

一种音视频处理方法、装置、设备及存储介质

相关申请的交叉引用

本公开要求于 2021 年 9 月 22 日提交的，申请名称为“一种音视频处理方法、装置、设备及存储介质”的、中国专利申请号为“202111109213.4”的优先权，该中国专利申请的全部内容通过引用结合在本公开中。

技术领域

本公开涉及数据处理领域，尤其涉及一种音视频处理方法、装置、设备及存储介质。

背景技术

随着互联网信息的日益丰富，观看音视频已成为人们日常生活中的一项娱乐活动。为了提高用户的观看体验，在发布各类音视频之前，对音视频进行剪辑是一个重要环节。

目前，在音视频剪辑过程中，对于无效词汇剪辑等一些细微改动，通常是用户反复收听音视频，同时微调时间起始点和结束点，以对音视频进行剪辑，操作繁琐，音视频剪辑的准确性有待提高。

技术解决方案

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，本公开实施例提供了一种音视频处理方法，能够提高音视频剪辑的精确性，简化用户操作。

第一方面，本公开提供了一种音视频处理方法，所述方法包括：

展示待编辑音视频对应的文本数据；其中，所述文本数据与所述待编辑音视频的音视频时间戳具有映射关系；

以及，按照时间轴轨道展示所述待编辑音视频；

响应于针对所述文本数据中的目标文本数据触发的预设操作，确定所述目标文本数据对应的音视频时间戳，作为目标音视频时间戳；

基于所述预设操作，对所述待编辑音视频中与所述目标音视频时间戳对应的音视频片段进行处理。

一种可选的实施方式中，所述方法还包括：

展示针对预设关键词或预设静音片段的第一编辑入口；

响应于针对所述第一编辑入口的触发操作，按照预设第二显示样式，对所述文本数据中的所述预设关键词或所述预设静音片段进行显示。

一种可选的实施方式中，所述第一编辑入口对应于第一编辑卡片，所述第一编辑卡片上设置有一键删除控件；所述响应于针对所述第一编辑入口的触发操作，按照预设第二显示样式，对所述文本数据中的所述预设关键词或所述预设静音片段进行显示之后，还包括：

响应于针对所述一键删除控件的触发操作，从所述文本数据中删除所述预设关键词或所述预设静音片段。

一种可选的实施方式中，所述方法还包括：

在第二编辑卡片上展示语音增强控件；

响应于针对所述语音增强控件的触发操作，对所述待编辑音视频中的人声进行增强处理。

一种可选的实施方式中，所述方法还包括：

基于所述待编辑音视频的音乐体裁和/或所述待编辑音视频对应的文本数据中的内容，确定所述待编辑音视频对应的配乐；

将所述配乐添加到所述待编辑音视频片段中。

一种可选的实施方式中，所述方法还包括：

在第三编辑卡片上展示响度均衡控件；

响应于针对所述响度均衡控件的触发操作，对所述待编辑音视频中音量的响度进行归一化处理。

一种可选的实施方式中，所述方法还包括：

在第四编辑卡片上展示智能片花控件；

响应于针对所述智能片花控件的触发操作，对所述待编辑音视频中的前预设时间段内的音视频片段中的音乐音量与人声音量进行调节，得到音量调节后音视频片段；其中，所述音量调节后音视频片段中的音乐音量与人声音量成反比。

一种可选的实施方式中，所述预设操作包括选中操作，所述基于所述预设操作，对所述待编辑音视频中与所述目标音视频时间戳对应的音视频片段进行处理，包括：

按照预设第一显示样式，对所述待编辑音视频中与所述目标音视频时间戳对应的音视频片段进行显示。

一种可选的实施方式中，所述预设操作包括删除操作，所述基于所述预设操作，对所述待编辑音视频中与所述目标音视频时间戳对应的音视频片段进行处理，包括：

基于所述删除操作，对所述待编辑音视频中与所述目标音视频时间戳对应的音视频片段进行删除。

一种可选的实施方式中，所述预设操作包括修改操作，所述基于所述预设操作，对所述待编辑音视频中与所述目标音视频时间戳对应的音视频片段进行处理，包括：

获取所述修改操作对应的修改后文本数据；

基于所述修改后文本数据和所述待编辑音视频中的音色信息生成音视频片段，作为待修改音视频片段；

利用所述待修改音视频片段，对所述待编辑音视频中与所述目标音视频时间戳对应的音视频片段进行替换处理。

一种可选的实施方式中，所述方法还包括：

当接收到在所述文本数据中针对第一文本数据的增加操作时，基于所述第一文本数据和所述待编辑音视频中的音色信息，生成第一音视频片段；

基于所述第一文本数据在所述文本数据中的位置信息，确定所述第一文本数据对应的第一音视频时间戳；

基于所述第一音视频时间戳，将所述第一音视频片段添加到所述待编辑音视频中。

第二方面，本公开还提供了一种音视频处理装置，所述装置包括：

第一展示模块，用于展示待编辑音视频对应的文本数据；其中，所述文本数据与所述待编辑音视频的音视频时间戳具有映射关系；

第二展示模块，用于按照时间轴轨道展示所述待编辑音视频；

确定模块，用于响应于针对所述文本数据中的目标文本数据触发的预设操作，确定所述目标文本数据对应的音视频时间戳，作为目标音视频时间戳；

编辑模块，用于基于所述预设操作，对所述待编辑音视频中与所述目标音视频时间戳对应的音视频片段进行处理。

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。

第五方面，本公开提供了一种计算机程序产品，所述计算机程序产品包括计算机程序/指令，所述计算机程序/指令被处理器执行时实现上述的方法。

本公开实施例提供的技术方案与相关技术相比具有如下优点：

本公开实施例提供了一种音视频处理方法，通过展示待编辑音视频对应的文本数据，响应于针对文本数据中的目标文本数据触发的预设操作，确定目标文本数据对应的音视频时间戳，作为目标音视频时间戳，以及基于预设操作，对待编辑音视频中与目标音视频时间戳对应的音视频片段进行处理。可见，本公开实施例提供的音视频处理方

法能够提高音视频剪辑的精确性，简化了用户操作，降低用户操作门槛。

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。

为了更清楚地说明本公开实施例或相关技术中的技术方案，下面将对实施例或相关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音视频处理方法的流程图；

图 2 为本公开实施例提供的一种音视频处理界面的示意图；

图 3 为本公开实施例提供的另一种音视频处理界面的示意图；

图 4 为本公开实施例提供的另一种音视频处理方法的流程图；

图 5 为本公开实施例提供的另一种音视频处理界面的示意图；

图 6 为本公开实施例提供的另一种音视频处理界面的示意图；

图 7 为本公开实施例提供的一种音视频处理装置的结构示意图；

图 8 为本公开实施例提供的一种音视频处理设备的结构示意图。

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。

本公开实施例提供了一种音视频处理方法，参考图 1，为本公开实施例提供的一种音视频处理方法的流程图，该方法包括：

S101：展示待编辑音视频对应的文本数据。

其中，文本数据与待编辑音视频的音视频时间戳具有映射关系，音视频时间戳用于指示每帧音视频的播放时间。

本公开实施例中，待编辑音视频包括但不限于录制得到的音视频、基于脚本得到的音视频等。文本数据可以是对待编辑音视频进行语音识别得到的，也可以是脚本，其中，文本数据为脚本的情况下，可以将文本数据与待编辑音视频匹配得到前述文本数据与待编辑音视频的音视频时间戳的映射关系，语音识别方法包括但不限于 ASR(Automatic Speech Recognition，自动语音识别) 技术。

本实施例中，可在界面上展示文本数据，作为一种示例，界面例如图 2 所示，图 2 中区域 P 示出了展示的文本数据。待编辑音视频包含不同用户的语音的情况下，可以确定不同用户的文本数据，如图 2 中展示的用户 a 和用户 b 的文本数据。

S102：按照时间轴轨道展示待编辑音视频。

本实施例中，可在界面上按照时间轴轨道展示待编辑音视频，作为一种示例，图 2 中区域 Q 示出了展示的待编辑音视频。

需要说明的是，步骤 102 的执行顺序不作具体限制。

S103：响应于针对文本数据中的目标文本数据触发的预设操作，确定目标文本数据对应的音视频时间戳，作为目标音视频时间戳。

本实施例中，预设操作包括但不限于选中操作、删除操作、修改操作。由于文本数据与待编辑音视频的音视频时间戳具有映射关系，因此针对文本数据中的目标文本数据，可以根据映射关系确定目标文本数据对应的目标音视频时间戳。

S104：基于预设操作，对待编辑音视频中与目标音视频时间戳对应的音视频片段进行处理。

本公开实施例中，基于音视频时间戳可确定待编辑音视频中的对应音视频片段，通过对待编辑音视频中与目标音视频时间戳对应的音视频片段进行处理，实现基于文本剪辑音视频，通过对文本的剪辑，联动的剪辑对应的音视频片段，能够实现对音视频精确性较高的剪辑。

一种可选的实施方式中，预设操作包括选中操作，则基于预设操作，对待编辑音视频中与目标音视频时间戳对应的音视频片段进行处理，包括：按照预设第一显示样式，对待编辑音视频中与目标音视频时间戳对应的音视频片段进行显示。

作为一种示例，第一显示样式例如是高亮显示，图 3 示出了另一种界面的示意图，参照图 3，基于选中操作，可以对目标文本数据进行高亮显示，以及基于时间轴轨道对目标音视频时间戳对应的音视频片段进行高亮显示，高亮显示部分如图 3 中虚线部分。

一种可选的实施方式中，预设操作包括删除操作，则基于预设操作，对待编辑音视频中与目标音视频时间戳对应的音视频片段进行处理，包括：基于删除操作，对待编辑音视频中与目标音视频时间戳对应的音视频片段进行删除。

其中，基于删除操作，可以对目标文本数据进行删除，以及基于对目标音视频时间戳对应的音视频片段进行删除。例如图 3 所示，在选中目标文本数据后，可以展示删除控件，响应于针对删除控件的触发操作，删除目标文本数据以及目标音视频时间戳对应的音视频片段。

一种可选的实施方式中，预设操作包括修改操作，则基于预设操作，对待编辑音视频中与目标音视频时间戳对应的音视频片段进行处理，包括：获取修改操作对应的修改后文本数据；基于修改后文本数据和待编辑音视频中的音色信息生成音视频片段，作为待修改音视频片段；利用待修改音视频片段，对待编辑音视频中与目标音视频时间戳对应的音视频片段进行替换处理。

其中，基于修改操作，可以对目标文本数据进行修改，例如图 3 所示，在选中目标文本数据后，可以展示修改控件，响应于针对修改控件的触发操作，并根据接收的修改内容生成修改后文本数据。以及基于修改后文本数据和音色信息生成待修改音视频片段，根据待修改音视频片段替换目标音视频时间戳对应的音视频片段，实现对待编辑音视频的修改。

本公开实施例提供的音视频处理方法中，通过展示待编辑音视频对应的文本数据，响应于针对文本数据中的目标文本数据触发的预设操作，确定目标文本数据对应的音视频时间戳，作为目标音视频时间戳，以及基于预设操作，对待编辑音视频中与目标音视频时间戳对应的音视频片段进行处理。可见，本公开实施例提供的音视频处理方法能够基于文本剪辑音视频，由于文本与音视频时间戳存在映射关系，通过对文本的剪辑，联动的剪辑对应的音视频片段，能够实现对音视频精确性较高的剪辑，并且，通过展示与音视频时间戳具有映射关系的文本数据，能够直观的展示音视频内容，相较于相关技术中用户剪辑音视频内容的方案，简化了用户操作，降低用户操作门槛。

基于上述实施例，在音视频处理场景中，为了提高听感体验，存在针对无效语气词剪辑、配乐、片花制作等多种功能的需求。根据本公开实施例的方法，可以便捷的实现上述功能，降低用户使用门槛，说明如下。

一种可选的实施方式中，在对话中通常会出现“嗯”“呃”“那个”等无效语气词以及静音片段，因此，为了保证对话的连贯性，存在对待编辑音视频进行编辑，以删除前述无效语气词及静音片段的需求。

因此，如图4所示，本公开实施例的音视频处理方法还包括：

步骤401，展示针对预设关键词或预设静音片段的第一编辑入口。

本实施例中，可以对展示的待编辑音视频对应的文本数据进行检测，确定文本数据中的预设关键词或预设静音片段，并在文本数据中存在预设关键词或预设静音片段的情况下，展示第一编辑入口。作为一种示例，图3中区域A所示控件为第一编辑入口，第一编辑入口上展示“修改建议01：去除无效语气词”信息。

其中，预设关键词可包括无效语气词等词汇，确定文本数据中预设关键词的实现方式有多种，例如可以通过匹配的方式确定文本数据中的预设关键词，再例如可以基于自然语言处理技术，确定文本数据中的预设关键词。

其中，预设静音片段根据两个相邻文字所对应的音视频时间戳之间的间隔确定，例如当该间隔大于预设阈值的情况下，确定两个相邻文字之间存在预设静音片段。静音片段可在界面上以空格的形式展示，可选地，可根据间隔的值，确定静音片段的展示长度。

步骤 402，响应于针对第一编辑入口的触发操作，按照预设第二显示样式，对文本数据中的预设关键词或预设静音片段进行显示。

其中，针对第一编辑入口的触发操作包括但不限于点击操作、语音指令、触摸轨迹。第二显示样式可以是高亮显示，也可以是其他形式的显示样式，此处不作具体限制。

图 5 示出了一种界面的示意图，图 5 中预设关键词“呃”“嗯”“那个”在界面中高亮显示，如虚线部分所示。

步骤 403，响应于针对一键删除控件的触发操作，从文本数据中删除预设关键词或预设静音片段。

本公开实施例中，第一编辑入口对应于第一编辑卡片，第一编辑卡片上设置有一键删除控件。响应于针对第一编辑入口的触发操作，展示第一编辑卡片，第一编辑卡片的展示方式包括但不限于下拉选项、悬浮窗等。

举例而言，参照图 5，第一编辑卡片如图 5 中区域 B 所示，可以针对每一个预设关键词统计出现次数，并在第一编辑卡片中展示预设关键词及对应的出现次数。

可选地，响应于针对预设关键词中目标关键词的触发操作，将目标关键词从预设关键词中去除，并同步修改第一编辑卡片中展示的预设关键词的出现次数，以使用户可以通过点击等操作，去除不属于无效语气词的关键词，以避免其被一键删除。

本实施例中，能够将预设关键词或预设静音片段删除操作以编辑卡片的形式呈现，提供了一键式操作，节省编辑时长，简化用户操作，降低用户使用门槛。

一种可选的实施方式中，音视频处理方法还包括：在第二编辑卡

片上展示语音增强控件；响应于针对语音增强控件的触发操作，对待编辑音视频中的人声进行增强处理。

本实施例中，展示针对待编辑音视频的第二编辑入口，第二编辑入口对应于第二编辑卡片，第二编辑卡片上设置语音增强控件。例如，可以基于待编辑音视频进行噪声检测，并在检测到噪声的情况下，展示第二编辑入口，作为一种示例，图2中区域C所示控件为第二编辑入口，第二编辑入口上展示“强化建议：语音增强”信息。进而，响应于针对第二编辑入口的触发操作，展示第二编辑卡片。

参照图6，第二编辑卡片如图6中区域D所示，第二编辑卡片中展示了语音增强控件“增强语音”，响应于对该语音增强控件的触发操作，对待编辑音视频中的人声进行增强处理，触发操作包括但不限于点击操作、语音指令、触摸轨迹。

本实施例中，能够将语音增强操作以编辑卡片的形式呈现，提供了一键式操作，能够增强用户人声满足听感体验，并简化用户操作，降低用户使用门槛。

一种可选的实施方式中，音视频处理方法还包括：基于待编辑音视频的音乐体裁和/或待编辑音视频对应的文本数据中的内容，确定待编辑音视频对应的配乐；将配乐添加到待编辑音视频片段中。

本实施例中，可以预先设置多个标签，每个标签与一个或多个配乐之间存在映射关系，基于待编辑音视频的音乐体裁和/或待编辑音视频对应的文本数据中的内容，确定与音乐体裁和/或文本数据中的内容对应的标签，基于标签与配乐之间的映射关系，确定待编辑音视频对应的配乐。

作为一种示例，针对待编辑音视频对应的文本数据中的内容，基于自然语言处理技术确定内容的主题为“运动”，进而确定“运动”标签对应的配乐，为待编辑音视频对应的配乐，并将该配乐添加到待编辑音视频片段中。

作为另一种示例，基于待编辑音视频的音乐体裁，确定对应的标

签，将该标签对应的配乐，作为待编辑音视频对应的配乐，并将该配乐添加到待编辑音视频片段中。

本实施例中，可以基于文本数据的内容、体裁智能推荐配乐，以满足添加配乐的场景需求，丰富听感的多样性、提高听感体验，并简化用户操作，降低用户使用门槛。

一种可选的实施方式中，音视频处理方法还包括：在第三编辑卡片上展示响度均衡控件；响应于针对响度均衡控件的触发操作，对待编辑音视频中音量的响度进行归一化处理。

本实施例中，展示针对待编辑音视频的第三编辑入口，第三编辑入口对应于第三编辑卡片，第三编辑卡片上设置响度均衡控件。例如，可以基于待编辑音视频进行音量的响度检测，并在检测到待编辑音视频不满足预设的响度均衡条件的情况下，展示第三编辑入口。进而，响应于针对第三编辑入口的触发操作，展示第三编辑卡片，响应于针对响度均衡控件的触发操作，对待编辑音视频中音量的响度进行归一化处理，例如使待编辑音视频中音量的响度处于预设范围内。

本实施例中，能够将响度均衡操作以编辑卡片的形式呈现，提供了一键式操作，能够提高听感体验，并简化用户操作，降低用户使用门槛。

一种可选的实施方式中，音视频处理方法还包括：在第四编辑卡片上展示智能片花控件；响应于针对智能片花控件的触发操作，对待编辑音视频中的前预设时间段内的音视频片段中的音乐音量与人声音量进行调节，得到音量调节后音视频片段。

本实施例中，展示针对待编辑音视频的第四编辑入口，第四编辑入口对应于第四编辑卡片，第四编辑卡片上设置智能片花控件。响应于针对第四编辑入口的触发操作，展示第四编辑卡片，响应于针对智能片花控件的触发操作，对待编辑音视频中的前预设时间段内的音视频片段中的音乐音量与人声音量进行调节，例如将人声音量增大第一音量值，将音乐音量降低第二音量值，或者在检测到人声的音视频片

段中，将音乐音量降低第三音量值，得到音量调节后音视频片段。

其中，音量调节后音视频片段中的音乐音量与人声音量成反比。

可选地，基于第四编辑卡片上展示的智能片花控件，还可以实现片头生成，例如，响应于针对智能片花控件的触发操作，确定当前选中的第二文本数据和与第二文本数据对应的第二音视频片段，将第二文本数据和第二音视频片段复制并粘贴至预设的片头区域，实现片花的效果。

本实施例中，能够将智能片花功能以编辑卡片的形式呈现，提供了一键式操作，实现片花的效果，并简化用户操作，降低用户使用门槛。

一种可选的实施方式中，音视频处理方法还包括：当接收到在文本数据中针对第一文本数据的增加操作时，基于第一文本数据和待编辑音视频中的音色信息，生成第一音视频片段；基于第一文本数据在文本数据中的位置信息，确定第一文本数据对应的第一音视频时间戳；基于第一音视频时间戳，将第一音视频片段添加到待编辑音视频中。

本实施例中，第一文本数据可以是响应于输入操作得到的，也可以是基于已有文本数据复制得到的。可根据待编辑音视频获取各用户的音色信息，当增加第一文本数据时，根据第一文本数据在文本数据中的位置信息确定对应的第一音视频时间戳，并在第一音视频时间戳的位置处添加第一音视频片段。

需要说明的是，前述编辑入口可以基于检测结果自动展示，也可以响应于触发操作在界面上展示。

本实施例中，采用音色克隆和语音播报技术，基于增加的文本克隆音色、智能生成音视频片段，实现了基于文本输入增加音视频片段，降低重新录制带来的时间成本和编辑成本，简化用户操作。

基于上述方法实施例，本公开还提供了一种音视频处理装置，参考图7，为本公开实施例提供的一种音视频处理装置的结构示意图，所述装置包括：

第一展示模块 701, 用于展示待编辑音视频对应的文本数据; 其中, 所述文本数据与所述待编辑音视频的音视频时间戳具有映射关系。

第二展示模块 702, 用于按照时间轴轨道展示所述待编辑音视频。

确定模块 703, 用于响应于针对所述文本数据中的目标文本数据触发的预设操作, 确定所述目标文本数据对应的音视频时间戳, 作为目标音视频时间戳。

编辑模块 704, 用于基于所述预设操作, 对所述待编辑音视频中与所述目标音视频时间戳对应的音视频片段进行处理。

一种可选的实施方式中, 音视频处理装置还包括:

第一处理模块, 用于展示针对预设关键词或预设静音片段的第一编辑入口; 响应于针对所述第一编辑入口的触发操作, 按照预设第二显示样式, 对所述文本数据中的所述预设关键词或所述预设静音片段进行显示。

一种可选的实施方式中, 第一编辑入口对应于第一编辑卡片, 所述第一编辑卡片上设置有一键删除控件; 第一编辑模块还用于, 响应于针对所述一键删除控件的触发操作, 从所述文本数据中删除所述预设关键词或所述预设静音片段。

一种可选的实施方式中, 音视频处理装置还包括:

第二处理模块, 用于在第二编辑卡片上展示语音增强控件; 响应于针对所述语音增强控件的触发操作, 对所述待编辑音视频中的人声进行增强处理。

一种可选的实施方式中, 音视频处理装置还包括:

第一添加模块, 用于基于所述待编辑音视频的音乐体裁和/或所述待编辑音视频对应的文本数据中的内容, 确定所述待编辑音视频对应的配乐; 将所述配乐添加到所述待编辑音视频片段中。

一种可选的实施方式中, 音视频处理装置还包括:

第三处理模块, 用于在第三编辑卡片上展示响度均衡控件; 响应于针对所述响度均衡控件的触发操作, 对所述待编辑音视频中音量的

响度进行归一化处理。

一种可选的实施方式中，音视频处理装置还包括：

第四处理模块，用于在第四编辑卡片上展示智能片花控件；响应于针对所述智能片花控件的触发操作，对所述待编辑音视频中的前预设时间段内的音视频片段中的音乐音量与人声音量进行调节，得到音量调节后音视频片段；其中，所述音量调节后音视频片段中的音乐音量与人声音量成反比。

一种可选的实施方式中，预设操作包括选中操作，编辑模块 704 具体用于：按照预设第一显示样式，对待编辑音视频中与目标音视频时间戳对应的音视频片段进行显示。

一种可选的实施方式中，预设操作包括删除操作，编辑模块 704 具体用于：基于所述删除操作，对所述待编辑音视频中与所述目标音视频时间戳对应的音视频片段进行删除。

一种可选的实施方式中，预设操作包括修改操作，编辑模块 704 具体用于：获取所述修改操作对应的修改后文本数据；基于所述修改后文本数据和所述待编辑音视频中的音色信息生成音视频片段，作为待修改音视频片段；利用所述待修改音视频片段，对所述待编辑音视频中与所述目标音视频时间戳对应的音视频片段进行替换处理。

一种可选的实施方式中，音视频处理装置还包括：

第二添加模块，用于当接收到在所述文本数据中针对第一文本数据的增加操作时，基于所述第一文本数据和所述待编辑音视频中的音色信息，生成第一音视频片段；基于所述第一文本数据在所述文本数据中的位置信息，确定所述第一文本数据对应的第一音视频时间戳；基于所述第一音视频时间戳，将所述第一音视频片段添加到所述待编辑音视频中。

前述实施例对音视频处理方法的解释说明，同样适用于本实施例的音视频处理装置，此处不再赘述。

本公开实施例提供的音视频处理装置中，通过展示待编辑音视频

对应的文本数据，响应于针对文本数据中的目标文本数据触发的预设操作，确定目标文本数据对应的音视频时间戳，作为目标音视频时间戳，以及基于预设操作，对待编辑音视频中与目标音视频时间戳对应的音视频片段进行处理。可见，本公开实施例提供的音视频处理方法能够基于文本剪辑音视频，由于文本与音视频时间戳存在映射关系，通过对文本的剪辑，联动的剪辑对应的音视频片段，能够实现对音视频精确性较高的剪辑，并且，通过展示与音视频时间戳具有映射关系的文本数据，能够直观的展示音视频内容，相较于相关技术中用户剪辑音视频内容的方案，简化了用户操作，降低用户操作门槛。

除了上述方法和装置以外，本公开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本公开实施例所述的音视频处理方法。

本公开实施例还提供了一种计算机程序产品，所述计算机程序产品包括计算机程序/指令，所述计算机程序/指令被处理器执行时实现本公开实施例所述的音视频处理方法。

另外，本公开实施例还提供了一种音视频处理设备，参见图 8 所示，可以包括：

处理器 801、存储器 802、输入装置 803 和输出装置 804。音视频处理设备中的处理器 801 的数量可以一个或多个，图 8 中以一个处理器为例。在本公开的一些实施例中，处理器 801、存储器 802、输入装置 803 和输出装置 804 可通过总线或其它方式连接，其中，图 8 中通过总线连接为例。

存储器 802 可用于存储软件程序以及模块，处理器 801 通过运行存储在存储器 802 的软件程序以及模块，从而执行音视频处理设备的各种功能应用以及数据处理。存储器 802 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 802 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 803 可用于接收输入的数字或字符信息，以及产生与音视频处理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 801 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 802 中，并由处理器 801 来运行存储在存储器 802 中的应用程序，从而实现上述音视频处理设备的各种功能。

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。

权 利 要 求 书

1、一种音视频处理方法，所述方法包括：

展示待编辑音视频对应的文本数据；其中，所述文本数据与所述待编辑音视频的音视频时间戳具有映射关系；

以及，按照时间轴轨道展示所述待编辑音视频；

响应于针对所述文本数据中的目标文本数据触发的预设操作，确定所述目标文本数据对应的音视频时间戳，作为目标音视频时间戳；

基于所述预设操作，对所述待编辑音视频中与所述目标音视频时间戳对应的音视频片段进行处理。

2、根据权利要求1所述的方法，其中，所述方法还包括：

展示针对预设关键词或预设静音片段的第一编辑入口；

响应于针对所述第一编辑入口的触发操作，按照预设第二显示样式，对所述文本数据中的所述预设关键词或所述预设静音片段进行显示。

3、根据权利要求2所述的方法，其中，所述第一编辑入口对应于第一编辑卡片，所述第一编辑卡片上设置有一键删除控件；所述响应于针对所述第一编辑入口的触发操作，按照预设第二显示样式，对所述文本数据中的所述预设关键词或所述预设静音片段进行显示之后，还包括：

响应于针对所述一键删除控件的触发操作，从所述文本数据中删除所述预设关键词或所述预设静音片段。

4、根据权利要求1所述的方法，其中，所述方法还包括：

在第二编辑卡片上展示语音增强控件；

响应于针对所述语音增强控件的触发操作，对所述待编辑音视频中的人声进行增强处理。

5、根据权利要求1所述的方法，其中，所述方法还包括：

基于所述待编辑音视频的音乐体裁和/或所述待编辑音视频对应的

文本数据中的内容，确定所述待编辑音视频对应的配乐；

将所述配乐添加到所述待编辑音视频片段中。

6、根据权利要求1所述的方法，其中，所述方法还包括：

在第三编辑卡片上展示响度均衡控件；

响应于针对所述响度均衡控件的触发操作，对所述待编辑音视频中音量的响度进行归一化处理。

7、根据权利要求1所述的方法，其中，所述方法还包括：

在第四编辑卡片上展示智能片花控件；

响应于针对所述智能片花控件的触发操作，对所述待编辑音视频中的前预设时间段内的音视频片段中的音乐音量与人声音量进行调节，得到音量调节后音视频片段；其中，所述音量调节后音视频片段中的音乐音量与人声音量成反比。

8、根据权利要求1所述的方法，其中，所述预设操作包括选中操作，所述基于所述预设操作，对所述待编辑音视频中与所述目标音视频时间戳对应的音视频片段进行处理，包括：

按照预设第一显示样式，对所述待编辑音视频中与所述目标音视频时间戳对应的音视频片段进行显示。

9、根据权利要求1所述的方法，其中，所述预设操作包括删除操作，所述基于所述预设操作，对所述待编辑音视频中与所述目标音视频时间戳对应的音视频片段进行处理，包括：

基于所述删除操作，对所述待编辑音视频中与所述目标音视频时间戳对应的音视频片段进行删除。

10、根据权利要求1所述的方法，其中，所述预设操作包括修改操作，所述基于所述预设操作，对所述待编辑音视频中与所述目标音视频时间戳对应的音视频片段进行处理，包括：

获取所述修改操作对应的修改后文本数据；

基于所述修改后文本数据和所述待编辑音视频中的音色信息生成音视频片段，作为待修改音视频片段；

利用所述待修改音视频片段，对所述待编辑音视频中与所述目标音视频时间戳对应的音视频片段进行替换处理。

11、根据权利要求 1 所述的方法，其中，所述方法还包括：

当接收到在所述文本数据中针对第一文本数据的增加操作时，基于所述第一文本数据和所述待编辑音视频中的音色信息，生成第一音视频片段；

基于所述第一文本数据在所述文本数据中的位置信息，确定所述第一文本数据对应的第一音视频时间戳；

基于所述第一音视频时间戳，将所述第一音视频片段添加到所述待编辑音视频中。

12、一种音视频处理装置，所述装置包括：

第一展示模块，用于展示待编辑音视频对应的文本数据；其中，所述文本数据与所述待编辑音视频的音视频时间戳具有映射关系；

第二展示模块，用于按照时间轴轨道展示所述待编辑音视频；

确定模块，用于响应于针对所述文本数据中的目标文本数据触发的预设操作，确定所述目标文本数据对应的音视频时间戳，作为目标音视频时间戳；

编辑模块，用于基于所述预设操作，对所述待编辑音视频中与所述目标音视频时间戳对应的音视频片段进行处理。

13、一种计算机可读存储介质，所述计算机可读存储介质中存储有指令，当所述指令在终端设备上运行时，使得所述终端设备实现如权利要求 1-11 任一项所述的方法。

14、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，实现如权利要求 1-11 任一项所述的方法。

15、一种计算机程序产品，所述计算机程序产品包括计算机程序/指令，所述计算机程序/指令被处理器执行时实现如权利要求 1-11 任一项所述的方法。