



(19)中華民國智慧財產局

(12)發明說明書公告本

(11)證書號數：TW I602437 B

(45)公告日：中華民國 106 (2017) 年 10 月 11 日

(21)申請案號：105100653

(22)申請日：中華民國 105 (2016) 年 01 月 11 日

(51)Int. Cl. : **H04N7/15 (2006.01)**

(30)優先權：2015/01/12 美國

62/102,533

(71)申請人：仁寶電腦工業股份有限公司 (中華民國) COMPAL ELECTRONICS, INC. (TW)
臺北市內湖區瑞光路 581 號

(72)發明人：詹婕妤 CHAN, CHIEH-YU (TW)；翁明哲 WENG, MING-CHE (TW)

(74)代理人：葉璟宗；詹東穎；劉亞君

(56)參考文獻：

TW 201223283A

CN 103096024A

US 2009/0213940A1

US 2012/0050456A1

審查人員：林建儒

申請專利範圍項數：20 項 圖式數：5 共 35 頁

(54)名稱

視訊與音訊處理裝置及其視訊會議系統

VIDEO AND AUDIO PROCESSING DEVICES AND VIDEO CONFERENCE SYSTEM

(57)摘要

本發明提出一種視訊會議系統，其包括至少一客戶端裝置以及可連線於客戶端裝置與遠端裝置的主機裝置。主機裝置將取得至少一筆候選視訊資料以及多筆待整合音訊資料，根據待整合音訊資料篩選出其中一筆候選視訊資料，並且整合待整合音訊資料，以及將整合後的音訊資料以及篩選後的視訊資料傳送至遠端裝置，其中待整合音訊資料包括客戶端音訊資料以及主機音訊資料，候選視訊資料包括至少其中一筆客戶端視訊資料以及主機視訊資料。本發明亦提出分別對應於客戶端裝置以及遠端裝置的視訊與音訊處理裝置。

A video conference system including at least one client device and a host device able to be connected to the client devices and a remote device is provided. The host device obtains at least one piece of candidate video data as well as multiple pieces of audio data to be integrated, selects one piece of video data from the candidate video data, integrates the audio data to be integrated, and sends the selected video data and the integrated audio data to the remote device, where the audio data to be integrated includes client audio data and host audio data, the candidate video data includes at least one piece of client video data and client video data. Video and audio processing devices corresponding to the client device and the remote device are also provided.

指定代表圖：

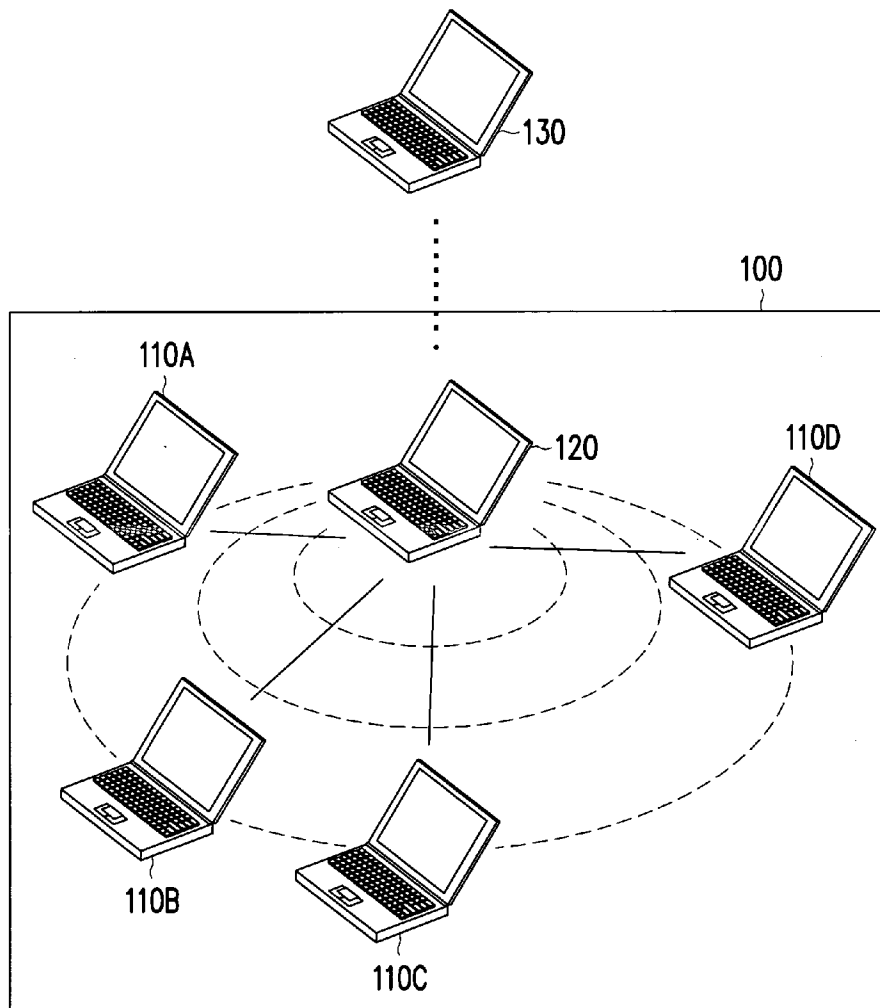
符號簡單說明：

100 . . . 視訊會議系統

110A~110D . . . 客戶端裝置

120 . . . 主機裝置

130 . . . 遠端裝置



【圖1A】



公告本

申請日: 105/01/11

IPC分類: H04N 7/15 (2006.01)

【發明摘要】

【中文發明名稱】

視訊與音訊處理裝置及其視訊會議系統

【英文發明名稱】

VIDEO AND AUDIO PROCESSING DEVICES AND VIDEO
CONFERENCE SYSTEM

【中文】本發明提出一種視訊會議系統，其包括至少一客戶端裝置以及可連線於客戶端裝置與遠端裝置的主機裝置。主機裝置將取得至少一筆候選視訊資料以及多筆待整合音訊資料，根據待整合音訊資料篩選出其中一筆候選視訊資料，並且整合待整合音訊資料，以及將整合後的音訊資料以及篩選後的視訊資料傳送至遠端裝置，其中待整合音訊資料包括客戶端音訊資料以及主機音訊資料，候選視訊資料包括至少其中一筆客戶端視訊資料以及主機視訊資料。本發明亦提出分別對應於客戶端裝置以及遠端裝置的視訊與音訊處理裝置。

【英文】A video conference system including at least one client device and a host device able to be connected to the client devices and a remote device is provided. The host device obtains at least one piece of candidate video data as well as multiple pieces of audio data to be integrated, selects one piece of video data from the

candidate video data, integrates the audio data to be integrated, and sends the selected video data and the integrated audio data to the remote device, where the audio data to be integrated includes client audio data and host audio data, the candidate video data includes at least one piece of client video data and client video data. Video and audio processing devices corresponding to the client device and the remote device are also provided.

【指定代表圖】圖1A。

【代表圖之符號簡單說明】

100：視訊會議系統

110A～110D：客戶端裝置

120：主機裝置

130：遠端裝置

【特徵化學式】

無

【發明說明書】

【中文發明名稱】

視訊與音訊處理裝置及其視訊會議系統

【英文發明名稱】

VIDEO AND AUDIO PROCESSING DEVICES AND VIDEO
CONFERENCE SYSTEM

【技術領域】

【0001】 本發明是有關於一種視訊與音訊處理的技術，且特別是有關於一種視訊與音訊處理裝置及其視訊會議系統。

【先前技術】

【0002】 隨著電子科技以及通訊網路不斷地發展，視訊的硬體設備、軟體技術及操作功能已大幅提昇。現今的視訊可隨時提供多個使用者透過例如是電腦、智慧型手機、平板電腦等電子裝置進行同步連線，以讓使用者之間在進行語音通話時，更可以透過螢幕看到其它使用者的動態影像，從而增強交流的真實感以及臨場感。基此，視訊會議已應用於商業溝通上，藉此企業得以進行跨地域的內外部溝通。

【0003】 然而，一般在進行多人對多人的視訊會議時，往往會因為視訊設備的鏡頭廣角或是麥克風的收音範圍不足而導致所有人必須緊靠在視訊設備，進而影響到了視訊會議的品質與效率。

【發明內容】

【0004】 有鑑於此，本發明提供一種視訊與音訊處理裝置及其視訊會議系統，其可提供具有主講者聚焦效果的多人視訊會議。

【0005】 本發明提出一種視訊與音訊處理裝置，包括通訊電路、影像擷取裝置、收音裝置、記憶體以及處理器。通訊電路用以連線至另一視訊與音訊處理裝置。影像擷取裝置用以擷取影像訊號。收音裝置用以擷取聲音訊號。記憶體用以儲存檔案資料。處理器耦接通訊電路、影像擷取裝置、收音裝置以及記憶體，並且用以執行下列步驟。首先，自影像擷取裝置以及收音裝置分別接收第一影像訊號以及第一聲音訊號，以產生第一視訊資料以及第一音訊資料。接著，判斷第一音訊資料是否符合至少一傳輸條件。若是，透過通訊電路將第一音訊資料以及第一視訊資料傳送至另一視訊與音訊處理裝置。若否，透過通訊電路僅將第一音訊資料傳送至另一視訊與音訊處理裝置。

【0006】 在本發明的一實施例中，視訊與音訊處理裝置更包括螢幕以及播音裝置。螢幕用以顯示畫面，而播音裝置用以播放聲音。處理器更耦接螢幕以及播音裝置，並且用以透過通訊電路自另一視訊與音訊處理裝置接收遠端視訊資料以及遠端音訊資料，以及分別透過螢幕以及播音裝置播放遠端視訊資料以及遠端音訊資料。

【0007】 在本發明的一實施例中，第一音訊資料包括音訊頻率，

而處理器用以判斷音訊頻率是否符合人聲頻率。

【0008】 在本發明的一實施例中，第一音訊資料更包括收音裝置的收音音量，而處理器更用以判斷收音音量是否大於音量門檻值。

【0009】 在本發明的一實施例中，第一音訊資料更包括收音裝置的收音雜訊比，而處理器更用以判斷該收音雜訊比是否大於雜訊比門檻值。

【0010】 本發明提出一種視訊與音訊處理裝置，包括通訊電路、影像擷取裝置、收音裝置、記憶體以及處理器。通訊電路用以連線至另一視訊與音訊處理裝置。影像擷取裝置用以擷取影像訊號。收音裝置用以擷取聲音訊號。記憶體用以儲存檔案資料。處理器接螢幕、通訊電路、影像擷取裝置、收音裝置、播音裝置以及記憶體，並且用以執行下列步驟。首先，取得至少一筆候選視訊資料以及多筆待整合音訊資料，其中各筆候選視訊資料對應於其中一筆所述待整合音訊資料。接著，根據各筆候選視訊資料所對應的待整合音訊資料，選擇其中一筆候選視訊資料，以產生篩選後的視訊資料，並且整合各筆待整合音訊資料，以產生整合後的音訊資料。透過通訊電路將整合後的音訊資料以及篩選後的視訊資料傳送至遠端裝置。

【0011】 在本發明的一實施例中，處理器更用以透過通訊電路自遠端裝置接收遠端視訊資料以及遠端音訊資料，並且透過通訊電路傳輸遠端視訊資料以及遠端音訊資料至各個其它視訊與音訊處理裝置。

【0012】 在本發明的一實施例中，視訊與音訊處理裝置更包括螢幕以及播音裝置。螢幕用以顯示畫面，而播音裝置用以播放所接收的聲音資料。處理器更耦接螢幕以及播音裝置，並且用以分別透過螢幕以及播音裝置播放遠端視訊資料以及遠端音訊資料。

【0013】 在本發明的一實施例中，各筆待整合音訊資料包括收音音量，而處理器用以選擇具有最大收音音量的待整合音訊資料所對應的候選視訊資料做為篩選後的視訊資料。

【0014】 在本發明的一實施例中，各筆待整合音訊資料包括收音雜訊比，而處理器用以選擇具有最大收音雜訊比的待整合音訊資料所對應的候選視訊資料做為篩選後的視訊資料。

【0015】 在本發明的一實施例中，各所述待整合音訊資料包括收音時間，而處理器用以選擇具有最長收音時間的待整合音訊資料所對應的候選視訊資料做為篩選後的視訊資料。

【0016】 在本發明的一實施例中，各所述待整合音訊資料包括收音時間，而處理器用以選擇收音時間大於時間門檻值的音訊資料所對應的候選視訊資料做為篩選後的視訊資料。

【0017】 在本發明的一實施例中，處理器用以針對各筆待整合音訊資料進行混音處理以及/或除噪處理，以產生整合後的音訊資料。

【0018】 在本發明的一實施例中，處理器更用以執行下列步驟。自影像擷取裝置以及收音裝置分別接收第二影像訊號以及第二聲音訊號，以產生第二視訊資料以及第二音訊資料。接著，判斷第二音訊資料是否符合至少一傳輸條件。若是，分別設定第二視訊

資料以及第二音訊資料為其中一筆候選視訊資料以及其中一筆待整合音訊資料。若否，僅設定第二音訊資料為其中一筆待整合音訊資料。

【0019】 在本發明的一實施例中，第二音訊資料包括音訊頻率，而處理器用以判斷音訊頻率是否符合人聲頻率。

【0020】 在本發明的一實施例中，第二音訊資料更包括收音裝置的收音音量，而處理器更用以判斷收音音量是否大於音量門檻值。

【0021】 在本發明的一實施例中，第二音訊資料更包括收音裝置的收音雜訊比，而處理器更用以判斷收音音量是否大於音量門檻值。

【0022】 本發明提出一種視訊與音訊處理裝置，包括至少一客戶端裝置以及可連線於客戶端裝置與遠端裝置的主機裝置。各個客戶端裝置將分別產生客戶端視訊資料以及客戶端音訊資料。主機裝置將取得至少一筆候選視訊資料以及多筆待整合音訊資料，根據待整合音訊資料選擇其中一筆候選視訊資料，以產生篩選後的視訊資料，並且整合各筆待整合音訊資料，以及將整合後的音訊資料以及篩選後的視訊資料傳送至遠端裝置，其中待整合音訊資料包括客戶端音訊資料以及主機音訊資料，候選視訊資料包括至少其中一筆客戶端視訊資料以及主機視訊資料，各筆候選視訊資料對應於其中一筆所述音訊資料。

【0023】 在本發明的一實施例中，對於各個客戶端裝置將判斷客戶端音訊資料是否符合至少一傳輸條件，若是，客戶端裝置將客

戶端音訊資料以及客戶端視訊資料傳送至主機裝置，若否，客戶端裝置僅將客戶端音訊資料傳送至主機裝置。

【0024】 在本發明的一實施例中，主機裝置更自遠端裝置接收遠端視訊資料以及遠端音訊資料，並且傳送遠端視訊資料以及遠端音訊資料至各所述客戶端裝置。

【0025】 基於上述，本發明所提出的視訊與音訊處理裝置及其視訊會議系統，主機裝置除了將整合本地端所產生的各筆音訊資料並且傳送整合後的音訊資料至遠端裝置外，更根據各筆音訊資料的聲音參數來傳送其中一筆視訊資料至遠端裝置，以使本地端在具有多個視訊會議參與者的前提下，遠端裝置的使用者視覺仍可聚焦於本地端的主講者上，以提升視訊會議的品質與效率。

【0026】 為讓本發明的上述特徵和優點能更明顯易懂，下文特舉實施例，並配合所附圖式作詳細說明如下。

【圖式簡單說明】

【0027】

圖 1A 繪示依據本發明一實施例之視訊會議系統的示意圖。

圖 1B 繪示依據本發明一實施例之客戶端裝置的示意圖。

圖 1C 繪示依據本發明一實施例之主機裝置的示意圖。

圖 2 繪示依據本發明一實施例之音訊與視訊處理方法的流程圖。

圖 3 繪示依據本發明一實施例之音訊與視訊處理方法的流程

圖。

圖 4 繪示依據本發明一實施例之音訊與視訊處理方法的流程圖。

圖 5A 繪示依據本發明一實施例之音訊與視訊處理方法的流程圖。

圖 5B 繪示圖 5A 的應用情境示意圖。

【實施方式】

【0028】 本發明的部份實施例接下來將會配合附圖來詳細描述，以下的描述所引用的元件符號，當不同附圖出現相同的元件符號將視為相同或相似的元件。這些實施例只是本發明的一部份，並未揭示所有本發明的可實施方式。更確切的說，這些實施例只是本發明的專利申請範圍中的裝置以及系統的範例。

【0029】 圖 1A 繪示依據本發明一實施例之視訊會議系統的示意圖。

【0030】 請參照圖 1A，視訊會議系統 100 包括客戶端裝置 110A～110D 以及主機裝置 120，其中主機裝置 120 可分別連線至客戶端裝置 110A～110D 以及遠端裝置 130。客戶端裝置 110A～110D 以及主機裝置 120 為本地端裝置，而客戶端裝置 110A～110D 可以是利用區域網路與主機裝置 120 進行無線連線，或者是直接與主機裝置 120 進行有線連結。此外，主機裝置 120 可以是利用網際網路與遠端裝置 130 進行連線。必須說明的是，視訊會議系統 100

中的客戶端裝置 110A~110D 僅是為了方便說明實作方式。在其它的實施例中，視訊會議系統 100 更可提供任意數量的客戶端裝置來與主機裝置 120 進行連線，本發明不在此設限。

【0031】 圖 1B 繪示依據本發明一實施例之客戶端裝置的示意圖。

【0032】 請參照圖 1B，客戶端裝置 110A~110D 皆包括螢幕 111、通訊電路 112、影像擷取裝置 113、播音裝置 114、收音裝置 115、記憶體 116 以及處理器 117。在本實施例中，客戶端裝置 110A~110D 例如是個人電腦、筆記型電腦、智慧型手機、平板電腦、個人數位助理等具有音訊與視訊處理功能的電子裝置，然而本發明不以此為限。為了方便說明，以下僅將針對客戶端裝置 110A 來進行描述。

【0033】 螢幕 111 用以顯示客戶端裝置 110A 輸出的畫面而提供給使用者觀看。在本實施例中，螢幕 111 例如是外接或是內建於客戶端裝置 110A 的液晶顯示器（Liquid Crystal Display，LCD）、發光二極體（Light-Emitting Diode，LED）顯示器、場發射顯示器（Field Emission Display，FED）或其他種類的顯示器。

【0034】 通訊電路 112 用以透過通訊網路與其它裝置連結，其可以例如是支援 WiMAX 通訊協定、Wi-Fi 通訊協定、3G 通訊協定、4G 通訊協定等無線網路連線，ADSL 寬頻或光纖網路等有線網路連線的元件，然而本發明不在此設限。

【0035】 影像擷取裝置 113 用以擷取其前方的影像，其可以是採用電荷耦合元件（charge coupled device, CCD）、互補性氧化金屬

半導體（Complementary Metal-Oxide Semiconductor，CMOS）元件或其他元件鏡頭的照相機。影像擷取裝置 113 可以是內建於客戶端裝置 110A，亦可以是外接於客戶端裝置 110A 的網路攝影機、數位相機、單眼相機、數位攝影機等。

【0036】 播音裝置 114 用以播放聲音，其包括喇叭。收音裝置 115 用以接收聲音，其包括麥克風。播音裝置 114 以及收音裝置 115 可以是內建於客戶端裝置 110A，亦可以是外接於客戶端裝置 110A。此外，當播音裝置 114 以及收音裝置 115 外接於客戶端裝置 110A 時，更可以整合為例如是耳機麥克風等單一裝置。

【0037】 記憶體 116 用以儲存檔案資料，其可以例如是任意型式的固定式或可移動式隨機存取記憶體（Random Access Memory，RAM）、唯讀記憶體（Read-Only Memory，ROM）、快閃記憶體（Flash memory）、硬碟或其他類似裝置或這些裝置的組合。

【0038】 處理器 117 耦接於螢幕 111、通訊電路 112、影像擷取裝置 113、播音裝置 114、收音裝置 115 以及記憶體 116，其係用以控制以及整合此些元件之間的作動。處理器 117 可以例如是中央處理單元（Central Processing Unit，CPU），或是其他可程式化之一般用途或特殊用途的微處理器（Microprocessor）、數位訊號處理器（Digital Signal Processor，DSP）、可程式化控制器、特殊應用積體電路（Application Specific Integrated Circuits，ASIC）、可程式化邏輯裝置（Programmable Logic Device，PLD）或其他類似裝置或這些裝置的組合。

【0039】 除此之外，在一實施例中，客戶端裝置 110A 亦可以是僅由通訊電路 112、影像擷取裝置 113、收音裝置 115、記憶體 116 以及處理器 117 所組成的電子裝置，並且可外接於包括螢幕 111 以及播音裝置 114 的另一電子裝置（例如是電視）。

【0040】 圖 1C 繪示依據本發明一實施例之主機裝置的示意圖。

【0041】 請參照圖 1C，主機裝置 120 包括螢幕 121、通訊電路 122、影像擷取裝置 123、播音裝置 124、收音裝置 125、記憶體 126 以及處理器 127。在本實施例中，主機裝置 120 亦可以例如是個人電腦、筆記型電腦、智慧型手機、平板電腦、個人數位助理等具有音訊與視訊處理功能的電子裝置，然而本發明不以此為限。螢幕 121、通訊電路 122、影像擷取裝置 123、播音裝置 124、收音裝置 125、記憶體 126 以及處理器 127 分別類似於圖 1B 中客戶端裝置 110A～110D 的螢幕 111、通訊電路 112、影像擷取裝置 113、播音裝置 114、收音裝置 115、記憶體 116 以及處理器 117，其相關說明請參照前述段落，於此不再贅述。

【0042】 在本實施例中，視訊會議系統 100 的多個本地端裝置（即客戶端裝置 110A～110D 以及主機裝置 120）與遠端裝置 130 進行視訊會議時，各個客戶端裝置 110A～110D 會將所產生的客戶端音訊資料傳送至主機裝置 120，更會利用客戶端音訊資料決定是否進一步將所產生的客戶端視訊資料傳送至主機裝置 120。主機裝置 120 本身亦會產生主機音訊資料以及主機視訊資料。之後，主機裝置 120 會整合音訊資料以及篩選視訊資料，並且將整合後的音訊

資料以及篩選後的視訊資料傳送到遠端裝置 130，據以實現具有主講者聚焦效果的視訊會議。以下將分別說明客戶端裝置 110A～110D 以及主機裝置 120 進行音訊與視訊處理的具體方式。

【0043】圖 2 繪示依據本發明一實施例之音訊與視訊處理方法的流程圖。本實施例的方法適用於圖 1B 的客戶端裝置 110A～110D，以下即搭配客戶端裝置 110A 中的各個元件說明其音訊與視訊處理方法之詳細步驟，而客戶端裝置 110B～110D 可依此類推。然而，就以實作上而言，本實施例的方法亦適用於僅包括以通訊電路 112、影像擷取裝置 113、收音裝置 115、記憶體 116 以及處理器 117 所組成的電子裝置，本發明不在此設限。

【0044】請同時參照圖 1B 以及圖 2，首先，客戶端裝置 110A 的處理器 117 將自影像擷取裝置 113 以及收音裝置 115 分別接收第一影像訊號以及第一聲音訊號，以產生第一視訊資料以及第一音訊資料（步驟 S202）。在此，第一影像訊號為影像擷取裝置 113 所擷取到客戶端裝置 110A 的使用者的動態影像，而第一聲音訊號為收音裝置 115 所擷取到客戶端裝置 110A 的周遭環境聲音。處理器 117 可將影像擷取裝置 113 以及收音裝置 115 所擷取到的第一影像訊號以及第一聲音訊號進行數位轉換，以產生第一視訊資料以及第一音訊資料。此外，處理器 117 亦可選擇性地壓縮第一視訊資料，以因應後續網路傳輸頻寬的不同需求，本發明不在此設限。

【0045】接著，處理器 117 將判斷第一音訊資料是否符合至少一

傳輸條件（步驟 S204），以藉由第一音訊資料的聲音參數來判定收音裝置 114 所擷取到的第一聲音訊號是否為客戶端裝置 110A 的使用者的說話聲以及判定客戶端裝置 110A 的使用者是否為視訊會議的可能主講者，從而判定第一音訊資料是否符合傳輸條件。在此的聲音參數可以是收音頻率、收音音量以及收音雜訊比等參數。

【0046】 具體來說，在一實施例中，由於第一聲音訊號為收音裝置 115 所擷取到的周遭環境聲音，因此處理器 117 會先判斷第一音訊資料的收音頻率是否符合人聲頻率。若是符合，則處理器 117 直接判定客戶端裝置 110A 的使用者為視訊會議的可能主講者。舉例而言，男性說話聲的頻率範圍為 85~180Hz，女性說話聲的頻率範圍為 165~255Hz，而處理器 117 可判斷第一音訊資料的收音頻率是否落入於此些區間內，藉以判定第一音訊資料是否對應於人聲。當第一音訊資料的收音頻率確實對應於人聲時，則處理器 117 將判定其符合傳輸條件。

【0047】 在一實施例中，處理器 117 在判定出第一音訊資料對應於人聲後，更會判斷第一音訊資料的收音音量是否大於預設的音量門檻值來判定客戶端裝置 110A 的使用者是否為視訊會議的可能主講者。一般而言，平常一般人說話的音量為 60dB 左右，而音量門檻值可以例如是預設為 55dB，以預留容許偵測誤差範圍。此外，處理器 117 亦可根據隨時根據視訊會議的周圍環境或者是客戶端裝置 110A 的使用者的手動調整來更改預設的音量門檻值。當第一音訊資料的收音頻率確實對應於人聲並且第一音訊資料的收

音音量大於音量門檻值時，則處理器 117 將判定其符合傳輸條件。

【0048】 在一實施例中，類似於收音音量的概念，處理器 117 在判定出第一音訊資料對應於人聲後，更會利用第一音訊資料的第一音訊資料的收音雜訊比是否大於預設的雜訊比門檻值來判定客戶端裝置 110A 的使用者是否為視訊會議的可能主講者，其中雜訊比門檻值可以例如是預設為 55dB。此外，處理器 117 亦可根據隨時根據客戶端裝置 110A 的周圍環境或者是客戶端裝置 110A 的使用者的手動調整來更改預設的雜訊比門檻值。當第一音訊資料的收音頻率確實對應於人聲並且第一音訊資料的雜訊比門檻值大於雜訊比門檻值時，則處理器 117 將判定其符合傳輸條件。

【0049】 在一實施例中，處理器 117 亦可同時利用第一音訊資料的收音頻率、收音音量以及收音雜訊比來判定客戶端裝置 110A 的使用者是否為視訊會議的可能主講者。當第一音訊資料的收音頻率確實對應於人聲、第一音訊資料的收音音量大於音量門檻值並且第一音訊資料的雜訊比門檻值大於雜訊比門檻值時，則處理器 117 將判定其符合傳輸條件。

【0050】 接著，當處理器 117 判定第一音訊資料符合傳輸條件時，則代表客戶端裝置 110A 的使用者為視訊會議的可能主講者，因此處理器 117 將透過通訊電路 112 傳送第一音訊資料以及第一視訊資料至主機裝置 120（步驟 S206）。另一方面，當處理器 117 判定第一音訊資料不符合任何傳輸條件時，則代表客戶端裝置 110A 的使用者並非為視訊會議的可能主講者，因此處理器 117 僅將透過

通訊電路 112 傳送第一音訊資料至主機裝置 120 (步驟 S208)，以減少本地端資料量的傳輸。必須先說明的是，在此傳送至主機裝置 120 的第一視訊資料僅為後續將傳送至遠端裝置 130 的其中一筆候選視訊資料，也就是說主機裝置 120 仍會在多筆候選視訊資料中篩選出其中一筆來傳送至遠端裝置 130。此外，傳送至主機裝置 120 的第一音訊資料亦會在後續與其它音訊資料進行整合。詳細細節將於稍後進行說明。

【0051】 主機裝置 120 除了將接收客戶端裝置 110A~110D 的音訊資料以及任何可能的視訊資料之外，本身亦會產生音訊資料以及視訊資料。詳細來說，圖 3 繪示依據本發明一實施例之音訊與視訊處理方法的流程圖。本實施例的方法適用於圖 1C 的主機裝置 120，以下即搭配主機裝置 120 中的各個元件說明其音訊與視訊處理方法之詳細步驟。然而，就以實作上而言，本實施例的方法亦適用於僅包括以通訊電路 122、影像擷取裝置 123、收音裝置 125、記憶體 126 以及處理器 127 所組成的電子裝置，本發明不在此設限。

【0052】 請同時參照圖 1C 以及圖 3，首先，主機裝置 120 的處理器 127 將自影像擷取裝置 123 以及收音裝置 125 接收第二影像訊號以及第二聲音訊號，以分別產生第二視訊資料以及第二音訊資料 (步驟 S302)，並且處理器 127 將判斷第二音訊資料是否符合至少一傳輸條件 (步驟 S304)。在此主機裝置 120 的處理器 127 執行步驟 S302 以及步驟 S304 的方式類似於客戶端裝置 110A~110D

執行步驟 S202 以及步驟 S204 的方式，相關說明請參照前述段落，於此不再贅述。

【0053】不同於圖 2 的實施例，由於主機裝置 120 本身將於稍後進行音訊資料的整合以及視訊資料的篩選，因此當處理器 127 判定第二音訊資料符合傳輸條件時，亦即主機裝置 120 的使用者為視訊會議的可能主講者，處理器 127 將第二視訊資料設定為其中一筆候選視訊資料以及將第二音訊資料設定為其中一筆待整合音訊資料（步驟 S306）。另一方面，當處理器 127 判定第二音訊資料不符合任何傳輸條件時，僅會將第二音訊資料設定為其中一筆待整合音訊資料（步驟 S308），亦即主機裝置 120 的使用者並非為視訊會議的可能主講者。

【0054】主機裝置 120 在取得本身以及客戶端裝置 110A～110D 所有待整合的音訊資料以及候選視訊資料後，將會分別進行整合以及篩選，以做為傳送至遠端裝置 130 的傳輸資料。詳細來說，圖 4 繪示依據本發明一實施例之音訊與視訊處理方法的流程圖。本實施例的方法適用於圖 1C 的主機裝置 120，以下即搭配主機裝置 120 中的各個元件說明其音訊與視訊處理方法之詳細步驟。

【0055】請同時參照圖 1C 以及圖 4，首先，主機裝置 120 的處理器 127 將取得至少一筆候選視訊資料以及多筆待整合音訊資料（步驟 S402）。在此的候選視訊資料以及待整合音訊資料即為客戶端裝置 110A～110D 在執行完圖 2 的流程以及主機裝置 120 在執行完圖 3 的流程後所取得的資料。

【0056】 接著，處理器 127 將根據各筆待整合音訊資料，自所有候選視訊資料中選擇其中之一者，以產生篩選後的視訊資料（步驟 S404）。詳細來說，候選視訊資料僅為視訊會議中的可能主講者的視訊資料，處理器 127 將比較所有候選視訊資料所對應的待整合音訊資料（亦即，與候選視訊資料相同來源裝置的待整合音訊資料），以進一步地選出將傳送至遠端裝置 130 的視訊資料。由於所有候選視訊資料所對應的待整合音訊資料已符合人聲頻率，因此處理器 127 將根據所有對應的待整合音訊資料的聲音參數來選擇所欲傳送的視訊資料。在此的聲音參數可以是收音時間、收音音量以及收音雜訊比等參數。

【0057】 在一實施例中，處理器 127 將選擇所有候選視訊資料所對應的待整合音訊資料之中具有最長收音時間者來做為篩選後的視訊資料。此實施例是則是直接將先說話的使用者視為視訊會議當下的主講者。

【0058】 在一實施例中，處理器 127 將選擇所有候選視訊資料所對應的待整合音訊資料之中收音時間大於時間門檻值者來做為篩選後的視訊資料，其中時間門檻值可以例如是 0.5 秒。此實施例是考量到多個使用者搶話的情形，因此處理器 127 僅將收音時間大於時間門檻值所對應的使用者視為視訊會議當下的主講者，以避免在後續遠端裝置 130 在極短的時間內不斷地切換顯示不同裝置使用者的視訊畫面。

【0059】 在一實施例中，處理器 127 將選擇所有候選視訊資料所

對應的待整合音訊資料之中具有最大收音音量者來做為篩選後的視訊資料。此實施例則是考量到其它候選視訊資料所對應的使用者可能在進行私下討論，而並非為視訊會議的主講者。

【0060】 在一實施例中，類似於收音音量的概念，處理器 127 將選擇所有候選視訊資料所對應的待整合音訊資料之中具有最大收音雜訊比者來做為篩選後的視訊資料。

【0061】 在一實施例中，處理器 127 亦可利用收音時間、收音音量以及收音雜訊比的不同組合來做為選擇依據，以使篩選後的結果更為精確。

【0062】 在一實施例中，當僅有單一筆候選視訊資料時，處理器 127 可直接將其設定為篩選後的視訊資料。在又另一實施例中，當沒有任何候選視訊資料時，即代表本地端的使用者並未發言，處理器 127 可持續地傳送前個時點的主講人的視訊資料來做為篩選後的視訊資料。

【0063】 另一方面，處理器 127 亦會一併整合所有待整合音訊資料，以產生整合後的音訊資料（步驟 S406）。詳細來說，處理器 127 可針對所有待整合音訊資料進行混音處理（Audio Mixing）以及/或除噪處理（Audio Denoising），以使整合後的音訊資料具有更好的品質。

【0064】 之後，處理器 127 將透過通訊電路 112 傳輸整合後的音訊資料以及篩選後的視訊資料至遠端裝置 130（步驟 S408）。換句話說，遠端裝置 130 自主機裝置 120 所接收的音訊資料以及視訊

資料為本地端整合後的聲音以及主講者的視訊畫面，從而達到具有主講者聚焦效果的視訊會議。

【0065】 附帶說明的是，主機裝置 120 同時將一併透過通訊電路 122 自遠端裝置 130 接收遠端視訊資料以及遠端音訊資料，並且分別透過螢幕 121 以及播音裝置 124 播放遠端視訊資料以及遠端音訊資料。此外，主機裝置 120 亦會傳送遠端視訊資料以及遠端音訊資料至客戶端裝置 110A~110D，而客戶端裝置 110A~110D 將透過螢幕 111 以及播音裝置 114 播放遠端視訊資料以及遠端音訊資料。

【0066】 圖 5A 繪示依據本發明一實施例之音訊與視訊處理方法的流程圖來說明圖 1A 的視訊會議系統 100 的應用情境。圖 5B 繪示圖 5A 的應用情境示意圖。

【0067】 請同時參照圖 1A 以及圖 5A，首先，視訊會議系統 100 的主機裝置 120 將取得至少一筆候選視訊資料以及多筆待整合音訊資料（步驟 S502）。在此的候選視訊資料為本地端的裝置使用者當中可能為主講者的視訊畫面，而待整合音訊資料為本地端的裝置使用者的說話聲。接著，主機裝置 120 將根據候選視訊資料所對應的待整合音訊資料來判斷本地端的主講者（步驟 S504），以產生主講者的視訊資料（步驟 S506）。另一方面，主機裝置 120 會將待整合音訊資料進行混音處理以及/或除噪處理（步驟 S508），以產生整合後的音訊資料（步驟 S510）。之後，主機裝置 120 會將主講者的視訊資料以及整合後的音訊資料傳送至遠端裝置（步驟

S512)。在此步驟 S502～S512 的細節請參照前述實施例的相關說明，於此不再贅述。

【0068】 請再參照圖 5B，在本實施例中，客戶端裝置 110A～110B 分別利用網路 LAN 與主機裝置 120 進行連線，而主機裝置 120 利用網路 WAN 與遠端裝置 130 進行連線。假設主機裝置 120 在圖 5A 中判定客戶端裝置 110A 的使用者 C1 為主講者後，會將使用者 C1 的視訊資料傳送至遠端裝置 130。此外，主機裝置 120 亦會將使用者 C1、C2 以及 H 的音訊資料 VC1、VC2 與 VH 所整合出的音訊資料 M 傳送至遠端裝置 130，而遠端裝置 130 將播放使用者 C1 的視訊畫面以及使用者 C1、C2 與 H 的聲音。此外，主機裝置 120 亦會接收並且傳送遠端裝置 130 的使用者 R 的視訊資料以及音訊資料 VC1 至客戶端裝置 110A～110B，而客戶端裝置 110A～110B 以及主機裝置 120 將同時播放使用者 R 的視訊畫面以及聲音。

【0069】 綜上所述，本發明所提出的視訊與音訊處理裝置及其視訊會議系統，主機裝置除了將整合本地端所產生的各筆音訊資料並且傳送整合後的音訊資料至遠端裝置外，更根據各筆音訊資料的聲音參數來傳送其中一筆視訊資料至遠端裝置，以使本地端在具有多個視訊會議參與者的前提下，遠端裝置的使用者視覺仍可聚焦於本地端的主講者上，以提升視訊會議的品質與效率。

【0070】 雖然本發明已以實施例揭露如上，然其並非用以限定本發明，任何所屬技術領域中具有通常知識者，在不脫離本發明的

精神和範圍內，當可作些許的更動與潤飾，故本發明的保護範圍當視後附的申請專利範圍所界定者為準。

【符號說明】

【0071】

100：視訊會議系統

110A～110D：客戶端裝置

120：主機裝置

130：遠端裝置

111、121：螢幕

112、122：通訊電路

113、123：影像擷取裝置

114、124：播音裝置

115、125：收音裝置

116、126：記憶體

117、127：處理器

S202～S208、S302～S308、S402～S408、S502～S512：

音訊與視訊處理方法的流程

C1、C2、H、R：使用者

VC1、VC2、VH、VR、M：音訊資料

LAN、WAN：網路

【發明申請專利範圍】

【第1項】 一種視訊與音訊處理裝置，包括：

通訊電路，用以連線至另一視訊與音訊處理裝置；

影像擷取裝置，用以擷取影像訊號；

收音裝置，用以擷取聲音訊號；

記憶體，用以儲存檔案資料；以及

處理器，耦接該通訊電路、該影像擷取裝置、該收音裝置以及該記憶體，用以執行下列步驟：

自該影像擷取裝置以及該收音裝置分別接收第一影像訊號以及第一聲音訊號，以產生第一視訊資料以及第一音訊資料；

判斷該第一音訊資料的聲音參數是否符合收音標準；

若是，透過該通訊電路傳送該第一音訊資料以及該第一視訊資料至該另一視訊與音訊處理裝置；以及

若否，透過該通訊電路僅傳送該第一音訊資料至該另一視訊與音訊處理裝置。

【第2項】 如申請專利範圍第1項所述的視訊與音訊處理裝置，更包括：

螢幕，用以顯示畫面；以及

播音裝置，用以播放聲音；

其中該處理器更耦接該螢幕以及該播音裝置，並且用以執行下列步驟：

透過該通訊電路自該另一視訊與音訊處理裝置接收遠端
視訊資料以及遠端音訊資料；以及

分別透過該螢幕以及該播音裝置播放該遠端視訊資料以
及該遠端音訊資料。

【第3項】 如申請專利範圍第1項所述的視訊與音訊處理裝置，其
中該第一音訊資料包括音訊頻率，而該處理器用以判斷該第一音
訊資料的該聲音參數是否符合該收音標準的步驟包括：

判斷該音訊頻率是否符合人聲頻率。

【第4項】 如申請專利範圍第3項所述的視訊與音訊處理裝置，其
中該第一音訊資料更包括該收音裝置的收音音量，而該處理器用
以判斷該第一音訊資料的該聲音參數是否符合該收音標準的步驟
更包括：

判斷該收音音量是否大於音量門檻值。

【第5項】 如申請專利範圍第3項所述的視訊與音訊處理裝置，其
中該第一音訊資料更包括該收音裝置的收音雜訊比，而該處理器
用以判斷該第一音訊資料的該聲音參數是否符合該收音標準的步
驟更包括：

判斷該收音雜訊比是否大於雜訊比門檻值。

【第6項】 一種視訊與音訊處理裝置，包括：

通訊電路，用以連接至至少一其它視訊與音訊處理裝置以及
遠端裝置；

影像擷取裝置，用以擷取影像訊號；

收音裝置，用以擷取聲音訊號；

記憶體，用以儲存檔案資料；以及

處理器，耦接該通訊電路、該影像擷取裝置、該收音裝置以及該記憶體，用以執行下列步驟：

取得至少一筆候選視訊資料以及多筆待整合音訊資料，其中各所述候選視訊資料對應於其中一筆所述待整合音訊資料；

根據各所述候選視訊資料所對應的該待整合音訊資料，選擇其中一筆所述候選視訊資料，以產生篩選後的視訊資料；

整合所述待整合音訊資料，以產生整合後的音訊資料；以及

透過該通訊電路傳輸該整合後的音訊資料以及該篩選後的視訊資料至該遠端裝置。

【第7項】如申請專利範圍第6項所述的視訊與音訊處理裝置，其中該處理器更用以執行下列步驟：

透過該通訊電路自該遠端裝置接收遠端視訊資料以及遠端音訊資料；以及

透過該通訊電路傳輸該遠端視訊資料以及該遠端音訊資料至各所述其它視訊與音訊處理裝置。

【第8項】如申請專利範圍第7項所述的視訊與音訊處理裝置，更包括：

螢幕，用以顯示畫面；以及

播音裝置，用以播放聲音；

其中該處理器更耦接至該螢幕以及該播音裝置，並且用以執行下列步驟：

分別透過該螢幕以及該播音裝置播放該遠端視訊資料以及該遠端音訊資料。

【第9項】 如申請專利範圍第6項所述的視訊與音訊處理裝置，其中各所述待整合音訊資料包括收音音量，而該處理器根據各所述候選視訊資料所對應的該待整合音訊資料，選擇其中一筆所述候選視訊資料，以產生該篩選後的視訊資料的步驟包括：

選擇具有最大收音音量的該待整合音訊資料所對應的該候選視訊資料做為該篩選後的視訊資料。

【第10項】 如申請專利範圍第6項所述的視訊與音訊處理裝置，其中各所述待整合音訊資料包括收音雜訊比，而該處理器根據各所述候選視訊資料所對應的該待整合音訊資料，選擇其中一筆所述候選視訊資料，以產生該篩選後的視訊資料的步驟包括：

選擇具有最大收音雜訊比的該待整合音訊資料所對應的該候選視訊資料做為該篩選後的視訊資料。

【第11項】 如申請專利範圍第6項所述的視訊與音訊處理裝置，其中各所述待整合音訊資料包括收音時間，而該處理器根據各所述候選視訊資料所對應的該待整合音訊資料，選擇其中一筆所述候選視訊資料，以產生該篩選後的視訊資料的步驟包括：

選擇具有最長收音時間的該待整合音訊資料所對應的該候選視訊資料做為該篩選後的視訊資料。

【第12項】 如申請專利範圍第6項所述的視訊與音訊處理裝置，其中各所述待整合音訊資料包括收音時間，而該處理器根據各所述音訊資料，選擇所述候選視訊資料其中之一者，以產生該篩選後的視訊資料的步驟包括：

選擇收音時間大於時間門檻值的該音訊資料所對應的該候選視訊資料做為該篩選後的視訊資料。

【第13項】 如申請專利範圍第6項所述的視訊與音訊處理裝置，其中該處理器用以整合所述待整合音訊資料，以產生整合後的音訊資料的步驟包括：

針對所述待整合音訊資料進行混音處理以及/或除噪處理，以產生該整合後的音訊資料。

【第14項】 如申請專利範圍第6項所述的視訊與音訊處理裝置，其中該處理器更用以執行下列步驟：

自該影像擷取裝置以及該收音裝置分別接收第二影像訊號以及第二聲音訊號，以產生第二視訊資料以及第二音訊資料；

判斷該第二音訊資料的聲音參數是否符合傳輸標準；

若是，分別設定該第二視訊資料以及該第二音訊資料為其中一筆所述候選視訊資料以及其中一筆所述待整合音訊資料；以及

若否，僅設定該第二音訊資料為其中一筆所述待整合音訊資料。

【第15項】 如申請專利範圍第14項所述的視訊與音訊處理裝置，其中該第二音訊資料包括音訊頻率，而該處理器用以判斷該第二音訊資料的該聲音參數是否符合該傳輸標準的步驟包括：

判斷該音訊頻率是否符合人聲頻率。

【第16項】 如申請專利範圍第14項所述的視訊與音訊處理裝置，其中該第二音訊資料更包括該收音裝置的收音音量，而該處理器用以判斷該第二音訊資料的該聲音參數是否符合該傳輸標準的步驟包括：

判斷該收音音量是否大於音量門檻值。

【第17項】 如申請專利範圍第14項所述的視訊與音訊處理裝置，其中該音訊資料更包括該收音裝置的收音雜訊比，而該處理器用以判斷該第二音訊資料的該聲音參數是否符合該傳輸標準的步驟更包括：

判斷該收音雜訊比是否大於收音雜訊比門檻值。

【第18項】 一種視訊會議系統，包括：

至少一客戶端裝置，其中各所述客戶端裝置分別產生客戶端視訊資料以及客戶端音訊資料；以及

主機裝置，連線於各所述客戶端裝置以及遠端裝置，其中該主機裝置取得至少一筆候選視訊資料以及多筆待整合音訊資料，根據各所述待整合音訊資料，選擇其中一筆所述候選視訊資料，以產生篩選後的視訊資料，整合所述待整合音訊資料，以產生整合後的音訊資料，傳輸該整合後的音訊資料以及該篩選後的視訊

資料至該遠端裝置，其中所述待整合音訊資料包括所述客戶端音訊資料以及該主機音訊資料，所述候選視訊資料包括至少其中一筆所述客戶端視訊資料以及主機視訊資料，各所述候選視訊資料對應於其中一筆所述音訊資料。

【第19項】 如申請專利範圍第18項所述的視訊會議系統，其中對於各所述客戶端裝置：

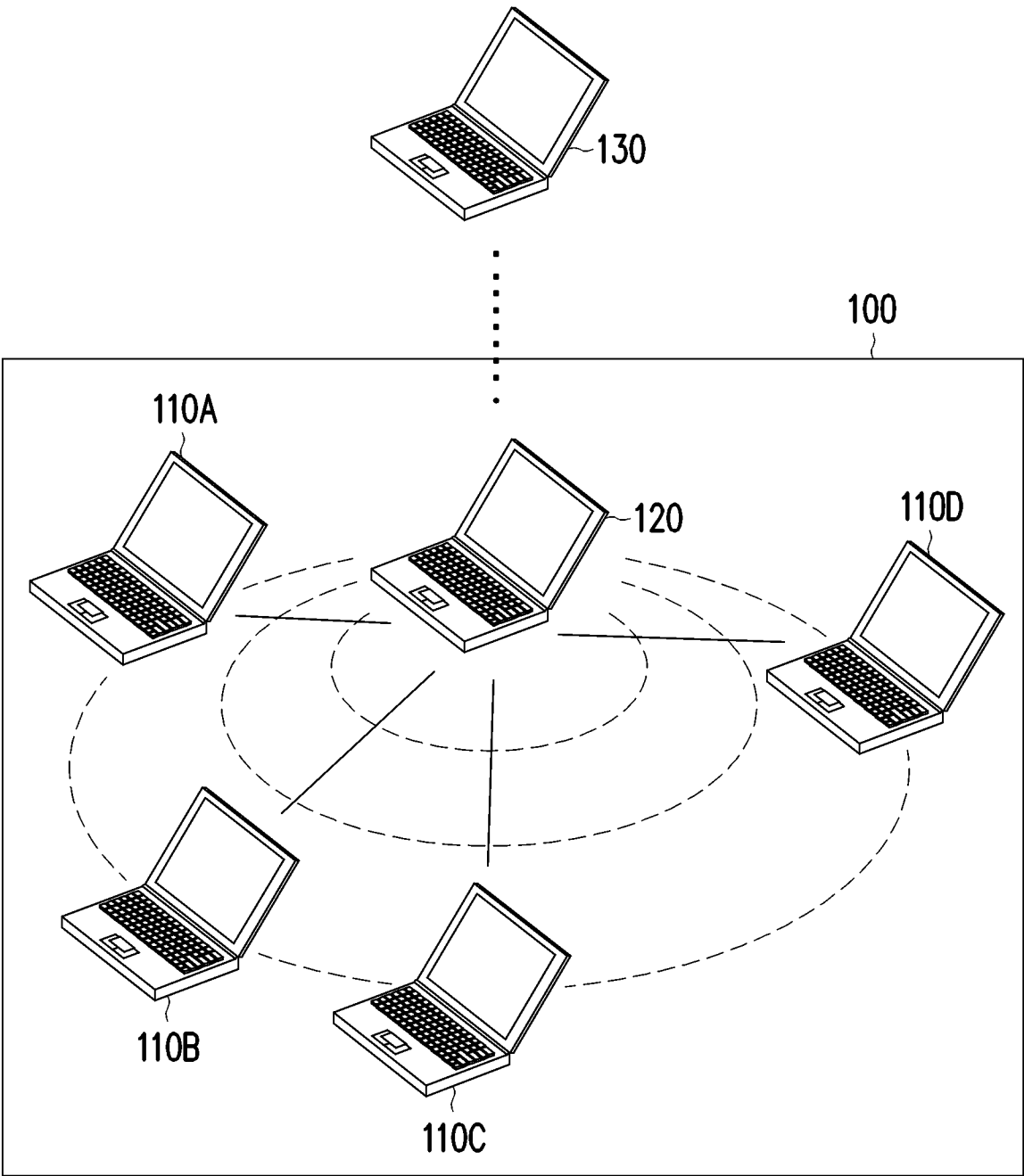
該客戶端裝置判斷該客戶端音訊資料的聲音參數是否符合收音標準；

若是，該客戶端裝置傳送該客戶端音訊資料以及該客戶端視訊資料至該主機裝置；以及

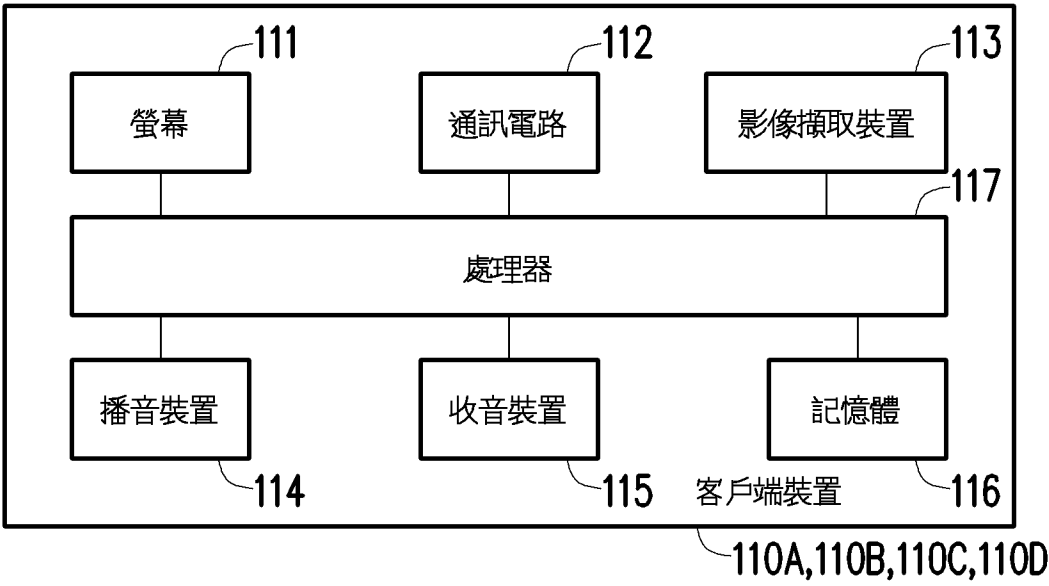
若否，該客戶端裝置僅傳送該客戶端音訊資料至該主機裝置。

【第20項】 如申請專利範圍第18項所述的視訊會議系統，其中該主機裝置更自該遠端裝置接收遠端視訊資料以及遠端音訊資料並且傳送該遠端視訊資料以及該遠端音訊資料至各所述客戶端裝置。

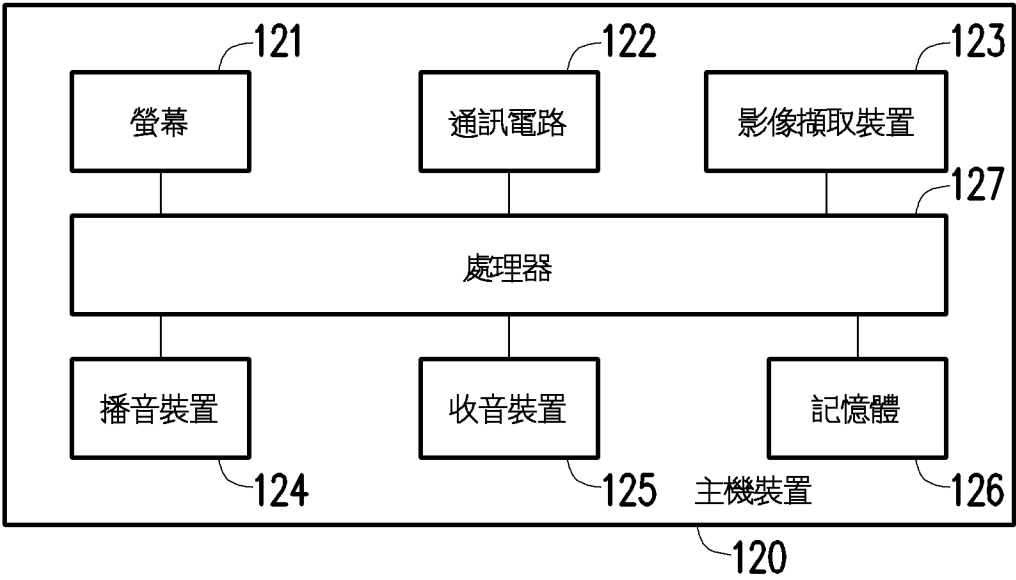
【發明圖式】



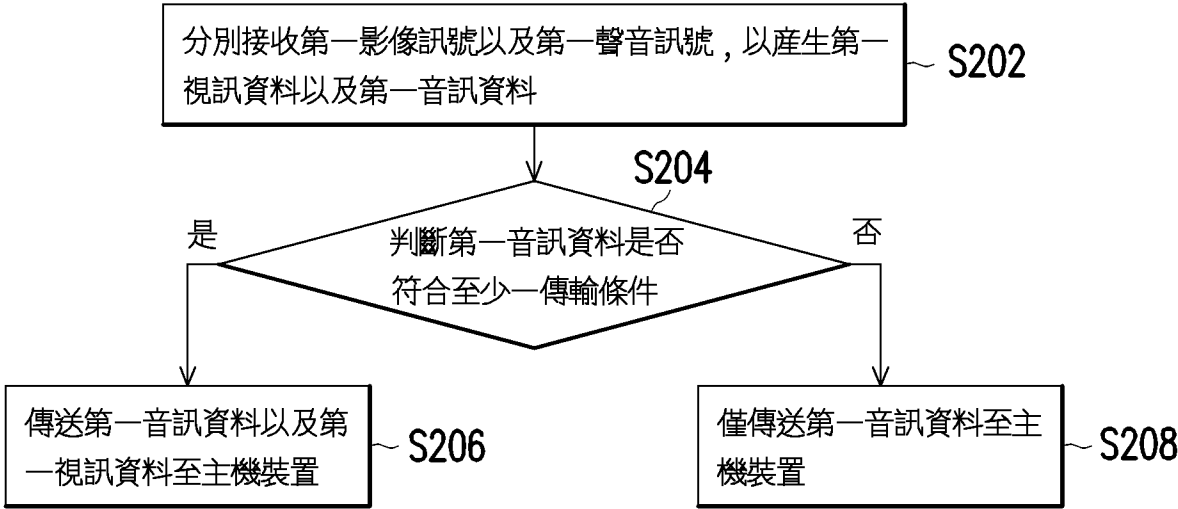
【圖1A】



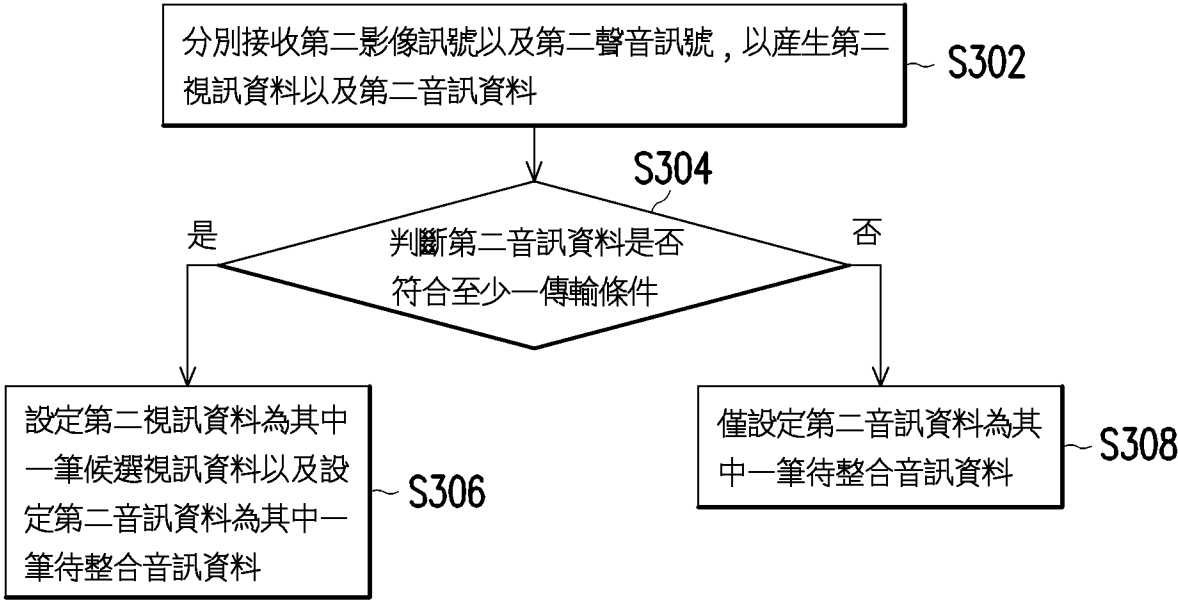
【圖1B】



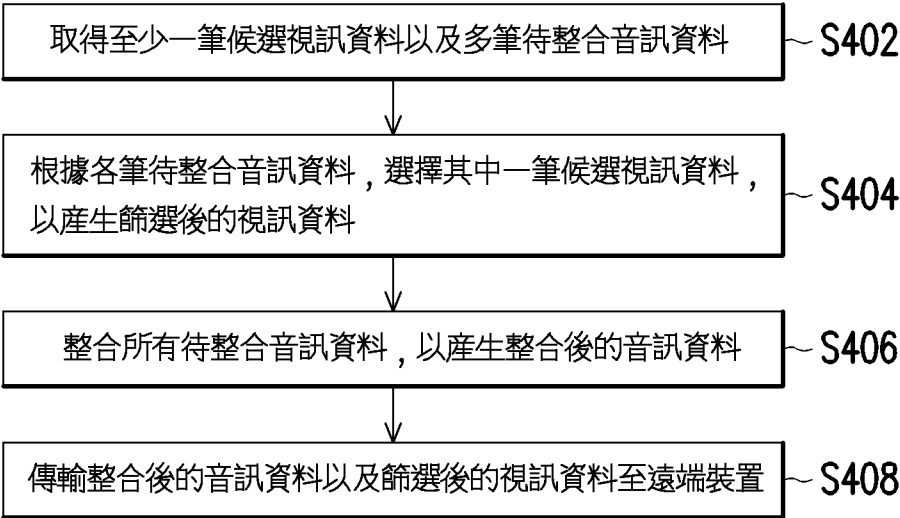
【圖1C】



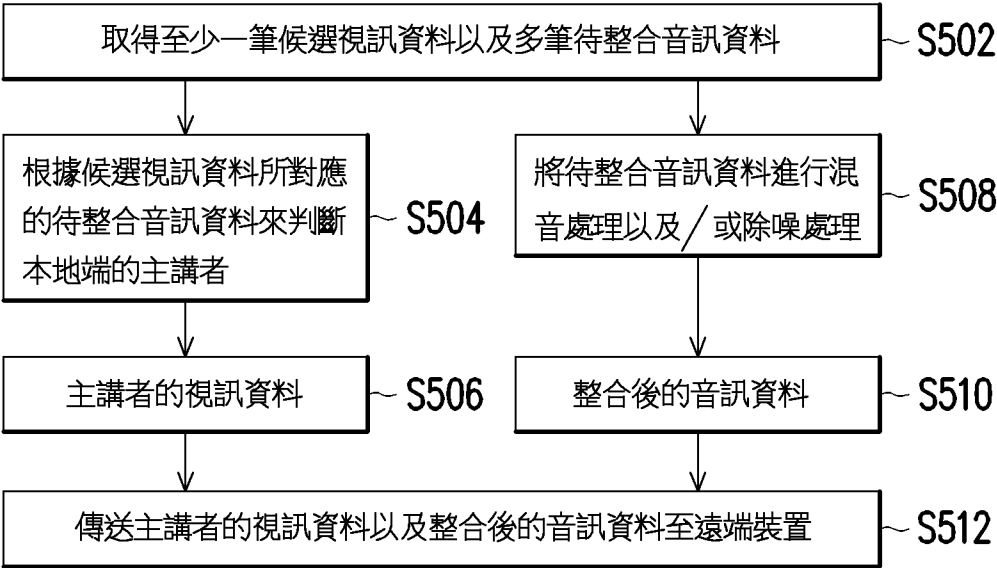
【圖2】



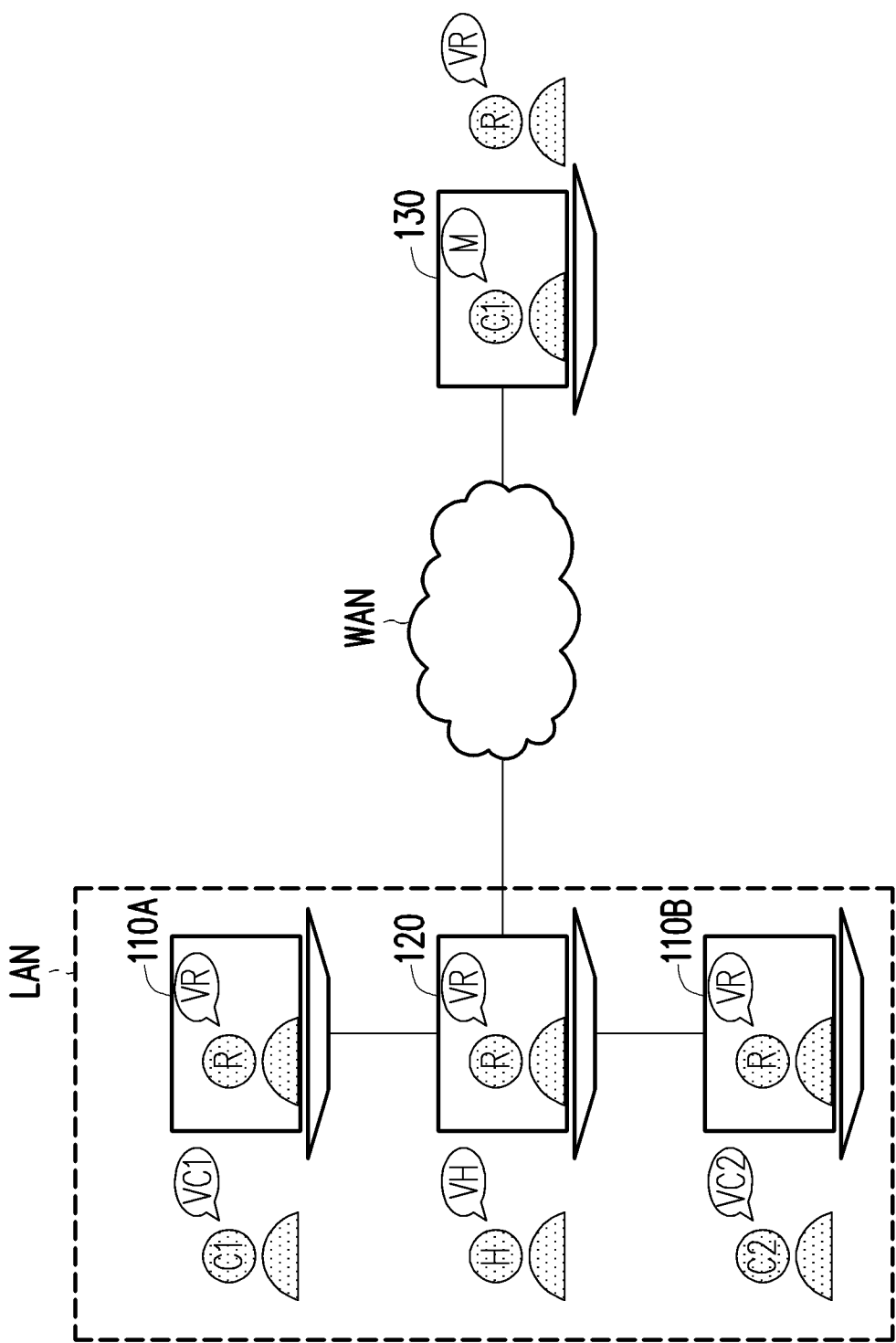
【圖3】



【圖4】



【圖5A】



【圖5B】