(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 108601519 B
(45)授权公告日 2021.04.06

(21)申请号 201780009674.X

(22)申请日 2017.01.27

(65)同一申请的已公布的文献号
申请公布号 CN 108601519 A

(43)申请公布日 2018.09.28

(30)优先权数据
15/013,537 2016.02.02 US

(85)PCT国际申请进入国家阶段日
2018.08.02

(86)PCT国际申请的申请数据
PCT/US2017/015463 2017.01.27

(87)PCT国际申请的公布数据
WO2017/136245 EN 2017.08.10

(73)专利权人 电子湾有限公司
地址 美国加利福尼亚州

(72)发明人 D·格拉斯哥 D·拉马德奇
M·麦克劳林 J·范温克尔
C·谢尔曼

(74)专利代理机构 北京集佳知识产权代理有限公司 11227
代理人 陈炜

(51)Int.Cl.
*A61B 3/00*(2006.01)

审查员 杨琼

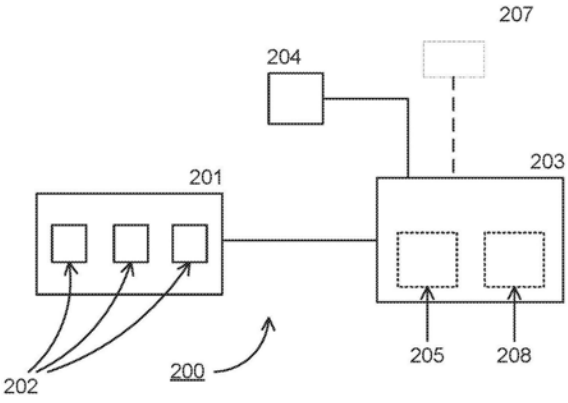
权利要求书2页 说明书7页 附图3页

(54)发明名称

个性化的实时音频处理

(57)摘要

用于实时音频处理的装置和方法采用注视检测传感器来检测用户注视方向,并输出与检测到的用户注视方向相对应的注视信号。数字信号处理单元响应于与在装置处接收到的多个声音相对应的多个信号,以及所确定的注视方向,以使用注视信号从多个信号中识别感兴趣的信号。处理感兴趣的信号以输出给用户。在实施例中,麦克风阵列提供多个信号。成像传感器可以与麦克风阵列或注视检测传感器一起工作以识别感兴趣的信号。




CN 108601519 B

1.一种装置,包括:

注视检测传感器,其配置成检测用户注视方向并输出与检测到的用户注视方向相对应的注视信号;

数字信号处理单元,其配置成:

响应于识别到多个声音信号和检测到的注视方向,所述多个声音信号对应于在所述装置处接收到的多个声音:

生成识别所述多个声音信号中的多个信号的列表;

通过图形界面呈现所述列表;

通过图形界面接收对所述多个声音信号中的多个信号中感兴趣的信号的选择;以及

处理感兴趣的信号以用于输出。

2.根据权利要求1所述的装置,其特征在于,还包括麦克风阵列,所述麦克风阵列配置成接收所述多个声音并输出所述多个信号,所述麦克风阵列还配置成相对于所述注视检测传感器的定位提供指示所述多个信号的方向性的信息。

3.根据权利要求2所述的装置,其特征在于,还包括:

成像传感器,其配置成检测图像并将与检测到的图像对应的图像信号输出到数字信号处理单元;

其中数字信号处理单元还配置成:

确定图像信号是否识别出在所述用户注视方向上说话的人;以及

使用图像信号从所述多个声音信号中识别所述多个信号。

4.根据权利要求3所述的装置,其特征在于,所述数字信号处理单元使用来自所述注视检测传感器和所述成像传感器的数据以识别所述多个声音信号中的多个信号。

5.根据权利要求3所述的装置,其特征在于,所述数字信号处理单元使用来自所述注视检测传感器和所述麦克风阵列的数据以识别所述多个声音信号中的多个信号。

6.根据权利要求3所述的装置,其特征在于,所述数字信号处理单元使用来自所述成像传感器和所述麦克风阵列的数据以识别所述多个声音信号中的多个信号。

7.根据权利要求1所述的装置,其特征在于,还包括覆盖开关,以覆盖所述注视检测传感器。

8.根据权利要求7所述的装置,其特征在于,响应于来自覆盖开关的信号,数字信号处理单元在有限的时间段中忽略来自注视检测传感器的输入。

9.根据权利要求1所述的装置,其特征在于,所述数字信号处理单元包括用于输出通信信号以指示来自用户的通信请求的设备。

10.根据权利要求1所述的装置,其特征在于,所述数字信号处理单元包括配置成从声源接收识别信号的设备,所述识别信号包括识别声源本身或与声源相关联的人的元数据。

11.一种方法,包括:

检测用户注视方向;

输出与检测到的用户注视方向对应的注视信号;

响应于识别到多个声音信号和检测到的注视方向,所述多个声音信号对应于接收到的多个声音:

生成识别所述多个声音信号中的多个信号的列表;

使得所述列表通过图形界面呈现；

接收对所述多个声音信号中的多个信号中感兴趣的信号的选择；和

处理感兴趣的信号以进行输出。

12.根据权利要求11所述的方法，其特征在于，还包括使用输出所述多个信号的麦克风阵列的相对定位，以及输出所述注视信号的注视检测传感器，以提供指示所述多个信号的方向性的信息。

13.根据权利要求12所述的方法，其特征在于，还包括：

检测在所述用户注视的方向上说话的人的图像；

响应于图像的检测，生成与检测到的图像相对应的图像信号；以及

使用图像信号从所述多个声音信号中识别所述多个信号。

14.根据权利要求13所述的方法，其特征在于，所述识别还包括使用所述注视信号和所述图像信号来识别所述多个声音信号中的多个信号。

15.根据权利要求13所述的方法，其特征在于，所述识别还包括使用所述注视信号以及所述麦克风阵列的相对定位来识别所述多个声音信号中的多个信号。

16.根据权利要求13所述的方法，其特征在于，所述识别还包括使用所述图像信号以及所述麦克风阵列的相对定位和所述注视信号来识别所述多个声音信号中的多个信号。

17.根据权利要求11所述的方法，其特征在于，还包括选择性地覆盖所述注视信号。

18.根据权利要求17所述的方法，其特征在于，还包括在有限的时间段中选择性地覆盖所述注视信号。

19.根据权利要求11所述的方法，其特征在于，还包括输出指示来自用户的通信请求的通信信号。

20.根据权利要求11所述的方法，其特征在于，还包括从声源接收识别信号，所述识别信号包括识别声源本身或与声源相关联的人的元数据。

# 个性化的实时音频处理

[0001] 相关申请的交叉引用

[0002] 本申请要求于2016年2月2日提交的、题为“PERSONALIZED,REAL-TIME AUDIO PROCESSING(个性化的实时音频处理)”、号为15/013,537的美国实用专利申请的优先权和权益,该专利申请以其全部内容并入本文。

## 技术领域

[0003] 本申请总体涉及音频处理领域,并且尤其涉及声音检测和增强领域。

## 背景技术

[0004] 在许多环境中,能够在环境中听到并区分特定噪音或声音的能力很重要。例如,在拥挤的演讲厅中,倾听说话者/演讲者的听众可能会遇到来自各种来源的噪音。这些可能包括厅内的噪音源,诸如空调或其他通风设备、其他听众成员的手机、在说话者/演讲者进行演示期间人们进行交谈或翻动纸张等等。噪音也可能来自演讲厅外(例如交通、走廊声音、工作人员操作清洁设备等)。在这样的环境中,听者辨别或听到来自他人的所需声音的能力会减弱。另外,当说话者或听者在演讲厅中移动时,声音条件可能会改变。例如,如果听者的座位太靠近说话者或一组音频扬声器,则声音水平可能过高。备选地,如果听者离声源太远,则声音水平可能过低。其他嘈杂的环境,诸如聚会、嘈杂的公共街道和体育场也会带来听力挑战。

[0005] 噪音消除系统基于现有的预定标准来处理噪音。例如,一些系统(诸如利用自动增益控制的那些系统)利用参考水平并确定噪音信号是否高于该水平。如果噪音信号不高于参考水平,则将其视为不需要的噪音并将其去除。这样的系统没有考虑这样的因子,这些因子例如可以使得本来的低水平信号成为期望的信号。

## 附图说明

[0006] 在附图的图示中,通过示例而非限制性的方式示出了本公开,其中:

[0007] 图1是描绘其中实时音频处理设备可以起作用的示例性环境的示意图。

[0008] 图2是示出根据实施例的实时音频处理设备的示意图。

[0009] 图3是示出实时音频处理方法的实施例的流程图。

[0010] 图4是示出实时音频处理方法的实施例的流程图。

## 具体实施方式

[0011] 参考图1,根据一个实施例,实时音频处理设备200可以位于演讲厅或礼堂101中,其中听者可能希望专注于来自说话者/演讲者或其他声源102的声音。相比其他声音/噪音源106,实时音频处理设备200可能更靠近声源102,可以接收多个声音103,其中一些声音可能是期望的,并且其中一些声音可能是噪音。多个声音可以包括期望的声音104、来自说话者的某个附近区域内的噪音源106的噪音105,和/或任何其他类型的噪音。

[0012] 其他环境中的相对定位和声音水平可以代表与在演讲厅或礼堂中遇到的那些问题类似的问题。任何一种拥挤的房间中，如例如在其中两个人可能希望在一个充满其他对话和无关噪音的房间里进行交谈的聚会场所或接待处；在嘈杂的公共街道或拥挤的公共交通工具上；或者甚至是一间带小隔间的办公室中，也会呈现出听力挑战。在此处所列出的任何环境中，听者希望专注于的实际声源可能不是附近最响亮的声源。

[0013] 现在参考图2，实时音频处理设备200可以包括具有一个或多个麦克风202的麦克风阵列201。在一个实施例中，麦克风阵列201以及实时音频处理设备200可以集成到用户的耳机中。麦克风阵列201检测多个声音并将声音转换为多个信号。麦克风阵列201内的每个麦克风202的位置连同来自每个麦克风的信号的幅度可以提供关于方向性和声音水平的信息。然后，麦克风阵列201将多个信号输出到实时音频处理设备200的信号处理单元203。

[0014] 在下面的论述中，信号处理单元203将被称为数字信号处理单元或DSP单元。已知DSP芯片以高效处理声音的方式设计、配置和/或以其他方式实施。然而，本文描述的技术和概念还可以在宽范围的适当地编程的处理器或微处理器、控制器或微控制器，或其他计算或处理设备中实施。

[0015] 在一个实施例中，DSP单元203可以使用关于哪个信号来自麦克风阵列201中的哪个麦克风202的信息来确定声音方向性和相对水平。使用该信息与信号一起，DSP单元203可以确定多个信号中的每个信号的幅度、频率和到达方向分量。DSP单元203可以执行一个或多个测向算法以确定每个信号的到达方向分量。通过非限制性示例的方式，这种算法可以包括到达时间差、三角测量和粒子速度测量。其他算法对于本领域普通技术人员而言是已知的。在一个实施例中，DSP单元203还确定多个信号中的每个信号的持久性水平。在一个实施例中，DSP单元203可以确定信号的幅度、频率或持续时间或全部三者的持久性。以这种方式，实时音频处理设备200在一方面可以在瞬态、突然和/或随机的声音与例如语音(其在水平和/或持续时间方面倾向于具有一定程度的持久性)之间进行区分。

[0016] 再次参考图2，实时音频处理设备200可以包括注视检测传感器204，以在实时音频处理设备的操作期间确定用户的注视方向。注视检测传感器204可以是陀螺仪传感器、加速度计或能够感测旋转运动或角速度变化的任何传感器装置。在这种实施方式中，基于注视检测传感器204相对于用户视线的位置，用户头部的运动和/或位置可以是用户的实际注视的替代物。

[0017] 在一个实施例中，注视检测传感器204可以使用类似于例如谷歌眼镜(Google Glass™)的技术来检测用户眼睛的实际位置。也可以使用开始出现在飞行员佩戴的头盔或护目镜中的技术，这些技术使得飞行员能够移动其眼睛来控制飞行器操作的各个方面，而不必将手从控制装置移开。在另一个实施例中，注视检测传感器204可以使用放置在用户头部或身体上时的麦克风阵列201中的麦克风的相对位置以及阵列的运动来获得关于用户注视方向的信息。然而，这种方法可能不太可靠，因为听者可能专注于的期望声源可能不是最响亮的声源。

[0018] 在一个实施例中，DSP单元203可以使用用户的注视来确定要强化或弱化哪个或哪些声音。在存在多个声源的情况下，例如当在正在播放音乐或者在附近有其他人对话的喧嚣聚会上与某人交谈时，可能的情况是：期望的声音(例如，与用户交谈的说话者的声音)可能不是听者所处位置处的最响亮的声源。因此，可能希望能够强化来自说话者方向的声音，

并且弱化来自其他方向的声音。现在将论述该方法的实施例的各方面。

[0019] 当用户的注视(或头部或身体位置)追踪说话者的方向时,注视检测传感器204可以生成与用户注视方向相对应的信号。DSP单元203可通过相对于来自说话者方向的声音比较麦克风阵列201或注视检测传感器204的运动来确认用户注视方向对应于说话者位置的方向。

[0020] 根据一个实施例,DSP单元203可以实施本领域普通技术人员已知的许多算法中的任何算法,用于检测和区分不需要的噪音以补充实时音频处理设备的声音处理能力。仅通过示例的方式,实时音频处理设备200可以利用声学波束形成技术,其可以作为单独的模块或算法,或者如图2所示作为DSP单元203内的模块205。使用声学波束形成是强化来自一个方向的声音同时弱化来自其他方向的声音的另一种方式。

[0021] 在一个实施例中,当DSP单元203已经确认用户的注视正在追踪说话者或其他期望的声源时,DSP单元203可以确定用户的注视是否正在继续追踪。例如,使用加速度计或其他运动传感器,DSP单元203可以确定头部运动的变化,和/或在相同或不同的一个或多个给定时间间隔内头部运动速度的变化。超过特定的(阈值)运动量可以指示专注点从说话者改变到另一个源。超过特定的(阈值)运动速度可以指示头部剧烈运动,而不是专注点的改变。取决于期望的效果,可以指示DSP单元203注意或忽略注视检测传感器204提示的注视方向的偏移。因此,例如,如果用户快速或缓慢地转动以倾听另一个说话者(例如,在演讲时坐在用户旁边的人),并且用户的注视在该另一个说话者的方向上保持足够长的时间,则DSP单元203可以确定用户已经改变了他/她的注视,并且可以确定用户可能专注于附近的说话者。

[0022] DSP单元203可以将用户注视方向与声音到达方向进行比较,该声音到达方向如由多个接收信号中的相应信号所表示。DSP单元203还可以将多个接收信号中的每个信号的幅度彼此比较以便预设幅度参考阈值。然后,DSP单元203可以确定来自多个接收信号中的哪些信号具有与用户注视方向匹配的到达方向分量,以及哪些信号具有满足或超过预设幅度阈值的幅度。具有与用户注视方向匹配的到达方向分量以及满足或超过幅度阈值的幅度的信号可以被认为表示期望的声音,而不具有匹配的到达方向分量或者不具有足够持久性的信号可被视为代表不需要的噪音。如果没有检测到具有与检测到的用户注视方向相对应的到达方向分量以及满足或超过幅度一致性阈值的幅度的信号,则实时音频处理设备将重新启动并尝试再次确定期望的声音。在这样的实施例中,与最强调用户注视的其他实施例相比可以向声音幅度赋予更多的重要性。

[0023] 在一个实施例中,DSP单元203可以将多个信号中的每个信号的持久性与持久性阈值进行比较。持久性可以是振幅(幅度)和持续时间的函数。不满足持久性阈值的信号可以被认为是不需要的噪音。上面论述的因子,包括到达方向、幅度和持久性值仅仅是示例性的,并且不应该被解释为限制DSP单元203在确定期望声音时可以考虑的因子的数量。

[0024] 在一个实施例中,DSP单元203可以将权重分配给到达方向、幅度和持久性因子。然后,DSP单元203可以获取多个信号中的每个信号的所有因子的加权和,并将该和与加权参考因子进行比较。如果多个信号中的任何信号的所有因子的加权和满足或超过加权参考因子,则DSP单元203可以确定该信号是期望的声音信号。不具有满足或超过相应加权参考因子的加权和的信号可以被认为是不需要的噪音。

[0025] 在确定多个接收信号中的哪个信号表示期望声音之后,DSP单元203可以去除(消

除)或弱化(削弱)表示不需要的噪音的任何信号。仅通过示例的方式,DSP单元203可以通过捕获不需要的噪音的频率分布、反转该频率分布并使其偏移来去除表示不需要的噪音的信号,这提供了与不需要的噪音信号相反的信号。然后输出相反的噪音信号以有效地消除不需要的噪音。在一个实施例中,DSP单元203可以为期望的声音提供额外的放大,并且相应地对不需要的噪音提供较少的放大,从而与不需要的噪音相比强化期望的声音。

[0026] 在一个实施例中,如图2中所示,实时音频处理设备可以包括成像传感器207以查找说话者的面部。注视检测传感器204使用用户注视方向来确定用户是否正在观看期望的声源,该声源可能不是最响亮的声源,但仍然是期望的。成像传感器207使用说话者的注视方向,并且可以与麦克风阵列201中的麦克风202一起工作,以确定最响亮的期望声源来自用户希望倾听的说话者。这样的实施例可能是有用的,其中听者简单地转动他/她的头部例如以对演讲者的话做笔记。在一个实施例中,成像传感器207可以与注视检测传感器204一起工作以确认声源方向,如例如当用户和说话者看着彼此时。

[0027] 有时可能需要覆盖(override)注视检测传感器204。例如,在嘈杂的环境中,诸如拥挤的礼堂或运动的地铁车厢中,锁定说话者以作为所需的声源可能是有用的,即使有时候用户将需要移开视线。例如,如前所述,在礼堂中,听者可能希望在演讲者说话时查看他/她的笔记本、计算机或平板电脑,或查看一些文本源。例如,在运动的地铁车厢中,用户可能需要查看在车厢中张贴的地图。在这种情况下转移视线可能会改变声源。但是如果用户可以覆盖注视检测传感器,则声源可以保持原样。覆盖可以是暂时的(例如,在短时间段例如5秒至10秒之后改变),或者它可以是永久性的,直到用户关闭覆盖为止。例如,可以通过覆盖开关208来实现这种覆盖,该覆盖开关208虽然在DSP单元203内示出,但是其可以是机载麦克风阵列201,或者与注视检测传感器204相关联,以方便用户。

[0028] 成像传感器207可以是视频传感器或红外传感器或具有类似功能的任何传感器设备。在一个实施例中,可以使用在数码相机和其他应用中使用的面部识别技术,其程度使得例如这种技术有助于观看说话者的眼睛。成像传感器207可以将与接收到的图像对应的信号输出到DSP单元203。DSP单元203可以监视成像传感器207的接收信号,并且可以利用包括面部识别或热识别等的算法来确定从成像传感器207接收的一个或多个信号是否对应于在用户的方向上说话的人的图像。在检测到有人正在朝向用户的方向说话时,DSP单元203可以强化来自该方向的信号,并且弱化或消除来自不同方向(例如,离开在说话者和听者之间延伸的确定的虚拟中心线超过一定度数的方向)的信号。然后,实时音频处理设备可以通过将信号的音量和频率参数修改为适当的音量和频率参数来进一步将来自说话者方向的信号处理至最优水平,如前所述。

[0029] 例如,在嘈杂的房间中可能存在这样的情况,其中两个需要交谈的人彼此在房间对面,而在他们之间的区域中有多组对话。如果两人看着彼此的眼睛,那么成像传感器207和/或注视检测传感器204可以将DSP单元203锁定到房间对面的人的语音,强化该声音以排除其他声音,使得两个人可以交谈,即使他们不是彼此相邻。在这种情景中,一个人可能希望抓住房间对面另一个人的目光。为了便于完成该操作,寻求注意的人可以对房间对面的人进行打招呼操作(ping)以提醒他/她需要注意和/或对话。在接收到打招呼信息之后,房间对面的人可以找到寻求专注的人并且进行眼神接触。在该情况下,实时音频处理设备200可以按如上所述方式起作用。如果房间对面的人希望忽略该请求,则他/她当然可以这样

做。

[0030] 打招呼操作可以按任何数量的方式进行，从设备(诸如各个人佩戴/使用的多个实时音频处理设备200)之间的信号广播到使用智能手机或其他手持设备等上的小应用发送信号都可以。

[0031] 在一个实施例中，用户可能希望向房间、大堂、礼堂等中的多个人进行通信，以表明他/她想要进行关于某事物的对话。为了实现该通信，用户可以能够向该区域中的人广播对该用户或该用户的声音源而言唯一的信号。在一个实施例中，可类似于实时音频处理设备200的其他用户的设备可以接收信号，并且可以通过锁定到该信号并因此锁定到发起者的声音/语音来进行响应。同样在一个实施例中，用户可以能够通过智能手机或平板电脑小应用访问各种可能的附近本地说话者的列表，并且可以选择那些说话者中的一个。在该情况下，小应用可以将信息传送到其相应的设备，从而锁定该信号。

[0032] 根据一个实施例，成像传感器207可以用作注视检测传感器204。在另一个实施例中，特别是在礼堂等中，成像传感器207可以通过辅助DSP单元203处理来自说话者的信号来补充注视检测传感器，即使在用户没有朝说话者方向看的情况下也是如此。DSP单元可以使注视检测传感器204优先于成像传感器207。以这种方式，当用户正在看演讲者时，例如，注视检测传感器204可以主要与DSP单元203一起工作以确保演讲者的语音被适当地处理。例如，当用户的注视向下指向以便记笔记时，可以给予成像传感器207优先权。备选地，可以认识到，当用户向下看以便记笔记时，注视检测传感器204将表明用户没有在声音的来源方向上看。在该情况下，DSP单元203可以忽略来自注视检测传感器204的输出。这可以是实现上述功能的一种方式。

[0033] 根据一个实施例，DSP单元203可以连接到音乐或视频播放器。在确定有人在用户的方向上说话时，DSP单元203可以被配置为自动暂停来自音乐或视频播放器的回放。以这种方式，用户/听者可以选择听与他/她交谈的人说话，而不必笨拙地寻找回放控制件。

[0034] 根据一个实施例，如果用户/听者在使用音乐或视频播放器的同时正在看说话者，则可以使实时音频处理设备确定用户/听者需要听到说话者而不是从音乐或视频播放器回放。设备可以响应于音频命令，或者可以使用DSP单元203中的注视检测传感器204以便指示设备中止或暂停回放，从而用户/听者可以听到说话者。

[0035] 在一个实施例中，可能希望能够识别声源而无需用户直接看它。例如，在交响乐队中，通常存在多个不同的弦乐器(例如小提琴、中提琴、大提琴、低音提琴)和/或多个木管乐器(例如双簧管、巴松管)和/或多个铜管乐器(例如小号、长号、萨克斯管、法国号、大号)和/或多个打击乐器(例如三角铁、低音鼓、定音鼓、军鼓)和/或键盘乐器(例如钢琴、风琴、大键琴、电子键盘、键盘)。这些类别中的列表并不意旨是广泛性的，而仅仅是用于例示不同乐器的范围。行进乐队可能有相似的乐器范围。

[0036] 在其中一个刚刚描述的情景中，听者可能需要专注于第一小提琴手，或者第二中提琴手(例如，在管弦乐队中进行演奏的朋友或亲戚不是更容易通过注视来辨认的主奏者或第一小提琴手的情况下)。或者，可能存在不属于这些的独奏者。在这样的场景中，每个表演者可以具有与他/她相关联的元数据，然后可以用表演者正在演奏的乐器来识别。例如，如果DSP单元具有包含该元数据的表格(其可以在演奏之前下载，可以在音乐会上或者预先通过小应用来下载)，那么在音乐会期间，用户可以输入指令以专注于与一个或多个乐器的

特定元数据相关联的一个或多个乐器，以便突出来自这些乐器的声音。

[0037] 现在参考图3，用于实时音频检测和处理的方法开始于301，此时实时音频处理设备接收多个声音并将它们处理成多个信号。在302处，确定用户注视方向。在303处，DSP单元203使用检测到的注视来识别所接收到的多个信号中的哪个(哪些)信号对应于期望的声音以及多个信号中的哪个(哪些)信号对应于不需要的噪音。该过程可循环执行，如图3中的304处所示，直到成功识别出期望的声音。在305处，使用上面论述的一个或多个算法去除或弱化不需要的噪音信号。在306处，处理期望的声音以使它们与预设的音量和频率参数一致。最后，在307处，在用户佩戴其中包含实时音频处理设备的音频回放设备的实施例中，可以暂停回放以便于用户听到说话者。步骤301至307中的每一个可以根据以上关于图2描述的任何实施例和技术来实施。

[0038] 图4表示图3的流程和随附描述的变型，考虑根据一个实施例结合注视检测传感器204使用成像传感器207。在401处，实时音频处理设备接收多个声音并将它们处理成多个信号。在402处，确定说话者的注视方向。在403处，DSP单元203使用检测到的说话者的注视来识别所接收到的多个信号中的哪个(哪些)信号对应于期望的声音以及多个信号中的哪个(哪些)信号对应于不需要的噪音。在404处，如果不存在匹配，则可以采取进一步的动作来确定是否应该使得成像传感器207优先于注视检测传感器204。如前所述，这样的一种情况可能发生在当听者/用户向下看例如以便记笔记时。如果用户没有向下看，则该过程可以循环执行，如图4中所示，直到成功识别出期望的声音。如果用户向下看，则可以沿着前面论述的方案给予成像传感器207或麦克风阵列201优先权。

[0039] 在一个实施例中，在405处，可以进行处理以确定正在实施的噪音消除算法是否成功工作，例如，以便查看是否成功识别出不来自说话者或所需声源的声音。如果不是，则在406处，该过程可以返回到起始处，如图4中所示，或返到流程图中的另一个地方，例如402或403。如果存在足够的相关性，即如果正确识别出不需要的声音，则在407处，使用上面论述的一种或多种算法来去除或弱化不需要的噪音信号。在408处，处理期望的声音以使它们与预设的音量和频率参数一致。步骤401至408中的每一个可以根据以上关于图2描述的任何实施例和技术来实施。

[0040] 在启动实时音频处理设备时，以上关于图3和图4论述的元素可以自动启动。备选地，设备可以响应合适的用户语音命令。用于响应用户语音命令的技术对于本领域普通技术人员而言将是已知的。

[0041] 所公开的实施例不将其适用性限于音乐厅、礼堂或演讲厅，或不将其适用性限于演讲者或说话者。在聚会或拥挤的房间中，在听者附近可能存在多个说话者，其中一些说话者甚至比用户可能希望倾听的说话者声音更大，看着具体说话者的能力可以有助于强化说话者的声音，并消除或弱化他人的声音或来自不同方向的声音。本发明实施例可以用于体育赛事、地铁、公共街道、餐馆，或者在声音的定向接收将有帮助的任何环境中。

[0042] 尽管已经参考若干示例性实施例描述了本公开，但是应该理解的是，已经使用的词语是描述和说明性的词语，而不是限制性的词语。在不脱离本公开的各方面的范围和精神的情况下，可以在所附权利要求的范围内进行改变，如现在所陈述和修改的那样。尽管已经参考特定装置、材料和实施例描述了本公开，但是本公开并非旨在限于所公开的细节；相反，本公开扩展到所有功能上等同的结构、方法和用途，诸如在所附权利要求的范围内的那

样。

[0043] 就本文中的方法或装置实施例中被描述为具有特定数量的元素而言，应该理解的是，定义完整的权利要求可以只需要少于所有元素。另外，各种实施例中描述的操作或功能顺序在实施任何所附权利要求时不要求或暗示要求这些顺序。可以按任何顺序执行操作或功能以实现所公开实施例的目标。例如，关于图3中的操作就是这种情况，其中，例如，可以首先处理用户的注视方向，而不是如图所示处理声音，因为可能优选的是首先确定用户的注视是否已经聚焦，然后继续处理所接收到的声音。

[0044] 提供的公开摘要是为了符合37C.F.R.§1.72(b)，其要求摘要将允许读者快速地确定技术公开的性质。应理解是摘要将不会用于解释或限制权利要求的范围或含义。另外，在前面的具体实施方式中，出于简化本公开的目的，各种特征可以组合在一起或在单个实施例中描述。本公开不应被解释为反映所要求保护的实施例需要比每个权利要求中明确记载的特征更多特征的意图。而是，如以下权利要求所反映的那样，主题可以涉及的特征少于任何所公开实施例的所有特征。因此，以下权利要求被并入具体实施方式中，其中每项权利要求自身作为限定单独要求保护的主题。

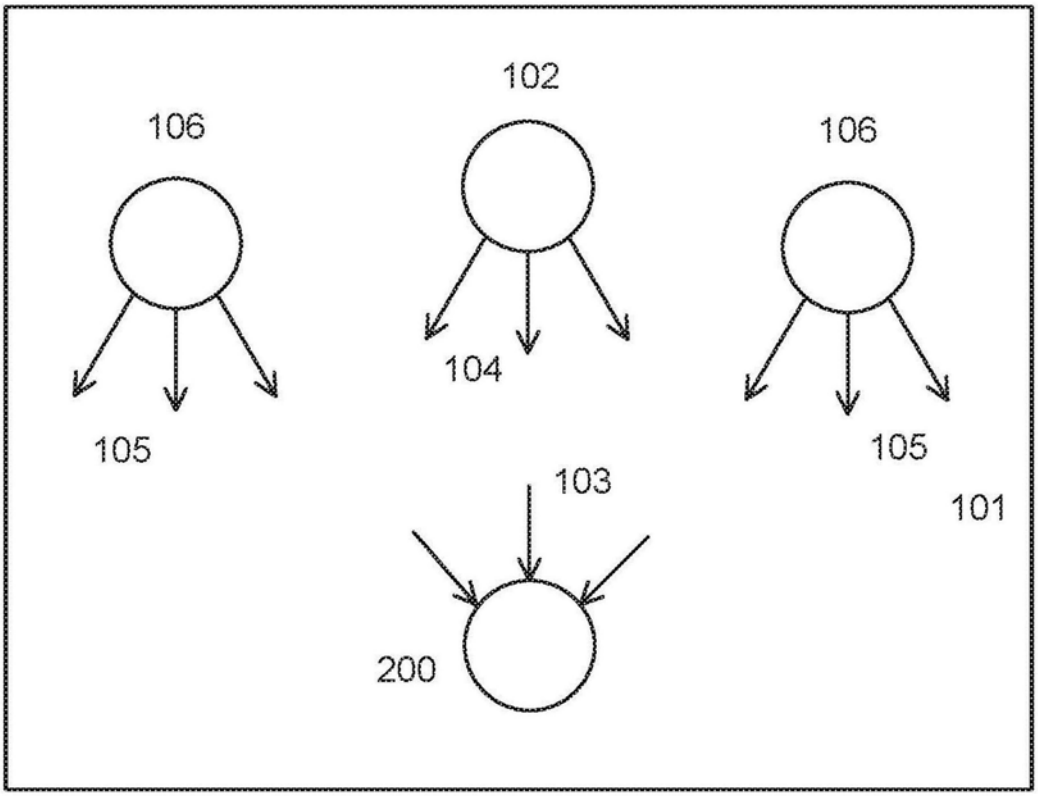

102
106
106
104
105
105
103
101
200

图1

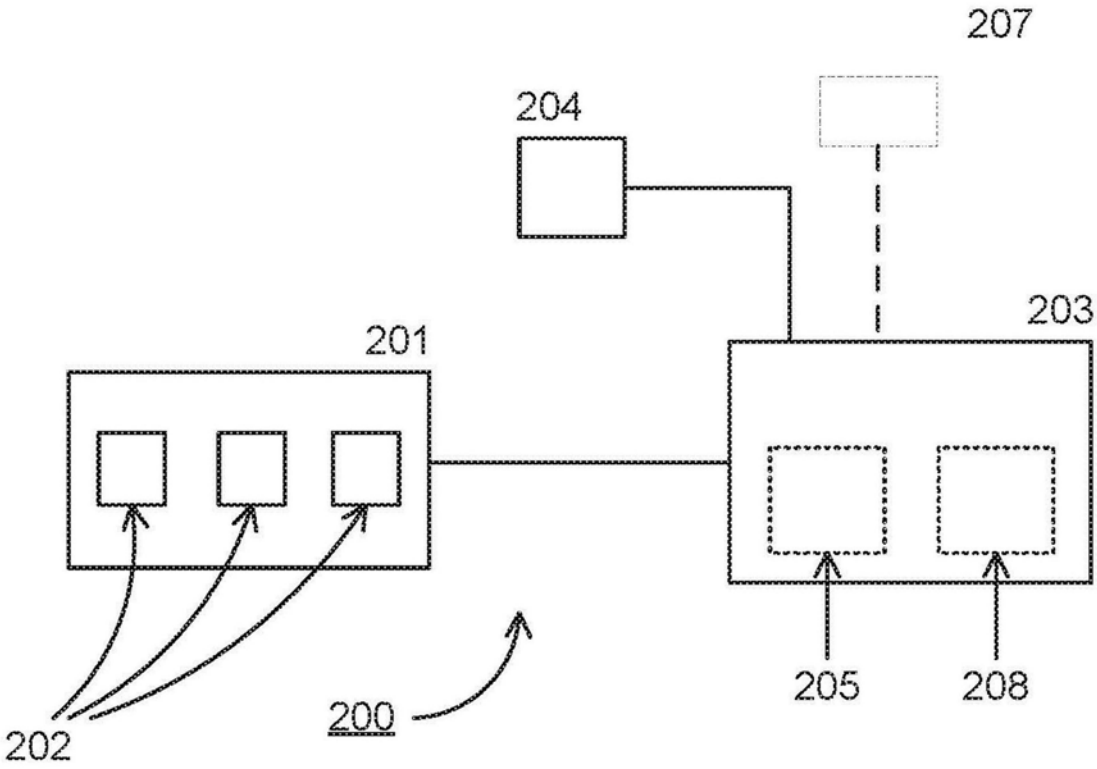

207
204
203
201
205
208
200
202

图2

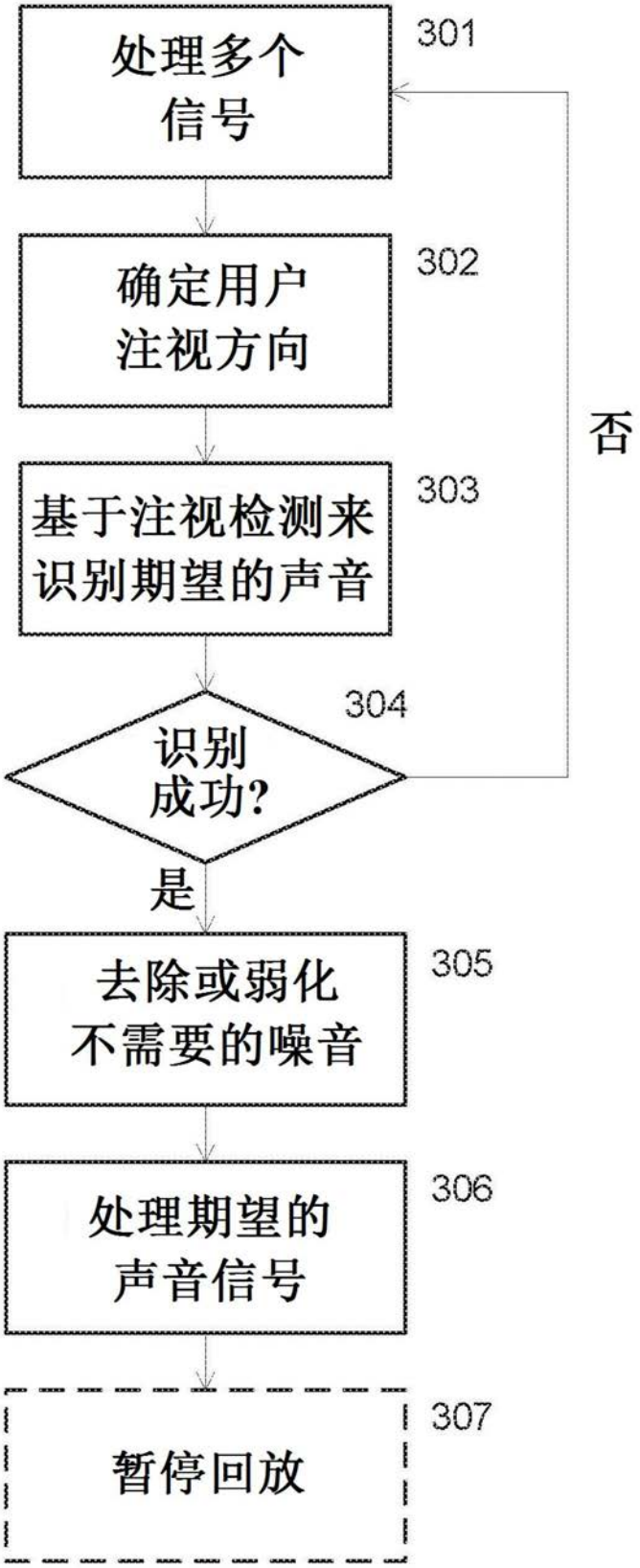

处理多个信号
301
确定用户注视方向
302
基于注视检测来识别期望的声音
303
识别成功?
304
否
是
去除或弱化不需要的噪音
305
处理期望的声音信号
306
暂停回放
307

图3

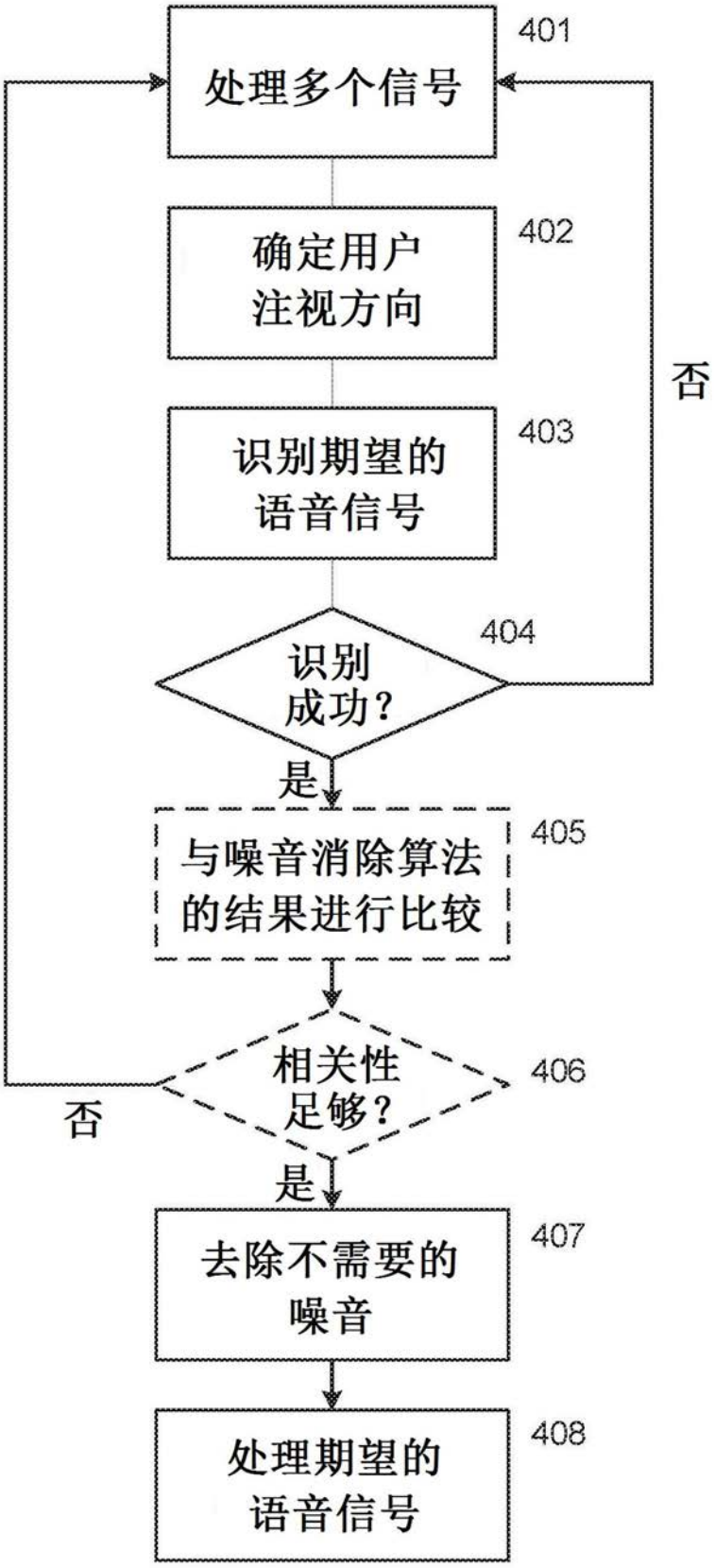

处理多个信号
401
确定用户
注视方向
402
识别期望的
语音信号
403
识别
成功？
404
否
是
与噪音消除算法
的结果进行比较
405
相关性
足够？
406
否
是
去除不需要的
噪音
407
处理期望的
语音信号
408

图4