

说明书

音频处理方法、装置、设备、存储介质和程序产品

本申请要求 2022 年 10 月 19 日递交的申请号为 202211280550.4、标题为“音频处理方法、装置、设备、存储介质和程序产品”的中国发明专利申请的优先权。该中国专利申请的全部内容通过引用并入本文中。

技术领域

本公开涉及音频处理技术领域，尤其涉及一种音频处理方法、装置、设备、存储介质和程序产品。

背景技术

随着互联网技术的发展，许多音乐播放器不仅支持海量歌曲的在线播放，还可以为用户提供 K 歌服务，K 歌服务是指由音乐播放器播放歌曲的音乐伴奏，用户跟随音乐伴奏进行演唱。在演唱过程中，该音乐播放器还可以对用户演唱的歌曲进行录音，以便于后续播放该用户的演唱的歌曲。

相关技术中，音频录制过程一般为：终端开启音乐播放器播放歌曲的音乐伴奏，当用户开始演唱时，终端同步开始录音，得到录制音频。当录制结束时，终端会基于根据音乐伴奏和录制音频进行音频混合，得到一首完成的歌曲。

但是，在终端播放音乐伴奏时会有播放时延，在录制音频的过程中会有采集时延，导致音乐伴奏和录制音频之间存在较大时延，在将音乐伴奏和录制音频混合的过程中，由于时延较大，导致音乐伴奏和录制音频不同步，严重影响听感，导致用户 K 歌体验差。

发明内容

为了解决上述技术问题，本公开提供了一种音频处理方法、装置、设备及存储介质，优化了在背景音乐场景下录音时延的方案，解决了录音时延过长、影响听觉效果的问题。

第一方面，本公开实施例提供一种音频处理方法，所述方法包括：

在向外播放第一音频的过程中，对外部声音采集得到第二音频；

基于当前系统时间和播放时延确定所述第一音频的播放时长，其中，所述播放时延为音频播放器启动时刻到所述音频播放器输出音频时刻之间的时长；

基于所述第一音频的播放时长和采集时延计算参考时间戳；其中，所述采集时延包括音频采集器启动时刻到所述第二音频被采集到第一个音频帧的时刻之间的时长；

基于所述参考时间戳对所述第二音频进行处理，得到第三音频；

将所述第一音频与所述第三音频进行混音处理，得到目标音频。

第二方面，本公开实施例提供一种音频处理装置，所述装置包括：

第二音频采集模块，用于在向外播放第一音频的过程中，对外部声音采集得到第二音频；

播放时长确定模块，用于基于当前系统时间和播放时延确定所述第一音频的播放时长，其中，所述播放时延为音频播放器启动时刻到所述音频播放器输出音频时刻之间的时长；

参考时间戳确定模块，用于基于所述第一音频的播放时长和采集时延计算参考时间戳；其中，所述采集时延包括音频采集器启动时刻到所述第二音频被采集到第一个音频帧的时刻之间的时长；

第三音频确定模块，用于基于所述参考时间戳对所述第二音频进行处理，得到第三音频；

目标音频确定模块，用于将所述第一音频与所述第三音频进行混音处理，得到目标音频。

第三方面，本公开实施例提供一种电子设备，所述电子设备包括：

一个或多个处理器；

存储装置，用于存储一个或多个程序；

当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现如上述第一方面中任一项所述的音频处理方法。

第四方面，本公开实施例提供一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现如上述第一方面中任一项所述的音频处理方法。

第五方面，本公开实施例提供一种计算机程序产品，该计算机程序产品包括计算机程序或指令，该计算机程序或指令被处理器执行时实现如上述第一方面中任一项所述的音频处理方法。

本公开涉及一种音频处理方法、装置、设备、存储介质和程序产品，所述方法包括：在向外播放第一音频的过程中，对外部声音采集得到第二音频；基于当前系统时间和播放时延确定所述第一音频的播放时长，其中，所述播放时延为音频播放器启动时刻到所述音频播放器输出音频时刻之间的时长；基于所述第一音频的播放时长和采集时延计算参考时间戳；其中，所述采集时延包括音频采集器启动时刻到所述第二音频被采集到第一个音频帧的时刻之间的时长；基于所述参考时间戳对所述第二音频进行处理，得到第三音频；将所述第一音频与所述第三音频进行混音处理，得到目标音频。本公开提供的技术方案在第一音频和第二音频进行混音的过程中，优化时延对音频的影响，使得播放的第一音频和录制的第二音频同步，降低时延对音频听感的影响，提高用户 K 歌体验。

附图说明

结合附图并参考以下具体实施方式，本公开各实施例的上述和其他特征、优点及方面将变得更加明显。贯穿附图中，相同或相似的附图标记表示相同或相似的元素。应当理解附图是示意性的，原件和元素不一定按照比例绘制。

图 1 为本公开实施例中的一种音频处理方法的流程示意图；

图 2 为本公开实施例中的录音时延详细内容的示意图；

图 3 为本公开实施例中的一种背景音乐和录音采集对齐方案的示意图；

图 4 为本公开实施例中的一种音频处理装置的结构示意图；以及

图 5 为本公开实施例中的一种电子设备的结构示意图。

具体实施方式

下面将参照附图更详细地描述本公开的实施例。虽然附图中显示了本公开的某些实施例，然而应当理解的是，本公开可以通过各种形式来实现，而且不应该被解释为限于这里阐述的实施例，相反提供这些实施例是为了更加透彻和完整地理解本公开。应当理解的是，本公开的附图及实施例仅用于示例性作用，并非用于限制本公开的保护范围。

应当理解，本公开的方法实施方式中记载的各个步骤可以按照不同的顺序执行，和/或并行执行。此外，方法实施方式可以包括附加的步骤和/或省略执行示出的步骤。本公开的范围在此方面不受限制。

本文使用的术语“包括”及其变形是开放性包括，即“包括但不限于”。术语“基于”是“至少部分地基于”。术语“一个实施例”表示“至少一个实施例”；术语“另一实施例”表示“至少一个另外的实施例”；术语“一些实施例”表示“至少一些实施例”。其他术语的相关定义将在下文描述中给出。

需要注意，本公开中提及的“第一”、“第二”等概念仅用于对不同的装置、模块或单元进行区分，并非用于限定这些装置、模块或单元所执行的功能的顺序或者相互依存关系。

需要注意，本公开中提及的“一个”、“多个”的修饰是示意性而非限制性的，本领域技术人员应当理解，除非在上下文另有明确指出，否则应该理解为“一个或多个”。

本公开实施方式中的多个装置之间所交互的消息或者信息的名称

仅用于说明性的目的，而并不是用于对这些消息或信息的范围进行限制。

下面将结合附图，对本申请实施例提出的音频处理方法进行详细介绍。

图 1 为本公开实施例中的一种音频处理方法的流程图，本实施例可适用于混音处理时，降低时延影响的情况，该方法可以由音频处理装置执行，该音频处理装置可以采用软件和/或硬件的方式实现，该音频处理装置可配置于电子设备中。所述电子设备包括具备拍照功能的智能终端，包括智能手机、笔记本电脑、平板电脑、数码相机/摄像机、游戏设备等。可选的，所述智能终端包括一触摸屏。

如图 1 所述，本公开实施例提供的音频处理方法主要包括如下步骤：

S101、在向外播放第一音频的过程中，对外部声音采集得到第二音频。

在一些具体的场景或者应用程序中，需要两种音频的混合，以达到某种目的效果，例如短视频拍摄中，背景音乐播出烘托出某种氛围，然后打开用户的录音设备采集用户朗读音频，达到用户表达效果。或者在 K 歌应用程序中，播放歌曲伴奏，采集用户歌唱音频，最终将歌曲伴奏和歌唱音频对齐合并至完整的歌曲。

具体的，从用户终端接收到音频播放指令到用户终端的音频播放装置输出音频，会有一个播放时延，以及从用户终端接收到音频录制指令到音频采集装置采集到首帧音频帧，也会存在一个采集时延。

在本公开实施例中，所述第一音频可以理解为用户终端的扬声器播放的音频。可选的，第一音频可以是指背景音乐（Back Ground Music, BGM）或者歌曲伴奏音频。第二音频是指用户终端采集到的外部声音形成的音频。

在本公开的一个实施方式中，使用用户终端的音频播放装置播放

第一音频，同时使用用户终端的音频采集装置采集第二音频，所述音频采集装置可以是拾音器、麦克风等中一种或多种。

在本公开的一个具体实施例中，在 K 歌场景中，响应于用户对 K 歌应用程序中录制控件的触发操作，用户终端的音频播放装置播放歌曲伴奏，同时，用户终端的音频采集装置采集外部的声音数据得到第二音频。或者，在视频录制场景中，响应于用户对视频录制应用程序中录制控件的触发操作，用户终端的音频播放装置播放背景音乐，同时，用户终端的音频采集装置采集外部的声音数据得到第二音频。

S102、基于当前系统时间和播放时延确定所述第一音频的播放时长，其中，所述播放时延为音频播放器启动时刻到所述音频播放器输出音频时刻之间的时长。

在本公开实施例中，上述系统时间是指用户终端的系统所运行显示的时间，所述当前系统时间可以理解为当前时刻获取到的系统运行的时间。其中，所述当前系统时间是指一个时间点。

在本公开实施例中，首先介绍一下本公开中涉及到一个时延参数。如图 2 所示，从接收到音频播放指令，到采集到音频，整个过程的时延 T1 主要包括：播放时延、传输时延 T_c 和采集时延，其中播放时延包括播放软件时延 T_a 和播放硬件时延 T_b ，采集时延包括：采集软件时延 T_e 和采集硬件时延 T_d 。需要说明的是，由于传输时延 T_c 很小，本公开实施例中不考虑传输时延 T_c 对两个音频混音的影响。

在本公开实施例中，所述播放时延为接收到一个音频帧的播放指令到音频播放器输出该音频帧之间的时长。其中，接收到音频帧的播放指令可以是响应前一个音频帧的播放指令后，自动生成并接收到当前音频帧的播放指令。例如：响应第 N 个音频帧的播放指令后，自动生成并接收到第 N+1 个音频帧的播放指令。接收到音频帧的播放指令还包括：响应于音频播放器启动指令之后，接收到首个音频帧的播放指令。其中，音频播放器启动指令由用户通过用户终端进行输入，例如：按压

用户终端上设置的物理按键，或者触控用户终端界面中显示的启动控件。

在本公开的一个实施方式中，在 K 歌应用程序中，如果用户设置了播放背景音乐，检测到用户对录制控件的触发操作后，接收到音频播放器启动指令。

在本公开的一个实施例中，将音频播放器输出首个音频帧的时刻与接收到首个音频帧的播放指令的时刻之间的差值作为播放时延。

其中，音频播放器输出首个音频帧时刻可以是采用音频监测装置监测音频播放器的状态，在监测到音频播放器输出第一音频中的首个音频帧时，获取此时的系统时间作为音频播放器输出首个音频帧时刻。

在本公开的一个实施方式中，所述播放时延包括播放软件时延和播放硬件时延，所述播放软件时延为所述音频播放器的一个缓冲区写入周期，所述播放硬件时延由系统时延函数确定。

具体的，播放软件时延 T_a 是音频播放器接收到一个音频帧并将该音频帧写入播放缓冲区的时长。其中，播放缓冲区内的音频帧用于发送给播放硬件设备进行播放。播放软件时延 T_a 为音频播放器的一个缓冲区写入周期；播放硬件时延 T_b 是音频播放器硬件设备接收到音频信号到音频播放器硬件设备输出音频的时长，播放硬件时延 T_b 可以通过音频系统中硬件延迟函数评估得到。例如算法“`AudioSystem: getOutputLatency`”，需要说明的是，本公开所提供的算法仅是一种示例，对具体算法不做具体限定。

具体的，缓冲区是设备内存空间预留了一定的存储空间，这些存储空间用于缓冲输入或输出数据，缓冲区的目的是减少磁盘的读写次数，因为计算机对于缓冲区的操作远远快于对磁盘的操作，可以提高计算机的运行速度。所述播放软件时延为音频播放器的一个缓冲区写入周期，可理解为在音频播放装置稳定运行的过程中，计算机对音频播放器的缓冲区的一次写入操作所消耗的时间。

在本公开的一个实施方式中，所述基于当前系统时间和播放时延确定所述第一音频的播放时长，包括：计算当前系统时间与所述音频播放器稳定时刻的第一差值；将所述第一差值与所述播放时延之间的第二差值作为所述第一音频的播放时长。

在本公开实施例中，计算当前系统时间与所述音频播放器稳定时刻的第一差值，即统计从音频播放器稳定时刻到当前系统时刻之间的时长。具体的，在音频播放器稳定时刻时，计时器从 0 开始计时，获取当前系统时间对应时刻，计时器的计时时长作为第一差值。或者，将当前系统时间减去音频播放器稳定时刻，得到第一差值。

进一步的，将所述第一差值与所述播放时延之间的第二差值作为所述第一音频的播放时长。其中，所述音频播放器稳定时刻可以理解为音频播放器启动后，经过冷启动可以进行稳定状态的时刻。

在本公开的一个实施方式中，将当前系统时间与音频播放器输出第一音频中的首个音频帧的时刻之间的时长作为第一播放时长。具体的，在音频播放器输出首个音频帧的时刻，计时器从 0 开始计时，获取当前系统时间对应时刻，此时，计时器的计时时长作为第一音频的播放时长。

在本公开的一个实施方式中，在实际应用中，在混音是在线实时处理的过程，因此，当前系统时间可以理解为系统采集到第二音频帧中的当前音频帧的时刻。

S103、基于所述第一音频的播放时长和采集时延计算参考时间戳；其中，所述采集时延包括音频采集器启动时刻到所述第二音频被采集到第一个音频帧的时刻之间的时长。

在本公开的一个实施方式中，所述采集时延包括：采集软件时延和采集硬件时延；所述采集软件时延为所述音频采集器的缓冲区的写入周期，所述采集硬件时延由第二音频中包括的音频帧对应的时间戳确定。

具体的，采集软件时延 T_e 是音频采集到音频信号至记录存储之间的时长，约为音频采集器的一个缓冲区写入周期；采集硬件时延 T_d 可以理解为音频采集器硬件的启动稳定时长，本公开采用时间提取函数来获得。采集时延为采集软件时延 T_e 与采集硬件时延 T_d 之和。

其中，所述采集软件时延为音频采集器的一个缓冲区写入周期，可理解为在音频采集器稳定运行的过程中，计算机对音频采集器的缓冲区的一次写入操作所消耗的时间。采集硬件时延主要指音频采集器中的硬件设备从监测到音频信号到对该音频信号进行处理，得到符合写入条件的音频信号之间的时长。本公开实施例中通过系统算法来获取，例如算法“AudioRecord#getTimestamp”，需要说明的是，本公开所提供的算法仅是一种示例，对具体算法不做具体限定。

在本公开的一个实施方式中，所述基于所述第一音频的播放时长和采集时延计算参考时间戳，包括：计算所述第一音频的播放时长与所述采集时延之间的第二差值；将所述第二差值作为所述参考时间戳。

其中，所述参考时间戳可以理解为对采集到的音频进行混音处理时的一个具备参考意义的时间戳。

在本公开的一个实施方式中，所述基于所述第一音频的播放时长和采集时延计算参考时间戳，包括：计算所述第一音频的播放时长与所述采集时延之间的第二差值；将所述第二差值作为所述参考时间戳。

在本公开实施例中，将上述步骤中计算得到的第一音频的播放时长，减去采集时延，得到一个时间点，将该时间点作为参考时间戳。

本公开实施例中，通过采集时延计算参考时间戳，优化了采集时延对整个时延过程的影响。

S104、基于所述参考时间戳对所述第二音频进行处理，得到第三音频。

在本公开实施例中，所述参考时间戳是通过上述实施例，计算得到的对第二音频进行处理的一个具有参考意义的数据。

在本公开实施例中，对外部声音采集得到的第二音频中包括一个或多个音频帧，针对每个音频帧，均采用步骤 S102 和步骤 S103 中的方式进行处理，得到每个音频帧和参考时间戳的对应关系。

在本公开的一个实施方式中，所述基于所述参考时间戳对所述第二音频进行处理，得到第三音频，包括：确定所述参考时间戳和所述音频帧的对应关系；在所述参考时间戳小于零时，将所述参考时间戳对应的音频帧进行丢包处理；在所述参考时间戳等于或大于零时，将所述等于或大于零的参考时间戳对应的音频帧作为第三音频。

在本公开的一个实施方式中，根据上述计算出的音频帧和参考时间戳的对应关系，在参考时间戳小于零时，将其对应的音频帧进行丢包处理，即不进行混音处理。

具体的，根据上述计算出的音频帧和参考时间戳的对应关系，在参考时间戳大于或者等于零时，将大于或者等于零的参考时间戳对应的音频帧作为第三音频帧。具体的，参考时间戳等于 0 时，将对应的音频帧作为第三音频的首个音频帧。参考时间戳大于 0 时，将参考时间戳按照由小到大的顺序，依次排列其对应的音频帧作为第三音频的音频帧。

S105、将所述第一音频与所述第三音频进行混音处理，得到目标音频。

其中，目标音频是所述第一音频和第三音频进行混音处理后得到的音频。具体的，所述第一音频的首个音频帧和第三音频的首个音频帧进行对齐后，将后续的音频进行混音处理，得到目标音频。

在本公开的一个具体实施方式中，如图 3 所示，第一音频在播放时，音频播放器接收到启动指令之后，会有一个冷启动的过程，才能是音频播放器达到稳定播放的状态，如图 3 所示，将音频播放器达到稳定播放的状态时，系统时钟的时间点标记为音频播放器的稳定时刻 T1，在音频播放器的硬件设备（例如：喇叭、扬声器等）输出首个音频

帧的时刻标记为 T_2 。其中，采集播放器经过冷启动后，开始采集外部声音，得到第二音频，其中第二音频的首个音频帧对应的系统时钟为 T_0 。

在本公开实施例中，从音频播放器输出首个音频帧的时刻 T_2 时，计时器从 0 开始计时，计时器的计时时长 T_{play} 为第一音频的播放时长。

$$T_{\text{play}} = T - T_2 = T - T_1 - T_{\text{play_latency}} \quad (1)$$

其中， T 为当前系统时间， $T_{\text{play_latency}}$ 为播放时延。

在本公开实施例中，音频采集器采集到第二音频中的音频帧的时间为 $T_{\text{sys_capture}}$ ；

$$T_{\text{sys_capture}} = T - T_{\text{capture_latency}} \quad (2)$$

其中， $T_{\text{capture_latency}}$ 为采集时延。

将上述公式 (1) 和公式 (2) 进行联合计算，可以得到参考时间戳 T_{capture} ：

$$T_{\text{capture}} = T_{\text{play}} - T_{\text{capture_latency}} \quad (3)$$

最后以参考时间戳 T_{capture} 作为第二音频的时间戳，如果参考时间戳 $T_{\text{capture}} < 0$ 则进行丢弃，不编码，如果参考时间戳等于 0，将其对应的音频帧作为第三音频的首个音频帧，并与第一音频的首个音频帧进行对齐，然后对第一音频和第三音频进行编码处理，得到目标音频。

图 4 为本公开实施例中的一种音频处理装置的结构示意图，本实施例可适用于音频处理的情况，该音频处理装置可以采用软件和/或硬件的方式实现，该音频处理装置可配置于电子设备中。所述电子设备包括具备音频处理功能的智能终端，包括智能手机、笔记本电脑、平板电脑、数码相机/摄像机、游戏设备等。可选的，所述智能终端包括一触摸屏。

如图 4 所述，本公开实施例提供的音频处理装置 40 主要包括：第

二音频采集模块 41、播放时长确定模块 42、参考时间戳确定模块 43、第三音频确定模块 44 和目标音频确定模块 45。

其中，第二音频采集模块 41，用于在向外播放第一音频的过程中，对外部声音采集得到第二音频；

播放时长确定模块 42，用于基于当前系统时间和播放时延确定所述第一音频的播放时长，其中，所述播放时延为音频播放器启动时刻到所述音频播放器输出音频时刻之间的时长；

参考时间戳确定模块 43，用于基于所述第一音频的播放时长和采集时延计算参考时间戳；其中，所述采集时延包括音频采集器启动时刻到所述第二音频被采集到第一个音频帧的时刻之间的时长；

第三音频确定模块 44，用于基于所述参考时间戳对所述第二音频进行处理，得到第三音频；

目标音频确定模块 45，用于将所述第一音频与所述第三音频进行混音处理，得到目标音频。

本公开实施例提供了一种音频处理装置，用于执行如下流程：在向外播放第一音频的过程中，对外部声音采集得到第二音频；基于当前系统时间和播放时延确定所述第一音频的播放时长，其中，所述播放时延为音频播放器启动时刻到所述音频播放器输出音频时刻之间的时长；基于所述第一音频的播放时长和采集时延计算参考时间戳；其中，所述采集时延包括音频采集器启动时刻到所述第二音频被采集到第一个音频帧的时刻之间的时长；基于所述参考时间戳对所述第二音频进行处理，得到第三音频；将所述第一音频与所述第三音频进行混音处理，得到目标音频。本公开提供的技术方案在第一音频和第二音频进行混音的过程中，优化时延对音频的影响，使得播放的第一音频和录制的第二音频同步，降低时延对音频听感的影响，提高用户 K 歌体验。

在一个可能的实施方式中，所述播放时延包括播放软件时延和播放硬件时延，所述播放软件时延为所述音频播放器的缓冲区写入周期，

所述播放硬件时延由系统时延函数确定。

在一个可能的实施方式中，所述采集时延包括：采集软件时延和采集硬件时延；所述采集软件时延为所述音频采集器的缓冲区的写入周期，所述采集硬件时延由第二音频中包括的音频帧对应的时间戳确定。

在一个可能的实施方式中，所述第二音频包括至少一个或多个音频帧，第三音频确定模块 44，包括：对应关系确定单元，用于确定所述参考时间戳和所述音频帧的对应关系；第一处理单元，用于在所述参考时间戳小于零时，将所述参考时间戳对应的音频帧进行丢包处理；第二处理单元，用于在所述参考时间戳等于或大于零时，将所述等于或大于零的参考时间戳对应的音频帧作为第三音频。

在一个可能的实施方式中，播放时长确定模块 42 包括：第一差值计算单元，用于计算当前系统时间与所述音频播放器稳定时刻的第一差值；第二差值计算单元，用于将所述第一差值与所述播放时延之间的第二差值作为所述第一音频的播放时长。

在一个可能的实施方式中，参考时间戳确定模块 43，具体用于计算所述第一音频的播放时长与所述采集时延之间的第二差值；将所述第二差值作为所述参考时间戳。

本公开实施例提供的音频处理装置，可执行本公开方法实施例所提供的音频处理方法中所执行的步骤，具备执行步骤和有益效果此处不再赘述。

图 5 为本公开实施例中的一种电子设备的结构示意图。下面具体参考图 5，其示出了适于用来实现本公开实施例中的电子设备 500 的结构示意图。本公开实施例中的电子设备 500 可以包括但不限于诸如移动电话、笔记本电脑、数字广播接收器、PDA（个人数字助理）、PAD（平板电脑）、PMP（便携式多媒体播放器）、车载终端（例如车载导航终端）、可穿戴终端设备等等的移动终端以及诸如数字 TV、台式计算机、智能家居设备等等的固定终端。图 5 示出的电子设备仅仅是一

个示例，不应对本公开实施例的功能和使用范围带来任何限制。

如图 5 所示，电子设备 500 可以包括处理装置（例如中央处理器、图形处理器等）501，其可以根据存储在只读存储器（ROM）502 中的程序或者从存储装置 508 加载到随机访问存储器（RAM）503 中的程序而执行各种适当的动作和处理以实现如本公开所述的实施例的图片渲染方法。在 RAM503 中，还存储有终端设备 500 操作所需的各种程序和数据。处理装置 501、ROM502 以及 RAM503 通过总线 504 彼此相连。输入/输出（I/O）接口 505 也连接至总线 504。

通常，以下装置可以连接至 I/O 接口 505：包括例如触摸屏、触摸板、键盘、鼠标、摄像头、麦克风、加速度计、陀螺仪等的输入装置 506；包括例如液晶显示器（LCD）、扬声器、振动器等的输出装置 507；包括例如磁带、硬盘等的存储装置 1008；以及通信装置 509。通信装置 509 可以允许终端设备 500 与其他设备进行无线或有线通信以交换数据。虽然图 5 示出了具有各种装置的终端设备 500，但是应理解的是，并不要求实施或具备所有示出的装置。可以替代地实施或具备更多或更少的装置。

特别地，根据本公开的实施例，上文参考流程图描述的过程可以被实现为计算机软件程序。例如，本公开的实施例包括一种计算机程序产品，其包括承载在非暂态计算机可读介质上的计算机程序，该计算机程序包含用于执行流程图所示的方法的程序代码，从而实现如上所述的页面跳转方法。在这样的实施例中，该计算机程序可以通过通信装置 509 从网络上被下载和安装，或者从存储装置 508 被安装，或者从 ROM 502 被安装。在该计算机程序被处理装置 501 执行时，执行本公开实施例的方法中限定的上述功能。

需要说明的是，本公开上述的计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质或者是上述两者的任意组合。计算机可读存储介质例如可以是——但不限于——电、磁、光、电磁、红外

线、或半导体的系统、装置或器件，或者任意以上的组合。计算机可读存储介质的更具体的例子可以包括但不限于：具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机访问存储器（RAM）、只读存储器（ROM）、可擦式可编程只读存储器（EPROM 或闪存）、光纤、便携式紧凑磁盘只读存储器（CD-ROM）、光存储器件、磁存储器件、或者上述的任意合适的组合。在本公开中，计算机可读存储介质可以是任何包含或存储程序的有形介质，该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。而在本公开中，计算机可读信号介质可以包括在基带中或者作为载波一部分传播的数据信号，其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式，包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读信号介质还可以是计算机可读存储介质以外的任何计算机可读介质，该计算机可读信号介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。计算机可读介质上包含的程序代码可以用任何适当的介质传输，包括但不限于：电线、光缆、RF（射频）等等，或者上述的任意合适的组合。

在一些实施方式中，客户端、服务器可以利用诸如 HTTP（HyperText Transfer Protocol，超文本传输协议）之类的任何当前已知或未来研发的网络协议进行通信，并且可以与任意形式或介质的数字数据通信（例如，通信网络）互连。通信网络的示例包括局域网（“LAN”），广域网（“WAN”），网际网（例如，互联网）以及端对端网络（例如，ad hoc 端对端网络），以及任何当前已知或未来研发的网络。

上述计算机可读介质可以是上述电子设备中所包含的；也可以是单独存在，而未装配入该电子设备中。

上述计算机可读介质承载有一个或者多个程序，当上述一个或者多个程序被该终端设备执行时，使得该终端设备：在向外播放第一音频的过程中，对外部声音采集得到第二音频；基于当前系统时间和播放时

延确定所述第一音频的播放时长，其中，所述播放时延为音频播放器启动时刻到所述音频播放器输出音频时刻之间的时长；基于所述第一音频的播放时长和采集时延计算参考时间戳，其中，所述采集时延包括音频采集器启动时刻到所述第二音频被采集到第一个音频帧的时刻之间的时长；基于所述参考时间戳对所述第二音频进行处理，得到第三音频；将所述第一音频与所述第三音频进行混音处理，得到目标音频。

可选的，当上述一个或者多个程序被该终端设备执行时，该终端设备还可以执行上述实施例所述的其他步骤。

可以以一种或多种程序设计语言或其组合来编写用于执行本公开的操作的计算机程序代码，上述程序设计语言包括但不限于面向对象的程序设计语言—诸如 Java、Smalltalk、C++，还包括常规的过程式程序设计语言—诸如“C”语言或类似的设计语言。程序代码可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中，远程计算机可以通过任意种类的网络——包括局域网(LAN)或广域网(WAN)——连接到用户计算机，或者，可以连接到外部计算机（例如利用因特网服务提供商来通过因特网连接）。

附图中的流程图和框图，图示了按照本公开各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上，流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分，该模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意，在有些作为替换的实现中，方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如，两个接连地表示的方框实际上可以基本并行地执行，它们有时也可以按相反的顺序执行，这依所涉及的功能而定。也要注意的，框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合，可以用执

行规定的功能或操作的专用的基于硬件的系统来实现，或者可以用专用硬件与计算机指令的组合来实现。

描述于本公开实施例中所涉及到的单元可以通过软件的方式实现，也可以通过硬件的方式来实现。其中，单元的名称在某种情况下并不构成对该单元本身的限定。

本文中以上描述的功能可以至少部分地由一个或多个硬件逻辑部件来执行。例如，非限制性地，可以使用的示范类型的硬件逻辑部件包括：现场可编程门阵列（FPGA）、专用集成电路（ASIC）、专用标准产品（ASSP）、片上系统（SOC）、复杂可编程逻辑设备（CPLD）等等。

在本公开的上下文中，机器可读介质可以是有形的介质，其可以包含或存储以供指令执行系统、装置或设备使用或与指令执行系统、装置或设备结合地使用的程序。机器可读介质可以是机器可读信号介质或机器可读储存介质。机器可读介质可以包括但不限于电子的、磁性的、光学的、电磁的、红外的、或半导体系统、装置或设备，或者上述内容的任何合适组合。机器可读存储介质的更具体示例会包括基于一个或多个线的电气连接、便携式计算机盘、硬盘、随机存取存储器（RAM）、只读存储器（ROM）、可擦除可编程只读存储器（EPROM 或快闪存储器）、光纤、便捷式紧凑盘只读存储器（CD-ROM）、光学储存设备、磁储存设备、或上述内容的任何合适组合。

根据本公开的一个或多个实施例，本公开提供了一种音频处理方法，包括：在向外播放第一音频的过程中，对外部声音采集得到第二音频；基于当前系统时间和播放时延确定所述第一音频的播放时长，其中，所述播放时延为音频播放器启动时刻到所述音频播放器输出音频时刻之间的时长；基于所述第一音频的播放时长和采集时延计算参考时间戳，其中，所述采集时延包括音频采集器启动时刻到所述第二音频被采集到第一个音频帧的时刻之间的时长；基于所述参考时间戳对所述第

二音频进行处理，得到第三音频；将所述第一音频与所述第三音频进行混音处理，得到目标音频。

根据本公开的一个或多个实施例，本公开提供了一种音频处理装置，所述装置包括：第二音频采集模块，用于在向外播放第一音频的过程中，对外部声音采集得到第二音频；播放时长确定模块，用于基于当前系统时间和播放时延确定所述第一音频的播放时长，其中，所述播放时延为音频播放器启动时刻到所述音频播放器输出音频时刻之间的时长；参考时间戳确定模块，用于基于所述第一音频的播放时长和采集时延计算参考时间戳；其中，所述采集时延包括音频采集器启动时刻到所述第二音频被采集到第一个音频帧的时刻之间的时长；第三音频确定模块，用于基于所述参考时间戳对所述第二音频进行处理，得到第三音频；目标音频确定模块，用于将所述第一音频与所述第三音频进行混音处理，得到目标音频。

根据本公开的一个或多个实施例，本公开提供了一种电子设备，包括：

一个或多个处理器；

存储器，用于存储一个或多个程序；

当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现如本公开提供的任一所述的音频处理方法。

根据本公开的一个或多个实施例，本公开提供了一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现如本公开提供的任一所述的音频处理方法。

本公开实施例还提供了一种计算机程序产品，该计算机程序产品包括计算机程序或指令，该计算机程序或指令被处理器执行时实现如上所述的音频处理方法。

以上描述仅为本公开的较佳实施例以及对所运用技术原理的说明。本领域技术人员应当理解，本公开中所涉及的公开范围，并不限于上述

技术特征的特定组合而成的技术方案，同时也应涵盖在不脱离上述公开构思的情况下，由上述技术特征或其等同特征进行任意组合而形成的其它技术方案。例如上述特征与本公开中公开的（但不限于）具有类似功能的技术特征进行互相替换而形成的技术方案。

此外，虽然采用特定次序描绘了各操作，但是这不应理解为要求这些操作以所示出的特定次序或以顺序次序执行来执行。在一定环境下，多任务和并行处理可能是有利的。同样地，虽然在上面论述中包含了若干具体实现细节，但是这些不应被解释为对本公开的范围的限制。在单独的实施例的上下文中描述的某些特征还可以组合地实现在单个实施例中。相反地，在单个实施例的上下文中描述的各种特征也可以单独地或以任何合适的子组合的方式实现在多个实施例中。

尽管已经采用特定于结构特征和/或方法逻辑动作的语言描述了本主题，但是应当理解所附权利要求书中所限定的主题未必局限于上面描述的特定特征或动作。相反，上面所描述的特定特征和动作仅仅是实现权利要求书的示例形式。

权 利 要 求 书

1.一种音频处理方法，其特征在于，所述方法包括：

在向外播放第一音频的过程中，对外部声音采集得到第二音频；

基于当前系统时间和播放时延确定所述第一音频的播放时长，其中，所述播放时延为音频播放器启动时刻到所述音频播放器输出音频时刻之间的时长；

基于所述第一音频的播放时长和采集时延计算参考时间戳，其中，所述采集时延包括音频采集器启动时刻到所述第二音频被采集到第一个音频帧的时刻之间的时长；

基于所述参考时间戳对所述第二音频进行处理，得到第三音频；

将所述第一音频与所述第三音频进行混音处理，得到目标音频。

2.根据权利要求1所述的方法，其特征在于，所述播放时延包括播放软件时延和播放硬件时延，所述播放软件时延为所述音频播放器的缓冲区写入周期，所述播放硬件时延由系统时延函数确定。

3.根据权利要求1所述的方法，其特征在于，所述采集时延包括：采集软件时延和采集硬件时延；所述采集软件时延为所述音频采集器的缓冲区的写入周期，所述采集硬件时延由第二音频中包括的音频帧对应的时间戳确定。

4.根据权利要求1所述的方法，其特征在于，所述第二音频包括至少一个或多个音频帧，所述基于所述参考时间戳对所述第二音频进行处理，得到第三音频，包括：

确定所述参考时间戳和所述音频帧的对应关系；

在所述参考时间戳小于零时，将所述参考时间戳对应的音频帧进行丢包处理；

在所述参考时间戳等于或大于零时，将所述等于或大于零的参考时间戳对应的音频帧作为第三音频。

5.根据权利要求1所述的方法，其特征在于，所述基于当前系统时

间和播放时延确定所述第一音频的播放时长，包括：

计算当前系统时间与所述音频播放器稳定时刻的第一差值；

将所述第一差值与所述播放时延之间的第二差值作为所述第一音频的播放时长。

6.根据权利要求1所述的方法，其特征在于，所述基于所述第一音频的播放时长和采集时延计算参考时间戳，包括：

计算所述第一音频的播放时长与所述采集时延之间的第二差值；

将所述第二差值作为所述参考时间戳。

7.一种视音频处理装置，其特征在于，包括：

第二音频采集模块，用于在向外播放第一音频的过程中，对外部声音采集得到第二音频；

播放时长确定模块，用于基于当前系统时间和播放时延确定所述第一音频的播放时长，其中，所述播放时延为音频播放器启动时刻到所述音频播放器输出音频时刻之间的时长；

参考时间戳确定模块，用于基于所述第一音频的播放时长和采集时延计算参考时间戳；其中，所述采集时延包括音频采集器启动时刻到所述第二音频被采集到第一个音频帧的时刻之间的时长；

第三音频确定模块，用于基于所述参考时间戳对所述第二音频进行处理，得到第三音频；

目标音频确定模块，用于将所述第一音频与所述第三音频进行混音处理，得到目标音频。

8.一种电子设备，其特征在于，所述电子设备包括：

一个或多个处理器；

存储装置，用于存储一个或多个程序；

当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现如权利要求1-6中任一项所述的方法。

9.一种计算机可读存储介质，其上存储有计算机程序，其特征在于，该程序被处理器执行时实现如权利要求1-6中任一项所述的方法。

10.一种计算机程序产品，该计算机程序产品包括计算机程序或指令，该计算机程序或指令被处理器执行时实现如权利要求 1-6 中任一项所述的方法。