



(12) 发明专利

(10) 授权公告号 CN 111527446 B

(45) 授权公告日 2022.05.17

(21) 申请号 201880084139.5

(22) 申请日 2018.11.19

(65) 同一申请的已公布的文献号  
申请公布号 CN 111527446 A

(43) 申请公布日 2020.08.11

(30) 优先权数据  
2017-250106 2017.12.26 JP  
2018-207633 2018.11.02 JP

(85) PCT国际申请进入国家阶段日  
2020.06.24

(86) PCT国际申请的申请数据  
PCT/JP2018/042696 2018.11.19

(87) PCT国际申请的公布数据  
W02019/130909 JA 2019.07.04

(73) 专利权人 佳能株式会社  
地址 日本东京都大田区下丸子3丁目30番2号

(72) 发明人 辻本悠贵 饭田吉信

(74) 专利代理机构 北京魏启学律师事务所  
11398

专利代理师 魏启学

(51) Int.Cl.  
G03B 17/00 (2021.01)  
G03B 15/00 (2021.01)  
G03B 17/38 (2021.01)  
G03B 17/56 (2021.01)  
H04N 5/232 (2006.01)

(56) 对比文件  
CN 104978956 A, 2015.10.14  
CN 106292991 A, 2017.01.04  
JP 2011188055 A, 2011.09.22  
JP 2013121078 A, 2013.06.17  
JP 09205574 A, 1997.08.05  
CN 102496195 A, 2012.06.13  
JP 2016156877 A, 2016.09.01  
US 2005114132 A1, 2005.05.26  
US 2016182800 A1, 2016.06.23  
JP 2009177480 A, 2009.08.06

审查员 张桑玲

权利要求书3页 说明书16页 附图16页

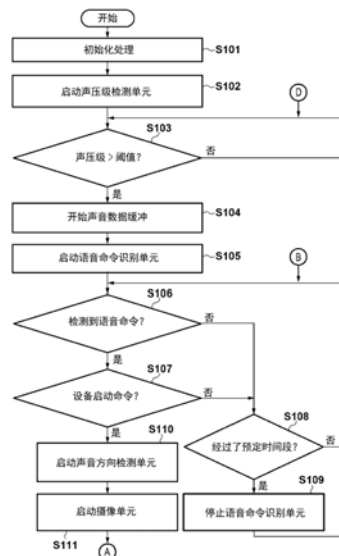
(54) 发明名称

摄像设备及其控制方法和记录介质

(57) 摘要

根据本发明,可以在无需用户进行特殊操作的情况下以用户期望的构图在用户期望的定时拍摄图像。为了实现此,本发明包括:摄像单元;声音收集单元,用于收集语音;检测单元,用于检测所述声音收集单元所收集到的语音的声压是否大于预定声压;以及识别单元,用于识别出所述声音收集单元所收集到的语音是用于指示利用所述摄像单元的拍摄的语音。在所述检测单元将所述声音收集单元所收集到的语音的声压检测为大于所述预定声压、并且所述识别单元将该语音识别为用于指示利用所述摄像单元的拍摄的语音的情况下,进行利用所述摄像单元的拍摄。

CN 111527446 B



1. 一种摄像设备,包括:

摄像单元;

声音收集单元;

检测单元,用于检测所述声音收集单元所收集到的语音的声压级;

驱动单元,用于使所述摄像单元的摄像方向移动;

电力供给单元;

识别单元,用于识别出所述声音收集单元所收集到的语音是用于利用所述摄像单元的拍摄的指示;以及

控制单元,用于响应于所述检测单元检测到声压级大于预定声压级的声压、并且所述识别单元将语音识别为用于利用所述摄像单元的拍摄的指示,所述控制单元控制所述摄像单元进行拍摄,

其中,响应于所述检测单元检测到声压级大于所述预定声压级的声压,所述控制单元启动所述识别单元;

其中,响应于所述识别单元识别出预定语音命令,所述控制单元启动所述摄像单元;

其中,所述声音收集单元包括多个麦克风,所述多个麦克风布置在与所述驱动单元所进行的平摇操作的转动轴垂直的平面上;

其中,所述多个麦克风包括第一麦克风和第二麦克风;

其中,在所述摄像设备由与所述摄像设备的底面垂直且包括所述摄像单元的光轴的平面划分成两个区域的情况下,所述第一麦克风布置在一个区域中,并且所述第二麦克风布置在另一区域中;

其中,所述多个麦克风还包括第三麦克风,所述第三麦克风布置在并非位于用于连接所述第一麦克风的点和所述第二麦克风的点的直线上的部分中;

其中,所述控制单元基于使用所述第一麦克风和所述第二麦克风所检测到的语音的声源的方向、以及使用所述第三麦克风和除所述第三麦克风以外的麦克风所检测到的声源的方向,来检测该语音的声源的方向;以及

其中,响应于所述控制单元检测到在所述电力供给单元不是向所述多个麦克风中的所有麦克风供给电力的状态下、语音的声压级超过了预定声压级,所述电力供给单元向所述声音收集单元的所有麦克风供给电力。

2. 根据权利要求1所述的摄像设备,其中,所述第一麦克风和所述第二麦克风是相对于所述摄像单元的光轴对称布置的。

3. 根据权利要求1所述的摄像设备,其中,在所述摄像单元拍摄运动图像的情况下,所述第一麦克风收集所述运动图像的L通道声音,并且所述第二麦克风收集所述运动图像的R通道声音。

4. 根据权利要求1所述的摄像设备,其中,在使用所述第一麦克风和所述第二麦克风所检测到的语音的声源的方向与用于连接所述第一麦克风的点和所述第二麦克风的点的直线垂直、并且使用所述第三麦克风和除所述第三麦克风以外的麦克风所检测到的语音的声源的方向与用于连接所述第三麦克风的点和除所述第三麦克风以外的麦克风的点的直线垂直的情况下,所述控制单元检测到声源存在于所述转动轴的轴方向上。

5. 根据权利要求1所述的摄像设备,

其中,所述控制单元控制所述驱动单元进行平摇操作,以使所述摄像单元的摄像方向移动以指向语音的声源的方向,

其中,在所述摄像方向指向语音的声源的方向的情况下,所述控制单元控制所述驱动单元进行俯仰操作,以使所述摄像单元的摄像方向移动以指向语音的声源的方向,以及

其中,在所述摄像单元检测到被摄体的情况下,所述控制单元控制所述驱动单元以停止所述摄像单元的俯仰操作。

6. 根据权利要求1所述的摄像设备,

其中,在所述驱动单元所进行的平摇操作所经过的平摇距离被划分成两个区间的情况下,所述控制单元控制所述驱动单元以在前半区间中以第一速度进行平摇操作,并且在后半区间中以从所述第一速度减速的速度进行平摇操作,以及

其中,所述摄像单元不在所述前半区间中进行调焦操作,而在所述后半区间中进行调焦操作。

7. 根据权利要求1所述的摄像设备,

其中,所述控制单元响应于用以拍摄的语音指示,控制所述驱动单元以使所述摄像单元的摄像方向移动以追踪被摄体,以及

其中,所述摄像单元进行静止图像或运动图像的拍摄和记录。

8. 一种用于控制摄像设备的控制方法,所述摄像设备包括:摄像单元;声音收集单元;检测单元,用于检测所述声音收集单元所收集到的语音的声压级;驱动单元,用于使所述摄像单元的摄像方向移动;电力供给单元;以及识别单元,用于识别出所述声音收集单元所收集到的语音是用于利用所述摄像单元的拍摄的指示,所述控制方法包括:

响应于所述检测单元检测到声压级大于预定声压级的声压、并且所述识别单元将语音识别为用于利用所述摄像单元的拍摄的指示,控制所述摄像单元进行拍摄,

其中,响应于所述检测单元检测到声压级大于所述预定声压级的声压,启动所述识别单元;

其中,响应于所述识别单元识别出预定语音命令,启动所述摄像单元;

其中,所述声音收集单元包括多个麦克风,所述多个麦克风布置在与所述驱动单元所进行的平摇操作的转动轴垂直的平面上;

其中,所述多个麦克风包括第一麦克风和第二麦克风;

其中,在所述摄像设备由与所述摄像设备的底面垂直且包括所述摄像单元的光轴的平面划分成两个区域的情况下,所述第一麦克风布置在一个区域中,并且所述第二麦克风布置在另一区域中;

其中,所述多个麦克风还包括第三麦克风,所述第三麦克风布置在并非位于用于连接所述第一麦克风的点和所述第二麦克风的点的直线上的部分中;

其中,基于使用所述第一麦克风和所述第二麦克风所检测到的语音的声源的方向、以及使用所述第三麦克风和除所述第三麦克风以外的麦克风所检测到的声源的方向,来检测该语音的声源的方向;以及

其中,响应于检测到在所述电力供给单元不是向所述多个麦克风中的所有麦克风供给电力的状态下、语音的声压级超过了预定声压级,所述电力供给单元向所述声音收集单元的所有麦克风供给电力。

9. 一种记录有程序的非暂时性记录介质,所述程序用于使得摄像设备执行控制方法,所述摄像设备包括:摄像单元;声音收集单元;检测单元,用于检测所述声音收集单元所收集到的语音的声压级;驱动单元,用于使所述摄像单元的摄像方向移动;电力供给单元;以及识别单元,用于识别出所述声音收集单元所收集到的语音是用于利用所述摄像单元的拍摄的指示,所述控制方法包括:

响应于所述检测单元检测到声压级大于预定声压级的声压、并且所述识别单元将语音识别为用于利用所述摄像单元的拍摄的指示,控制所述摄像单元进行拍摄,

其中,响应于所述检测单元检测到声压级大于所述预定声压级的声压,启动所述识别单元;

其中,响应于所述识别单元识别出预定语音命令,启动所述摄像单元;

其中,所述声音收集单元包括多个麦克风,所述多个麦克风布置在与所述驱动单元所进行的平摇操作的转动轴垂直的平面上;

其中,所述多个麦克风包括第一麦克风和第二麦克风;

其中,在所述摄像设备由与所述摄像设备的底面垂直且包括所述摄像单元的光轴的平面划分成两个区域的情况下,所述第一麦克风布置在一个区域中,并且所述第二麦克风布置在另一区域中;

其中,所述多个麦克风还包括第三麦克风,所述第三麦克风布置在并非位于用于连接所述第一麦克风的点和所述第二麦克风的点的直线上的部分中;

其中,基于使用所述第一麦克风和所述第二麦克风所检测到的语音的声源的方向、以及使用所述第三麦克风和除所述第三麦克风以外的麦克风所检测到的声源的方向,来检测该语音的声源的方向;以及

其中,响应于检测到在所述电力供给单元不是向所述多个麦克风中的所有麦克风供给电力的状态下、语音的声压级超过了预定声压级,所述电力供给单元向所述声音收集单元的所有麦克风供给电力。

## 摄像设备及其控制方法和记录介质

### 技术领域

[0001] 本发明涉及摄像设备及其控制方法和记录介质。

### 背景技术

[0002] 在使用诸如照相机等的摄像设备拍摄静止图像或运动图像时,用户通常在经由取景器等确定拍摄目标、并且亲自确认拍摄状况并调整要拍摄的图像的取景之后,拍摄图像。这样的摄像设备配备有如下的功能:在检测到错误时,向用户通知用户所进行的操作错误,或者检测外部环境并向用户通知处于不适合拍摄的环境中。此外,已知有控制照相机以进入适合于拍摄的状态的机制。

[0003] 与根据用户操作来执行拍摄的这样的摄像设备相比,存在在无需用户给出拍摄指示的情况下间断地和连续地进行拍摄的生活记录照相机(专利文献1)。

[0004] 现有技术文献

[0005] 专利文献

[0006] 专利文献1:日本特表2016-536868

### 发明内容

[0007] 发明要解决的问题

[0008] 然而,由于已知的附着至用户的身体的类型的生活记录照相机定期地进行自动拍摄,因此存在所获得的图像与用户的意图无关的情况。

[0009] 本发明是有鉴于上述问题而做出的,并且目的在于提供如下的技术,该技术使得能够在无需用户进行特殊操作的情况下,以用户期望的构图在用户期望的定时拍摄图像。

[0010] 用于解决问题的方案

[0011] 为了解决该问题,例如,本发明的摄像设备包括以下的结构。也就是说,一种摄像设备,其特征在于,包括:摄像部件;声音收集部件,用于收集语音;检测部件,用于检测所述声音收集部件所收集到的语音的声压是否大于预定声压;以及识别部件,用于识别出所述声音收集部件所收集到的语音是用于指示利用所述摄像部件的拍摄的语音,其中,在所述检测部件将所述声音收集部件所收集到的语音的声压检测为大于所述预定声压、并且所述识别部件将该语音识别为用于指示利用所述摄像部件的拍摄的语音的情况下,进行利用所述摄像部件的拍摄。

[0012] 发明的效果

[0013] 根据本发明,可以在无需用户进行特殊操作的情况下,以用户期望的构图在用户期望的定时拍摄图像。

[0014] 本发明的其它特征和优点将在以下参考附图的说明中显而易见。注意,在附图中,相同或相似的构成元件被指派相同的附图标记。

## 附图说明

[0015] 附图包括在说明书中并构成说明书,例示本发明的实施例,并用于连同说明书的描述一起说明本发明的原理。

[0016] 图1是根据实施例的摄像设备的框图。

[0017] 图2是根据实施例的声音输入单元和声音信号处理单元的详细框图。

[0018] 图3A是根据实施例的摄像设备的正视图和俯视图。

[0019] 图3B是示出实施例中的摄像设备的使用示例的图。

[0020] 图3C是示出实施例中的摄像设备的使用示例的图。

[0021] 图3D是示出实施例中的摄像设备的使用示例的图。

[0022] 图3E是示出实施例中的摄像设备的使用示例的图。

[0023] 图4是示出根据实施例的摄像设备的平摇和俯仰操作的图。

[0024] 图5A是示出实施例中的中央控制单元的处理过程的流程图。

[0025] 图5B是示出实施例中的中央控制单元的处理过程的流程图。

[0026] 图6是示出图5B中的语音命令处理的详情的流程图。

[0027] 图7是示出实施例中的语音命令的含义与语音命令之间的关系的图。

[0028] 图8是实施例中的从启动起直到操作拍摄开始命令为止的时序图。

[0029] 图9A是示出根据实施例的声音方向检测方法的图。

[0030] 图9B是示出根据实施例的声音方向检测方法的图。

[0031] 图9C是示出根据实施例的声音方向检测方法的图。

[0032] 图10A是示出在摄像设备的正上方存在声源的情况下的检测方法的图。

[0033] 图10B是示出在摄像设备的正上方存在声源的情况下的检测方法的图。

[0034] 图11A是示出第二实施例中的中央控制单元的处理过程的流程图。

[0035] 图11B是示出第二实施例中的中央控制单元的处理过程的流程图。

[0036] 图12是示出第二实施例中的摄像设备的操作的图。

## 具体实施方式

[0037] 以下将参考附图来详细说明实施例。

[0038] [第一实施例]

[0039] 图1是根据第一实施例的摄像设备1的框结构图。摄像设备1由可动摄像单元100和支撑单元200构成,该可动摄像单元100包括光学透镜单元,其中进行摄像的方向(光轴方向)是可变的,该支撑单元200包括中央控制单元(CPU),该CPU进行可动摄像单元100的驱动控制,并且控制摄像设备整体。

[0040] 注意,支撑单元200设置有与可动摄像单元100的面接触的、包括压电元件的多个驱动单元11~13。可动摄像单元100通过控制这些驱动单元11~13的振动来进行平摇和俯仰操作。注意,结构可以如下:使用伺服马达等来实现平摇和俯仰操作。

[0041] 可动摄像单元100包括镜头单元101、摄像单元102、镜头致动器控制单元103和声音输入单元104。

[0042] 镜头单元101由包括变焦透镜、光圈/快门和调焦透镜等的拍摄光学系统构成。摄像单元102包括诸如CMOS传感器或CCD传感器等的图像传感器,将由镜头单元101形成的光

学图像光电转换成电信号,并输出该电信号。镜头致动器控制单元103包括马达驱动器IC,并且驱动镜头单元101的变焦透镜、光圈/快门和调焦透镜等所用的各种致动器。基于从后面将说明的支撑单元200中的中央控制单元201接收到的致动器驱动指示数据来驱动各种致动器。声音输入单元104是包括麦克风(以下称为麦克风(mic))的声音输入单元,并由多个麦克风(在本实施例中为四个麦克风)构成,并且将声音转换成电信号,将该电信号转换成数字信号(声音数据),并输出该数字信号。

[0043] 另一方面,支撑单元200包括用于控制摄像设备1整体的中央控制单元201。中央控制单元201由CPU、存储有CPU所要执行的程序的ROM和用作CPU的工作区的RAM构成。此外,支撑单元200包括摄像信号处理单元202、视频信号处理单元203、声音信号处理单元204、操作单元205、存储单元206和显示单元207。支撑单元200还包括外部输入/输出端子单元208、声音再现单元209、电源单元210、电源控制单元211、位置检测单元212、枢转控制单元213、无线通信单元214、以及上述驱动单元11~13。

[0044] 摄像信号处理单元202将从可动摄像单元100的摄像单元102输出的电信号转换成视频信号。视频信号处理单元203根据用途来对从摄像信号处理单元202输出的视频信号进行处理。视频信号的处理包括图像的剪切、通过旋转处理所实现的电子图像稳定操作、以及用于检测被摄体(面部)的被摄体检测处理。

[0045] 声音信号处理单元204对从声音输入单元104输出的数字信号进行声音处理。在声音输入单元104是用于输出模拟信号的麦克风时,声音信号处理单元204可以包括用于将模拟信号转换成数字信号的构成元件。注意,后面将使用图2来说明包括声音输入单元104的声音信号处理单元204的详情。

[0046] 操作单元205用作摄像设备1和用户之间的用户界面,并且由各种开关和按钮等构成。存储单元206存储诸如通过拍摄所获得的视频信息等的各种数据。显示单元207包括诸如LCD等的显示器,并且基于从视频信号处理单元203输出的信号来根据需要显示图像。此外,显示单元207通过显示各种菜单等作用用户界面的一部分。外部输入/输出端子单元208相对于外部设备接收/输出通信信号和视频信号。声音再现单元209包括扬声器,将声音数据转换成电信号,并再现声音。电源单元210是驱动摄像设备的整体(构成元件)所需的电源,并且在本实施例中假定为可再充电电池。

[0047] 电源控制单元211根据摄像设备1的状态来控制从电源单元210向上述各个构成元件的电力的供给/切断。根据摄像设备1的状态,存在不使用的构成元件。电源控制单元211在中央控制单元201的控制下,通过根据摄像设备1的状态切断向不使用的构成元件的电力来执行抑制电力消耗的功能。注意,电力供给/切断将通过后面给出的说明变得清楚。

[0048] 位置检测单元212由陀螺仪、加速度传感器和GPS等构成,并且检测摄像设备1的移动。位置检测单元212也用于应对用户将摄像设备1附着至他/她的身体的情况。枢转控制单元213根据来自中央控制单元201的光轴方向的指示来生成用于驱动驱动单元11~13的信号,并输出这些信号。驱动单元11~13的压电元件根据从枢转控制单元213施加的驱动信号而振动,并且移动可动摄像单元100的光轴方向。结果,可动摄像单元100在中央控制单元201所指示的方向上进行平摇和俯仰操作。

[0049] 无线通信单元214按照诸如WiFi(注册商标)或BLE(蓝牙(Bluetooth(注册商标)))低功耗)等的无线标准进行图像数据等的的数据发送。

[0050] 接着,将参考图2来说明本实施例中的声音输入单元104和声音信号处理单元204的结构、以及声音方向检测处理。图2示出声音输入单元104和声音信号处理单元204的结构、以及声音信号处理单元204、中央控制单元201和电源控制单元211之间的连接关系。

[0051] 声音输入单元104由四个非定向麦克风(麦克风104a、104b和104c以及麦克风104d)构成。各麦克风包括A/D转换器,按预先设置的采样率(命令检测和方向检测处理:16kHz,运动图像记录:48kHz)收集声音,使用内部A/D转换器将所收集到的声音的声音信号转换成数字声音数据,并输出该数字声音数据。注意,在本实施例中,声音输入单元104由四个数字麦克风构成,但也可以由具有模拟输出的麦克风构成。在模拟麦克风的情况下,在声音信号处理单元204中仅需设置相应的A/D转换器。此外,本实施例中的麦克风的数量是四个,但数量仅需为三个或更多个。

[0052] 麦克风104a在摄像设备1的电源接通的情况下被无条件地供给电力,并且进入可收集声音状态。另一方面,其它麦克风104b、104c和104d在中央控制单元201的控制下是电源控制单元211的电力供给/切断的目标,并且在摄像设备1的电源接通之后的初始状态下,切断向这些其它麦克风的电力。

[0053] 声音信号处理单元204由声压级检测单元2041、语音存储器2042、语音命令识别单元2043、声音方向检测单元2044、运动图像声音处理单元2045和命令存储器2046构成。

[0054] 在由从麦克风104a输出的声音数据表示的声压级超过预先设置的阈值时,声压级检测单元2041将表示检测到了声音的信号供给至电源控制单元211和语音存储器2042。

[0055] 电源控制单元211在从声压级检测单元2041接收到表示检测到了声音的信号时,向语音命令识别单元2043供给电力。

[0056] 语音存储器2042是在中央控制单元201的控制下利用电源控制单元211的电力供给/切断的目标其中之一。此外,语音存储器2042是暂时存储从麦克风104a输出的声音数据的缓冲存储器。在麦克风104a的采样率为16kHz、针对各采样输出两个字节(16位)的声音数据、并且假定最长的语音命令为5秒的情况下,语音存储器2042需要具有约160千字节( $\cong 5 \times 16 \times 1000 \times 2$ )的容量。此外,在语音存储器2042的容量充满来自麦克风104a的声音数据时,旧的声音数据被新的声音数据覆盖。结果,语音存储器2042保持最近的预定时间段(在以上示例中为5秒)的声音数据。此外,语音存储器2042在被从声压级检测单元2041接收到表示检测到了声音的信号所触发的情况下,开始将来自麦克风104a的声音数据存储于采样数据区域中。

[0057] 命令存储器2046由非易失性存储器构成,并且在命令存储器2046中预先存储(登记)了与摄像设备所识别出的语音命令有关的信息。尽管后面将说明详情,但例如,要存储在命令存储器2046中的语音命令的类型是如图7所示的。与包括“启动命令”的多个类型的命令有关的信息被存储在命令存储器2046中。

[0058] 语音命令识别单元2043是在中央控制单元201的控制下利用电源控制单元211的电力供给/切断的目标其中之一。注意,语音识别本身是已知的技术,因此这里省略了对语音识别的说明。语音命令识别单元2043通过参考命令存储器2046来进行用于识别语音存储器2042中所存储的声音数据的处理。此外,语音命令识别单元2043判断通过由麦克风104a进行的聲音收集所获得的声音数据是否是语音命令,并且还判断命令存储器2046中所存储的语音命令中的哪个语音命令与声音数据匹配。此外,语音命令识别单元2043在检测到与



命令存储器2046中所存储的语音命令其中之一匹配的声音数据时,将表示确定了这些语音命令中的哪个语音命令的信息、以及语音存储器2042内所存储的声音数据中的用于确定该语音命令的声音数据的开始和结束地址(或接受语音命令的定时)供给至中央控制单元201。

[0059] 声音方向检测单元2044是在中央控制单元201的控制下利用电源控制单元211的电力供给/切断的目标其中之一。此外,声音方向检测单元2044基于来自四个麦克风104a~104d的声音数据来周期性地用于检测存在声源的方向的处理。声音方向检测单元2044包括内部缓冲存储器2044a,并将表示所检测到的声源方向的信息存储在缓冲存储器2044a中。注意,声音方向检测单元2044进行声音方向检测处理的周期(例如,16kHz)可以充分长于麦克风104a的采样周期。注意,假定缓冲存储器2044a具有足以存储与可以存储在语音存储器2042中的声音数据的持续时间相同的持续时间的声音方向信息的容量。

[0060] 运动图像声音处理单元2045是在中央控制单元201的控制下利用电源控制单元211的电力供给/切断的目标其中之一。运动图像声音处理单元2045接收来自四个麦克风中的麦克风103a和104b的两个声音数据作为立体声音数据,并对该立体声音数据进行诸如各种滤波处理、风声消除、立体声感增强、驱动声音去除、ALC(自动电平控制)和压缩处理等的针对运动图像声音的声音处理。尽管详情将通过后面给出的说明变得清楚,但在本实施例中,麦克风104a用作立体声麦克风中的L通道麦克风,并且麦克风104b用作R通道麦克风。

[0061] 注意,在图2中,考虑到电力消耗和电路结构,示出声音输入单元104的各麦克风与声音信号处理单元204中所包括的各块之间的针对四个麦克风的最小连接数。然而,结构也可以如下:在电力和电路结构允许的范围内,多个麦克风由声音信号处理单元204中所包括的各块共享使用。此外,在本实施例中,麦克风104a作为基准麦克风而连接,但任何麦克风均可以是基准麦克风。

[0062] 将参考图3A~3E来说明摄像设备1的外观图和使用示例。图3A示出根据本实施例的摄像设备1的外观的俯视图和正视图。摄像设备1的可动摄像单元100具有大致半球形状,并且包括第一壳体150,该第一壳体150包括从-20度到表示垂直方向的90度的范围内的切出窗口,并且在由该图所示的箭头A表示的水平面中可枢转360度,其中与底面平行的水平面为0度。此外,可动摄像单元100包括第二壳体151,该第二壳体151可以在从水平方向到如由该图所示的箭头B所示的垂直方向的范围内,连同镜头单元101和摄像单元102一起沿着切出窗口枢转。这里,箭头A所示的第一壳体150的枢转操作对应于平摇操作,并且箭头B所示的第二壳体151的枢转操作对应于俯仰操作,并且这些操作是通过驱动单元11~13来实现的。注意,如上所述,假定本实施例中的摄像设备的可俯仰范围为从-20度到+90度的范围。

[0063] 麦克风104a和104b以夹持第一壳体150的切出窗口的方式布置在前面侧的位置。此外,麦克风104c和104d布置在第一壳体150的后面侧。如图3A所示,即使在第二壳体151固定的状态下在沿着箭头A的任何方向上进行第一壳体150的平摇操作,麦克风104a和104b相对于镜头单元101和摄像单元102的相对位置也不会改变。也就是说,麦克风104a相对于摄像单元102的摄像方向始终位于左侧,并且麦克风104b始终位于右侧。此外,麦克风104a和104b相对于摄像单元102的摄像方向对称布置,因此麦克风104a用于立体声麦克风的L通道

输入,并且麦克风104b用于立体声麦克风的R通道输入。因此,可以在由通过摄像单元102所进行的拍摄而获得的图像表示的空间与麦克风104a和104b所获取到的声场之间保持固定关系。

[0064] 注意,如图3A所示,本实施例中的四个麦克风104a、104b、104c和104d布置在摄像设备1的俯视图中的矩形的各顶点的位置。此外,假定这四个麦克风位于图3A中的一个水平面上,但允许小的位置偏移。

[0065] 麦克风104a与麦克风104b之间的距离大于麦克风104a与麦克风104c之间的距离。注意,相邻麦克风之间的距离期望在从约10mm到30mm的范围内。此外,在本实施例中,麦克风的数量是四个,但麦克风的数量可以是三个或更多个,只要满足麦克风不布置在直线上的条件即可。此外,图3A所示的麦克风104a~104d的布置位置是示例性的,并且布置方法可以根据机械限制和设计限制等适当改变。

[0066] 图3B~3E示出本实施例中的摄像设备1的使用模式。图3B是示出摄像设备1放置在桌子等上、并且拍摄者本人和拍摄者周围的被摄体是拍摄目标的使用模式的图。图3C是示出摄像设备1从拍摄者的颈部悬挂、并且拍摄者前方的被摄体在他/她移动时是拍摄目标的使用模式的图。图3D是示出摄像设备1固定到拍摄者的肩部、并且前后和右侧的周围被摄体是拍摄目标的使用模式的图。此外,图3E是示出如下的使用模式的图:摄像设备1固定至用户所把持的杆的端部,其目的是使摄像设备1移动到用户期望的拍摄位置(高位置、手不能到达的位置)并进行拍摄。

[0067] 将参考图4来进一步详细地说明本实施例的摄像设备1的平摇和俯仰操作。这里,将假定如图3B所示、摄像设备1被放置成站立的示例性使用情况来进行说明,但这同样可以适用于其它使用情况。

[0068] 图4中的4a表示镜头单元101指向水平方向的状态。图4的4a所表示的状态被定义为初始状态,并且在如从上方观看到的、第一壳体150沿逆时针方向上进行90度的平摇操作时,进入图4中的4b所表示的状态。另一方面,在第二壳体151从图4中的4a所表示的初始状态起进行90度的俯仰操作时,进入图4中的4c所表示的状态。如上所述,第一壳体150和第二壳体151的枢转是通过由枢转控制单元213驱动的驱动单元11~13的振动来实现的。

[0069] 接着,将按照图5A和5B所示的流程图来说明摄像设备1的中央控制单元201所进行的处理的过程。图5A和5B所示的处理示出在摄像设备1的主电源接通的情况下中央控制单元201所进行的处理。

[0070] 中央控制单元201在步骤S101中进行摄像设备1的初始化处理。在该初始化处理中,中央控制单元201将可动摄像单元100中的摄像单元102的摄像方向的水平面中的当前方向分量确定为平摇操作的基准角度(0度)。

[0071] 在下文,在进行可动摄像单元100的平摇操作之后的摄像方向的水平面中的分量由从该基准角度起的相对角度表示。此外,声音方向检测单元2044所检测到的声源方向的水平面中的分量也由相对于基准角度的相对角度表示。此外,尽管后面将说明详情,但声音方向检测单元2044还进行与在摄像设备1正上方的方向(平摇操作的转动轴的轴方向)上是否存在声源有关的判断。

[0072] 注意,在该阶段,切断向语音存储器2042、声音方向检测单元2044、运动图像声音处理单元2045、以及麦克风104b~104d的电力。

[0073] 在初始化处理结束时,在步骤S102中,中央控制单元201通过控制电源控制单元211来开始向声压级检测单元2041和麦克风104a供给电力。结果,声压级检测单元2041基于从麦克风104a输出的声音数据来执行用于检测在被转换成该声音数据之前的声音的声压级的处理,并且在判断为声音的声压级超过预先设置的阈值时,向中央控制单元201通知该事实。注意,该阈值例如被设置为60dB SPL(声压级),但该阈值可以由摄像设备1根据环境等改变,或者可以关注于所需频带中的声音分量。

[0074] 中央控制单元201在步骤S103中等待声压级检测单元2041检测到声压级超过阈值的的声音。在检测到声压级超过阈值的的声音时,在步骤S104中,声音存储器2042开始用于接收并存储来自麦克风104a的声音数据的处理。

[0075] 此外,在步骤S105中,中央控制单元201通过控制电源控制单元211来开始向语音命令识别单元2043供给电力。结果,语音命令识别单元2043开始用于参考命令存储器2046来识别语音存储器2042中所存储的声音数据的处理。此外,语音命令识别单元2043进行用于识别语音存储器2042中所存储的声音数据的处理,并且在识别出与命令存储器2046中的语音命令其中之一匹配的语音命令时,向中央控制单元201通知如下的信息,该信息包括用于指定所识别的语音命令的信息、以及与语音存储器2042中的用于确定该所识别的语音命令的声音数据的开始地址和结束地址(或接受语音命令的定时)有关的信息。

[0076] 在步骤S106中,中央控制单元201判断是否从语音命令识别单元2043接收到表示识别出了语音命令的信息。如果判断为“否”,则中央控制单元201使处理进入步骤S108,并且判断从语音命令识别单元2043的启动起所经过的时间是否超过了预先设置的阈值。此外,只要所经过的时间为阈值或更小,则中央控制单元201等待语音命令识别单元2043识别语音命令。然后,如果在经过了由阈值表示的时间时、语音命令识别单元2043未识别出语音命令,则中央控制单元201使处理进入步骤S109。在步骤S109中,中央控制单元201通过控制电源控制单元211来切断向语音命令识别单元2043的电力。然后,中央控制单元201使处理返回到步骤S103。

[0077] 另一方面,中央控制单元201在从语音命令识别单元2043接收到表示识别出了语音命令的信息时,使处理进入步骤S107。在步骤S107中,中央控制单元201判断所识别出的语音命令是否对应于图8所示的启动命令。此外,中央控制单元201在判断为所识别出的语音命令是除启动命令以外的命令时,使处理进入步骤S108。此外,如果所识别的语音命令是启动命令,则中央控制单元201使处理从步骤S107进入步骤S110。

[0078] 在步骤S110中,中央控制单元201通过控制电源控制单元211来开始向声音方向检测单元2044和麦克风104b~104d供给电力。结果,声音方向检测单元2044开始用于基于同一时间点的来自四个麦克风104a~104d的声音数据来检测声源方向的处理。用于检测声源方向的处理是按预定周期进行的。此外,声音方向检测单元2044将表示所检测到的声音方向的声音方向信息存储在内部缓冲存储器2044a中。这里,声音方向检测单元2044将声音方向信息存储在缓冲存储器2044a中,使得确定声音方向信息所使用的声音数据的定时可以与声音存储器2042中所存储的声音数据的定时相关联。通常,声音方向和声音存储器2042中的声音数据的地址可以存储在缓冲存储器2044a中。注意,声音方向信息是表示水平面中的、表示声源方向与上述基准角度的差的角度角度的信息。此外,尽管后面将说明详情,但在声源位于摄像设备1正上方的情况下,将表示声源在正上方的方向上的信息设置为声音方向

信息。

[0079] 在步骤S111中,中央控制单元201通过控制电源控制单元211来开始向摄像单元102和镜头致动器控制单元103供给电力。结果,可动摄像单元100开始用作摄像设备。

[0080] 接着,在步骤S151中,中央控制单元201判断从语音命令识别单元2043是否接收到表示识别出了语音命令的信息。如果判断为“否”,则中央控制单元201使处理进入步骤S152,并且判断与来自用户的指示相对应的作业当前是否在执行中。尽管详情将通过图6的流程图的说明变得清楚,但运动图像拍摄和记录、以及追踪处理等对应于作业。这里,在假定没有正在执行这样的作业的情况下继续该说明。

[0081] 在步骤S153中,判断从识别出前一语音命令起所经过的时间是否超过预先设置的阈值。如果判断为“否”,则中央控制单元201使处理返回到步骤S151,并且等待识别出语音命令。然后,如果没有正在执行作业、并且即使从识别出前一语音命令起所经过的时间超过阈值、也未识别出新的语音命令,则中央控制单元201使处理进入步骤S154。在步骤S154中,中央控制单元201通过控制电源控制单元211来切断向摄像单元102和镜头致动器控制单元103的电力供给。另外,在步骤S155中,中央控制单元201还通过控制电源控制单元211切断向声音方向检测单元2044的电力供给,并且使处理返回到步骤S106。

[0082] 假定中央控制单元201从语音命令识别单元2043接收到表示接收到了语音命令的信息。在这种情况下,语音命令识别单元2043使处理从步骤S151进入步骤S156。

[0083] 本实施例中的中央控制单元201在执行与所识别出的语音命令相对应的作业之前,进行用于使说出了语音命令的人物进入可动摄像单元100的摄像单元102的视野的处理。然后,中央控制单元201在人物在摄像单元102的视野中的状态下,执行基于所识别的语音命令的作业。

[0084] 为了实现上述技术,在步骤S156中,中央控制单元201从声音方向检测单元2044的缓冲存储器2044a获取与语音命令识别单元2043所识别出的语音命令同步的声音方向信息。语音命令识别单元2043在识别出语音命令时,向中央控制单元201通知语音存储器2042中的语音命令的开始和结束这两个地址。然后,中央控制单元201从缓冲存储器2044a获取在由这两个地址表示的时间段中检测到的声音方向信息。可能存在如下情况:在由这两个地址表示的时间段中存在多个声音方向信息。在这种情况下,中央控制单元201从缓冲存储器2044a获取在时间上最近的声音方向信息。这是因为,在时间上最近的声音方向信息表示说出了语音命令的人物的当前位置的概率高。

[0085] 在步骤S157中,中央控制单元201判断由所获取到的声音信息表示的声源方向是否是摄像设备正上方的方向。注意,后面将说明与声音方向是否是摄像设备正上方的方向有关的判断的详情。

[0086] 如果在摄像设备1正上方的方向上存在声源,则中央控制单元201使处理进入步骤S158。在步骤S158中,中央控制单元201通过控制枢转控制单元213,来使可动摄像单元100的第二壳体151枢转,使得如由图4的4c所示,镜头单元101和摄像单元102的摄像方向为正上方向。在摄像单元102的摄像方向被设置为正上方向的情况下,在步骤S159中,中央控制单元201从视频信号处理单元203接收拍摄图像,并且判断在拍摄图像中是否存在可以是声源的对象(人物的面部)。如果判断为“否”,则中央控制单元201使处理返回到步骤S151。另一方面,如果在拍摄图像中存在对象,则中央控制单元201使处理进入步骤S164,并且执行

与已识别出的语音命令相对应的作业。注意,后面将使用图6来说明步骤S164中的处理的详情。

[0087] 在步骤S157中,中央控制单元201在判断为由声音信息表示的方向是除正上方向以外的方向时,使处理进入步骤S160。在步骤S160中,中央控制单元201通过控制枢转控制单元213来进行可动摄像单元100的平摇操作,使得摄像单元102的水平面中的当前角度与由声音信息表示的水平面中的角度一致。然后,在步骤S161中,中央控制单元201从视频信号处理单元203接收拍摄图像,并且判断在拍摄图像中是否存在可以是声源的对象(面部)。如果判断为“否”,则中央控制单元201使处理进入步骤S162,并且通过控制枢转控制单元213来进行可动摄像单元100向着目标对象的俯仰操作。然后,在步骤S163中,中央控制单元201判断摄像单元102的摄像方向的俯仰方向的角度是否达到俯仰操作的上限(在本实施例中为从水平方向起的90度)。如果判断为“否”,则中央控制单元201使处理返回到步骤S161。以这种方式,中央控制单元201在进行俯仰操作的同时,判断在来自视频信号处理单元203的拍摄图像中是否存在可以是声源的对象(面部)。然后,如果即使摄像单元102的摄像方向的俯仰方向的角度达到了俯仰操作的上限、也没有检测到对象,则中央控制单元201使处理从步骤S163返回到步骤S151。另一方面,如果在拍摄图像中存在对象,则中央控制单元201使处理进入步骤S164,并且执行与已识别的语音命令相对应的作业。

[0088] 接着,将基于图6的流程图和图7所示的语音命令表来说明步骤S164中的处理的详情。与图7的语音命令表中所示的诸如“你好,照相机”等的语音命令相对应的语音模式数据被存储在命令存储器2046中。注意,在图7中示出代表性的语音命令。注意,语音命令不限于此。此外,应当注意,以下说明中的语音命令是在图5B的步骤S151的定时检测到的语音命令。

[0089] 首先,在步骤S201中,中央控制单元201判断语音命令是否是启动命令。

[0090] 启动命令是用于使得摄像设备1转变为可以进行摄像的状态的语音命令。启动命令是在图5A的步骤S107中判断出的命令,而不是用于执行与摄像有关的作业的命令。因此,如果所识别出的语音命令是启动命令,则中央控制单元201忽略该命令并使处理返回到步骤S151。

[0091] 在步骤S202中,中央控制单元201判断语音命令是否是停止命令。停止命令是用于使得状态从可以进行一系列摄像的状态转变为等待启动命令的输入的状态的命令。因此,如果所识别出的语音命令是停止命令,则中央控制单元201使处理进入步骤S211。在步骤S211中,中央控制单元201通过控制电源控制单元211来切断向摄像单元102、声音方向检测单元2044、语音命令识别单元2043、运动图像声音处理单元2045、以及麦克风104b~104d等的电力,并且停止这些单元。然后,中央控制单元201使处理返回到启动时的步骤S103。

[0092] 在步骤S203中,中央控制单元201判断语音命令是否是静止图像拍摄命令。静止图像拍摄命令是用于请求摄像设备1执行一个静止图像的拍摄/记录作业的命令。因此,中央控制单元201在判断为语音命令是静止图像拍摄命令时,使处理进入步骤S212。在步骤S212中,中央控制单元201将通过由摄像单元102进行的拍摄所获得的一个静止图像数据例如作为JPEG文件记录在存储单元206中。注意,静止图像拍摄命令的作业是通过进行一个静止图像的拍摄和记录来完成的,因此该作业不是上述图5B的步骤S152中的判断目标作业。

[0093] 在步骤S204中,中央控制单元201判断语音命令是否是运动图像拍摄命令。运动图

像拍摄命令是用于请求摄像设备1拍摄并记录运动图像的命令。中央控制单元201在判断为语音命令是运动图像拍摄命令时,使处理进入步骤S213。在步骤S213中,中央控制单元201使用摄像单元102来开始运动图像的拍摄和记录,并使处理返回到步骤S151。在本实施例中,所拍摄到的运动图像被存储在存储单元206中,但所拍摄到的运动图像可以经由外部输入/输出端子单元208被发送至网络上的文件服务器。运动图像拍摄命令是用于使得运动图像的拍摄和记录继续的命令,因此该作业是上述图5B的步骤S152中的判断目标作业。

[0094] 在步骤S205中,中央控制单元201判断语音命令是否是运动图像拍摄结束命令。如果语音命令是运动图像拍摄结束命令、并且实际正进行运动图像的拍摄/记录,则中央控制单元201结束记录(作业)。然后,中央控制单元201使处理返回到步骤S151。

[0095] 在步骤S206中,中央控制单元201判断语音命令是否是追踪命令。追踪命令是用于向摄像设备1请求使得用户连续位于摄像单元102的摄像方向上的命令。中央控制单元201在判断为语音命令是追踪命令时,使处理进入步骤S215。然后,在步骤S215中,中央控制单元201开始控制枢转控制单元213,使得对象连续位于视频信号处理单元203所获得的视频的中心位置。此外,中央控制单元201使处理返回到步骤S151。结果,可动摄像单元100通过进行平摇操作或俯仰操作来追踪移动的用户。注意,尽管进行了用户的追踪,但未进行拍摄图像的记录。此外,在进行追踪时,作业是上述图5B的步骤S152中的判断目标作业。然后,在接收到追踪结束命令时,中央控制单元201最终结束运动图像的拍摄和记录。注意,静止图像拍摄命令和运动图像拍摄命令的作业例如可以在进行追踪时执行。

[0096] 在步骤S207中,中央控制单元201判断语音命令是否是追踪结束命令。如果语音命令是追踪结束命令、并且实际上正进行追踪,则中央控制单元201结束追踪(作业)。然后,中央控制单元201使处理返回到步骤S151。

[0097] 在步骤S208中,中央控制单元201判断语音命令是否是自动运动图像拍摄命令。中央控制单元201在判断为语音命令是自动运动图像拍摄命令时,使处理进入步骤S217。在步骤S217中,中央控制单元201开始利用摄像单元102的运动图像拍摄和记录,并且返回到步骤S151。通过自动运动图像拍摄命令所执行的作业与通过上述的运动图像拍摄命令所执行的作业的不同之处在于:每次用户说话时,在镜头单元101的摄像方向指向语音的声源方向的同时,进行运动图像的拍摄/记录。例如,在存在多个说话者的会议的环境中,在进行平摇和俯仰操作的同时记录运动图像,以便在每次讲话时,使说话者进入镜头单元101的视角中。注意,在这种情况下,在正执行自动运动图像拍摄命令的作业时,不接受用于结束该作业的语音命令。假定通过对操作单元205中所设置的预定开关进行操作来结束该作业。此外,在正执行该作业时,中央控制单元201停止语音命令识别单元2043。此外,中央控制单元201参考声音方向检测单元2044在声压级检测单元2041检测到了超过阈值的声压级的定时所检测到的声音方向信息,来进行可动摄像单元104的平摇和俯仰操作。

[0098] 注意,尽管在图6中未示出,但是如果所识别出的语音命令是放大命令,则中央控制单元201通过控制镜头致动器控制单元103来使当前变焦倍率增大预先设置的值。此外,如果所识别出的语音命令是缩小命令,则中央控制单元201通过控制镜头致动器控制单元103来使当前变焦倍率减小预先设置的值。注意,如果镜头单元101已在远摄端或广角端,则不能进一步增大放大率或缩小率,因此在进行这样的语音命令时,中央控制单元201忽略该语音命令。

[0099] 以上进行了说明。除上述语音命令以外的语音命令将在步骤S207之后的步骤中执行,但这里将省略其说明。

[0100] 这里,将按照图8所示的时序图来说明本实施例中的从在摄像设备1中主电源接通时起的序列的示例。

[0101] 在摄像设备1的主电源接通时,声压级检测单元2041开始用于检测来自麦克风104a的声音数据的声压级的处理。假定用户在定时T601开始说出启动命令“你好,照相机”。结果,声压级检测单元2041检测到超过阈值的声压。在被该检测触发的情况下,在定时T602,语音存储器2042开始存储来自麦克风104a的声音数据,并且语音命令识别单元2043开始识别语音命令。在用户说完启动命令“你好,照相机”时,在定时T603,语音命令识别单元2043识别出该语音命令,并指定所识别出的语音命令是启动命令。

[0102] 在定时T603,中央控制单元201在被启动命令的识别触发的情况下,开始向声音方向检测单元2044的电力供给。此外,中央控制单元201还在定时T604开始向摄像单元102的电力供给。

[0103] 假定用户在定时T606开始说出例如“动画开始”。在这种情况下,将说话的开始的定时的声音数据从定时T607起顺次存储在语音存储器2042中。此外,在定时T608,语音命令识别单元2043将声音数据识别为表示“动画开始”的语音命令。语音命令识别单元2043向中央控制单元201通知语音存储器2042中的表示“动画开始”的声音数据的开始和结束地址、以及识别结果。中央控制单元201将由所接收到的开始和结束地址表示的范围确定为有效范围。此外,中央控制单元201从声音方向检测单元2044的缓冲存储器2044a内的有效范围中提取最新的声源方向信息,并且在定时T609,基于所提取的信息,通过控制枢转控制单元213来开始可动摄像单元100的平摇和俯仰操作。

[0104] 在可动摄像单元100正进行平摇和俯仰操作期间、摄像信号处理单元202在定时T612在使用摄像单元102所生成的图像中检测到被摄体(对象:面部)的情况下,中央控制单元201停止平摇和俯仰操作(定时T613)。此外,在定时T614,中央控制单元201向运动图像声音处理单元2045供给电力,以进入利用麦克风104a和104b收集立体声的状态。此外,中央控制单元201在定时T615开始拍摄和记录带有声音的运动图像。

[0105] 接着,将说明本实施例中的声音方向检测单元2044所进行的用于检测声源方向的处理。该处理是在图5A中的步骤S110之后周期性地且连续地进行的。

[0106] 首先,将使用图9A来说明使用两个麦克风(即,麦克风104a和104b)的简单声音方向检测。在图9A中,假定麦克风104a和104b布置在平面上(与平摇操作的转动轴垂直的平面上)。麦克风104a和104b之间的距离由 $d[a-b]$ 表示。假定摄像设备1和声源之间的距离相对于距离 $d[a-b]$ 足够大。在这种情况下,可以通过比较麦克风104a和104b所收集到的声音来指定麦克风104a和104b之间的声音的延迟时间。

[0107] 可以通过将到达延迟时间乘以声速(在空气中约为340m/s)来指定距离 $l[a-b]$ 。结果,可以使用以下的等式来指定声源方向角度 $\theta[a-b]$ 。

[0108]  $\theta[a-b] = \arccos(l[a-b]/d[a-b])$

[0109] 然而,在所获得的声源方向 $\theta[a-b]$ 和 $\theta[a-b]'$ (图9A)之间不能区分通过使用两个麦克风所获得的声音方向。也就是说,不能指定两个方向中的哪个方向。

[0110] 因而,将如下使用图9B和9C来说明本实施例中的声源方向的检测方法。具体地,由

于存在使用两个麦克风可以估计的两个声源方向,因此这两个方向被视为临时方向。此外,使用另外两个麦克风来获得声源方向,并且获得两个临时方向。然后,将在这些临时方向之间共同的方向确定为要获得的声源方向。注意,假定图9B和9C中的上方向是可动摄像单元100的摄像方向。可动摄像单元100的摄像方向也可被改述为镜头单元101的光轴方向(主轴方向)。

[0111] 图9B示出使用三个麦克风的方法。将使用麦克风104a、104b和104c来给出说明。在如图3A所示的布置中,与麦克风104a和104b排列的方向垂直的方向是镜头单元101的摄像方向。

[0112] 如参考图9A所述,根据麦克风104a和104b的位置,距离 $d[a-b]$ 是已知的,因此可以根据声音数据指定距离 $I[a-b]$ ,则可以指定 $\theta[a-b]$ 。此外,由于麦克风104a和104c之间的距离 $d[a-c]$ 是已知的,因此也可以根据声音数据指定距离 $I[a-c]$ ,并且可以指定 $\theta[a-c]$ 。如果可以计算出 $\theta[a-b]$ 和 $\theta[a-c]$ ,则可以将在与麦克风104a、104b和104c布置于的平面相同的二维平面上(与平摇操作的转动轴垂直的平面上)的这些角度之间共同的角度确定为准确的声音产生方向。

[0113] 将使用图9C来说明使用四个麦克风确定声源方向的方法。由于图3A所示的麦克风104a、104b、104c和104d的布置,因此与麦克风104a和104b排列的方向垂直的方向是镜头单元101的摄像方向(光轴方向)。在使用四个麦克风的情况下(即,在使用各自位于对角线上的麦克风104a和104d的配对以及麦克风104b和104c的配对这两个配对的情况下),可以准确地计算出声源方向。

[0114] 由于麦克风104a和104d之间的距离 $d[a-d]$ 是已知的,因此可以根据声音数据指定距离 $I[a-d]$ ,并且也可以指定 $\theta[a-d]$ 。

[0115] 此外,由于麦克风104b和104c之间的距离 $d[b-c]$ 是已知的,因此可以根据声音数据指定距离 $I[b-c]$ ,并且也可以指定 $\theta[b-c]$ 。

[0116] 因此,一旦 $\theta[a-d]$ 和 $\theta[b-c]$ 是已知的,可以在与麦克风布置于的平面相同的二维平面上准确地检测到声音产生方向。

[0117] 此外,通过增加诸如 $\theta[a-b]$ 和 $\theta[c-d]$ 等的检测角度的数量,也可以提高方向角度的检测精度。

[0118] 为了进行上述处理,如图3A所示,麦克风104a和104b以及麦克风104c和104d布置在矩形的四个顶点处。注意,麦克风的数量无需是四个,并且可以是三个,只要三个麦克风没有排成直线即可。

[0119] 上述方法的缺点是仅可以检测到同一二维平面上的声音方向。因此,在声源位于摄像设备1的正上方的情况下,不能检测到方向。因此,接着,将参考图10A和10B来说明声音方向检测单元2044中的与存在声源的方向是否是正上方向有关的判断的原理。

[0120] 图10A是示出使用三个麦克风的方法的图。将使用麦克风104a、104b和104c来给出说明。在麦克风如图3A所示布置时,与麦克风104a和104b排列的方向垂直的方向是镜头单元101的摄像方向(光轴方向)。麦克风104a和104b排列的方向是连接麦克风104a的中心点和麦克风104b的中心点的直线的方向。

[0121] 将说明声音在与声音输入单元104布置于的平面相交的直线的方向上(即,从上方)进入麦克风104a、104b和104c的情况。



[0122] 这里,在声源位于摄像设备1的正上方的情况下,可以认为麦克风104a和104b相对于声源在等距离处。也就是说,在来自声源的声音在两个麦克风104a和104b之间的到达时间方面不存在差异。因此,可以认识到,在与连接麦克风104a和104b的直线垂直相交的方向上存在声源。

[0123] 此外,同样可以认为,麦克风104a和104c相对于声源在等距离处,因此在来自声源的声音在两个麦克风104a和104c之间的到达时间方面也不存在差异。因此,可以认识到,在与连接麦克风104a和104c的直线垂直相交的方向上存在声源。

[0124] 也就是说,在麦克风104a和104b所检测到的声音的时间差的绝对值由 $\Delta T1$ 表示、麦克风104a和104c所检测到的声音的时间差的绝对值由 $\Delta T2$ 表示、并且与预先设置的足够小的阈值 $\epsilon$ 的关系满足以下条件的情况下,可以判断为声源位于摄像设备1的正上方。

[0125] 条件: $\Delta T1 < \epsilon$  且  $\Delta T2 < \epsilon$

[0126] 将参考图10B来说明使用四个麦克风104a、104b、104c和104d的位于摄像设备1的正上方的声源的检测方法。如图3A所示,将考虑麦克风104a和104d的配对以及麦克风104b和104c的配对。

[0127] 在摄像设备1的正上方存在声源的情况下,麦克风104a和104d相对于声源在等距离处,这些麦克风104a和104d所检测到的声音的时间差的绝对值 $\Delta T3$ 为零或极小的值。也就是说,认识到,在与连接麦克风104a和104d的直线垂直相交的方向上存在声源。

[0128] 此外,由于麦克风104b和104c也相对于声源在等距离处,因此这些麦克风104b和104c所检测到的声音的时间差的绝对值 $\Delta T4$ 也为零或极小的值。也就是说,认识到,在与连接麦克风104b和104c的直线垂直相交的方向上存在声源。因此,如果满足以下条件,则可以判断为声源位于摄像设备1的正上方。

[0129] 条件: $\Delta T3 < \epsilon$  且  $\Delta T4 < \epsilon$

[0130] 如上所述,针对三个或更多个麦克风中的两对麦克风获得声音的到达时间差的绝对值,并且在这两个绝对值都小于足够小的阈值的情况下,可以判断为存在声源的方向是正上方向。注意,在确定了两对时,允许任何组合,只要这两对的方向不彼此平行即可。

[0131] 以上说明了第一实施例。根据上述实施例,防止了除说出语音命令的人物(其面部)以外的对象被误识别为被摄体。此外,可以执行说出了语音命令的人物所期望的作业。

[0132] 此外,如上述实施例所述,在中央控制单元201的控制下,向各个麦克风104a~104d和构成声音信号处理单元204的各元件的电力是在进入实际使用该元件的阶段之后供给的,因此与所有的构成元件都处于可操作状态的情况相比,可以抑制电力消耗。

[0133] [第二实施例]

[0134] 将说明第二实施例。假定第二实施例中的设备的结构与第一实施例的该结构相同。此外,假定语音命令识别单元2043已将运动图像拍摄命令识别为语音命令、并且声音方向检测单元2044已检测到说出语音命令的用户存在于的方向,来给出说明。

[0135] 在第一实施例中,摄像设备1在将运动图像拍摄命令识别为语音命令时,进行可动摄像单元100的平摇和俯仰操作,使得语音命令的源存在于的方向与镜头单元101的摄像方向(光轴方向)一致。此外,摄像设备1在进行平摇和俯仰操作时进行被摄体检测处理。此外,最后,摄像设备1在摄像单元102所拍摄到的图像中存在被摄体的条件下,开始被摄体的运动图像的拍摄和记录。

[0136] 因此,即使用户说出了运动图像拍摄命令,用户也需要等待小段时间(等待时间),直到执行了作业为止。此外,还存在如下的问题:由于在进行平摇操作期间、从摄像单元102获得的图像沿水平方向流动,因此难以进行镜头单元101的调焦。

[0137] 因此,在第二实施例中,将说明如下的示例:在缩短等待时间的同时,可以抑制利用摄像单元102在失焦状态下的作业的开始。具体地,中央控制单元201使得枢转控制单元213以高速的第一平摇操作速度进行平摇操作,以面向声音方向检测单元2044所检测到的声音方向(以下称为“目标方向”)。以该第一平摇操作速度进行驱动的时间段例如是直到达到紧挨在开始该平摇操作之前的镜头单元101的摄像方向(光轴方向)与目标方向之间的角度差的一半为止的时间段,并且摄像设备1在该时间段(以第一平摇操作速度进行驱动的时间段)中不进行调焦控制。另外,在镜头单元101的摄像方向(光轴方向)与目标方向之间的角度差已减小到识别出语音时的这两者之间的角度差的一半的情况下,中央控制单元201控制枢转控制单元213以按第二平摇操作速度进行平摇,该第二平摇操作速度比第一平摇操作速度慢,并且在该第二平摇操作速度下可以进行调焦。此外,中央控制单元201在以第二平摇操作速度进行平摇操作的时间段内,通过控制镜头致动器控制单元103来开始镜头单元101的调焦控制。此外,在满足以下条件时,中央控制单元201开始运动图像的拍摄和记录。

[0138] 条件:摄像单元102所拍摄到的图像处于聚焦状态,并且镜头单元101的摄像方向(光轴方向)与目标方向之间的角度差为预先设置的阈值 $\alpha$ 或更小

[0139] 这里,阈值 $\alpha$ 是由摄像单元102可以拍摄的视角的左端和右端及其中心形成的角度。这是因为,如果该角度减小到预先设置的阈值 $\alpha$ 或更小,则可以估计为被摄体将开始进入摄像单元102所拍摄的图像。这里,第二平摇操作速度 $v_2$ 是从作为初始速度的第一平摇操作速度 $v_1$ 起以预定速率减速的速度。

[0140] 图11A和11B是示出通过第二实施例中的语音命令识别单元2043识别出语音命令而触发了开始的处理的流程图。与该流程图有关的程序存储在中央控制单元201内所包括的ROM中。图12是示出在第二实施例中识别出语音命令之后的处理的序列的图。以下将参考这些图来说明第二实施例中的中央控制单元201的处理。另外,这里将假定运动图像拍摄命令已被识别为语音命令来给出说明。

[0141] 首先,在步骤S1001中,中央控制单元201从声音方向检测单元2044的缓冲存储器2044a获取在由语音命令识别单元2043识别出语音命令所基于的声音数据的开始和结束地址所表示的有效时间段中(也在说话时间段中)检测到的声音方向信息。另外,在步骤S1002中,中央控制单元201判断由所获取到的声音方向信息表示的方向是否在可动摄像单元100的摄像单元102的当前视角中。

[0142] 如果由所获取到的声音方向信息表示的方向在摄像单元102的当前视角中,则中央控制单元201使处理进入步骤S1003,并且开始所识别出的运动图像的拍摄和记录。

[0143] 另一方面,如果由所获取到的声音方向信息表示的方向在可动摄像单元100的当前视角外,则中央控制单元201使处理进入步骤S1004。在步骤S1004中,中央控制单元201根据镜头单元101的当前摄像方向(光轴方向)A和步骤S1001中所获取到的声源方向C来计算要进行的平摇操作的角度 $\theta$ ,并且计算在如下情况下进行平摇操作所需的时间(= $t_4-t_0$ ):在从摄像方向A起直到整个平摇操作的角度的一半的角度为止的区域(前半区间)中以第一

平摇操作速度 $v_1$ 进行平摇操作,并且在该角度的其余部分的区域(后半区间)中使平摇操作以预定速率减速。

[0144] 将假定镜头单元101使用利用如下方法的PI(光遮断器)来给出说明,在该方法中,通过受光部检测被物体遮断的来自发光部的光来检测位置。在这种情况下,当前镜头单元101的方向由A表示,目标方向由C表示,并且由镜头单元101的方向A和目标方向C形成的角度的一半角度处的方向由B表示。根据由镜头单元101的方向A和目标方向C形成的角度 $\theta$ 来确定PI的步数PIAC。中央控制单元201根据该步数PIAC,计算以第一平摇操作速度进行驱动的前半区间的距离DAB、以及以第二平摇操作速度进行驱动的后半区间的距离DBC。可以通过 $DAB/v_1$ 来获得在前半区间以第一平摇操作速度进行驱动的时间段( $t_1-t_0$ )。此外,在后半区间以第二平摇操作速度 $v_2$ 进行驱动的时间段( $t_4-t_1$ )是 $DBC/v_2$ (注意,在本实施例中, $DBC=DAB$ )。这里,如上所述,第二平摇操作速度 $v_2$ 是从作为初始速度的第一平摇操作速度 $v_1$ 起以预定速率减速的速度。

[0145] 在步骤S1006中,中央控制单元201通过控制枢转控制单元213,开始可动摄像单元100以第一平摇操作速度 $v_1$ 向着目标方向的平摇操作。此外,在步骤S1007中,中央控制单元201继续该平摇操作,直至判断为平摇操作的时间段达到了 $DAB/v_1$ 为止。

[0146] 然后,在从开始平摇操作起的时间段达到 $DAB/v_1$ 的情况下,在步骤S1008中,中央控制单元201将平摇操作速度设置为第二平摇操作速度 $v_2$ ,直至结束平摇操作的时刻 $t_4$ 为止。另外,在步骤S1009中,中央控制单元201判断由镜头单元101的当前摄像方向(光轴方向)和目标方向形成的角度是否为预先设置的阈值 $\alpha$ 或更小。如果该角度为预先设置的阈值 $\alpha$ 或更小,则被摄体可能在视角中,因此在步骤S1010中,中央控制单元201确定调焦状态值AF。另外,在步骤S1011中,中央控制单元201判断所获得的调焦状态值AF是否为预定阈值 $AF_{th}$ 或更小。在调焦状态值AF为预定阈值 $AF_{th}$ 或更小的情况下,进入聚焦状态。因此,即使平摇操作未完成,中央控制单元201也使处理进入步骤S1003,以开始基于所识别出的运动图像拍摄命令的作业。

[0147] 另一方面,如果调焦状态值AF超过预定阈值 $AF_{th}$ ,则中央控制单元201使处理进入步骤S1012,并且判断是否经过了最初规划的平摇操作时间段。注意,在步骤S1012中,中央控制单元201可以判断镜头单元101的摄像方向(光轴方向)是否达到目标方向。另外,如果步骤S1012的判断为否定(“否”),则中央控制单元201使处理返回到步骤S1008。

[0148] 另一方面,如果步骤S1012的判断为肯定(“是”),则中央控制单元201使处理进入步骤S1013。在步骤S1013中,中央控制单元201通过控制枢转控制单元213,将利用枢转驱动单元的驱动从平摇操作切换到俯仰操作,并且开始俯仰操作。

[0149] 另外,在步骤S1014中,中央控制单元201继续步骤S1013的俯仰操作,直到判断为检测到被摄体或达到俯仰上限为止。如果检测到被摄体(定时 $t_5$ ),则中央控制单元201停止俯仰操作,并且使处理进入步骤S1003。另一方面,如果在未检测到被摄体的情况下达到了俯仰上限(定时 $t_6$ ),则中央控制单元201进行错误处理。对于该错误,错误处理可以是不执行所识别出的静止图像拍摄开始命令等的处理。

[0150] 尽管以上进行了说明,但如果将根据第二实施例的处理应用于第一实施例,则步骤S160及其之后的处理仅需被第二实施例中的步骤S1002及其之后的处理替代。在这种情况下,第二实施例中的步骤S1015的处理对应于步骤S163的处理。

[0151] 注意,在上述第二实施例中,说明了如下的示例:将用于使得镜头单元101的摄像方向(光轴方向)与目标方向一致的平摇操作时间段划分为两个时间段,并且在前一时间段中设置第一平摇操作速度,且在后一时间段中,将平摇操作速度切换到比第一平摇操作速度慢的第二平摇操作速度。然而,可以采用如下的结构:在经过了以第一平摇操作速度的驱动时间段之后,可以按固定速率进行减速,使得在目标方向上平摇操作速度为零。

[0152] 如上所述,根据第二实施例,即使在产生特定语音的位置在当前视角A外、并且需要通过进行平摇和俯仰驱动以改变视角来拍摄被摄体的情况下,也可以在不需要的平摇和俯仰驱动或调焦之后开始运动图像记录。此外,作为切换平摇操作的速度的结果,可以进行调焦,并且开始运动图像记录的定时也可被设置成比第一实施例中的该定时提前。

[0153] 在上述实施例中,说明了声压级检测单元2041、语音命令识别单元2043、声音方向检测单元2044和运动图像声音处理单元2045等是与中央控制单元201分开的处理单元的示例。然而,结构可以如下:中央控制单元201通过执行程序来替换这些处理单元中的全部或部分。

[0154] (其它实施例)

[0155] 还可以通过读出并执行记录在存储介质(还可被更完整地称为“非暂时性计算机可读存储介质”)上的计算机可执行指令(例如,一个或多个程序)以进行上述实施例中的一个或多个的功能以及/或者包括用于进行上述实施例中的一个或多个的功能的一个或多个电路(例如,专用集成电路(ASIC))的系统或设备的计算机和通过下面的方法来实现本发明的实施例,其中,该系统或设备的计算机通过例如从存储介质读出并执行计算机可执行指令以进行上述实施例中的一个或多个的功能以及/或者控制该一个或多个电路以进行上述实施例中的一个或多个的功能来进行上述方法。

[0156] 本公开内容不限于上述实施例,并且可以在本发明的精神和范围内进行各种改变和修改。因此,为了公开本发明的范围,添加了所附的权利要求书。

[0157] 本申请要求2017年12月26日提交的日本专利申请2017-250106和2018年11月2日提交的日本专利申请2018-207633的优先权,这两个申请的全部内容通过引用而被包含于此。

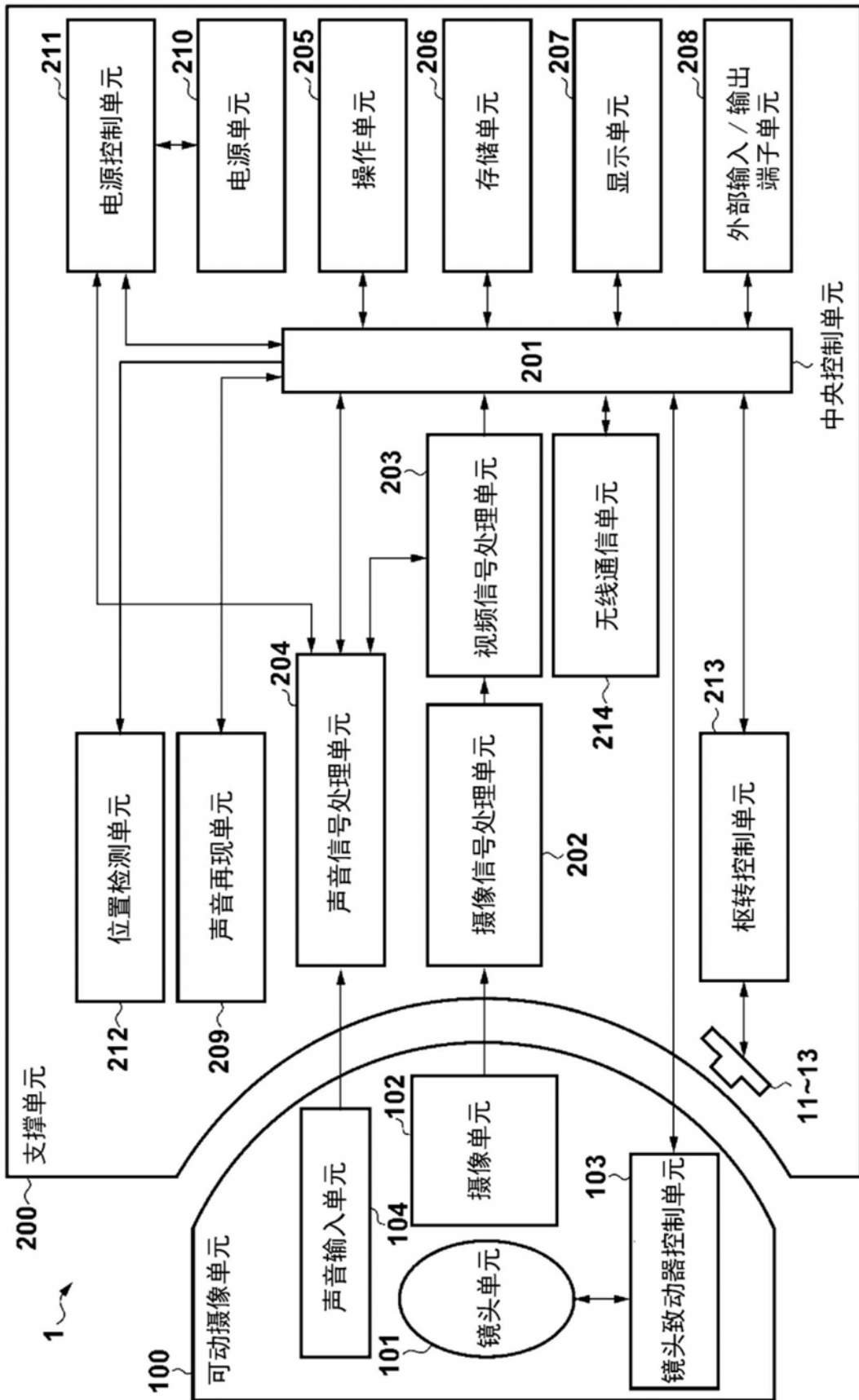


图1

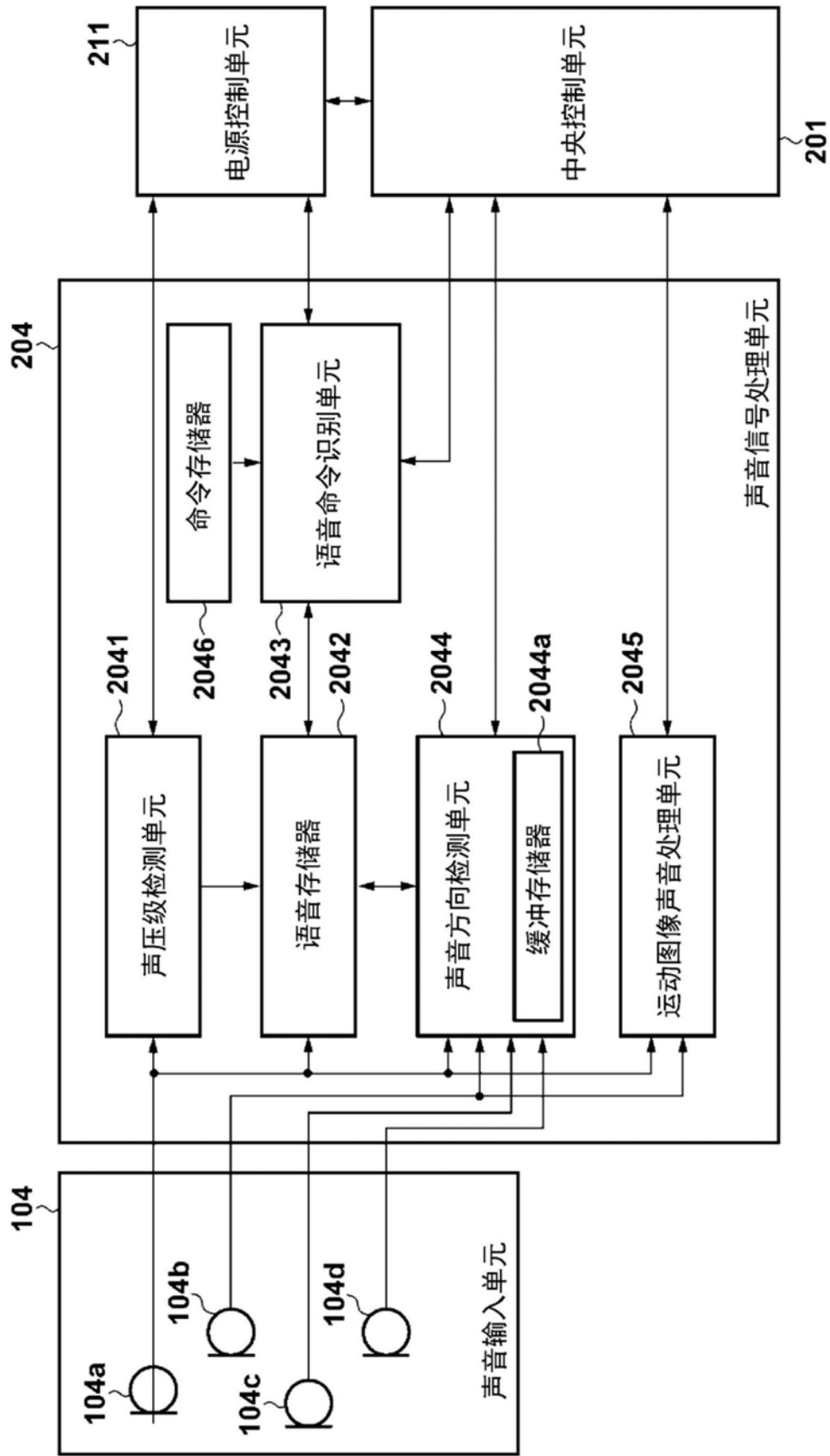


图2

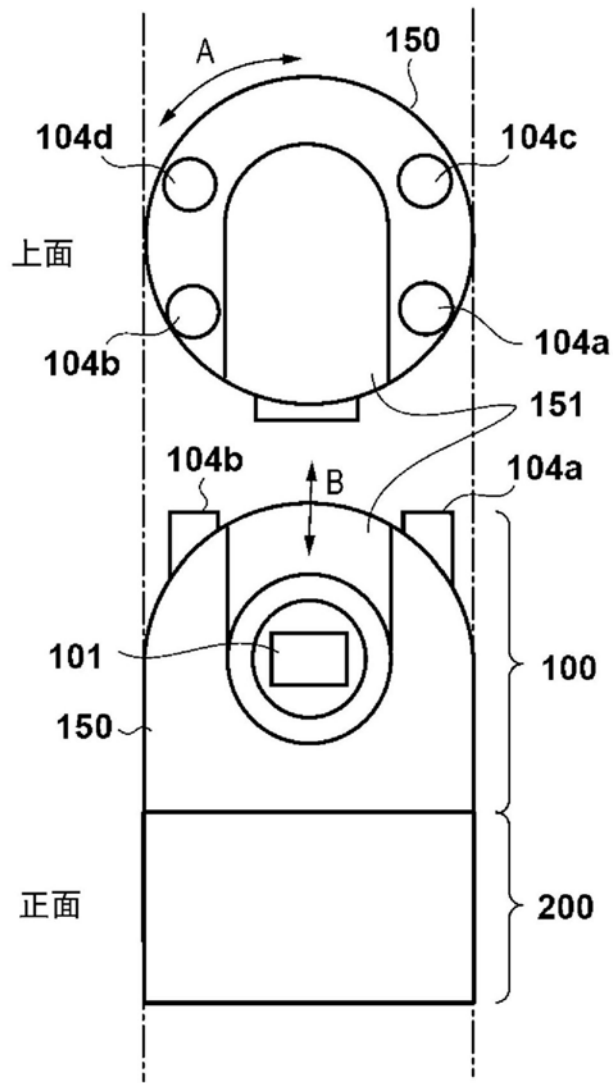


图3A

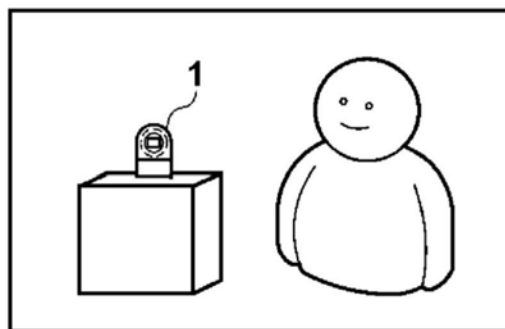


图3B



图3C



图3D

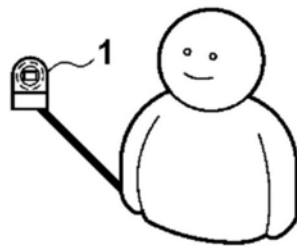


图3E



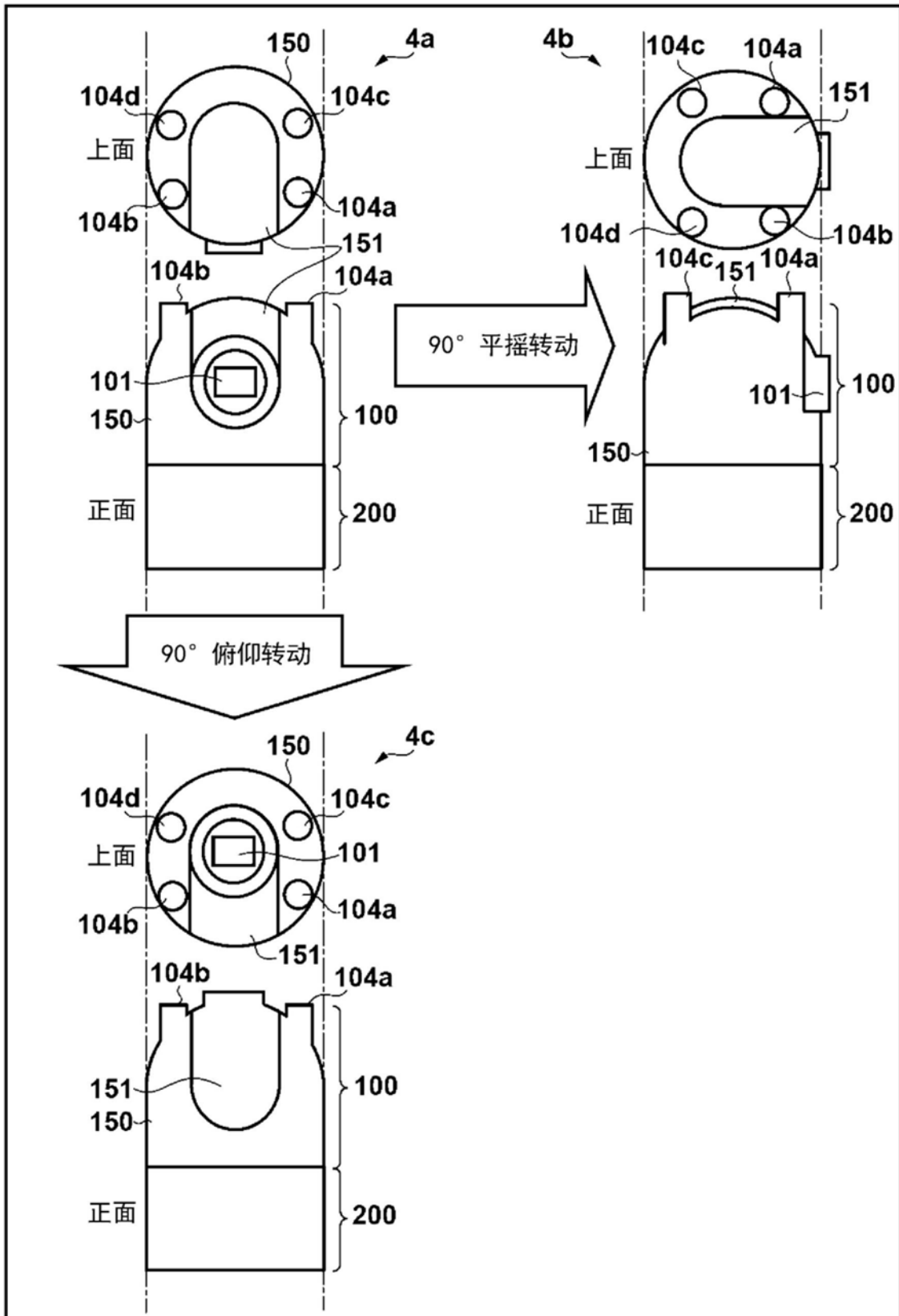


图4

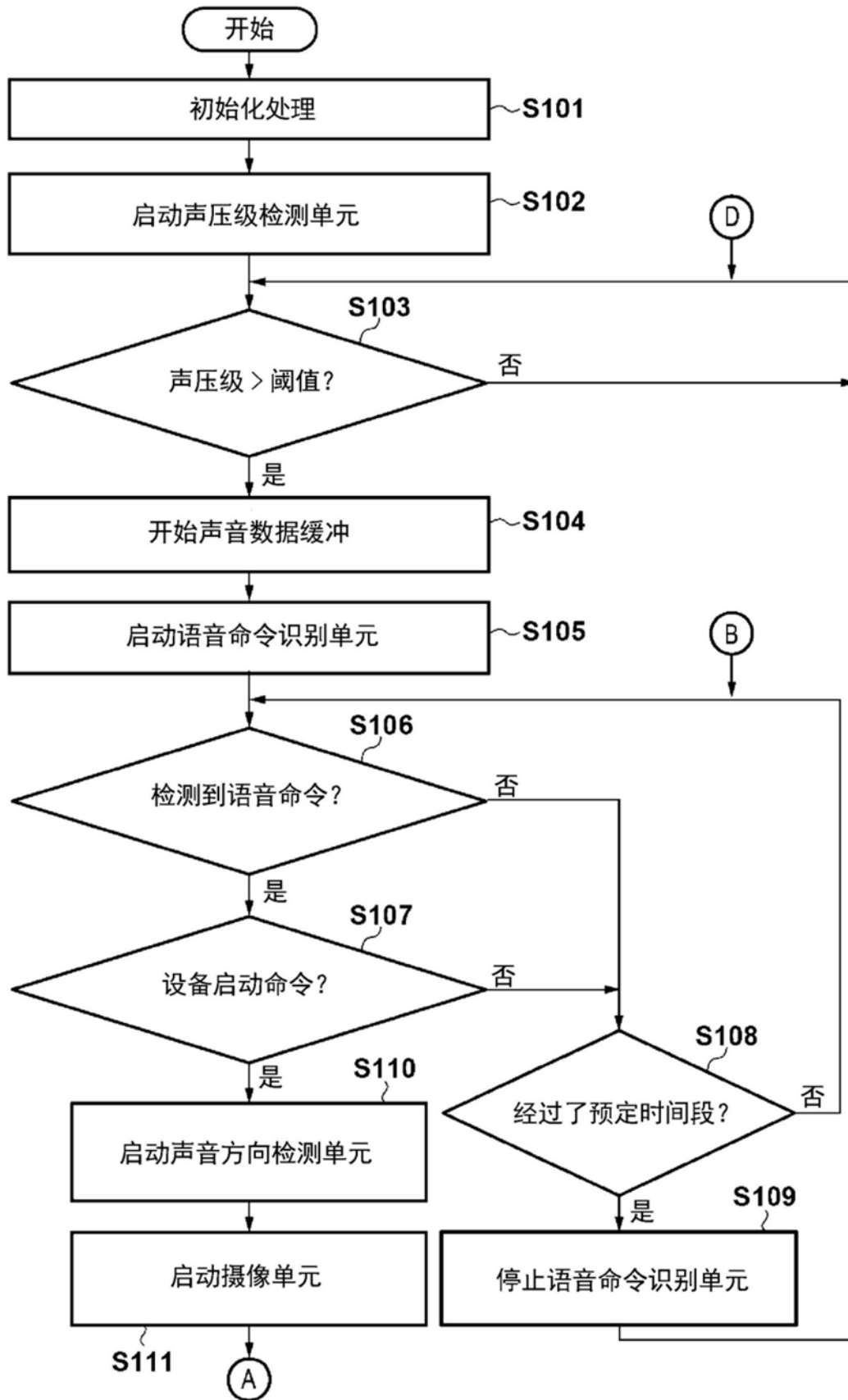


图5A

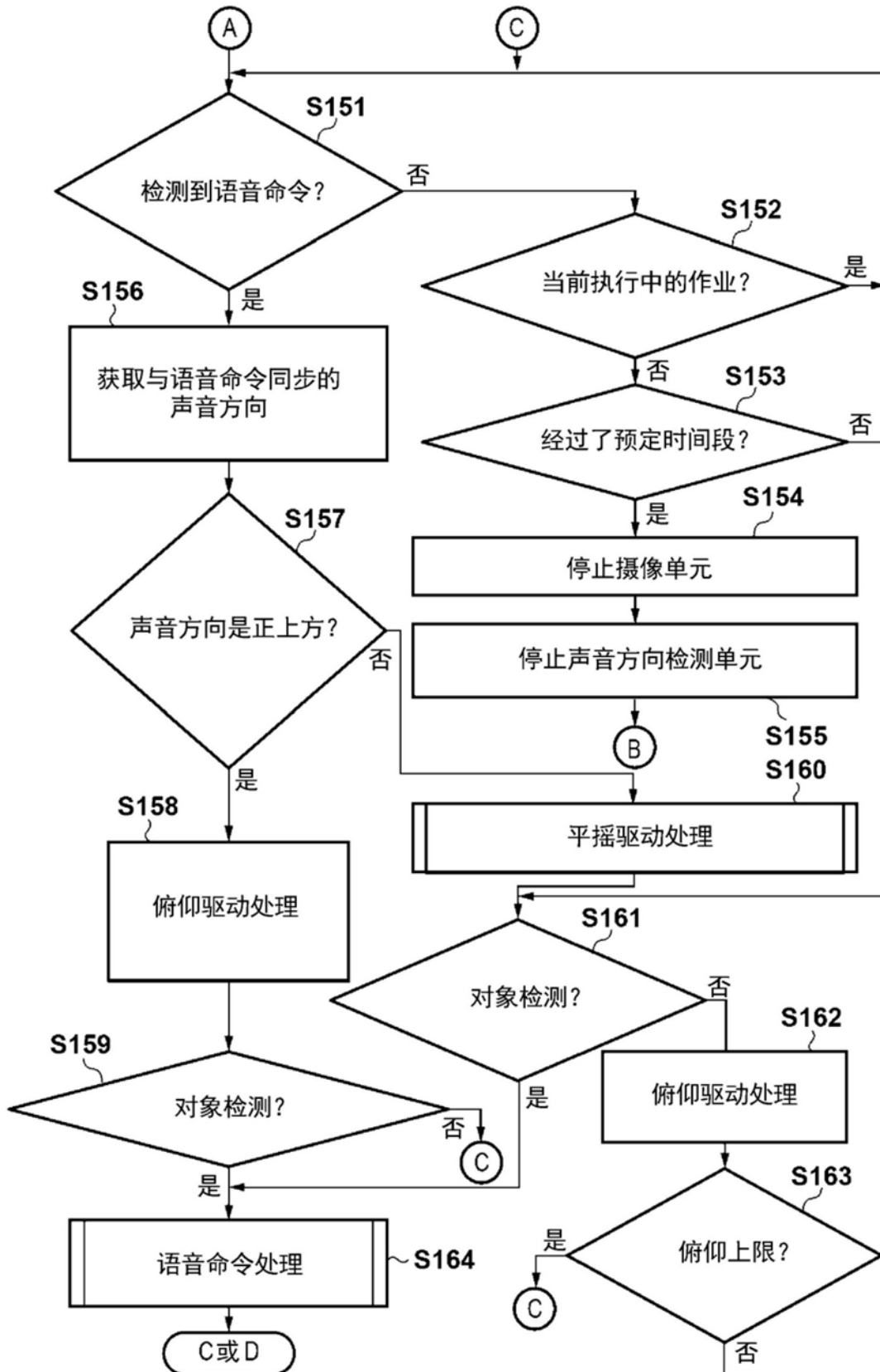


图5B

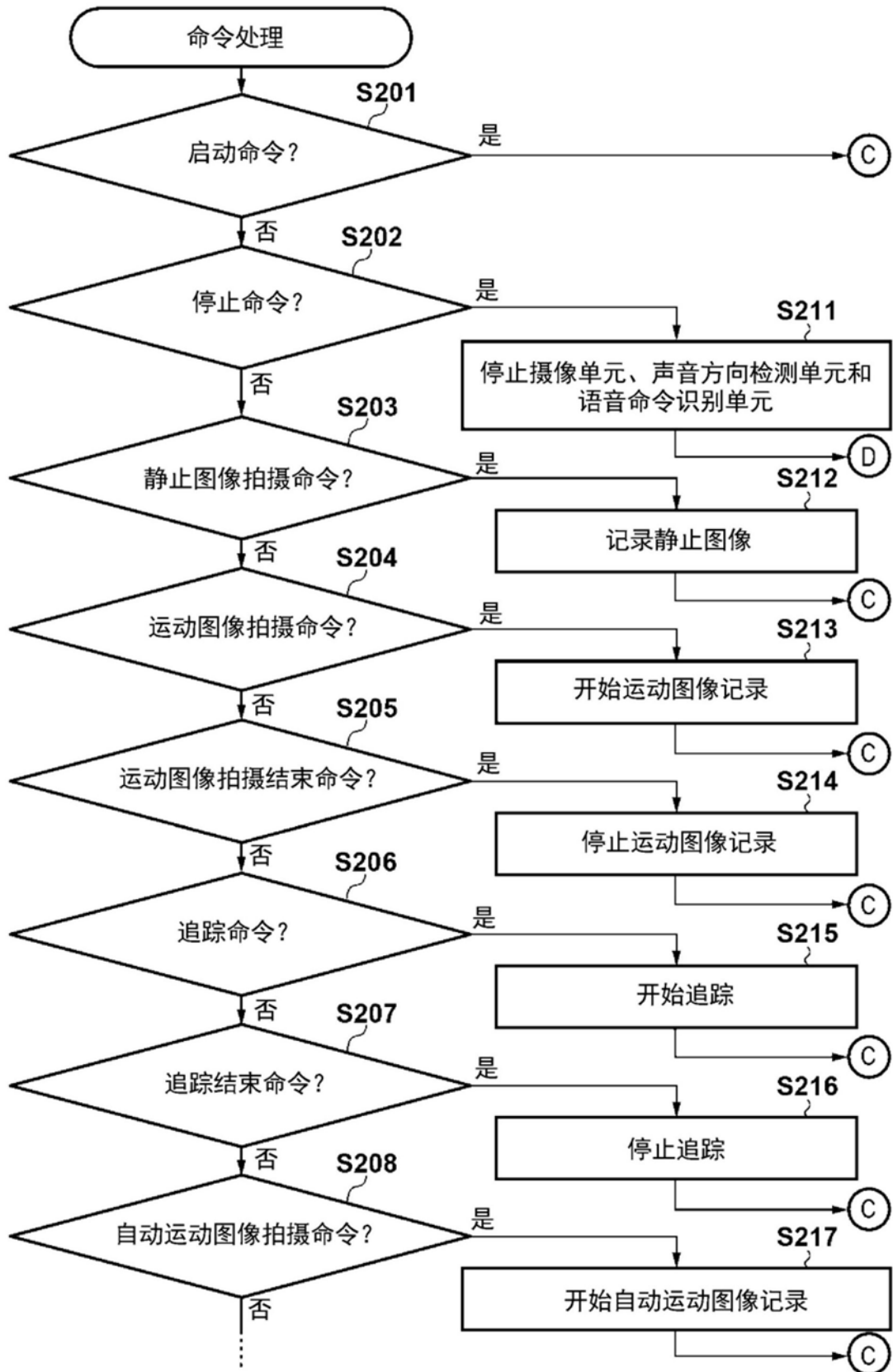


图6

## 命令

启动	“你好，照相机”、“醒来，照相机”
停止	“再见，照相机”、“拜拜，照相机”
拍摄静止图像	“对我拍快照”、“拍照”
开始运动图像拍摄	“开始动画”、“拍视频”
结束运动图像拍摄	“停止动画”
开始追踪	“跟随我”、“追踪我”
结束追踪	“停止跟随”、“停止追踪”
⋮	
开始自动运动 图像拍摄	“开始自动动画”
⋮	
放大命令	“放大”
缩小命令	“缩小”

图7

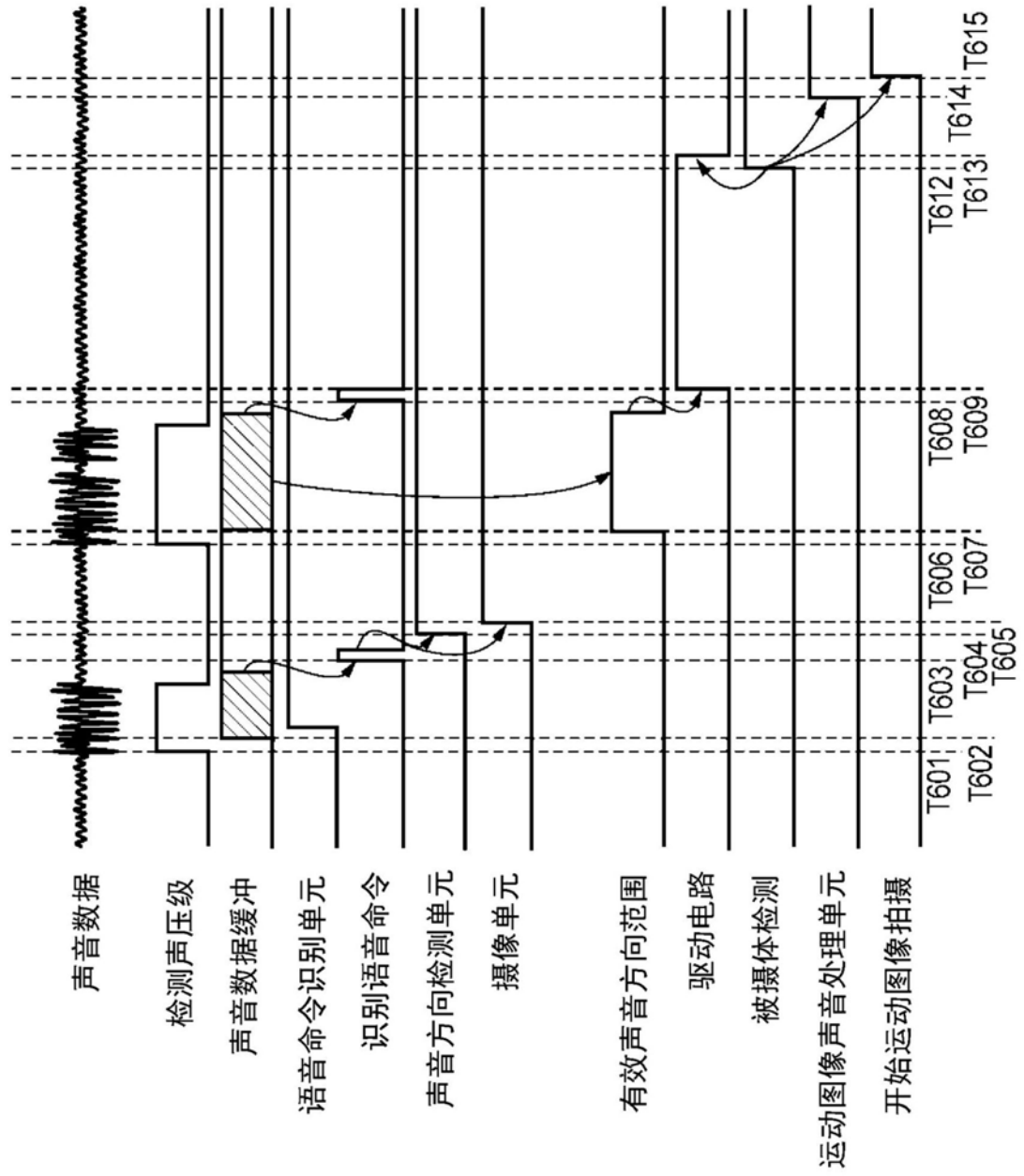


图8

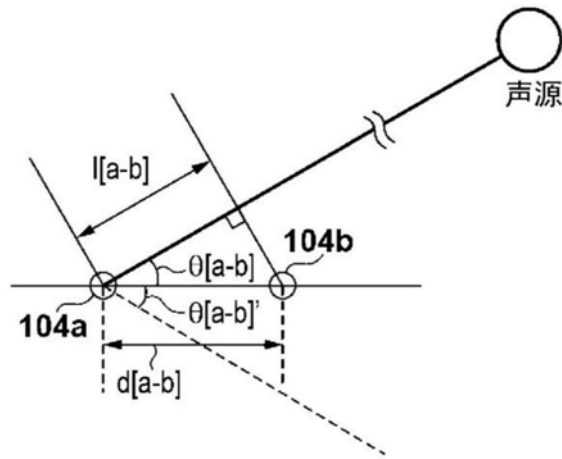


图9A

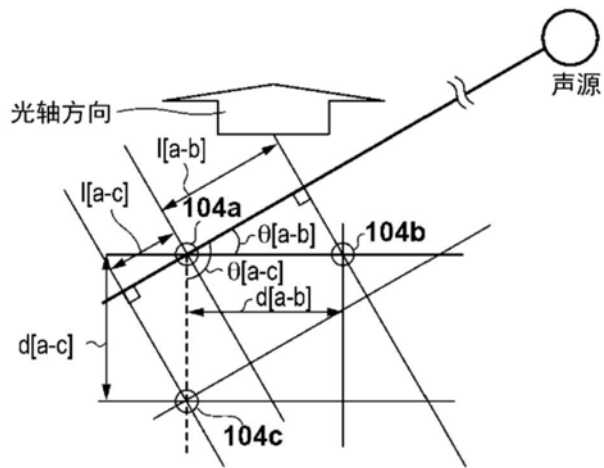


图9B

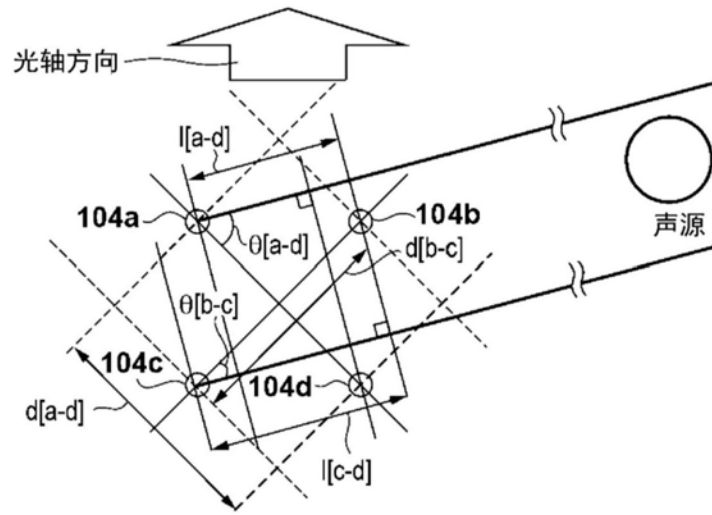


图9C

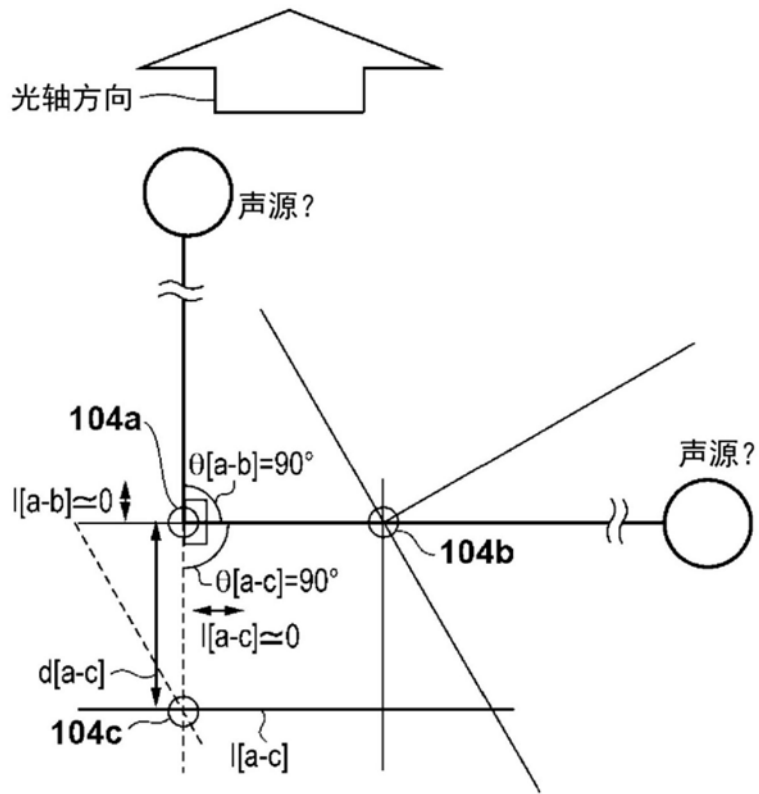


图10A



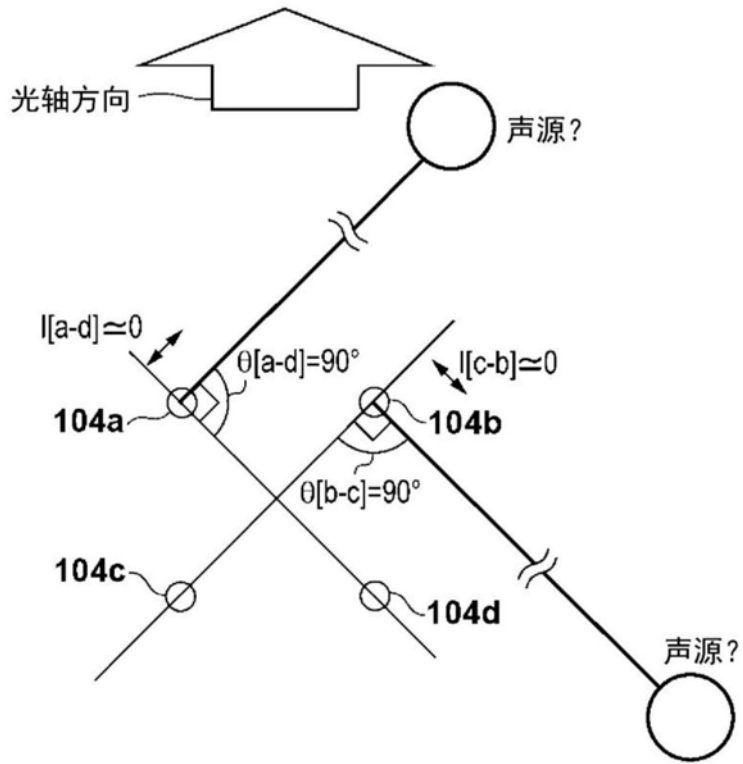


图10B

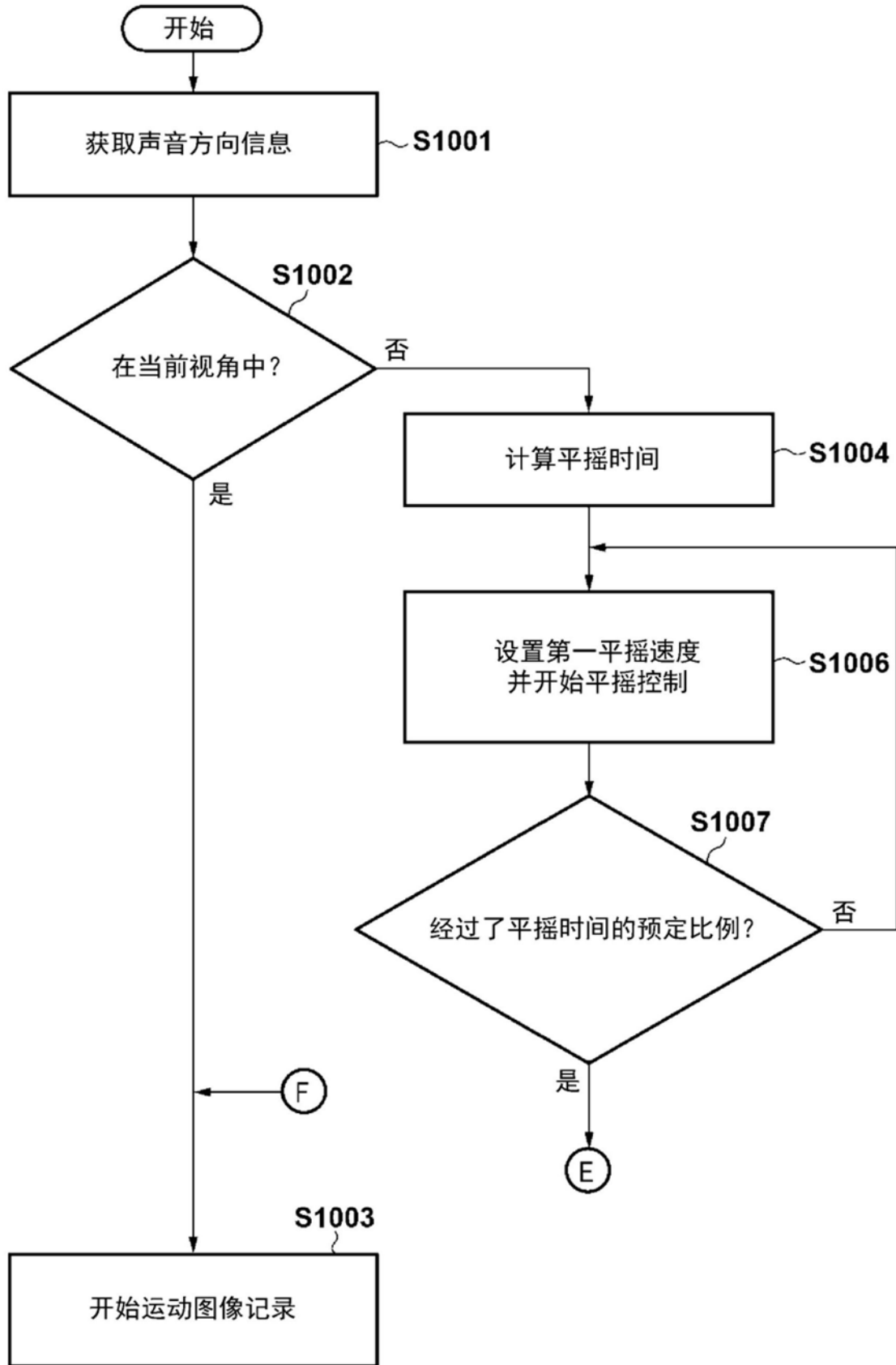


图11A

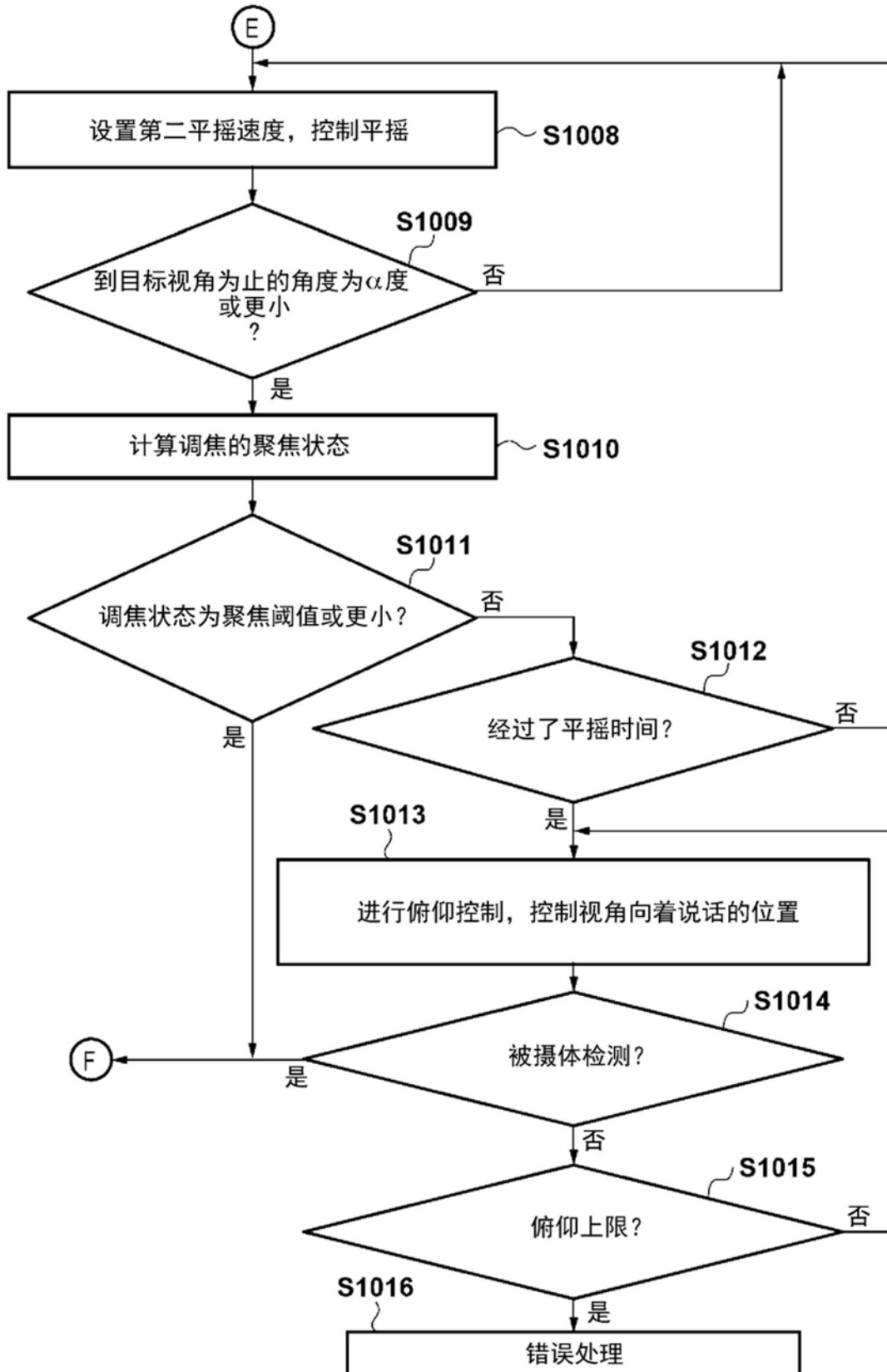


图11B

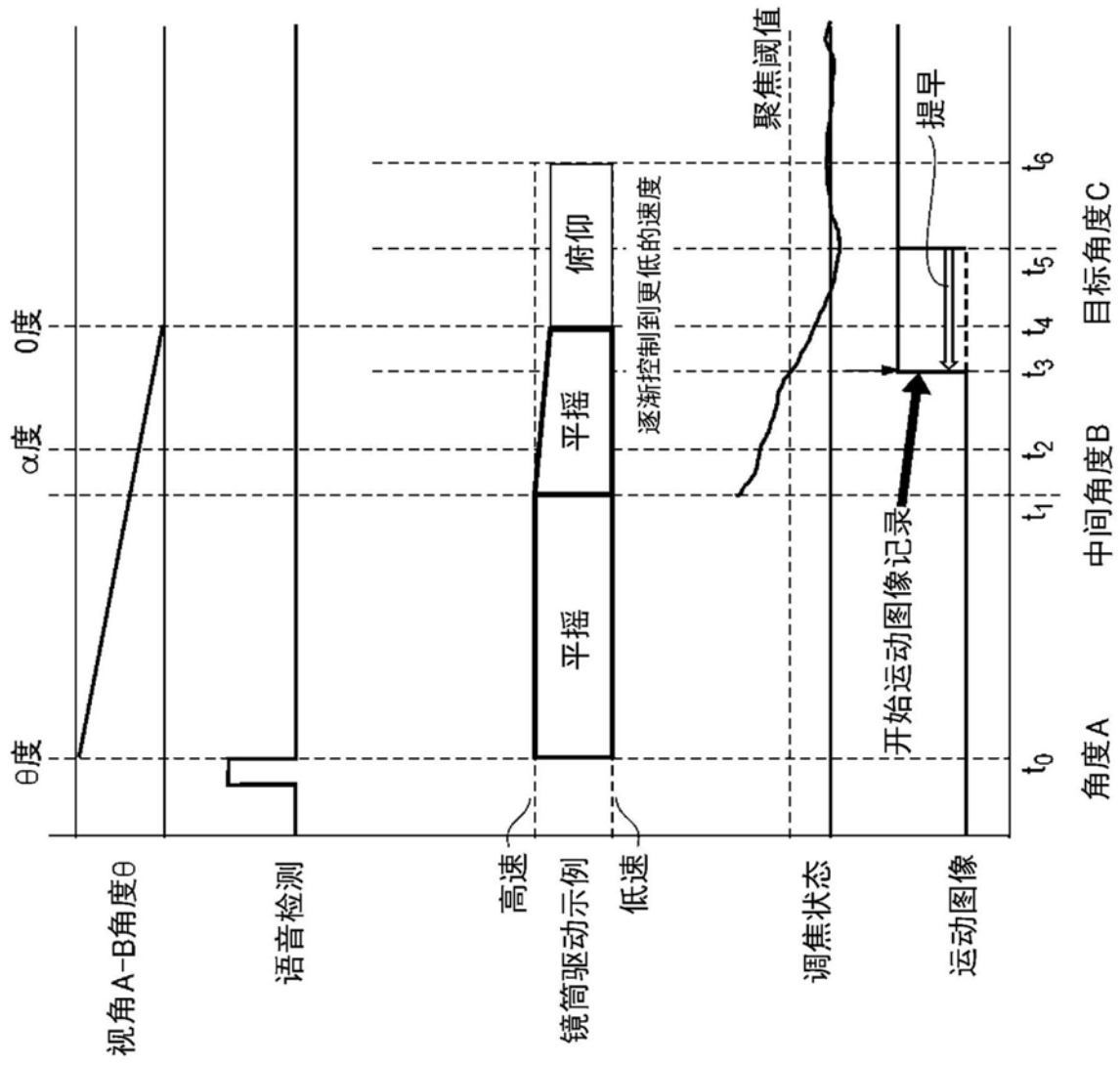


图12