr conference enjoy an unprecedented audio and video quality conference, there is a need to enable other non-ViPr users to join a ViPr conference. The system **10** enables a unicast voice-only telephone call (i.e. PSTN, Mobile phones and SIP phones) to be added to a multi-party ViPr conference.

[0424] The current ViPr system provides support for telephony systems through SIP-based analog and digital telephony gateways. This functionality enables ViPr users to make/receive point-to-point calls to/from telephone users. However, they do not allow a ViPr user to add a telephone call to a ViPr conference. This is due to the unicast nature of telephone calls and the inability of the telephony gateways to convert them to PMP/multicast streams. The ViPr UAM will enhance the ViPr system's support for telephony by enabling ViPr users to add unicast telephone calls to ViPr conferences.

[0425] In order to support this functionality, the ViPr UAM adds seamless conferencing functionality between the ViPr terminals and telephone users (i.e. PSTN, Mobile phones and SIP phones) by converting an upstream unicast telephone audio stream to point-to-multipoint audio streams (i.e. PMP-SVC or IP Multicast) and mixing/converting downstream PMP/multicast ViPr audio streams to unicast telephone audio streams as well as performing downstream audio transcoding of ViPr audio from the wideband 16 bit/16 KHz PCM encoding to G.711 or G.722.

[0426] An additional functionality provided by the UAM is that of an Intermedia gateway that converts IP/UDP audio streams to ATM SVC audio streams and vice-versa. This functionality enables the interoperability between ViPr systems deployed in ATM environments and SIP-based Voice-over-IP (VoIP) telephony gateways on Ethernet networks.

[0427] The UAM allows one or more ViPr phones to work with one or more phone gateways.

[0428] The UAM will support ViPr Conference calls with unicast audio devices present in following configurations:

- **[0429]** Type 1: Support one conference call with only one audio unicast device present as a participant.
- **[0430]** Type 2: Support multiple conference calls. Each conference call could potentially have multiple audio Unicast devices present as a participant.
- **[0431]** Type 3: Support multiple conference calls with each conference call having exactly one audio unicast device present as a participant.

[0432] Preferably, 20 participants (unicast devices plus ViPr phones) can be serviced by a single Unicast Manager application.

[0433] The unicast device will be used in the configuration shown in FIG. **1**.

[0434] As shown in FIG. **1**, all calls to and from a unicast device to a ViPr are always sent to the UAM. The UAM implements a B2B SIP UA to connect the unicast device to a ViPr.

Example: User A at POTS1 calls user B at ViPr V**1**. The following sequence of events takes place:

- **[0435]** 1. UD**1** (Mediatrics or whatever unicast device) receives the request from User_A to connect to User_B.

[0436] 2. UD**1** sends an INVITE to UAM. The To field or the Display Name in the INVITE identifies the call is for User_B.

[0437] 3. UAM receives INVITE as incoming call C**1**.

[0438] 4. UAM extracts the sip address of User_B from the INVITE on C**1** and initiates a call C**2** to this user by sending out an INVITE to V**1**.

[0439] 5. UAM also cross connects C**1** to C**2**.

[0440] 6. V**1** sees an incoming INVITE from UAM, which is identified by the SDP as a ViPr class device. Thus software on V**1** knows that the peer software is capable of supporting all the functionality expected of a ViPr device including Replaces/Refers etc.

[0441] 7. Say User_B at V**1** replies back to INVITE with OK.

[0442] 8. The UAM will mark the connection C**2** as up. It then sends OK on C**1**.

Media Streams in this Example

[0443] The media streams between V**1** and UD**1** are sent in either of following ways:

[0444] 1. The media is sent directly from V**1** to UD**1**. This can be done by UAM writing the right SDP. Thus while sending INVITE to V**1** it puts the IP address, port of UD**1** for receive. And while sending OK to UD**1** it puts the IP address, port of V**1** as receive address.

[0445] 2. The media is relayed by UAM. In this case, UAM relays data from V**1** to UD**1** and vice-a-versa. It is easy to see that if UAM and ViPr communicate are connected via an ATM cloud, then an SVC between V**1** and UAM could be set up. Thus, the UAM acts as an ATM to Ethernet gateway for media traffic.

[0446] Extending the example 1 further, User_A decides to join User_B at V**2** into the conference. The following events happen:

[0447] 1. The Sip connection between UAM and V**1** is replaced by A conference call C**3** with V**1**, V**2** and UAM as participants. Thus, the B2B UA is now cross connecting a conference call (C**3**) with a unicast call (C**1**).

[0448] 2. UAM always relays traffic between C**3** and C**4**. Option 11 above. It mixes the traffic from V**1** and V**2** and relays it to UD**1**. It also multicasts traffic from UD**1** to V**1** and V**2**.

[0449] The functionality performed by the UAM can be broken into following components:

[0450] SIP B2B UA Unit [SBU]. This unit performs the sip signaling required to implement the B2B SIP UA.

[0451] Media Cross Connect and Mixer [MCMU].

[0452] The UAM functionality will be decided across three processes: SBU, Unicast Mixer Manager and Sip stack, as shown in FIG. **2**.

[0453] The SipServer process will implement the SIP functionality and would provide the SBU with an abstracted signaling API (Interface Ia). Interface Ia also stays unchanged.

[0454] The SBU implements the call control and glue logic for implementing the B2B UA. This unit derives from Callmanager/Vupper code base. The SBU is responsible for setting up the right mixer streams too. For this purpose, SBU interfaces with the UMM process through RPC.

[0455] UMM implements the functionality for cross-connecting media streams as well as implementing the audio mixing functionality.

[0456] The SBU implements the call control and glue logic for implementing the B2B UA. The SBU is responsible for setting up the right mixer streams too. For this purpose, SBU interfaces with the UMM process through RPC.

```
Session
Class MediaSession
{
    int         SelfID        // Self ID
    CVString    GUID          // Conference Call ID
    CVList      XIDList;      // List of cross connects
    GUID
}
SIPB2BCrossConnect
Class SIPB2BCrossConnect
{
    int         SelfID        // Self ID
    int         SessionID     // Of session of which it is a member
    Int         ViPrLegID     // SiPCallLeg connected to ViPr
    Int         UDLegID       // Leg connected to unicast device.
}
SIPB2BCallLeg
Class SIPB2BCrossConnect
{
    int         SelfID        // Self ID - returned by callmanager
    int         XID           // ID of Cross Connect who owns this
leg
    SipCallLeg  ViPrLeg       // Leg connected to ViPr
    SipCallLeg  UDLeg         // Leg connected to unicast device.
}
```

[0457] The SBU unit is internally structured as follows:

[0458] As can be seen from FIG. **3**, the design for SBU reuses and extend the SIP/Media Stream interface offered by the CallManager to implement the signaling call control logic for UAM.

[0459] The following text presents the flow of control when the user A initiates a call to User_B.

[0460] In the following SipServer refers to SipServer at UAM, SBU refers to SBU at UAM and UMM refers to UMM at UAM.

[0461] To clarify the example further, assume the following:

[0462] The entire network is Ethernet network

[0463] IP address of V**1** is 172.19.64.101

[0464] IP address of V**2** I 172.19.64.101

[0465] IP address of interface of UAM which is connected to V**1**/V**2** cloud is 172.19.64.51, IP interface of UAM connected to UD**1** cloud is 169.144.50.100

[0466] IP address of UD**1** is 169.144.50.48

[0467] Address is represented as <IpAddress, port> tuple

[0468] All the addresses and ports in the example are illustrative, they are not required to be fixed but are rather allocated by OS.

[0469] In the following example, all the SIP events received by SBU (at UAM) are actually received by SipServer and than passed to SBU. However, the Sipserver receiving the event and passing it to SBU is not shown for brevity.

Flow of control for a P2P call between UD1 and V1

#	Loc	Action
1	UD1	INVITE sent from UD1 to SD1. This invite contains the Address <169.144.50.48, 50000> for receiving stream from UD1 for this call.
2	SBU	SBU gets an incoming call C1. SBU examines the call and sees it is from a Unicast device. It then performs the following actions. Extracts the address (User_B) of final destination UD1 is trying to reach. It allocates address <172.19.64.51, 40002> for receiving media stream from V1. It initiates an outgoing call (C2) to User_B by asking sipserver to send an INVITE to User_B. This invite contains the address <172.19.64.51, 40002>. It also allocates a sip cross connect (XID = 1) and binds C1 and C2 to XID = 1. At this point sip cross connect XID = 1 C1 and C2 as a back-to-back calls. It also stores XID = 1 in the calls C1 and C2. This is to enable retrieving XID from Call ID.
3	V1	V1 receives an incoming INVITE and accepts the call by sending an OK to UAM. The OK contains address <172.19.64.101, 10002> for receiving traffic from UAM.
4	SBU	SBU Gets OK (call accept event) on C2. It the performs following steps: Receives the cross connect (XID = 1) of which C2 is a member. Allocates an address for use of C2. <169.144.50.100, 40001> Instructs SipServer to send OK On call C2. This OK contains address <1169.144.50.100, 40001> for receiving media from UD1. Allocates a Session with ID (say, SID = 100). This session ID is stored in Sip Cross connect XID = 1. The SipCross connect with XID = 1 is also added to the list of Cross-connects part of this session. At this time, there is just one SIP cross connect in the list. SBU then allocates a media channel to be used for receiving and sending data from UD1, say with CHID = 0. SBU allocates a media channel to be used for sending and receiving data from V1, say CHID-1. SBU then informs UMM to setup channels for sending and receiving data from V1 and UD1 as follows: SBU informs UMM that channel = 0 should be used to send/receive data to/from UD1. This is done by asking UMM to associate channel = 0 with send address <169.144.50.48, 50000> and Receive address <169.144.50.100, 40001>. SBU informs UMM that channel = 1 should be used to send/receive data to/from V1. This is done by asking UMM to associate channel = 0 with send address <172.19.64.101, 10001> and Receive address <172.19.64.51, 40002>. SBU then instructs the UMM to construct a media cross connect by informing UMM that Channels CID = 0 and CID = 1 are part of same session SID = 100. It should be noted that UMM is not informed (nor does it care) about the SIP calls C1 and C2.
5	UD1	Receives an OK from UAM. It knows from OK that for sending audio media to UAM it must use the address <169.144.50.100, 40001>.

[0470] The above table explains what happens for a pass through call. The following is the control flow when this call is converted into a conference call. In this case, say User B conferences User_C at V**2** into the call.

[0471] Further assume the following:

[0472] IP address of V**2** is 171.19.64.102

Initiating a conference with a user on unicast device.

#	Loc	Action
6	V1	V1 # Sends an INVITE to Conference Host H (at V1) to initiate conference. The INVITE contains the multicast IP address <239.192.64.101, 10002> on which V1 would multicast its audio stream.

-continued

Initiating a conference with a user on unicast device.

#	Loc	Action
7	H	Host Gets an INVITE to start a conference call. It sends an OK back to V1. H also constructs a globally unique ID for this conference call. (say, GUID = 900).
8	V1	Refers UAM into the conference (with Replaces = C2).
9	H	Sends an INVITE to UAM with following information: GUID = 900 Replaces = C1 Stream information for V1 (User_B) <239.192.64.101, 10002>
10	SBU	On getting Invite for a conference call (C3) SBU performs following: Sees that Replace ID = C2. It thus knows that V1 wants to bring POTS1 (UD1) into Conference GUID = 100. It Retrieves the SIP Cross-connect XID = 1 from C2. It retrieves the Session ID from the SipCross Connect, SID = 100. And sets the GUID member of the Session to GUID = 900. It Sets the GUID in Sip Cross-connect XID = 1 to GUID = 100. It releases the sip connection C2 by informing SipServer to send a Bye on C2. Removes C2 from SIP Cross-connect XID = 1 and replaces it with C3. It also sets the SIP cross connect ID in C3 to XID = 1. It also sets the XID member within C3 to point to XID = 1. It allocates address <239.192.64.51, 40003> for transmitting data on behalf of UD1. It informs UMM to delete channel CID = 1. Thus UMM will now stop transmitting media to address <172.19.64.101, 10001> and stop receiving media at address <172.19.64.51, 40002>. It sends an OK back to the Host. The OK contains information that everyone on the conference should send receive media streams from POTS1 (UD1) on address <239.192.64.51, 40003>. SBU then instructs UMM to set up the right audio streams for conference (GUID = 900) with V1 and UD1 present as participants as follows: SBU informs that channel = 2 should be used to send/receive data to/from V1. Thus channel = 2 is associated with send address <239.192.64.51, 40003> and Receive address <239.192.64.101, 10002>. SBU informs UMM to associate channel = 2 with Session SID = 100. SBU informs the UMM to set the retransmit address field for channel = 0 <239.192.64.51, 40003>. It should again be noted that UMM is not aware of either the presence of SIP calls C1 and C3, nor does not it know that there is a conference call with GUID = 900. Internally, UMM does not really look at the send address in channel = 2 to relay data from UD1 to conference. Rather, it looks at the retransmit address in the Channel ID = 2.
11	Host	Gets OK from UAMD. It sends a RE_INVITE to V1 indicating the presence of stream from User_A at <239.192.64.51, 40003>.
12	V1	Refers User_C at V2 into the conference.
13	H	Sends an INVITE to V2 indicating presence of streams from User_A at and User_B.
14	V2	V2 sends an OK. The OK contains the multicast IP address <239.192.64.102, 20001> on which V1 would multicast its audio stream. At this point, User_C can start listening to audio from User_A and User_B by registering to appropriate multicast addresses.
15	H	Sends a RE_INVITE to V1 and UAMD indicating presence of a new participant User_C sending audio at <239.192.64.102, 20001>.
16	V1	Gets a RE_INVITE and sees that party User_C is now on the call. It sends an OK back to H.
17	SBU	Gets a RE_INVITE and sees that a new party User_C is also on conference call with GUID = 900. It then performs following steps: Sends an OK back to the Host through sip server. Allocates a media channel CID = 3 for receiving traffic from User_C. Informs UMM to join media from User_C into the conference call identified by GUID = 900 as follows: SBU informs UMM that channel = 3 should be used to send/receive data to/from (User_C) at V2. Thus, channel = 3 is associated with send address <239.192.64.51, 40003> and Receive address <239.192.64.102, 20001>. SBU informs UMM to associate channel = 2 with Session SID = 100. It should be noted again that all UMM knows is that there are three channels (CID = 0, 2 and 3) which all belong to the same session. UMM knows that CID = 2 and 3 are streams from ViPr phone and CID = 0 are from a unicast device. Thus, UMM reads multicast data from channels CID = 2 (<239.192.64.102, 20001> and CID = 3 (<239.192.64.101, 10002>) mixes them

-continued

Initiating a conference with a user on unicast device.		
#	Loc	Action
		and sends it on channel = 0<169.144.50.48, 50000>. Also the data read from channel CID = 0, is retransmitted on retransmit address associated with CID = 0 <239.192.64.51, 40003>. The details of how UMM performs this appropriate mixing are in a different section.
18	H	Gets the OK for re-invites sent in step 16. The conference call is now up.

[0473] To add another ViPr user to the conference, steps 12 through 18 are repeated. Consider the steps that are required to another Unicast Device user say User_D on POTS2.

[0474] Assume the following:

[0475] User_C on ViPr V**2** decides to conference in User_D on POTS2 into the conference.

Flow of control for adding second unicast user to a conference.		
#	Loc	Action
19	V2	Refers User__D at POTS2 into the conference.
20	H	Sends an INVITE to UAM with following information: User__A, User__B and User__C call along with the addresses on which they are generating media streams. GUID = 900
21	SBU	Gets Request for an incoming conference call (C4) with GUID = 900 To address = Address of User__D It then performs following tasks: It allocates a SIP Cross-connect with ID, XID = 2. It adds C4 to the sip cross connect XID = 2. It also sets the XID member within C4 to XID = 2. It searches all the Session structures to see if there is a session with GUID = 900. It finds that a session with ID = 100 is associated with this conference call. It then adds SIP cross connect with XID = 2, to the list of cross connects attached to Session SID = 100. At this point there are two SIP cross connects (XID = 1, and XID = 2) which are part of the SIP session SID = 100. It also stores information within sip cross connect XID = 2, to indicate it is associated with Session = 100. It allocates an address <169.144.50.51, 40011> for receiving traffic from User__D. It allocates a media channel CHID = 4 for receiving traffic from User__D <239.192.64.51, 40012>. It initiates a connection C5 by sending an INVITE to UD1 for User__D. The INVITE contains the information that UD1 should send audio media streams for this call at <169.144.50.51, 40004>. It adds C5 to the sip cross connect of XID = 2. Thus XID = 2 is now connecting CID = 4 and CID = 5 as back to back SIP calls. It also sets XID member of C5 to XID = 2.
22	UD1	Receives INVITE from UAM and sends back an OK to UAM. It indicates in the OK that the address on which it should be sent data for call C5 is <169.144.50.48, 50002>.
23	SBU	Receives OK from UAM for C5. It then performs following steps: It retrieves the sip cross connect of which C5 is a member, XID = 2. It retrieves the session from sip cross connect, SID = 100. It then allocates an address <239.192.64.51, 40012> to relay data received on User__D into the conference, GUID = 900. It then sends an OK to Host indicating that User__D would be generating traffic on <239.192.64.51, 40012>. It then allocates channels for receiving traffic User__A (CHID = 5), User__B (CHID = 6) and (CHID = 7). It then asks UMM to add User__D into the conference as follows: SBU informs UMM that channel = 4 should be used to send/receive data to/from User__D. Thus channel = 3 is associated with send address <169.144.50.51, 40011> and Receive address <169.144.50.48, 50002>. SBU also informs UMM to set the retransmit address of CHID = 4 to <239.192.64.51, 40012>.

-continued

Flow of control for adding second unicast user to a conference.

#	Loc	Action
		SBU informs UMM that Channel = 5, 6 and 7 should be used to exchange traffic with User_A, User_B and User_C. The following information is provided for these channels. CHID = 5 [Rx = <239.192.64.102, 20001>, Tx = <239.192.64.51, 40012> CHID = 6 [Rx = <239.192.64.101, 10001>, Tx = <239.192.64.51, 40012> CHID = 7 [Rx = <<239.192.64.51, 40012>, Tx = <239.192.64.51, 40012> SBU informs UMM to associate channel = 4, 5, 6, 7 with Session SID = 100 {Please note that CHID = 5 the information for receiving packets from User_A is same as one present in CHID = 2 and would seem like a waste and troublesome but this has in fact has a desirable effect of not requiring any change in call manager and also eliminates needs for book keeping in SBU. Same holds for CHID = 3 and CHID = 6. The UMM would never receive anything on CHID = 7 because multicasts are not received by the host which transmitted them.} In the UMM there are two channels CHID = 2 and 5 which are referring to the same receive multicast address, now since both the channels belong to the same session = 100, it is not a problem. Since the UMM will not read packets from duplicate channels. However, if Channel = 2 is deleted then UMM will go and read packets from CHID = 5.
24	H	Host receives the OK on C5 (from UAM) with information added to receive audio streams from User_D. It Sends a Re-Invite to User_A, User_B and User_C indicating presence of a new stream from User_D.
25	SBU	Gets a REINVITE on C3 indicating presence of another user User_D transmitting on multicast address <239.192.64.51, 40012> It then performs following tasks: Sends an OK back to host on C3 through sip server. It retrieves the sip cross connect of which C3 is a member, XID = 1. It retrieves the session SID = 100 from sip cross connect XID = 1 It allocates channel CHID = 8 to receive audio from the User_D. It then instructs UMM to receive and mix traffic from User_D into the Session SID = 100. as follows: SBU informs UMM that channel = 8 should be used to send/receive data to/from User_D. Thus channel = 8 is associated with send address and Receive address <239.192.64.51, 40012>. SBU also sets the session ID for channel CHID = 8 to SID = 100. [NOTE: Since UAMD programs the IP sockets to never receive packets it has transmitted on a multicast address, no traffic would be received on CHID = 8. Which is exactly what is desired.].
26	V1 and V2	Sends an OK to re-invite sent by Host
27	H	Receives OK from all the participants, the conference call now has 4 parties on call. Two of which are unicast devices.

[0476] UMM implements the functionality for cross-connecting media streams as well as implementing the audio mixing functionality.

Deployment Scenario 1:

[0477] Referring to FIG. **4**, this scenario covers two cases:

[0478] A ViPr user in a multi-party ViPr audio/video conference adding a unicast audio-only telephone user to the conference:

[0479] In this case, ViPr users in multi-party ViPr conference decide to add a unicast telephone user to the conference. As a result, one of the participants initiates a call to the destination telephone number. The ViPr SIP server redirects the call to the ViPr UAM. The ViPr UAM terminates the ViPr audio-only call and establishes a back-to-back call to the destination telephone via the telephony gateway.

[0480] Once the call is established, the ViPr UAM converts the unicast G.711/G.722 audio stream received from the telephone into a PMP/multicast stream and forwards it to the ViPr terminals without any transcoding. On the other hand, the ViPr UAM performs transcoding and mixing of the wideband 16 bit/16 KHz PCM ViPr audio streams received from the various ViPr terminals into one G.711 or G.722 unicast audio stream and forwards it to the telephone destination.

[0481] A ViPr user in point-to-point audio-only conference with a telephone user adding another ViPr user to the conference:

[0482] In this case, a ViPr user (V**1**) in point-to-point audio-only call with a telephone user (T) decides to add another ViPr user (V**2**) to the conference. As a result, the ViPr user V**1** initiates an audio/video call to the destination ViPr user V**2**. The ViPr system tears down the established

point-to-point call between V**1** and the ViPr UAM and re-establishes a PMP/multicast call between V**1**, V**2** and the ViPr UAM.

[0483] The ViPr UAM terminates the new ViPr audio/video call and bridges it to the already established back-to-back telephone call. Throughout this process, the telephone call remains active and the switching is transparent to the telephone user.

[0484] Once the call is established, the ViPr UAM converts the unicast G.711/G.722 audio stream received from the telephone into a PMP/multicast stream and forwards it to the ViPr terminals without any transcoding. On the other hand, the ViPr UAM performs transcoding and mixing of the wideband 16 bit/16 KHz PCM ViPr audio streams received from the various ViPr terminals into one G.711 or G.722 unicast audio stream and forwards it to the telephone destination.

[0485] ViPr uses Session Initiation Protocol (SIP) as a means of establishing, modifying and clearing multi-stream multi-media sessions. The UAM will add conferencing capabilities between the ViPr terminals and telephone users (i.e. PSTN, Mobile phones and SIP phones) by converting upstream unicast voice-only telephone streams into point-to-multipoint streams (i.e. PMP-SVC or IP Multicast) and converting downstream ViPr multicast/PMP audio streams to unicast telephone voice-only streams as well as performing downstream audio transcoding of ViPr audio from wideband 16 bit/16 KHz PCM encoding to G.711 or G.722.

Scenario 2:

[0486] Referring to FIG. **5**, this scenario covers two cases:

A telephone User Calling a ViPr User:

[0487] In this case, a telephone user initiates a call (audio only) to a ViPr user. The telephony gateway redirects the call to the ViPr UAM. The ViPr UAM terminates the telephone call and establishes a back-to-back ViPr audio-only call to the destination ViPr terminal.

[0488] Once the call is established, the ViPr UAM forwards the G.711/G.722 audio stream received from the telephone to the ViPr terminal without any transcoding. On the other hand, the ViPr UAM performs transcoding of the ViPr audio stream from wideband 16 bit/16 KHz PCM to G.711 or G.722 and forwards it to the telephone destination.

A ViPr User Calling a Telephone User:

[0489] In this case, a ViPr user initiates a call to a telephone user. The ViPr SIP server redirects the call to the ViPr UAM. The ViPr UAM terminates the ViPr audio-only call and establishes a back-to-back PSTN call to the destination telephone via the telephony gateway. Transcoding is done in the same way as described in the previous paragraph.

[0490] FIG. **6** gives a typical usage context for UAM. The features provided by the UAM are the following.

Feature 1

[0491] Say that ViPr V**1** and V**2** are in a point-to-point call and they wish to engage Unicast Device UD**1** in a conference call. Put in other words the intent is to form a conference call with UD**1**, V**1** and V**2** in conference. Say user at V**1** requests that user at UD**1** be joined into the conference call with V**1** and V**2** as other parties. This request is forwarded by one of the SIP servers to the UAM.

UAM then Performs the Following Tasks:

[0492] It joins the conference call on behalf of UD**1**. Call this conference call C**1**.

[0493] It also makes a point-to-point call with the Unicast Device. Call this conference call C**2**.

[0494] It relays audio data received on C**2** to C**1**.

[0495] It accepts the audio data from V**1** and V**2** parties in call C**2**, mixes and forwards this data to UD.

Feature 2

[0496] Consider the case where vipr-net in the figure above is ATM and UD-net is an IP network. Also, suppose that it is desired that to the extent possible only SVCs be used over the ATM network for audio rather than LANE/CLIP. This could be for security concerns or for performance issues.

[0497] In this case, if a ViPr V**1** on vipr-net wishes to engage a unicast device (UD**1**) in an audio conversation, than UAM is used to provide functionality to use SVC in the ATM network and IP in the IP network.

[0498] To do this all call from V**1** to UD**1** is broken into two calls from V**1** to UAMD and from UAMD to V**2**.

[0499] The configuration required for features supported by UAM can be broken into following categories:

[0500] Configuration for ViPr to UD calls.

[0501] Configuration for UD to ViPr calls.

[0502] General configuration

General Configuration

[0503] The B2BUA SIP UA is made to run on any desired port (other than 5060). This is done by modifying the vipr.ini file to include following parameter:

SIP_Port=7070[any valid port number]

Configuration for ViPr to UD Calls

[0504] For a typical ViPr call when a user dials a "number" its "call-request" is sent to SIP Server which than forwards it to the appropriate destinations. However, this case is different. In this case, when a user says I wish to talk to unicast device (UD**1**) the SIP Server forwards the request to UAM. In addition, it also puts information in the request to identify that this call should be forwarded to UD**1**. Thus, the SIP Server is programmed to route calls made to the SIP-URIs serviced by the UAM devices to the appropriate UAMD Server.

[0505] It is also possible to specify a default unicast device SIP address to which to forward all calls received by the UAM. This default address can be specified in vipr.ini file by adding following lines:

UD_SERVER_ADDRESS=169.144.50.48

X_FORWARD_AVAILABLE=0

[0506] It should be noted that when a call is made from a unicast device to a ViPr, the call has to be delivered to the

UAM. To do this, appropriate configuration is performed at unicast device, please refer to unicast device specific documentation for this.

Configuration for UD to ViPr Call

[0507] The calls originating at the UD for a ViPr are routed to the UAM. One way to achieve this is by programming the UD to direct/forward all calls to UAM. Also, the eventual destination of the calls (say V**1**) is specified in the call request to UAM. Typically, this address will be the To field in the SIP message. These configurations are performed at the UD or the SIP Server.

[0508] In addition, when UAM receives a call request from a UD, it forwards it to a gateway Marshall server for performing sanity checks on the called party. This gateway address can be specified in the vipr.ini file

GatewayMarshallServer=sip.eng.fore.com:5065

LIST OF ACRONYMS

[0509] ATM Asynchronous Transfer Mode
[0510] ISDN Integrated Services Digital Network
[0511] IP Internet Protocol
[0512] LAN Local Area Network
[0513] MC Multicast (IP)
[0514] MCMU Media Cross Connect and Mixer
[0515] MCU Media Conferencing Unit
[0516] PBX Private Branch Exchange (private telephone switchboard)
[0517] PCM Pulse-Code Modulation
[0518] PMP Point-to-Multipoint (ATM)
[0519] POTS "Plain Old Telephone System"
[0520] PRI Primary Rate Interface (ISDN)
[0521] PSTN Public Switched Telephone Network
[0522] SBU SIP back-to-back user agent
[0523] SIP Session Initiation Protocol
[0524] SVC Switched Virtual Circuit (ATM)
[0525] UAM Unicast Audio Mixer
[0526] ViPr™ Virtual Presence System
[0527] WAN Wide Area Network

[0528] Although the invention has been described in detail in the foregoing embodiments for the purpose of illustration, it is to be understood that such detail is solely for that purpose and that variations can be made therein by those skilled in the art without departing from the spirit and scope of the invention except as it may be described by the following claims.

1. A teleconferencing system comprising:
a network; and
a plurality of nodes in communication with each other to form a conference of live scenes at each node, each node having a video display with a display layout, at least one of the nodes individually controlling at least partially the display layout of each node in the conference with a specific format able to be unique for each node.

2. A system as described in claim **1** wherein each node is forced to display video in the specific format.

3. A system as described in claim **2** wherein each node is locked into the specific format.

4. A system as described in claim **3** wherein each node is forced to display certain video streams from the other nodes of the conference at specific locations on the display.

5. A system as described in claim **4** wherein each node controls what is displayed on any portion of the screen not controlled by the one of the nodes.

6. A system as described in claim **5** wherein the one of the nodes completely controls the display layout of each node.

7. A method for providing a teleconference comprising the steps of:
forming a conference with a plurality of nodes in communication with each other through a network of live scenes at each node, each node having a video display with a display layout; and
controlling individually with at least one of the nodes at least partially the display layout of each node in the conference with a specific format able to be unique for each node.

8. A teleconferencing node for a network with other nodes comprising:
a network interface which communicates with the other nodes to form a conference of live scenes at each node; and
a controller individually controlling at least partially the display layout of each node in the conference with a specific format able to be unique for each node.

9. A method for conducting a telecommunications conference between at least three nodes comprising the steps of:
establishing a conference between the nodes of live scenes at each node;
making a change to the conference; and
communicating only the change to the nodes.

10. A method as described in claim **9** wherein the communicating step includes the step of communicating only the change only to the nodes affected by the change.

11. A method as described in claim **10** wherein the making step includes the step of making the change to one of the nodes' status.

12. A method as described in claim **11** wherein the making step includes the step of making the change to the conference's state.

13. A method as described in claim **12** including the step of sending a directed message from one of the nodes to only certain of but less than all of the parties.

14. A method as described in claim **13** wherein the establishing step includes the step of establishing the conference based on SIP NOTIFY/OK techniques.

15. A teleconferencing system comprising:
a network; and
a plurality of nodes in communication with each other through the network to form a conference of live scenes at each node, each node communicating only a change to the other nodes when the change occurs.

16. A system as described in claim **15** wherein each node communicates only the change only to the nodes affected by the change.

17. A system as described in claim **16** wherein the change to the conference includes a change to one of the nodes' status.

18. A system as described in claim **17** wherein the change to the conference includes a change to the conference's state.

19. A system as described in claim **18** wherein one of the nodes sends a directed message to only certain of but less than all of the nodes.

20. A system as described in claim **19** wherein the plurality of nodes establishes the conference based on SIP NOTIFY/OK techniques.

21. A system as described in claim **20** wherein each node has a controller and a Network Interface in communication with the controller and the network, the controller effecting any change to its node and sending the change through the network interface to through the network to the other nodes of the conference.

22. A teleconferencing node for a network with other nodes comprising:

a network interface which communicates with the other nodes to form a live conference of live scenes at each node; and

a controller which communicates only a change to the other nodes when the change occurs.

* * * * *