

视频配音方法、装置、设备及存储介质

本申请要求于 2020 年 7 月 23 日递交的中国专利申请第 202010728035.2
5 号的优先权，该中国专利申请的全文以引入的方式并入以作为本申请的一部分。

技术领域

本公开涉及数据处理领域，尤其涉及一种视频配音方法、装置、设备及
10 存储介质。

背景技术

视频配音，是指通过设备录制声音，为视频的画面配音。目前，通常是
基于单独的录音窗口对音频数据进行录制，得到包含音频数据的录音文件，
15 然后将录音文件添加到目标视频的音频轨道上，人工进行时间轴的匹配，最终完成对目标视频的配音。

但是，对于用户而言，要完成对目标视频的配音，不仅需要录音，还需要
人工将录音文件与目标视频的音频轨道的时间轴进行匹配，显然是一项操作
繁杂的工作，另外，还可能存在因人工匹配时间轴导致的视频配音结果不
20 准确的问题。

发明内容

为了解决上述技术问题或者至少部分地解决上述技术问题，本公开提供
了一种视频配音方法、装置、设备及存储介质，无需人工匹配时间轴，提高
25 了音频配音结果的准确性。

第一方面，本公开提供了一种视频配音方法，所述方法包括：

响应于针对目标视频的第一时间点的录音开始触发操作，从所述第一时间
点对应的视频画面开始，基于时间轴播放所述目标视频以及接收音频数
据；

30 响应于针对所述目标视频的第二时间点的录音结束触发操作，生成包含
从所述第一时间点到所述第二时间点的所述音频数据的录音文件，其中，所

述录音文件用于为视频片段配音，所述视频片段以与所述第一时间点对应的视频画面为起始帧且以与所述第二时间点对应的视频画面为结束帧的，所述录音文件与所述视频片段的时间轴具有联动关系。

一种可选的实施方式中，所述响应于针对所述目标视频的第二时间点的录音结束触发操作，生成包含从所述第一时间点到所述第二时间点的所述录音数据的录音文件之前，还包括：

在基于时间轴播放所述目标视频以及接收音频数据时，控制音频轨道上的指针跟随正在播放的视频画面移动，其中，所述指针用于指示当前录音进度。

10 一种可选的实施方式中，所述方法还包括：

在视频配音暂停的状态下，响应于针对所述指针的拖动操作，利用所述拖动操作的目的帧对应的时间点，更新所述第一时间点。

一种可选的实施方式中，所述响应于针对所述目标视频的第二时间点的录音结束触发操作，生成包含从所述第一时间点到所述第二时间点的所述录音数据的录音文件之前，还包括：

15 在基于时间轴播放所述目标视频以及接收音频数据时，基于所述时间轴显示所述音频数据的波形图。

一种可选的实施方式中，所述响应于针对目标视频的第一时间点的录音开始触发操作，从所述第一时间点对应的视频画面开始，基于时间轴播放所述目标视频以及接收音频数据之前，还包括：

20 在目标视频的第一时间点的位置，显示闪烁动画，其中，所述闪烁动画用于提示基于所述第一时间点准备录音。

一种可选的实施方式中，所述响应于针对所述目标视频的第二时间点的录音结束触发操作，生成包含从所述第一时间点到所述第二时间点的所述音频数据的录音文件之后，还包括：

25 对时间轴具有联动关系的所述录音文件与所述视频片段进行预览播放。

第二方面，本公开还提供了一种视频配音装置，所述装置包括：

接收模块，用于响应于针对目标视频的第一时间点的录音开始触发操作，从所述第一时间点对应的视频画面开始，基于时间轴播放所述目标视频以及接收音频数据；

30 生成模块，用于响应于针对所述目标视频的第二时间点的录音结束触发

操作,生成包含从所述第一时间点到所述第二时间点的所述音频数据的录音文件;其中,所述录音文件用于为视频片段配音,所述视频片段以与所述第一时间点对应的视频画面为起始帧且以与所述第二时间点对应的视频画面为结束帧的,所述录音文件与所述视频片段的时间轴具有联动关系。

5 一种可选的实施方式中,所述装置还包括:

控制模块,用于在基于时间轴播放所述目标视频以及接收音频数据时,控制音频轨道上的指针跟随正在播放的视频画面移动,其中,所述指针用于指示当前录音进度。

第三方面,本公开提供了一种计算机可读存储介质,所述计算机可读存储
10 介质中存储有指令,当所述指令在终端设备上运行时,使得所述终端设备实现上述的方法。

第四方面,本公开提供了一种设备,包括:存储器,处理器,及存储在所述存储器上并可在所述处理器上运行的计算机程序,所述处理器执行所述计算机程序时,实现上述的方法。

15 本公开实施例提供的技术方案与现有技术相比具有如下优点:

本公开实施例提供了一种视频配音方法,在接收针对目标视频的第一时间点的录音开始触发操作时,从第一时间点对应的视频画面开始,基于时间轴播放目标视频以及接收音频数据。当接收到针对目标视频的第二时间点的录音结束触发操作时,生成包含从第一时间点到第二时间点的音频数据的录音文件,同时该录音文件与视频片段的时间轴具有联动关系,该视频片段以
20 第一时间点对应的视频画面为起始帧且以与第二时间点对应的视频画面为结束帧。由于本公开实施例在播放目标视频的同时,基于时间轴录音,生成的录音文件与对应的视频片段的时间轴具有联动关系,也就是说,该录音文件与该视频片段的时间轴无需重新匹配,即可得到较准确的视频配音结果,
25 避免了因人工匹配时间轴导致的操作繁杂且视频配音结果不准确的问题。

附图说明

此处的附图被并入说明书中并构成本说明书的一部分,示出了符合本公开的实施例,并与说明书一起用于解释本公开的原理。

30 为了更清楚地说明本公开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,对于

本领域普通技术人员而言，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

图 1 为本公开实施例提供的一种视频配音方法的流程图；

图 2 为本公开实施例提供的一种未开始录音状态下的视频录音的界面示意图；

图 3 为本公开实施例提供的一种正在录音状态下的视频录音的界面示意图；

图 4 为本公开实施例提供的一种视频配音装置结构框图；

图 5 为本公开实施例提供的一种视频配音设备结构框图。

具体实施方式

为了能够更清楚地理解本公开的上述目的、特征和优点，下面将对本公开的方案进行进一步描述。需要说明的是，在不冲突的情况下，本公开的实施例及实施例中的特征可以相互组合。

在下面的描述中阐述了很多具体细节以便于充分理解本公开，但本公开还可以采用其他不同于在此描述的方式来实施；显然，说明书中的实施例只是本公开的一部分实施例，而不是全部的实施例。

目前，通常基于单独的录音窗口对视频进行配音，在得到录音文件后，需要人工将录音文件与视频的时间轴进行匹配，最终完成视频配音。

但是，人工将录音文件与视频的时间轴进行匹配的方式，不仅存在操作繁杂的问题，还可能会导致视频配音结果不准确。

为此，本公开提供了一种视频配音方法，在接收针对目标视频的第一时间点的录音开始触发操作时，从第一时间点对应的视频画面开始，基于时间轴播放目标视频以及接收音频数据。当接收到针对目标视频的第二时间点的录音结束触发操作时，生成包含从第一时间点到第二时间点的音频数据的录音文件，同时该录音文件与视频片段的时间轴具有联动关系，该视频片段以与第一时间点对应的视频画面为起始帧且以与第二时间点对应的视频画面为结束帧。

由于本公开在播放目标视频的同时，基于时间轴录音，生成的录音文件与对应的视频片段的时间轴具有联动关系，也就是说，该录音文件与该视频片段的时间轴无需重新匹配，即可得到较准确的视频配音结果，避免了因人

工匹配时间轴导致的操作繁杂且视频配音结果不准确的问题。

基于此，本公开实施例提供了一种视频配音方法，参考图 1，为本公开实施例提供的一种视频配音方法的流程图，该方法包括：

5 S101：响应于针对目标视频的第一时间点的录音开始触发操作，从所述第一时间点对应的视频画面开始，基于时间轴播放所述目标视频以及接收音频数据。

例如，目标视频可以是预先获得的任意一段视频，以作为本公开实施例的处理对象。

10 例如，第一时间点可以为目标视频的起始时间点，也可以是目标视频中的任意一个时间点，第一时间点可以根据用户的录音需求确定。

实际应用中，在针对目标视频进行配音之前，首先展示目标视频的音频轨道，如图 2 所示，为本公开实施例提供的一种未开始录音状态下的视频录音的界面示意图，例如，用户可以在音频轨道上确定目标视频中开始录音的第一时间点，然后针对该第一时间点，触发开始录音操作。具体的，可以通过点击录音开始按钮的操作触发从第一时间点开始录音操作。

15 相应的，系统接收到针对目标视频的第一时间点的录音开始触发操作，从目标视频中第一时间点对应的视频画面开始，基于时间轴播放目标视频，并且基于时间轴接收用户输入的音频数据，实现视频录音。

20 本公开实施例中，基于同一时间轴，从第一时间点对应的视频画面开始播放目标视频，同时接收用户输入的音频数据，因此，最终得到的录音文件与目标视频的时间轴具有联动关系，后续无需对录音文件与目标视频进行时间轴的匹配。

25 一种可选的实施方式中，在确定目标视频中的第一时间点之后，在目标视频的第一时间点的位置，视频录音的界面显示闪烁动画，例如，闪烁动画为如图 2 所示的“准备录音...”的闪烁动画，用于提示用户基于第一时间点准备录音，例如，图 2 中指针所在的位置即为目标视频的第一时间点的位置。

例如，显示闪烁动画的第一时间点的位置，可以包括目标视频的音频轨道上第一时间点附近的位置。

30 S102：响应于针对所述目标视频的所述目标音轨上第二时间点的录音结束触发操作，生成包含从所述第一时间点到所述第二时间点的所述音频数据的录音文件。

例如，所述录音文件用于为视频片段配音，该视频片段为以所述第一时间点对应的视频画面为起始帧且以与所述第二时间点对应的视频画面为结束帧，所述录音文件与所述视频片段的时间轴具有联动关系。

例如，第二时间点可以为目标视频的结束时间点，也可以为目标视频中的任意一个时间点，第二时间点可以根据用户的视频录音需求确定。例如，目标视频中的第二时间点在第一时间点之后。

实际应用中，用户可以在音频轨道上确定目标视频中结束录音的第二时间点，然后针对该第二时间点，触发结束录音操作。具体的，可以通过点击录音结束按钮的操作触发从第一时间点到第二时间点的录音结束触发操作。

10 本公开实施例中，在接收到录音结束触发操作后，生成包含有第一时间点到第二时间点期间接收到的音频数据的录音文件，例如，该录音文件用于为以第一时间点对应的视频画面为起始帧且以第二时间点对应的视频画面为结束帧的视频片段配音。

15 由于录音文件与视频片段的时间轴具有联动关系，因此，本公开实施例后续无需对录音文件与视频片段的时间轴进行匹配，即可自动化的得到较准确的视频配音结果。

一种可选的实施方式中，为了便于用户获知当前配音进度，可以在基于时间轴播放目标视频以及接收音频数据的过程中，控制音频轨道上的指针跟随正在播放的视频画面移动，该指针的位置用于指示当前录音进度。如图 3 20 所示，为本公开实施例提供的一种正在录音状态下的视频录音的界面示意图，例如，音频轨道上的指针根据正在播放的视频画面从图 2 中指针所在的起始位置移动到图 3 中的位置。

另一种可选的实施方式中，为了使用户更直观的感受已配音部分的录音数据成功被系统接收，即录音有效，本公开实施例在基于时间轴播放目标 25 视频以及接收音频数据的过程中，基于时间轴显示接收到的音频数据的波形图。通过波形图的展示使得用户更直观的感受已配音部分的音频数据有效。如图 3 所示，在已配音部分的音频轨道上展示音频数据的波形图，同时在波形图下方还可以展示“正在录音...”的提示，用于提示用户当前处于录音状态。

30 另一种可选的实施方式中，可以通过暂停配音按钮，触发视频配音暂停，在视频配音暂停状态下，目标视频停止播放视频画面，同时停止接收用户输

入的音频数据。在视频配音暂停状态下，用户可以通过拖动音频轨道上的指针至目标视频中的某个视频帧对应的时间点，触发对第一时间点的更新，即对视频配音的开始时间的更新。

具体的，在视频配音暂停的状态下，接收针对音频轨道上的指针的拖动操作，利用该拖动操作的目的帧对应的时间点，更新视频配音的第一时间点。例如，拖动操作的目的帧可以为拖动操作对应的松开鼠标时刻的视频帧，将第一时间点更新为该视频帧对应的时间点。后续在确定第二时间点后，可以触发针对更新后的第一时间点到第二时间点的视频片段的配音操作。

另一种可选的实施方式中，在完成视频配音之后，可以针对视频配音对应的录音文件和视频片段进行预览，具体的，对时间轴具有联动关系的录音文件和视频片段进行预览播放，以使用户感受完成配音的视频播放画面。

本公开实施例提供的视频配音方法中，在接收针对目标视频的上第一时间点的录音开始触发操作时，从第一时间点对应的视频画面开始，基于时间轴播放目标视频以及接收音频数据。当接收到针对目标视频的第二时间点的录音结束触发操作时，生成包含从第一时间点到第二时间点的音频数据的录音文件，同时该录音文件与以第一时间点对应的视频画面为起始帧且以与第二时间点对应的视频画面为结束帧的视频片段的时间轴具有联动关系。由于本公开实施例在播放目标视频的同时，基于时间轴录音，生成的录音文件与对应的视频片段的时间轴具有联动关系，也就是说，该录音文件与该视频片段的时间轴无需重新匹配，即可得到较准确的视频配音结果，避免了因人工匹配时间轴导致的操作繁杂且视频配音结果不准确的问题。

与上述方法实施例基于同一个发明构思，本公开还提供了一种视频配音装置，参考图 4，为本公开实施例提供的一种视频配音装置，所述装置包括：

接收模块 401，用于响应于针对目标视频的第一时间点的录音开始触发操作，从所述第一时间点对应的视频画面开始，基于时间轴播放所述目标视频以及接收音频数据；

生成模块 402，用于响应于针对所述目标视频的第二时间点的录音结束触发操作，生成包含从所述第一时间点到所述第二时间点的所述音频数据的录音文件，例如，所述录音文件用于为视频片段配音，所述视频片段以与所述第一时间点对应的视频画面为起始帧且以与所述第二时间点对应的视频画面为结束帧，所述录音文件与所述视频片段的时间轴具有联动关系。

一种可选的实施方式中，所述装置还包括：

控制模块，用于在基于时间轴播放所述目标视频以及接收音频数据时，控制音频轨道上的指针跟随正在播放的视频画面移动，所述指针用于指示当前录音进度。

5 一种可选的实施方式中，所述装置还包括：

更新模块，用于在视频配音暂停的状态下，响应于针对所述指针的拖动操作，利用所述拖动操作的目的帧对应的时间点，更新所述第一时间点。

一种可选的实施方式中，所述装置还包括：

10 第一显示模块，用于在基于时间轴播放所述目标视频以及接收音频数据时，基于所述时间轴显示所述音频数据的波形图。

一种可选的实施方式中，所述装置还包括：

第二显示模块，用于在目标视频的第一时间点的位置，显示闪烁动画，所述闪烁动画用于提示基于所述第一时间点准备录音。

一种可选的实施方式中，所述装置还包括：

15 预览模块，用于对所述录音文件与所述视频片段进行预览播放。

本公开实施例提供的视频配音装置，在接收针对目标视频的第一时间点的录音开始触发操作时，从第一时间点对应的视频画面开始，基于时间轴播放目标视频以及接收音频数据。当接收到针对目标视频的第二时间点的录音结束触发操作时，生成包含从第一时间点到第二时间点的音频数据的录音文件，同时该录音文件与以第一时间点对应的视频画面为起始帧且以与第二时间点对应的视频画面为结束帧的视频片段的时间轴具有联动关系。由于本公开实施例在播放目标视频的同时，基于时间轴录音，生成的录音文件与对应的视频片段的时间轴具有联动关系，也就是说，该录音文件与该视频片段的时间轴无需重新匹配，即可得到较准确的视频配音结果，避免了因人工匹配时间轴导致的操作繁杂且视频配音结果不准确的问题。

20

25

另外，本公开实施例还提供了一种视频配音设备，参见图 5 所示，可以包括：

处理器 501、存储器 502、输入装置 503 和输出装置 504。视频配音设备中的处理器 501 的数量可以一个或多个，图 5 中以一个处理器为例。在本发明的一些实施例中，处理器 501、存储器 502、输入装置 503 和输出装置 504 可通过总线或其它方式连接，其中，图 5 中以通过总线连接为例。

30

存储器 502 可用于存储软件程序以及模块，处理器 501 通过运行存储在存储器 502 的软件程序以及模块，从而执行视频配音设备的各种功能应用以及数据处理。存储器 502 可主要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、至少一个功能所需的应用程序等。此外，存储器 502 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、闪存器件、或其他易失性固态存储器件。输入装置 503 可用于接收输入的数字或字符信息，以及产生与视频配音设备的用户设置以及功能控制有关的信号输入。

具体在本实施例中，处理器 501 会按照如下的指令，将一个或一个以上的应用程序的进程对应的可执行文件加载到存储器 502 中，并由处理器 501 来运行存储在存储器 502 中的应用程序，从而实现上述视频配音设备的各种功能。

需要说明的是，在本文中，诸如“第一”和“第二”等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且，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、物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、物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，并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

以上所述仅是本公开的具体实施方式，使本领域技术人员能够理解或实现本公开。对这些实施例的多种修改对本领域的技术人员来说将是显而易见的，本文中所定义的一般原理可以在不脱离本公开的精神或范围的情况下，在其它实施例中实现。因此，本公开将不会被限制于本文所述的这些实施例，而是要符合与本文所公开的原理和新颖特点相一致的最宽的范围。