(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局

(43) 国际公布日
2024 年 4 月 4 日 (04.04.2024)

WIPO | PCT

(10) 国际公布号
WO 2024/066751 A1

(51) 国际专利分类号:
G06F 3/01 (2006.01)

(21) 国际申请号: PCT/CN2023/111770

(22) 国际申请日: 2023 年 8 月 8 日 (08.08.2023)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202211211572.5 2022年9月30日 (30.09.2022) CN

(71) 申请人:歌尔股份有限公司(**GOERTEK INC.**) [CN/CN]; 中国山东省潍坊市高新技术产业开发区东方路268号, Shandong 261031 (CN)。

(72) 发明人: 王文 (**WANG, Wen**); 中国山东省潍坊市高新技术产业开发区东方路 268 号, Shandong 261031 (CN)。 李昱锋 (**LI, Yufeng**); 中国山东省潍坊市高新技术产业开发区东方路 268 号, Shandong 261031 (CN)。 李佳明 (**LI, Jiaming**); 中国山东省潍坊市高新技术产业开发区东方路268号, Shandong 261031 (CN)。

(81) 指定国(除另有指明，要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD,

(54) **Title:** AR GLASSES AND AUDIO ENHANCEMENT METHOD AND APPARATUS THEREFOR, AND READABLE STORAGE MEDIUM

(54) 发明名称: 一种AR眼镜及其音频增强方法和装置、可读存储介质

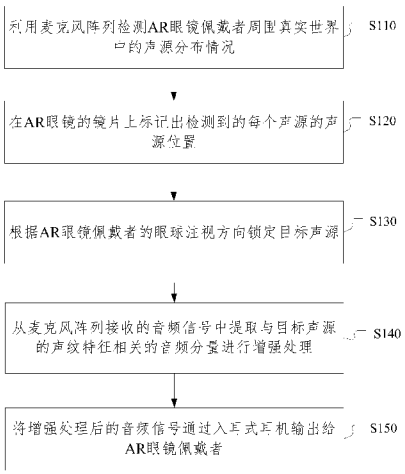

图 1

S110 Use a microphone array to detect a sound source distribution condition in a real world around a wearer of AR glasses
S120 Mark the sound source position of each detected sound source on lenses of the AR glasses
S130 Lock a target sound source according to the eyeball gaze direction of the wearer of the AR glasses
S140 Extract, from an audio signal received by the microphone array, an audio component related to a voiceprint feature of the target sound source to perform enhancement processing
S150 Output the enhanced audio signal to the wearer of the AR glasses by means of an in-ear headphone

(57) **Abstract:** Disclosed in the embodiments of the present invention are AR glasses and an audio enhancement method and apparatus therefor, and a readable storage medium. The audio enhancement method for AR glasses comprises: using a microphone array to detect a sound source distribution condition in a real world around a wearer of AR glasses; marking the sound source position of each detected sound source on lenses of the AR glasses; locking a target sound source according to the eyeball gaze direction of the wearer of the AR glasses; extracting, from an audio signal received by the microphone array, an audio component related to a voiceprint feature of

WO 2024/066751 A1

[见续页]

SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国(除另有指明，要求每一种可提供的地区保护): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

the target sound source to perform enhancement processing; and outputting the enhanced audio signal to the wearer of the AR glasses by means of an in-ear headphone. By using the solution in the embodiments of the present invention, the enhancement of directional audio can be realized, so as to assist a wearer of AR glasses in better acquiring a sound of a sound source on which the wearer of the AR glasses focuses, thereby reducing the interference of sounds of other sound sources, achieving a certain noise reduction effect and truly implementing the effect of augmented reality of the AR glasses.

(57) 摘要: 本发明实施例公开了一种AR眼镜及其音频增强方法和装置、可读存储介质。所述的AR眼镜的音频增强方法包括：利用麦克风阵列检测AR眼镜佩戴者周围真实世界中的声源分布情况；在AR眼镜的镜片上标记出检测到的每个声源的声源位置；根据AR眼镜佩戴者的眼球注视方向锁定目标声源；从麦克风阵列接收的音频信号中提取与目标声源的声纹特征相关的音频分量进行增强处理；将增强处理后的音频信号通过入耳式耳机输出给AR眼镜佩戴者。采用本发明实施例的方案，可以实现定向音频的强化，以此辅助AR眼镜佩戴者更好的获取其所关注声源的声音，降低其他声源声音的干扰，具有一定的降噪效果，真正实现AR眼镜增强现实的作用。

一种 AR 眼镜及其音频增强方法和装置、可读存储介质

本申请要求于2022年09月30日提交中国专利局、申请号为202211211572.5、发明名称为“一种AR眼镜及其音频增强方法和装置、可读存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本发明涉及增强现实 (Augmented Reality，AR)技术领域，尤其涉及一种 AR 眼镜及其音频增强方法和装置、可读存储介质。

背景技术

AR 技术是利用计算机生成一种逼真的视、听、力、触和动等感觉的虚拟信息，将虚拟信息放在真实世界中展现，并且让人和虚拟信息进行互动。

基于 AR 技术的 AR 眼镜，其佩戴者不仅可以看见真实世界对象，还可以看见产生的虚拟对象，如同在真实世界中存在一样。除了视觉上的增强现实外，AR 眼镜佩戴者还可以通过多个麦克风收音来获取周围真实世界中的环境声音，而且还可以使用虚拟音频来改善真实世界。

针对真实世界中多个声源的嘈杂环境，AR 眼镜佩戴者可能会希望仅获取其中的一个或部分声源的声音，忽略掉其他声源的声音。虽然可以利用定向麦克风优先接收来自位于特定方向上和/或特定距离处的真实声源的声音，同时消除来自其它位置处声源的噪声，但是定向麦克风对声源的最大灵敏度的方向和/或距离不一定对应于用户关注的方向和/或用户关注的距离。因此，存在辅助 AR 眼镜佩戴者更好的获取其所关注声源的声音的需求。

发明内容

本发明实施例提供了一种 AR 眼镜及其音频增强方法和装置、可读存储介质，旨在辅助 AR 眼镜佩戴者更好的获取其所关注声源的声音。

依据本发明的第一方面，提供了一种 AR 眼镜的音频增强方法，包括：

利用麦克风阵列检测 AR 眼镜佩戴者周围真实世界中的声源分布情况；

在 AR 眼镜的镜片上标记出检测到的每个声源的位置；

根据AR眼镜佩戴者的眼球注视方向锁定目标声源；

从麦克风阵列接收的音频信号中提取与所述目标声源的声纹特征相关的音频分量进行增强处理；

将增强处理后的音频信号通过入耳式耳机输出给AR眼镜佩戴者。

依据本发明的第二方面，提供了一种AR眼镜的音频增强装置，包括：

声源分布检测单元，用于利用麦克风阵列检测AR眼镜佩戴者周围真实世界中的声源分布情况；

声源位置标记单元，用于在AR眼镜的镜片上标记出检测到的每个声源的声源位置；

目标声源锁定单元，用于根据AR眼镜佩戴者的眼球注视方向锁定目标声源；

音频增强单元，用于从麦克风阵列接收的音频信号中提取与所述目标声源的声纹特征相关的音频分量进行增强处理；

音频输出单元，用于将增强处理后的音频信号通过入耳式耳机输出给AR眼镜佩戴者。

依据本发明的第三方面，提供了一种AR眼镜，包括麦克风阵列、眼动仪、入耳式耳机、存储器和处理器，所述存储器中存储有计算机程序，所述计算机程序由所述处理器加载并执行，以实现上述的AR眼镜的音频增强方法。

依据本发明的第四方面，提供了一种可读存储介质，所述可读存储介质存储一个或多个计算机程序，所述一个或多个计算机程序当被处理器执行时，实现前述的AR眼镜的音频增强方法。

本发明实施例提供的技术方案可以实现以下有益效果：

本发明实施例提供的AR眼镜及其音频增强方法和装置、可读存储介质，针对真实世界中多个声源的嘈杂环境，通过将AR眼镜佩戴者的眼球注视方向附近区域内的声源声音优化放大，对其他区域声源声音进行抑制消除，可以实现定向音频的强化，以此辅助AR眼镜佩戴者更好的获取其所关注声源的声音，降低其他声源声音的干扰，具有一定的降噪效果，提升AR眼镜佩戴者的用户体验，真正实现AR眼镜增强现实的作用。

附图说明

为了更清楚地说明本发明实施例中的技术方案，下面将对实施例描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明中记载的一些实施例，对于本领域普通技术人员来讲，还可以根据这些附图获得其他的附图。在附图中：

图 1 是本发明实施例提供的 AR 眼镜的音频增强方法的流程示意图；

图 2 是 AR 眼镜上设置 4 个麦克风的一种位置示意图；

图 3 是图 2 所示的 4 个麦克风在 AR 眼镜佩戴时的位置示意图；

图 4 是图 3 所示的 4 个麦克风的平面位置示意图；

图 5 是 AR 眼镜佩戴者周围 3 个声源的分布示意图；

图 6 是图 5 所示 3 个声源的平面位置示意图；

图 7 是图 6 所示场景在声源 1 单独发声时，4 个麦克风（MIC1~4）拾取的声音信号波形示意图；

图 8 是图 6 所示场景下 AR 眼镜佩戴者通过位置移动确定声源位置的示意图；

图 9 是 AR 眼镜佩戴者周围 3 个声源在世界坐标系中的坐标位置示意图；

图 10 是图 9 所示的 3 个声源转换为相机坐标系中的坐标位置示意图；

图 11 是本发明实施例给出的一种在 AR 眼镜镜片上标记声源位置的示意图；

图 12 是利用眼动仪获取眼球注视方向的原理示意图；

图 13 是本发明实施例给出的一种在 AR 眼镜镜片上区别标记目标声源的示意图；

图 14 是本发明实施例提供的 AR 眼镜的音频增强装置的结构示意图；

图 15 是本发明实施例提供的 AR 眼镜的功能模块结构示意图。

具体实施方式

下面将参照附图更详细地描述本发明实施例。提供这些实施例是为了能够更透彻地理解本发明，并且能够将本发明的范围完整的传达给本领域的技术人员。虽然附图中显示了本发明的示例性实施例，然而应当理解，可以以各种形式实现本发明而不应被这里阐述的实施例所限制。

本发明实施例提供了一种 AR 眼镜的音频增强方法。图 1 是本发明实施例提供的 AR 眼镜的音频增强方法的流程示意图。参见图 1 所示，包括步骤 S110~ S150：

步骤 S110，利用麦克风阵列检测 AR 眼镜佩戴者周围真实世界中的声源分布情况。

本步骤是基于双耳定位原理，利用 AR 眼镜上设置的多个麦克风构成的麦克风阵列，来检测 AR 眼镜佩戴者周围真实世界中的声源分布情况。

双耳定位是指双耳听觉所具有的判断声源方位的属性。声源相对于听音者在不同的位置发声时，声波到达听者两耳的强度和时间不同，听觉系统根据这些信息可以确定声源的位置。与此类似，可以根据两个或多个麦克风组成的麦克风阵列获取的声源信息进行计算，以确定声源

位置。下面以 4 个麦克风构成的麦克风阵列为示例对确定声源位置的原理进行说明。

图 2 是 AR 眼镜上设置 4 个麦克风的一种位置示意图。可以理解，AR 眼镜上设置的麦克风数量不限于 4 个，且不局限于图 2 所示的位置设置方式，一般而言，构成麦克风阵列的麦克风数量越多，位置设置的越对称，可以越准确地确定声源位置。

图 3 是图 2 所示的 4 个麦克风在 AR 眼镜佩戴时的位置示意图。图 4 是图 3 所示的 4 个麦克风的平面位置示意图。

假设 AR 眼镜佩戴者周围真实世界中存在 3 个声源，且声源分布情况如图 5 所示。图 5 是 AR 眼镜佩戴者周围 3 个声源的分布示意图。图 6 是图 5 所示 3 个声源的平面位置示意图。

当 AR 眼镜佩戴者在第一位置处时，根据麦克风阵列中每个麦克风拾取的不同强度声音信号，可以获取到每个声源的第一方向线。

以图 6 所示场景为例，图 7 是图 6 所示场景在声源 1 单独发声时，4 个麦克风（MIC1~4）拾取的声音信号波形示意图。当声源 1 单独发声时，4 个麦克风（MIC1~4）可以拾取到不同强度的声音信号，4 个麦克风拾取的声音信号波形如图 7 所示，通过比对声音信号的强度即可定位出声源方向。

在图 7 中可以看到，MIC3 和 MIC4 拾取到的声音信号强度大于 MIC1 和 MIC2，可以初步判断声源 1 的大体位置。再根据 MIC3 和 MIC4 之间的微小差异，可以精确判断出声源 1 在 MIC3 和 MIC4 坐标系中的第一方向线。

即使在多个声源同时发声的情形下，仍可以结合麦克风阵列声源定位算法，例如基于到达时延的声源定位算法,或基于高分辨率谱估计的声源定位算法,或基于波束形成的声源定位算法等，获取到每个声源相对于 AR 眼镜佩戴者所处位置的第一方向线。

当 AR 眼镜佩戴者移动到第二位置处时,根据麦克风阵列中每个麦克风拾取的不同强度声音信号，基于同样的定位算法或原理，可以获取到每个声源的第二方向线。进而根据每个声源的第一方向线和第二方向线的交点位置，确定出每个声源的位置。

图 8 是图 6 所示场景下 AR 眼镜佩戴者通过位置移动确定声源位置的示意图。如图 8 所示，当 AR 眼镜佩戴者在位置 1 处时，可以获取到声源 1 相对于位置 1 的方向线 11，当 AR 眼镜佩戴者在位置 2 处时，可以获取到声源 1 相对于位置 2 的方向线 12。计算两组方向线 11 和 12 的交点位置，即可确定出声源 1 所在位置。

采用上述方法可以确定出检测到的每个声源位置,从而获知 AR 眼镜佩戴者周围真实世界中

的声源分布情况。

步骤 S120，在 AR 眼镜的镜片上标记出检测到的每个声源的声源位置。本步骤 S120 可以具体为：

首先，以 AR 眼镜佩戴者的头部中心为坐标原点建立世界坐标系，对检测到的每个声源可以用(x,y,z)坐标表示其在世界坐标系中的坐标位置。以 AR 眼镜佩戴者周围真实世界中存在 3 个声源的场景为例。图 9 是 AR 眼镜佩戴者周围 3 个声源在世界坐标系中的坐标位置示意图。

接着，以 AR 眼镜佩戴者的瞳孔为坐标原点建立相机坐标系，根据相机标定算法得到的转换公式，将检测到的每个声源在世界坐标系中的坐标位置转换为相机坐标系中的坐标位置。

建立的相机坐标系为平面坐标系，对检测到的每个声源可以用(x,y)坐标表示其在相机坐标系中的坐标位置。图 10 是图 9 所示的 3 个声源转换为相机坐标系中的坐标位置示意图。

然后，在 AR 眼镜的镜片上对检测到的每个声源采用第一标记方式标记出其在相机坐标系中的坐标位置。

在 AR 眼镜的镜片上可以采用多种方式标记出声源位置。图 11 是本发明实施例给出的一种在 AR 眼镜镜片上标记声源位置的示意图。图 11 所示的标记方式是以相机坐标系中声源的坐标位置为中心，显示一个设定半径（例如 2 厘米）的圆形虚线框，将检测到的每个声源的位置在 AR 眼镜的镜片上标记出来。当然，虚线框还可以采用红色、黄色等线条颜色进行醒目标记，圆形框还可以设计成三角形、方形、椭圆形等形状。

步骤 S130，根据 AR 眼镜佩戴者的眼球注视方向锁定目标声源。

本步骤 S130 可以具体为：利用眼动仪获取 AR 眼镜佩戴者的眼球注视方向，并将该眼球注视方向转换成相机坐标系中的坐标位置；当眼球注视方向与某个声源在相机坐标系中的坐标距离小于预设距离值时，将该声源确定为目标声源；在 AR 眼镜的镜片上对所述目标声源采用不同于所述第一标记方式的第二标记方式进行区别标记，以此锁定目标声源。

眼动仪采用的是眼动追踪（Eye-tracking）技术追踪眼睛的运动，通过图像处理技术，定位瞳孔位置，获取瞳孔中心坐标，并通过相应算法，计算人的注视点。

利用眼动仪可以获取 AR 眼镜佩戴者的眼球注视方向。图 12 是利用眼动仪获取眼球注视方向的原理示意图。如图 12 所示，当瞳孔从眼坐标系的原点 O(0,0)点移动到 E(ex,ey)点时，此时可以计算得到 OE 线的方向向量，该方向向量即为 AR 眼镜佩戴者的眼球注视方向。

需要说明的是，眼坐标系(Eye Coordinates)也可以是以 AR 眼镜佩戴者的瞳孔为坐标原点建立，因此该眼坐标系同于相机坐标系。在计算机图形中眼坐标系是一个右手性的正交标准坐标系，并将眼睛解读为正看向这个坐标系的负 Z 轴同时拍摄相片。

对于一个 1920*1080 的相机镜头显示范围来讲，当瞳孔从眼坐标系的原点 O(0,0)点移动到 E(ex,ey)点时，可以通过如下公式将眼球注视方向 OE 转换成相机坐标系中的坐标 C(cx,cy)：

$$\begin{cases} cx=(ex*1920)/EX \\ cy=(ey*1080)/EY \end{cases}$$

其中，EX，EY 分别取值为人眼部的长度值和宽度值，通常情况下这两个值采用的是默认的统计数值。

当$\sqrt{(x'-c_x)^2+(y'-c_y)^2}<m$时，也即眼球注视方向与某个声源在相机坐标系中的坐标距离小于预设距离值 m，此时可以认为眼球注视焦点坐标落在某个声源的范围内，AR 眼镜佩戴者正在注视该声源方向，将该声源确定为目标声源，将其他所有非注视的声源确定为非目标声源。可以理解，通过降低 m 值可以提升目标声源的确定精度。

一旦确定出目标声源，就可以在 AR 眼镜的镜片上区别标记出该目标声源，以此锁定目标声源。

图 13 是本发明实施例给出的一种在 AR 眼镜镜片上区别标记目标声源的示意图。对比图 11 和图 13 所示，声源 1 所在位置处使用的是圆形实线框标记，而其他两个声源所在位置处仍使用的是圆形虚线框标记，也即通过将虚线框变换为实线框区别标记出来声源 1，以此将声源 1 锁定为目标声源。当然，还可以辅以不同颜色、不同形状等方式与未被选中的声源进行区别标记。

步骤 S140，从麦克风阵列接收的音频信号中提取与目标声源的声纹特征相关的音频分量进行增强处理。

本步骤 S140 可以具体为：根据目标声源在相机坐标系中的坐标位置，查找声纹数据库获取目标声源的声纹特征；从麦克风阵列当前接收的音频信号中提取与目标声源的声纹特征相关的音频分量；对提取的音频分量的增益进行放大，和/或对其他未提取的音频分量的增益进行降低或关闭。

声纹(Voiceprint)，是用电声学仪器显示的携带言语信息的声波频谱，是由波长、频率以及强度等百余种特征维度组成的生物特征，具有稳定性、可测量性、唯一性等特点。

为了能够获取目标声源的声纹特征，需要在上述步骤 S110 之后本步骤 S140 之前，对检测到

的每个声源分别提取声纹特征，并将声纹特征与对应的声源位置进行关联，建立声纹数据库。

也即，对检测到的每个声源除了进行定位外，也要记录声音信号，从中提取每个声源的声纹特征，并且将每个声源的声纹特征与其对应的声源位置进行关联，建立声纹数据库，以便后续从混杂的声音信号中提取出特定声源的声纹。

现已有多种技术或手段可以在多个声源环境下提取目标声源的声纹。因此对检测到的每个声源分别提取声纹特征，不是本发明讨论的内容，并且可以采用多种现有技术实现。

将每个声源的声纹特征与其声源位置进行关联，建立的声纹数据库可以以声源为编号采用数组的形式进行保存。锁定目标声源后，可以根据目标声源在相机坐标系中的坐标位置，查找声纹数据库获取目标声源的声纹特征。例如，根据目标声源在相机坐标系中的坐标位置确定声源编号为 1，然后在声纹数据库中查找得到对应声源编号 1 的声纹特征。

接着就可以根据目标声源的声纹特征，对麦克风阵列的音频信号进行处理，从麦克风阵列当前接收的音频信号中提取与目标声源的声纹特征相关的音频分量。然后根据用户设定倍数对提取的音频分量的增益进行放大，也可以同时对其他未提取的音频分量的增益进行降低或关闭，以达到突出目标声源，弱化非目标声源的目的。

步骤 S150，将增强处理后的音频信号通过入耳式耳机输出给 AR 眼镜佩戴者。

AR 眼镜的扬声器或音频接口设置在正常佩戴时靠近人耳的镜腿处，当使用本发明实施例上述的音频增强方法时，需将扬声器拉出变成入耳式耳机进行佩戴，或者通过音频接口外接双声道入耳式耳机进行佩戴。

本步骤 S150 中，AR 眼镜佩戴者为了更好的获取外界真实世界中其所关注的声源的声音，需要隔绝外界环境的声音，不能通过耳朵直接接听扬声器播放的声音，而是要通过入耳式耳机来接听处理后的麦克风阵列的音频信号，实现定向音频的强化。

由以上各步骤说明可知，针对真实世界中多个声源的嘈杂环境，采用本发明实施例的 AR 眼镜的音频增强方法，通过将 AR 眼镜佩戴者的眼球注视方向附近区域内的声源声音优化放大，对其他区域声源声音进行抑制消除，可以实现定向音频的强化，以此辅助 AR 眼镜佩戴者更好的获取其所关注声源的声音，降低其他声源声音的干扰，具有一定的降噪效果，提升 AR 眼镜佩戴者的用户体验，真正实现 AR 眼镜增强现实的作用。

与前述的方法同属于一个技术构思，本发明实施例还提供了一种 AR 眼镜的音频增强装置。

图 14 是本发明实施例提供的 AR 眼镜的音频增强装置的结构示意图，如图 14 所示，本发明实施例的 AR 眼镜的音频增强装置包括：

声源分布检测单元 141，用于利用麦克风阵列检测 AR 眼镜佩戴者周围真实世界中的声源分布情况；

声源位置标记单元 142，用于在 AR 眼镜的镜片上标记出检测到的每个声源的声源位置；

目标声源锁定单元 143，用于根据 AR 眼镜佩戴者的眼球注视方向锁定目标声源；

音频增强单元 144，用于从麦克风阵列接收的音频信号中提取与所述目标声源的声纹特征相关的音频分量进行增强处理；

音频输出单元 145，用于将增强处理后的音频信号通过入耳式耳机输出给 AR 眼镜佩戴者。

在一些实施例中，上述的声源分布检测单元 141 具体用于：

当 AR 眼镜佩戴者在第一位置处时，根据麦克风阵列中每个麦克风拾取的不同强度声音信号，获取检测到的每个声源的第一方向线；当 AR 眼镜佩戴者在第二位置处时，根据麦克风阵列中每个麦克风拾取的不同强度声音信号，获取检测到的每个声源的第二方向线；根据检测到的每个声源的第一方向线和第二方向线的交点位置，确定出检测到的每个声源的位置。

在一些实施例中，上述的声源位置标记单元 142 具体用于：

以 AR 眼镜佩戴者的头部中心为坐标原点建立世界坐标系，对检测到的每个声源确定出其在世界坐标系中的坐标位置；以 AR 眼镜佩戴者的瞳孔为坐标原点建立相机坐标系，根据相机标定算法得到的转换公式，将检测到的每个声源在世界坐标系中的坐标位置转换为相机坐标系中的坐标位置；在 AR 眼镜的镜片上对检测到的每个声源采用第一标记方式标记出其在相机坐标系中的坐标位置。

在一些实施例中，上述的目标声源锁定单元 143 具体用于：

利用眼动仪获取 AR 眼镜佩戴者的眼球注视方向，并将所述眼球注视方向转换成相机坐标系中的坐标位置；当所述眼球注视方向与某个声源在相机坐标系中的坐标距离小于预设距离值时，将所述声源确定为目标声源；在 AR 眼镜的镜片上对所述目标声源采用不同于所述第一标记方式的第二标记方式进行区别标记，以此锁定所述目标声源。

在一些实施例中，如图 14 所示，本发明实施例的 AR 眼镜的音频增强装置还可以包括：

声纹特征提取单元 146，用于对所述声源分布检测单元 141 检测到的每个声源分别提取声纹特征，并将声纹特征与对应的声源位置进行关联，建立声纹数据库。

在一些实施例中，上述的音频增强单元144具体用于：

根据所述目标声源在相机坐标系中的坐标位置，查找所述声纹数据库获取所述目标声源的声纹特征；从麦克风阵列当前接收的音频信号中提取与所述目标声源的声纹特征相关的音频分量；对提取的音频分量的增益进行放大，和/或对其他未提取的音频分量的增益进行降低或关闭。

本发明实施例的AR眼镜的音频增强装置中的各个模块或单元的实现过程，可以参见上述方法实施例，在此不再赘述。

与前述的AR眼镜的音频增强方法同属于一个技术构思，本发明实施例还提供了一种AR眼镜。图15是本发明实施例提供的AR眼镜的功能模块结构示意图。参见图15，本发明实施例提供的AR眼镜包括：麦克风阵列、眼动仪、入耳式耳机、存储器和处理器，其中，存储器可以是内存，例如高速随机存取存储器(Random-Access Memory，RAM)，也可以是非易失性存储器（non-volatile memory），例如至少一个磁盘存储器等。存储器中存储有计算机程序，该计算机程序由处理器加载并执行，以实现前述的AR眼镜的音频增强方法。

在硬件层面上，该AR眼镜还可以选择性地包括通信模块等。扬声器、入耳式耳机、麦克风阵列、眼动仪、存储器、处理器、通信模块等可以通过内部总线相互连接，该内部总线可以是ISA(Industry Standard Architecture，工业标准体系结构）总线、PCI(Peripheral Component Interconnect，外设部件互连标准)总线或EISA(Extended Industry Standard Architecture，扩展工业标准结构）总线等。总线可以分为地址总线、数据总线、控制总线等。为便于表示，图15中仅用一个双向箭头表示，但并不表示仅有一根总线或一种类型的总线。

最后，本发明实施例还提出了一种可读存储介质，该可读存储介质存储一个或多个计算机程序，该一个或多个计算机程序当被处理器执行时，实现前述的AR眼镜的音频增强方法。

可读存储介质包括永久性和非永久性、可移动和非可移动媒体，可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。可读存储介质的例子包括，但不限于相变内存 (PRAM)、静态随机存取存储器 (SRAM)、动态随机存取存储器 (DRAM)、其他类型的随机存取存储器 (RAM)、只读存储器 (ROM)、电可擦除可编程只读存储器 (EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器 (CD-ROM)、数字多功能光盘 (DVD) 或其他光学存储、磁盒式磁带，磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质，可用于存储可以被计算设备访问的信息。

本领域内的技术人员应明白，本发明的方案可提供为方法、装置、或计算机程序产品。因此，本发明可采用完全硬件实施例、完全软件实施例、或软件和硬件相结合实施例的形式。而且，本发明可采用在一个或多个包含有计算机程序的可读存储介质上实施的计算机程序产品的形式。

还需要说明的是，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、商品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、商品或者设备所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括要素的过程、方法、商品或者设备中还存在另外的相同要素。

以上仅为本发明的实施例而已，并不用于限制本发明。对于本领域技术人员来说，本发明可以有各种更改和变化。凡在本发明的精神和原理之内所作的任何修改、等同替换、改进等，均应包含在本发明的权利要求范围之内。

权 利 要 求

1.一种 AR 眼镜的音频增强方法，其特征在于，包括：

利用麦克风阵列检测 AR 眼镜佩戴者周围真实世界中的声源分布情况；

在 AR 眼镜的镜片上标记出检测到的每个声源的声源位置；

根据 AR 眼镜佩戴者的眼球注视方向锁定目标声源；

从麦克风阵列接收的音频信号中提取与所述目标声源的声纹特征相关的音频分量进行增强处理；

将增强处理后的音频信号通过入耳式耳机输出给 AR 眼镜佩戴者。

2.根据权利要求 1 所述方法，其特征在于，所述利用麦克风阵列检测 AR 眼镜佩戴者周围真实世界中的声源分布情况包括：

当 AR 眼镜佩戴者在第一位置处时，根据麦克风阵列中每个麦克风拾取的不同强度声音信号，获取检测到的每个声源的第一方向线；

当 AR 眼镜佩戴者在第二位置处时，根据麦克风阵列中每个麦克风拾取的不同强度声音信号，获取检测到的每个声源的第二方向线；

根据检测到的每个声源的第一方向线和第二方向线的交点位置，确定出检测到的每个声源的位置。

3.根据权利要求 1 所述方法，其特征在于，所述在 AR 眼镜的镜片上标记出检测到的每个声源的声源位置包括：

以 AR 眼镜佩戴者的头部中心为坐标原点建立世界坐标系，对检测到的每个声源确定出其在世界坐标系中的坐标位置；

以 AR 眼镜佩戴者的瞳孔为坐标原点建立相机坐标系，根据相机标定算法得到的转换公式，将检测到的每个声源在世界坐标系中的坐标位置转换为相机坐标系中的坐标位置；

在 AR 眼镜的镜片上对检测到的每个声源标记出其在相机坐标系中的坐标位置。

4.根据权利要求 3 所述方法，其特征在于，所述根据 AR 眼镜佩戴者的眼球注视方向锁定目标声源包括：

利用眼动仪获取 AR 眼镜佩戴者的眼球注视方向，并将所述眼球注视方向转换成相机坐标系中的坐标位置；

当所述眼球注视方向与某个声源在相机坐标系中的坐标距离小于预设距离值时，将所述声

源确定为目标声源；

在AR眼镜的镜片上对所述目标声源进行区别标记，以此锁定所述目标声源。

5.根据权利要求1-4任一项所述方法，其特征在于，所述方法还包括：对检测到的每个声源分别提取声纹特征，并将声纹特征与对应的声源位置进行关联，建立声纹数据库。

6.根据权利要求5所述方法，其特征在于，所述从麦克风阵列接收的音频信号中提取与所述目标声源的声纹特征相关的音频分量进行增强处理包括：

根据所述目标声源在相机坐标系中的坐标位置，查找所述声纹数据库获取所述目标声源的声纹特征；

从麦克风阵列当前接收的音频信号中提取与所述目标声源的声纹特征相关的音频分量；

对提取的音频分量的增益进行放大，和/或对其他未提取的音频分量的增益进行降低或关闭。

7.一种AR眼镜的音频增强装置，其特征在于，包括：

声源分布检测单元，用于利用麦克风阵列检测AR眼镜佩戴者周围真实世界中的声源分布情况；

声源位置标记单元，用于在AR眼镜的镜片上标记出检测到的每个声源的声源位置；

目标声源锁定单元，用于根据AR眼镜佩戴者的眼球注视方向锁定目标声源；

音频增强单元，用于从麦克风阵列接收的音频信号中提取与所述目标声源的声纹特征相关的音频分量进行增强处理；

音频输出单元，用于将增强处理后的音频信号通过入耳式耳机输出给AR眼镜佩戴者。

8.根据权利要求7所述装置，其特征在于，所述装置还包括：

声纹特征提取单元，用于对所述声源分布检测单元检测到的每个声源分别提取声纹特征，并将声纹特征与对应的声源位置进行关联，建立声纹数据库。

9.一种AR眼镜，包括麦克风阵列、眼动仪、入耳式耳机、存储器和处理器，所述存储器中存储有计算机程序，所述计算机程序由所述处理器加载并执行，以实现权利要求1~6任一项所述的AR眼镜的音频增强方法。

10. 一种可读存储介质，所述可读存储介质存储一个或多个计算机程序，所述一个或多个计算机程序当被处理器执行时，实现权利要求1~6任一项所述的AR眼镜的音频增强方法。

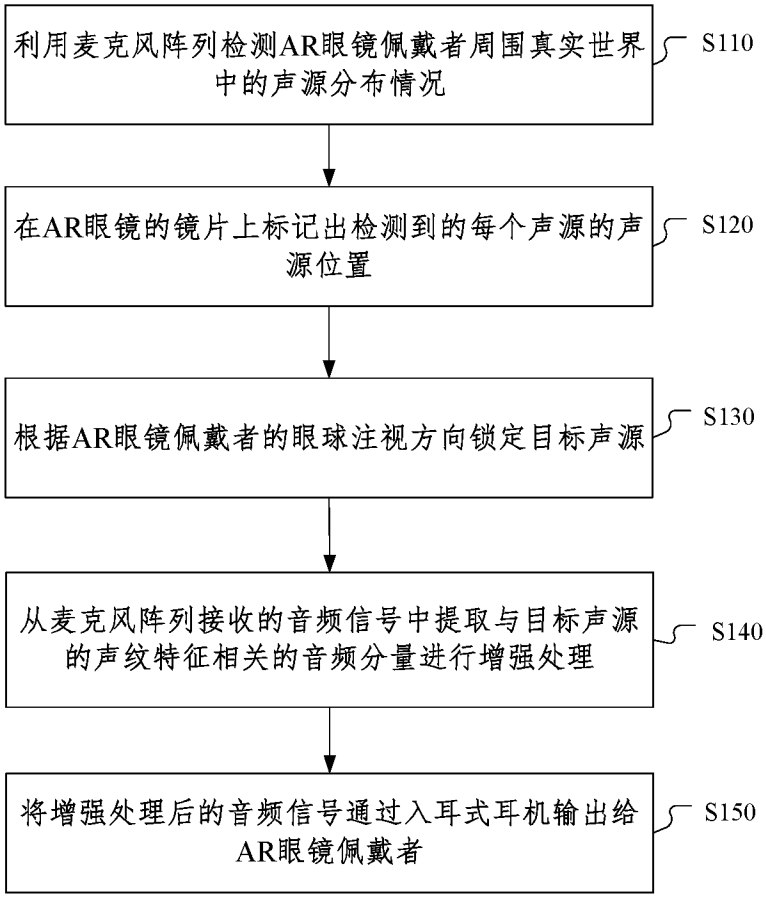
利用麦克风阵列检测AR眼镜佩戴者周围真实世界中的声源分布情况
S110
在AR眼镜的镜片上标记出检测到的每个声源的声源位置
S120
根据AR眼镜佩戴者的眼球注视方向锁定目标声源
S130
从麦克风阵列接收的音频信号中提取与目标声源的声纹特征相关的音频分量进行增强处理
S140
将增强处理后的音频信号通过入耳式耳机输出给AR眼镜佩戴者
S150

图 1

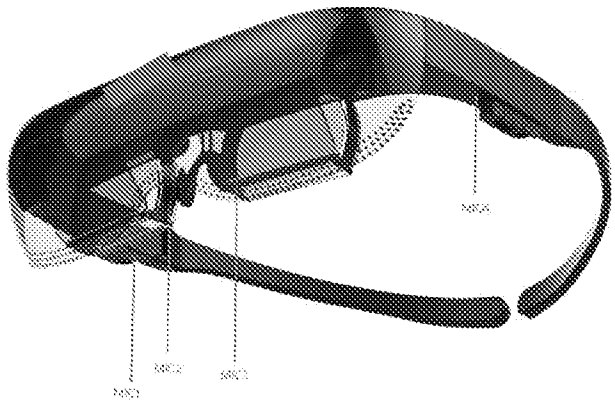

图 2

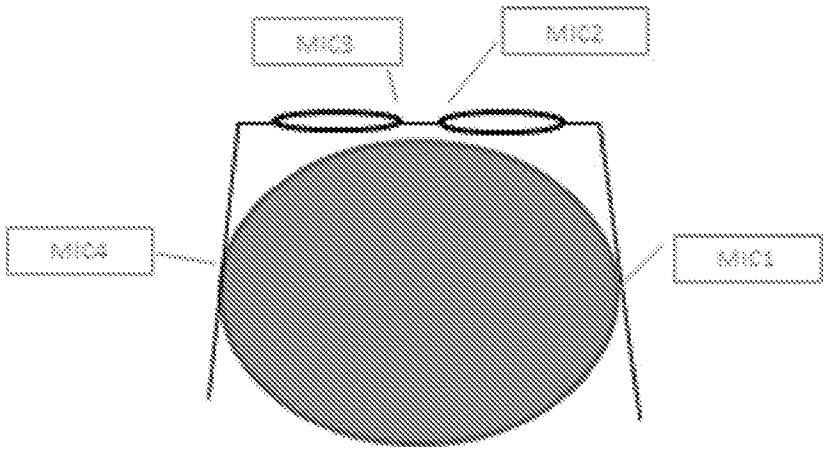

图 3

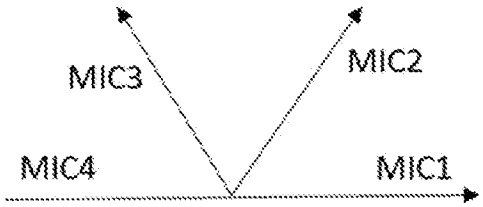

图 4

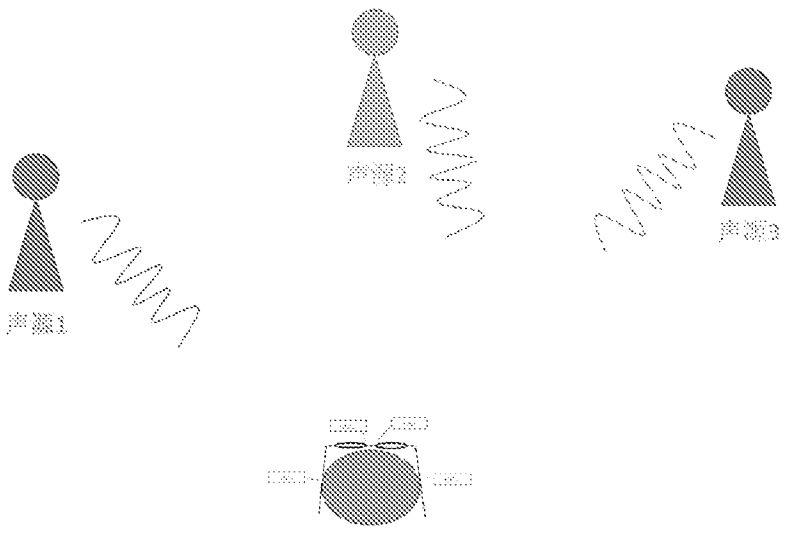

图 5

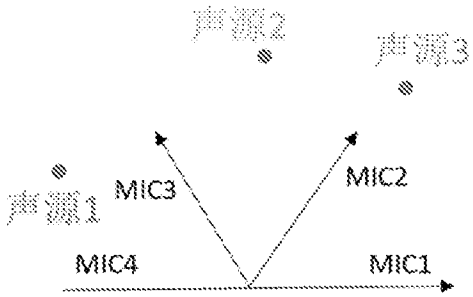
声源2
声源3
声源1
MIC3
MIC2
MIC4
MIC1

图 6

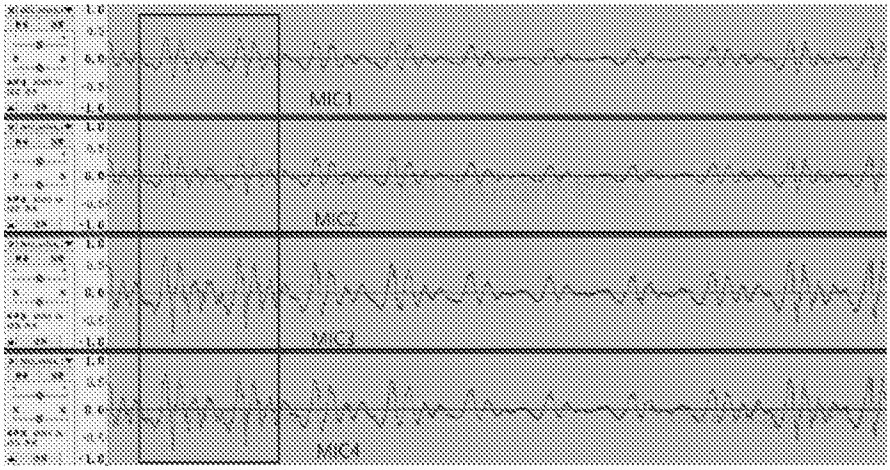
MIC1
MIC2
MIC3
MIC4

图 7

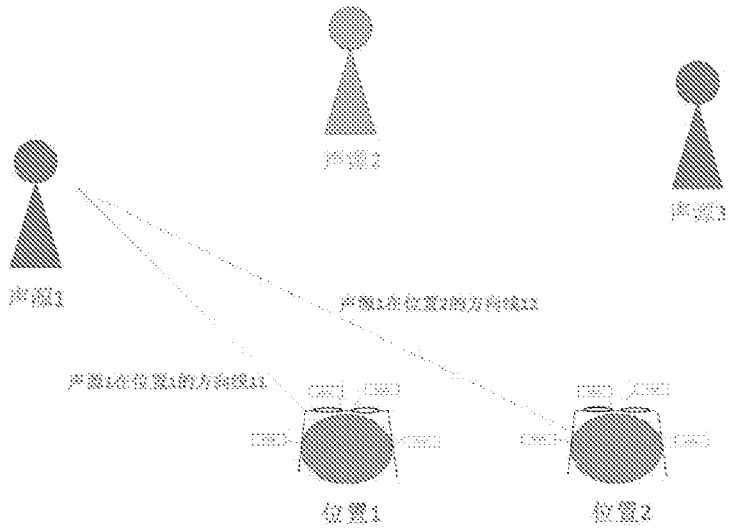
声源2
声源3
声源1
位置1
位置2

图 8

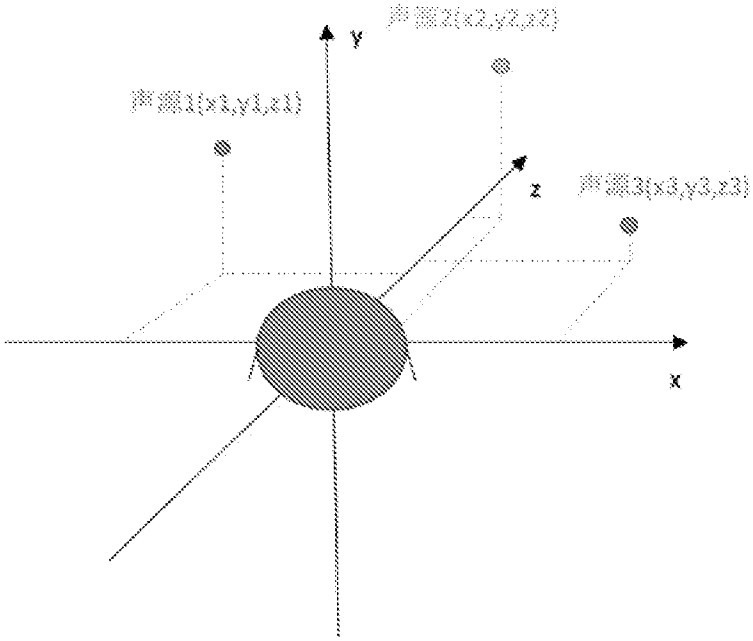
y
声源2(x2,y2,z2)
声源1(x1,y1,z1)
z
声源3(x3,y3,z3)
x

图 9

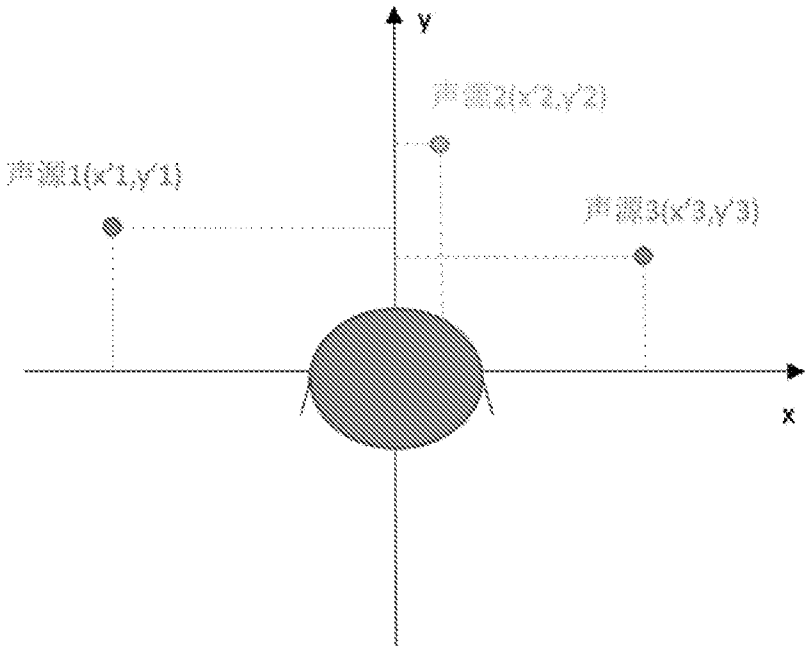
y
声源2(x'2,y'2)
声源1(x'1,y'1)
声源3(x'3,y'3)
x

图 10

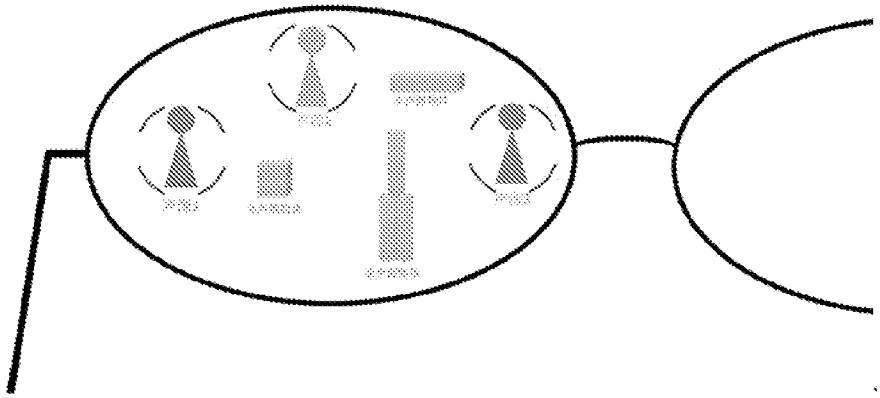

图 11

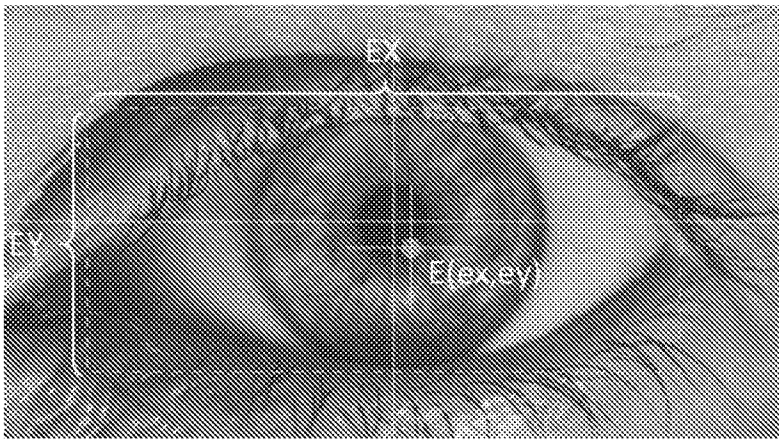

图 12

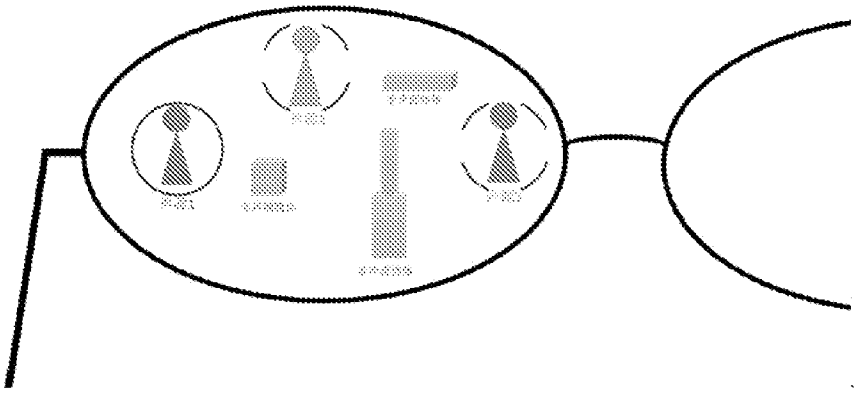

图 13

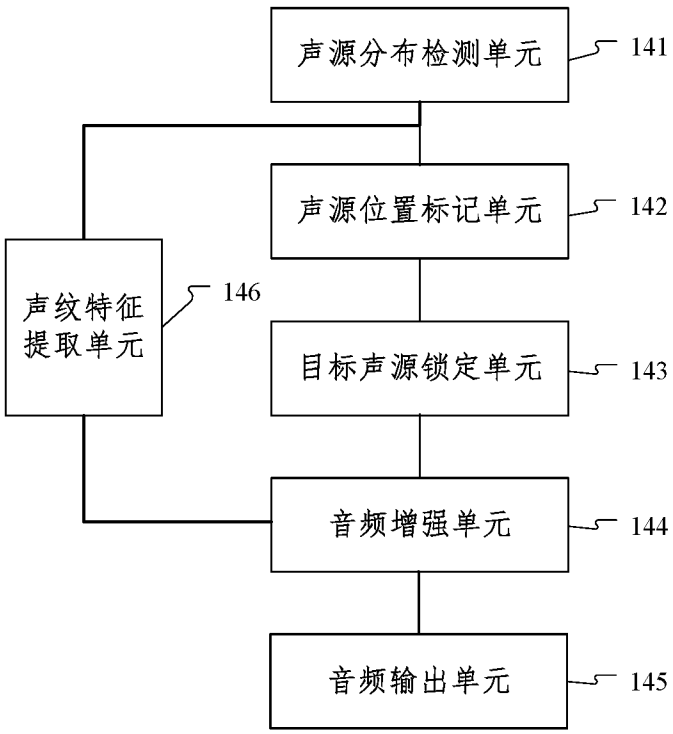
声源分布检测单元
141
声源位置标记单元
142
声纹特征提取单元
146
目标声源锁定单元
143
音频增强单元
144
音频输出单元
145

图 14

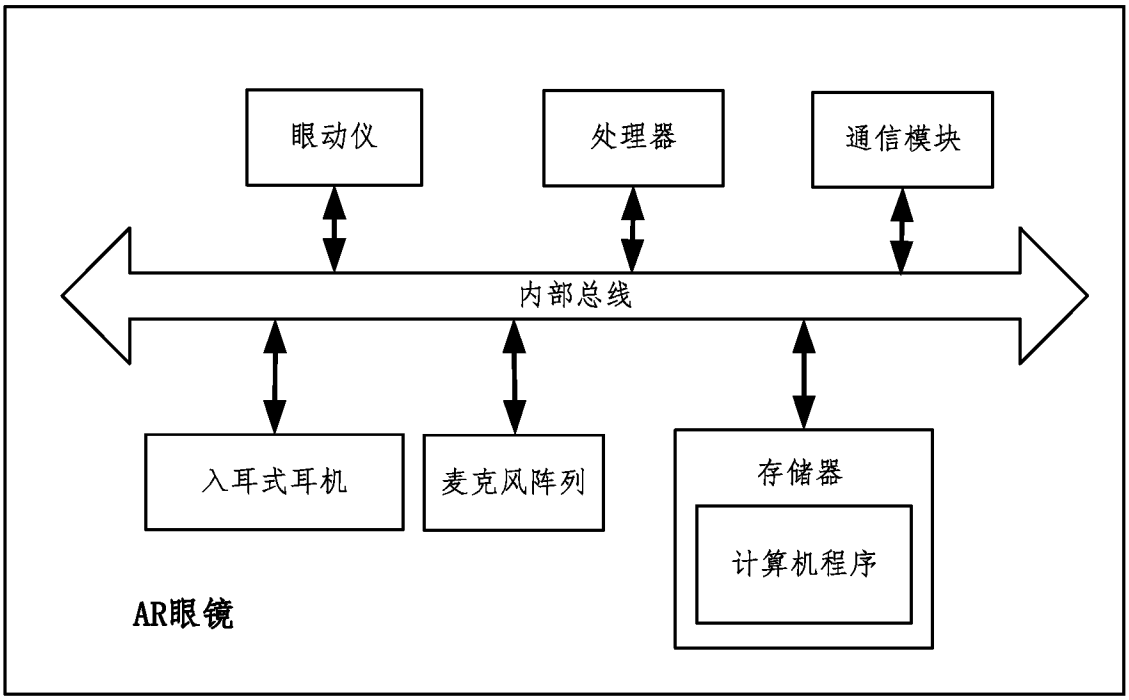
眼动仪
处理器
通信模块
内部总线
入耳式耳机
麦克风阵列
存储器
计算机程序
AR眼镜

图 15

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2023/111770

A. CLASSIFICATION OF SUBJECT MATTER

G06F3/01(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC:G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS; CNTXT; VEN; ENTXT; EPTXT; WOTXT; CNKI: 虚拟现实, 智能, 眼镜, 音频, 增强, 声源, 音源, 显示, 标记, 视线, 凝视, 世界坐标, 目标, 对象, virtual reality, smart, glasses, audio, augmentation, sound source, voice source, display, marker, sight line, gaze, worldcoordinate, target, object

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 116126132 A (GOERTEK TECHNOLOGY CO., LTD.) 16 May 2023 (2023-05-16) claims 1-10	1-10
Y	US 2021281956 A1 (IBM) 09 September 2021 (2021-09-09) description, paragraphs 17-84, and figures 1-3	1-10
Y	CN 112986914 A (AUTOMATION RESEARCH INSTITUTE OF CHINA SOUTH INDUSTRIES GROUP CORPORATION) 18 June 2021 (2021-06-18) description, paragraphs 52-90, and figure 2	1-10
Y	US 2016080874 A1 (FULLAM, Scott) 17 March 2016 (2016-03-17) description, paragraphs 9-63	1-10
A	CN 113596673 A (NINGBO QIXIN ELECTRONIC TECHNOLOGY CO., LTD.) 02 November 2021 (2021-11-02) entire document	1-10
A	CN 112669578 A (SHADOW CREATOR INFORMATION TECHNOLOGY CO., LTD.) 16 April 2021 (2021-04-16) entire document	1-10

☑ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"D" document cited by the applicant in the international application
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
05 November 2023	**09 November 2023**

Name and mailing address of the ISA/CN	Authorized officer
China National Intellectual Property Administration (ISA/CN) **China No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088**	Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2023/111770

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2021174823 A1 (SPECTRUM ACCOUNTABLE CARE COMPANY) 10 June 2021 (2021-06-10) entire document	1-10

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2023/111770

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	116126132	A	16 May 2023	None			
US	2021281956	A1	09 September 2021	DE	112021001516	T5	05 January 2023
				US	11134349	B1	28 September 2021
				AU	2021235335	A1	04 August 2022
				JP	2023519495	A	11 May 2023
				KR	20220120633	A	30 August 2022
				WO	2021181232	A1	16 September 2021
				CN	115136619	A	30 September 2022
				GB	2614121	A	28 June 2023
CN	112986914	A	18 June 2021	None			
US	2016080874	A1	17 March 2016	WO	2016044034	A1	24 March 2016
CN	113596673	A	02 November 2021	None			
CN	112669578	A	16 April 2021	CN	112669578	B	19 April 2022
US	2021174823	A1	10 June 2021	None			

国际检索报告

国际申请号
PCT/CN2023/111770

A. 主题的分类

G06F3/01(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

IPC:G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称，和使用的检索词（如使用）)

CNABS;CNTXT;VEN;ENTXT;EPTXT;WOTXT;CNKI:虚拟现实，智能，眼镜，音频，增强，声源，音源，显示，标记，视线，凝视，世界坐标，目标，对象，virtual reality,smart, glasses,audio,augmentation,sound source,voice source,display,marker,sight line, gaze,worldcoordinate,target,object

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
PX	CN 116126132 A (歌尔科技有限公司) 2023年5月16日 (2023 - 05 - 16) 权利要求1-10	1-10
Y	US 2021281956 A1 (IBM) 2021年9月9日 (2021 - 09 - 09) 说明书第17-84段，图1-3	1-10
Y	CN 112986914 A (中国兵器装备集团自动化研究所) 2021年6月18日 (2021 - 06 - 18) 说明书第52-90段，图2	1-10
Y	US 2016080874 A1 (FULLAM,Scott) 2016年3月17日 (2016 - 03 - 17) 说明书第9-63段	1-10
A	CN 113596673 A (宁波旗芯电子科技有限公司) 2021年11月2日 (2021 - 11 - 02) 全文	1-10
A	CN 112669578 A (上海影创信息科技有限公司) 2021年4月16日 (2021 - 04 - 16) 全文	1-10

☑ 其余文件在C栏的续页中列出。

☑ 见同族专利附件。

* 引用文件的具体类型：

"A" 认为不特别相关的表示了现有技术一般状态的文件

"D" 申请人在国际申请中引证的文件

"E" 在国际申请日的当天或之后公布的在先申请或专利

"L" 可能对优先权要求构成怀疑的文件，或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件（如具体说明的）

"O" 涉及口头公开、使用、展览或其他方式公开的文件

"P" 公布日先于国际申请日但迟于所要求的优先权日的文件

"T" 在申请日或优先权日之后公布，与申请不相抵触，但为了理解发明之理论或原理的在后文件

"X" 特别相关的文件，单独考虑该文件，认定要求保护的发明不是新颖的或不具有创造性

"Y" 特别相关的文件，当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时，要求保护的发明不具有创造性

"&" 同族专利的文件

国际检索实际完成的日期	国际检索报告邮寄日期
2023年11月5日	2023年11月9日
ISA/CN的名称和邮寄地址 中国国家知识产权局 中国北京市海淀区蓟门桥西土城路6号 100088	受权官员 谭岳峰 电话号码 (+86) 020-28958950

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
A	US 2021174823 A1 (SPECTRUM ACCOUNTABLE CARE COMPANY) 2021年6月10日 (2021 - 06 - 10) 全文	1-10

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2023/111770

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	116126132	A	2023年5月16日	无			
US	2021281956	A1	2021年9月9日	DE	112021001516	T5	2023年1月5日
				US	11134349	B1	2021年9月28日
				AU	2021235335	A1	2022年8月4日
				JP	2023519495	A	2023年5月11日
				KR	20220120633	A	2022年8月30日
				WO	2021181232	A1	2021年9月16日
				CN	115136619	A	2022年9月30日
				GB	2614121	A	2023年6月28日
CN	112986914	A	2021年6月18日	无			
US	2016080874	A1	2016年3月17日	WO	2016044034	A1	2016年3月24日
CN	113596673	A	2021年11月2日	无			
CN	112669578	A	2021年4月16日	CN	112669578	B	2022年4月19日
US	2021174823	A1	2021年6月10日	无			