



[12] 发明专利申请公开说明书

[21] 申请号 03147129.3

[43] 公开日 2004年3月3日

[11] 公开号 CN 1479525A

[22] 申请日 2003.6.27 [21] 申请号 03147129.3

[30] 优先权

[32] 2002. 6.27 [33] US [31] 10/184,499

[71] 申请人 微软公司

地址 美国华盛顿州

[72] 发明人 R·卡特勒 I·塔斯海弗

A·古普塔 芮 勇

[74] 专利代理机构 上海专利商标事务所

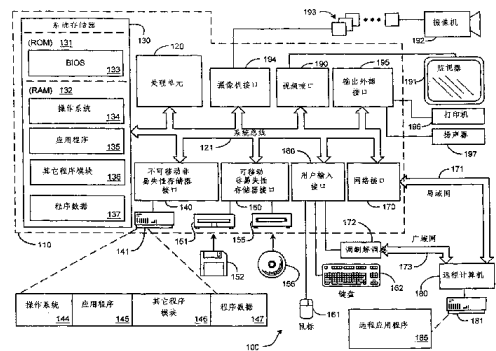
代理人 李家麟

权利要求书 4 页 说明书 12 页 附图 6 页

[54] 发明名称 全向摄像机与话筒阵列的集成设计

[57] 摘要

提出一种集成了话筒阵列的全向摄像机(360°摄像机)。这种摄像机主要用于电视会议和会议录制,并把该装置设计成置于会议桌上。话筒阵列系平面结构,话筒尽量靠近桌面以消除桌面的声反射。摄像机用细柱棒接至话筒阵列基座,在50~4000Hz频率范围内,细柱棒在声学上不被话筒阵列看到,这样从言者到阵列中所有话筒提供了直接通路,因而可用于声源定位(测定发言者位置)和束形成(通过只滤波来自特定方向的声音而提高发言者的音质)。摄像机阵列从桌面升高,对与会者提供正面近视点。



1. 一种捕获音视频数据的系统，其特征在于包括：
一台或多台捕获视频图像数据的摄像机；
捕获音频信号的话筒阵列；和
把所述一台或多台摄像机接到所述话筒阵列的柱体，所述柱体在人话音频率范围内呈声学透明。
2. 如权利要求 1 所述的系统，其特征在于，所述一台或多台摄像机在 360° 内拍摄图像。
3. 如权利要求 1 的系统，其特征在于，还包括话筒阵列基座，其中所述话筒装到所述话筒阵列基座上。
4. 如权利要求 3 所述的系统，其特征在于，所述话筒阵列基座的高度小得足以将所述话筒阵列定位成尽量减小该阵列从任何表面拾取的声反射。
5. 如权利要求 1 所述的系统，其特征在于，柱体在声学上透明的频率范围为 50~4000Hz。
6. 如权利要求 1 所述的系统，其特征在于，将音频信号输入计算机，所述计算机利用所述音频信号作声源定位，所述声源定位测定声音产生音频信号的方向。
7. 如权利要求 1 所述的系统，其特征在于，将音频信号输入计算机，所述计算机利用所述音频信号作束形成，所述束形成通过滤除来自所有其它方向的声音，提高了来自一个方向的声音的音质。
8. 如权利要求 1 所述的系统，其特征在于，将音频信号输入计算机，所述计算机用人检测与跟踪算法精密地探测视频图像数据中捕获的人，并把该人的视频图像数据与发言者产生的音频信号关联起来。
9. 如权利要求 1 所述的系统，其特征在于，将视频图像数据输入计算机，所述计算机利用所述视频图像数据缝合诸图像。
10. 如权利要求 1 所述的系统，其特征在于，话筒阵列中的话筒为全向。
11. 如权利要求 1 所述的系统，其特征在于，话筒阵列中的话筒为单向。
12. 如权利要求 1 所述的系统，其特征在于，话筒阵列含 3 只话筒。
13. 如权利要求 1 所述的系统，其特征在于，话筒阵列含 8 只话筒。
14. 如权利要求 1 所述的系统，其特征在于，话筒阵列含多只在围绕平面

话筒基座周边的圆中等边设置的话筒。

15. 如权利要求 14 所述的系统，其特征在于，把至少一只话筒装在橡胶箱里以防外来的声反射。

16. 如权利要求 14 所述的系统，其特征在于，在至少一只话筒下面安置隔音材料，以防外来的声反射。

17. 如权利要求 1 所述的系统，其特征在于，所述摄像机还包括镜头盖，在正常操作模式中向上，在保密模式中向下。

18. 如权利要求 17 所述的系统，其特征在于，话筒在摄像机处于被唤醒的所述保密模式时被切断。

19. 如权利要求 1 所述的系统，其特征在于，摄像机上的灯在摄像机工作时发亮。

20. 如权利要求 1 所述的系统，其特征在于，音频信号在网上发射。

21. 如权利要求 1 所述的系统，其特征在于，音频信号保存到计算机可读媒体上。

22. 如权利要求 1 所述的系统，其特征在于，视频图像数据在网上发射。

23. 如权利要求 1 所述的系统，其特征在于，视频图像数据保存到计算机可读媒体上。

24. 如权利要求 1 所述的系统，其特征在于，视频信号用 1394 总线传给计算机。

25. 如权利要求 1 所述的系统，其特征在于，音频信号用模拟电缆传给计算机。

26. 如权利要求 1 所述的系统，其特征在于，所述摄像机是 IEEE1394 摄像机。

27. 一种对电话会议和会议录制捕获音视频的处理，其特征在于包括以下处理动作：

用全向摄像机拍摄与会者图像，

用话筒阵列捕获所述会议中出现的声音的音频信号，所述话筒阵列装在基座中，后者用声学透明棒接至所述全向摄像机。

28. 如权利要求 27 所述的处理，其特征在于，所述话筒阵列置于桌上，所述与会者围绕所述桌上就座。

29. 如权利要求 28 所述的处理，其特征在于，话筒阵列尽量靠近所述桌

子，以尽量减少桌子的声反射被阵列拾取。

30. 如权利要求 27 所述的处理，其特征在于，话筒阵列包括 3 只或更多的话筒。

31. 如权利要求 30 所述的处理，其特征在于，所述话筒阵列包括在话筒基座周边附近圆周内等间隔的话筒。

32. 如权利要求 27 所述的处理，其特征在于，全向摄像机包括多个视频传感器。

33. 如权利要求 32 所述的处理，其特征在于，所述视频传感器围绕一圆周等间隔设置。

34. 如权利要求 27 所述的处理，其特征在于，全向摄像机包括单个配有捕获 360° 光线的双曲面镜头的视频传感器。

35. 如权利要求 27 所述的处理，其特征在于，所述声学透明棒为空心。

36. 如权利要求 27 所述的处理，其特征在于，所述声学透明棒附接于话筒基座，而摄像机电缆通过声学透明棒，故所述电缆不造成声障。

37. 如权利要求 27 所述的处理，其特征在于，所述声学透明棒的外径为 2 厘米或更小。

38. 如权利要求 27 所述的处理，其特征在于，声学透明棒的直径使来自发言者的 50~4000Hz 音频绕过声学透明棒几乎无干扰地到达话筒阵列中的所有话筒。

39. 如权利要求 27 所述的声学透明棒，其特征在于，所述棒高度可调节。

40. 一种集成式全向摄像机与话筒阵列，其特征在于包括：

全向摄像机；

声学透明棒，用于升高所述摄像机而提供优化摄像机覆盖；和

话筒阵列。

41. 如权利要求 40 所述的集成式全向摄像机与话筒阵列，其特征在于，所述全向摄像机用多个视频传感器实现 360° 摄像机覆盖。

42. 如权利要求 40 所述的集成式全向摄像机与话筒阵列，其特征在于，所述全向摄像机应用一个视频传感器和一块捕获 360° 光线而实现全景覆盖的双曲面镜头。

43. 如权利要求 40 所述的集成式全向摄像机与话筒阵列，其特征在于，声学透明柱棒升高所述摄像机，对围桌就座的人提供正面图像。

44. 如权利要求 40 所述的集成式全向摄像机与话筒阵列，其特征在于，话筒阵列包括多只绕圆周等边距离设置的话筒。

45. 如权利要求 40 所述的集成式全向摄像机与话筒阵列，其特征在于，话筒阵列置于表面，并尽量接近表面以减少来自所述表面的声发射。

46. 如权利要求 41 所述的集成式摄像机与话筒阵列，其特征在于，用视场较窄长的视频传感器拍摄较长距离的图像，用视场较短宽的视频传感器拍摄较短距离图像。

47. 一种全向话筒阵列，其特征在于包括：

多只在绕平面话筒基座周边的圆内等边设置的话筒；而其中

所述话筒基座低得足以把所述话筒阵列定位得尽量减少来自话筒基座落坐的任何表面为阵列拾取的声反射。

48. 如权利要求 47 所述的话筒阵列，其特征在于，把所述多只话筒捕获的音频信号输入计算机，所述计算机用所述音频信号作声源定位，所述声源定位测定产生音频信号的声音方向。

49. 如权利要求 47 所述的话筒阵列，其特征在于，把所述多只话筒捕获的音频信号输入计算机，所述计算机用所述音频信号作束形成，所述束形成通过过滤除来自所有其它方向的声音而提高来自一个方向的声音的音质。

50. 如权利要求 47 所述的话筒阵列，其特征在于，话筒阵列的话筒为全向型。

51. 如权利要求 47 所述的话筒阵列，其特征在于，话筒阵列的话筒为单向型。

52. 如权利要求 47 所述的话筒阵列，其特征在于，话筒阵列含 8 只话筒。

53. 如权利要求 47 所述的话筒阵列，其特征在于，至少一只话筒装在橡胶箱里以防外来声反射。

54. 如权利要求 47 所述的话筒阵列，其特征在于，在至少一只话筒下面安置隔音材料，以防外来声反射。

全向摄像机与话筒阵列的集成设计

背景

技术领域

本发明针对一种集成式全向摄像机与话筒阵列，尤其针对能用于电话会议与会议记录的集成式全向摄像机与话筒阵列。

背景技术

电视会议系统的商业成功为数不多，这有多种原因，尤其是这些系统通常有多种技术上的缺陷。摄像机视点不佳和图像分辨率不足，使与会者难以看清发言者。这样再加上不准确的发言者探测（对配备全景—倾斜—变焦摄像机的系统尤其如此），使摄像机对不准发言者。另外，低劣的视频压缩技术常常导致电视像质差和图像显示“不连贯”。

用于电话会议的系统的拍摄装置，倾向于关注少数几个对电视会议和会议观点有价值的数据源，包括视频数据、音频数据和计算机监视器上示出的电子文件或说明。假定有共享文件与说明的各种软件方法，则以改进的方法捕获音视频数据尤受关注。

捕获视频数据有三种不同方法：全景/倾斜/变焦（PTZ）摄像机。基于反光镜的全向摄像机和摄像机阵列。尽管目前大多选用 PTZ 摄像机，但是有两大限制。首先，它们只能拍摄有限的视场，若它们变焦得太近，会丢失会议室场景；若变焦得过远，就看不出人的表情。其次，因控制电机要花移动摄像机的时间，故摄像机对会议的反应（如发言者切换）很慢。实际上 PTZ 摄像机不能移动过多过快，否则会分散人们收看会议的注意力。

假定这些缺点和当前的技术在基于反光镜/棱镜的全向视觉传感器中有了改进，研究人员已开始重新考虑捕获与分析视频的方法，如在娱乐、新闻与体育网上广播中，BeHere 公司提供了 360° 因特网视频技术。运用它的接口，远程用户可以控制个人化的 360° 摄像机角度而得到“身临其景”的体验，与其它观众无关。虽然该法克服了 PTZ 摄像机面临的视场有限和摄像机反应慢这两大难题，但是按当今的技术和市场需求来衡量，制作这类装置太昂贵了。此外，

这些基于反光镜/棱镜的全向摄像机还存在分辨率低（即使有 1MP 个传感器）与散焦的问题，导致视频质量低下。

在另一方法中，把多个廉价的摄像机或视频传感器组装成全向摄像机阵列，如一种已知系统用 4 台美国国家电视系统委员会（NTSC）摄像机构建会议室全景。但这种设计存在若干缺点。首先，NTSC 摄像机提供质量相对低的视频信号。另外，在能对信号分析、发射或记录之前，4 台摄像机要用 4 块视频拍摄板对信号数字化。要用 4 块视频拍摄板增大了这类系统的成本和复杂性，更难制造和保养。

除了视频拍摄的问题外，在会议室中捕获高质量音频也提出了挑战。音频捕获系统要求除去各种噪声与混响，还须调整不同输入信号电平的增益。一般有三种方法可满足这些要求。最简方法是使用紧贴话筒（如通过头戴式话筒），但对用户/发言者有点不便。第二种方法是在会议桌上放一话筒，这样避免了多条声路，是目前最常用的会议音频录制法。这些系统用若干（一般 3 只）超心形话筒提供全向特性。第三种方法设置于桌上电话会议系统，该法在指向发言者的 PTZ 摄像机顶部装一单向话筒，摄像机/话筒组由计算机控制，计算机用一组独立的话筒作声源定位。但该法要用两组独立的话筒。

发明内容

本发明针对一种可克服电视会议与会议记录系统中上述诸局限性的系统与处理，尤其是本系统与方法用集成式全向摄像机与话筒阵列完成这一任务。

在最一般的意义上，本发明包括一柱形棒，它们细得足以在声学上不被人讲话频率范围（50~4000Hz）看见，并将摄像机阵列接至话筒阵列，因而消除了声音绕射与声影。

该集成式摄像机与话筒阵列应用的 360° 摄像机，设计成可克服上述各电视会议问题。360° 摄像机位于会议桌中央，与典型的电视会议系统（摄像机在会议室一端）相比，与会者的摄像机视点甚佳。摄像机从桌面升高，可提供与会者接近的正面视点。另外，集成式摄像机与话筒阵列提供足够的分辨率，让远程观众看清与会者的面部表情（如在一工作实施例中，分辨度为 300×480）。摄像机可任一全向型，或应用摄像机阵列，或应用带双曲反光镜的单一视频传感器。

话筒阵列是一种平面结构，话筒较佳地装在话筒阵列基座里，尽量靠近桌

面以消除会议桌声反射。如前所述，摄像机用细柱棒接至话筒阵列基座，对人话音频率范围（即约 50~4000Hz）而言，细柱棒对话筒阵列在声学上不可见。这样从发言者到阵列里的所有话筒形成一直达通路，可更好地实现声源定位（测定发言者位置）和束形成（通过滤除非来自发言者方向的声音而提高发言者声音品质）。集成话筒阵列用于实行实时声源定位，而摄像机阵列与基于计算机视觉的人检测与跟踪一起用来准确地探测发言者处于图像中的位置。基于音视频的发言者探测可用于自动摄像机管理，大大提高了视频压缩（如对面部区域比背景应用更多的位）。

集成式摄像与话筒阵列的输出较佳地接到 PC，此类应用像图像缝合与压缩一样，可实行声源定位、束形成和摄像机管理。

集成式摄像机与话筒阵列的一工作实施例，利用 1394 总线将视频传给 PC，并用模拟电缆把音频传给个人计算机（PC）。本例应用了 5 台提供优质视频的 IEEE1394 摄像机，只需单个 1394 卡。另一实施例对所有摄像机与话筒应用单块印刷电路板（PCB），使所有音视频都通过单根 1394 电缆发射。该 1394 电缆还供电力，故摄像机与 PC 间仅需单根电缆。

应用的话筒可以是全向或单向型，但全向型为优选，因其能对所有感兴趣的声角提供均一响应。话筒最小需求数为 3 个，尽管本发明一较佳实施例用了 8 只，以提高声源定位精度，整个音频系统的束形成与耐用性更佳。话筒较佳地在围绕圆形平面话筒基座四周的圆圈内等边的安置，虽然其它配置也行。使用的话筒越多，全向音频覆盖和信噪比越佳。然而，要考虑到成本与大量话筒的复杂性。再者，话筒多了。音频信号处理就更复杂。为减少桌面噪声，可将话筒装在橡胶箱里，话筒下面安置隔声体。

摄像机可应用一块镜头屏，正常操作模式时朝上，保密模式时朝下。或者在保密模式时，可以关闭摄像机传感器的快门，或以电子技术隔离摄像机将其切断。在唤醒保密模式时，最好也关闭话筒。录制期间，摄像机上的灯光发亮，让用户知道摄像机在工作。当摄像机处于保密模式时，关闭灯光。

该集成式全向摄像机与话筒设计可以有各种其它实施例，其部分原因在于系统的模块化特征。如在一实施例中，所用的全向摄像机用多个视频传感器实现 360° 摄像机覆盖。而在本发明另一实施例中，应用的一台全向摄像机，利用一个视频传感器和一块捕获 360° 光的双曲面镜头实现全景覆盖。再者，它们自己已可以应用在声学透明的柱棒上升高的任一种此类摄像机装置，以提供

与会者的正视图。或者它们能与上述的话筒阵列集成在一起。或者，其它摄像机设计也可与柱棒结合使用。连接摄像机与话筒阵列的棒不一定为柱形，只要细得在 50~4000Hz 范围内不绕射声音就行。

同样地，如前所述，在一实施例中，为对室内任一发言者实现通畅的通路且声音离桌面最小的反射，话筒阵列包含的话筒围绕圆周以等边距离设置得尽量靠近桌面。然而，利用声学透明棒能与全向摄像机装置集成的其它话筒配置也可行。另外，可以使用刚才讨论的全向话筒阵列而不用任何摄像机实现 360° 声覆盖。这种覆盖因减小或消除了多径问题，故尤其适用于声源定位与束成形。

应用本发明的摄像机与话筒阵列的一实施例，用计算机优化图像数据和音频信号。摄像机的数字图像输出和话筒阵列的音频输出（经模/数转换器）传入计算机，后者以各种功能增强并运用图像与音频输入，如全景图像滤波器把全向摄像机中各传感器拾取的图像缝合在一起。另外，图像数据经压缩，使它更适合在网上广播（如因特网），或较佳地通过将准备发送和/或录制的音视频输出分离的分离器，保存到计算机可读媒体。或者，还可把图像数据输入个人检测器/跟踪器以改进摄像机管理，如图像/视频包含该发言者的部分经识别，与音频信号相关联，从而在他们说话时把电视会议显示的摄像机视图对准该发言者。另外，使面部区域比背景具有更大的分辨率，可用发言者位置改进视频压缩。

音频输入还可用于各种用的，如将音频用于声源定位，从而在任何指定时刻对发言者方向优化音频。另外，可在计算机中用束形成模块改善音频的束形状，从而改善对指定方向音频的滤波。还可用减噪与自动增益控制模块改善信噪比，其方法是相对室内背景噪声，减少噪声并调节增益，以更好地捕获来自发言者的音频信号。这些图像和音频处理模块可以单独或组合使用，或根本不用。

音视频信号无论增强与否，都到播送到另一电视会议地点或因特网，也可保存到计算机可读媒体供以后观看。

上述集成式摄像机与话筒阵列主要用于电视会议和会议录制。把话筒阵列与全向摄像机集成起来，明显简化了音视频之间必要的校正（精密制造的摄像机和话筒阵列无须校正），可用单个装置获取会议室的音视频信息。

附图简介

通过下面的描述、所附如权利要求和附图，能更好地理解本发明的专用特征、方面和优点，其中：

图 1 示出构成实施本发明一示例系统的通用计算装置。

图 2 示出集成式摄像机与话筒阵列在会议桌上较佳的定位。

图 3A 是集成式摄像机与话筒阵列一实施例的侧视图。

图 3B 是图 3A 所示集成摄像机与话筒阵列实施例的俯视图。

图 4 是集成式摄像机与话筒阵列另一实施例的透视图。

图 5 是本发明用计算机增强捕获的音视频数据的一工作实施例。

较佳实施例的详细描述

1.0 示例的工作环境

在下述本发明诸较佳实施例中，要参照构成本发明一部分的诸附图，图中以示例方式示出可实施本发明的特定实施例。应理解，也可应用其它实施例并作出结构上的变化而不违背本发明的范围。

图 1 示出一例可对其实施本发明的合适的计算系统环境 100，该环境 100 仅是一例合适的计算环境，并不对发明的应用范围或功能提出任何限制。计算环境 100 既不解释为具有任何相依性，也不解释为相对示例工作环境 100 所示任一元件或元件组合有任何要求。

本发明与各种其它通用或专用计算系统环境或配置工作。适用本发明的已知计算系统、环境和/或配置的实例，包括但不限于个人计算机、伺服计算机、手持或膝上装置、多处理器系统、基于微处理器的系统、机顶盒、可编程消费类电子产品、网络 PC、小型计算机、主机架计算机、包括以上任一系统或装置的分布计算环境等。

本发明以计算机可执行指令的一般范围来描述，如计算机执行的程序模块。程序模块通常包括执行特定任务或实施特定抽象数据类型的子程序、程序、目标、元件、数据结构等。本发明还可在分布计算环境中实施，其中由通过通信网连接的远程处理装置执行诸任务。在分布计算环境中，程序模块可设置在本地和远程计算机存贮媒体里，包括存储器存贮装置。

参照图 1，实施本发明的示例系统包括计算机 110 形式的通用计算装置。计算机 110 的部件包括但不限于处理单元 120、系统存储器 130 和系统总线 121，总线 121 把包括系统存储器的各种系统部件耦至处理单元 120。系统总线 121

可以是若干类总线结构的任一种，包括存储器总线或存储器控制器、外设总线和使用任一种总线结构的本地总线。举例来说，但不作限制，这类结构包括工业标准结构（ISA）总线、精微通道结构（MCA）总线、增强型 ISA（EISA）总线、视频电子学标准协会（VESA）本地总线以及同时称为 Mezzanine 总线的外设部件互连（PCI）总线。

计算机 110 一般包括各种计算机可读媒体，可以是任一种计算机 110 能存储的现有媒体，包括易失与非易失媒体、可卸与不可卸媒体。举个例但不作限制，计算机可读媒体包括计算机存储媒体和通信媒体。计算机存储媒体包括以任一方法与技术构制实现的易失与非易失、可卸与不可卸媒体，用于存储计算机可读指令、数据结构、程序模块或其它数据等信息。计算机存储媒体包括但不限于 RAM、ROM、EEPROM、闪速存储器或其它存储器技术、CD-ROM、数字通用盘（DVD）或其它光盘存储器、盒式磁带、磁带、磁盘存储器或其它磁存储装置，或者任何其它可用于存储所需信息并为计算机 110 存储的媒体。通信媒体一般实施计算机可读指令、数据结构、程序模块或载波等调制的数据信号或其它传输机构中的其它数据，包括任何信息提供媒体。“调制的数据信号”表示一种具有其一组或多组特性或以编码信号中信息的方式变化的信号。举个例但不作限制，通信媒体包括有线网或直接导线连接等有线媒体和声学、射频、红外等无线媒体与其它无线媒体。以上的任一种组合也包括在计算机可读媒体范围内。

系统存储器 130 包括易失和/或非易失存储器形式的计算机存储媒体，如只读存储器（ROM）131 与随机存取存储器（RAM）132。基本输入/输出系统 133（BIOS）一般贮存在 ROM131 里，它包含的基本子程序帮助在计算机 110 内诸单元之间传递信息，如在启动期间。RAM132 通常包含的数据和/或程序模块，可直接存取和/或现在由处理单元 120 对其起作用。举个例但不作限制，图 1 示出了操作系统 134、应用程序 135、其它程序模块 136 和程序数据 137。

计算机 110 还可包括其它可卸/不可卸、易失/非易失计算机存储媒体。仅举个例，图 1 示出 3 对不可卸非易失磁性媒体读或写的硬盘驱动器 141、对可卸非易失磁盘 152 读或写的磁盘驱动器 151，以及对 CD-ROM 等可卸非易失光盘 156 或其它光媒体读或写的光盘驱动器 155。能用于示例工作环境的其它可卸/不可卸、易失/非易失计算机存储媒体，包括但不限于盒式磁带、闪速存储器卡、数字通用盘、数字像带、固态 RAM、固态 ROM 等。硬盘驱动器 141 一般通

过接口 140 等不可卸存储器接口接至系统总线 121，而磁盘驱动器 151 和光盘驱动器 155 通常靠接口 150 等可卸存储器接口接至系统总线 121。

上述图 1 所示的驱动器及其有关的计算机存贮媒体，为计算机 110 存贮了计算机可读指令、数据结构、程序模块和其它数据。如在图 1 中，硬盘驱动器 141 被示为存贮操作系统 144、应用程序 145、其它程序模块 146 和程序数据 147。注意，这些部件可与操作系统 134、应用程序 135、其它程序模块 136 和程序数据 137 相同或不同。这里对操作系统 144、应用程序 145、其它程序模块 146 和程序数据 147 给出不同的标号，至少示明它们是不同的复制品。用户通过输入装置把指令和信息送入计算机 110，如键盘 162 和常称为鼠标、转球或触板的定向装置 161。其它输入装置（未示出）包括话筒、摇杆、游戏板、卫星盘、扫描器等。各种输入装置往往通过耦接系统总线 121 的用户输入接口 160 接至处理单元 120，也可用其它接口和总线结构如平行口、游戏口或通用串行总线（USB）等连接。通过视频接口 190 等接口，也将监视器 191 或其它类显示装置接至系统总线 121。除了监视器，计算机还包括扬声器 197 与打印机 196 等其它外围输出装置，可通过输出外围接口 195 连接。对本发明尤为重要，还将能拍摄一系列图像 164 的摄像机 163（如数字/电子静像或视频摄像机或胶卷/照相扫描器）包括为个人计算机 110 的输入装置。再者，虽然只示出一台摄像机，但是可包括多台作为个人计算机 110 的输入装置的摄像机。来自一台或多台摄像机的图像 164 经有关摄像机接口 165 输入计算机 110，接口 165 接系统总线 121，由此让图像传到并贮存于 RAM132 或一个与计算机 110 关联的其它数据存贮装置。但要指出，图像数据也可从上述任一计算机可读媒体输入计算机 110，不要求使用摄像机 163。

利用对一台或多台远程计算机如远程计算机 180 的逻辑连接，计算机 110 可工作于网络环境。远程计算机 180 可以是个人机、服务器、路由器、网络 PC、同等装置或其它公共网节点，一般包括许多或全部以上对计算机 110 描述的单元，尽管图 1 只示出存储器存贮装置 181。图 1 中的逻辑连接包括局域网（LAN）171 和广域网（WAN）173，但还可包括其它网。这类网络环境在办公室、企事业计算机网、内联网和因特网中很常见。

在 LAN 网环境中使用时，计算机 110 经网接口或适配器 170 接 LAN171。在 WAN 网环境里使用时，计算机 110 一般包括在 WAN173 如因特网内建立通信的调制解调器 172 或其它装置，而调制解调器 172 可以是内部或外接，可通过用户

输入接口 160 或其它有关机构接系统总线 121。在网环境中，对计算机 110 或其部分示出的程序模块，可贮存在远程存储器存贮装置里。举个例但不作限制，图 1 示出的远程应用程序逻辑 85 驻留于存储器装置 181。图示的网连接显然是示例性的，可以使用在计算机之间建立通信链路的其它装置。

现已讨论了示例的工作环境，描述的其余部分将着重描绘实施本发明的程序模块。

3.0 集成式全向摄像机与话筒阵列

本节讨论经声学透明棒连接的集成式全向摄像机与话筒阵列。

3.1 概述

本发明解决了全向摄像机与话筒阵列的优化集成问题。设计目标是：

1. 话筒阵列设计应提供一条从发言者到阵列中所有话筒的通畅的通路，若做不到，则使用的任何声源定位和束形成算法就变得极其困难，结果劣化。
2. 话筒阵列设计应将话筒尽量置于靠近桌面或其它表面，以防声音从表面反射，否则会劣化任何声源定位和束形成结果。
3. 应升高摄像机阵列以提供与会者的正面近像。摄像机阵列应小得不令人注目。
4. 话筒阵列几何尺寸（话筒之间的位置和距离）应让音频处理算法实现良好的束形状，在工作频段内指向发言者，提供优质声音。

3.2 集成设计

集成式摄像机与话筒阵列应用了把话筒基座接至摄像机阵列的柱杆。在人话音频率范围内（50~400Hz），该柱杆在声学上不可见。

如图 2 所示，集成式摄像机与话筒阵列 202 预定置于会议桌 204 的中央。

该设计提供一条从任一指定发言者或声源到所有话筒的通畅的通路，而话筒阵列靠近桌面以防声音从桌子反射而造成多径问题。另外，该设计将摄像机从桌面升高，对所有与会者都提供正面视像或近像。

集成式摄像机与话筒阵列保证了良好的束形状，通过滤波只来自一个方向的声音，可提高发言者的音质。再者，摄像机与话筒的集成特征不需要重复校正，因而是有利的。由于摄像机与话筒集成可单一装置，故只需一次初始校正。而且由于集成式摄像机与话筒是一种小型的固定式设计，所以与要求独立的电缆和会议桌上额外空间的两种分离的摄像机与话筒部件相比，更显得无障碍。

3.3 系统部件

集成式全向摄像机与话筒阵列的一实施例示于图 3A 与 3B，该设计的诸部件包括全向摄像机 302、柱体 304、话筒基座 306、话筒 308、话筒前置放大器 310 和模/数 (A/D) 转换器 (未示出)。

3.3.1 全向摄像机

全向摄像机技术有多种，其中一种摄像机类型里的多个视频传感器以背对背方式紧密组装在一起。另一全向摄像机类型应用单个视频传感器，配用的双曲面镜头捕获 360° 的光线。本发明的集成式摄像机与话筒阵列设计可应用任一此类全向摄像机。较佳地，摄像机头 302 应小得在置于会议桌或其它表面上时无妨碍。

若采用多传感器摄像机配置，可应用多个摄像机或视频传感，最好为 8 个。这些传感器较佳地背对背安置，使各传感器的投射中心等角距分开。例如，若使用 8 只传感器，则各传感器与其邻近的传感器分开 45°。但若需拍摄不同距离的图像，可应用不同的镜头和不同的摄像机安置，如会议桌为矩形或椭圆形就是如此。距离较长，可以用视场窄长的镜头，距离较短，可用视场宽短的镜头拍摄图像，此时，摄像机传感器可以不围绕摄像机头等边设置。宽视场的摄像机传感器可以离窄视场的摄像机传感器更远。或者，也可应用视场可变的摄像机 (旋转和推拉变焦距镜头而调到指定状态)。

图 4 所示本发明一工作实施例，采用了密微的多传感器设计，其中把多个超小型摄像机传感器 402 围绕圆形摄像机头 404 周边背对背配置。摄像机 404 底部较佳地装有吸音材料，防止声音从桌子反射到摄像机与话筒。该例使用 8 只摄像机传感器。这种密微设计将摄像机传感器背对背地紧密封装在一起，改善了图像缝合。本例中，摄像头 404 的直径约 50mm，把摄像头接至基座的柱体 406 高约 10 英寸，这一高度把摄像机 404 置于可拍摄全体与会者的正面像。或者，摄像头 404 可以低于该高度，将摄像机传感器 402 稍微朝上指，也可使摄像机无大碍。

3.3.2 柱体

参照图 3A 与 3B 的实施例，全向摄像机 302 附接较佳的空心柱体 304，后者附接于话筒基座 306。柱体直径 D1 应细得足以让柱体 304 对相应于预定应用的频率在声学上呈透明。较佳地，摄像机电缆通过柱体 304 敷设而不造成任何附加的声障。柱体高度 H1 应高得足以对围绕会议桌就坐的每个人提供正面图像，但又不高得使摄像机显得突出。柱体高度还可选成可调节，如在电视会议

场合，可按不同的桌子和与会者高度调柱高。

再参照图 3A 与 3B 的工作实施例，柱体 304 的直径 $D1$ 为 2cm 或不到，使对应于人语音的约 50~4000Hz 音频从任一发言者基本上无干扰地通过该柱体传到基座里的所有话筒。本例中的柱高 $H1$ 为 14cm。

3.3.3 话筒基座

话筒基座通常保持话筒、话筒前置放大器和 A/D 转换器，它接至柱体，对摄像机电缆提供连接出口。话筒基座为低型面，以尽量缩小桌面与话筒的距离。基座使各话筒与与会者有一直接通路。

在图 3A 与 3B 的集成式摄像机与话筒阵列工作实施例中，话筒基座 306 的直径 $D3$ 比摄像头 302 的直径 $D2$ 更宽，使集成式摄像机与话筒阵列具有稳定性，防止容易被翻倒。基座高度 $H3$ 相对很小，该尺寸较佳地低得足以使埋置式话筒 308 保持接近桌面而避免多径问题，却高得足以让话筒 308 埋入基座。对该工作例， $D3$ 为 16cm， $H3$ 为 1.5cm， $H2$ 为 6cm， $D2$ 为 10cm。

3.3.4 话筒

使用的话筒可以是全向或单向型，但优选全向话筒，因它们对所有关注的声角都给出均一的响应。所需的最少话筒数为 3 个，但本发明该实施例用了 8 只，旨在提高声源定位精度，更好的束形成和整个音频系统的耐用性。

为减少桌面噪声，可把话筒装入橡胶箱，同样地在话筒下面放置隔音材料。

再参照图 3A 与 3B 的工作实施例，话筒 308 围绕平面话筒基座 306 上的圆周等边安置。本例应用 8 只话筒 308。一般，使用的话筒越多，全向音频覆盖与信噪比越好，但要考虑大量话筒的成本与复杂性。另外，若话筒更多，音频信号处理就更复杂。在图 3A 与 3B 的工作实施例中，柱体 306 与各话筒 308 的中心距离 $D5$ 为 7cm。

3.3.5 话筒前置放大器、A/D 转换器

话筒前置放大器 310 和模/数 (A/D) 转换器 (未示出) 较佳地汇集入话筒基座 306，如图 3B 所示。本例中，前置放大器 310 的宽度 $D4$ 为 5.901cm。话筒前置放大器放大来自话筒的信号，对后接的 A/D 转换器将信号幅值归一化。A/D 转换器把来自摄像机的模拟信号转换为数字信号。

本例中，对来自话筒的信号的信号取样相互同步在 1 微秒内，以利声源定位和束形成。

3.4 保密模式

摄像机应用镜头罩，在正常操作模式中打开，在保密模式中闭合。或在保密模式中关闭摄像机传感器的光闸，或以电子技术将摄像机隔离成关机。在唤醒保密模式时，最好也关闭话筒。录制期间，摄像机顶灯发亮，让用户知道摄像机在工作。接通保密模式时，灯关闭。

4.0 其它模块化特征在实施例

集成式全向摄像机与话筒设计的各种其它实施例是可行的，部分原因在于系统的模块化特征。

例如，可以应用各种摄像机实施例。在一实施例中，使用的全向摄像机应用了实现 360° 摄像机覆盖的多个视频传感器。而在本发明另一实施例中，使用的全向摄像机应用一个视频传感器和一个捕获 360° 光线以实现全景覆盖的双曲面镜头。再者，这些摄像机可在声学透明柱棒上自己升高，对与会者提供正面图像。这些摄像机可与话筒阵列集成。或者，其它全向摄像机设计也可结合使用柱棒和/或话筒阵列。

同样地，可以采用各种话筒配置法。在一实施例中，话筒阵列包括的诸话筒围绕圆周等边距设置，并尽量靠近桌面，对室内任一发言者引成通畅的通路。然而，其它话筒配置法可用声学透明棒与摄像机集成起来。或者，可应用刚才讨论的全向话筒阵列，无须摄像机实现优化的 360° 声覆盖。这种覆盖由于减少或消除了多径问题，在声源定位和束形成方面尤其有用。

在集成式摄像机与话筒阵列的一实施例中，在 PC 上作图像缝合与压缩。另一实施例则在摄像机中以现场可编程门阵列 (FPGA) 或其它门阵列执行图像缝合与压缩，该设计用 USB 接口使摄像机与 PC 接口，并让 PC 用更多的 CPU 循环执行其它任务，如图像压缩和录播会议。

5.0 示例工作实施例

图 5 示出应用本发明摄像机 502 和话筒阵列 504 的一工作实施例，摄像机 502 的图像输出和话筒阵列 504 的音频输出，经模/数转换器 506 传给计算机 508，后者执行各种功能以增强并利用图像与音频输入，如全景滤波模块 510 把全向摄像机 502 中各种传感器拾取的图像缝合在一起。另外，图像数据被压缩模块 512 压缩，使它更适于在网上（如因特网）广播 514 或保存到计算机可读媒体 516（较佳地通过分离器 520）。或者，也可将图像数据输入个人检测器/跟踪器模块 522，以改进摄像机管理（524），如图像/视频中含发言者的部分经识别，使之与音频信号相关联，而将指向发言者的摄像机/传感器拍摄的

图像将作广播或保持到盘上。

音频输入还可作它用，如将音频输入声源定位模块 526，隔离来自发言者的音频。另在计算机 508 中可用束形成模块 528 改善音频的束形状。减噪与自动增益控制模块 530 通过降低噪声和调节增益，以相对于室内背景噪声更好地捕获来自发言者的音频信号，也可提高信噪比。

如上所述，可将音视频信号播送到另一电视会议地点或因特网，还可保存到计算机可读媒体供以后观看。

上述的本发明用于示例和说明，但并非无遗漏或把本发明限于揭示的精密形式，如上述集成式摄像机与话筒阵列诸例可用于监视系统。可按上述说明作出多种修改变化。本发明范围不受该详述的限制，而由所附如权利要求限定。

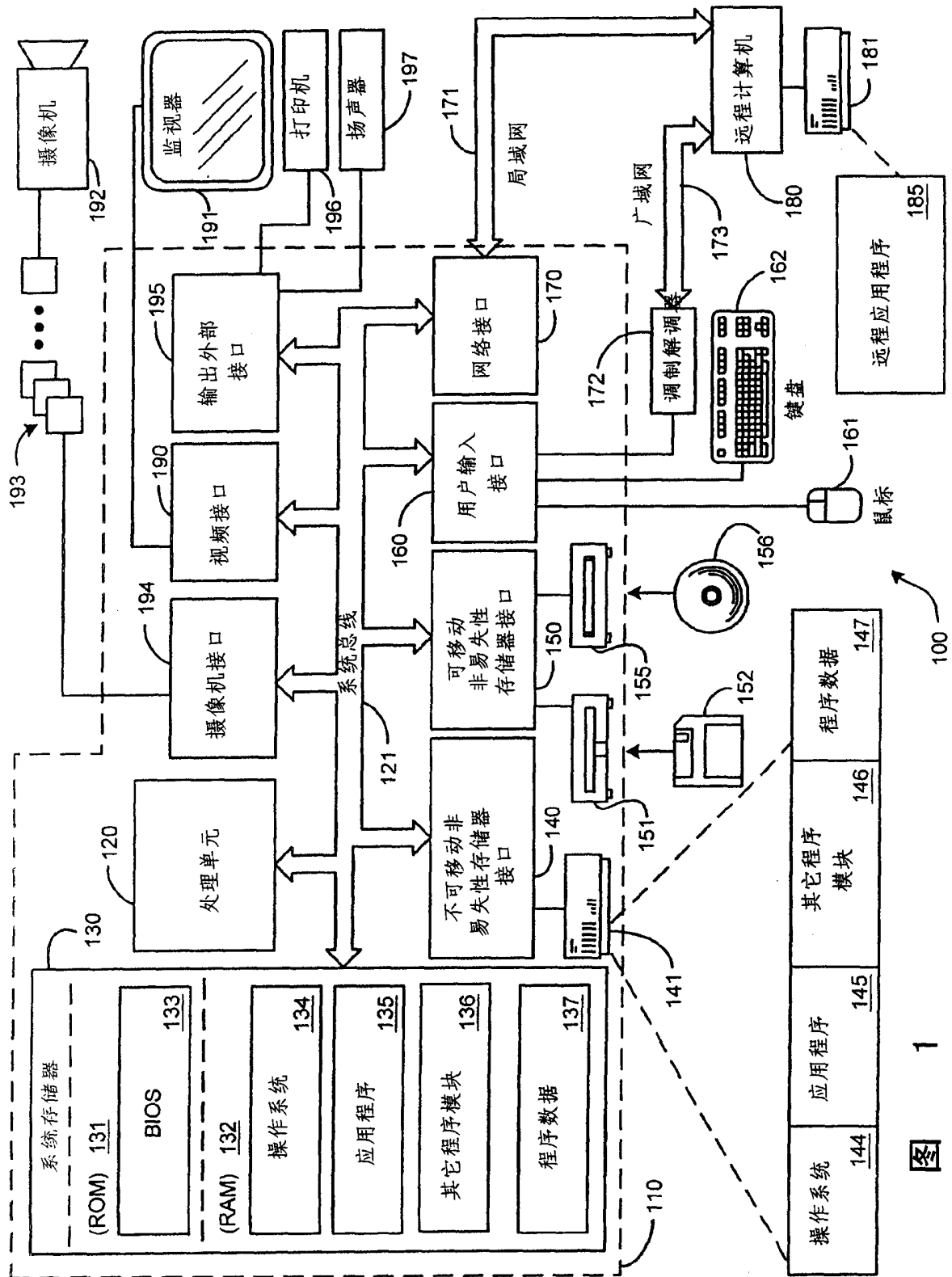


图 1

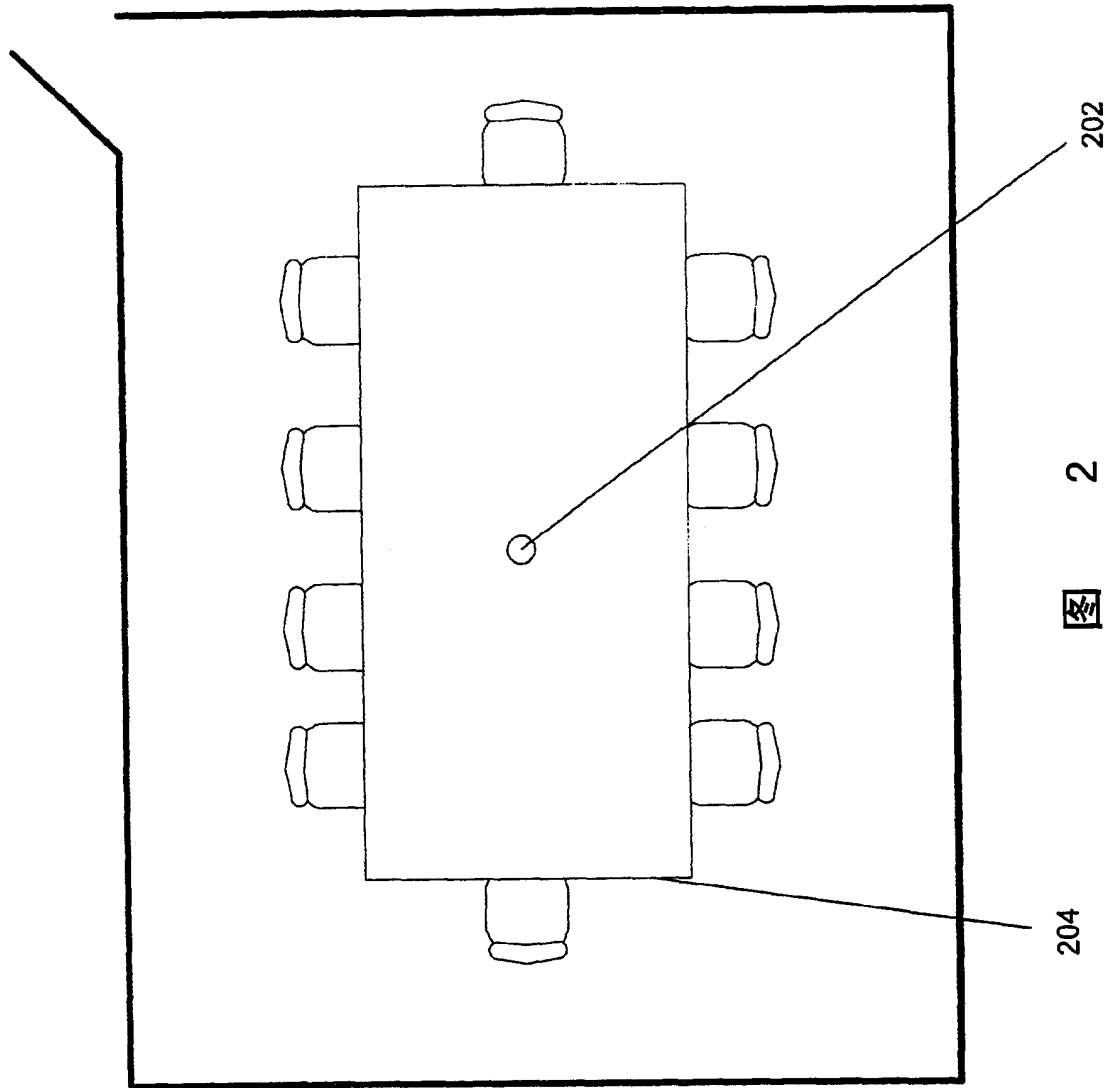


图 2

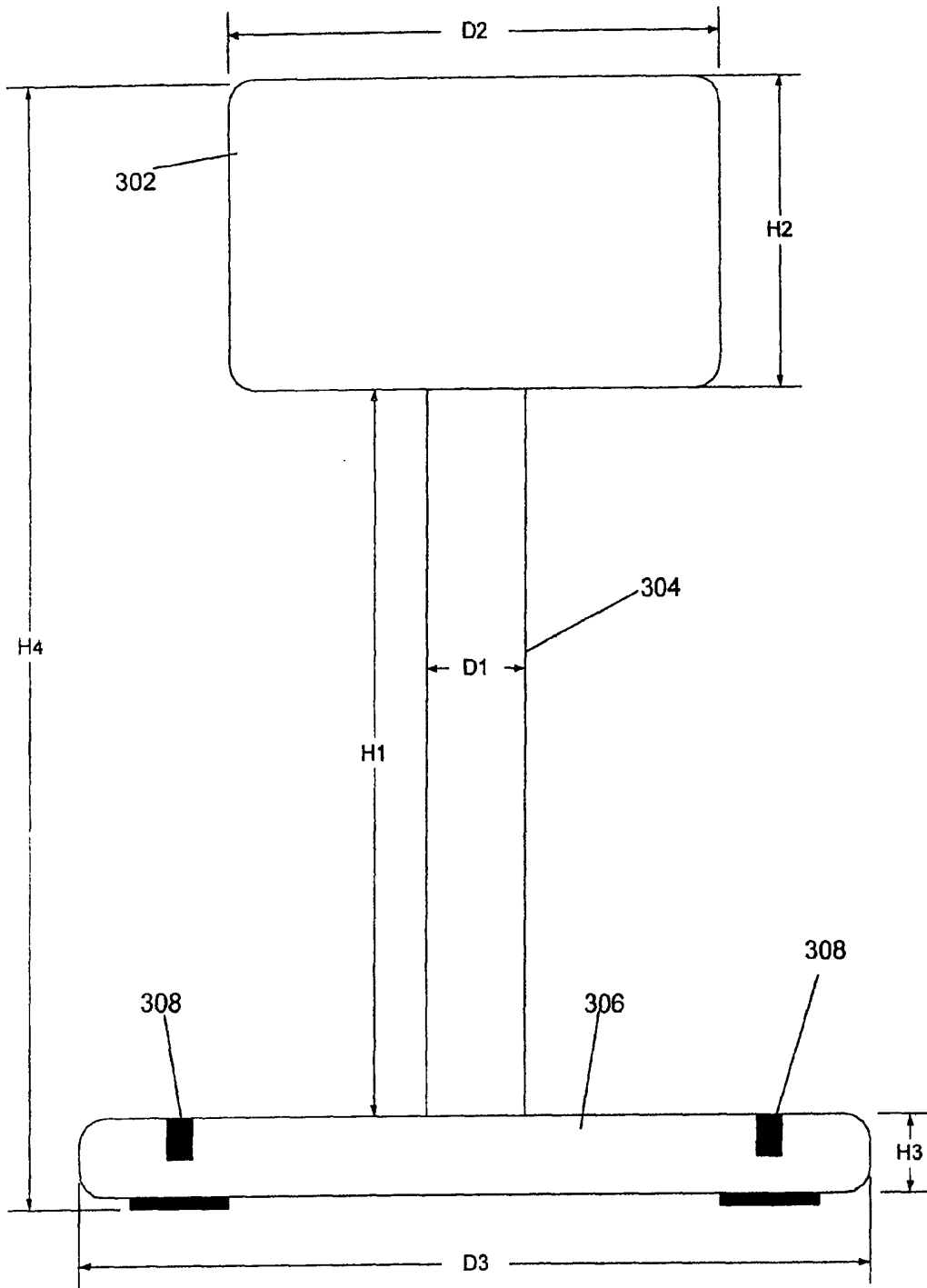


图 3A

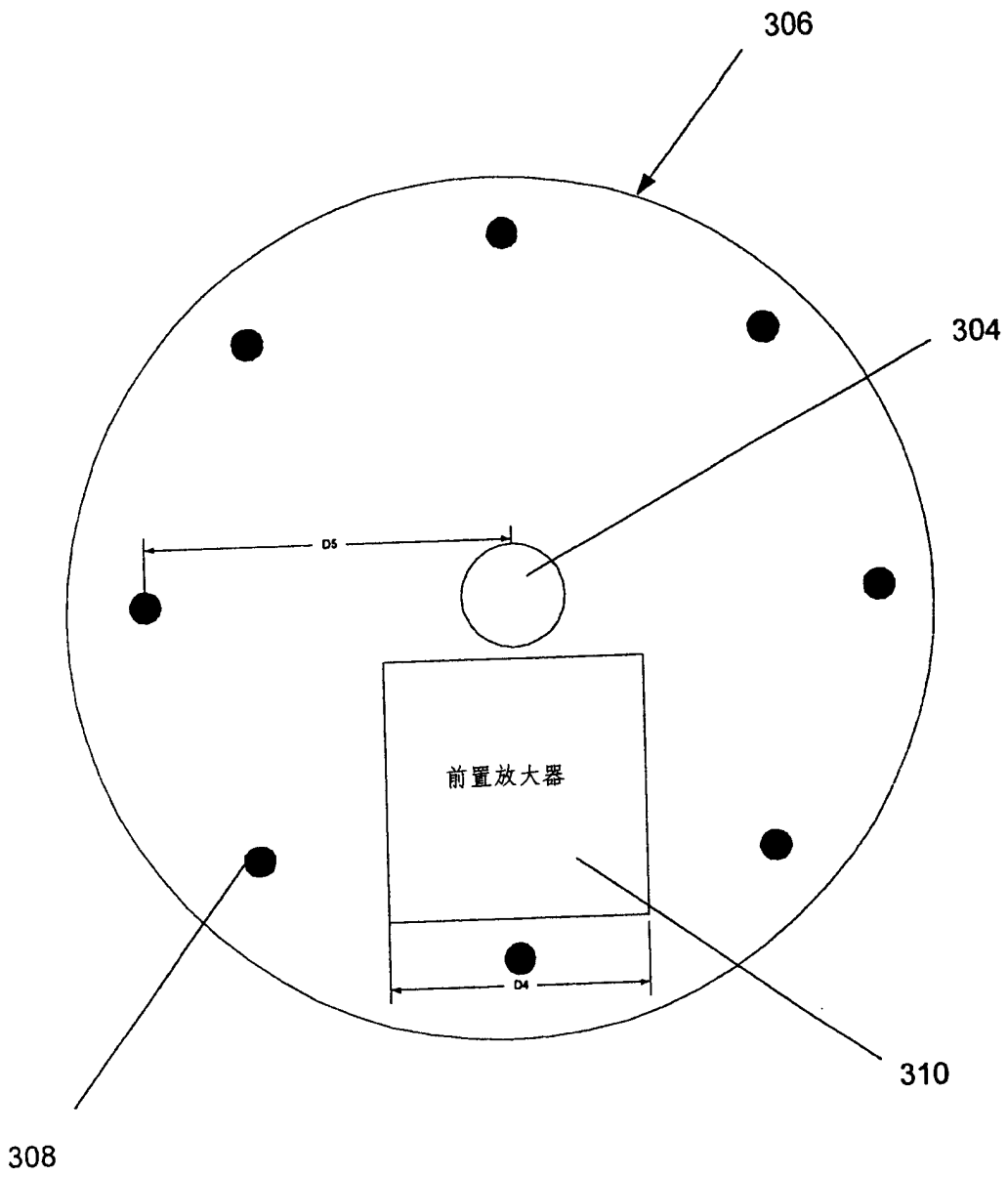


图 3B

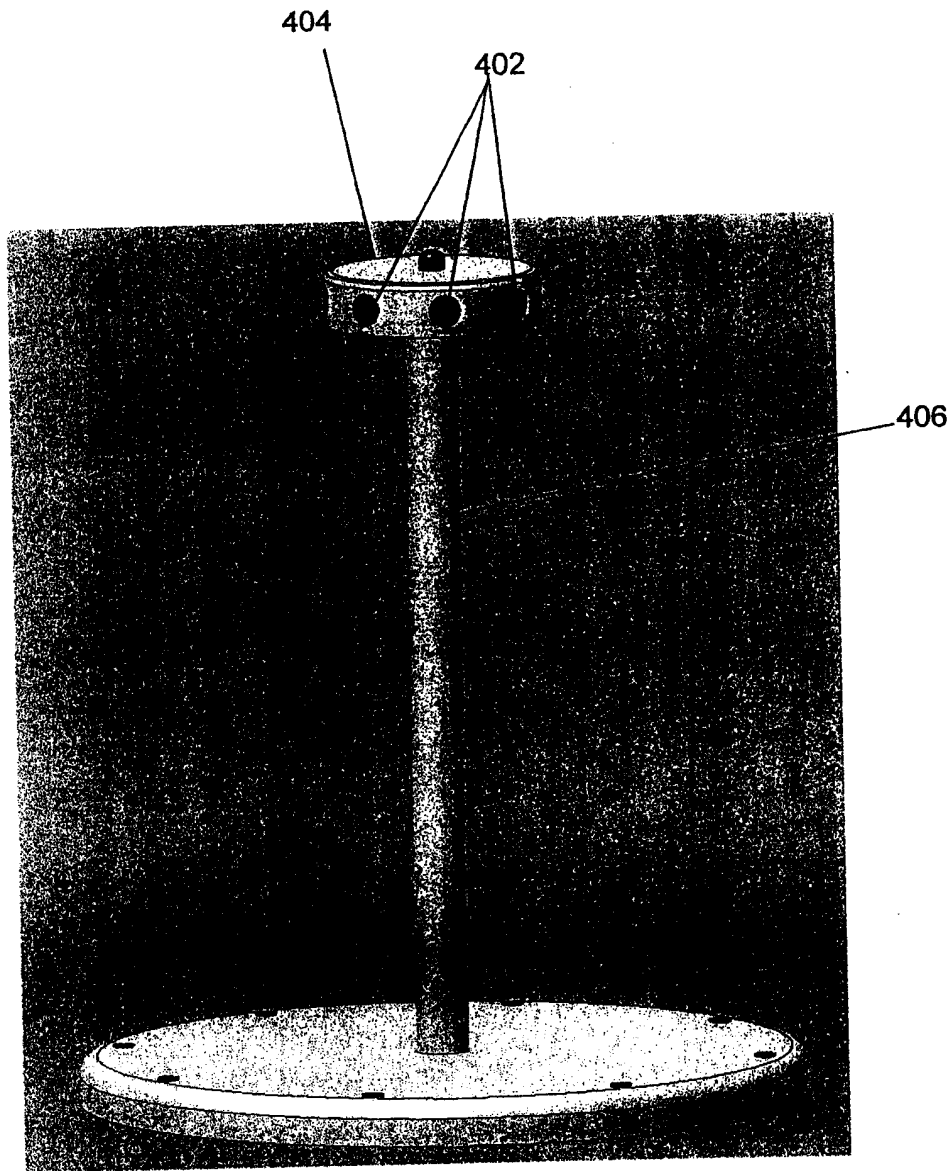


图 4

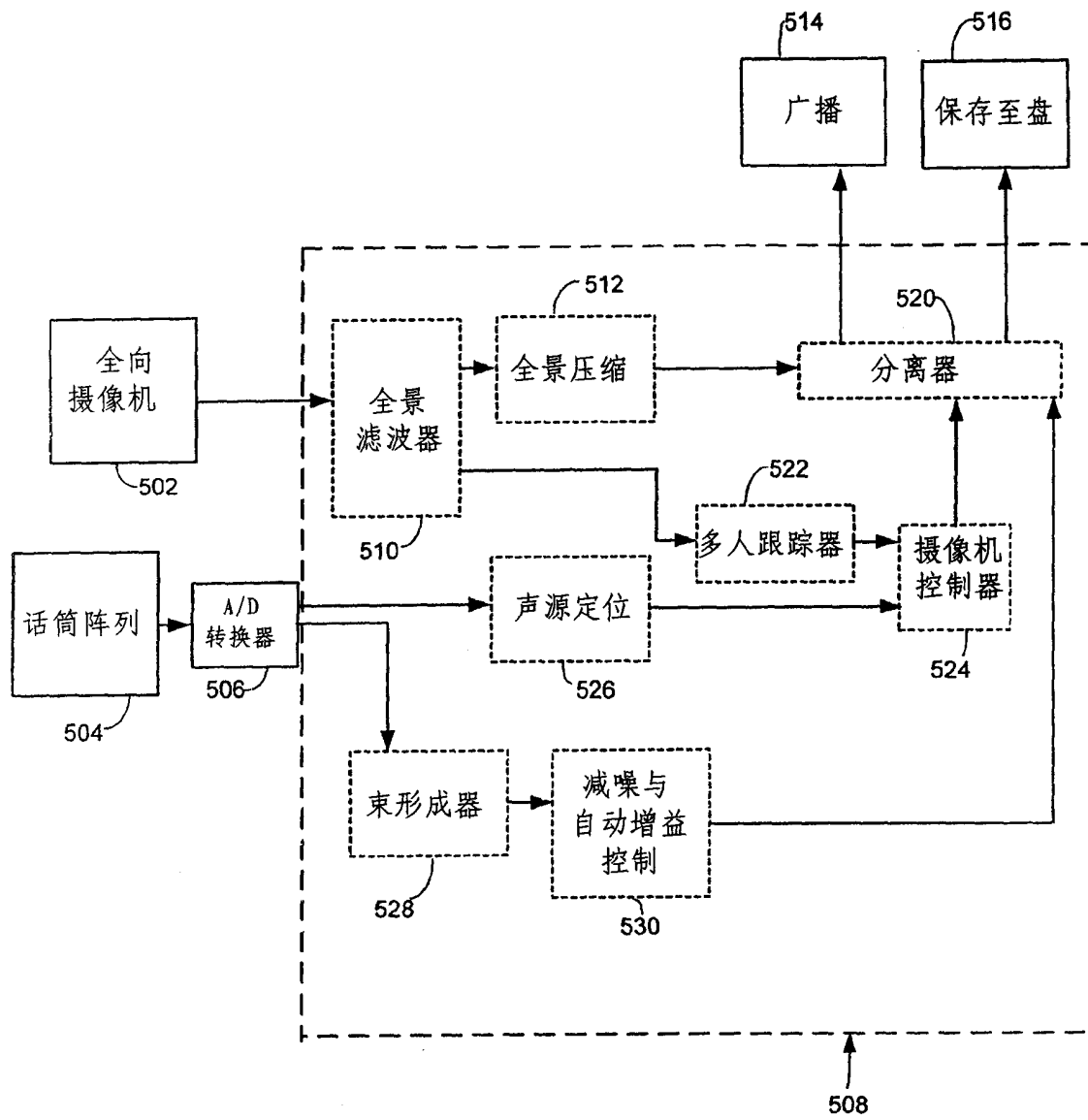


图 5