(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 113113034 A
(43)申请公布日 2021.07.13

(21)申请号 202110023469.7

(22)申请日 2021.01.08

(30)优先权数据
16/740297 2020.01.10 US

(71)申请人 辛纳普蒂克斯公司
地址 美国加利福尼亚州

(72)发明人 A·马斯纳迪-施拉兹 F·内斯塔

(74)专利代理机构 中国专利代理(香港)有限公司 72001
代理人 冯夏雨 陈岚

(51)Int.Cl.
G10L 21/02(2013.01)
G10L 21/0216(2013.01)
G10L 25/87(2013.01)
G01S 3/80(2006.01)

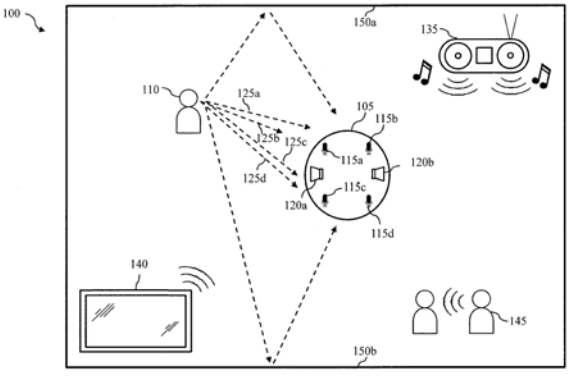
权利要求书3页 说明书10页 附图7页

(54)发明名称

用于平面麦克风阵列的多源跟踪和语音活动检测

(57)摘要

本文中描述的实施例提供了组合的多源到达时间差(TDOA)跟踪和语音活动检测(VAD)机制,其可适用于通用阵列几何形状,例如,位于平面上的麦克风阵列。组合的多源TDOA跟踪和VAD机制扫描麦克风对中的麦克风阵列的方位角和仰角角度,基于此可以在多个麦克风对的多维TDOA空间中形成物理上可允许的TDOA的平面轨迹。以此方式,多维TDOA跟踪通过分离地针对每个维度执行TDOA搜索而减少了传统TDOA中通常涉及的计算的数量。



CN 113113034 A

1.一种用于通过多源跟踪和语音活动检测来增强多源音频的方法,包括:

经由音频输入电路从麦克风阵列接收一个或多个多源音频信号;

根据导向最小方差(STMV)波束形成器、基于在由来自所述麦克风阵列的多个麦克风对定义的多维空间中构造的TDOA轨迹信息,针对所述一个或多个多源音频信号计算TDOA检测数据;

基于直到当前时间-步长的所述计算的TDOA检测数据来更新多个音频轨道;

基于所述计算的TDOA检测数据,针对所述多个音频轨道中的每个构造语音活动检测(VAD)数据;以及

使用所述更新的多个音频轨道和所述构造的VAD数据生成一个或多个增强的多源音频信号。

2.根据权利要求1所述的方法,其中由来自所述麦克风阵列的多个麦克风对定义的所述多维空间通过以下步骤形成:

从所述麦克风阵列选择第一麦克风作为参考麦克风;以及

将来自所述麦克风阵列的每个剩余的麦克风与所述参考麦克风配对。

3.根据权利要求2所述的方法,其中所述TDOA轨迹信息在启动阶段基于所述多个麦克风对的空间信息通过以下步骤计算一次:

针对每个麦克风对,基于相应对中的两个麦克风之间的距离和角度,计算与入射声线的特定方位角角度和特定仰角角度对应的TDOA位置;以及

通过变化所有麦克风对上的所述入射声线的所述特定方位角角度和所述特定仰角角度来形成TDOA位置点的网格。

4.根据权利要求3所述的方法,其中当所述麦克风阵列物理地位于现实中的第二平面上时,所述TDOA位置点的网格位于所述多维空间中的第一平面上,所述多维空间具有等于麦克风对的总数量的维度数量。

5.根据权利要求2所述的方法,其中针对所述一个或多个多源音频信号计算TDOA检测数据还包括:

针对每个麦克风对:

针对每一个频带使用来自所述相应麦克风对的所述一个或多个多源音频信号的时频表示来计算协方差矩阵;

基于TDOA位置针对每一个频带构造导向矩阵,所述TDOA位置针对对应于所述相应麦克风对的方位角和仰角角度的不同扫描;

基于所述构造的导向矩阵和所述计算的协方差矩阵,构造跨所有频带对齐的方向协方差矩阵;以及

基于所述构造的方向协方差矩阵,确定使波束功率最小化的伪似然解。

6.根据权利要求5所述的方法,还包括:

通过取跨所有麦克风对的所有确定的伪似然解的乘积来计算STMV联合伪似然度;

确定使所述STMV联合伪似然度最大化的一对方位角和仰角角度;以及

将所述确定的一对方位角和仰角角度转换为代表所述TDOA检测数据的极坐标表示。

7.根据权利要求6所述的方法,其中基于所述构造的导向矩阵和所述计算的协方差矩阵来构造跨所有频带对齐的所述方向协方差矩阵在所有麦克风对以及方位角和仰角角度

的所有扫描上被重复。

8.根据权利要求6所述的方法,其中基于所述构造的导向矩阵和所述计算的协方差矩阵来构造跨所有频带对齐的所述方向协方差矩阵以减少的重复通过以下步骤执行:

将所述多维空间划分为多个段,其中段的数量小于所述多维空间的维度的总数量;

将每个TDOA位置点从TDOA位置点的网格映射到最接近的段;以及

使用所述段的数量以及所述TDOA位置点的网格与所述段的数量之间的映射关系而不是根据方位角和仰角角度的所有扫描建立的所述TDOA位置点的网格来计算所述方向协方差矩阵。

9.根据权利要求1所述的方法,其中基于直到所述当前时间-步长的所述计算的TDOA检测数据来更新所述多个音频轨道还包括:

识别对应于当前时间-步长的TDOA检测和先前已经建立直到所述当前时间-步长的一组现有音频轨道;以及

基于所述TDOA检测与所述现有音频轨道的门之间的比较来确定是将所述TDOA检测并入到所述现有音频轨道中的一个中还是建立新的音频轨道。

10.根据权利要求1所述的方法,其中基于所述计算的TDOA检测数据针对所述多个音频轨道中的每个构造VAD数据还包括:

当相应音频轨道最接近于所述TDOA检测时,将第一值分配给所述相应音频轨道;以及

将第二值分配给其他音频轨道。

11.一种用于通过多源跟踪和语音活动检测来增强多源音频的音频处理设备,包括:

音频输入电路,其被配置成从麦克风阵列接收一个或多个多源音频信号;

到达时间差(TDOA)估计器,其被配置成根据导向最小方差(STMV)波束形成器、基于在由来自所述麦克风阵列的多个麦克风对定义的多维空间中构造的TDOA轨迹信息,针对所述一个或多个多源音频信号计算TDOA检测数据;

多源音频跟踪器,其被配置成基于直到当前时间-步长的所述计算的TDOA检测数据来更新多个音频轨道,以及基于所述计算的TDOA检测数据,针对所述多个音频轨道中的每个构造语音活动检测(VAD)数据;以及

音频增强引擎,其被配置成使用所述更新的多个音频轨道和所述构造的VAD数据生成一个或多个增强的多源音频信号。

12.根据权利要求11所述的音频处理设备,其中由来自所述麦克风阵列的多个麦克风对定义的所述多维空间通过以下步骤形成:

从所述麦克风阵列选择第一麦克风作为参考麦克风;以及

将来自所述麦克风阵列的每个剩余的麦克风与所述参考麦克风配对。

13.根据权利要求12所述的音频处理设备,其中所述TDOA轨迹信息在启动阶段基于所述多个麦克风对的空间信息通过以下步骤计算一次:

针对每个麦克风对,基于相应对中的两个麦克风之间的距离和角度,计算与入射声线的特定方位角角度和特定仰角角度对应的TDOA位置;以及

通过变化所有麦克风对上的所述入射声线的所述特定方位角角度和所述特定仰角角度来形成TDOA位置点的网格。

14.根据权利要求13所述的音频处理设备,其中当所述麦克风阵列物理地位于现实中

的第二平面上时，所述TDOA位置点的网格位于所述多维空间中的第一平面上，所述多维空间具有等于麦克风对的总数量的维度数量。

15.根据权利要求12所述的音频处理设备，其中所述TDOA估计器被配置成通过以下步骤来计算所述TDOA检测数据：

针对每个麦克风对：

针对每一个频带使用来自所述相应麦克风对的所述一个或多个多源音频信号的时频表示来计算协方差矩阵；

基于TDOA位置针对每一个频带构造导向矩阵，所述TDOA位置针对对应于所述相应麦克风对的方位角和仰角角度的不同扫描；

基于所述构造的导向矩阵和所述计算的协方差矩阵，构造跨所有频带对齐的方向协方差矩阵；以及

基于所述构造的方向协方差矩阵，确定使波束功率最小化的伪似然解。

16.根据权利要求15所述的音频处理设备，其中所述TDOA估计器还被配置成通过以下步骤计算所述TDOA检测数据：

通过取跨所有麦克风对的所有确定的伪似然解的乘积来计算STMV联合伪似然度；

确定使所述STMV联合伪似然度最大化的一对方位角和仰角角度；以及

将所述确定的一对方位角和仰角角度转换为代表所述TDOA检测数据的极坐标表示。

17.根据权利要求16所述的音频处理设备，其中所述TDOA估计器还被配置成基于所述构造的导向矩阵和所述计算的协方差矩阵，在所有麦克风对以及方位角和仰角角度的所有扫描上重复构造跨所有频带对齐的所述方向协方差矩阵的操作。

18.根据权利要求16所述的音频处理设备，其中所述TDOA估计器还被配置成通过以下步骤以减少的重复基于所述构造的导向矩阵和所述计算的协方差矩阵来构造跨所有频带对齐的所述方向协方差矩阵：

将所述多维空间划分为多个段，其中段的数量小于所述多维空间的维度的总数量；

将每个TDOA位置点从TDOA位置点的网格映射到最接近的段；以及

使用所述段的数量以及所述TDOA位置点的网格与所述段的数量之间的映射关系而不是根据方位角和仰角角度的所有扫描建立的所述TDOA位置点的网格来计算所述方向协方差矩阵。

19.根据权利要求11所述的音频处理设备，其中所述多源音频跟踪器被配置成通过以下步骤基于直到所述当前时间-步长的所述计算的TDOA检测数据来更新所述多个音频轨道：

识别对应于当前时间-步长的TDOA检测和先前已经建立直到所述当前时间-步长的一组现有音频轨道；以及

基于所述TDOA检测与所述现有音频轨道的门之间的比较来确定是将所述TDOA检测并入到所述现有音频轨道中的一个中还是建立新的音频轨道。

20.根据权利要求11所述的音频处理设备，其中所述多源音频跟踪器被配置成通过以下步骤基于所述计算的TDOA检测针对所述多个音频轨道中的每个构造VAD数据：

当相应音频轨道最接近于所述TDOA检测时，将第一值分配给所述相应音频轨道；以及

将第二值分配给其他音频轨道。

用于平面麦克风阵列的多源跟踪和语音活动检测

技术领域

[0001] 根据一个或多个实施例，本公开总体上涉及音频信号处理，并且更特别地例如，涉及用于通用平面麦克风阵列的多源跟踪和多流语音活动检测的系统和方法。

背景技术

[0002] 近年来，智能扬声器和其他语音控制的设备和装置已经获得普及。智能扬声器通常包括用于从环境接收音频输入(例如，用户的口头命令)的麦克风阵列。当在音频输入中检测到目标音频(例如，口头命令)时，智能扬声器可以将检测到的目标音频转化为一个或多个命令，并基于该命令执行不同的任务。

[0003] 这些智能扬声器的一个挑战是要在操作环境中高效地以及有效地将目标音频(例如，口头命令)与噪声或其他激活的扬声器隔离。例如，在存在一个或多个噪声源的情况下，一个或多个扬声器可以是激活的。当目标是增强特定的扬声器时，该扬声器被称为目标扬声器，而其余扬声器可以被视为干扰源。现有的话音增强算法大多使用多个输入通道(麦克风)来利用源的空间信息，诸如与独立成分分析(ICA)相关的盲源分离(BSS)方法以及空间滤波或波束形成方法。

[0004] 然而，BSS方法主要是为批处理而设计的，其由于大的响应延迟而在真实应用中可能通常是不合意的或甚至不适用的。另一方面，空间滤波或波束形成方法通常要求在作为要被最小化的成本函数的语音活动检测(VAD)下进行监督，这可能过度依赖于与仅噪声/干扰段有关的协方差矩阵的估计。

[0005] 因此，存在对用于在多流音频环境中检测和处理(一个或多个)目标音频信号的改进的系统和方法的需要。

附图说明

[0006] 参考以下附图和后面的具体实施方式，可以更好地理解本公开的各方面及其优点。应当领会，相同的参考标号用于标识在一个或多个附图中图示的相同的元件，其中在附图中的示图是为了图示本公开的实施例的目的而不是为了限制本公开的实施例的目的。附图中的部件不一定是按比例的，而是将重点放在清楚地图示本公开的原理上。

[0007] 图1图示了根据本公开的一个或多个实施例的用于音频处理设备的示例操作环境。

[0008] 图2是根据本公开的一个或多个实施例的示例音频处理设备的框图。

[0009] 图3是根据本公开的一个或多个实施例的用于多轨道音频增强的示例音频信号处理器的框图。

[0010] 图4是根据本公开的各种实施例的用于处理来自通用麦克风阵列的多个音频信号的示例多轨道活动检测引擎的框图。

[0011] 图5A是图示根据本公开的一个或多个实施例的麦克风对的示例几何形状的图。

[0012] 图5B是图示根据本公开的一个或多个实施例的多维空间中的与不同的麦克风阵

列几何形状对应的到达时间差(TDOA)轨迹信息的示例网格的图。

[0013] 图6是根据本公开的各种实施例的用于通过多源跟踪和活动检测来增强多源音频信号的示例方法的逻辑流程图。

[0014] 图7是根据本公开的各种实施例的用于使用麦克风对在多维空间中计算TDOA轨迹信息的示例过程的逻辑流程图。

具体实施方式

[0015] 本公开提供了用于在多流音频环境中检测和处理(一个或多个)目标音频信号的改进的系统和方法。

[0016] 语音活动检测(VAD)可用于在利用从多个输入通道得到的源的空间信息的过程中监督目标音频的话音增强。VAD可以允许在期望的扬声器的无声时段期间诱发干扰/噪声源的空间统计,使得当期望的扬声器变得激活时,噪声/干扰的影响随后可以被消除。例如,通过以下操作可以推导每个源的VAD来以到达时间差(TDOA)或到达方向(DOA)的形式跟踪源的空间信息:通过确定检测何时出现在现有轨道附近来构造VAD和利用检测的历史。此过程通常称为测量到轨道(M2T)分配。这样,可以针对所有感兴趣的源推导多个VAD。

[0017] 具体地,现有的DOA方法通常基于方位角和仰角角度的闭合形式映射来为整个麦克风阵列构造单个导向矢量,其可以用来利用线性或圆形阵列的特殊几何形状。这样的DOA方法无法扩展到麦克风阵列的通用或任意几何形状。另外,这些基于闭合形式映射的DOA方法通常需要在多维空间中的广泛搜索。对于任意几何形状,可以使用现有的基于TDOA的方法,该方法可能不限于特定的阵列几何形状,并且可以为每个麦克风对构造多个导向矢量,以形成多维TDOA矢量(每对一个维度)。然而,这些现有方法承担引入由来自每个TDOA对的谱的峰的交叉相交而形成的TDOA重影的风险。因此,通常需要涉及特定阵列几何形状的进一步后处理来去除TDOA重影。

[0018] 鉴于对不受特别阵列几何形状约束的多流VAD的需要,本文中描述的实施例提供了可适用于通用阵列几何形状(例如,位于平面上的麦克风阵列)的组合的多源TDOA跟踪和VAD机制。通过对每个维度分离地执行TDOA搜索,组合的多源TDOA跟踪和VAD机制可以减少传统TDOA中通常涉及的计算的数量。

[0019] 在一些实施例中,采用了用于位于平面上的通用阵列几何形状的多维TDOA方法,该方法避免了不想要的重影TDOA。在一个实施例中,获得通用地配置的麦克风的笛卡尔坐标,可以选择该麦克风中的一个作为参考麦克风。可以扫描麦克风的方位角和仰角角度,基于此可以在多个麦克风对的多维TDOA空间中形成物理上可允许的TDOA的平面轨迹。这样,形成的平面轨迹避免了重影TDOA的形成,因此不需要进一步的后处理来去除重影TDOA。而且,与全DOA扫描方法相比,本文中公开的多维TDOA方法通过分离地执行在与每个维度有关的成对TDOA域中的搜索、而不是在全多维空间中的搜索,来降低计算复杂度。

[0020] 图1图示了根据本公开的各种实施例的音频处理系统可以在其中操作的示例操作环境100。操作环境100包括音频处理设备105、目标音频源110和一个或多个噪声源135-145。在图1中所图示的示例中,操作环境100被图示为房间,但是可以设想的是,操作环境可以包括其他区域,诸如车辆内部、办公室会议室、家庭房间、室外体育场或机场。根据本公开的各种实施例,音频处理设备105可以包括两个或更多音频感测部件(例如,麦克风)115a-

115d以及,可选地,一个或多个音频输出部件(例如,扬声器)120a-120b。

[0021] 音频处理设备105可以被配置成经由音频感测部件115a-115d感测声音并且生成包括两个或更多音频输入信号的多通道音频输入信号。音频处理设备105可以使用本文中公开的音频处理技术来处理音频输入信号,以增强从目标音频源110接收的音频信号。例如,可以将处理的音频信号传输到音频处理设备105内的其他部件(诸如话音识别引擎或语音命令处理器),或者传输到外部设备。因此,音频处理设备105可以是处理音频信号的独立设备,或者可以是将处理的音频信号转换成其他信号(例如,命令、指令等)以用于与外部设备交互或控制外部设备的设备。在其他实施例中,音频处理设备105可以是通信设备,诸如移动电话或实现IP语音(VoIP)的设备,并且处理的音频信号可以通过网络传输到另一设备以用于输出到远程用户。通信设备还可以从远程设备接收处理的音频信号,并且经由音频输出部件120a-120b输出处理的音频信号。

[0022] 目标音频源110可以是产生可由音频处理设备105检测的声音的任何源。可以基于由用户或系统要求指定的准则来定义要由系统检测的目标音频。例如,目标音频可以被定义为人类话音、由特别的动物或机器发出的声音。在所图示的示例中,目标音频被定义为人类话音,并且目标音频源110是人。除了目标音频源110之外,操作环境100可以包括一个或多个噪声源135-145。在各种实施例中,不是目标音频的声音可以被处理为噪声。在所图示的示例中,噪声源135-145可以包括:播放音乐的扬声器135;播放电视节目、电影或体育赛事的电视140;以及非目标扬声器145之间的背景对话。将领会的是,不同的噪声源可存在于各种操作环境中。

[0023] 注意的是,目标音频和噪声可能从不同的方向以及在不同的时间到达音频处理设备105的音频感测部件115a-115d。例如,噪声源135-145可以在操作环境100内的不同位置处产生噪声,并且目标音频源(人)110可以在操作环境100内的位置之间移动时讲话。此外,目标音频和/或噪声可以从房间100内的固定物(例如,墙壁)反射。例如,考虑目标音频可以从目标音频源110经过以到达每个音频感测部件115a-115d的路径。如由箭头125a-125d所指示,目标音频可以直接从目标音频源110分别传播到音频感测部件115a-115d。另外,目标音频可以从墙壁150a和150b反射,以及从目标音频源110间接到达音频感测部件115a-115d,如由箭头130a-130b所指示。在各种实施例中,音频处理设备105可以使用一种或多种音频处理技术来估计并施加房间脉冲响应,以进一步增强目标音频并抑制噪声。

[0024] 图2图示了根据本公开的各种实施例的示例音频处理设备200。在一些实施例中,音频处理设备200可以被实现为图1的音频处理设备105。音频处理设备200包括音频传感器阵列205、音频信号处理器220和主机系统部件250。

[0025] 音频传感器阵列205包括两个或更多传感器,每个传感器可以实现为将具有声波形式的音频输入转换为音频信号的换能器。在所图示的环境中,音频传感器阵列205包括多个麦克风205a-205n,每个麦克风生成音频输入信号,该音频输入信号被提供给音频信号处理器220的音频输入电路222。在一个实施例中,音频传感器阵列205生成多通道音频信号,其中每个通道对应于来自麦克风205a-n中的一个的音频输入信号。

[0026] 音频信号处理器220包括音频输入电路222、数字信号处理器224和可选的音频输出电路226。在各种实施例中,音频信号处理器220可以被实现为包括模拟电路、数字电路和数字信号处理器224的集成电路,其可操作以执行存储在固件中的程序指令。音频输入电路

222,例如,可以包括到音频传感器阵列205的接口、抗混叠滤波器、模数转换器电路、回声消除电路以及其他音频处理电路和部件。数字信号处理器224可操作以处理多通道数字音频信号来生成增强的音频信号,该增强的音频信号被输出到一个或多个主机系统部件250。在各种实施例中,数字信号处理器224可以可操作以执行回声消除、噪声消除、目标信号增强、后滤波和其他音频信号处理功能。

[0027] 可选的音频输出电路226处理从数字信号处理器224接收的音频信号,以用于输出到至少一个扬声器,诸如扬声器210a和210b。在各种实施例中,音频输出电路226可以包括将一个或多个数字音频信号转换为模拟音频信号的数模转换器和用于驱动扬声器210a-210b的一个或多个放大器。

[0028] 音频处理设备200可以被实现为可操作以接收和增强目标音频数据的任何设备,诸如,例如,移动电话、智能扬声器、平板电脑、膝上型计算机、台式计算机、语音控制的装置或汽车。主机系统部件250可以包括用于操作音频处理设备200的各种硬件和软件部件。在所图示的实施例中,主机系统部件250包括处理器252、用户接口部件254、用于与外部设备和网络(诸如网络280(例如,英特网、云、局域网络或蜂窝网络)和移动设备284)进行通信的通信接口256以及存储器258。

[0029] 处理器252和数字信号处理器224可以包括处理器、微处理器、单核处理器、多核处理器、微控制器、可编程逻辑器件(PLD)(例如,现场可编程门阵列(FPGA))、数字信号处理(DSP)设备或其他逻辑器件(其可以通过硬接线、执行软件指令或两者的组合来配置)中的一个或多个,以执行针对本公开的实施例的本文中讨论的各种操作。主机系统部件250被配置成诸如通过总线或其他电子通信接口与音频信号处理器220和其他主机系统部件250对接和通信。

[0030] 将领会的是,尽管音频信号处理器220和主机系统部件250被示出为并入硬件部件、电路和软件的组合,但是在一些实施例中,硬件部件和电路可操作以执行的至少一些或全部功能性可被实现为软件模块,该软件模块响应于存储在数字信号处理器224的固件或存储器258中的软件指令和/或配置数据而由处理器252和/或数字信号处理器224执行。

[0031] 存储器258可以被实现为可操作以存储包括音频数据和程序指令的数据和信息的一个或多个存储设备。存储器258可以包括一个或多个各种类型的存储设备,包括易失性和非易失性存储设备,诸如RAM(随机存取存储器)、ROM(只读存储器)、EEPROM(电可擦除只读存储器)、闪速存储器、硬盘驱动器和/或其他类型的存储器。

[0032] 处理器252可以可操作以执行存储在存储器258中的软件指令。在各种实施例中,话音识别引擎260可操作以处理从音频信号处理器220接收的增强的音频信号,包括识别和执行语音命令。诸如通过移动或蜂窝电话网络上的语音呼叫或IP网络上的VoIP呼叫,语音通信部件262可以可操作以促进与诸如移动设备284或用户设备286之类的一个或多个外部设备的语音通信。在各种实施例中,语音通信包括将增强的音频信号传输到外部通信设备。

[0033] 用户接口部件254可以包括显示器、触摸板显示器、小键盘、一个或多个按钮和/或其他输入/输出部件,其可操作以使用户能够直接与音频处理设备200交互。

[0034] 通信接口256促进音频处理设备200与外部设备之间的通信。例如,通信接口256可以实现音频处理设备200与一个或多个本地设备(诸如移动设备284或诸如通过网络280提供对远程服务器282的网络访问的无线路由器)之间的Wi-Fi(例如,802.11)或蓝牙连接。在

各种实施例中，通信接口256可以包括其他有线和无线通信部件，其促进音频处理设备200与一个或多个其他设备之间的直接或间接通信。

[0035] 图3图示了根据本公开的各种实施例的示例音频信号处理器300。在一些实施例中，音频信号处理器300体现为一个或多个集成电路，其包括模拟和数字电路以及由数字信号处理器(诸如图2的音频信号处理器220)实现的固件逻辑。如所图示，音频信号处理器300包括音频输入电路315、子带频率分析器320、多轨道VAD引擎325、音频增强引擎330和合成器335。

[0036] 音频信号处理器300接收来自多个音频传感器(诸如包括至少两个音频传感器305a-n的传感器阵列305)的多通道音频输入。音频传感器305a-305n可以包括与诸如图2的音频处理设备200之类的音频处理设备集成的或与连接到其的外部部件集成的麦克风。根据本公开的各种实施例，音频信号处理器300可以知道或不知道音频传感器305a-305n的布置。

[0037] 音频信号可以初始地由音频输入电路315处理，该音频输入电路315可以包括抗混叠滤波器、模数转换器和/或其他音频输入电路。在各种实施例中，音频输入电路315输出数字、多通道、时域音频信号，其中M是传感器(例如，麦克风)输入的数量。将多通道音频信号输入到子带频率分析器320，其将多通道音频信号划分为连续的帧，并将每个通道的每一帧分解为多个频率子带。在各种实施例中，子带频率分析器320包括傅立叶变换过程并且输出多个频率窗口。然后分解的音频信号被提供给多轨道VAD引擎325和音频增强引擎330。

[0038] 多轨道VAD引擎325可操作以分析一个或多个音频轨道的帧并生成指示当前帧中是否存在目标音频活动的VAD输出。如上所讨论，目标音频可以是要由音频系统识别的任何音频。当目标音频是人类话音时，多轨道VAD引擎325可以被具体实现为用于检测语音活动。在各种实施例中，多轨道VAD引擎325可操作以接收音频数据的帧，并且针对每个音频轨道生成关于对应于音频数据的帧的相应音频轨道上的目标音频的存在或不存在的VAD指示输出。关于图4中的400进一步图示了多轨道VAD引擎325的详细部件和操作。

[0039] 音频增强引擎330从子带频率分析器320接收子带帧，并从多轨道VAD引擎325接收VAD指示。根据本公开的各种实施例，音频增强引擎330配置成基于接收的多轨道VAD指示来处理子带帧，以增强多轨道音频信号。例如，音频增强引擎330可以增强被确定为来自目标音频源的方向的音频信号的部分，并且抑制被确定为噪声的音频信号的其他部分。

[0040] 在增强目标音频信号之后，音频增强引擎330可以将处理的音频信号传递给合成器335。在各种实施例中，合成器335通过组合子带以形成增强的时域音频信号来以逐帧为基础重构一个或多个多通道音频信号。然后，增强的音频信号可以被变换回到时域，并被发送到系统部件或外部设备以用于进一步处理。

[0041] 图4图示了根据本公开的各种实施例的用于处理来自通用麦克风阵列的多个音频信号的示例多轨道VAD引擎400。多轨道VAD引擎400可以被实现为由数字信号处理器执行的数字电路和逻辑的组合。在一些实施例中，可以将多轨道VAD引擎400安装在诸如图3中的300之类的音频信号处理器中。多轨道VAD引擎400可以向图3中的多轨道VAD引擎325提供进一步的结构和功能细节。

[0042] 根据本公开的各种实施例，多轨道VAD引擎400包括子带分析模块405、基于块的TDOA估计模块410、TDOA轨迹计算模块420以及多源跟踪和多流VAD估计模块430。

[0043] 子带分析模块405接收由$x_m(t),\ m=1,...,M$表示的多个音频信号402，即在对于总共M个麦克风(例如，类似于图3中的音频传感器305a-n)而言的第m个麦克风处记录的采样的时域音频信号。可以经由图3中的音频输入电路315来接收音频信号$x_m(t),\ m=1,...,M$。

[0044] 子带分析模块405被配置成获得音频信号402并将其变换为时频域表示404，其表示为与原始时域音频信号$x_m(t)$对应的$X_m(l,k)$，其中l指示子带时间索引，以及k指示频带索引。例如，子带分析模块405可以类似于图3中的子带频率分析器320，其执行傅立叶变换以将输入时域音频信号转换为频域表示。子带分析模块405然后可以将生成的时频域表示404发送到基于块的TDOA估计模块410以及多源跟踪和多流VAD估计模块430。

[0045] TDOA轨迹计算模块420被配置成扫描通用麦克风阵列(例如，形成通用阵列几何形状的音频传感器305a-n)。例如，对于平面上给定的任意麦克风阵列几何形状，在系统启动时计算一次可准许的TDOA位置的轨迹。点的该轨迹可以避免重影形成。

[0046] 对于M个麦克风的阵列，可以选择第一麦克风作为参考麦克风，其转而给出全部相对于第一麦克风的M-1个麦克风对。例如，图5A图示了示例麦克风对。索引为第i-1对的麦克风对包括麦克风i 502和用于入射声线505的参考麦克风1 501，该入射声线505具有从远源(假设远场模型)发射的方位角角度θ和为零的仰角角度。501和502的麦克风对之间的距离以及两个麦克风之间的角度分别表示为d_{i-1}和ψ_{i-1}，其可以在给出第i麦克风502的笛卡尔坐标的情况下计算。对于当入射声线505以方位角θ和仰角φ成角度时的一般情况，第(i-1)麦克风对的TDOA可以被计算为

$$\tau_{i-1}(\theta,\phi)=\frac{d_{i-1}}{c}cos(\theta-\psi_{i-1})cos\phi \qquad (1)$$

其中c是传播速度。

[0047] 在扫描了不同的仰角和方位角角度之后，TDOA轨迹计算模块420可以构造可准许的TDOA的网格。当所有M个麦克风位于平面上时，产生的TDOA轨迹(针对θ和φ的所有扫描的$(\tau_1(\theta,\phi),\ldots,\tau_p(\theta,\phi),\ldots,\tau_{M-1}(\theta,\phi))$也位于(M-1)维空间中的平面上。M个麦克风的不同布局可能导致(M-1)维空间中的不同平面。

[0048] 例如，图5B图示了两个不同的示例麦克风布局连同它们相应的TDOA网格。在510处示出了一组M=4个麦克风，其中第一和第三麦克风之间的距离为8 cm，以及可准许的TDOA的产生的网格在如515处示出的M-1=3维空间中。当第一和第三麦克风之间的距离增加到16 cm时，如在520处所示，在525处示出了可准许的TDOA的产生的网格。

[0049] 回到参考图4，TDOA轨迹计算模块420然后可以将(M-1)维TDOA 403发送到基于块的TDOA估计模块410。基于块的TDOA估计模块410接收TDOA 403和多源音频的时频域表示404，基于所述时频域表示404，TDOA估计模块410使用从连续帧获得的数据来提取源麦克风(例如，图3中所示出的音频传感器305a-n)的TDOA信息。

[0050] 在一个实施例中，基于块的TDOA估计模块410采用导向最小方差(STMV)波束形成器来从多源音频的时频域表示404获得TDOA信息。具体地，基于块的TDOA估计模块410可以选择麦克风作为参考麦克风，然后通过将剩余的M-1个麦克风与参考麦克风配对来指定总共M-1个麦克风对。麦克风对由$p=1,...,M-1$索引。

[0051] 例如，可以选择第一麦克风作为参考麦克风，并且因此，$X_1(l,k)$表示来自参考麦克风的音频的时频表示。对于第p对麦克风，基于块的TDOA估计模块410以矩阵形式将第p对的

频率表示计算为$\boldsymbol{X}_p(k,l)=\left[X_1(l,k)\ X_{p+1}(l,k)\right]^T$，其中$()^T$代表转置。然后，基于块的TDOA估计模块410计算每个频带k的第p个输入信号对的协方差矩阵：

$$R_p(k)=\sum_l \boldsymbol{X}_p(k,l)\boldsymbol{X}_p(k,l)^H \quad (2)$$

其中$()^H$代表厄米转置。

[0052] 在一些实现方式中，在一定数量的连续帧的块上实现计算$R_p(k)$中的求和。为简洁起见，此处忽略块的索引。

[0053] 然后，基于块的TDOA估计模块410可以为每个对和频带如下构造导向矩阵：

$$T_k(\tau_p)=diag([1\ e^{-j2\pi f_k\tau_p}]^T) \quad (3)$$

其中，τ_p是在对θ和φ的不同扫描(为简洁起见忽略)之后从TDOA轨迹计算模块420获得的第p对的TDOA；f_k是频带k处的频率；以及diag([a, b])表示具有对角元素a和b的2×2对角矩阵。

[0054] 对于每个麦克风对p，基于块的TDOA估计模块410通过以下方式构造跨所有频带相干地对齐的方向协方差矩阵：

$$C_p(\tau_p)=\sum_k T_k(\tau_p)R_p(k)T_k(\tau_p)^H \quad (4)\ 。$$

[0055] 方向协方差矩阵$C_p(\tau_p)$的计算在所有麦克风对p和针对τ_p的方位角/仰角(θ,ϕ)的所有扫描上被重复。为了减少所有扫描上的计算，将对应于第p个麦克风对的每个维度p的TDOA空间线性地量化成q段。在处理开始时(在系统启动时)，通过扫描每个方位角和仰角角度(θ,ϕ)获得的TDOA轨迹点$(\tau_1,\ldots\tau_p,\ldots,\tau_{M-1})$被映射到对于每个维度而言最接近的量化点。对于每个方位角/仰角(θ,ϕ)，$(\theta,\phi)\rightarrow(ind_1(\theta,\phi),\ldots,ind_{M-1}(\theta,\phi))$的映射保存在存储器中，其中$1\le ind_p(\theta,\phi)\le q$是与扫描角度θ和φ有关的维度p的量化的TDOA索引。

[0056] 例如，如果存在M=4个麦克风，并且方位角和仰角扫描分别为$\Theta=0^o:5:355^o$ $\Phi=0^o:10:80^o$。需要执行的$C_p(\tau_p)$的不同计算的数量为

$length(\Theta)\times length(\Phi)\times(M-1)=72\times9\times3=1944$。当TDOA轨迹点$(\tau_1,\ldots\tau_p,\ldots,\tau_{M-1})$被量化时，因为TDOA维度中的一些可以量化为q个量化段之中的相同段，因此不需要执行所有计算。因此，例如，如果$q=50$，则计算$C_p(\tau_p)$所需的不同计算的最大数量被减少为$q\times(M-1)=50\times3=150$。利用TDOA量化来执行$C_p(\tau_p)$的计算的伪代码可以在算法1中展示如下：

算法1 使用TDOA量化来计算$C_p(\tau_p)$

$R_{avg} = zeros(2,2,length(\theta),length(\Phi),M-1)$
for $k>0$ **do**
  $V = zeros(2,2,\boldsymbol{q},M-1)$
  for $p = 1:M-1$ **do**
    for $i = 1:length(\theta)$ **do**
      for $j = 1:length(\Phi)$ **do**
        if $V(:,:,ind_p(\theta(i),\Phi(j)),p) = \mathbf{0}$ **then**
          $T_k = diag([1 \quad e^{-j2\pi f_k \tau_p(\theta(i),\Phi(j))}]^T)$
          $V(:,:,ind_p(\theta(i),\Phi(j)),p) = T_k R_p(k) T_k{}^H$
        end if
        $R_{avg}(:,:,i,j,p) = R_{avg}(:,:,i,j,p) + V(:,:,ind_p(\theta(i),\Phi(j)),p)$
      end for
    end for
  end for
end for

。

[0057] 接下来，对于每一对p，服从于无失真准则(具有其等效伪似然解)对波束功率进行最小化的方向被计算如下：

$$\mathcal{L}_p^{STMV}(\tau_p) = \frac{1}{\mathbf{1}^T C_p^{-1}(\tau_p)\mathbf{1}} \qquad (5)$$

其中$\mathbf{1} = [1\ 1]^T$。然后，基于块的TDOA估计模块410可以针对所有M-1对麦克风将STMV联合(joint)伪似然度计算为：

$$\mathcal{L}^{STMV}(\tau_1,\ldots,\tau_{M-1}) = \prod_{p=1}^{M-1} \mathcal{L}_p^{STMV}(\tau_p) \qquad (6)$$ 。

[0058] 然后识别产生所有M-1个对的最大STMV联合伪似然度的方位角和仰角，由以下等式表示

$$\theta^*,\phi^* = argmax_{\theta,\phi}\ \mathcal{L}^{STMV}(\tau_1,\ldots,\tau_{M-1}) \qquad (7)$$

然后可以将方位角和仰角对θ^*、ϕ^*用于多源跟踪和多流VAD估计。一种可能的解决方案可以包括直接跟踪每个麦克风对的两个麦克风之间的角度。然而，由于360度中的方位角的环绕(wrap-around)效应，如果直接对成对的麦克风之间的角度进行跟踪，则当麦克风源跨过0°朝向360°时以及反之亦然，可能发生轨道丢失。因此，为了避免这样的紊乱，使用极性变换以圆形方式基于成对的麦克风之间的角度将检测z计算如下：

$$z = \alpha_{scale}[cos\theta^* cos\phi^* \quad sin\theta^* cos\phi^*]^T \qquad (8)$$

其中$\alpha_{scale} > 1$是缩放常数，其可以扩展测量空间，从而允许利用与有意义的概念(如角度)有关的参数进行跟踪。

[0059] 然后，基于块的TDOA估计模块410将计算的检测z发送到多源跟踪和多流VAD估计模块430。如果存在轨道的为$\mathcal{N}$的最大数量，则从基于块的TDOA估计模块410获得的TDOA要通过以递归方式更新从先前步骤获得的轨道而被跟踪。具体地，如果在块(时间-步长)n-1处获得的检测由z_{n-1}表示，并且直到那时存在t_{n-1}个轨道，则对于在时间-步长n处出现的新的检测z_n 406，多源跟踪和多流VAD估计模块430基于现有轨道的门对新的检测z_n如下进行处理：

如果z_n仅落入先前t_{n-1}个轨道中的一个的门中，则更新特别轨道以并入检测z_n。

[0060] 如果z_n落入多个先前t_{n-1}个轨道的重叠门中，则更新最接近于检测z_n的轨道以并入

检测z_n。

[0061] 如果z_n不落入任何先前t_{n-1}个轨道的门中，并且未达到轨道的最大数量$\mathcal{N}$（例如，$t_{n-1}<\mathcal{N}$），则发动新的轨道以并入检测z_n并在时间-步长n处更新现有轨道的数量，例如，$t_n = t_{n-1} + 1$。

[0062] 如果z_n不落入任何先前t_{n-1}个轨道的门中，并且已达到轨道的最大数量$\mathcal{N}$（例如，$t_{n-1}=\mathcal{N}$），则在现有的$\mathcal{N}$个轨道之中具有最低功率的轨道被停止，并利用新的轨道来代替以并入检测z_n。

[0063] 对于不被更新、发动或代替(如先前步骤中那样)的所有其他轨道而言，则这些轨道以相同的平均值被更新，但每个相应轨道的方差增加，以(例如，基于随机行走模型)计及不确定性。每个相应轨道的功率还被衰减，使得未来出现的源有机会被发动。以此方式，可以从模块430输出在时间-步长n处并入最新检测406的跟踪结果408，其由$z_n^{Tr}(1),\ldots z_n^{Tr}(\mathcal{N})$表示。

[0064] 当所有音频轨道已经被更新时，模块430使用最近的邻近物M2T分配来生成多流VAD 412。具体地，在时间-步长n处，可以通过将1分配给最接近于检测z_n的轨道以及将0分配给其他轨道来执行M2T分配。在一些实现方式中，可在VAD在先前时间-步长中为1之后被完全分配为零之前将释放延迟(hangover)应用于VAD以具有中间值，例如，-1，等。以此方式，从模块430，例如，向图3中的音频增强引擎330输出由$VAD_1,\ldots,VAD_N$表示的多流VAD 412以用于音频增强，每个多流VAD 412代表在相应轨道中是否找到任何语音活动检测。

[0065] 图6图示了根据本公开的各种实施例的用于通过多源跟踪和VAD来增强多源音频信号的示例方法600。在一些实施例中，方法600可以由音频信号处理器300中的一个或多个部件和/或多轨道VAD引擎400的一个或多个部件来执行。

[0066] 方法600以步骤602开始，在步骤602，可以基于麦克风阵列的空间信息来计算TDOA轨迹信息。例如，TDOA轨迹信息可以在系统启动时通过利用变化的方位角和仰角角度的入射声线扫描麦克风阵列来计算一次。如关于图7进一步描述的那样，可以在通过将来自麦克风阵列的麦克风配对而构成的多维空间中以降低的复杂度来执行计算。

[0067] 参考图7，其为步骤602提供了进一步的详细步骤，在步骤702，可以从麦克风阵列选择第一麦克风作为参考麦克风。在步骤704，可以将来自麦克风阵列的每个剩余的麦克风与参考麦克风配对。在步骤706，对于每个麦克风对，可以基于相应对中的两个麦克风之间的距离和角度(例如，根据关于图4描述的等式(1))来计算对应于入射声线的特定方位角角度和特定仰角角度的TDOA位置。在图5A中还示出了具有入射声线的特定方位角角度和特定仰角角度的示例麦克风对。

[0068] 在步骤708，如果存在更多的麦克风对要处理，则该方法在步骤710检索(retrieve)下一个麦克风对并在步骤706重复，直到已经计算了所有麦克风对的TDOA位置。

[0069] 在步骤712，如果存在方位角和仰角角度的更多扫描，则该方法在步骤714检索方位角和仰角角度的下一次扫描并在步骤706重复，直到计算出方位角和仰角角度的所有扫描的TDOA位置。

[0070] 在步骤712，当没有方位角和仰角角度的更多扫描要处理时(例如，已经在方位角和仰角角度的所有扫描上针对所有麦克风对计算了TDOA位置)，则可以在步骤716形成TDOA位置点的网格。图5B中示出了对应于麦克风阵列的不同几何形状的TDOA位置点的示例网

格。

[0071] 回到参考图6，在系统启动时计算TDOA轨迹信息时，方法600前进到步骤604。在步骤604，可以从麦克风阵列接收一个或多个多源音频信号。例如，可以经由图3中的音频输入电路315来接收图4中的多源音频信号402的时域采样。

[0072] 在步骤606，可以将一个或多个多源音频信号从时域表示变换成时频表示。例如，子带分析模块405可以将时域信号变换成时频域表示，如关于图4所描述的那样。

[0073] 在步骤608，可以基于计算的TDOA轨迹，根据STMV波束形成器针对一个或多个多源音频信号计算TDOA检测数据。例如，对于每个麦克风对，可以针对每一个频带使用来自相应麦克风对的一个或多个多源音频信号的时频表示(例如，根据关于图4描述的等式(2))来计算协方差矩阵。然后，可以基于TDOA位置针对每一个频带(例如，根据关于图4描述的等式(3))构造导向矩阵，该TDOA位置针对对应于相应麦克风对的方位角和仰角角度的不同扫描。可以基于构造的导向矩阵和计算的协方差矩阵(例如，根据关于图4描述的等式(4))来构造跨所有频带对齐的方向协方差矩阵。可以基于构造的方向协方差矩阵(例如，根据关于图4描述的等式(5))确定使波束功率最小化的伪似然解。然后可以通过取跨所有麦克风对的所有确定的伪似然解的乘积来(例如，根据关于图4描述的等式(6))计算STMV联合伪似然度。然后可以(例如，根据关于图4描述的等式(7))确定使STMV联合伪似然度最大化的一对方位角和仰角角度。然后可以(例如，根据关于图4描述的等式(8))将确定的一对方位角和仰角角度变换为代表TDOA检测数据的极坐标表示。

[0074] 在步骤610，基于直到当前时间-步长的计算的TDOA检测数据可以更新多个音频轨道并且可以构造VAD数据。例如，可以识别对应于当前时间-步长的TDOA检测和先前已经建立直到当前时间-步长的一组现有音频轨道。然后，方法600可以基于现有音频轨道的门与TDOA检测之间的比较来确定是将TDOA检测并入到现有音频轨道中的一个中还是建立新的音频轨道，如关于图4中的模块430所描述的那样。对于另一个示例，方法600可以在相应音频轨道最接近于TDOA检测时将第一值分配给相应音频轨道的VAD，以及将第二值分配给其他音频轨道的VAD，如关于图4中的模块430所描述的那样。

[0075] 在步骤612，可以使用更新的多个音频轨道和构造的VAD数据来生成一个或多个增强的多源音频信号。例如，增强的多源音频信号然后可以被传输到各种设备或部件。对于另一个示例，增强的多源音频信号可以被打包并通过网络传输到另一个音频输出设备(例如，智能电话、计算机等)。增强的多源音频信号也可以被传输到诸如自动话音识别部件之类的语音处理电路以进行进一步处理。

[0076] 前述公开不旨在将本发明限制为所公开的精确形式或特别的使用领域。因此，设想的是，根据本公开，无论在本文中明确描述还是暗示，对本公开的各种替代实施例和/或修改是可能的。例如，本文中描述的实施例可用于提供环境中的多个声音源的位置，以便(例如，在并入来自诸如视频流、3D相机、激光雷达等之类的其他模态的附加信息的应用中)监督人机交互任务。已经像这样描述了本公开的实施例，本领域普通技术人员将认识到优于常规方法的优点，并且可以在形式和细节上做出改变而不背离本公开的范围。因此，本公开仅由权利要求限制。

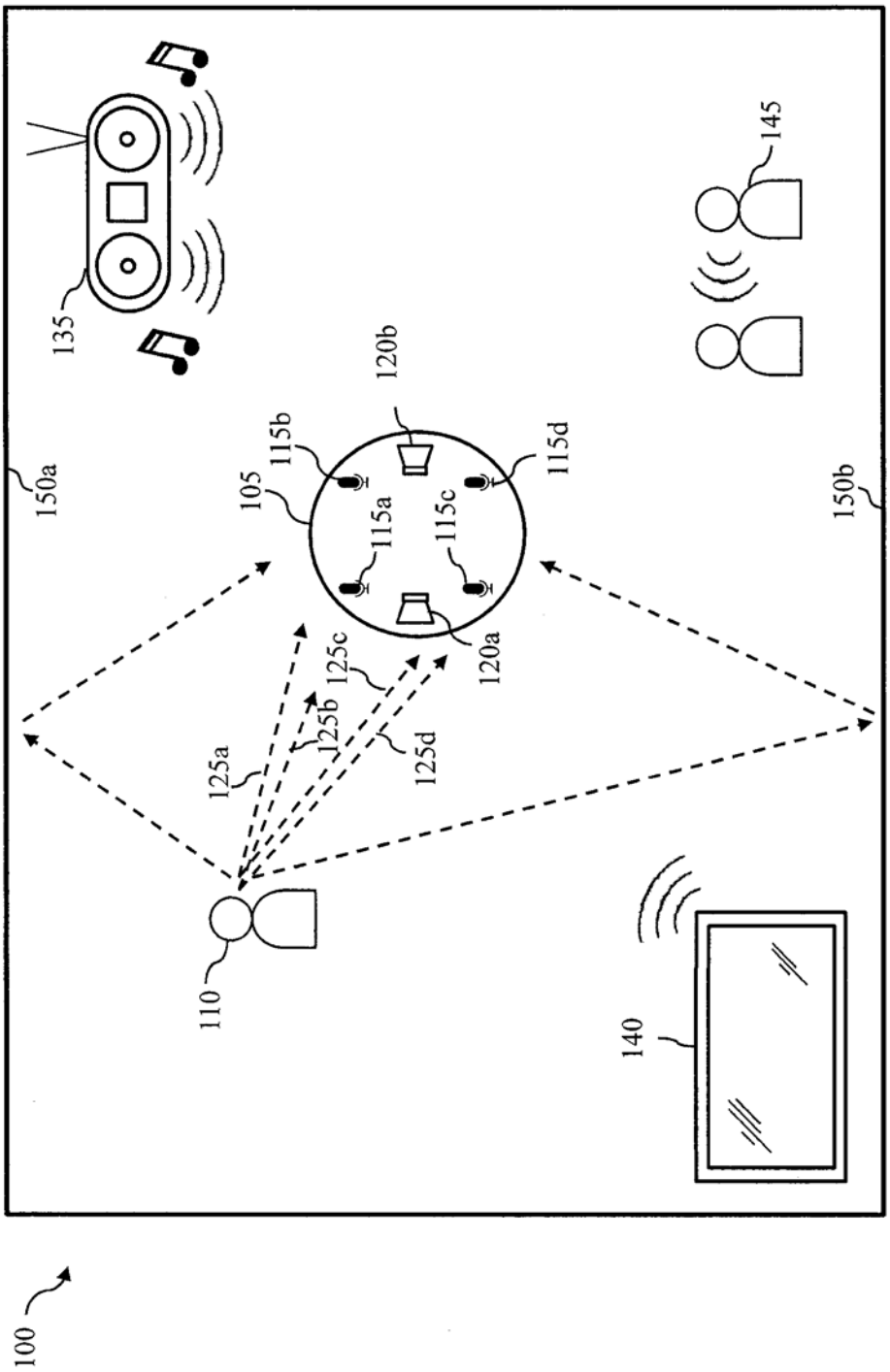

135
145
120b
115b
105
115d
150a
150b
115a
115c
120a
125c
125b
125a
125d
110
140
100

图 1

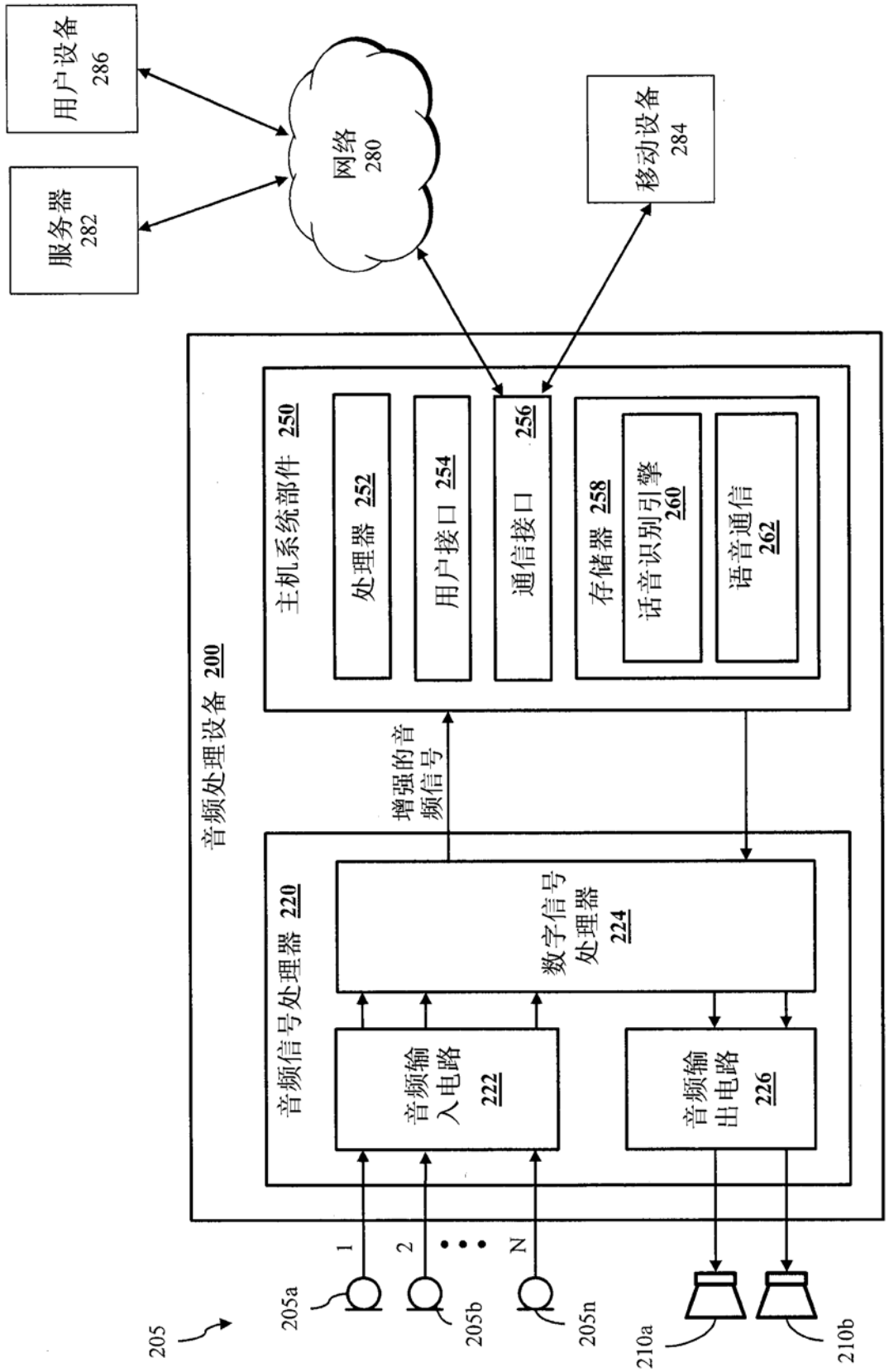

用户设备
286
服务器
282
网络
280
移动设备
284
音频处理设备 200
主机系统部件 250
处理器 252
用户接口 254
通信接口 256
存储器 258
话音识别引擎
260
语音通信
262
增强的音
频信号
音频信号处理器 220
数字信号
处理器
224
音频输
入电路
222
音频输
出电路
226
1
2
N
205
205a
205b
205n
210a
210b

图 2

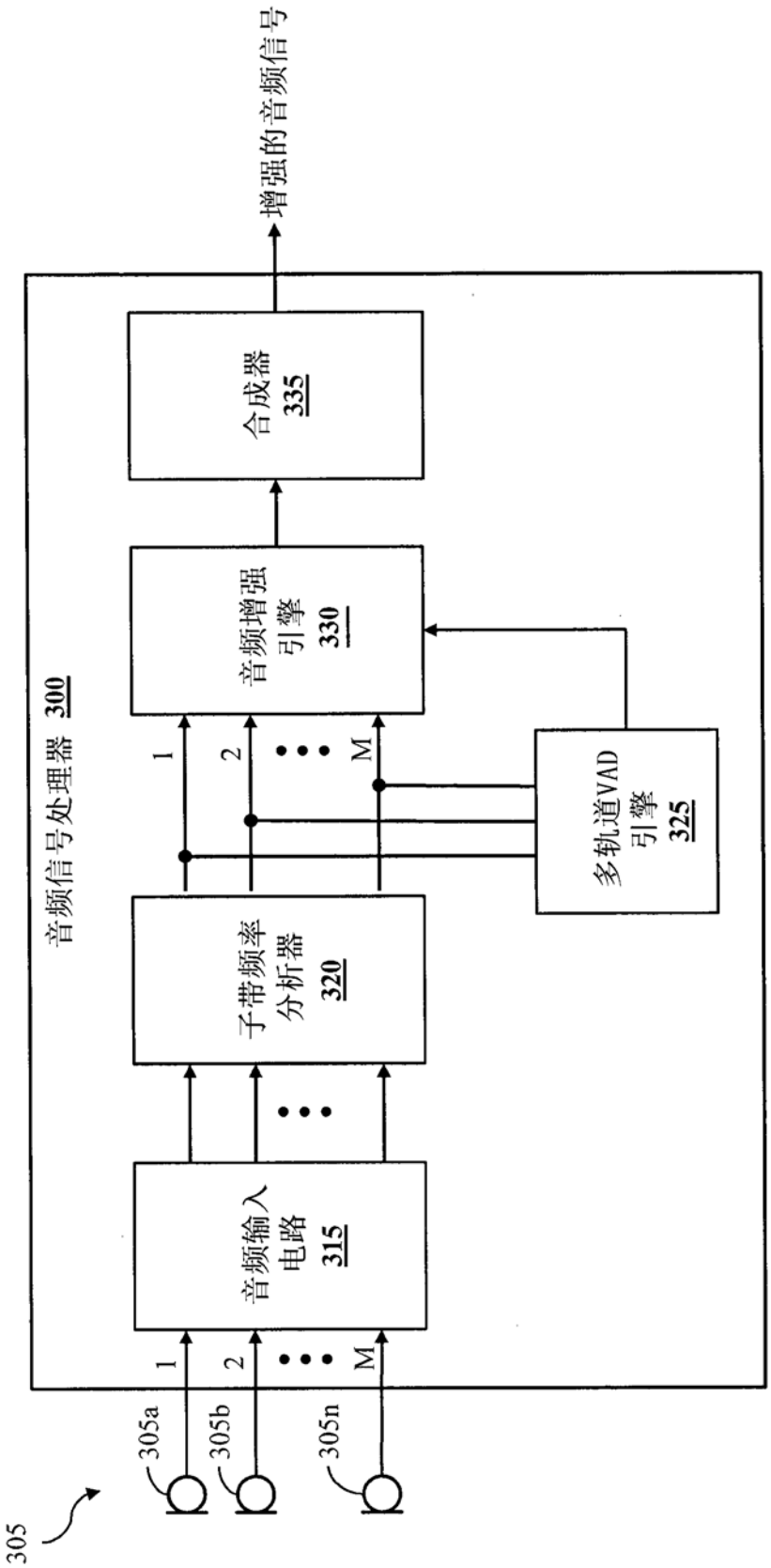

增强的音频信号
音频信号处理器 300
合成器 335
音频增强引擎 330
子带频率分析器 320
音频输入电路 315
多轨道VAD引擎 325
1
2
M
305a
305b
305n
305

图 3

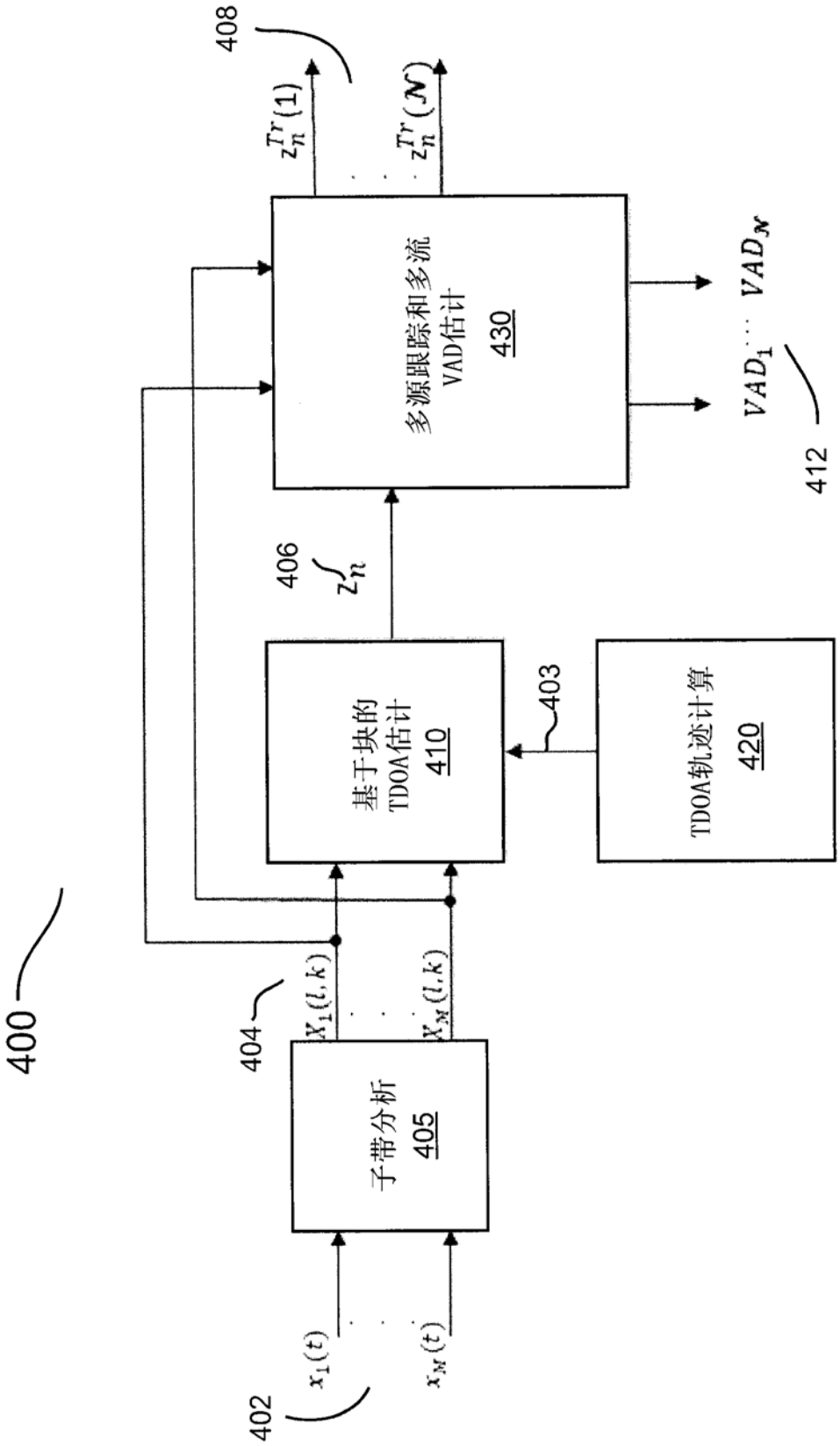

400
402
$x_1(t)$
$x_M(t)$
404
$X_1(l,k)$
$X_M(l,k)$
子带分析
405
基于块的
TDOA估计
410
403
TDOA轨迹计算
420
406
z_n
多源跟踪和多流
VAD估计
430
408
$z_n^{Tr}(1)$
$z_n^{Tr}(N)$
$VAD_1 \cdots VAD_N$
412

图 4

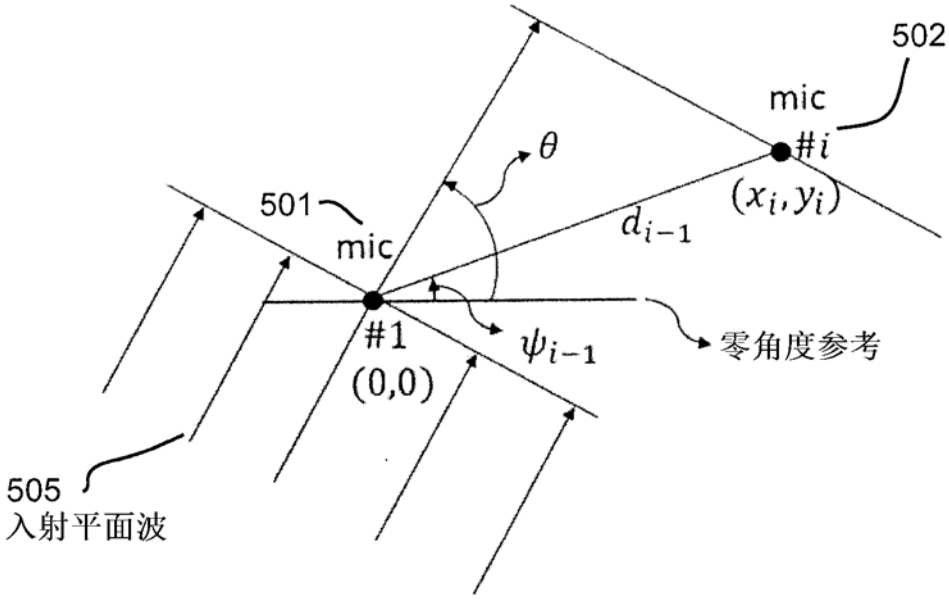

502
mic
#i
(x_i, y_i)
θ
501
mic
d_{i-1}
#1
(0,0)
ψ_{i-1}
零角度参考
505
入射平面波

图 5A

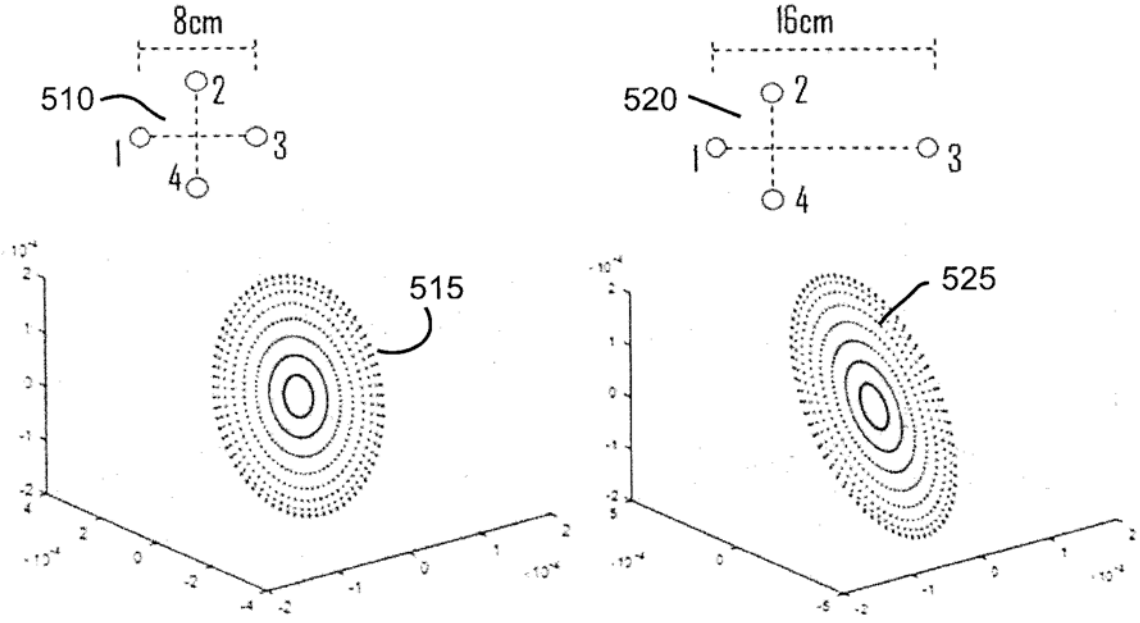

8cm
510
2
1
3
4
16cm
520
2
1
3
4
515
525

图 5B

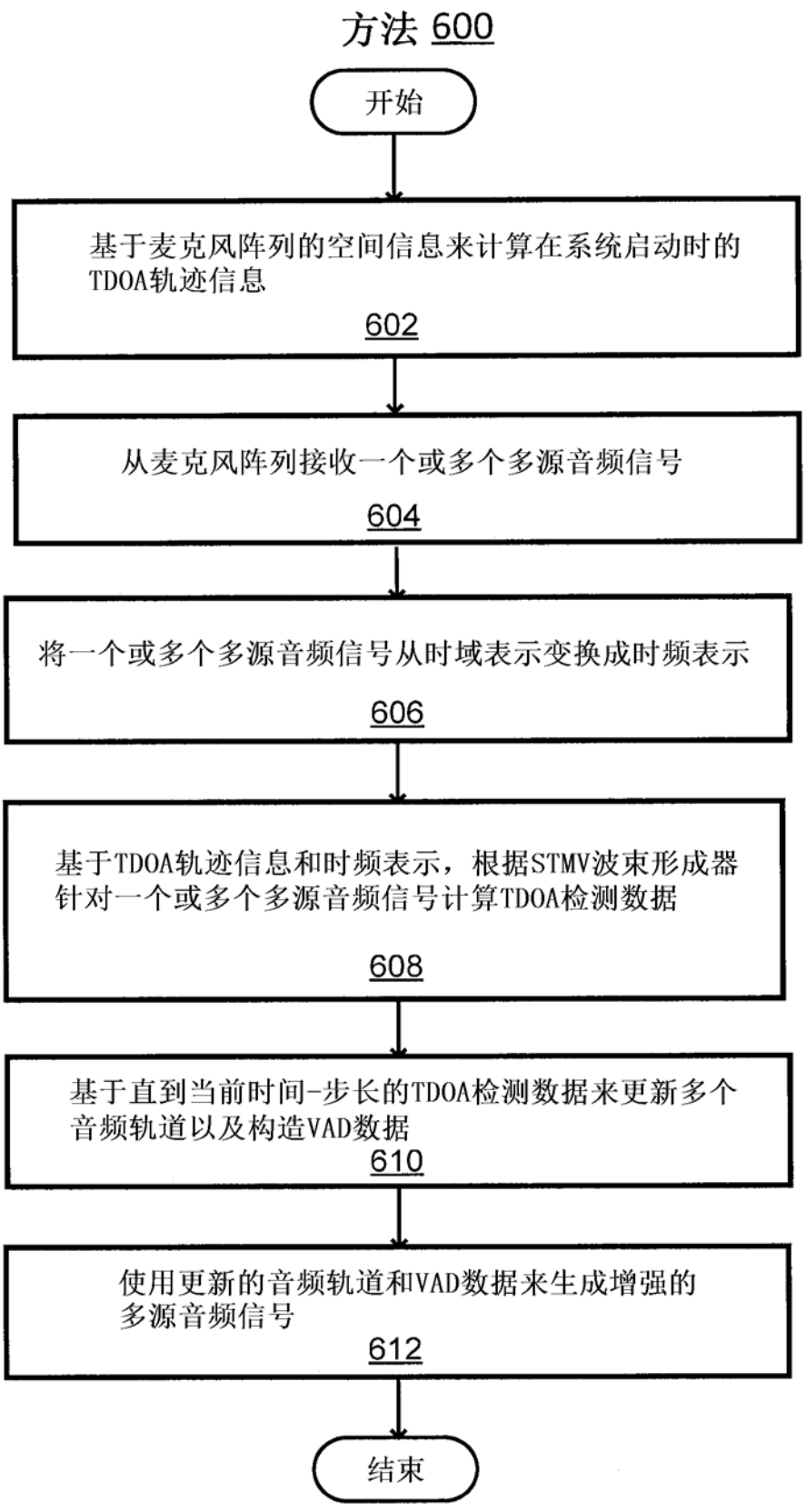

方法 600
开始
基于麦克风阵列的空间信息来计算在系统启动时的TDOA轨迹信息
602
从麦克风阵列接收一个或多个多源音频信号
604
将一个或多个多源音频信号从时域表示变换成时频表示
606
基于TDOA轨迹信息和时频表示，根据STMV波束形成器针对一个或多个多源音频信号计算TDOA检测数据
608
基于直到当前时间-步长的TDOA检测数据来更新多个音频轨道以及构造VAD数据
610
使用更新的音频轨道和VAD数据来生成增强的多源音频信号
612
结束

图 6

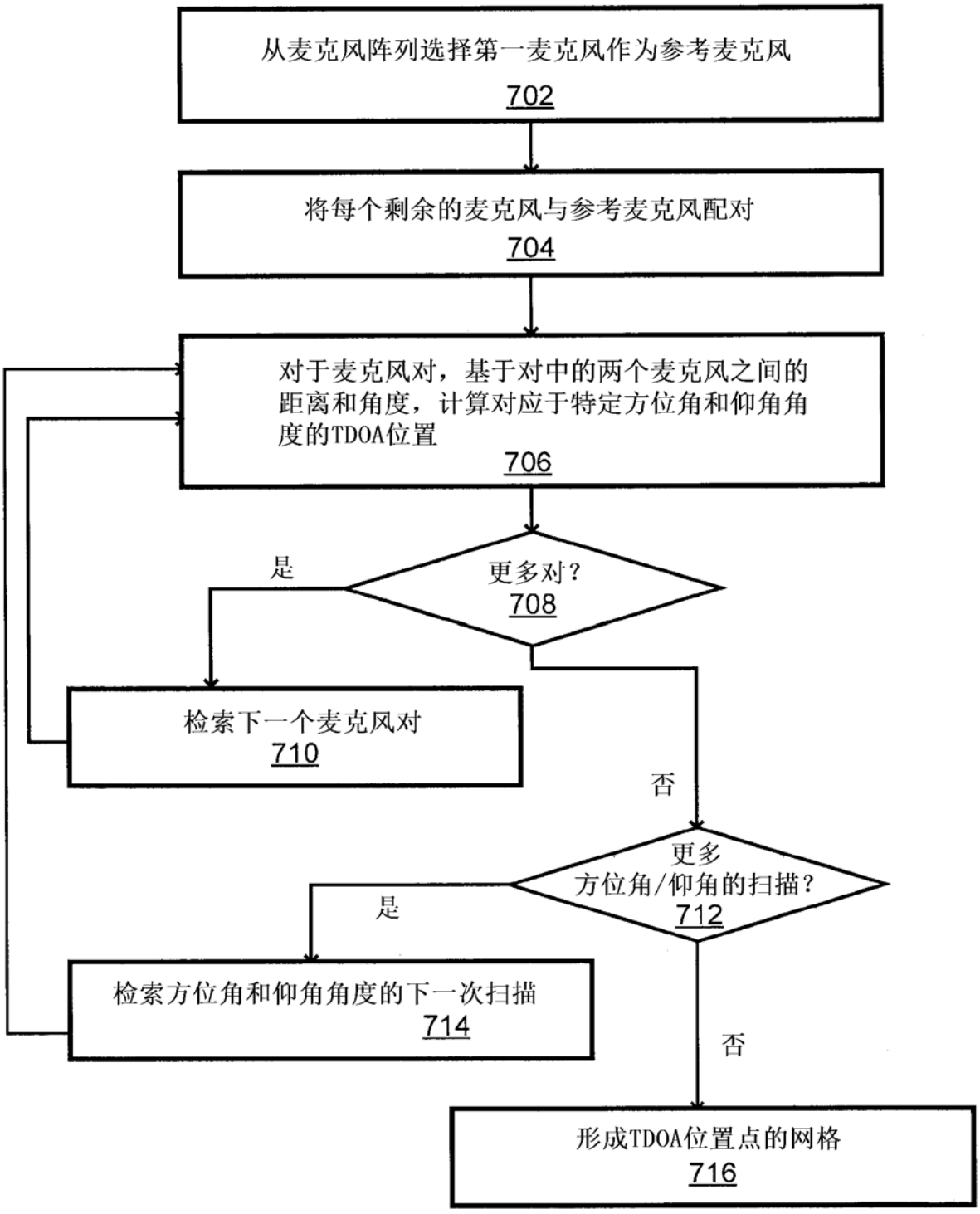

从麦克风阵列选择第一麦克风作为参考麦克风
702
将每个剩余的麦克风与参考麦克风配对
704
对于麦克风对，基于对中的两个麦克风之间的距离和角度，计算对应于特定方位角和仰角角度的TDOA位置
706
更多对？
708
是
检索下一个麦克风对
710
否
更多方位角/仰角的扫描？
712
是
检索方位角和仰角角度的下一次扫描
714
否
形成TDOA位置点的网格
716

图 7