



US 20060248210A1

(19) **United States**(12) **Patent Application Publication**
Kenoyer(10) **Pub. No.: US 2006/0248210 A1**(43) **Pub. Date: Nov. 2, 2006**(54) **CONTROLLING VIDEO DISPLAY MODE IN
A VIDEO CONFERENCING SYSTEM****Publication Classification**(51) **Int. Cl.**
G06F 15/16 (2006.01)(75) Inventor: **Michael L. Kenoyer**, Austin, TX (US)(52) **U.S. Cl.** **709/231**

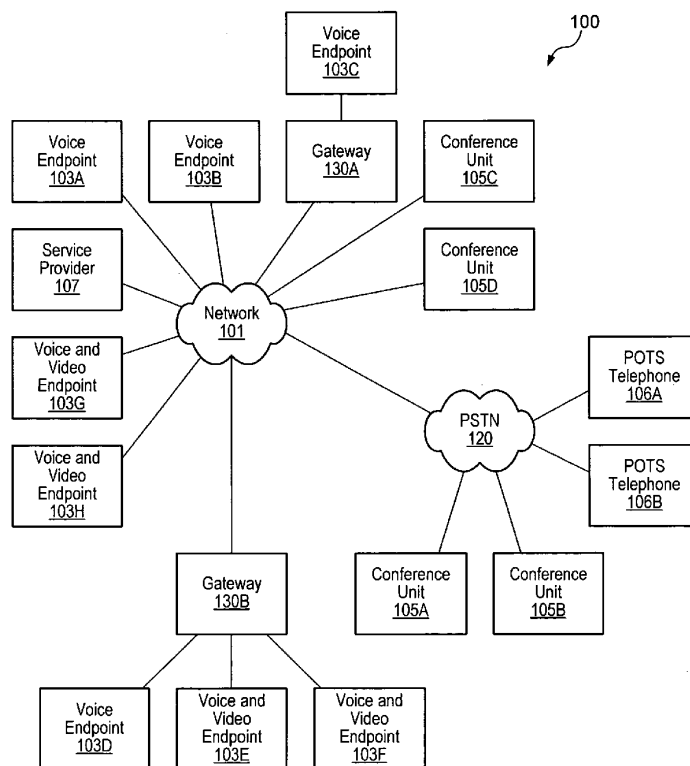
Correspondence Address:

**MEYERTONS, HOOD, KIVLIN, KOWERT &
GOETZEL, P.C.**
700 LAVACA, SUITE 800
AUSTIN, TX 78701 (US)(57) **ABSTRACT**

System and method for controlling video display modes in a video conferencing system. An audio signal from each of a plurality of video conferencing system locations may be received. An accumulated amount of audio signal may be determined from each of one or more of the audio signals. Subsequently, a display mode of two or more possible display modes may be determined for at least one of the video conferencing system locations based on the determined accumulated audio signal. Determining the accumulated audio signal may comprise determining a signal metric for each of one or more of the audio signals using an integrated form of the signal. The method may include comparing accumulated amounts of audio signal from one or more audio signals with at least one accumulation threshold. The display mode may also be determined based on the comparison between the accumulated audio signal and at least one accumulation threshold.

(73) Assignee: **LifeSize Communications, Inc.**(21) Appl. No.: **11/348,217**(22) Filed: **Feb. 6, 2006****Related U.S. Application Data**

(60) Provisional application No. 60/676,918, filed on May 2, 2005.



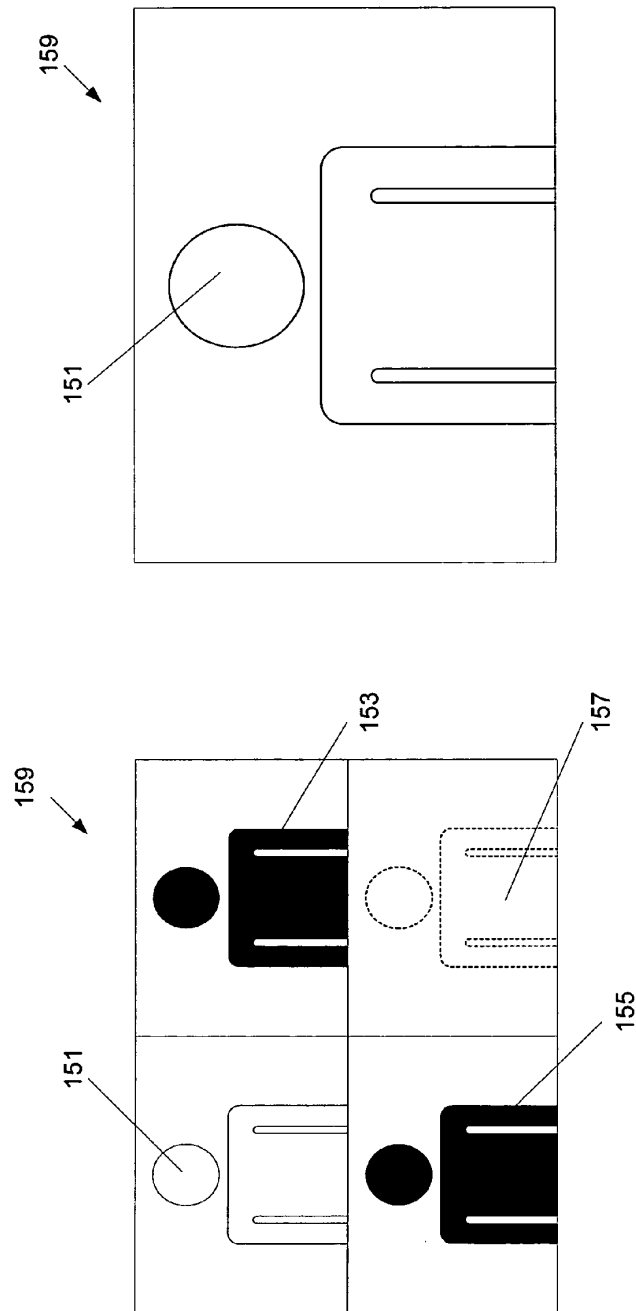


FIG. 1b (Prior Art)

FIG. 1a (Prior Art)

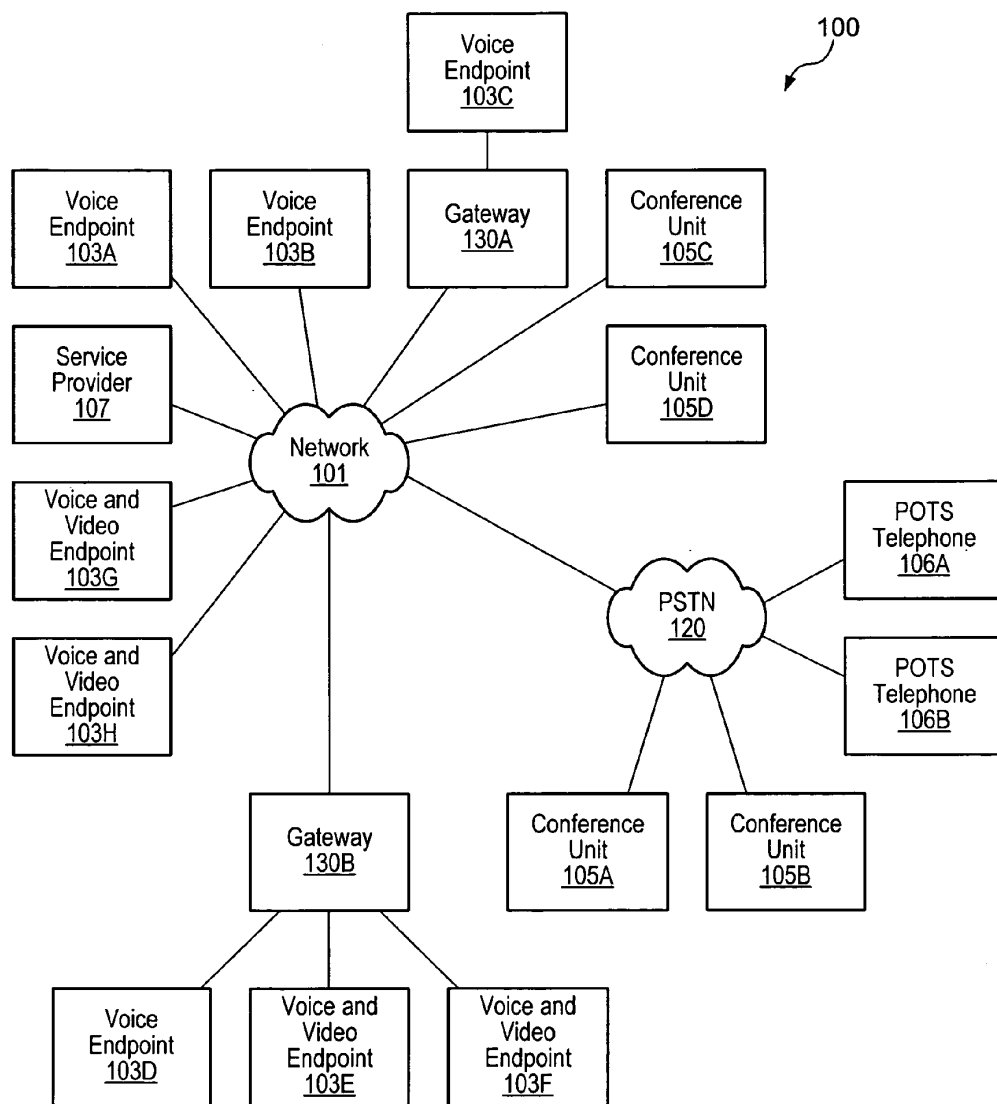


FIG. 2

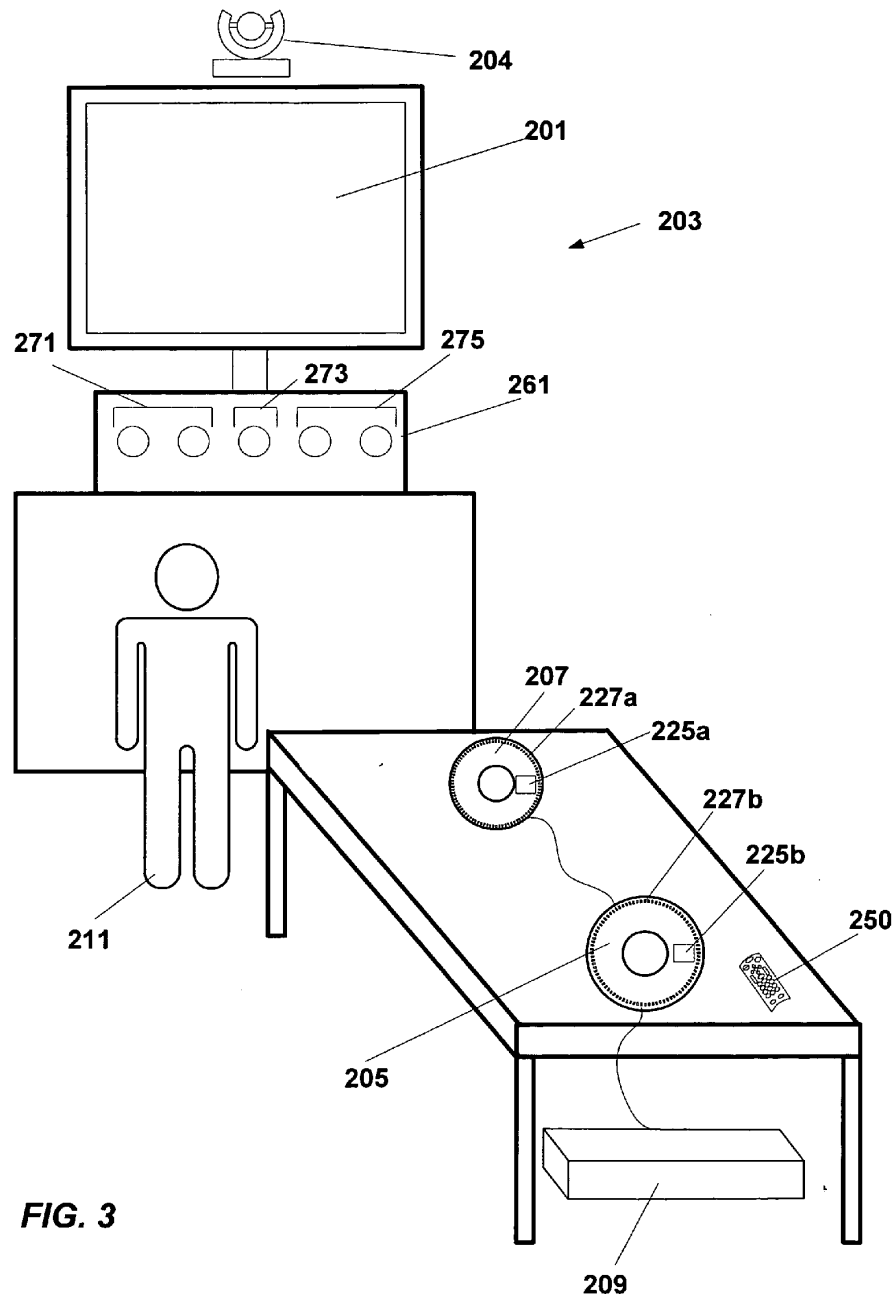


FIG. 3

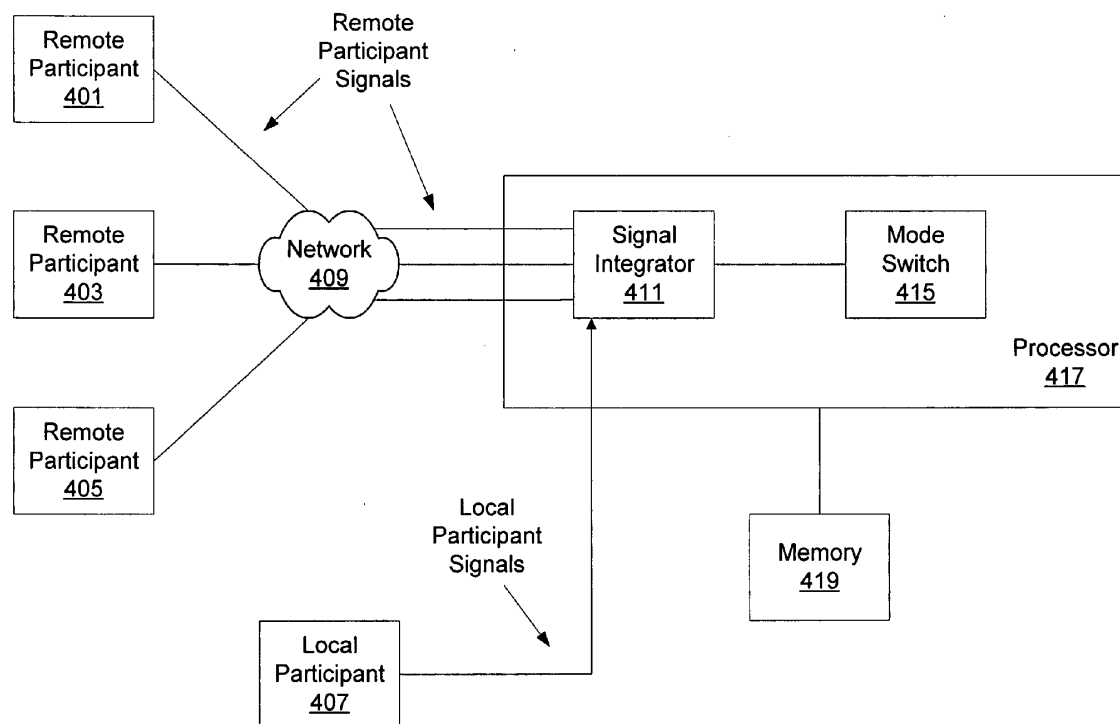


FIG. 4

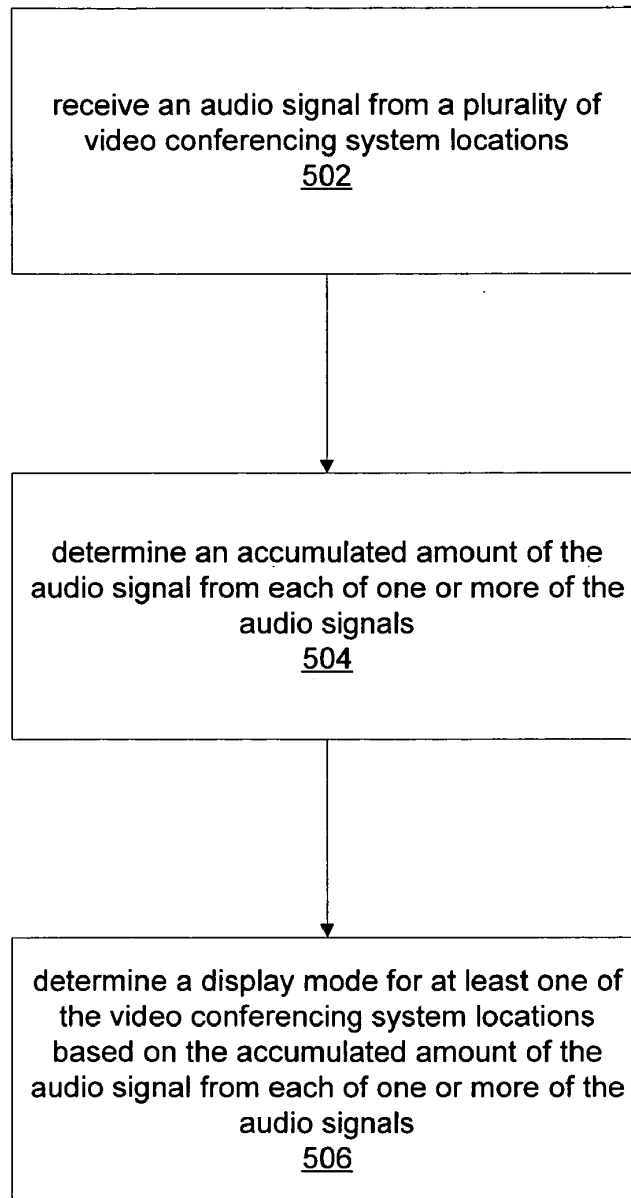


FIG. 5

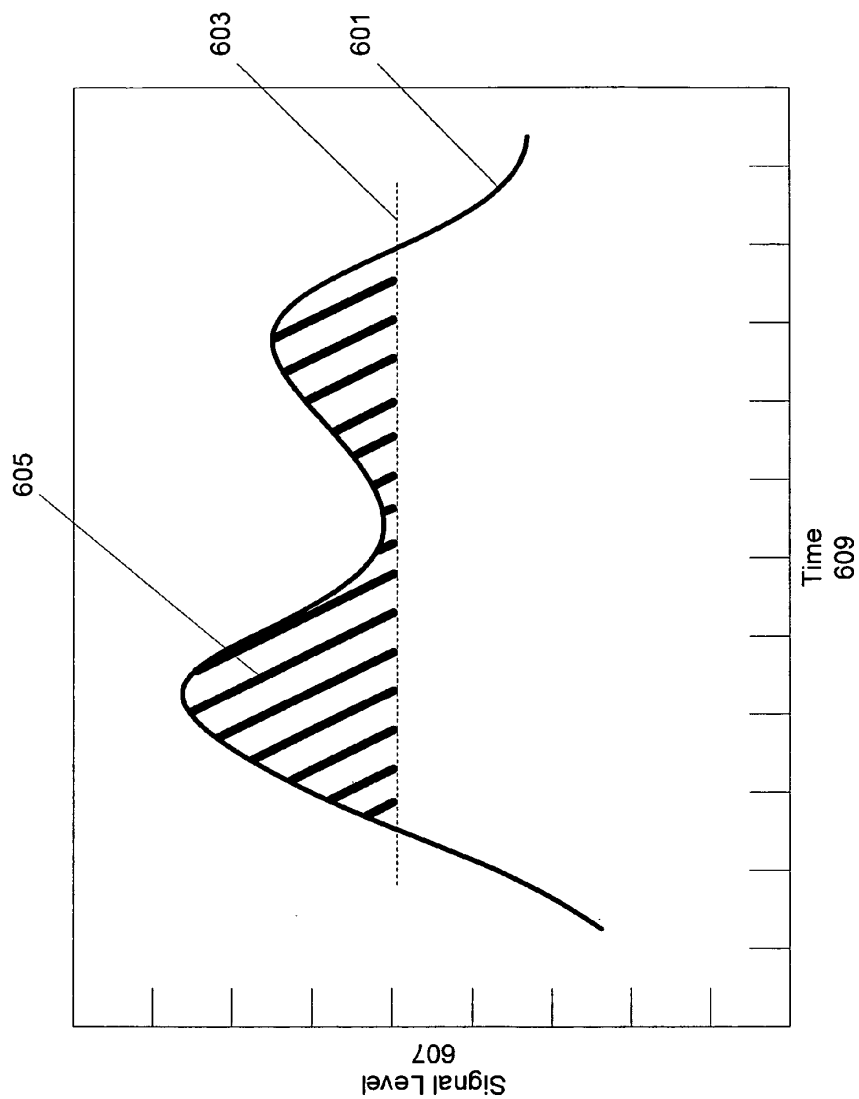
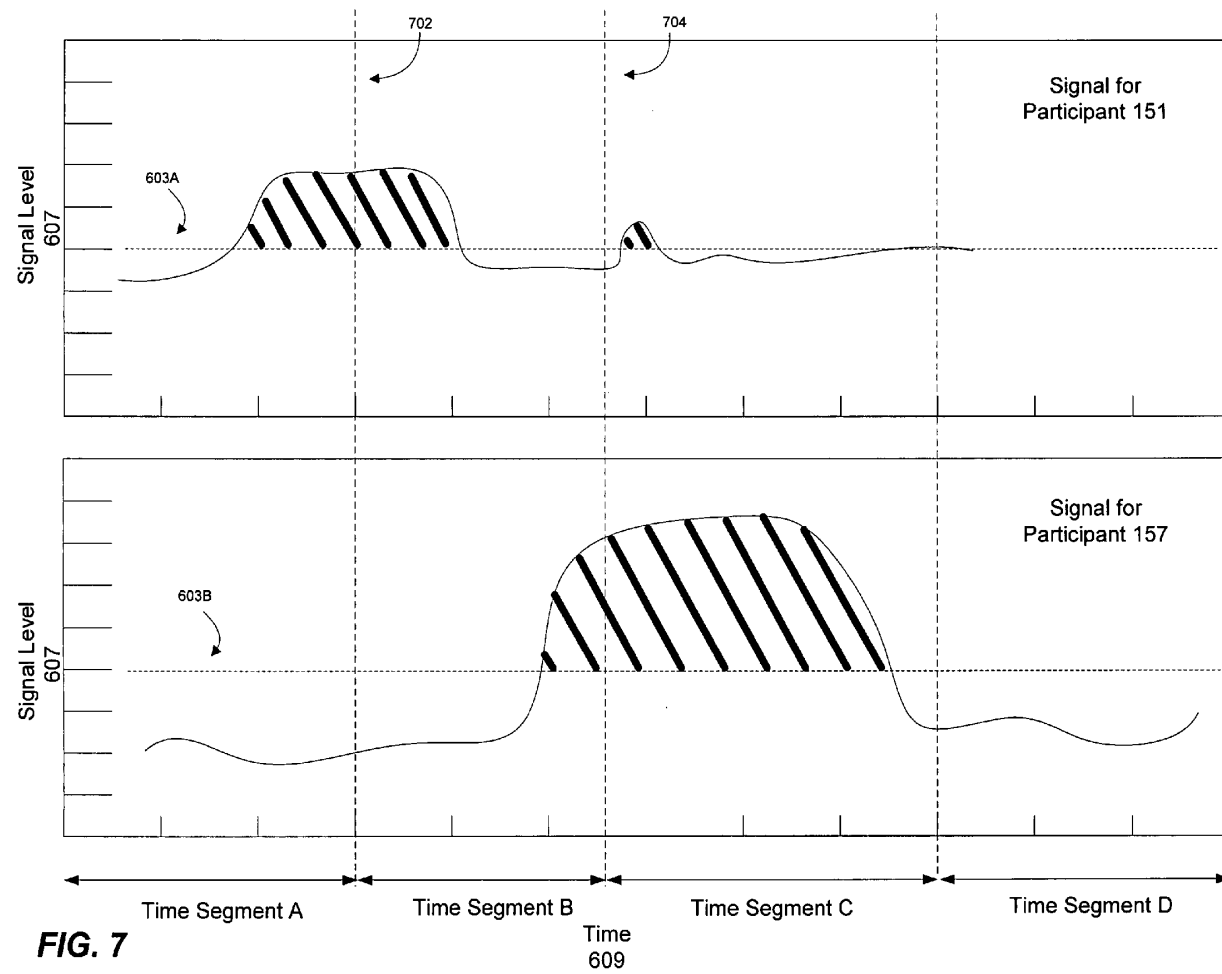


FIG. 6



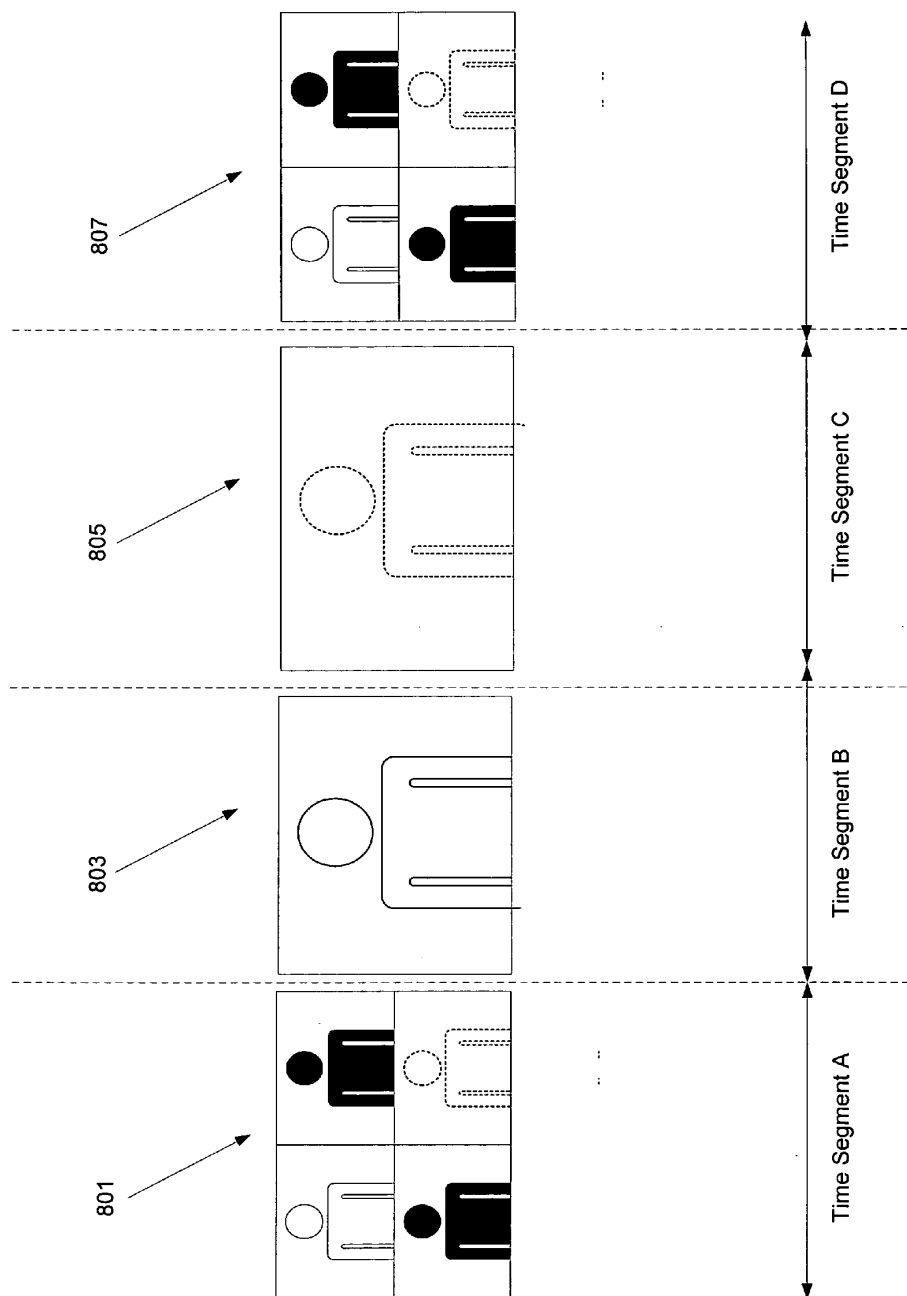


FIG. 8

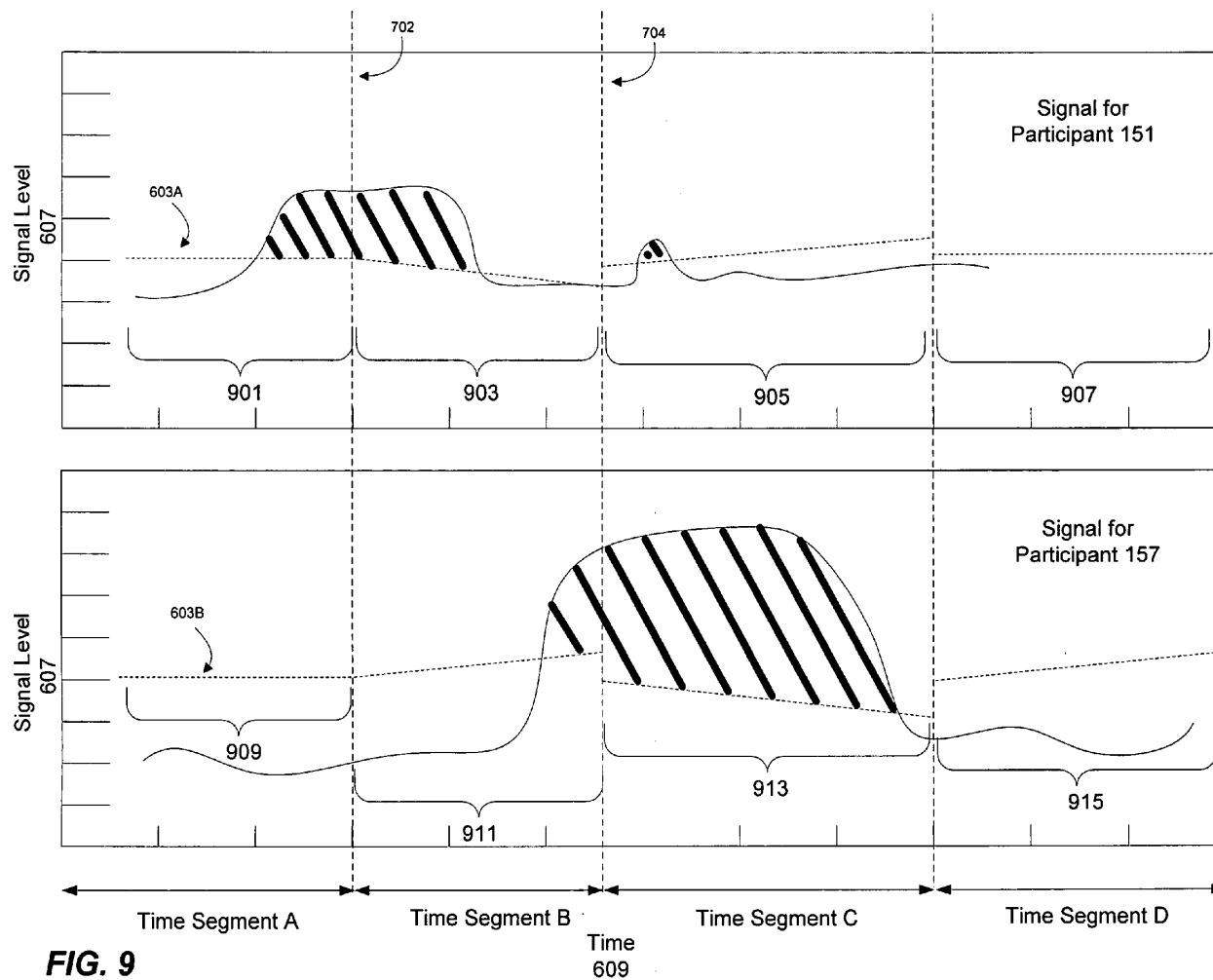


FIG. 9

Other Participant <u>1001a</u>	Other Participant <u>1001b</u>	Other Participant <u>1001c</u>
Other Participant <u>1001d</u>	Dominant Participant <u>1003</u>	Other Participant <u>1001e</u>
Other Participant <u>1001f</u>	Other Participant <u>1001g</u>	Other Participant <u>1001h</u>

FIG. 10a

Other Participant <u>1111a</u>	Other Participant <u>1111b</u>	Other Participant <u>1111c</u>	Other Participant <u>1111d</u>
Other Participant <u>1111e</u>	Dominant Participant <u>1113</u>		Other Participant <u>1111f</u>
Other Participant <u>1111g</u>			Other Participant <u>1111h</u>
Other Participant <u>1111i</u>	Other Participant <u>1111j</u>	Other Participant <u>1111k</u>	Other Participant <u>1111l</u>

FIG. 10b

Dominant Participant <u>1123</u>			Other Participant <u>1121a</u>
			Other Participant <u>1121b</u>
			Other Participant <u>1121c</u>
Other Participant <u>1121d</u>	Other Participant <u>1121e</u>	Other Participant <u>1121f</u>	Other Participant <u>1121g</u>

FIG. 10c

CONTROLLING VIDEO DISPLAY MODE IN A VIDEO CONFERENCING SYSTEM

PRIORITY CLAIMS

[0001] This application claims priority to U.S. Provisional Application No. 60/676,918 titled "Audio and Video Conferencing", which was filed May 2, 2005, whose inventors are Michael L. Kenoyer, Wayne Mock and Patrick D. Vanderwilt which is hereby incorporated by reference in its entirety as though fully and completely set forth herein.

BACKGROUND OF THE INVENTION

[0002] 1. Field of the Invention

[0003] The present invention relates generally to video conferencing and, more specifically, to automatically switching between display modes within a video conference.

[0004] 2. Description of the Related Art

[0005] Video conferencing may be used to allow two or more people to communicate using both video and audio. A video conferencing system may include a camera and microphone at each participant's location to collect video and audio from a respective participant to send to the other participant(s). A speaker and display at each respective participant location may reproduce the audio and video, respectively, from the other participant(s). The video conferencing system may also allow for use of a computer system to allow additional functionality into the video conference, such as data conferencing (including displaying and/or modifying a document for participants during the conference).

[0006] A video conferencing system may support multiple video display modes. In a continuous presence mode, a plurality or all of the participants may be presented on the display at a respective location, as shown in **FIG. 1a**. Thus, continuous presence mode allows a viewer to see a plurality or all of the participants, whose images are typically tiled on the display as shown in **FIG. 1a**. In a single speaker display mode, a participant may view video of the currently talking speaker, as shown in **FIG. 1b**.

[0007] It may be desirable for a video conferencing system to automatically switch the display between a single speaker mode and a continuous presence mode. For example, U.S. Pat. No. 6,744,460 (the '460 Patent) titled "Video Display Mode Automatic Switching System and Method" relates to a system that uses a timer to determine how long a participant has been speaking. When a respective participant has been speaking for a length of time greater than a threshold, as determined by the timer, the system may switch to single speaker mode displaying that respective participant. When no participants are speaking for greater than a time threshold, then the system displays video signals of all of the participants in continuous presence mode. The '460 Patent teaches the "duration of the signals from each of the endpoints are continuously monitored by the timer . . ." Based on the duration of these signals, the system switches between single speaker mode and multiple speaker mode.

[0008] The method described in the '460 Patent has several disadvantages. For example, the system of the '460 patent only considers speaking time, and does not consider the intensity or amplitude of the participants' voices. For

example, if one of the participants begins talking more loudly or shouting during the conference, the system or the '460 Patent will take as long to switch to that person as to switch to someone who is quietly talking. It would be desirable to provide a video conferencing system that more intelligently switches between single speaker and continuous presence mode.

SUMMARY OF THE INVENTION

[0009] In various embodiments, a video conferencing system switches between single speaker and continuous presence mode based on the amount of accumulated audio signal of various ones of the participants. For example, when a first speaker begins speaking, the method may begin accumulating, e.g., via integration, the audio signal of the first speaker. When the accumulated audio signal of the first speaker becomes greater than a certain accumulation threshold, the video conferencing system may automatically switch to single speaker mode presenting the video image of the first speaker. Thus, if the first speaker is speaking more loudly or even yelling during the video conference, the system may switch to single speaker mode faster than if the first speaker were talking normally. Conversely, if the first speaker begins speaking softly, the system may switch to single speaker mode after a greater amount of time has passed. Thus, the method does not switch between video display modes based on time, but rather switches based on the amount of accumulated audio signal of respective participants.

[0010] In some embodiments, the system may receive audio signals from a plurality of participants in a video conference. An audio signal may be generated by a single speaker at a respective participant location or by multiple speakers at that participant location. The accumulated amount of the audio signal may then be determined from each of one or more of the audio signals. Determining the accumulated amount of audio signal may be performed by determining a signal metric for each of one or more of the respective audio signals using an integrated form of the respective signal. More specifically, determining the accumulated amount of audio signal may include integrating each of the one or more audio signals from the plurality of video conferencing systems to generate respective accumulated amounts of audio signal. In some embodiments, the signal metric may be constrained to utilize certain types of audio signals, such as human voices and/or to reject other types of audio signals, such as fan noise or paper shuffling.

[0011] Said another way, the system may operate to analyze incoming signals in order to determine the accumulated amount of audio signal for each participant or participant location. In some embodiments, the signal may be manipulated through various available methods to provide desirable processed signals. For example, incoming audio signals may be processed such that they are always positive. The signals may be integrated using any suitable methods for determining an accumulated amount of audio signal.

[0012] In some embodiments, the signals may only be processed and/or integrated when exceeding a minimum audio level. The level above which the signal may be integrated is herein referred to as an audio threshold. Thus, determining the accumulated amount of the audio signal may occur after the audio signal has exceeded an audio threshold.

[0013] In one embodiment, the audio signal may be accumulated only while the audio signal is continuous and uninterrupted, or substantially uninterrupted. In other words, the accumulation of a respective audio signal may be restarted each time the audio signal stops, e.g., when the level of the respective audio signal goes below the audio threshold for a certain time period or accumulation amount. Said another way, the system may begin accumulating an audio signal when the speaker begins to talk and end the accumulation of the audio signal when the respective speaker stops speaking or is interrupted. Thus, in a video conference with a lot of “back and-forth” talking, where the participants do not exceed their respective accumulation thresholds before being interrupted, the system may remain in continuous presence mode.

[0014] In some embodiments, an interruption may have to exceed an interruption threshold to end the accumulation of the audio signal of the currently speaking participant. For example, in a video conference where one participant begins to speak, and another participant coughs or interjects a brief comment, e.g., “yes”, “I agree”, etc., the system may continue to integrate the speaking participant’s signal because the noise or comment from the other participant did not exceed the interruption threshold. Thus, interjections below the interjection threshold may not hinder the system from switching from the previous display mode, e.g., continuous presence mode, to the new display mode, e.g., the single window display of the currently speaking participant. Thus, the system may intelligently filter interruptions and integrate audio signals in a desirable manner. The interruption threshold may be based on the accumulated audio signal of the interruption or may be time based.

[0015] In some embodiments, a display mode from two or more possible display modes for at least one of the video conferencing system locations may be determined based on the accumulated amount of the audio signal from each of one or more of the audio signals. In other words, the system may choose from a plurality of display modes for each of the participants based on the uninterrupted accumulated amount of audio signal being generated by the participants. In some embodiments, the possible display modes comprise a single window display mode and a multiple window (continuous presence) display mode. The multiple window display mode may comprise a display with a subset or all of the participants in the video conference as will be described in more detail below.

[0016] The method may also include comparing an accumulated amount of the audio signal from one or more of the audio signals with at least one accumulation threshold, where the display mode may be determined based on the comparing. For example, if a participant begins to talk, the system may switch the other participants’ displays to the speaking participant only after the speaking participant has accumulated enough audio signal to exceed the accumulation threshold. The accumulation threshold will be discussed in more detail hereinbelow.

[0017] In some embodiments, if the accumulated amount of audio signal corresponding to a first location exceeds an accumulation threshold, video signals from the first location may be displayed on each of a plurality of video conferencing systems in the single window mode. In other words, if a participant’s accumulated signal exceeds some value, e.g.,

if the participant speaks enough to surpass his respective accumulation threshold, each of the other participants, i.e., the listening participants, may view that single speaker. The talking participant, however, may view a continuous presence mode, e.g., he may see all of the other participants or, alternatively, a subset therefrom.

[0018] Alternatively, if the accumulated amount of audio signal corresponding to any location does not exceed the accumulation threshold, video signals from a plurality of locations may be displayed on each of a plurality of video conferencing systems in a continuous presence mode. Said another way, if no one in the video conference is speaking in an uninterrupted manner for a certain threshold amount of audio signal (e.g., energy of the audio signal), the participants may view a continuous presence display mode comprising a subset or all of the participants on their display.

[0019] In one embodiment, if the accumulated amount of audio signal corresponding to a subset of locations repeatedly exceeds the accumulation threshold, video signals from that respective subset of locations may be displayed on each of a plurality of video conferencing systems in a continuous presence mode. In other words, if participants from a certain subset of participant locations are doing all of the talking, i.e., exceeding a common (or respective) accumulation threshold(s), this subset of the talking participants may be displayed on each of the participants’ displays. Alternatively, the participants’ displays may show each of the talking participants singly, and intelligently switch between each of the talking participants throughout the conversation.

[0020] In embodiments utilizing an accumulation threshold, the method may also include modifying, e.g., raising, a respective accumulation threshold for a video conferencing system when an accumulated amount of audio signal has not exceeded the respective accumulation threshold within a predetermined amount of time. The method may also modify, e.g., lower, a respective accumulation threshold for a video conferencing system when an accumulated amount of audio signal has recently exceeded the respective accumulation threshold within a predetermined amount of time. In other words, the accumulation thresholds may be variable, i.e. may dynamically change, throughout the duration of the video conference. For example, the accumulation threshold variables may vary differently depending on whether the respective participant has spoken within some predetermined amount of time.

[0021] The accumulation thresholds may also vary with respect to each participant, i.e., each participant may have his own threshold that may vary independently from the other participants’ thresholds. In one embodiment, each participant’s threshold may be normalized with respect to the average audio level of each participant. For example, quieter participants may have lower thresholds than louder participants. Such an example will be described in more detail hereinbelow.

BRIEF DESCRIPTION OF THE DRAWINGS

[0022] A better understanding of the present invention may be obtained when the following detailed description is considered in conjunction with the following drawings, in which:

[0023] **FIGS. 1a and 1b** illustrate examples of continuous presence and single speaker modes for video conference displays;

[0024] FIG. 2 illustrates a video conferencing system, according to one embodiment;

[0025] FIG. 3 illustrates a participant location or conferencing unit, according to one embodiment;

[0026] FIG. 4 illustrates a network and local system for use in video conferencing, according to one embodiment;

[0027] FIG. 5 is a flowchart illustrating an exemplary method for controlling video display modes in a video conferencing system, according to one embodiment;

[0028] FIG. 6 illustrates an audio signal integrated above a threshold, according to one embodiment;

[0029] FIG. 7 illustrates two respective audio signals integrated above a fixed threshold, according to one embodiment;

[0030] FIG. 8 illustrates a display mode according to the integrated audio signals, according to one embodiment;

[0031] FIG. 9 illustrates two respective audio signals integrated above a variable audio threshold, according to one embodiment; and

[0032] FIGS. 10a-c illustrate various embodiments of continuous presence screens.

[0033] While the invention is susceptible to various modifications and alternative forms, specific embodiments thereof are shown by way of example in the drawings and will herein be described in detail. It should be understood, however, that the drawings and detailed description thereto are not intended to limit the invention to the particular form disclosed, but on the contrary, the intention is to cover all modifications, equivalents, and alternatives falling within the spirit and scope of the present invention as defined by the appended claims. Note, the headings are for organizational purposes only and are not meant to be used to limit or interpret the description or claims. Furthermore, note that the word “may” is used throughout this application in a permissive sense (i.e., having the potential to, being able to), not a mandatory sense (i.e., must). The term “include”, and derivations thereof, mean “including, but not limited to”. The term “coupled” means “directly or indirectly connected”.

DETAILED DESCRIPTION OF THE EMBODIMENTS

INCORPORATION BY REFERENCE

[0034] U.S. Pat. No. 6,744,460 titled “Video Display Mode Automatic Switching System and Method” is hereby incorporated by reference as though fully and completely set forth herein.

[0035] U.S. Patent Application titled “Speakerphone”, Ser. No. 11/251,084, which was filed Oct. 14, 2005, whose inventor is William V. Oxford is hereby incorporated by reference in its entirety as though fully and completely set forth herein.

[0036] U.S. Patent Application titled “Video Conferencing System Transcoder”, Ser. No. 11/252,238, which was filed Oct. 17, 2005, whose inventors are Michael L. Kenoyer and Michael V. Jenkins, is hereby incorporated by reference in its entirety as though fully and completely set forth herein.

[0037] U.S. Patent Application titled “Speakerphone Supporting Video and Audio Features”, Ser. No. 11/251,086, which was filed Oct. 14, 2005, whose inventors are Michael L. Kenoyer, Craig B. Malloy and Wayne E. Mock is hereby incorporated by reference in its entirety as though fully and completely set forth herein.

[0038] U.S. Patent Application titled “High Definition Camera Pan Tilt Mechanism”, Ser. No. 11/251,083, which was filed Oct. 14, 2005, whose inventors are Michael L. Kenoyer, William V. Oxford, Patrick D. Vanderwilt, Hans-Christoph Haenlein, Branko Lukic and Jonathan I. Kaplan, is hereby incorporated by reference in its entirety as though fully and completely set forth herein.

FIG. 2—Video Conferencing System

[0039] FIG. 2 illustrates an embodiment of a video conferencing system 100. Video conferencing system 100 may include a network 101, endpoints 103A-103H (e.g., audio and/or video conferencing systems), gateways 130A-130B, a service provider 107 (e.g., a multipoint control unit (MCU)), a public switched telephone network (PSTN) 120, conference units 105A-105D, and plain old telephone system (POTS) telephones 106A-106B. Endpoints 103C and 103D-103H may be coupled to network 101 via gateways 130A and 130B, respectively, and gateways 130A and 130B may each include a firewall, a network address translator (NAT), a packet filter, and/or proxy mechanisms, among others. Conference units 105A-105B and POTS telephones 106A-106B may be coupled to network 101 via PSTN 120. In some embodiments, conference units 105A-105B may each be coupled to PSTN 120 via an Integrated Services Digital Network (ISDN) connection, and each may include and/or implement H.320 capabilities. In various embodiments, video and audio conferencing may be implemented over various types of networked devices.

[0040] In some embodiments, endpoints 103A-103H, gateways 130A-130B, conference units 105C-105D, and service provider 107 may each include various wireless or wired communication devices that implement various types of communication, such as wired Ethernet, wireless Ethernet (e.g., IEEE 802.11), IEEE 802.16, paging logic, RF (radio frequency) communication logic, a modem, a digital subscriber line (DSL) device, a cable (television) modem, an ISDN device, an ATM (asynchronous transfer mode) device, a satellite transceiver device, a parallel or serial port bus interface, and/or other type of communication device or method.

[0041] In various embodiments, the methods and/or systems described may be used to implement connectivity between or among two or more participant locations or endpoints, each having voice and/or video devices (e.g., endpoints 103A-103H, conference units 105A-105D, POTS telephones 106A-106B, etc.) that communicate through various networks (e.g., network 101, PSTN 120, the Internet, etc.).

[0042] Endpoints 103A-103C may include voice conferencing capabilities and include or be coupled to various audio devices (e.g., microphones, audio input devices, speakers, audio output devices, telephones, speaker telephones, etc.). Endpoints 103D-103H may include voice and video communications capabilities (e.g., video conferencing capabilities) and include or be coupled to various audio

devices (e.g., microphones, audio input devices, speakers, audio output devices, telephones, speaker telephones, etc.) and include or be coupled to various video devices (e.g., monitors, projectors, displays, televisions, video output devices, video input devices, cameras, etc.). In some embodiments, endpoints 103A-103H may comprise various ports for coupling to one or more devices (e.g., audio devices, video devices, etc.) and/or to one or more networks.

[0043] Conference units 105A-105D may include voice and/or video conferencing capabilities and include or be coupled to various audio devices (e.g., microphones, audio input devices, speakers, audio output devices, telephones, speaker telephones, etc.) and/or include or be coupled to various video devices (e.g., monitors, projectors, displays, televisions, video output devices, video input devices, cameras, etc.). In some embodiments, endpoints 103A-103H and/or conference units 105A-105D may include and/or implement various network media communication capabilities. For example, endpoints 103A-103H and/or conference units 105C-105D may each include and/or implement one or more real time protocols, e.g., session initiation protocol (SIP), H.261, H.263, H.264, H.323, among others.

[0044] In various embodiments, a codec may implement a real time transmission protocol. In some embodiments, a codec (which may be short for “compressor/decompressor”) may comprise any system and/or method for encoding and/or decoding (e.g., compressing and decompressing) data (e.g., audio and/or video data). For example, communication applications may use codecs to convert an analog signal to a digital signal for transmitting over various digital networks (e.g., network 101, PSTN 120, the Internet, etc.) and to convert a received digital signal to an analog signal. In various embodiments, codecs may be implemented in software, hardware, or a combination of both. Some codecs for computer video and/or audio may include MPEG, Indeo, and Cinepak, among others.

[0045] At least one of the participant locations may include a camera for acquiring high resolution or high definition (e.g., HDTV compatible) signals. At least one of the participant locations may include a high definition display (e.g., an HDTV display), for displaying received video signals in a high definition format. In one embodiment the network 101 may be 1.5 MB or less (e.g., T1 or less). In another embodiment, the network is 2 MB or less.

FIG. 3—Participant Location

[0046] FIG. 3 illustrates an embodiment of a participant location, also referred to as an endpoint or conferencing unit (e.g., a video conferencing system). In some embodiments, the video conference system may have a system codec 209 to manage both a speakerphone 205/207 and a video conferencing system 203. For example, a speakerphone 205/207 and a video conferencing system 203 may be coupled to the integrated video and audio conferencing system codec 209 and may receive audio and/or video signals from the system codec 209.

[0047] In some embodiments, the participant location may include a high definition camera 204 for acquiring high definition images of the participant location. The participant location may also include a high definition display 201 (e.g., a HDTV display). High definition images acquired by the camera may be displayed locally on the display and may also be encoded and transmitted to other participant locations in the video conference.

[0048] The participant location may also include a sound system 261. The sound system 261 may include multiple speakers including left speakers 271, center speaker 273, and right speakers 275. Other numbers of speakers and other speaker configurations may also be used. In some embodiments, the video conferencing system may include a camera 204 for capturing video of the conference site. In some embodiments, the video conferencing system may include one or more speakerphones 205/207 which may be daisy chained together.

[0049] The video conferencing system components (e.g., the camera 204, display 201, sound system 261, and speakerphones 205/207) may be coupled to a system codec 209. The system codec 209 may receive audio and/or video data from a network. The system codec 209 may send the audio to the speakerphone 205/207 and/or sound system 261 and the video to the display 201. The received video may be high definition video that is displayed on the high definition display. The system codec 209 may also receive video data from the camera 204 and audio data from the speakerphones 205/207 and transmit the video and/or audio data over the network to another conferencing system. In some embodiments, the conferencing system may be controlled by a participant through the user input components (e.g., buttons) on the speakerphone and/or remote control 250. Other system interfaces may also be used.

[0050] FIG. 4 illustrates an exemplary embodiment of a video conferencing system comprising a plurality of participants located at respective endpoints. As shown, the video conferencing system includes a local participant 407 and one or more remote participants 401, 403 and 405. Each participant 401-407 may be at a respective location or endpoint. Each location may include video conferencing equipment, such as the equipment described regarding FIG. 3.

[0051] The various participants in the video conference may communicate over a transmission medium or network 409. The network 409 may be any of various types suitable for transmission of video and audio data between the participant locations. In one embodiment, the network is or includes a wide area network, such as the Internet. The network 409 may also include various other types of communication systems, such as ISDN (Integrated Services Digital Network), the PSTN (Public Switched Telephone Network), LANs (local area networks) and/or other types of WANs.

[0052] Each of the participants may be coupled to a control unit, e.g., a multipoint control unit (MCU). The MCU may comprise processor 417 and memory 419. In one embodiment, the MCU may be coupled to memory 419 via transmission media. Note that the system and method described herein may utilize suitable types of control units other than the MCU; the MCU is exemplary only, and in fact, other control units are envisioned.

[0053] In some embodiments, the MCU may be comprised in a server. Each of the participant's endpoints may be coupled to the MCU via a network such as network 101. In one embodiment, the server may be an internet hosted web-server capable of providing video conferencing services to end-users.

[0054] Alternatively, at least one of the participant locations may comprise the MCU. The MCU may operate to

receive audio and video signals from each of the participant locations and selectively combine the signals for output to the various participant locations. In some embodiments, the MCU may operate to selectively provide different combinations of signals for different display modes. For example, in a single speaker display mode, where a participant from one location is talking, the MCU may operate to send the video signal of that participant to each of a subset or all of the participant locations. In a continuous presence display mode, where multiple participants are conversing, the MCU may operate to combine the video signals of a subset of the participants and provide this combined signal to each of the participant locations.

[0055] In one embodiment, the system is operable to intelligently select a video display mode based on the received audio signals from one or more of the participant locations.

[0056] FIG. 5 is a flowchart illustrating an exemplary method for controlling video display modes in a video conferencing system, according to one embodiment. It should be noted that in various embodiments of the methods described below, one or more of the elements described may be performed concurrently, in a different order than shown, or may be omitted entirely. Other additional elements may also be performed as desired.

[0057] In 502, an audio signal from each of a plurality of video conferencing system locations may be received. The audio signal may be from a single speaker at a respective party location or from multiple speakers at that party location. In one embodiment, the audio signals may be received by an MCU, and the MCU may be operable to perform the reception via network cables or other transmission media as described above. For example, in FIG. 4, the MCU may receive audio signals from each of local participant 407 and remote participants 401, 403, and 405.

[0058] In 504, an accumulated amount of the audio signal may be determined from each of one or more of the audio signals. Determining the accumulated amount of audio signal may be performed by determining a signal metric for each of one or more of the respective audio signals using an integrated form of the respective signal. More specifically, determining the accumulated amount of audio signal may include integrating each of the one or more audio signals from the plurality of video conferencing systems to generate respective accumulated amounts of audio signal. In some embodiments, the MCU may implement signal integrator 411 to perform the determination of the accumulated amount of the audio signal.

[0059] Said another way, the MCU and coupled components may operate to analyze incoming audio signals in order to determine the accumulated amount of audio signal for each participant or participant location. In some embodiments, the signal may be manipulated through various available methods to provide desirable processed signals. For example, incoming audio signals may be processed such that they are always positive. FIG. 6 illustrates such a signal. As further examples, the absolute value, the root-mean square (rms), or the square of the signal (providing the signal's energy), may be taken to provide positively valued signals. As another example, the signals may be smoothed to facilitate integration or accumulation computations.

[0060] The processed or unprocessed signals may be integrated using any suitable methods for integration. For

example, the signal might be sampled at given lengths or intervals or by other suitable methods, approximated using Riemann, trapezoidal, or Simpson sums, or processed using other appropriate techniques as desired. In some embodiments, the accumulated amount of audio signal may be determined using other methods. For example, the volume or intensity of the signal may be measured via averaging methods, e.g., average amplitude or decibels. Note that in the systems and methods disclosed herein, integration is not limited to those methods described above, and in fact, may refer to any suitable methods for measuring accumulated audio signal. In other words, determining the accumulated amount of audio signal may comprise performing various other signal processing methods on the received audio signal.

[0061] Thus, determining the accumulated amount of audio signal may include integrating (or approximating the integration of) various forms of the signal to provide accumulated energy, power, rms, absolute value, intensity, or other desirable signal metrics of the audio signal. As another example, changes in amplitude may be integrated and/or tracked (e.g., the changes in amplitude of a person's voice may be integrated).

[0062] In some embodiments, the signals may only be processed and/or integrated when exceeding an audio level. More specifically, the signal integrator may begin measuring (or accumulating) the accumulated audio signal once a minimum audio level has been reached. The level above which the signal may be integrated is herein referred to as an audio threshold. FIG. 6 illustrates an exemplary signal exceeding an audio threshold. The signal, shown in FIG. 6 in a signal level 607 versus time 609 plot, exceeds audio threshold 603 and may be integrated over the area 605. As FIG. 6 further shows, signals below the audio threshold, such as 601, may not be integrated. Thus, determining the accumulated amount of the audio signal may occur after the audio signal has exceeded an audio threshold. In some embodiments, the audio signal may only continue to be accumulated while the audio signal remains above the audio threshold without "significant" interruption.

[0063] In one embodiment, the accumulated amount of audio signal from each of one or more of the audio signals may be an uninterrupted accumulated amount of audio signal. In other words; the accumulation of a respective audio signal may be restarted each time the audio signal stops, e.g., when the level of the respective audio signal goes below the audio threshold for a certain time period. Said another way, the system may begin accumulating an audio signal when the speaker begins to talk and end the accumulation of the audio signal when the respective speaker stops speaking or is interrupted. Thus, in a video conference with a lot of back and forth talking where the participants do not exceed their respective accumulation thresholds before being interrupted, the system may remain in continuous presence mode.

[0064] In some embodiments, an interruption may have to exceed an interruption threshold to end the accumulation of the audio signal of the currently speaking participant. In other words, the audio signal may continue to be accumulated as long as no "significant" interruption occurs. For example, in a video conference where one participant begins to speak, and another participant coughs or interjects a brief

comment, e.g., “yes”, “I agree”, etc., the system may continue to integrate the speaking participant’s signal because the noise or comment from the other participant did not exceed the interruption threshold. Thus, interjections below the interjection threshold may not hinder the system from switching from the previous display mode, e.g., continuous presence mode, to the new display mode, e.g., the single window display of the currently speaking participant. Thus, the system may intelligently filter interruptions and integrate audio signals in a desirable manner.

[0065] The interruption threshold may be based on the accumulated audio signal of the interruption, or may be time based. Thus in one embodiment if the accumulated audio signal of the “interruption” is less than an interruption threshold then the “interruption” is ignored, and the audio signal currently being accumulated continues to be accumulated. In another embodiment, if the “interruption” is less than an interruption threshold time period, then the “interruption” is ignored, and the audio signal currently being accumulated continues to be accumulated. As used herein, the term “significant interruption” may refer to an amount of interruption, which in some embodiments is less than or equal to a certain percentage (2%, 4%, 5%, 7%, etc.) of the accumulation threshold for determining display mode. Alternatively, the term “significant interruption” may refer to an amount of accumulated energy equivalent to 2 seconds of normal talking voice, or 1.5 seconds of a raised talking voice.

[0066] In some embodiments, rules may be used (e.g., predetermined and/or provided by a conference participant) to determine when to accumulate energy. Rules may be threshold based. For example, when the audio is below a first threshold, no energy is integrated. When above the first threshold but below a second threshold, a percentage of the audio is integrated, etc. Rules may also be based on how quickly (or slowly) the audio is fluctuating between various thresholds. For example, if a participant’s voice suddenly shifts above a high threshold, the audio may be integrated at a higher percentage (which may exceed 100% in some embodiments). This may allow more emphasis to be given to a participant who suddenly begins shouting. In some embodiments, audio exceeding a threshold may not be integrated above the threshold. For example, the audio may be integrated under the threshold but not over it. This may prevent the system from switching too quickly to naturally loud speakers. In some embodiments, the system may adapt the rules throughout the conference based on factors such as a time averaged participant audio levels.

[0067] In some embodiments, the signal metric may be constrained to utilize certain types of audio signals (such as human voices) and/or to reject other types of audio signals (such as fan noise or paper shuffling). For example, the audio may be processed to detect human voices and the corresponding signal metric may be comprised of the human voice component. This may allow human voices to be tracked and integrated without including extraneous noise. For example, a loud air conditioner switching on at a remote conference site may be ignored by the system because the dominant frequencies of the air conditioner noise do not match human voice frequencies. In some embodiments, the system may integrate only audio of frequencies in a certain range (e.g., a range dominated by human voice). In some embodiments, the system may integrate audio that com-

prises fundamental harmonics (e.g., characteristic of human voice). In some embodiments, the system may identify and track the voices of different participants. In some embodiments, different weights may be given to different voices for the integration. For example, the voice of the leader of the conference may be weighted during integration so the system switches to him/her (or stays on them) more often.

[0068] In 506, a display mode from two or more possible display modes for at least one of the video conferencing system locations may be determined based on the accumulated amount of the audio signal from each of one or more of the audio signals. In other words, the system may choose from a plurality of display modes for each of the participants based on the accumulated amount of audio signal being generated by the participants. In some embodiments, the possible display modes comprise a single window display mode and a multiple window display mode. The multiple window display mode may comprise the continuous presence display mode described hereinabove. As noted above, the continuous presence display mode may comprise a display with a subset or all of the participants in the video conference as will be described in more detail below.

[0069] In some embodiments, the method may compare an accumulated amount of the audio signal from one or more of the audio signals with at least one accumulation threshold, where the display mode may be determined based on the comparing. Said another way, the MCU may use signal integrator 411 to determine if a participant has accumulated audio signal above a certain level, i.e., an accumulation threshold. As used herein, the accumulation threshold corresponds to the level of accumulated audio signal after which the display mode is changed. For example, if a participant begins to talk, the system may switch the other participants’ displays to the speaking participant only after the speaking participant has accumulated enough audio signal to exceed the accumulation threshold. The accumulation threshold will be discussed in more detail below with regard to FIGS. 7 and 8.

[0070] The value of the accumulation threshold may be static, may be set by an administrator or moderator, or may be set by one participant, or may be set by each participant. In one embodiment, the value of the accumulation threshold may be set to approximate a normal talking voice with an 8 second time duration, or a loud talking voice with a 6 second time duration.

[0071] In some embodiments, if the accumulated amount of audio signal corresponding to a first location exceeds an accumulation threshold, video signals from the first location may be displayed on each of a plurality of video conferencing systems in the single window mode. In other words, if a participant’s accumulated signal exceeds some value, e.g., if the participant speaks enough to surpass his respective accumulation threshold, each of the other participants, i.e., the listening participants, may view that single speaker. Alternatively, the other participants may view that speaker in combination with a subset of the other participants. The talking participant, however, may view a continuous presence mode, e.g., he may see all of the other participants or, alternatively, a subset therefrom. In one embodiment, a subset or any of the participants may be able to choose the subset of the participants that may be viewed or may set a desired display mode independent of any determination of

the accumulated audio signal. The MCU may utilize a mode switch 415 function to implement the display change for each of the participants.

[0072] Alternatively, if the accumulated amount of audio signal corresponding to any location does not exceed the accumulation threshold, video signals from a plurality of locations may be displayed on each of a plurality of video conferencing systems in a continuous presence mode. Said another way, if no one in the video conference is speaking in an uninterrupted manner for a certain threshold amount of audio signal (e.g. energy), the participants may view a continuous presence display mode comprising a subset or all of the participants on their display.

[0073] In one embodiment, if the accumulated amount of audio signal corresponding to a subset of locations are determined to be repeatedly exceeding the accumulation threshold, video signals from that respective subset of locations may be displayed on each of a plurality of video conferencing systems in a continuous presence mode. In other words, if participants from a certain subset of participant locations are doing all of the talking, i.e., exceeding a common (or respective) accumulation threshold, this subset of the talking participants may be displayed on each of the participants' displays. Alternatively, the participants' displays may show each of the talking participants singly, and intelligently switch between each of the talking participants throughout the conversation. In some embodiments, the talking participants and the listening participants may view different displays. For example, the talking participants may view all of listening participants, the other talking participants, or all of the participants in the video conference. Similarly, the listening participants may view all of the talking participants, the currently talking participant, or a subset or all of the participants in the video conference. Note that the displays for the talking and listening participants are not limited to the displays described above, and in fact, other displays are contemplated. In some embodiments, the talking and listening participants may be able to manually choose between a plurality of views to be displayed. In one embodiment, only one audio signal may exceed the accumulation threshold at any given time.

[0074] In embodiments utilizing an accumulation threshold, the method may also include modifying, e.g., raising, a respective accumulation threshold for a video conferencing system when an accumulated amount of audio signal has not exceeded the respective accumulation threshold within a predetermined amount of time. The method may also modify, e.g., lower, a respective accumulation threshold for a video conferencing system when an accumulated amount of audio signal has recently exceeded the respective accumulation threshold within a predetermined amount of time. In other words, the accumulation thresholds may be variable, i.e. may dynamically change, throughout the duration of the video conference. Additionally, the accumulation threshold variables may vary differently depending on whether the respective participant has spoken within some predetermined amount of time.

[0075] The accumulation thresholds may vary with respect to each participant, i.e., each participant may have his own threshold that may vary independently from the other participants' thresholds. In one embodiment, each participant's threshold may be normalized with respect to

each participant. For example, quieter participants may have lower thresholds than louder participants. Such an example will be described in more detail hereinbelow.

[0076] As described above, the accumulated amount of audio signal from each participant may be measured to determine when the video conferencing system should switch between two speakers and/or switch between single speaker mode and continuous presence (multiple speaker) mode. **FIGS. 7 and 8** illustrate an example where the use of accumulated audio signal 605, rather than time 609, provides improvements to display mode switching as outlined below.

[0077] In some embodiments, the system may determine when a single speaker is presumed to be talking, e.g., when the volume or amplitude level 607 of the audio signal from one participant location is above a certain audio threshold 603, or greater than the other locations by a certain threshold or ratio. When a single speaker is determined to be talking, as illustrated in **FIG. 7** in time segment A, the system may begin to integrate or sample the audio or voice signal received from that user or that location. When a certain amount of audio signal has been generated or accumulated by the integration, such as in **FIG. 7** in the integrated area before 702 for participant 151, the system may presume that the user has been talking for a sufficient amount (e.g., of accumulated audio signal) and that he may be a single talking user. At this point, the system may switch from continuous presence mode, illustrated in **FIG. 8** during time segment A as 801, where a subset or all of the participants are displayed, to a single speaker mode 803, where only the single speaker, in this case participant 151, may be displayed. The display of the talking participant may remain in continuous presence mode to allow the talking participant to view a plurality of other participants. However, the displays of the other participants may be switched to the location of the talking participant.

[0078] Note that this method does not measure the amount of time that a participant has been speaking, but rather measures the amount of accumulated audio signal generated by the remote location. Thus, if a participant is speaking very loudly or more loudly than normal if the thresholds are normalized, the system may switch the other participants' displays to the talking participant faster than if the talking participant was speaking more softly. Such a situation is illustrated in the transition 704 to time segment C in **FIGS. 7 and 8**. In this instance, participant 157 generates the threshold amount of accumulated audio signal in a smaller amount of time than that of participant 151 during time segment A. In this case, the single window display mode transfers to participant 157 more quickly than it had previously for participant 151 because of participant 157's louder speaking volume. Moreover, if a participant begins shouting in the video conference, the system will switch the other participants' displays to the shouting participant even faster. This occurs because the system measures the accumulated audio signal, essentially the amount of audio signal produced, as opposed to the prior art method which simply measures the length of time a participant speaks.

[0079] Finally, when no participant speaks, as illustrated in time segment D of **FIGS. 7 and 8**, the system may switch the participants' displays back to continuous presence mode. Thus, the present method provides a significant improve-

ment over prior time based methods, in that the method switches the participants' displays to a participant speaking loudly more quickly.

[0080] In some embodiments, the system may adjust the accumulation threshold of each participant based on the participant's total accumulated audio signal, i.e., the sum of all the accumulated audio signals from that participant. Thus, participants who are speaking more in the video conference may have their accumulation threshold lowered, while other participants who are speaking less or not at all in the video conference may have their accumulation threshold raised. Moreover, the system may switch to, i.e., switch a plurality of participants' displays to, those participants who are speaking more or more often in a video conference in a faster or more responsive manner than participants who are speaking less in the video conference. Consequently, the system may switch to those participants who are speaking less in the video conference in a slower or less responsive manner, presuming that these less-talking participants may not be speaking very long or often. In some embodiments, the accumulation thresholds may be adjusted each time the system switches to a new speaker. The thresholds may also be adjusted after a predetermined amount of time for each participant, e.g., long enough to predict the participant's long-term behavior.

[0081] In situations where two of the participants are having a dialog, e.g., two people or two participant locations are in a discussion, conversation focus tends to go back and forth between those two people or locations. In one embodiment, the system tracks which participants are talking.

[0082] If the system determines that two of the plurality of participants (or participant locations) are engaging in a conversation, the system may lower the accumulation threshold required to display a single talking participant. Thus, when a first participant of these two participants begins talking, the system may show the first participant more quickly. Similarly, after the first participant stops talking and the second participant begins to talk, the system may switch to that second participant in single presentation mode more quickly. Thus, the system may essentially ping-pong back and forth between each of the two talking participants. In other words, after one of the participants stops talking and the other participant starts talking, the system may switch to the single presentation mode of the talking participant substantially immediately, e.g., within a second or two.

[0083] In another embodiment, as noted above, when the system detects that two (or a subset) of the participants are doing all of the talking (i.e., only their audio signals are exceeding the accumulation threshold) the system may show these subset of participants in continuous presence mode. Thus, in some embodiments, when the system detects that two of the participants are having a dialog, the system may display these two participants in a dual split display mode. Thus, if six different participant locations are participating in the video conference, but participants A and B are dominating the conversation, the accumulation threshold for participants A and B may be lowered. Therefore, when either of participants A or B begins speaking, the system may quickly switch to this dual display mode. In some embodiments, participants A and B may have two associated accumulation thresholds. The system may display the dual display mode

using the lowered accumulation threshold, and then later switch to the single speaker mode for one of the participants upon reaching a second higher accumulation threshold. In other words, the system may intelligently switch to a single speaker mode if the second participant in the two-person dialog is no longer responding to the conversation.

[0084] When another one of the participants (participant C) begins speaking, a first (greater) accumulation threshold may be required to switch from the dual-display mode to a continuous presence mode (where the three talking participants or all six participants are displayed). A second (and even greater) accumulation threshold may be required to switch from continuous presence mode to single speaker mode for participant C.

[0085] Thus, the algorithm may be intelligent, e.g., using heuristics, to know that, for example, two speakers in the past ten minutes have been the dominant speakers; so, when one accumulates even a small amount of accumulated audio signal, the system may switch to that single speaker, or switch to a dual-speaker mode, much more quickly.

[0086] As described above, others of the participants that are not engaged in this two-person dialog may not have this lowered accumulation threshold. Thus, when a third participant that is not part of this two-person dialog begins speaking, this third participant must generate a greater amount of audio signal energy before the display switches to either a continuous presence mode or single speaker mode view of this third participant. As noted above, participants other than the two dominant participants may also have two different accumulation thresholds, a first to go from dual display mode of the two dominant speakers to continuous presence mode, and a second accumulation threshold to go to single speaker mode for that participant.

[0087] In some embodiments, each participant may have independent audio thresholds. For example, participant 151 may have audio threshold 603A, and participant 157 may have audio threshold 603B. Independent thresholds may be desirable for situations when a first participant is in a noisy environment. In such environments, a larger audio threshold may allow the MCU to properly determine when the first participant is speaking; i.e., the larger audio threshold may prevent the MCU from mistaking background noise for the participant's voice. However, if a second participant is in a quiet environment, it may be desirable for the second participant to have a much lower audio threshold than the first participant. In some embodiments, each participant's independent audio threshold may be normalized with respect to each participant. For example, in some situations, a first participant may have a louder normal speaking volume than a second participant. In this case, the first participant may have a higher audio threshold than the second participant. Thus, the quieter participants, such as the second participant, may not have to speak louder than normal to exceed their respective audio thresholds. Thus, independent audio thresholds may be desirable.

[0088] In some embodiments, similar to the accumulation threshold, the audio threshold for each participant may vary throughout a video conference. **FIG. 9** illustrates two respective audio signals integrated above a variable audio threshold, according to one embodiment. In some embodiments, the thresholds may be continuous. In other embodiments, the thresholds may be defined as a piece-wise func-

tion such as that in **FIG. 9**. In one embodiment, the threshold may vary with respect to whether the participant is speaking; for example, participant **151**'s threshold may decrease while participant **151** is speaking, as in **903**, and may increase while participant **151** is listening, as in **905**. Similarly, participant **157**'s threshold may also decrease while he is speaking, **913**, and increase while listening, **911** and **915**.

[0089] **FIGS. 10a-c** illustrate various embodiments of continuous presence screens. In some embodiments, the system may determine a dominant speaker **1003**, **1113**, or **1123** for display in a central and/or larger area of the display than corresponding other participants (e.g., other participants **1001a-h**; **1111a-l**; and **1121a-g**). Other continuous presence displays are also contemplated.

[0090] Thus, various embodiments of the systems and methods described above may facilitate intelligent control of video display modes in a video conferencing system.

[0091] Embodiments of these methods may be implemented from a memory medium. A memory medium may include any of various types of memory devices or storage devices. The term "memory medium" is intended to include an installation medium, e.g., a CD-ROM, floppy disks, or tape device; a computer system memory or random access memory such as DRAM, DDR RAM, SRAM, EDO RAM, Rambus RAM, etc.; or a non-volatile memory such as a magnetic media, e.g., a hard drive, or optical storage. The memory medium may comprise other types of memory as well, or combinations thereof. In addition, the memory medium may be located in a first computer in which the programs are executed, or may be located in a second different computer that connects to the first computer over a network, such as the Internet. In the latter instance, the second computer may provide program instructions to the first computer for execution. The term "memory medium" may include two or more memory mediums that may reside in different locations, e.g., in different computers that are connected over a network. In some embodiments, a carrier medium may be used. A carrier medium may include a memory medium as described above, as well as signals such as electrical, electromagnetic, or digital signals, conveyed via a communication medium such as a bus, network and/or a wireless link.

[0092] In some embodiments, a method may be implemented from memory medium(s) on which one or more computer programs or software components according to one embodiment may be stored. For example, the memory medium may comprise an electrically erasable programmable read-only memory (EEPROM), various types of flash memory, etc. which store software programs (e.g., firmware) that is executable to perform the methods described herein. In some embodiments, field programmable gate arrays may be used. Various embodiments further include receiving or storing instructions and/or data implemented in accordance with the foregoing description upon a carrier medium.

[0093] Further modifications and alternative embodiments of various aspects of the invention may be apparent to those skilled in the art in view of this description. Accordingly, this description is to be construed as illustrative only and is for the purpose of teaching those skilled in the art the general manner of carrying out the invention. It is to be understood that the forms of the invention shown and described herein are to be taken as embodiments. Elements and materials may

be substituted for those illustrated and described herein, parts and processes may be reversed, and certain features of the invention may be utilized independently, all as would be apparent to one skilled in the art after having the benefit of this description of the invention. Changes may be made in the elements described herein without departing from the spirit and scope of the invention as described in the following claims.

What is claimed is:

1. A method, comprising:

receiving an audio signal from each of a plurality of video conferencing system locations;

determining an accumulated amount of the audio signal from each of one or more of the audio signals; and

determining a display mode for at least one of the video conferencing system locations based on said determining, wherein the display mode is determined from two or more possible display modes.

2. The method of claim 1, wherein said determining the accumulated amount of audio signal comprises:

determining a signal metric for each of one or more of the respective audio signals using an integrated form of the respective signal.

3. The method of claim 1,

wherein said determining the accumulated amount of the audio signal comprises integrating each of the one or more audio signals from the plurality of video conferencing systems to generate respective accumulated amounts of audio signal.

4. The method of claim 1,

wherein the two or more possible display modes comprise a single window display mode and a multiple window display mode.

5. The method of claim 1, further comprising:

comparing the accumulated amount of the audio signal from one or more of the audio signals with at least one accumulation threshold;

wherein the display mode is determined based on said comparing.

6. The method of claim 5, wherein:

if the accumulated amount of audio signal corresponding to a first location exceeds an accumulation threshold, displaying video signals from the first location on a display of each of a plurality of video conferencing systems in a single window mode;

if the accumulated amount of audio signal corresponding to any location does not exceed the accumulation threshold, displaying video signals from that respective locations on a display of each of a plurality of video conferencing systems in a continuous presence mode.

7. The method of claim 5, wherein:

if the accumulated amount of audio signal corresponding to a subset of locations exceeds the accumulation threshold, displaying video signals from that respective subset of locations on a display of each of a plurality of video conferencing systems in a continuous presence mode.

8. The method of claim 5, further comprising:
modifying a respective accumulation threshold for a video conferencing system when an accumulated amount of audio signal has not exceeded the respective accumulation threshold within a predetermined amount of time.
9. The method of claim 5, further comprising:
modifying a respective accumulation threshold for a video conferencing system when an accumulated amount of audio signal has recently exceeded the respective accumulation threshold within a predetermined amount of time.
10. The method of claim 5,
wherein said comparing comprises comparing the accumulated amounts of audio signal from each of the audio signals with the at least one accumulation threshold.
11. The method of claim 1,
wherein said determining the accumulated amount of the audio signal comprises determining the accumulated amount of the audio signal after the audio signal has exceeded an audio threshold.
12. The method of claim 1,
wherein the accumulated amount of audio signal from each of one or more of the audio signals is an uninterrupted accumulated amount of audio signal.
13. A computer accessible memory medium comprising program instructions for determining a display mode in a video conferencing system, wherein the program instructions are executable to implement:
receiving an audio signal from each of a plurality of video conferencing system locations;
determining an accumulated amount of the audio signal from each of one or more of the audio signals; and
determining a display mode for at least one of the video conferencing system locations based on said determining, wherein the display mode is determined from two or more possible display modes.
14. The memory medium of claim 13,
wherein the accumulated amount of audio signal from each of one or more of the audio signals is an uninterrupted accumulated amount of audio signal.
15. The memory medium of claim 13,
wherein said determining the accumulated amount of the audio signal comprises integrating each of the one or more audio signals from the plurality of video conferencing systems to generate respective accumulated amounts of audio signal.
16. The memory medium of claim 13, wherein the program instructions are further executable to implement:
comparing the accumulated amount of the audio signal from one or more of the audio signals with at least one accumulation threshold;
wherein the display mode is determined based on said comparing.
17. A method for automatically determining a display mode for a display device comprising the steps of:
(a) receiving a signal from each of multiple endpoints;
(b) monitoring an amount of audio signal from each of the multiple endpoints;
(c) comparing the amount of audio signal from each of the multiple endpoints with predefined parameters; and
(d) determining a display mode from available display modes, wherein available display modes are single-window display and multiple-window display, based on step (c);
18. The method of claim 17, further comprising:
(e) wherein when the determined display mode is different than a current display mode of the display device, transmitting a display mode command signal based on a determination in step (d), the display mode command signal affecting the display mode of the display device; and
19. The method of claim 18,
wherein step (e) comprises a command signal to specify the multiple-window display upon the duration from each of the multiple endpoints not exceeding a first predefined parameter.
20. The method of claim 18, wherein step (e) comprises a command signal to specify the single-window display to display video images originating from one of the multiple endpoints from which the duration exceeds a predefined parameter and upon none of the durations from each of the other multiple endpoints exceeds the predefined parameter.
21. The method of claim 18, wherein step (e) comprises a command signal to specify the multiple-window display upon the durations from at least two of the multiple endpoints exceeding a predefined parameter.
22. The method of claim 17, wherein the display device is coupled to a video conferencing device or application.
23. A system, comprising:
a plurality of video conferencing systems, wherein the plurality of video conferencing systems are coupled through a network and wherein the plurality of video conferencing systems provide video and audio signals of participants using the respective systems;
a signal integrator, wherein the signal integrator determines an amount of accumulated audio signal for each of the plurality of video conferencing systems; and
a mode switch coupled to the signal integrator and operable to select a display mode based on the amount of accumulated audio signal for each of the plurality of video conferencing systems
24. The system of claim 23,
wherein if the amount of accumulated audio signal of a first video conferencing system exceeds an accumulation threshold, the mode switch directs a display coupled to at least one of the video conferencing systems to display the video signals provided by the first video conferencing system with the amount of accumulated audio signal that exceeds the accumulation threshold.
25. The system of claim 23, wherein if each of the amounts of accumulated audio signal of the plurality of video conferencing systems do not exceed the accumulation threshold, a display on at least one of the plurality of video

conferencing system displays at least two of the plurality of video conferencing systems video signals.

26. The system of claim 23, wherein if two or more of the plurality of video conferencing systems each exceed the accumulation threshold, a display on at least one of the plurality of video conferencing system displays the two or more of the plurality of video conferencing systems.

27. A switching system for automatically determining a display mode for a video display device comprising:

- an integrator configured to determine an amount of audio signal of each of a plurality of audio signals, the signals being from a source at each of multiple endpoints; and
- a switching processor coupled to the integrator and to a video switching module, configured to determine an appropriate display mode from the available display modes, wherein available display modes are single-

window display and multiple-window display, based upon a comparison of the integrated audio signal energy of each of the signals with at least one pre-defined parameter.

28. The switching system of claim 27,

wherein upon a determination that the appropriate display mode is different than the current display mode the switching processor transmits to the video switching module a display mode command, the display mode command being chosen from a single-window display command to effect the single-window display and a multiple-window display command to effect the multiple-window display.

* * * * *