

一种合唱音频的处理方法、设备及存储介质

本申请要求于 2021 年 04 月 27 日提交中国专利局、申请号为 202110460280.4、发明名称为“一种合唱音频的处理方法、设备及存储介质”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。

5 技术领域

本申请涉及计算机应用技术领域，特别是涉及一种合唱音频的处理方法、设备及存储介质。

背景技术

10 随着计算机技术的快速发展，音频类、视频类、办公类等各类软件逐渐增多，给人们的生活带来了很多便利。使用音频类软件，用户可以进行听歌、唱歌等体验。

目前，为了给用户提供演唱会大合唱的听觉体验，多是将多人歌唱数据进行直接叠加处理。但是，这种经过简单的叠加处理得到的音频，在听
15 感上，声场聚集在人头中心，具有头中效应，声场不够宽阔，听觉体验较差。

发明内容

本申请的目的是提供一种合唱音频的处理方法、设备及存储介质，以
20 避免声场聚集在人头中心而产生头中效应，使得声场更加宽阔，提升听觉体验。

为解决上述技术问题，本申请提供如下技术方案：

一种合唱音频的处理方法，包括：

分别获得多个演唱者对同一目标歌曲进行演唱的干声音频；

25 对获得的多个所述干声音频进行时间对齐处理；

对进行时间对齐处理后的多个所述干声音频进行虚拟声像定位，以将多个所述干声音频定位到多个虚拟声像上；其中，多个所述虚拟声像位于预先建立的虚拟声像坐标系中，所述虚拟声像坐标系以人头为中心，以左

右耳所在直线中点为坐标原点，第一坐标轴的正方向表示人头正前方，第二坐标轴的正方向表示人头从左耳到右耳的侧方，第三坐标轴的正方向表示人头正上方，每个所述虚拟声像与所述坐标原点的距离在设定距离范围内，每个所述虚拟声像相对于所述第一坐标轴和所述第二坐标轴构成的平面的俯仰角在设定角度范围内；

- 5 基于进行虚拟声像定位后的多个所述干声音频，生成合唱音频；
在获取到基于所述目标歌曲演唱的主唱音频的情况下，将所述主唱音频、所述合唱音频和相应的伴奏进行合成后，输出大合唱效果音频。

- 在本申请的一种具体实施方式中，所述对获得的多个所述干声音频进行时间对齐处理，包括：

- 确定所述目标歌曲对应的参考音频；
针对获得的每个所述干声音频，分别提取当前干声音频和所述参考音频的音频特征，所述音频特征为指纹特征或基频特征；
将所述当前干声音频与所述参考音频的音频特征相似度最大值对应的
15 时间确定为音频对齐时间；

基于所述音频对齐时间，对所述当前干声音频进行时间对齐处理。

在本申请的一种具体实施方式中，还包括：

- 分别对获得的多个所述干声音频进行带通滤波处理，得到多个低音数据；
20 相应的，所述基于进行虚拟声像定位后的多个所述干声音频，生成合唱音频，包括：

基于进行虚拟声像定位后的多个所述干声音频和多个所述低音数据，生成合唱音频。

在本申请的一种具体实施方式中，还包括：

- 25 分别对获得的多个所述干声音频进行混响模拟处理；
相应的，所述基于进行虚拟声像定位后的多个所述干声音频，生成合唱音频，包括：

基于进行虚拟声像定位后的多个所述干声音频和混响模拟处理后的多个所述干声音频，生成合唱音频。

在本申请的一种具体实施方式中，所述分别对获得的多个所述干声音频进行混响模拟处理，包括：

利用梳状滤波器和全通滤波器的级联分别对获得的多个所述干声音频进行混响模拟处理。

- 5 在本申请的一种具体实施方式中，在所述对进行时间对齐处理后的多个所述干声音频进行虚拟声像定位之后，还包括：

分别对进行虚拟声像定位后的多个所述干声音频进行混响模拟处理；

相应的，所述基于进行虚拟声像定位后的多个所述干声音频，生成合唱音频，包括：

- 10 基于进行虚拟声像定位，且进行混响模拟处理后的多个所述干声音频，生成合唱音频。

在本申请的一种具体实施方式中，还包括：

分别对获得的多个所述干声音频进行双声道模拟处理；

- 15 相应的，所述基于进行虚拟声像定位后的多个所述干声音频，生成合唱音频，包括：

基于进行虚拟声像定位后的多个所述干声音频和双声道模拟处理后的多个所述干声音频，生成合唱音频。

在本申请的一种具体实施方式中，在所述分别对获得的多个所述干声音频进行双声道模拟处理之后，还包括：

- 20 对进行双声道模拟处理后的多个所述干声音频进行混响模拟处理；

相应的，所述基于进行虚拟声像定位后的多个所述干声音频和双声道模拟处理后的多个所述干声音频，生成合唱音频，包括：

基于进行虚拟声像定位后的多个所述干声音频、双声道模拟处理及混响模拟处理后的多个所述干声音频，生成合唱音频。

- 25 在本申请的一种具体实施方式中，所述对进行时间对齐处理后的多个所述干声音频进行虚拟声像定位，包括：

按照虚拟声像的个数，将获得的进行时间对齐处理后的多个所述干声音频进行分组，组数与虚拟声像的个数相同；

将各组干声音频分别定位到对应的虚拟声像上，不同组干声音频对应

不同虚拟声像。

在本申请的一种具体实施方式中，

- 多个所述虚拟声像中，位于人头后方的虚拟声像相对于所述第一坐标轴和所述第二坐标轴构成的平面的仰角大于位于人头前方的虚拟声像相对于所述第一坐标轴和所述第二坐标轴构成的平面的仰角；

或者，

每个所述虚拟声像均匀分布在所述第一坐标轴和所述第二坐标轴构成的平面的一周。

- 在本申请的一种具体实施方式中，所述将所述主唱音频、所述合唱音频和相应的伴奏进行合成，包括：

对所述主唱音频和所述合唱音频分别进行音量调整，和/或，对所述主唱音频和所述合唱音频进行混响模拟处理；

将进行音量调整和/或进行混响模拟处理后的所述主唱音频、所述合唱音频和相应的伴奏进行合成。

- 一种合唱音频的处理装置，其特征在于，包括：

干声音频获得模块，用于分别获得多个演唱者对同一目标歌曲进行演唱的干声音频；

对齐处理模块，用于对获得的多个所述干声音频进行时间对齐处理；

- 虚拟声像定位模块，用于对进行时间对齐处理后的多个所述干声音频进行虚拟声像定位，以将多个所述干声音频定位到多个虚拟声像上；其中，多个所述虚拟声像位于预先建立的虚拟声像坐标系中，所述虚拟声像坐标系以人头为中心，以左右耳所在直线中点为坐标原点，第一坐标轴的正方向表示人头正前方，第二坐标轴的正方向表示人头从左耳到右耳的侧方，第三坐标轴的正方向表示人头正上方，每个所述虚拟声像与所述坐标原点的距离在设定距离范围内，每个所述虚拟声像相对于所述第一坐标轴和所述第二坐标轴构成的平面的俯仰角在设定角度范围内；

合唱音频生成模块，用于基于进行虚拟声像定位后的多个所述干声音频，生成合唱音频；

大合唱效果音频输出模块，用于在获取到基于所述目标歌曲演唱的主

唱音频的情况下，将所述主唱音频、所述合唱音频和相应的伴奏进行合成后，输出大合唱效果音频。

一种合唱音频的处理设备，包括：

存储器，用于存储计算机程序；

5 处理器，用于执行所述计算机程序时实现上述任一项所述的合唱音频的处理方法的步骤。

一种计算机可读存储介质，所述计算机可读存储介质上存储有计算机程序，所述计算机程序被处理器执行时实现上述任一项所述的合唱音频的处理方法的步骤。

10 应用本申请实施例所提供的技术方案，分别获得多个演唱者对同一目标歌曲进行演唱的干声音频后，对获得的多个干声音频进行时间对齐处理，并对对齐后的多个干声音频进行虚拟声像定位，以将多个干声音频定位到多个虚拟声像上，多个虚拟声像位于以人头为中心的虚拟声像坐标系中，与坐标原点的距离在设定距离范围内，环绕人耳，基于虚拟声像定位后的
15 多个干声音频，生成合唱音频，并在获取到基于目标歌曲演唱的主唱音频的情况下，将主唱音频、合唱音频和相应的伴奏进行合唱，得到并输出大合唱效果音频。将多个干声音频定位到环绕人耳的多个虚拟声像上，可以使得生成的合唱音频具有声场环绕音效，在听感上，可以有效避免最终输出的大合唱效果音频的声场聚集在人头中心而产生的头中效应，使得声场
20 更加宽阔。

附图说明

为了更清楚地说明本申请实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，
25 下面描述中的附图仅仅是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图1为本申请实施例中一种合唱音频的处理方法的实施流程图；

图2为本申请实施例中虚拟声像定位坐标系展示声像方位的示意图；

- 图3为本申请实施例中一种虚拟声像定位示意图；
图4为本申请实施例中定位后的虚拟声像的示意图；
图5为本申请实施例中一种空间声场过程组成示意图；
图6为本申请实施例中一种梳状滤波器和全通滤波器的级联形式示意图；
5 图；
图7为本申请实施例中一种混响脉冲响应示意图；
图8为本申请实施例中一种双声道模拟过程示意图；
图9为本申请实施例中一种合唱音频的处理系统的框架示意图；
图10为本申请实施例中一种合唱音频的处理系统的具体结构示意图；
10 图11为本申请实施例中一种合唱音频的处理装置的结构示意图；
图12为本申请实施例中一种合唱音频的处理设备的结构示意图。

具体实施方式

本申请的核心是提供一种合唱音频的处理方法。分别获得多个演唱者
15 对同一目标歌曲进行演唱的干声音频后，对获得的多个干声音频进行时间
对齐处理，并对对齐后的多个干声音频进行虚拟声像定位，以将多个干声
音频定位到多个虚拟声像上，多个虚拟声像位于以人头为中心的虚拟声像
坐标系中，与坐标原点的距离在设定距离范围内，环绕人耳，基于虚拟声
像定位后的多个干声音频，生成合唱音频，并在获取到基于目标歌曲演唱
20 的主唱音频的情况下，将主唱音频、合唱音频和相应的伴奏进行合唱，得
到并输出大合唱效果音频。将多个干声音频定位到环绕人耳的多个虚拟声
像上，可以使得生成的合唱音频具有声场环绕音效，在听感上，可以有效
避免最终输出的大合唱效果音频的声场聚集在人头中心而产生的头中效
应，使得声场更加宽阔。

25 在实际应用中，本申请实施例所提供的方法可以应用在想要获得大合
唱音效的各种场景中，可以通过服务器与客户端的交互进行具体方案的实
施。

举例而言，在场景 1 中，服务器可以预先获得多个演唱者，如演唱者
1、2、3、4...对同一目标歌曲进行演唱的干声音频，对获得的干声音频

进行时间对齐处理，并在对齐后对多个干声音频进行虚拟声像定位，将多个干声音频定位到多个虚拟声像上，多个虚拟声像可以环绕人耳，基于虚拟声像定位后的多个干声音频，生成合唱音频。当用户 X 想要使得自己演唱的歌曲实现大合唱音效时，可以通过客户端对目标歌曲进行演唱，服务器通过客户端得到用户 X 演唱的主唱音频，将主唱音频、合唱音频和相应的伴奏进行合成，可以得到大合唱效果音频，将大合唱效果音频通过客户端输出出来，可以使得用户 X 感受到大合唱音效。

在场景 2 中，几个好朋友（用户 1、2、3、4、5）在同一时间段但不同空间对目标歌曲进行演唱，想要达到大合唱音效。在任意一个用户的角度来看，可以将当前用户作为主唱。如在用户 1 的角度，服务器可以分别获得用户 2、3、4、5 对目标歌曲进行演唱的干声音频，对获得的干声音频进行时间对齐处理，对齐后将多个干声音频定位到多个虚拟声像上，多个虚拟声像环绕人耳，基于虚拟声像定位后的多个干声音频，生成合唱音频。服务器在获取到用户 1 通过客户端基于目标歌曲演唱的主唱音频的情况下，将主唱音频、合唱音频和相应的伴奏进行合成，得到大合唱效果音频，通过客户端输出给用户 1，这样用户 1 可以感受到大合唱音效。

上述仅为示例性地描述了应用场景，在实际应用中，本申请技术方案还可以应用于更多的场景，如多人合唱、多人小乐队等的音效处理场景。

为了使本技术领域的人员更好地理解本申请方案，下面结合附图和具体实施方式对本申请作进一步的详细说明。显然，所描述的实施例仅仅是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

参见图 1 所示，为本申请实施例所提供的一种合唱音频的处理方法的实施流程图，该方法可以包括以下步骤：

S110：分别获得多个演唱者对同一目标歌曲进行演唱的干声音频。

在本申请实施例中，可以根据实际需要获得多个干声音频。多个干声音频可以是不同演唱者对于同一目标歌曲演唱得到的音频数据，不同演唱者可以处于相同或不同环境中。

S120: 对获得的多个干声音频进行时间对齐处理。

分别获得多个演唱者对同一目标歌曲进行演唱的干声音频，因为多个干声音频可能为不同演唱者在不同时间演唱的，可能存在延迟等不对齐现象。为了后续达到较好的大合唱音效，可以先对获得的多个干声音频进行时间对齐处理，使得进行时间对齐处理后的干声音频不存在严重抢拍或慢拍，如提前或滞后 1 秒以上的音频。具体的，可以利用对齐工具对获得的多个干声音频最相同起始位置的时间对齐。

在本申请的具体实施方式中，在对获得的多个干声音频进行时间对齐处理之前，还可以对获得的多个干声音频进行初步筛选，如通过音质检测等工具进行筛选，剔除音频本身带有杂音、有伴奏回踩、音频长度太短、音频能量太小、爆音等音质差的音频。然后再对筛选后保留的干声音频进行时间对齐处理及后续步骤的操作。

S130: 对进行时间对齐处理后的多个干声音频进行虚拟声像定位，以将多个干声音频定位到多个虚拟声像上。

其中，多个虚拟声像位于预先建立的虚拟声像坐标系中，虚拟声像坐标系以人头为中心，以左右耳所在直线中点为坐标原点，第一坐标轴的正方向表示人头正前方，第二坐标轴的正方向表示人头从左耳到右耳的侧方，第三坐标轴的正方向表示人头正上方，每个虚拟声像与坐标原点的距离在设定距离范围内，每个虚拟声像相对于第一坐标轴和第二坐标轴构成的平面的俯仰角在设定角度范围内。

在本申请实施例中，可以预先建立一个虚拟声像坐标系，用于展示声像方位。虚拟声像坐标系具体可以是笛卡尔坐标系。如图 2 所示，虚拟声像坐标系可以以人头为中心，以左右耳所在直线中点为坐标原点，第一坐标轴即 x 轴的正方向表示人头正前方，第二坐标轴即 y 轴的正方向表示人头从左耳到右耳的侧方，第三坐标轴即 z 轴的正方向表示人头正上方，即头顶方向，声像在空间中具有一定的方位角 (azimuth) 和仰角 (elevation)，可以用 (θ, φ) 来表示，rad 表示当前声像与坐标原点的距离。

一般声信号为单路信号，可视为声像在 $(\theta, \varphi) = (0^\circ, 0^\circ)$ 位置上，为了得到某一虚拟声像，可以利用 HRTF (Head Related Transfer Function, 头相关变

换函数) 进行数据卷积实现定位操作。虚拟声像定位示意图如图 3 所示, 其中, X 表示某一真实音源 (单路信号), Y_L 、 Y_R 分别表示左耳和右耳听到的声信号, $HRTF$ 表示声信号从音源位置到双耳的传输路径的传递函数。基于 $HRTF$ 技术可以将真实音源 (单路信号) 通过某一 (θ, φ) 位置上左右耳的 $HRTF$ 滤波, 获得双路声信号。

左、右耳接收声信号的频域特性可表示为:

$$\begin{cases} Y_L=X \cdot H_L \\ Y_R=X \cdot H_R \end{cases}。$$

可以简单的认为人耳听到的声信号是声源 X 经 $HRTF$ 滤波的结果。因此, 在进行虚拟声像定位时, 可以将声信号经过对应位置的 $HRTF$ 进行滤波。在虚拟声像坐标系中可以设定多个虚拟声像, 每个虚拟声像与坐标原点的距离可以在设定距离范围内, 如 1 米范围, 每个虚拟声像相对于虚拟声像坐标系的第一坐标轴和第二坐标轴构成的平面的俯仰角可以在设定角度范围内, 如 10° 范围, 使得多个虚拟声像环绕人耳。

具体的, 多个虚拟声像的每个虚拟声像可以均匀分布在第一坐标轴和第二坐标轴构成的平面的一周。即以相同的间隔角度环绕人耳水平面一周。该间隔角度可以根据实际情况或者对历史数据的分析进行设定, 如设定为 30° 。如果设定间隔角度为 30° , 则以 30° 为间隔环绕人耳水平面一周, 可以定位出 12 个虚拟声像, 这 12 个虚拟声像的仰角为 0° , 方位角分别为: 0° 、 30° 、 60° 、...、 330° 。当然, 间隔角度还可以设定为其他值, 如 15° 、 60° 等。

在另一种实施方式中, 多个虚拟声像中, 位于人头后方的虚拟声像相对于第一坐标轴和第二坐标轴构成的平面的仰角可以大于位于人头前方的虚拟声像相对于第一坐标轴和第二坐标轴构成的平面的仰角。即多个虚拟声像中, 位于人头后方的虚拟声像的仰角可以大于位于人头前方的虚拟声像的仰角。这样可以增强定位效果, 减少虚拟声像的前后镜像问题。如可以将位于人头后方的虚拟声像的仰角上调 10° , 即位于人头前方的虚拟声像的仰角 $\theta=0^{\circ}$, 位于人头后方的虚拟声像的仰角 $\theta=10^{\circ}$ 。

如图 4 所示, 多个虚拟声像以 30° 为间隔环绕人耳水平面一周, 位于

人头前方的虚拟声像的仰角 $\theta=0^{\circ}$ ，位于人头后方的虚拟声像的仰角 $\theta=10^{\circ}$ 。

需要说明的是，设定的多个虚拟声像在虚拟声像坐标系中的位置不限于上面所提到的几种，还可以根据实际需要进行具体设定，只需要满足每个虚拟声像与坐标原点的距离在设定距离范围内，每个虚拟声像相对于第一坐标轴和第二坐标轴构成的平面的俯仰角在设定角度范围内即可。如多个虚拟声像中一部分虚拟声像以 30° 为间隔环绕人耳平面一周，仰角为 0° ，另一部分虚拟声像以 60° 为间隔环绕人耳平面一周，仰角为 10° ，这两部分虚拟声像与坐标原点的距离可以相同或不同，但都在设定距离范围内，这样将会增强后续生成的合唱音频的环绕效果。

10 对进行时间对齐处理后的多个干声音频进行虚拟声像定位，将多个干声音频定位到多个虚拟声像上后，可以继续执行后续步骤的操作。

S140: 基于虚拟声像定位后的多个干声音频，生成合唱音频。

在分别获得多个演唱者对同一目标歌曲进行演唱的干声音频，对多个干声音频进行时间对齐处理，并对对齐后的干声音频进行虚拟声像定位，
15 将多个干声音频定位到多个虚拟声像上之后，可以使得多个干声音频中每个干声音频分别经过对应的虚拟声像位置的 HRTF 滤波处理，在每个虚拟声像处可得到对应的音频数据。基于虚拟声像定位后的多个干声音频，可以生成合唱音频。具体的，可以将经过多个虚拟声像位置的 HRTF 滤波处理后得到的对应的音频数据进行叠加，或者加权叠加，得到合唱音频。
20 得到的合唱音频的音效具有三维声场听感。

S150: 在获取到基于目标歌曲演唱的主唱音频的情况下，将主唱音频、合唱音频和相应的伴奏进行合成后，输出大合唱效果音频。

在本申请实施例的一个应用场景中，生成合唱音频后，可以将合唱音频存储于数据库中，待需要时使用。比如，某个用户想要自己演唱的歌曲
25 具有合唱效果，这种情况下就可以利用合唱音频达到相应效果。

可以获取当前用户基于目标歌曲演唱的音频，将该音频作为主唱音频，然后将主唱音频、合唱音频和相应的伴奏进行合成，得到大合唱效果音频，输出大合唱效果音频，当前用户即可享受到大合唱音效。

对于主唱音频、合唱音频和相应的伴奏的合成，可以通过多种方式实

现,如先将主唱音频和相应的伴奏进行合成,再与合唱音频进行合成,或者,先将合唱音频和相应的伴奏进行合成,再与主唱音频进行合成,又或者,先将主唱音频和合唱音频进行合成,再与相应的伴奏进行合成,如可以对主唱音频和合唱音频进行均衡调整后,按照设定的声伴比叠加相应的伴奏。不同实现方式得到的大合唱音效会有所差别,可以根据实际情况选择具体实现方式。

应用本申请实施例所提供的方法,分别获得多个演唱者对同一目标歌曲进行演唱的干声音频后,对获得的多个干声音频进行时间对齐处理,并对对齐后的多个干声音频进行虚拟声像定位,以将多个干声音频定位到多个虚拟声像上,多个虚拟声像位于以人头为中心的虚拟声像坐标系中,与坐标原点的距离在设定距离范围内,环绕人耳,基于虚拟声像定位后的多个干声音频,生成合唱音频,并在获取到基于目标歌曲演唱的主唱音频的情况下,将主唱音频、合唱音频和相应的伴奏进行合唱,得到并输出大合唱效果音频。将多个干声音频定位到环绕人耳的多个虚拟声像上,可以使

10 得生成的合唱音频具有声场环绕音效,在听感上,可以有效避免最终输出的大合唱效果音频的声场聚集在人头中心而产生的头中效应,使得声场更加宽阔。

在本申请的一个实施例中,步骤 S120 对获得的多个干声音频进行时间对齐处理,可以包括以下步骤:

- 20 第一个步骤:确定目标歌曲对应的参考音频;
- 第二个步骤:针对获得的每个干声音频,分别提取当前干声音频和参考音频的音频特征,音频特征为指纹特征或基频特征;
- 第三个步骤:将当前干声音频与参考音频的音频特征相似度最大值对应的时间确定为音频对齐时间;
- 25 第四个步骤:基于音频对齐时间,对当前干声音频进行时间对齐处理。
- 为便于描述,将上述几个步骤结合起来进行说明。

在本申请实施例中,分别获得多个演唱者对同一目标歌曲进行演唱的干声音频后,在对获得的多个干声音频进行时间对齐处理的过程中,可以先确定目标歌曲对应的参考音频。具体的,可以从获得的多个干声音频中

选择出音质较好的一个干声音频，作为参考音频。还可以将目标歌曲的原唱干声音频确定为参考音频。

- 针对获得的每个干声音频，可以分别提取当前干声音频和参考音频的音频特征，音频特征为指纹特征或基频特征。如可以通过多频带滤波提取
- 5 Mel 频段信息、Bark 频段信息、Erb 频段功率等，然后通过半波整流、二值判断等得到指纹特征。又如可以通过 pyin、crepe、harvest 等基频提取工具提取基频特征。参考音频的音频特征可以在提取一次后保存，在有需要时直接调用即可。

- 将当前干声音频与参考音频的音频特征进行比较，可以通过相似度曲线等表征，可以将相似度最大值对应的时间确定为音频对齐时间。然后基
- 10 于音频对齐时间，对当前干声音频进行时间对齐处理。

对于获得的每个干声音频均通过与参考音频的音频特征的比较，得到相应的音频对齐时间，并进行时间对齐处理后，即可得到时间对齐处理后的多个干声音频。

- 15 在本申请的一个实施例中，该方法还可以包括以下步骤：

分别对获得的多个干声音频进行带通滤波处理，得到多个低音数据；
相应的，基于进行虚拟声像定位后的多个干声音频，生成合唱音频，
包括：

- 基于进行虚拟声像定位后的多个干声音频和多个低音数据，生成合唱
- 20 音频。

在本申请实施例中，在分别获得多个演唱者对同一目标歌曲进行演唱的干声音频后，可以分别对获得的多个干声音频进行带通滤波处理，如对多个干声音频进行截止频率为[33,523]Hz 的带通滤波处理，得到多个低音数据。

- 25 基于进行虚拟声像定位后的多个干声音频和多个低音数据，可以生成合唱音频。具体的，可以将得到的多个低音数据与基于进行虚拟声像定位后的多个干声音频进行叠加或者加权叠加等处理，生成合唱音频。叠加低音信号后，可以增强声信号的厚重感。

在本申请的一个实施例中，该方法还可以包括以下步骤：

分别对获得的多个干声音频进行混响模拟处理;

相应的, 基于进行虚拟声像定位后的多个干声音频, 生成合唱音频, 包括:

5 基于进行虚拟声像定位后的多个干声音频和混响模拟处理后的多个干声音频, 生成合唱音频。

通常, 声场中声源发出的声信号会经过直达声、反射、混响等过程。图 5 所示为一种典型的空间声场过程组成示意图。在该图中, 幅度最大的声信号为直达声, 紧接着的声信号是声波在距离听者最近的物体上反射得到的反射声信号, 其具有明显的方向性, 之后的一段密集的声信号是由声波经过周围物体的多次反射后叠加得到的混响声信号, 是大量不同方位反
10 射声的叠加, 没有方向性。

根据已知的房间脉冲响应特性, 混响声为多路反射声的叠加, 特点是能量弱、没有方向性, 因为它是大量来自不同方位的后期反射声的叠加, 具有较高的回声密度, 所以可以利用混响产生具有包围感的环绕音效。

15 在本申请实施例中, 在分别获得多个演唱者对同一目标歌曲进行演唱的干声音频后, 可以分别对获得的多个干声音频进行混响模拟处理。具体的, 可以利用梳状滤波器和全通滤波器的级联分别对获得的多个干声音频进行混响模拟处理。

图 6 所示为梳状滤波器和全通滤波器的一种级联形式, 其中, 四个梳状滤波器并联后与两个全通滤波器串联。实际模拟得到的混响脉冲响应如
20 图 7 所示。

需要说明的是, 图 6 所示仅为一种具体形式, 在实际应用中, 可以有其他更多形式, 梳状滤波器、全通滤波器的个数和级联方式都可以根据实际需要调整。

25 分别对获得的多个干声音频进行混响模拟处理, 及进行虚拟声像定位, 将多个干声音频定位到多个虚拟声像上之后, 可以基于进行虚拟声像定位后的多个干声音频和混响模拟处理后的多个干声音频, 生成合唱音频。具体的, 可以将进行虚拟声像定位后的多个干声音频和混响模拟处理后的多个干声音频进行叠加或者加权叠加处理, 生成合唱音频。这样可以增强声

信号的空间音效，进一步抑制头中效应，扩展声场。

在本申请的一个实施例中，在对进行时间对齐处理后的多个干声音频进行虚拟声像定位之后，该方法还可以包括以下步骤：

分别对进行虚拟声像定位后的多个干声音频进行混响模拟处理；

- 5 相应的，基于进行虚拟声像定位后的多个干声音频，生成合唱音频，包括：

基于进行虚拟声像定位，且进行混响模拟处理后的多个干声音频，生成合唱音频。

- 10 在本申请实施例中，在分别获得多个演唱者对同一目标歌曲进行演唱的干声音频，对获得的多个干声音频进行时间的对齐处理，并进行虚拟声像定位之后，可以进一步分别对进行虚拟声像定位后的多个干声音频进行混响模拟处理，混响模拟处理过程可以参考上一实施例的混响模拟处理过程，这里不再赘述。

- 15 基于进行虚拟声像定位，且进行混响模拟处理后的多个干声音频，可以生成合唱音频。具体的，可以将进行虚拟声像定位，且进行混响模拟处理后的多个干声音频进行叠加或者加权叠加等处理，生成合唱音频。

对进行虚拟声像处理后的多个干声音频进行混响模拟处理，可以增强声信号的空间音效，进一步抑制头中效应，扩展声场。

在本申请的一个实施例中，该方法还可以包括以下步骤：

- 20 分别对获得的多个干声音频进行双声道模拟处理；

相应的，基于进行虚拟声像定位后的多个干声音频，生成合唱音频，包括：

基于进行虚拟声像定位后的多个干声音频和双声道模拟处理后的多个干声音频，生成合唱音频。

- 25 在本申请实施例中，在分别获得多个演唱者对同一目标歌曲进行演唱的干声音频，且对获得的多个干声音频进行时间对齐处理后，可以分别对多个干声音频进行双声道模拟处理。通过延迟来降低两声道信号相关性，尽量扩展声场得到双路输出。

如图8所示，多个干声音频可以通过左右各8组不同的延迟加权实现

双声道模拟，其中， d 表示延迟， g 表示权重。由于一般房间脉冲响应取80ms作为混响时间，所以延迟参数可以选择21ms~79ms中不等的16个参数。利用幅度衰减来表示声波因反射作用而造成的能量损失，由此可以降低两路环境信息的相关性。即可以将干声音频分别做拷贝得到信息相同的两路信号，两路信号完全相关，再利用不同的延迟与幅度进行衰减，降低两路信号的相关性，以得到伪立体声信号。

需要说明的是，图8所示仅为一种具体示例，可以根据实际需要，设置较少组或更多组不同的延迟实现双声道模拟。

基于进行虚拟声像定位后的多个干声音频和双声道模拟处理后的多个干声音频，可以生成合唱音频。具体的，可以将进行虚拟声像定位后的多个干声音频和双声道模拟处理后的多个干声音频进行叠加或者加权叠加等处理，生成合唱音频。

在本申请的一个实施例中，在分别对获得的多个干声音频进行双声道模拟处理之后，该方法还可以包括以下步骤：

15 对进行双声道模拟处理后的多个干声音频进行混响模拟处理；

相应的，基于进行虚拟声像定位后的多个干声音频和双声道模拟处理后的多个干声音频，生成合唱音频，包括：

基于进行虚拟声像定位后的多个干声音频、双声道模拟处理及混响模拟处理后的多个干声音频，生成合唱音频。

20 在本申请实施例中，在分别获得多个演唱者对同一目标歌曲进行演唱的干声音频，对多个干声音频进行时间对齐处理，且分别对多个干声音频进行双声道模拟处理之后，可以进一步对进行双声道模拟处理后的多个干声音频进行混响模拟处理，以增强声信号的空间效应，抑制头中效应，扩展声场。

25 对多个干声音频进行虚拟声像定位，将多个干声音频定位到多个虚拟声像上之后，可以基于进行虚拟声像定位后的多个干声音频、双声道模拟处理及混响模拟处理后的多个干声音频，生成合唱音频。具体的，可以对虚拟声像定位后的多个干声音频、双声道模拟处理及混响模拟处理后的多个干声音频进行叠加或者加权叠加等处理，生成合唱音频。

在实际应用中，在获得多个演唱者对同一目标歌曲进行演唱的干声音频后，可以先对获得的多个干声音频进行时间对齐处理，然后再基于进行时间对齐处理后的多个干声音频进行虚拟声像定位、低音增强、混响模拟、双声道模拟等处理，具体的处理可以综合上面各实施例进行，通过对多个干声音频进行虚拟声像定位、低音增强、混响模拟、双声道模拟，使得最后生成的合唱音频具有声场环绕音效，可以对较大范围的声音不对齐具有较高的鲁棒性，如果要将合唱音频与主唱音频进行叠加，即使主唱音频延迟落差较大，也可以保证用户具有和谐的听觉体验。

图 9 所示为对进行时间对齐处理后的多个干声音频进行处理的系统框架示意图，其中包括低音增强单元、虚拟声像定位单元、双声道模拟单元和混响模拟单元。低音增强单元用于对多个干声音频进行带通滤波处理，得到低音数据；虚拟声像定位单元用于对多个干声音频进行虚拟声像定位，以将多个干声音频定位到多个虚拟声像上；双声道模拟单元用于对多个干声音频进行双声道模拟处理；混响模拟单元用于对多个干声音频进行混响模拟处理。虚拟声像定位单元和双声道模拟单元均可与混响模拟单元连接，通过虚拟声像定位单元对多个干声音频进行虚拟声像定位后，可以进一步通过混响模拟单元进行混响模拟处理，同样，通过双声道模拟单元对多个干声音频进行双声道模拟处理之后，可以进一步通过混响模拟单元进行混响模拟处理。最终可以对经过这些单元处理后的音频数据进行加权叠加，得到合唱音频。

图 10 所示为对多个干声音频进行处理的一种具体示例，H 表示 HRTF 滤波的传递函数，通过该传递函数的处理可以对多个干声音频进行虚拟声像定位，将多个干声音频定位到环绕人耳水平面的 12 个虚拟声像上，REV 表示混响模拟单元，BASS 表示低音增强单元，REF 表示双声道模拟单元。这里的混响模拟单元可以使用同一参数，还可以根据实际需求，对不同的混响模拟单元配置不同参数，得到灵活的混响调制。

本申请实施例最后生成的合唱音频的大合唱效果与真实演唱会合唱的听感更加逼近。在实际应用中，在主唱音频基础上添加伴奏，同时混入合唱音频，可以让用户在听感上具有身临其境的演唱会体验，获得更加震撼

的沉浸式声场包围体验。

在本申请的一个实施例中，对进行时间对齐处理后的多个干声音频进行虚拟声像定位，可以包括以下步骤：

步骤一：按照虚拟声像的个数，将获得的进行时间对齐处理后的多个干声音频进行分组，组数与虚拟声像的个数相同；

步骤二：将各组干声音频分别定位到对应的虚拟声像上，不同组干声音频对应不同虚拟声像。

为便于描述，将上述两个步骤结合起来进行说明。

在本申请实施例中，分别获得多个演唱者对同一目标歌曲进行演唱的干声音频，并对获得的多个干声音频进行时间对齐处理之后，可以按照虚拟声像的个数，将获得的进行时间对齐处理后的多个干声音频进行分组，分得的组数与虚拟声像的个数相同，同一个组内包括若干个干声音频。如果获得的干声音频数量较多，则可以使得同一干声音频仅在一个组中，如果获得的干声音频数量较少，则可以使得同一干声音频在多个组中，以更好地实现大合唱音效。

将多个干声音频进行分组后，可以将各组干声音频分别定位到对应的虚拟声像上，不同组干声音频对应不同虚拟声像。实现对多个干声音频的虚拟声像的定位处理，增强大合唱音效。

在本申请的一个实施例中，将主唱音频、合唱音频和相应的伴奏进行合成，可以包括以下步骤：

对主唱音频和合唱音频分别进行音量调整，和/或，对主唱音频和合唱音频进行混响模拟处理；

将进行音量调整和/或进行混响模拟处理后的主唱音频、合唱音频和相应的伴奏进行合成。

在本申请实施例中，在获取到基于目标歌曲演唱的主唱音频后，可以对主唱音频和合唱音频分别进行音量调整，使得主唱音频和合唱音频的音量相当，或者主唱音频的音量大于合唱音频的音量。同时，还可以对主唱音频和合唱音频进行混响模拟处理，以得到具有包围感的环绕音效。

再将进行音量调整和/或进行混响模拟处理后的主唱音频、合唱音频和

相应的伴奏进行合成，使得最后输出的大合唱效果音频为用户带来更好的听感体验。

相应于上面的方法实施例，本申请实施例还提供了一种合唱音频的处理装置，下文描述的合唱音频的处理装置与上文描述的合唱音频的处理方法可相互对应参照。

参见图 11 所示，该装置可以包括以下模块：

干声音频获得模块 1110，用于分别获得多个演唱者对同一目标歌曲进行演唱的干声音频；

10 时间对齐处理模块 1120，用于对获得的多个干声音频进行时间对齐处理；

15 虚拟声像定位模块 1130，用于对进行时间对齐处理后的多个干声音频进行虚拟声像定位，以将多个干声音频定位到多个虚拟声像上；其中，多个虚拟声像位于预先建立的虚拟声像坐标系中，虚拟声像坐标系以人头为中心，以左右耳所在直线中点为坐标原点，第一坐标轴的正方向表示人头正前方，第二坐标轴的正方向表示人头从左耳到右耳的侧方，第三坐标轴的正方向表示人头正上方，每个虚拟声像与坐标原点的距离在设定距离范围内，每个虚拟声像相对于第一坐标轴和第二坐标轴构成的平面的俯仰角在设定角度范围内；

20 合唱音频生成模块 1140，用于基于进行虚拟声像定位后的多个干声音频，生成合唱音频；

大合唱效果音频获得模块 1150，用于在获取到基于目标歌曲演唱的主唱音频的情况下，将主唱音频、合唱音频和相应的伴奏进行合成后，输出大合唱效果音频。

25 应用本申请实施例所提供的装置，分别获得多个演唱者对同一目标歌曲进行演唱的干声音频后，对获得的多个干声音频进行时间对齐处理，并对对齐后的多个干声音频进行虚拟声像定位，以将多个干声音频定位到多个虚拟声像上，多个虚拟声像位于以人头为中心的虚拟声像坐标系中，与坐标原点的距离在设定距离范围内，环绕人耳，基于虚拟声像定位后的多

个干声音频，生成合唱音频，并在获取到基于目标歌曲演唱的主唱音频的情况下，将主唱音频、合唱音频和相应的伴奏进行合唱，得到并输出大合唱效果音频。将多个干声音频定位到环绕人耳的多个虚拟声像上，可以使得生成的合唱音频具有声场环绕音效，在听感上，可以有效避免最终输出的大合唱效果音频的声场聚集在人头中心而产生的头中效应，使得声场更加宽阔。

- 5 在本申请的一种具体实施方式中，时间对齐处理模块 1120，用于：
确定目标歌曲对应的参考音频；
针对获得的每个干声音频，分别提取当前干声音频和参考音频的音频
10 特征，音频特征为指纹特征或基频特征；
将当前干声音频与参考音频的音频特征相似度最大值对应的时间确定为音频对齐时间；
基于音频对齐时间，对当前干声音频进行时间对齐处理。
- 15 在本申请的一种具体实施方式中，还包括低音数据获得模块，用于：
分别对获得的多个干声音频进行带通滤波处理，得到多个低音数据；
相应的，合唱音频生成模块 1140，用于：
基于进行虚拟声像定位后的多个干声音频和多个低音数据，生成合唱
音频。
- 20 在本申请的一种具体实施方式中，还包括混响模拟处理模块，用于：
分别对获得的多个干声音频进行混响模拟处理；
相应的，合唱音频生成模块 1140，用于：
基于进行虚拟声像定位后的多个干声音频和混响模拟处理后的多个干
声音频，生成合唱音频。
- 25 在本申请的一种具体实施方式中，混响模拟处理模块，用于：
利用梳状滤波器和全通滤波器的级联分别对获得的多个干声音频进行
混响模拟处理。
- 在本申请的一种具体实施方式中，混响模拟处理模块，还用于：
在对进行时间对齐处理后的多个干声音频进行虚拟声像定位之后，分
别对进行虚拟声像定位后的多个干声音频进行混响模拟处理；

相应的，合唱音频生成模块 1140，用于：

基于进行虚拟声像定位，且进行混响模拟处理后的多个干声音频，生成合唱音频。

在本申请的一种具体实施方式中，还包括双声道模拟处理模块，用于：

5 分别对获得的多个干声音频进行双声道模拟处理；

相应的，合唱音频生成模块 1140，用于：

基于进行虚拟声像定位后的多个干声音频和双声道模拟处理后的多个干声音频，生成合唱音频。

在本申请的一种具体实施方式中，混响模拟处理模块，还用于：

10 在分别对获得的多个干声音频进行双声道模拟处理之后，对进行双声道模拟处理后的多个干声音频进行混响模拟处理；

相应的，合唱音频生成模块 1140，用于：

基于进行虚拟声像定位后的多个干声音频、双声道模拟处理及混响模拟处理后的多个干声音频，生成合唱音频。

15 在本申请的一种具体实施方式中，虚拟声像定位模块 1130，用于：

按照虚拟声像的个数，将获得的进行时间对齐处理后的多个干声音频进行分组，组数与虚拟声像的个数相同；

将各组干声音频分别定位到对应的虚拟声像上，不同组干声音频对应不同虚拟声像。

20 在本申请的一种具体实施方式中，多个虚拟声像中，位于人头后方的虚拟声像相对于第一坐标轴和第二坐标轴构成的平面的仰角大于位于人头前方的虚拟声像相对于第一坐标轴和第二坐标轴构成的平面的仰角；或者，每个虚拟声像均匀分布在第一坐标轴和第二坐标轴构成的平面的一周。

25 在本申请的一种具体实施方式中，大合唱效果音频获得模块 1150，用于：

对主唱音频和合唱音频分别进行音量调整，和/或，对主唱音频和合唱音频进行混响模拟处理；

将进行音量调整和/或进行混响模拟处理后的主唱音频、合唱音频和相应的伴奏进行合成。

相应于上面的方法实施例，本申请实施例还提供了一种合唱音频的处理设备，包括：

- 存储器，用于存储计算机程序；
- 5 处理器，用于执行计算机程序时实现上述合唱音频的处理方法的步骤。

如图 12 所示，为合唱音频的处理设备的组成结构示意图，合唱音频的处理设备可以包括：处理器 10、存储器 11、通信接口 12 和通信总线 13。处理器 10、存储器 11、通信接口 12 均通过通信总线 13 完成相互间的通信。

- 在本申请实施例中，处理器 10 可以为中央处理器（Central Processing Unit, CPU）、特定应用集成电路、数字信号处理器、现场可编程门阵列或者其他可编程逻辑器件等。

处理器 10 可以调用存储器 11 中存储的程序，具体的，处理器 10 可以执行合唱音频的处理方法的实施例中的操作。

- 存储器 11 中用于存放一个或者一个以上程序，程序可以包括程序代码，程序代码包括计算机操作指令，在本申请实施例中，存储器 11 中至少存储有用于实现以下功能的程序：

- 分别获得多个演唱者对同一目标歌曲进行演唱的干声音频；
 - 对获得的多个干声音频进行时间对齐处理；

- 对进行时间对齐处理后的多个干声音频进行虚拟声像定位，以将多个干声音频定位到多个虚拟声像上；其中，多个虚拟声像位于预先建立的虚拟声像坐标系中，虚拟声像坐标系以人头为中心，以左右耳所在直线中点为坐标原点，第一坐标轴的正方向表示人头正前方，第二坐标轴的正方向表示人头从左耳到右耳的侧方，第三坐标轴的正方向表示人头正上方，每个虚拟声像与坐标原点的距离在设定距离范围内，每个虚拟声像相对于第一坐标轴和第二坐标轴构成的平面的俯仰角在设定角度范围内；

- 基于进行虚拟声像定位后的多个干声音频，生成合唱音频；

在获取到基于目标歌曲演唱的主唱音频的情况下，将主唱音频、合唱音频和相应的伴奏进行合成后，输出大合唱效果音频。

在一种可能的实现方式中，存储器 11 可包括存储程序区和存储数据

区，其中，存储程序区可存储操作系统，以及至少一个功能（比如音频播放功能、音频合成功能）所需的应用程序等；存储数据区可存储使用过程中所创建的数据，如声像定位数据、音频合成数据等。

此外，存储器 11 可以包括高速随机存取存储器，还可以包括非易失性存储器，例如至少一个磁盘存储器件或其他易失性固态存储器件。

通信接口 12 可以为通信模块的接口，用于与其他设备或者系统连接。

当然，需要说明的是，图 12 所示的结构并不构成对本申请实施例中合唱音频的处理设备的限定，在实际应用中合唱音频的处理设备可以包括比图 12 所示的更多或更少的部件，或者组合某些部件。

10

相应于上面的方法实施例，本申请实施例还提供了一种计算机可读存储介质，计算机可读存储介质上存储有计算机程序，计算机程序被处理器执行时实现上述合唱音频的处理方法的步骤。

15 本说明书中各个实施例采用递进的方式描述，每个实施例重点说明的都是与其它实施例的不同之处，各个实施例之间相同或相似部分互相参见即可。

20 专业人员还可以进一步意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、计算机软件或者二者的结合来实现，为了清楚地说明硬件和软件的可互换性，在上述说明中已经按照功能一般性地描述了各示例的组成及步骤。这些功能究竟以硬件还是软件方式来执行，取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本申请的范围。

25 结合本文中所公开的实施例描述的方法或算法的步骤可以直接用硬件、处理器执行的软件模块，或者二者的结合来实施。软件模块可以置于随机存储器（RAM）、内存、只读存储器（ROM）、电可编程 ROM、电可擦除可编程 ROM、寄存器、硬盘、可移动磁盘、CD-ROM、或技术领域内所公知的任意其它形式的存储介质中。

本文中应用了具体个例对本申请的原理及实施方式进行了阐述，以上实施例的说明只是用于帮助理解本申请的技术方案及其核心思想。应当指出，对于本技术领域的普通技术人员来说，在不脱离本申请原理的前提下，还可以对本申请进行若干改进和修饰，这些改进和修饰也落入本申请权利要求5要求的保护范围内。

权 利 要 求

- 1、一种合唱音频的处理方法，其特征在于，包括：
分别获得多个演唱者对同一目标歌曲进行演唱的干声音频；
对获得的多个所述干声音频进行时间对齐处理；
- 5 对进行时间对齐处理后的多个所述干声音频进行虚拟声像定位，以将多个所述干声音频定位到多个虚拟声像上；其中，多个所述虚拟声像位于预先建立的虚拟声像坐标系中，所述虚拟声像坐标系以人头为中心，以左右耳所在直线中点为坐标原点，第一坐标轴的正方向表示人头正前方，第二坐标轴的正方向表示人头从左耳到右耳的侧方，第三坐标轴的正方向表示人头正上方，每个所述虚拟声像与所述坐标原点的距离在设定距离范围内，每个所述虚拟声像相对于所述第一坐标轴和所述第二坐标轴构成的平面的俯仰角在设定角度范围内；
- 10 基于进行虚拟声像定位后的多个所述干声音频，生成合唱音频；
在获取到基于所述目标歌曲演唱的主唱音频的情况下，将所述主唱音频、所述合唱音频和相应的伴奏进行合成后，输出大合唱效果音频。
- 15 2、根据权利要求1所述的合唱音频的处理方法，其特征在于，所述对获得的多个所述干声音频进行时间对齐处理，包括：
确定所述目标歌曲对应的参考音频；
针对获得的每个所述干声音频，分别提取当前干声音频和所述参考音频的音频特征，所述音频特征为指纹特征或基频特征；
- 20 将所述当前干声音频与所述参考音频的音频特征相似度最大值对应的时间确定为音频对齐时间；
基于所述音频对齐时间，对所述当前干声音频进行时间对齐处理。
- 25 3、根据权利要求1所述的合唱音频的处理方法，其特征在于，还包括：
分别对获得的多个所述干声音频进行带通滤波处理，得到多个低音数据；
相应的，所述基于进行虚拟声像定位后的多个所述干声音频，生成合唱音频，包括：
基于进行虚拟声像定位后的多个所述干声音频和多个所述低音数据，

生成合唱音频。

4、根据权利要求1所述的合唱音频的处理方法，其特征在于，还包括：
分别对获得的多个所述干声音频进行混响模拟处理；

5 相应的，所述基于进行虚拟声像定位后的多个所述干声音频，生成合唱音频，包括：

基于进行虚拟声像定位后的多个所述干声音频和混响模拟处理后的多个所述干声音频，生成合唱音频。

5、根据权利要求4所述的合唱音频的处理方法，其特征在于，所述分别对获得的多个所述干声音频进行混响模拟处理，包括：

10 利用梳状滤波器和全通滤波器的级联分别对获得的多个所述干声音频进行混响模拟处理。

6、根据权利要求1所述的合唱音频的处理方法，其特征在于，在所述对进行时间对齐处理后的多个所述干声音频进行虚拟声像定位之后，还包括：

15 分别对进行虚拟声像定位后的多个所述干声音频进行混响模拟处理；

相应的，所述基于进行虚拟声像定位后的多个所述干声音频，生成合唱音频，包括：

基于进行虚拟声像定位，且进行混响模拟处理后的多个所述干声音频，生成合唱音频。

20 7、根据权利要求1所述的合唱音频的处理方法，其特征在于，还包括：
分别对获得的多个所述干声音频进行双声道模拟处理；

相应的，所述基于进行虚拟声像定位后的多个所述干声音频，生成合唱音频，包括：

25 基于进行虚拟声像定位后的多个所述干声音频和双声道模拟处理后的多个所述干声音频，生成合唱音频。

8、根据权利要求7所述的合唱音频的处理方法，其特征在于，在所述分别对获得的多个所述干声音频进行双声道模拟处理之后，还包括：

对进行双声道模拟处理后的多个所述干声音频进行混响模拟处理；

相应的，所述基于进行虚拟声像定位后的多个所述干声音频和双声道

模拟处理后的多个所述干声音频，生成合唱音频，包括：

基于进行虚拟声像定位后的多个所述干声音频、双声道模拟处理及混响模拟处理后的多个所述干声音频，生成合唱音频。

- 5 9、根据权利要求1所述的合唱音频的处理方法，其特征在于，所述对进行时间对齐处理后的多个所述干声音频进行虚拟声像定位，包括：

按照虚拟声像的个数，将获得的进行时间对齐处理后的多个所述干声音频进行分组，组数与虚拟声像的个数相同；

将各组干声音频分别定位到对应的虚拟声像上，不同组干声音频对应不同虚拟声像。

- 10 10、根据权利要求1所述的合唱音频的处理方法，其特征在于，多个所述虚拟声像中，位于人头后方的虚拟声像相对于所述第一坐标轴和所述第二坐标轴构成的平面的仰角大于位于人头前方的虚拟声像相对于所述第一坐标轴和所述第二坐标轴构成的平面的仰角；

或者，

- 15 每个所述虚拟声像均匀分布在所述第一坐标轴和所述第二坐标轴构成的平面的一周。

11、根据权利要求1至10之中任一项所述的合唱音频的处理方法，其特征在于，所述将所述主唱音频、所述合唱音频和相应的伴奏进行合成，包括：

- 20 对所述主唱音频和所述合唱音频分别进行音量调整，和/或，对所述主唱音频和所述合唱音频进行混响模拟处理；

将进行音量调整和/或进行混响模拟处理后的所述主唱音频、所述合唱音频和相应的伴奏进行合成。

- 25 12、一种合唱音频的处理设备，其特征在于，包括：

存储器，用于存储计算机程序；

处理器，用于执行所述计算机程序时实现如权利要求1至11任一项所述的合唱音频的处理方法的步骤。

13、一种计算机可读存储介质，其特征在于，所述计算机可读存储介质上存储有计算机程序，所述计算机程序被处理器执行时实现如权利要求

1 至 11 任一项所述的合唱音频的处理方法的步骤。