

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

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
22 July 2010 (22.07.2010)

(10) International Publication Number
WO 2010/082208 A2

PCT

(51) International Patent Classification:

H04N 5/44 (2006.01) **G06F 3/00** (2006.01)
H04N 7/173 (2006.01)

(21) International Application Number:

PCT/IN2009/000281

(22) International Filing Date:

13 May 2009 (13.05.2009)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

1029/MUM/2008	13 May 2008 (13.05.2008)	IN
1028/MUM/2008	13 May 2008 (13.05.2008)	IN
2035/MUM/2008	24 September 2008 (24.09.2008)	IN
63/MUM/2009	9 January 2009 (09.01.2009)	IN

(71) Applicant (for all designated States except US): **TATA CONSULTANCY SERVICES LIMITED** [IN/IN]; Nirmal Building, 9th Floor, Nariman Point, Mumbai 400 021, Maharashtra (IN).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **ARPAN, Pal** [IN/IN]; Bengal Intelligent Park, Bldg.-D, Plot No. A2 M2 & N2, Block EP & GP, Salt Lake Electronics Complex, Sector-V, Kolkata 700 091, West Bengal (IN). **CHIRABRATA, Bhaumik** [IN/IN]; Bengal Intelligent Park, Bldg.-D, Plot No. A2 M2 & N2, Block EP & GP, Salt Lake Electronics Complex, Sector-V, Kolkata 700 091, West Bengal (IN). **AVIK, Ghose** [IN/IN]; Bengal Intelligent Park, Bldg.-D, Plot No. A2 M2 & N2, Block EP & GP, Salt Lake Electronics Complex, Sector-V, Kolkata 700 091, West Bengal (IN). **ANIRUDDHA, Sinha** [IN/IN]; Bengal Intelligent Park, Bldg.-D, Plot No. A2 M2 & N2, Block EP & GP, Salt Lake Electronics Complex, Sector-V, Kolkata 700 091, West Bengal (IN). **JASMA, Shukla** [IN/IN]; Bengal Intelligent Park, Bldg.-D, Plot No. A2 M2 & N2, Block EP & GP, Salt Lake Electronics Complex, Sector-V, Kolkata 700 091, West Bengal (IN). **DEBNARAYAN, Kar** [IN/IN]; Bengal Intelli-

gent Park, Bldg.-D, Plot No. A2 M2 & N2, Block EP & GP, Salt Lake Electronics Complex, Sector-V, Kolkata 700 091, West Bengal (IN). **GHOSHDASTIDAR, Somnath** [IN/IN]; Bengal Intelligent Park, Bldg.-D, Plot No. A2 M2 & N2, Block EP & GP, Salt Lake Electronics Complex, Sector-V, Kolkata 700 091, West Bengal (IN).

(74) Agent: **DEWAN, Mohan**; R. K. Dewan & Company, Podar Chambers, S.A. Brelvi Road, Fort, Mumbai 400 001 Maharashtra (IN).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- as to the identity of the inventor (Rule 4.17(i))
- of inventorship (Rule 4.17(iv))

Published:

- without international search report and to be republished upon receipt of that report (Rule 48.2(g))

(54) Title: METHODS AND SYSTEMS FOR WIRELESS/WIRED TRANSMISSION

(57) Abstract: A multimedia system consisting of viewing mechanisms adapted to receive communication from a communication means, wherein said communication means is a screen-viewable input device adapted to provide full-fledged input options viewable and effected from a screen, via a remote control unit of said screen, said communication means comprising pre-defined number of keys on said remote control, said number of keys being less than the number of keys viewable on screen; on-screen viewable keys, distributed in a pre-defined format, adapted to be displayed on said display means; and correlation means adapted to correlate said remote control keys, in a pre-defined manner, with said on-screen viewable keys.



WO 2010/082208 A2

METHODS AND SYSTEMS FOR WIRELESS/WIRED TRANSMISSION

Field of the Invention:

This invention relates to methods and systems for wireless/wired transmission.

Particularly, this invention relates to apparatus and method for control, transmission and integration of content.

Still particularly, this invention relates to method and apparatus for accessing and integrating TV broadcast, Internet, Content from storage media and sensor device and enabling rich communication.

Background of the invention:

The world is witnessing phenomenal convergence in terms of content, access technology, access devices, and interaction and interfacing systems. Several categories of contents like voice, data and multimedia content can be converged on and accessed on the same device. Content from multiple applications including voice call, video call, television program, Internet, video conference, short message service and instant messages, which can be accessed using a single access device which can be desktop computer, laptop computer, mobile phone, PDA, Television or another suitable device.

As Television is widely available in many homes, it is worthwhile to use it as the display device to bring this vast range of converged services and applications cost-effectively to vast population of humanity. This can be done cost-effectively by adding a Set-top Box to the Television. Television screens are usually bigger than some of the access device screens and this makes it ideally suitable for accessing large range of applications and content, especially multimedia content.

However it has been a challenge to provide cost-effective and easy-to-use keyboard for accessing services like Internet and messaging services from Television, because the keyboard has to be wire-less in order to enable viewing from a distance. A full-fledged

separate wireless or Infra-red keyboard adds up to the cost. One way to counter this problem is to provide an On-screen keyboard on TV screen which the user can operate with a simple Infra-red remote control through the Set-top Box. However, since the infra-red remote has only the arrow keys to navigate, the available on-screen keyboards require a large number of keystrokes for typing text on the screen, which in turn, reduces the ease of use.

Multimedia processing platforms are available which are capable of relaying/providing a variety of entertainment, collaboration and information applications that allows an end-user to:

- a) view broadcast/on-demand tv content;
- b) playback/record audio-video files and view images either from local storage or from internet;
- c) collaborate/virtually connect over the IP network (video Chat, Audio Chat, Send/Receive SMS, MMS);
- d) share audio-video/image content with others;
- e) browse information over the internet;

As the seams between a monitor for use with a processing unit, a television, an internet facilitating means, an internet interaction means all begin to vanish, integrated platforms are developed which cater to the multimodal functionalities of providing all services on a single platform, thus relegating the need for different mechanisms for different applications.

For the purposes of this specification, a 'viewing mechanism' includes a television, a monitor, a display unit and the like stand-alone or processor based display means such as a computer.

The prior art also includes applications, which are being developed on a single platform that required a single user interface to tie the variety of applications together. Thus, upon seamless blending of the available broadcast/on-demand TV content (available from local

cable TV provider or satellite TV provider through a separate box or Internet TV content through IPTV) with information and media content available in local storage and Internet for doing value added interactive applications on various media such as the TV network.

It has also been another challenge to provide an effective interaction mechanism between multimedia systems, and particularly to adapt the current television sets to provide video conferencing capability in a seamless fashion i.e. to provide a channel and bandwidth independent Video conferencing system.

A particular form of video conferencing is conducted over a 1xRTT CDMA wireless channel. A major challenge in this form of video conferencing is to maintain the video and audio quality in a time varying wireless channel where the uplink and downlink bandwidth changes varies from 9.6 kbps to 56 kbps (kilo bits per second).

Now a days, digital set-top boxes (STB) are becoming popular. These boxes provide Internet browsing and multimedia experience along with viewing selected TV channels.

Detail of the existing video conferencing solution over wireless channel is provided here below:

Patent - US7225267 discloses Reactive bandwidth control for streaming data in video conferencing. This document talks about detecting the inception network congestion and appropriately taking action on the streaming data. This system uses some probe packets and the RTT to evaluate certain parameters.

Again, Patent - US7136066 discloses a system and method for scalable portrait video in video conferencing. This disclosure talks about the generation, coding and transmission of an effective video form, scalable portrait video. It mentions about the bit-rate range of 20-40 kbps in CDMX-1X channel. It also mentions about the usage in video conferencing scenario.

Patent - US7133368 discloses a Peer-to-peer method of quality of service (QoS) probing and analysis and infrastructure which is applicable in video conferencing. A peer-to-peer (P2P) probing/network quality of service (QoS) analysis system is disclosed which utilizes a UDP-based probing tool for determining latency, bandwidth, and packet loss ratio between peers in a network. The system disclosed in this document uses probe packets to get the channel condition. The system disclosed therein also discusses about probe packet structures and bandwidth calculation.

Patent -US7130268 discloses an end-to-end bandwidth estimation for congestion control in packet switching networks suitable for video conferencing. This document discloses the bandwidth estimation technique in a packet network for RTP/UDP over IP to choose the server that satisfies a client's request.

Patent - US6201791 discloses a method and apparatus for measuring the flow capacity of and determining the optimal window size of a communications network suitable for video conferencing. This disclosure illustrates a method to find the idle bandwidth of a channel by sending some packets in a window. This disclosure also shows a method to estimate the optimum window size.

Patent - US5627970 discloses methods and apparatus for achieving and maintaining optimum transmission rates and preventing data loss in a processing system network inter-alia suitable for video conferencing. The disclosure talks about achieving and maintaining data transmission rates in processing system networks, independent of communication between the node and the processing system network, and includes techniques for data transmission initialization, data retransmission, and buffer management.

Patent application US20040233844 discloses a technique for bi-level and full-color video combination for video communication. The disclosure talks about a video-encoding scheme based on the estimated channel bandwidth.

Patent application US20050111487 discloses a method and apparatus for quality of service determination suitable for use in video conferencing. The method discloses therein estimates the bandwidth capacity, available bandwidth and utilization along a path in an IP network. ICMP time-stamp requests are sent from a source host on the edge or inside the network to all routers on the end-to-end path to a desired destination.

Patent application US20070081462 discloses a Congestion level and signal quality based estimator for bit-rate and automated load balancing for wlans which is suitable for use in the video conferencing process. The disclosure therein determines the congestion of the wireless access points.

Similarly, Patent application WO2005022845A1 discloses a rate based congestion control for packet networks. The disclosure talks about a method to compute an end-to-end fashion the sending rate of data, audio and video flows over a packet switching network such as the Internet Protocol Network.

Again, Patent application WO2007140429A2 discloses a video rate adaptation to reverse link conditions. The disclosure relates to video rate adaptation techniques that may use information from a medium access control (MAC) layer and radio link protocol (RLP) layer. This disclosure also talks about the video throughput, which is estimated based on the size of the video flow RLP queue at an access terminal and relates to adaptive quality control for video chat, making a brief mention of a framework. However, it does not disclose an on-screen keyboard.

In the prior art to develop a system for interactive television, one needs to invoke each application separately and configure them based on use case scenarios. This requires the core architect to design separate core modules for separate applications. What is observed is that these modules have a lot of commonality that can be utilized. Further, third-party applications like web browser, chat etc. also need to be integrated under the same user interface. In order to control the third-party applications in a generic way, there is also the need to control the inputs to the applications (which are normally keyboard and mouse

events) and output of the applications (which are normally graphics drivers). This leads to the requirement of having a generic media framework which can be dynamically configured to carry out different functionalities. At the same time it should allow the integration of third-party applications with minimal interactions with the same.

Objects of the Invention:

An object of this invention is to provide a multimedia and sensor system consisting of television sets or viewing mechanisms for provisioning broadcast from a plurality of sub-systems, for providing control of said television set or said viewing mechanism, and for providing seamless transfer of content between said television sets or said viewing mechanisms.

One object of this invention is to provide a universal inputting mechanism for a converged service access mechanism.

Another object of this invention is to provide a multi-functional inputting mechanism.

Another object of this invention is to provide an on-screen keyboard to a computing/multimedia device like Set-top Box of a television set such that is operable by remote control on Television (TV) and other display device without attached or inbuilt keyboard or mouse.

Yet another object of this invention is to provide a method of using the on-screen keyboard for browsing internet, emailing, chatting and sending SMS on a Set-top Box using TV as display and Infra-red remote as user-control.

Still another object of this invention is to provide a novel way of using the on-screen keyboard having with normal low cost remotes which do not have an alphabet keypad.

An additional object of this invention is to provide a system and method of providing adaptive quality in a video conferencing solution, between at least two television sets or

at least two computers or at least two television sets connected to set-top boxes or computers, or such viewing mechanisms, based on the channel condition.

Yet an additional object of this invention is to provide a system and method for providing channel independent and bandwidth independent video conferencing.

Still an additional object of this invention is to provide a system and method for dynamically adapting the quality of the video conference based on instantaneous channel condition and bandwidth availability.

Still an additional object of this invention is to provide a system for detecting the channel condition and adjusting the video and audio packet delays and packet sizes and transfers in the process of video conferencing.

Summary of the Invention:

This invention discloses a multimedia and sensor system for interaction, said multimedia and sensor system comprising novel input mechanisms and novel features for interfacing, and a novel way of integrating television content with content from the Internet and from a local storage sub-system.

According to this invention, there is provided a multimedia system consisting of television sets as viewing mechanisms, for provisioning broadcast from a plurality of sub-systems and for providing seamless transfer of content between at least two television sets, said multimedia and sensor system comprises:

- communication means adapted to provide a communication interface between a user and said television set of said multimedia system;
- integration means adapted to integrate content from a plurality of subsystems to be viewable on said television of said multimedia system;
- interaction means adapted to provide a network independent and bandwidth independent seamless interaction mechanism between users of said multimedia system; and

- display means adapted to display said integrated content and said interaction content.

According to a first embodiment of this invention, a communication means for controlling a system is envisaged where an on-screen keyboard is displayed on the monitor of a Television, by a Computer or Set-top Box or similar device and is operated by a remote control which has among other things eight (8) special keys for performing navigation and selection of character or character set. The character set is organized in blocks with each block containing up to a maximum of 4 characters. Characters are organized into character-blocks according to a mathematical formulation proposed in this invention. Hierarchical navigation and selection method is used across and within the said specially organized character blocks in such a way as to reduce the number of keystrokes required for navigation and selection. Round robin navigation is used across each row and columns of the specially organized character-blocks to reduce the number of keystrokes further in some special cases.

Typically, said communication means is a screen-viewable input device adapted to provide full-fledged input options viewable and effected from a screen, via a remote control unit of said screen, said communication means comprising:

- pre-defined number of keys on said remote control, said number of keys being less than the number of keys viewable on screen;
- on-screen viewable keys, distributed in a pre-defined format, adapted to be displayed on a display means; and
- correlation means adapted to correlate said remote control keys, in a pre-defined manner, with said on-screen viewable keys.

Preferably, said pre-defined number of keys comprises navigation keys.

Preferably, said pre-defined number of keys comprises character selection keys.

According to this invention there is provided a method and apparatus for hierarchical navigation and typing on device without attached or inbuilt physical keyboard and mouse by use of an on-screen keyboard system which comprises:

- an on-screen keyboard with characters, symbols, character-sets optimally organized into key-blocks wherein each key-block consists one or many characters, symbols or character-sets and the said optimal organization is such that it reduces the number of keystrokes for navigation and typing; and

- a remote control device, among its other keys, having four (4) special keys for navigation across the key-blocks and another four (4) special keys for selection of characters, symbols, character-sets within the key-blocks.

This invention envisages a novel inputting mechanism for inputting text, or graphic on an input receiving mechanism, typically on a viewing mechanism. Typically, the inputting mechanism is a virtual inputting mechanism.

Typically, said communication means comprises page formulation means adapted to formulate a plurality of pages for viewing of said correlated on-screen keys, in a divided and formulated manner.

The on-screen keyboard system typically can have different organization for a few characters, symbols and or character-sets to provide them greater prominence and or visibility and the said organization is accomplished by a key-block having less than regular maximum number of cells and or some cells in a key-block not being used.

Typically, said correlation means is a shortcut defining means adapted to collate and correlate a combination of keywords to a specific key of the remote control.

The one or more cells in the key-block can typically contain a character-set containing more than one character so that after navigating to the said key-block, the character-set can be typed with a single keystroke.

The on-screen keyboard solution in accordance with this invention is specially designed for internet browsing and chatting. Instead of pressing 3 times 'w', a user shortcut key named as 'www.' and '.com' are given in the keyboard. Many smileys have been put as shortcuts along with normal keyboard symbols like '\$' and '#'.

Typically, said formulation means includes a priority based variable viewing means adapted to appropriate priority to a set of characters for on-screen viewing depending upon a) most recently used set of key-strokes; b) more frequently used set of key-strokes; and c) user-defined set of key-strokes.

In accordance with a preferred embodiment of the invention, a given number of N total characters can be optimally organized into R rows and C columns or C rows and R columns of key-blocks on the on-screen keyboard so that any character, symbol or character-set on the same screen will maximum be $(R+C-1)$ keystrokes away. By making use of hot key, any character, symbol or character-set on a different screen will be maximum $(R+C)$ key strokes way. The alphabets are arranged from [a-z] in descending order in each line for ease of search even by computer illiterate persons. Each block has 4 alphabets, and user moves from block to block using up, down, right and left arrow keys. Then using diagonal arrow keys, alphabet of interest is chosen. So for typing any alphabet, user does not have to use more than $(R+C)$ keystrokes. This eliminates problem of traversing long distances while using QWERTY on-screen keyboard.

More particularly, this invention brings values to systems, which do not have wireless mouse implementation. Even with wireless mouse, positioning the cursor on a on-screen keyboard is troublesome for computer illiterate users.

The on-screen keyboard solution in accordance with this invention has 3 different layouts – one for small letters, one for capital letters and the other is for rest of the keys in a standard computer keyboard. Transitioning between these screens are according to hot keys in standard remotes. So in all totality $(8 + \text{No. of different key layouts})$ buttons of remotes are required to give full computer keyboard facility to all TV applications. For

example, in the current embodiment, a total of 11 buttons are required on the remote to switch, navigate and select between three different on-screen key-board layouts.

In particular this invention relates to reduced amount of button press in a remote control for going to any key from any key within a on-screen keyboard.

Typically, some keys have a constant function on each screen and some keys have variable functions of various screens.

Typically, a quadrangle may be divided into 4 quadrants, each of said quadrants may cater to an individual button/function of a keyboard such as an alphabet or a numerical or a punctuation mark or the like. A plurality of such quadrangles may be dispersed in a pre-defined space to form the inputting mechanism of this invention.

Typically, said input means is adapted to individually select each of the virtual buttons or selection applications from a remote location i.e. spaced apart from the viewing mechanism. Said input means may be hosted by a set-top box of a TV, or a toggle mechanism adapted to select each of the functionalities offered by toggling from one functionality to another, or a remote control mechanism or a pointer mechanism, or a touch screen mechanism or the like.

In accordance with yet another embodiment of this invention said virtual application may be loaded onto said viewing mechanism by means of a set-top box. Thus the pre-programming of the set-top box may provide for the functionalities offered on the viewing mechanism through said set-top box, and hence, the virtual application means may be simultaneously configured in accordance with the functionalities provided/collaborated on said viewing mechanism.

In accordance with still another embodiment of this invention, there is provided a transmission means adapted to transmit control from said input means to said viewing mechanism.

In accordance with an additional embodiment of this invention, there is provided a receiving means adapted to receive control from said input means.

Although the invention relates to a virtual inputting means, a real physical keyboard may also be constructed in accordance with the layout as described in accordance with this invention. Thus enables reduction in size of the keyboard, enhanced output configuration adapted to be provided by the same keyboard.

According to a second aspect of this invention, said system envisages a novel way of integrating broadcast TV content available from existing set top boxes with the information and media content available from a local storage and from the Internet. The blending of the broadcast TV content with the information and media content available in local storage and internet also opens up possibilities of various value-added applications where the broadcast TV content can be made interactive.

Typically, said integration means comprises:

- media centric subsystems adapted to host or relay corresponding media;
- media centric processing means adapted to process varied applications from said media centric subsystems to achieve a common substrate for relaying said media;
- application development means adapted to provide a user development interface for applying user-based developments;
- input event decision manager adapted to manage distribution of user inputs to specific modules of said user development interface;
- collation means adapted to collate said processed media and said user based developments to obtain integrated multimedia content; and
- display means adapted to display integrated multimedia content.

The invention provides a mechanism to blend already available broadcast TV content (available from local cable TV provider or satellite TV provider through a separate box) with information and media content available in local storage and internet for doing value-added interactive applications on TV.

In order to achieve the above requirement, this invention also aims at the creation of a generic media framework consisting of a Data Source Sub-system, a Data Process Sub-system, a Data Sink Sub-system and a Control Sub-system, which can be dynamically configured to bring about a wide-range of media and data centric interactive applications.

In particular, this invention envisages a subsystem based approach for processing of digital media and sensor data in order to make all applications come under a common architecture, using which any application can be developed and even third party application executables can also be integrated provided they have a well defined set of user inputs.

The framework solution in accordance with this invention provides an Input Event Decision Manager (IEDM) which works in tandem with a user interface to maintain the application state. Based on this application state, the IEDM forwards certain user inputs, received from kernel event drivers like that of keyboard, mouse, IR etc. to certain user modules. Using this mechanism any third party application can be controlled by filtering the user events that it wishes to receive.

The framework solution in accordance with this invention further provides means to launch a third-party application (like web browser) from the framework along with running Display Enhancement algorithms like TV anti-flicker through the display driver.

According to a third aspect of this invention, the system discloses a method of providing adaptive quality in a video conferencing solution based on the channel condition.

Typically, said interaction means is adapted to provide interaction by means of video conferencing, which is bandwidth independent and communication channel independent, said interaction means comprising:

- transmitter means further comprising:
 - capturing means adapted to capture multimedia;
 - encoding means adapted to encode said captured multimedia;

- tampering means adapted to tamper said encoded multimedia for ease of transmission;
- size determination means adapted to determine size of said tampered multimedia;
- tagging means adapted to tag said tampered multimedia for appropriate transmission;
- gap determination means adapted to send probe packets to determine gaps in transmission of said multimedia; and
- channel determination means adapted to determine status of communication channel;
- receiver means further comprising:
 - means to receive transmitted multimedia;
 - means to respond to probe packets of said gap determination means;
 - extraction means adapted to extract said tagged information for appropriate transmission;
 - decoding means adapted to decode said received multimedia; and
 - playing means adapted to play said multimedia.

Typically, said capturing means comprises:

- video capturing means adapted to capture video data;
- audio capturing means adapted to capture audio data; and
- text capturing means adapted to capture input text data.

Typically, said encoding means comprises:

- video encoding means adapted to encode said captured video data;
- audio encoding means adapted to encode said captured audio data; and
- text encoding means adapted to encode said text data.

Typically, said interaction means includes channel bandwidth estimation means based on said size determination means and said channel determination means.

Typically, said tampering means comprises:

- Video packet generation means adapted to generate video data packets, based on output of bandwidth estimation means adapted to estimate channel bandwidth, by fragmenting a single video frame;
- Audio packet generation means adapted to generate audio data packets, based on output of bandwidth estimation means adapted to estimate channel bandwidth, by aggregating multiple video frames; and
- Text packet generation means adapted to generate text data packets, based on output of bandwidth estimation means adapted to estimate channel bandwidth, by aggregating pre-defined bits of text data.

Typically, said size determination means comprises:

- video packet size determination means adapted to determine size of said video packet;
- audio packet size determination means adapted to determine size of said audio packet; and
- text packet size determination means adapted to determine size of said text packet;

Typically, said tagging means comprises:

- video header generation and insertion means adapted to generate and insert headers for video packets;
- audio header generation and insertion means adapted to generate and insert headers for audio packets; and
- text header generation and insertion means adapted to generate and insert headers for text packets.

Typically, said gap determination means comprises:

- probe packet generation means adapted to send probe packets in said communication channel to obtain silence period;
- video discontinuation means adapted to discontinue transmission of video packets based on the detection of said silence period;

- audio discontinuation means adapted to discontinue transmission of audio packets based on the detection of said silence period; and
- text discontinuation means adapted to discontinue transmission of text packets based on the detection of said silence period.

Typically, said channel determination means comprises:

- video parsing means adapted to parse video packets for authenticating transmission;
- audio parsing means adapted to parse audio packets for authenticating transmission; and
- text parsing means adapted to parse text packets for authenticating transmission;

Typically, said interaction means comprises:

- invitation means adapted to provide an invitation from a first user of said interaction means to a second and subsequent user(s) of said interaction means by means of online or offline communication medium.

Typically, said invitation means is by means of a communication means selected from a plurality of communication means consisting of real-time chat interface means, e-mail client, and off-line SIM based SMS means.

Typically, interaction means comprises a resolution configuration means adapted to configure resolution

Typically, said interaction means includes means to select mode of communication from full-duplex communication mode and half-duplex communication mode.

Typically, said interaction means includes means to generate delays between packets of transmission depending on channel determination means, bandwidth determination means and gap determination means.

Particularly, this invention envisages a video conferencing solution using the PPP (Point to Point Protocol).

In particular, this invention envisages a novel way of detecting the channel condition and dynamically adjusting the video and audio packet sizes.

Still particularly, this invention envisages a novel way of detecting the channel condition and adjusting the video and audio packet delays in the process of video conferencing.

In particular, this invention envisages video conferencing over a CDMA 1xRTT (Code Division Multiple access 1x Real Time Transmission) channel and wireline ADSL/LAN (Asynchronous Digital Subscriber Line / Local Area Network) using IP (Internet Protocol). However, the method in accordance with this invention is also applicable for any communication channel where the instantaneous bandwidth capacity changes with time. This invention is also directly applicable to any type of wireless channel including WiMAX, WLAN. This invention is further also applicable to wireline network if the QoS (Quality of Service) is not guaranteed by the network service providers.

In particular, this invention envisages a video conferencing solution where the transmitter apparatus captures the video using an analog camera or web camera and audio using a microphone on a DaVinci based platform, encodes the video using a video encoder and audio using an audio encoder and finally the encoded data along with video conferencing specific header is encapsulated using UDP (Universal Datagram Protocol) and transmitted over IP. A receiver apparatus receives the UDP data over wireless or wireline channel, extracts the header and encoded data, decodes and displays the video data and decodes and plays the audio data. .

Typically, the video is encoded using H.264 baseline encoder.

Typically, the audio is encoded using NB-AMR (Narrow Band – Adaptive MultiRate). The invention is also applicable to any speech (G.711, G.723, G.729) or audio codec (MP3, AAC).

In accordance with a preferred embodiment of the invention, the data is used to change the video and audio packet sizes, delays and sometimes make it half-duplex.

This invention also proposes video and audio half duplex along with video packet size and delay changes.

The method in accordance with this invention does a bandwidth estimation based on probe packets and controls the video and audio packet size and delays.

This invention aims at optimum video conferencing solution by automatic configuration of the packets sizes for the audio and video data and automatic configuration of the video packet interval based on the instantaneous channel condition. The solution consists of transmitter and receiver apparatus.

The invention consists of following modules in the transmitter apparatus:

- i. Video data capture from camera
- ii. Audio data capture from microphone
- iii. H.264 baseline video encoder
- iv. NB-AMR encoder
- v. Generation of video packets by fragmenting a single video frame
- vi. Generation of audio packets by aggregating multiple audio frames
- vii. Automatic determination of the video and audio packet size
- viii. Automatic determination of the video packet interval
- ix. Generation and insertion of video and audio headers for video and audio packets respectively
- x. Transmission of video and audio packets using UDP/IP

- xi. Optional discontinuation of the transmission of audio, based on the detection of silence period in the audio
- xii. Optional discontinuation of the transmission of video, based on the detection of silence period in the audio
- xiii. Generation of the probe packets to determine the instantaneous channel condition and bandwidth
- xiv. Determination of the instantaneous channel condition and bandwidth by parsing the PPP statistics

The invention consists of following modules in the receiver apparatus:

- i. Reception of video and audio packets over UDP/IP
- ii. Response of the probe packets
- iii. Extraction of video and audio payloads by parsing the headers
- iv. Aggregation of video packets to form a complete frame
- v. Decoding of video frame
- vi. Display of video frame
- vii. Separation of audio frames from an audio packet
- viii. Decoding of audio frame
- ix. Playing of the audio data on the speakers

The video conferencing solution in accordance with this invention, discloses a transmitter and a receiver apparatus which typically can be one of the following:

- Personal Computers (PC),
- Digital Set Top Boxes (DSTB)
- Notebook (NB),
- Personal digital assistance (PDA),
- Mobile phone,
- Any consumer device having video conferencing facility.

The video conferencing solution in accordance with this invention has means to send an invitation for video chat from one device to another device using TCP/IP (Transmission

Control Protocol / Internet Protocol). The invitation can be sent using the IP address or a name of the device. The name of the device is uploaded to the video conferencing server once the device boots up.

The video conferencing solution in accordance with this invention is applicable to device involved in video conferencing application which may reside in heterogeneous networks. One may reside in wireless 1xRTT channel and the other may reside in ADSL.

The video conferencing solution in accordance with this invention also has means whereby an inviter can send an SMS to the invitee if the invitee is offline.

The video conferencing solution in accordance with this invention can be set to different modes, namely fast video, normal video, slow video and video freeze. In the last mode, only audio will be transmitted.

The video conferencing solution in accordance with this invention has means whereby the modes of the video conferencing solution can be set dynamically by the user while the video conferencing is on going.

The video conferencing solution in accordance with this invention has means whereby the modes of the video conferencing solution can be set automatically based on the channel condition and available bandwidth.

The video conferencing solution in accordance with this invention has means whereby the resolution of the video encoded data can be configured by the user or automatically configured based on the channel bandwidth.

The video conferencing solution in accordance with this invention has means whereby, the spatial quality of the video encoded data can be configured by the user or automatically configured based on the channel bandwidth.

The video conferencing solution in accordance with this invention has means whereby the video encoded data is packetized into N bytes of fragments for a single frame.

The video conferencing solution in accordance with this invention has means whereby the audio encoded data is aggregated into M frames before a packet is transmitted.

The video conferencing solution in accordance with this invention has means whereby the delay between successive video encoded packets is derived from the current channel condition.

The video conferencing solution in accordance with this invention has means whereby the audio transmission will be discontinued after D seconds during the silence period to save the bandwidth due to UDP/IP overheads.

The video conferencing solution in accordance with this invention has means whereby the audio transmission will automatically change from full duplex to half-duplex based on the channel condition.

The video conferencing solution in accordance with this invention has means whereby the video transmission will be discontinued after D seconds during the silence period of audio to save the bandwidth.

Brief Description of the Accompanying Drawings:

The invention will now be described with reference to the accompanying drawings, in which:

Figure 1 illustrates the architecture of the system in accordance with this invention;

Figure 2 illustrates a schematic of a control/communication mechanism;

Figure 3 illustrates a typical organization of keyboard according to proposed scheme of the control/communication mechanism, when mostly lower case alphabets are selected;

Figure 4 illustrates a typical organization of keyboard according to proposed scheme of the control/communication mechanism, when mostly capital alphabets are selected;

Figure 5 illustrates a typical organization of keyboard according to proposed scheme of the control/communication mechanism, when mostly symbols are used;

Figure 6 illustrates a typical layout of remote control of the control/communication mechanism, which can be used to operate the on-screen keyboard;

Figure 7 illustrates a particular representation of algorithm of the control/communication mechanism, to elaborate the way of organizing the characters, symbols or character-sets into key-blocks;

Figure 8 illustrates an exemplary embodiment of the plurality of options viewable at the viewing mechanism for providing input from the control/communication mechanism,;

Figure 9 illustrates a schematic of the interaction mechanism in accordance with this invention;

Figure 10 illustrates a schematic block diagram of the transmission apparatus of the interaction mechanism in accordance with this invention;

Figure 11 illustrates a schematic block diagram of the reception apparatus of the interaction mechanism for the video-conference solution in accordance with this invention;

Figure 12 illustrates a scheme for effective bandwidth calculation for the interaction mechanism, using round trip delay of probe packets;

Figure 13 illustrates a schematic diagram of the integration mechanism to show how various applications can be mapped using the media framework; and

Figure 14 illustrates a schematic diagram of the integration mechanism to describe the flow of content in an interactive TV application.

Detailed Description of Invention:

Figure 1 illustrates the architecture of the system in accordance with this invention. It illustrates the multimedia system consisting of Television sets (TV1, TV2) as viewing mechanisms for provisioning broadcast from a plurality of sub-systems such as local storage (LS), Internet (I), and Set top box (STB) and for providing seamless transfer of content between at least two television sets (TV1, TV2), which further includes communication/control means (CM), integration means (IGM) and interaction means (ICM).

A first aspect of the invention (as shown in Figures 2-10) relate to typing on the monitor of display device like Television or any other device through a computing device like Set-top Box etc. which does not have a separate keyboard and mouse.

According to Figure 2 of the accompanying drawings, this invention envisages a novel inputting mechanism (IM) for inputting text, or graphic on an input receiving mechanism, typically on a viewing mechanism (VM). Typically, the inputting mechanism (IM) is a virtual inputting mechanism.

In accordance with an embodiment of this invention, there is provided a viewing mechanism (VM) such that said virtual inputting means (IM) can be viewed on said viewing mechanism. Typically, said viewing mechanisms (VM) may offer a plurality of options such as web-browsing services, chat services, and the like. Typically, for each of the services offered, there is provided a virtual button or selection application.

In accordance with yet another embodiment of this invention said virtual application may be loaded onto said viewing mechanism (VM) by means of a set-top box (S). Thus the pre-programming of the set-top box (S) may provide for the functionalities offered on the viewing mechanism through said set-top box, and hence, the virtual application means may be simultaneously configured in accordance with the functionalities provided/collaborated on said viewing mechanism.

In accordance with still another embodiment of this invention, there is provided a transmission means (TX) adapted to transmit control from said input means (IM) to said viewing mechanism (VM).

In accordance with an additional embodiment of this invention, there is provided a receiving means (RX) adapted to receive control from said input means (IM).

Typing on the monitor of the device is required for accessing Internet and for accessing many other services. On-screen keyboard is used in this kind of situation. However the currently available on-screen keyboards require a large number of keystrokes. The current Invention propose a on-screen keyboard which can be operated with a purpose-built remote control using reduced number of key-strokes.

The characters in the on-screen is organized into blocks of upto 4 characters or symbols, or character-sets in a block. However it is possible to have fewer than 4 characters, or symbols or character-sets in a key-block for providing greater prominence to some characters, or symbols or character-sets and thus better ease-of-use. The algorithm in Figure 7 gives a typical method for organizing a given number of characters, symbols or character-sets into an optimum number of key-blocks organized horizontally and vertically. The number of key-blocks in horizontal direction is called columns and number of key-blocks in the vertical direction is called rows.

In Figure 3, the top-left key-block below the space bar key-block contain the characters a, b, g, h as below.

a	b
g	h

Each of the individual portions in a key-block is called a cell. Thus the key-block above contain 4 cells and the cells contain a, b, g and h. A cell can contain a single character, a symbol or a set of characters. The character sets can be commonly used sequence in a given domain like “www.” or “.com” in Internet. The advantage of this kind of character-sets is that the complete character-set can be typed with a single selection reducing the typing effort even further. In the above-mentioned key-block the letter a is in the up-left cell in top-left corner, the letter b is in the up-right cell in top-right corner, the letter g is in the down-left cell in bottom-left corner and the letter h is in the down-right cell in bottom-right corner.

The on-screen keyboard is operated by a remote control and a typical representation of the keyboard is provided in Figure 6 of the accompanying drawings. The Up (101), Down (102), Left (103) and Right (104) keys in the remote control is used to move across the key-blocks in vertical and horizontal direction. As these 4 keys are used to navigate across key-blocks they are called Navigational keys.

The remote control also has Select-Up-Left (105) key, Select Up-Right (106) key, Select Down-Left (107) key and Select Down-Right (108) key as diagonal arrow keys. They are used to select one of the available character, symbols or character-sets in a key-block and hence they are called selection keys.

Navigation between key blocks can be done using the Up (101), Down (102), Left (103) and Right (104) keys. After navigating to a key-block, Select Up-Left (105) key can be used to select the character, symbol or character-set in the up-left cell, Select Up-Right (106) key can be used to select the character, symbol or character-set in the up-right cell, Select Down-Left (107) key can be used to select the character, symbol or character-set in the down-left cell, and Select Down-Right (108) key can be used to select the character, symbol or character-set in the down-right cell.

A key block can have fewer than 4 cells to give prominence to and ensure greater visibility of some characters, symbols or character-sets. Some typical layout of key-blocks are drawn below:



Where a key-blocks has only 2 cells organized vertically, either of Select Up-Left (105) and Select Up-Right (106) key can be used to select content in upper cell and either of Select Down-Left (107) and Select Down-Right (108) can be used to select content in lower cell.



Where a key-blocks has only 2 cells organized horizontally, either of Select Up-Left (105) and Select Down-Left (107) key can be used to select content in left cell and either of Up Right (106) and Down Right (108) can be used to select content in right cell.



When a key-block contain only one cell, any of the 4 Selection keys can be used to select its content.

A special meaning can be attached to a cell in that case selecting the said cell will manifest some special behavior rather than typing the content on that cell. In representative Figure 1, the cell containing the sequence “Caps” has a special meaning in the sense that selecting it will make available the upper case version of the on-screen key-board with a typical representation given in Figure 3 of the accompanying drawings.

The switching between Small letter screen, Capital letter screen, Symbol screen and any other types of screens can also be achieved through specially assigned hot keys in the remote control or through specially assigned buttons on the On-screen keyboard.

Typically, the navigation across the key-blocks is in Round Robin mode in both vertical and horizontal direction. In the vertical direction when the cursor is in the bottom-most key-block and if Down (102) key is pressed then the cursor moves to the top-most key-block in that column. In similar manner Round Robin behavior can be observed in all direction.

In one embodiment of this invention, a screen layout resembling the layout in Figure 3 of the accompanying drawings can be used when mostly lower case English alphabets are required. In this embodiment, when the cell containing "Caps" is selected, the keyboard containing the upper case letters appears and it allows typing in upper case. In this embodiment, when the cell containing "Symbol" is selected, the keyboard containing some symbols appears and it allows typing those symbols. The symbols can be single character or multiple characters. A typical representation of screen-layout with the symbols is presented in Figure 5 of the accompanying drawings.

Figure 7 of the accompanying drawings provides the flow-chart for a typical representation as a way of illustration for one way of organizing a given numbers of characters or symbols or character-sets optimally so that number of required keystrokes is reduced in accordance with the current invention. Number of characters or symbols or character-sets or any combination thereof is taken as an input for the algorithm and is represented by numChars variable in the flow chart.

This total set of numChars characters, symbols or character-set can be organized into blocks of C columns and R rows where each key-block have a maximum of 4 cells.

The method determines the square number with even square root which is equal to numChars or the next square number after numChars. The reason for locating the square

number with even number square root is that in each direction (vertical or horizontal) 2 cells can be placed in a key-block. In the flow-chart in Figure 7, the variable named num1 contain the said even square root number. Since square of num1 may be significantly more than numChars, there is a high probability that multiplication of num1 and its preceding even number can be more than or equal to numChars. It can be deduced from this algorithm that optimum number for columns and rows of key-block for optimal navigation cab be either of $(\text{num1})/2$ columns and $(\text{num1})/2$ rows, or $(\text{num1})/2$ columns and $(\text{num1} - 2)/2$ rows or $(\text{num1} - 2)/2$ columns and $(\text{num1})/2$ rows.

A normal onscreen keyboard may contain n columns and k rows of keys. It will take an user a maximum of $(n+k-1)$ keystrokes to navigate from one edge (say in top-left) to other edge (say bottom-right). In the proposed scheme, $\text{numChars} = n*k$. The maximum no. of keystrokes to navigate in this scenario will be $(\text{num1} - 1)$, where num1 is calculated as per the flowchart diagram in Figure 7 of the accompanying drawings. It is trivial to prove that num1 will always be less than $n + k$ and hence the proposed keyboard will always perform better in terms of navigation keystrokes required compared to traditional keyboard.

Even with a non-optimal organization of columns and rows of key-blocks, some degree of benefit of reduced keystrokes can be obtained if the basic principle of 8 key based key-block navigation is adopted.

Typically, said novel inputting mechanism (IM), of Figure 2 of the accompanying drawings, has a pre-defined layout of keys (as seen in Figures 3, 4, 5 and 10).

In accordance with an embodiment of this invention, the input mechanism (IM) typically invokes a plurality of screens (Figure 3, Figure 4, Figure 5), wherein each screen caters to a host of inputting options. Typically, one screen (Figure 3) may showcase all the alphabets and punctuation marks is small. Typically, another screen (Figure 5) may showcase further punctuation marks and special characters. The keys are thus encoded to produce a plurality of outputs. The primary (most common) outputs are the alphabets of a

language, typically, the English language, as well as the numbers. Secondary outputs include punctuation marks, special characters and special insert-able objects.

In accordance with another embodiment of this invention, said input mechanism (IM) is adapted to be configured such that it caters to tertiary outputs too.

Typically, some keys have a constant function on each screen (such as the spacebar key as seen in Figures 3, 4, and 5 respectively) and some keys have variable functions on various screens (such as the A, B, C, D keys).

Typically, a quadrangle may be divided into 4 quadrants, each of said quadrant may cater to an individual button/function of a keyboard such as an alphabet or a numerical or a punctuation mark or the like. A plurality of such quadrangles may be dispersed in a pre-defined space to form the inputting mechanism of this invention.

Typically, said input means (IM) is adapted to individually select each of the virtual buttons or selection applications from a remote location i.e. spaced apart from the viewing mechanism. Said input means (IM) may be hosted by a set-top box of a TV, or a toggle mechanism adapted to select each of the functionalities offered by toggling from one functionality to another, or a remote control mechanism or a pointer mechanism, or a touch screen mechanism or the like.

For the purposes of this specification, a communication channel includes Code Division Multiple access channel, wireline Asynchronous Digital Subscriber Line, wireline Local Area Network Line.

Referring to the drawings 9, 10, and 11 of the accompanying drawings, the video conferencing solution implemented on the interactive TV STB consists of following components as shown in Figure 9:

- i. Video conferencing application – This is the application, responsible to take user inputs to initiate, accept, reject or terminate videoconference.

- ii. STB Framework – This is a generic framework responsible to accept commands from application or user inputs (keyboard, mouse etc.) and send messages to different processing modules.
- iii. Encode control thread – This is the thread to accept commands from the framework for controlling the video and audio encoding.
- iv. Decode control thread – This is the thread to accept commands from the framework for controlling the video and audio decoding.
- v. Video encode thread – This is the thread responsible to take the video frames from the camera, resize to QCIF (Quarter Common Intermediate Format) or Sub-QCIF and encode using H.264 video encoder. The encoded data is packetized and transmitted using UDP/IP. However, the proposed method is applicable to any video resolution and has no dependency on the video resolution. We have tested using the H.264 video codec. However, the proposed solution is applicable for any video codec (MPEG1, MPEG2, MPEG4 etc.). The video payload will get changed during transmission. The packet structure will not change.
- vi. Audio encode thread – This is the thread responsible to take the audio samples from the microphone sampled at 8000 Hz (Hertz) and encode 200 msec frames using NB-AMR encoder. The encoded data is packetized and transmitted using UDP/IP. This thread is also responsible to send the probe packets.
- vii. Video decode thread – This is the thread responsible to receive the video packets from the network, accumulate the packets to form a complete video frame and decode video frames using H.264 video decoder. However, the proposed solution is applicable for any video codec (MPEG1, MPEG2, MPEG4 etc.).
- viii. Audio decode thread – This is the thread responsible to receive the audio packets from the network, extract the audio frames from the packets and decode audio frames using NB-AMR audio decoder. The invention is applicable to any speech (G.711, G.723, G.729) or audio codec (MP3, AAC).
- ix. Probe decode thread – This is the thread responsible to receive the probe packets and determine the instantaneous channel condition. This information is used to determine the video packet size and video packet interval.

Transmission Apparatus

The modules of the transmission apparatus are shown in figure 10 of the accompanying drawings.

Video and audio data capture

The video is captured in CIF (Common Intermediate Format), resized to QCIF and send to display. This is the display for the self-view.

The audio is captured using 8000 Hz sampling rate at 16 bits per samples.

Video encoding and packetization

The captured CIF video is resized to QCIF or sub-QCIF and encoded using H.264 video encoder.

The encoded video frames are broken into packets of N bytes. A 9-byte header is added to each video packet. Each video packet is transmitted at an interval of D_v seconds.

$N = \text{function (Eff BW)}$

$D_v = \text{function (Eff BW)}$

Configurable parameters of video encoder are given below:

- i. Input resolution (width/height) – user configurable
- ii. Video packet size – user and automatic configurable
- iii. Video packet interval – user and automatic configurable
- iv. Maximum video bit rate – user configurable
- v. Maximum video frame rate – user configurable

Details of the 9-byte header for a video packet is given below:

- i. Frame type (1 byte) – This is the indication of the type of frame whether it is I (Independent) or P (Predictive) frame.
- ii. Total sub-sequence number (1 byte) – This is the number indicating the total number of packets generated from a video frame.
- iii. Sub sequence number (1 byte) – This is the packet number.
- iv. Sequence number (4 bytes) – This is the video frame number.
- v. Video payload size (2 bytes) – This contains the number of video bytes in the current packet.

Audio encoding and packetization

The 20 msec audio frames are encoded using NB-AMR audio encoder. M audio frames are aggregated and send in a single UDP packet. The DTX (Discontinuous Transmission) is enabled in the NB-AMR encoder, which indicates the silence period in the audio. If the silence period for more than D seconds then the audio transmission is discontinued. If the silence period for more than D seconds then the video transmission is discontinued optionally.

In the current implementation, M is 10 and D is configurable based on the channel condition. For a good channel, D is large (more 10 seconds), for a bad channel, D is small (3 to 5 seconds).

Probe packet and channel bandwidth calculation

The probe packets are transmitted by the transmission apparatus and returned back by the receiver apparatus. The details of transmission of probe packets is shown in figure 12 of the accompanying drawings.. Each time two probe packets are transmitted P1(t_i) and P2(t_i). The time interval (T) between sending probe packets is a function of the RTT (Round trip time) of the probe packets. The value of T is initialized to 2 seconds in the beginning.

The packet structure of the probe packet is given below:

Packet header – 9 bytes

Packet payload size – 63 bytes

Packet UDP/IP overhead – 28 bytes

Total probe packet size – 100 bytes

The effective bandwidth of the channel is calculated based on the following equation.

Eff BW = function (RTT(P1, t_i), RTT (P2, t_i))

T = function (RTT(P1, t_i), RTT (P2, t_i)), where $T = t_{i+1} - t_i$

Where,

Eff BW is Effective Bandwidth of the channel

RTT (P1, t_i) is the round trip time of the first probe packet transmitted at time t_i

RTT (P2, t_i) is the round trip time of the second probe packet transmitted at time t_i

Reception Apparatus

The modules of the reception apparatus are shown in figure 11 of the accompanying drawings.

Audio frame extraction

The payload of the audio UDP packets are received by the audio decoder thread. The audio packets are broken into encoded audio frames. Each audio encoded frame is decoded by the NB-AMR decoder and the 20 msec audio samples are played out in the speakers with 8000 Hz as the sampling rate.

Video frame extraction

The payload of the video UDP packets are received by the video decoder thread. The video packets are accumulated to form a complete video frame. Each video frame is decoded by the H.264 video decoder. The decoded frames are displayed on the screen. Due to erroneous channel conditions, if all the video packets of a frame is not received, then the complete frame is discarded and the decoder waits for the next I frame.

Probe packet extraction

The probe packets are send back to the transmitter immediately after it is received.

Referring to the drawings, the video conferencing solution implemented on the interactive TV STB consists of following components as shown in Figure 9 of the accompanying drawings:

1. Video conferencing application – This is the application, responsible to take user inputs to initiate, accept, reject or terminate videoconference.
2. STB Framework – This is a generic framework responsible to accept commands from application or user inputs (keyboard, mouse etc.) and send messages to different processing modules.
3. Encode control thread – This is the thread to accept commands from the framework for controlling the video and audio encoding.
4. Decode control thread – This is the thread to accept commands from the framework for controlling the video and audio decoding.

5. Video encode thread – This is the thread responsible to take the video frames from the camera, resize to QCIF (Quarter Common Intermediate Format) or Sub-QCIF and encode using H.264 video encoder. The encoded data is packetized and transmitted using UDP/IP. However, the proposed method is applicable to any video resolution and has no dependency on the video resolution. We have tested using the H.264 video codec. However, the proposed solution is applicable for any video codec (MPEG1, MPEG2, MPEG4 etc.). The video payload will get changed during transmission. The packet structure will not change.
6. Audio encode thread – This is the thread responsible to take the audio samples from the microphone sampled at 8000 Hz (Hertz) and encode 200 msec frames using NB-AMR encoder. The encoded data is packetized and transmitted using UDP/IP. This thread is also responsible to send the probe packets.
7. Video decode thread – This is the thread responsible to receive the video packets from the network, accumulate the packets to form a complete video frame and decode video frames using H.264 video decoder. However, the proposed solution is applicable for any video codec (MPEG1, MPEG2, MPEG4 etc.).
8. Audio decode thread – This is the thread responsible to receive the audio packets from the network, extract the audio frames from the packets and decode audio frames using NB-AMR audio decoder. The invention is applicable to any speech (G.711, G.723, G.729) or audio codec (MP3, AAC).
9. Probe decode thread – This is the thread responsible to receive the probe packets and determine the instantaneous channel condition. This information is used to determine the video packet size and video packet interval.

Transmission Apparatus

The modules of the transmission apparatus are shown in figure 10 of the accompanying drawings.

Video and audio data capture

The video is captured in CIF (Common Intermediate Format), resized to QCIF and send to display. This is the display for the self-view.

The audio is captured using 8000 Hz sampling rate at 16 bits per samples.

Video encoding and packetization

The captured CIF video is resized to QCIF or sub-QCIF and encoded using H.264 video encoder.

The encoded video frames are broken into packets of N bytes. A 9-byte header is added to each video packet. Each video packet is transmitted at an interval of D_v seconds.

$N = \text{function}(\text{Eff BW})$

$D_v = \text{function}(\text{Eff BW})$

Configurable parameters of video encoder are given below:

- vi. Input resolution (width/height) – user configurable
- vii. Video packet size – user and automatic configurable
- viii. Video packet interval – user and automatic configurable
- ix. Maximum video bit rate – user configurable
- x. Maximum video frame rate – user configurable

Details of the 9-byte header for a video packet is given below:

- vi. Frame type (1 byte) – This is the indication of the type of frame whether it is I (Independent) or P (Predictive) frame.
- vii. Total sub-sequence number (1 byte) – This is the number indicating the total number of packets generated from a video frame.
- viii. Sub sequence number (1 byte) – This is the packet number.
- ix. Sequence number (4 bytes) – This is the video frame number.
- x. Video payload size (2 bytes) – This contains the number of video bytes in the current packet.

Audio encoding and packetization

The 20 msec audio frames are encoded using NB-AMR audio encoder. M audio frames are aggregated and send in a single UDP packet. The DTX (Discontinuous Transmission) is enabled in the NB-AMR encoder, which indicates the silence period in the audio. If the silence period for more than D seconds then the audio transmission is discontinued. If the

silence period for more than D seconds then the video transmission is discontinued optionally.

In the current implementation, M is 10 and D is configurable based on the channel condition. For a good channel, D is large (more 10 seconds), for a bad channel, D is small (3 to 5 seconds).

Probe packet and channel bandwidth calculation

The probe packets are transmitted by the transmission apparatus and returned back by the receiver apparatus. The details of transmission of probe packets is shown in figure 12 of the accompanying drawings.. Each time two probe packets are transmitted P1(t_i) and P2(t_i). The time interval (T) between sending probe packets is a function of the RTT (Round trip time) of the probe packets. The value of T is initialized to 2 seconds in the beginning.

The packet structure of the probe packet is given below:

Packet header – 9 bytes

Packet payload size – 63 bytes

Packet UDP/IP overhead – 28 bytes

Total probe packet size – 100 bytes

The effective bandwidth of the channel is calculated based on the following equation.

Eff BW = function (RTT(P1, t_i), RTT (P2, t_i))

$T = \text{function (RTT(P1, } t_i), \text{RTT (P2, } t_i)), \text{ where } T = t_{i+1} - t_i$

Where,

Eff BW is Effective Bandwidth of the channel

RTT (P1, t_i) is the round trip time of the first probe packet transmitted at time t_i

RTT (P2, t_i) is the round trip time of the second probe packet transmitted at time t_i

Reception Apparatus

The modules of the reception apparatus are shown in figure 11 of the accompanying drawings.

Audio frame extraction

The payload of the audio UDP packets are received by the audio decoder thread. The audio packets are broken into encoded audio frames. Each audio encoded frame is decoded by the NB-AMR decoder and the 20 msec audio samples are played out in the speakers with 8000 Hz as the sampling rate.

Video frame extraction

The payload of the video UDP packets are received by the video decoder thread. The video packets are accumulated to form a complete video frame. Each video frame is decoded by the H.264 video decoder. The decoded frames are displayed on the screen. Due to erroneous channel conditions, if all the video packets of a frame is not received, then the complete frame is discarded and the decoder waits for the next I frame.

Probe packet extraction

The probe packets are send back to the transmitter immediately after it is received.

According to a third aspect of this invention, there is provided an integration mechanism to integrate content from a set-top box, from a local storage means, and from the Internet on to a viewing mechanism, typically on to a television set. This can be seen in Figure 13 and 14 of the accompanying drawings.

Various applications have been mapped using the current framework. Let us first consider an interactive TV application.

The media workflow of interactive TV is as follows:

Video source is A/V in and sink is display, processing is rendered/blended with graphics.

Audio source is line-in and sink is line out, processing is bypassed.

- Broadcast TV feed from existing STB goes into the interactive STB via analog input.
- The interactive STB has it own graphics media content, which is blended with the broadcast TV video.
- The final display to the user from the interactive STB is a “hybrid” of broadcast and interactive contents

The following are some of the use cases of the interactive TV.

Use Case I -- Interactive Advertisements on TV

Using this application, a user can view “click-able” advertisement content on broadcast TV. To achieve this, the web browser is transparently blended with the TV and the links are placed on “hotspots” in the content. The coordinate and URL details of these hotspots are delivered with the content meta-data. A user can click on these contents and the website for buying the product or more information on the same opens up.

Use Case II -- SMS Voting on TV

Interactive reality game shows on TV require the user to send votes using Short Messaging Service (SMS) to some well known SMS server. Further these numbers and messaging details are flashed onto the TV channel at the bottom of the screen.

Using the interactive TV application, a user can invoke a semi-transparent graphics window on the TV. The user reads the SMS details from the screen and types in the same into this SMS window. The user then sends an SMS to vote for his or her favorite contestant or answer interactive quizzes.

Use Case III – Local Display of Captured Medical Sensor Data

This use case is related to a home or remote health care center scenario where patient data is captured using medical sensor devices connected directly to the STB (using USB interface). The data source in this case is the USB driver of the medical device, the data processing block decodes the sensor data, based on device proprietary protocol or any well known medical data exchange protocols like DICOM and display using the resident user interface on the STB.

Use Case IV –Remote Upload of Captured Medical Sensor Data

This use case is related to a home or remote health care center scenario where patient data is captured using medical sensor devices connected directly to the STB (using USB interface). The data source in this case is the USB driver of the medical device, the data processing block in this case bypasses any processing and the sink is network whereby causing a HTTP/FTP upload of the binary data to desired server. This data is processed at

the server and can be viewed by an expert medical practitioner on-demand for remote consultancy.

Similarly, many other applications can also be mapped to the same infrastructure. A few other application workflows are given below:

Video Conference workflow

Flow of media for a Peer-to-Peer Video Conference application can be similarly established, where on one side the local video source is camera and sink is networking, the processing done is encoding and on the other side, the remote video source is network and sink is display, the processing done is decoding.

Similarly for audio, the local audio source is microphone and sink is network, the processing done is encoding and the remote audio source is network, sink is earphone out, the processing done is decoding.

- Video from user's camera is encoded and streamed over network to peer.
- Peer's video coming over network is decoded and displayed
- The speech data from user is captured from microphone, encoded and streamed over network to peer
- Peer's speech coming from remote end is decoded and played back to earphone out.

PVR workflow

A typical PVR scenario can be captured as follows:

Video source is A/V in, sink is storage and optionally display and processing is encoding.

Audio source is line in, sink is storage and optionally line out, and processing is encoding.

- Analog TV feed is captured, encoded and stored.
- The same feed can be sent to display and line out in parallel.

Place-shifting workflow

The place-shifting application allows remote user to log into his home Customer-premises-equipment (CPE) from a remote computer over the internet. Using this connection he can watch his favorite TV shows sitting anywhere in the world. The workflow for place-shifting application is as follows.

Video source is A/V in, sink is network as well as display, and processing is encoding.

Audio source is line in, sink is network as well as line out, and processing is encoding.

- Analog TV feed is captured and sent to display and line out
- The same feed is in-parallel encoded and sent to network to be decoded and viewed by remote host

IPTV Workflow

IPTV application can be realized as follows:

Video source is network, processing is de-multiplexing and decoding, and sink is display.

Audio source is network, processing is de-multiplexing and decoding, and sink is line out.

- In case of IPTV, the audio and video feeds are coming from network over IP.
- The content is de-multiplexed and decoded and sent to output channels.

Although the invention has been described in terms of particular embodiments and applications, one of ordinary skill in the art, in light of this teaching, can generate additional embodiments and modifications without departing from the spirit of or exceeding the scope of the claimed invention. Accordingly, it is to be understood that the drawings and descriptions herein are proffered by way of example to facilitate comprehension of the invention and should not be construed to limit the scope thereof.

Claims:

- 1.A multimedia system consisting of viewing mechanisms adapted to receive communication from a communication means, wherein said communication means is a screen-viewable input device adapted to provide full-fledged input options viewable and effected from a screen, via a remote control unit of said screen, said communication means comprising:
- pre-defined number of keys on said remote control, said number of keys being less than the number of keys viewable on screen;
 - on-screen viewable keys, distributed in a pre-defined format, adapted to be displayed on said display means; and
 - correlation means adapted to correlate said remote control keys, in a pre-defined manner, with said on-screen viewable keys.
- 2.A system as claimed in claim 1 wherein, said pre-defined number of keys comprises navigation keys.
- 3.A system as claimed in claim 1 wherein, said pre-defined number of keys comprises character selection keys.
- 4.A system as claimed in claim 1 wherein, said communication means comprises page formulation means adapted to formulate a plurality of pages for viewing of said correlated on-screen keys, in a divided and formulated manner.
- 5.A system as claimed in claim 1 wherein, said correlation means is a shortcut defining means adapted to collate and correlate a combination of keywords to a specific key of the remote control.
- 6.A system as claimed in claim 1 wherein, said formulation means includes a priority based variable viewing means adapted to appropriate priority to a set of characters for

on-screen viewing depending upon a) most recently used set of key-strokes; b) more frequently used set of key-strokes; and c) user-defined set of key-strokes.

7.A system as claimed in claim 1 wherein, said communication means includes:

- transmission means adapted to transmit control signals from said communication means to said viewing mechanism; and
- receiver means adapted to receive said transmitted control signals from said transmission means.

8.A multimedia system consisting of viewing mechanisms adapted to integrate content from a plurality of subsystems, said integration means comprising:

- media centric subsystems adapted to host or relay corresponding media or sensor data;
- media centric processing means adapted to process varied applications from said media centric subsystems to achieve a common substrate for relaying said media or sensor data;
- application development means adapted to provide a user development interface for applying user-based developments;
- input event decision manager adapted to manage distribution of user inputs to specific modules of said user development interface;
- collation means adapted to collate said processed media or sensor data and said user based developments to obtain integrated multimedia content; and
- display means adapted to display integrated multimedia and sensor data content.

9.A system as claimed in claim 8 wherein, said media centric subsystem includes subsystems selected from a plurality of subsystems consisting of set-top enabled television subsystem, local storage subsystem adapted to store static as well as dynamic multimedia content and/or sensor data, sensor device used in various domains internet effected subsystem adapted to store web related content.

10. A system as claimed in claim 8 wherein, said system includes an input event decision management engine adapted to decide the flow of user input to be targeted to a specific subsystem.
11. A system as claimed in claim 8 wherein, said system includes a remote login interface means adapted to transport the settings of various subsystems to a remotely located viewing means for availing the functionalities of said integrated subsystem.
12. A multimedia system comprising an interaction means adapted to provide interaction by means of video conferencing which is bandwidth independent and communication channel independent, said interaction means comprising:
- transmitter means further comprising:
 - capturing means adapted to capture multimedia;
 - encoding means adapted to encode said captured multimedia;
 - tampering means adapted to tamper said encoded multimedia for ease of transmission;
 - size determination means adapted to determine size of said tampered multimedia;
 - time interval determination means adapted to determine the interval between two said tampered multimedia;
 - tagging means adapted to tag said tampered multimedia for appropriate transmission;
 - gap determination means adapted to send probe packets to determine gaps in transmission of said multimedia; and
 - channel determination means adapted to determine status of communication channel;
 - receiver means further comprising:
 - means to receive transmitted multimedia;
 - means to respond to probe packets of said gap determination means;
 - extraction means adapted to extract said tagged information for appropriate transmission;

- decoding means adapted to decode said received multimedia; and
 - playing means adapted to play said multimedia.
13. A system as claimed in claim 12 wherein, said capturing means comprises:
- video capturing means adapted to capture video data;
 - audio capturing means adapted to capture audio data; and
 - text capturing means adapted to capture input text data.
14. A system as claimed in claim 12 wherein, said encoding means comprises:
- video encoding means adapted to encode said captured video data;
 - audio encoding means adapted to encode said captured audio data; and
 - text encoding means adapted to encode said text data.
15. A system as claimed in claim 12 wherein, said interaction means includes channel bandwidth estimation means based on said size determination means and said channel determination means.
16. A system as claimed in claim 12 wherein, said tampering means comprises:
- Video packet generation means adapted to generate video data packets, based on output of bandwidth estimation means adapted to estimate channel bandwidth, by fragmenting a single video frame;
 - Audio packet generation means adapted to generate audio data packets, based on output of bandwidth estimation means adapted to estimate channel bandwidth, by aggregating multiple video frames; and
 - Text packet generation means adapted to generate text data packets, based on output of bandwidth estimation means adapted to estimate channel bandwidth, by aggregating pre-defined bits of text data.
17. A system as claimed in claim 12 wherein, said size determination means comprises:
- video packet size determination means adapted to determine size of said video packet;

- audio packet size determination means adapted to determine size of said audio packet; and
- text packet size determination means adapted to determine size of said text packet;

18. A system as claimed in claim 12 wherein, said time interval determination means comprises:

- video packet interval determination means adapted to determine the interval between two said video packets;
- audio packet time interval determination means adapted to determine interval between two said audio packets; and
- text packet time interval determination means adapted to determine interval between two said text packets;

19. A system as claimed in claim 12 wherein, said tagging means comprises:

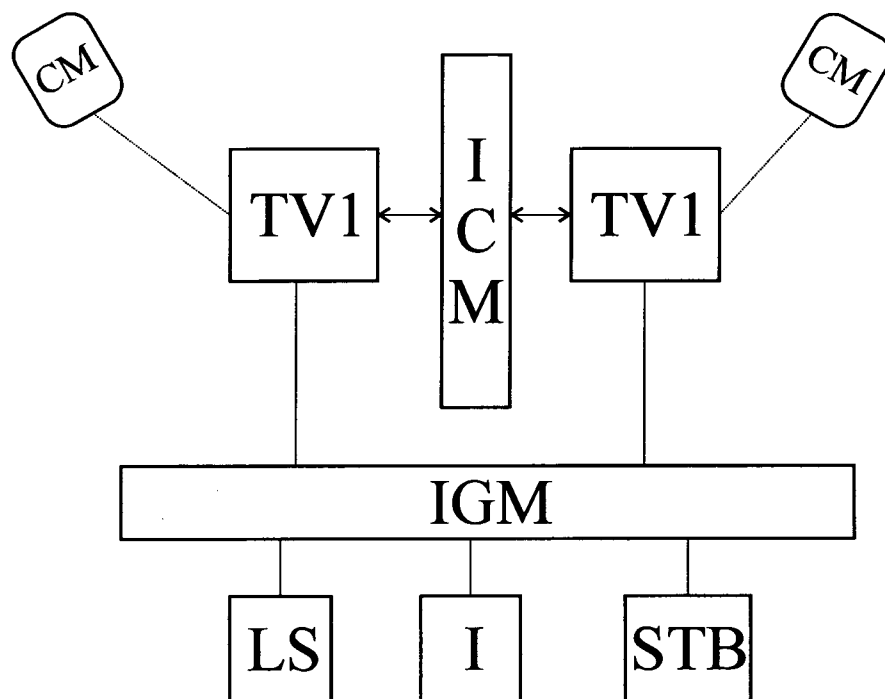
- video header generation and insertion means adapted to generate and insert headers for video packets;
- audio header generation and insertion means adapted to generate and insert headers for audio packets; and
- text header generation and insertion means adapted to generate and insert headers for text packets.

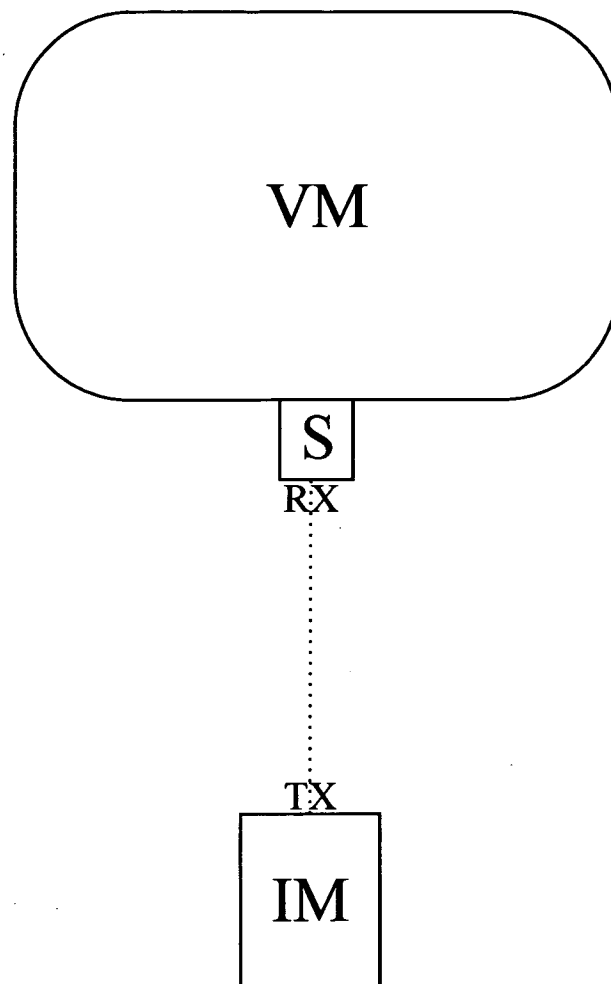
20. A system as claimed in claim 12 wherein, said gap determination means comprises:

- probe packet generation means adapted to send probe packets in said communication channel to obtain silence period;
- video discontinuation means adapted to discontinue transmission of video packets based on the detection of said silence period;
- audio discontinuation means adapted to discontinue transmission of audio packets based on the detection of said silence period; and
- text discontinuation means adapted to discontinue transmission of text packets based on the detection of said silence period;

21. A system as claimed in claim 12 wherein, said channel determination means comprises:
- video parsing means adapted to parse video packets for authenticating transmission;
 - audio parsing means adapted to parse audio packets for authenticating transmission; and
 - text parsing means adapted to parse text packets for authenticating transmission;
22. A system as claimed in claim 12 wherein, said interaction means comprises:
- invitation means adapted to provide an invitation from a first user of said interaction means to a second user of said interaction means by means of online or offline communication medium.
23. A system as claimed in claim 22 wherein said invitation means is by means of a communication means selected from a plurality of communication means consisting of real-time chat interface means, e-mail client, and off-line SIM based SMS means.
24. A system as claimed in claim 12, wherein interaction means comprises a resolution configuration means adapted to configure resolution
25. A system as claimed in claim 12, wherein said interaction means includes means to select mode of communication from full-duplex communication mode and half-duplex communication mode.
26. A system as claimed in claim 12, wherein said interaction means includes means to generate delays between packets of transmission depending on channel determination means, bandwidth determination means and gap determination means.
27. A multimedia system consisting of viewing mechanisms, for provisioning broadcast from a plurality of sub-systems and for providing seamless transfer of content between at least two viewing mechanism, said multimedia system comprising:

- communication means adapted to provide a communication interface between a user and said television set of said multimedia system;
- integration means adapted to integrate content from a plurality of subsystems to be viewable on said television of said multimedia system;
- interaction means adapted to provide a network independent and bandwidth independent seamless interaction mechanism between users of said multimedia system; and
- display means adapted to display said integrated content and said interaction content.

**FIGURE 1**

**FIGURE 2**

Space bar			
a b	c d	e f	Home End
g h	i j	k l	Bksp Del
m n	o p	q r	↓
s t	u v	w x	
y z	! &	, -	()
= ?	\ /	: ;	- _
www.	.com	Caps	Symbols

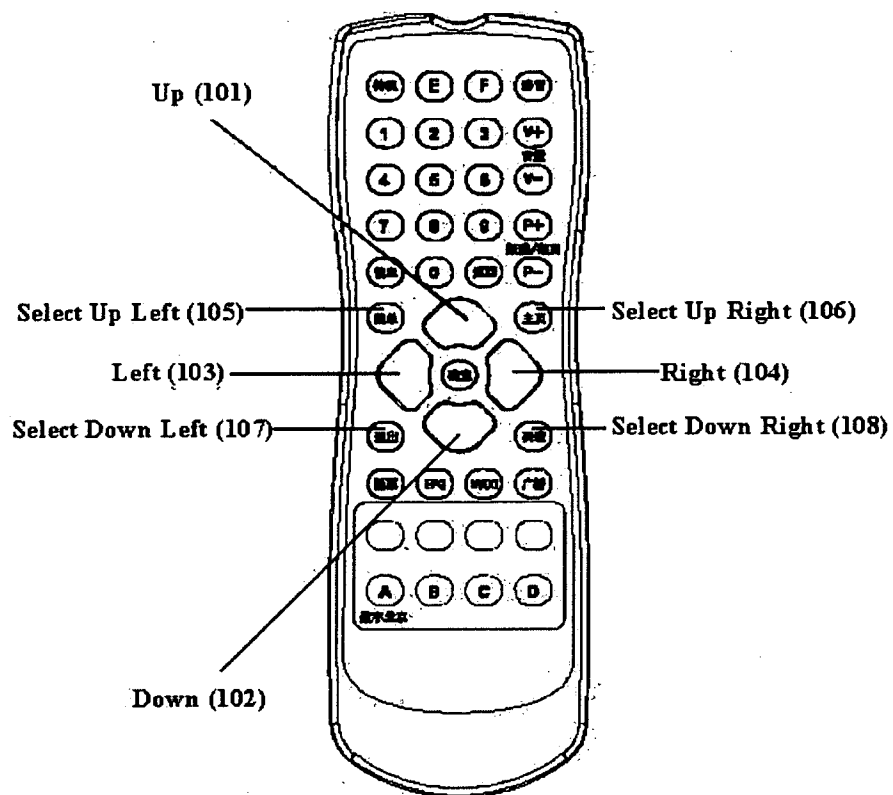
FIGURE 3

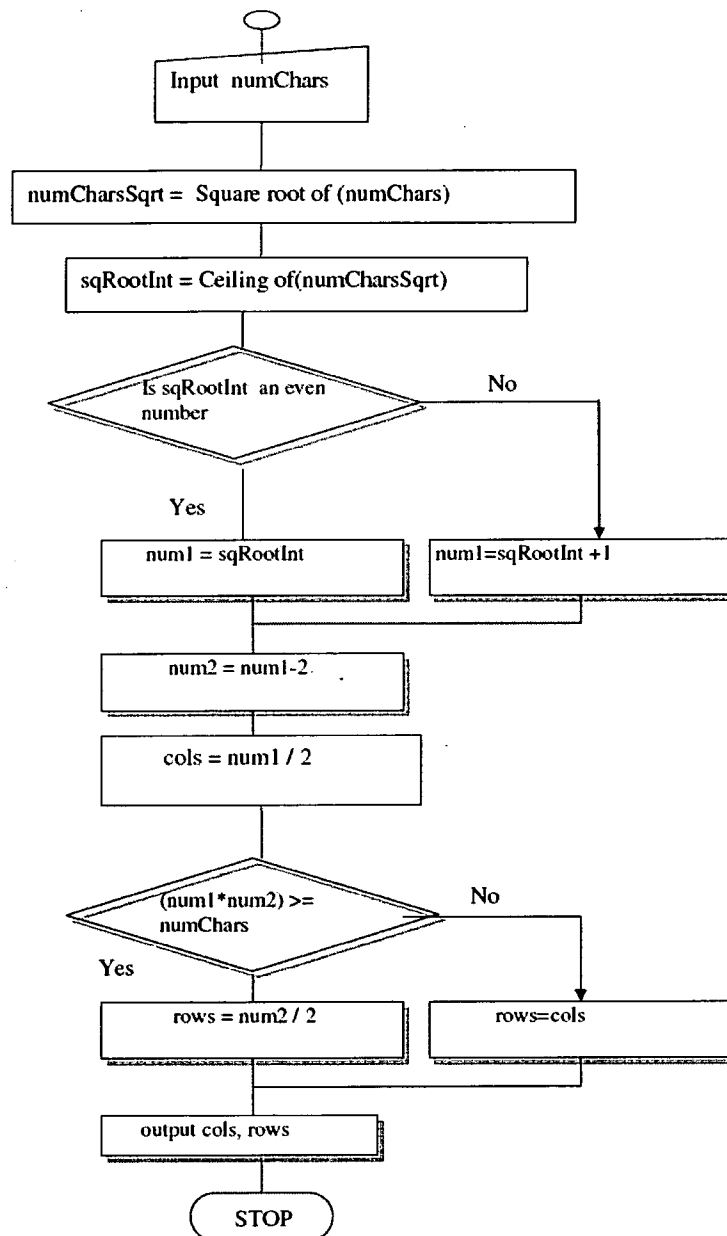
Space bar							
A	B	C	D	E	F	Home	End
G	H	I	J	K	L	Bksp	Del
M	N	O	P	Q	R		
S	T	U	V	W	X		
Y	Z	!	&	,	.	(	)
=	?	\	/	:	;	-	_
www.		.com		Small		Symbols	

FIGURE 4

Space bar			
% ^	_ @	+	Home End
# \$	< >	" '	Bksp Del
{ }	~ !	:-) :-D	↑
[]	* &	;-) :-(-	↓
www. .com	:(':	>:-< :-@	~I :Q
:-P O:-)	:- :-[	:-\ :-O	:-* \~o
www.	.com	Caps	Small

FIGURE 5

**FIGURE 6**

**FIGURE 7**

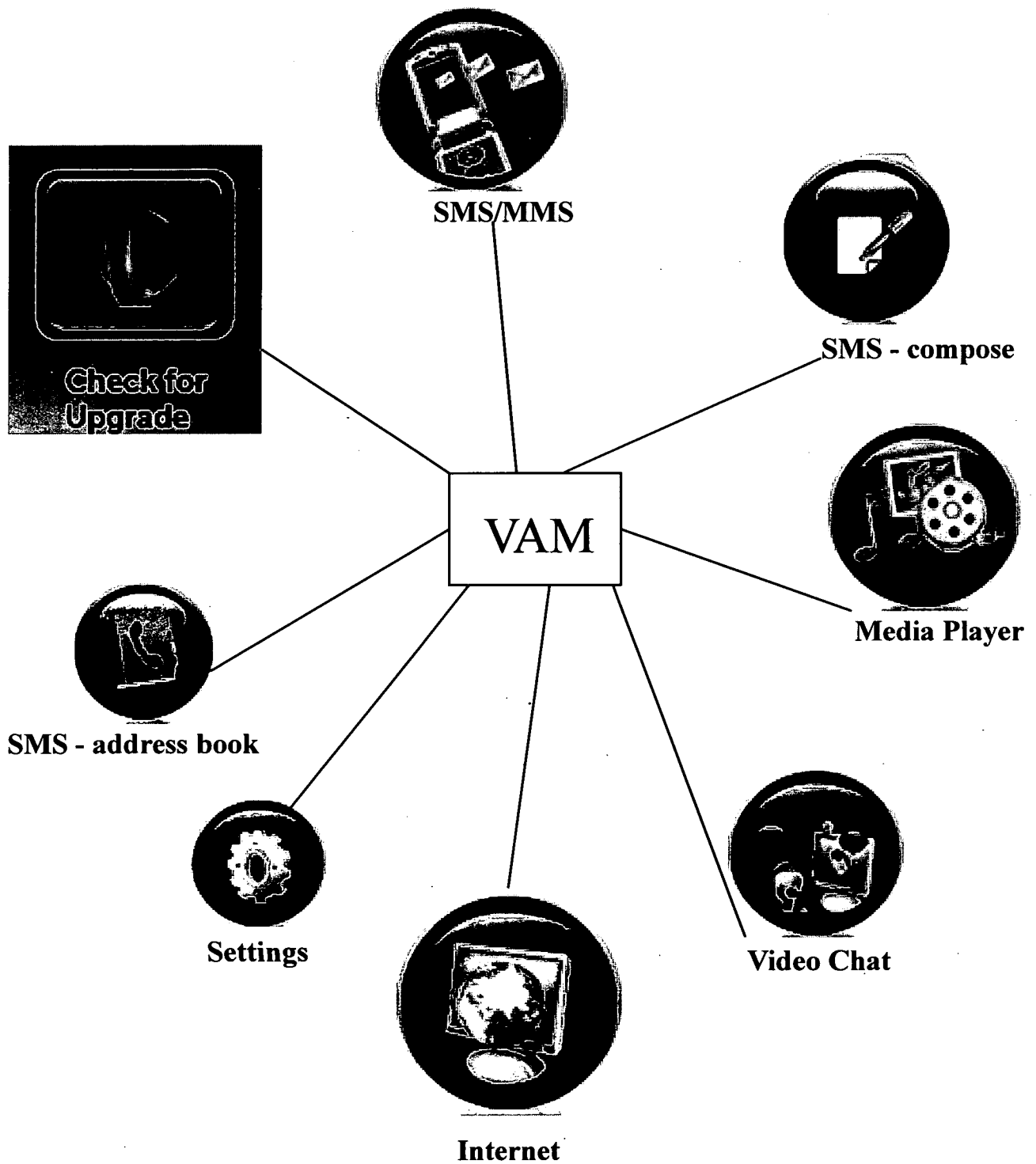
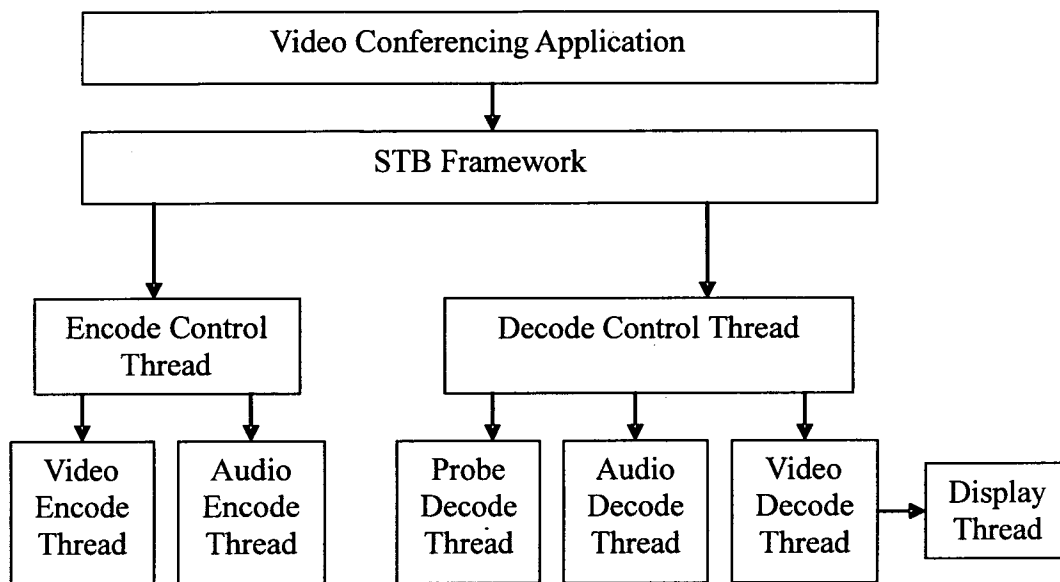
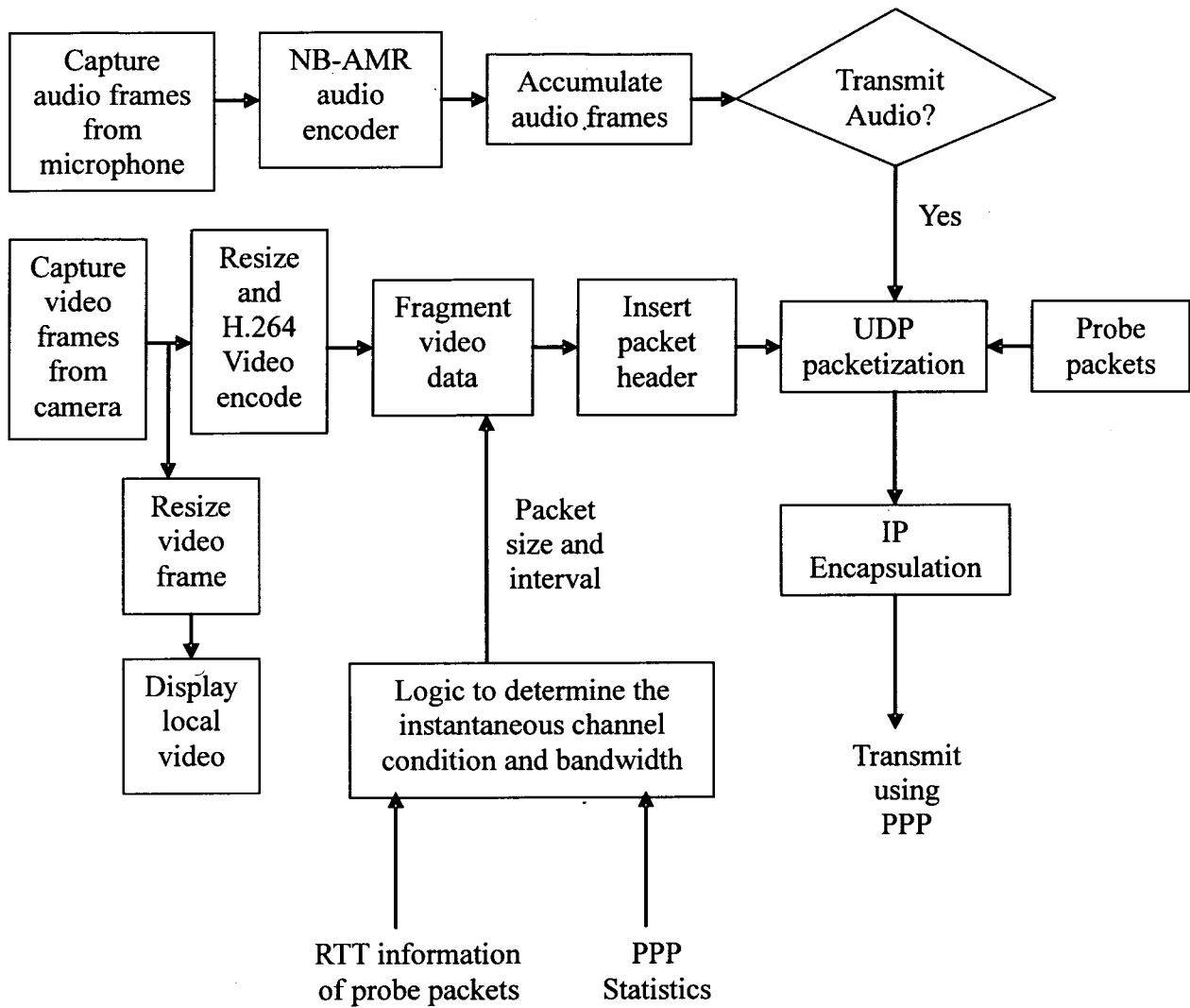


FIGURE 8

**FIGURE 9**

**FIGURE 10**

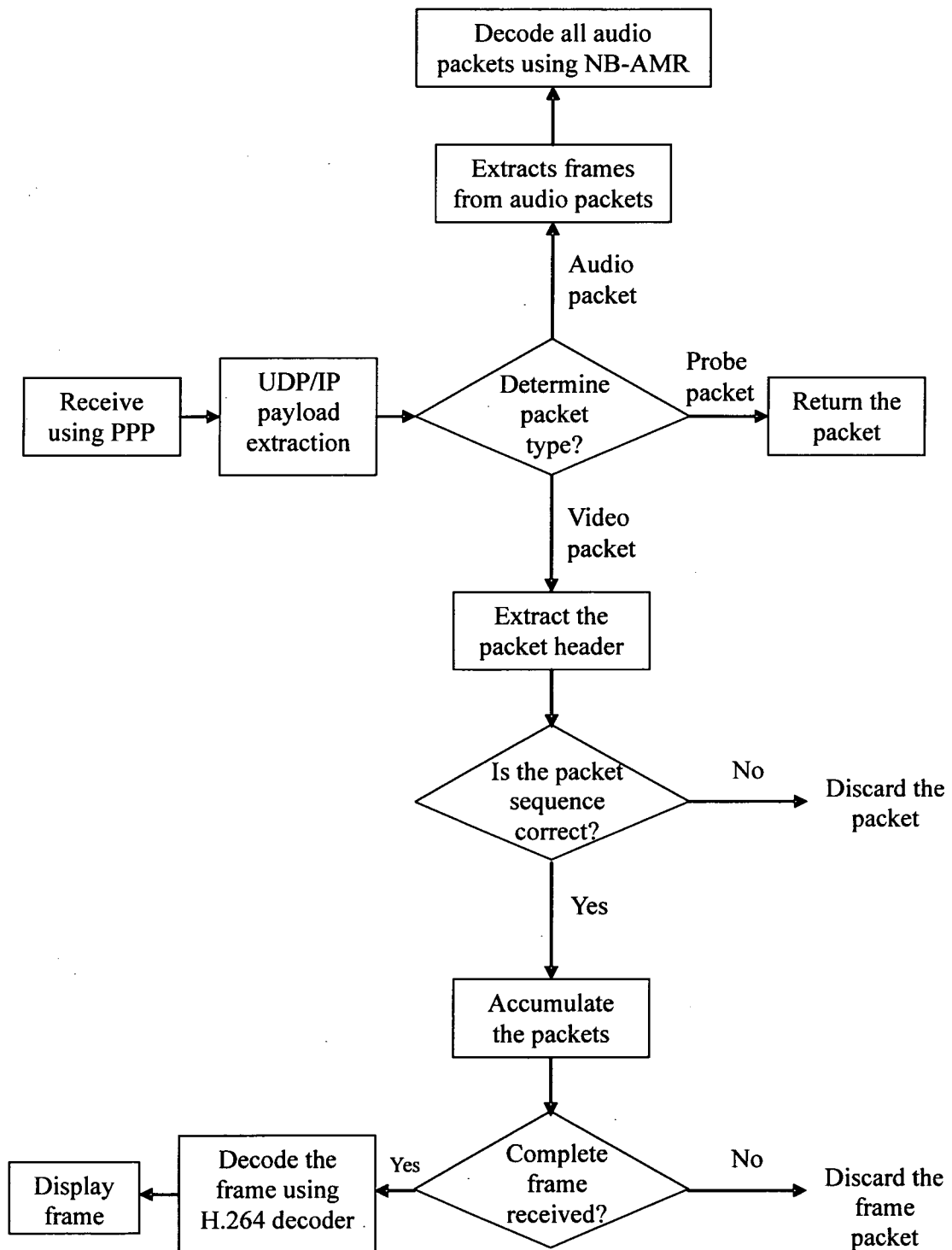
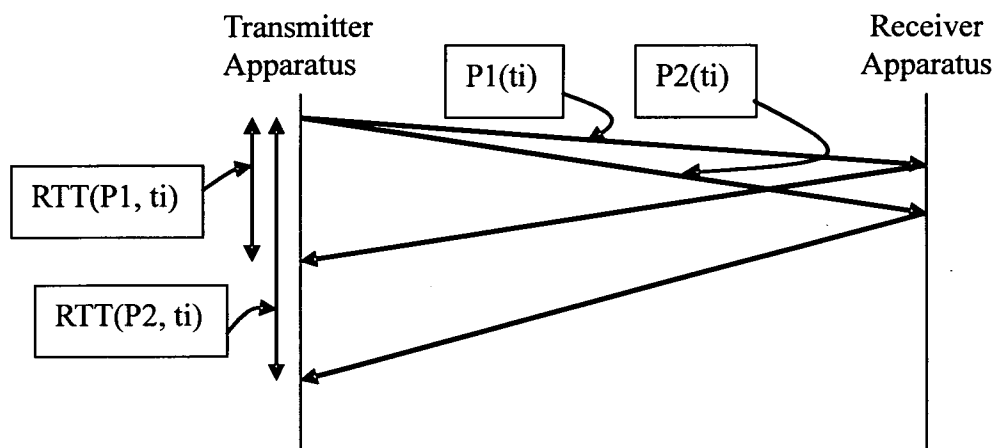


FIGURE 11

**FIGURE 12**

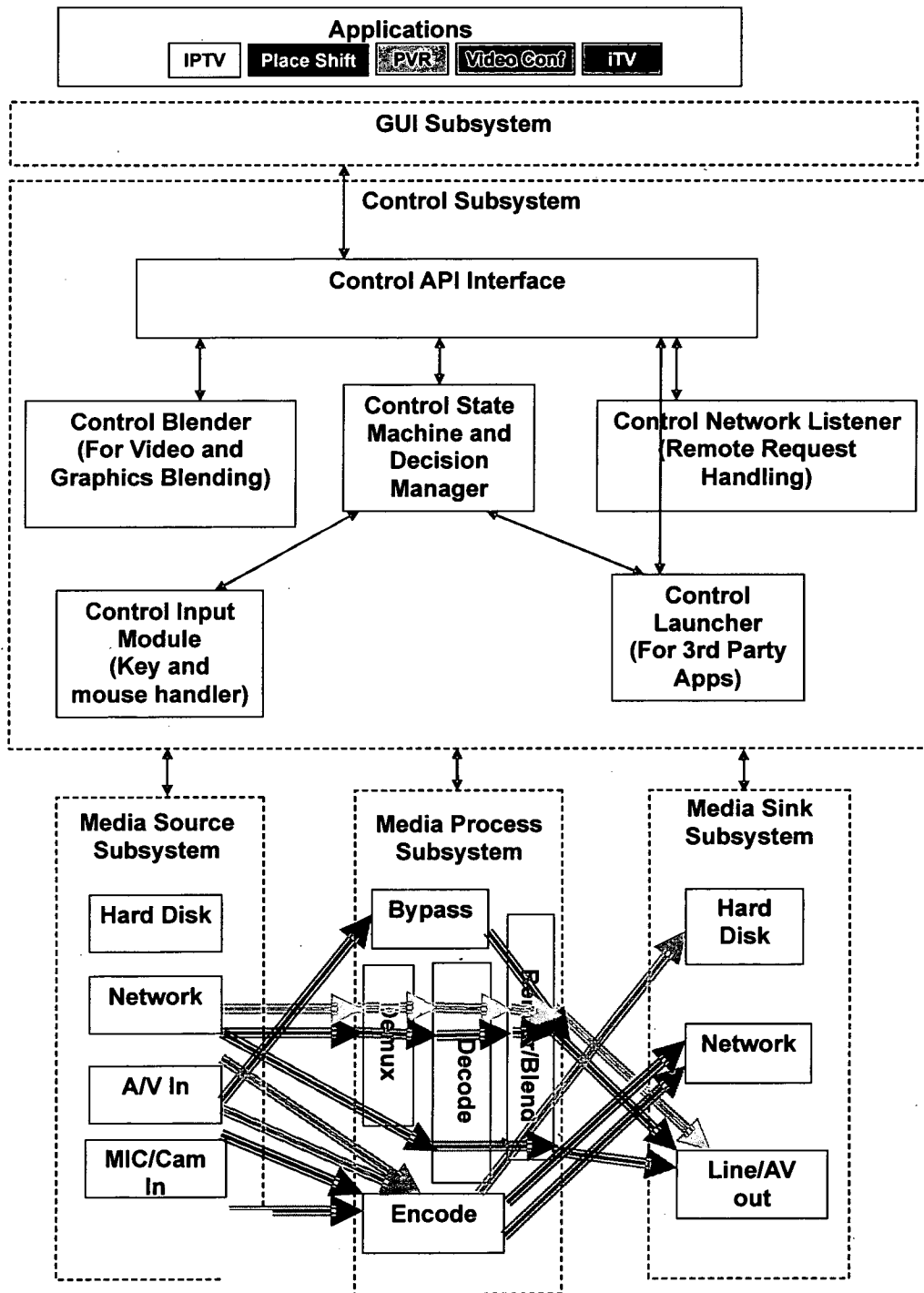
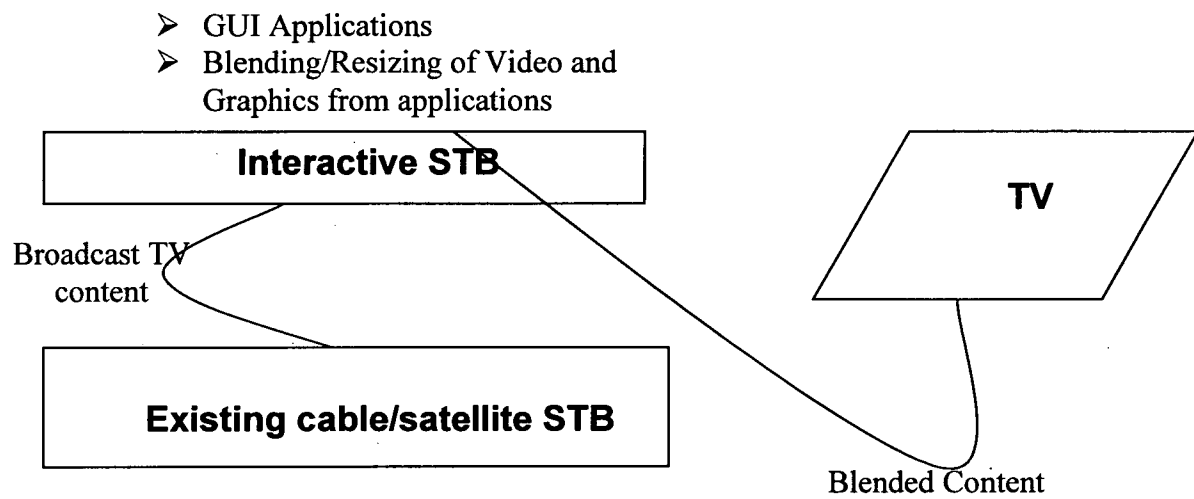


FIGURE 13

**FIGURE 14**