

字幕处理方法及装置

5 本申请要求于 2022 年 9 月 14 日递交的中国专利申请第 202211117721.1 号的优先权，在此全文引用上述中国专利申请公开的内容以作为本申请的一部分。

技术领域

本公开的实施例涉及一种字幕处理方法及装置。

10

背景技术

视频中的字幕能够起到辅助理解视频内容的作用，因此，在对视频进行编辑时常常会添加字幕。

15 目前，通常采用人工输入字幕文本或者使用字幕识别工具对相应的音频进行识别得到字幕文本，然后，再通过反复试听音频对字幕文本进行调整分割得到大量的文本片段，将这些文本片段与视频合成从而为视频添加字幕。针对字幕这类批量文本场景，若用户想要实现某种预期字幕效果，需要反复对字幕文本的分割结果进行调整、合成以及预览字幕效果，采用上述方式进行字幕编辑效率很低。

20

发明内容

为了解决上述技术问题，本公开提供了一种字幕处理方法及装置。

第一方面，本公开实施例提供了一种字幕处理方法，包括：

25 在对多媒体素材进行编辑的过程中，对所述多媒体素材对应的音频进行语音识别得到所述音频对应的字幕文本以及所述字幕文本包括的各文本元素分别对应音频片段的时间戳信息；

根据各所述文本元素分别对应音频片段的时间戳信息与所述多媒体素材中各素材单元进行匹配，确定所述多媒体素材中与各所述文本元素分别匹配的素材片段；其中，与所述文本元素相匹配的素材片段在编辑时间线上的
30 时间与所述文本元素对应的音频片段在编辑时间线上的时间一致；

将各所述文本元素分别与相匹配的时间范围内的素材片段合成得到有字幕文本逐字跳出动画效果的目标多媒体素材。

第二方面，本公开实施例提供了一种字幕处理装置，包括：

语音识别模块，用于在对多媒体素材进行编辑的过程中，对所述多媒体
5 素材对应的音频进行语音识别得到所述音频对应的字幕文本以及所述字幕
文本包括的各文本元素分别对应音频片段的时间戳信息；

匹配模块，用于根据各所述文本元素分别对应音频片段的时间戳信息与
所述多媒体素材中各素材单元进行匹配，确定所述多媒体素材中与各所述文
本元素分别匹配的素材片段；其中，与所述文本元素相匹配的素材片段在编
10 辑时间线上的时间与所述文本元素对应的音频片段在编辑时间线上的时间
一致；

字幕合成模块，用于将各所述文本元素分别与相匹配的时间范围内的素
材片段合成得到有字幕文本逐字跳出动画效果的目标多媒体素材。

第三方面，本公开实施例提供了一种电子设备，包括：存储器和处理器；
15 所述存储器被配置为存储计算机程序指令；所述处理器被配置为执行所述计
算机程序指令，使得所述电子设备实现第一方面所述的字幕处理方法。

第四方面，本公开实施例提供一种可读存储介质，包括：计算机程序指
令，电子设备的至少一个处理器执行所述计算机程序指令，使得所述电子设
备实现如第一方面所述的字幕处理方法。

20 第五方面，本公开实施例提供一种计算机程序产品，电子设备执行所述
计算机程序产品，使得所述电子设备实现第一方面所述的字幕处理方法。

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
25 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的技术方案，下面将对实施例所需要使
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 为本公开一实施例提供的字幕处理方法的流程图；

30 图 2 为本公开另一实施例提供的字幕处理方法的流程图；

图 3 为本公开另一实施例提供的字幕处理方法的流程图；
图 4A 至图 4I 为本公开提供的人机交互界面示意图；以及
图 5 为本公开一实施例提供的字幕处理装置的结构示意图。

5 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。

10 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。

字幕能够辅助用户理解视频内容，不同的字幕效果还能够表示更多维度的内容，例如，视频中的音频说到某个字时对应的文本才出现，这样的字幕效果在剧情演绎的垂类常常用来表示画外音，在 talking 垂类常常用来表示说话者的自信激昂的情绪。实现上述特定的字幕效果常通过用户手动输入字幕将字幕文本分割成一个个的字，之后再反复试听语音进行调整。此外，还可以通过用户输入一个完成的语句，使用关键帧蒙版的方式实现文本逐个出现的效果。不仅字幕编辑效率较低，且在移动端操作极其不便。

20 基于此，本公开实施例提供一种字幕处理方法及装置，其中，该方法包括：在对多媒体素材进行编辑的过程中，通过对多媒体素材对应的音频进行语音识别得到音频对应的字幕文本以及字幕文本中各文本元素对应的音频片段的时间戳信息；根据各文本元素对应的音频片段的时间戳信息，确定多媒体素材片段中与文本元素相匹配的素材片段；再将各文本元素分别与相匹配的时间内的素材片段合成得到有字幕文本逐字跳出动画效果的目标多媒体素材。本公开中，与文本元素相匹配的视频帧图像的时间范围的起始时刻与该文本元素对应的音频片段的起始时刻一致，能够实现说到某个词时相应文本字幕出现的字幕动画效果；此外，用户输入指令可实现动态字幕自动生成，用户操作简单，有利于提升用户体验，且本公开的方法能够适用于各种类型的设备，适用范围较广。

30 本公开提供的方法可以由电子设备执行，电子设备可以但不限于是平板

5 电脑、手机（如折叠屏手机、大屏手机等）、可穿戴设备、车载设备、增强现实（augmented reality, AR）/虚拟现实（virtual reality, VR）设备、笔记本电脑、超级移动个人计算机（ultra-mobile personal computer, UMPC）、上网本、个人数字助理（personal digital assistant, PDA）等，本公开对电子设备的具体类型不作任何限制。

图 1 为本公开一实施例提供的字幕处理方法的流程示意图。以电子设备执行本实施例的方法为例进行举例说明，电子设备中安装有编辑类应用程序，用户可以通过应用程序编辑多媒体素材。请参阅图 1 所示，本实施例的方法包括：

10 S101、在对多媒体素材进行编辑的过程中，对多媒体素材对应的音频进行语音识别得到音频对应的字幕文本以及字幕文本包括的各文本元素分别对应音频片段的时间戳信息。

多媒体素材可以为用户实时录制的视频素材，也可以为之前编辑过的视频素材，也可以为电子设备中存储的视频素材，本公开对此不做限定，也可以为音频素材、图像素材等等，本公开对于多媒体素材的类型不做限定。且
15 本公开对于多媒体素材的数量不做限定，若有多个多媒体素材，多个多媒体素材可以按照导入顺序进行排列，可以视作一个整体。

其中，对多媒体素材进行编辑的过程可以理解为提前录制或者导入带有音频的多媒体素材或者音频素材，或者，为多媒体素材（如视频素材、图像
20 素材）添加背景音乐。当然，编辑的方式不仅限于此。

字幕文本是通过当前正在编辑的多媒体素材对应的音频进行文本识别得到的，其中，多媒体素材对应的音频可以为多媒体素材包括的原始音频或者用户向多媒体素材添加的背景音乐，背景音乐可以是应用程序中的一个音频，如一首完整的歌曲，或者，一首歌曲的部分片段，或者，一个剪切后的
25 音频片段等等，本公开对此不做限定。当多媒体素材为音频素材时，可以对多媒体素材本身进行语音识别。

在一些实施例中，应用程序可以通过电子设备向中台服务发送音频，中台服务调用字幕识别工具对音频进行文本识别，得到相应的字幕文本以及字幕文本中各文本元素分别对应的音频片段的时间戳信息，时间戳信息可以包
30 括音频片段的起始时刻和结束时刻。

例如，多媒体素材片段对应的音频的总时长为 7 秒，对音频进行语音识别得到的字幕文本为：我今天很开心呀，总共 7 个文本元素，每个文本元素对应音频片段的时间为 1 秒，因此，各文本元素与相应的音频片段的时间戳信息的对应关系如下表 1 所示：

5

表 1

文本元素	音频片段的时间戳信息
我	00: 00—00: 01
今	00: 01—00: 02
天	00: 02—00: 03
很	00: 03—00: 04
开	00: 04—00: 05
心	00: 05—00: 06
呀	00: 06—00: 07

上述示例以音频所采用的语言种类为中文进行示例，相应的文本元素是以字为单位；若音频采用其他语言种类时，文本元素以相应的单词为单位，例如，音频采用英文时，文本元素以英文单词为单位。

10 在一些实施例中，应用程序可以响应用户输入的指令执行语音识别。本公开对于触发语音识别的指令的实现方式不做限定。在一些实施例中，语音识别的指令可以包括但不限于点击、双击、长按、滑动等类型的操作。例如，在应用程序的一个页面上设置有一个为多媒体素材添加识别字幕对应的区域/控件时，语音识别的指令可以为该区域/控件上接收到的操作。

15 S102、根据各文本元素分别对应音频片段的时间戳信息与多媒体素材中各素材单元进行匹配，确定多媒体素材中与各文本元素分别匹配的素材片段；其中，与文本元素相匹配的素材片段在编辑时间线上的时间与文本元素对应的音频片段在编辑时间线上的时间一致。

20 若多媒体素材为图像素材/视频素材时，多媒体素材中与文本元素相匹配的素材片段可以理解为图像片段/视频片段，素材片段中包括与文本元素进行合成的图像帧/视频帧。若多媒体素材为音频素材时，多媒体素材中与文本相匹配的素材片段可以理解为音频片段，素材片段中包括与文本元素进行合成的一个或多个语音发音单元。

由于本公开提供的字幕处理方法旨在实现说到某个词时相应文本字幕出现的字幕效果，因此，根据各文本元素分别对应的音频片段的时间戳信息确定相匹配的素材片段时，文本元素对应的音频片段在编辑时间线上的时间与素材片段在编辑时间线上的时间一致。

- 5 此处编辑时间线上一致可以理解为与文本元素相匹配的素材片段在编辑时间线上的起始时刻与文本元素对应的音频片段在编辑时间线上的起始时刻一致。

而字幕中文本元素消失的时间可以是灵活多变的，可以在其对应的音频片段结束时消失，也可以在其所属的语句（或者指定长度的文本片段）达到
10 结束位置时消失，或者，还可以在其对应的音频片段结束后经过预设时长消失，本公开不做限定。

因此，在编辑时间线上，与文本元素相匹配的素材片段对应时间的结束时刻可以等于该文本元素对应的音频片段的结束时刻，采用该方式，字幕文本具有逐字跳出动画效果，且前面出现的文本元素会随着其对应的音频片段
15 达到结束时刻而消失。

与文本元素相匹配的视频帧图像所属时间范围的结束时刻可以晚于该文本元素对应的音频片段的结束时刻，采用该方式，文本元素会逐个出现，且前面出现的文本元素会在其相应的音频片段结束后保留一段时间再消失。

其中，文本元素切换的速度取决于音频中发音对象的语速。

- 20 S103、将各文本元素分别与相匹配的时间范围内的素材片段合成得到有字幕文本逐字跳出动画效果的目标多媒体素材。

其中，将文本元素与相匹配的素材片段合成时，可以为文本元素使用预先设定的第一字幕动画样式，为多媒体素材自动添加的字幕在生成时便自动携带第一字幕动画样式对应的字幕效果，满足用户对于字幕效果的需求，减少用户
25 在后期操作。第一字幕动画样式可以包括文本元素的入场样式、出场样式以及循环样式中的一项或多项。

步骤 S102 和步骤 S103 可以通过调用动态字幕资源包（也可以称为字幕动画资源包）自动实现，将字幕文本以及字幕文本包括的各文本元素的时间戳信息传入动态字幕资源包，动态字幕资源包为各文本元素批量应用预先
30 设定的字幕动画样式，并将带有预设的字幕动画样式的文本元素叠加在相匹配

的素材片段中，从而为多媒体素材添加具有字幕文本采用第一字幕动画样式逐字跳出动画效果的字幕。

本实施例的方法，能够实现说到某个词时相应文本字幕出现的字幕动画效果；此外，用户输入指令可实现动态字幕自动生成，用户操作简单，有利于提升用户体验，且实施例的方法能够适用于各种类型的设备，适用范围较广，在批量文本场景中，对于屏幕较小的移动设备也能够快速地为多媒体素材片段添加指定效果的字幕。

通过图 1 所示实施例的方法为多媒体素材片段添加字幕后，用户还可以对字幕文本的内容进行再次编辑，再次编辑可以但不限于：删除文本元素、插入新增文本元素、替换文本元素。图 2 为本公开另一实施例提供的字幕处理方法的流程图。请参阅图 2 所示，本实施例的方法在图 1 所示实施例的基础上，还包括：

S104、响应文本删除指令，从所述字幕文本中删除相应的文本元素得到更新后的字幕文本。

其中，删除字幕文本中的文本元素，保留剩余的文本元素以及剩余文本元素的时间戳信息即可，得到更新后的字幕文本以及更新后的字幕文本中各文本元素的时间戳信息。

示例性地：假设，删除前字幕文本为：今（00:00-00:01）天（00:01-00:02）真（00:02-00:04）开（00:04-00:05）心（00:05-00:06）啊（00:06-00:07），括号内表示文本元素对应的音频片段的时间戳信息。

删除最后一个文本元素“啊”之后，更新后的字幕文本为：今（00:00-00:01）天（00:01-00:02）真（00:02-00:04）开（00:04-00:05）心（00:05-00:06），括号内表示文本元素对应的音频片段的时间戳信息。

若是删除其他位置的文本元素，采用类似的方式处理即可。

S105、响应文本插入指令，在所述字幕文本中插入新增文本元素得到更新后的字幕文本。

本步骤所执行的文本插入为不删除字幕文本中现有的文本元素的情况下插入新增文本元素。在一些实施例中，可根据新增文本元素的插入位置不同配置不同的处理方式。在一些实施例中，若插入新增文本元素的位置为字幕文本的中间或者文本末端，则将新增文本元素与相邻的前一个文本元素合

并，共用相邻的前一个文本元素对应的音频片段的时间戳；若插入新增文本元素的位置为字幕文本的最前端，新增文本元素与字幕文本的第一个文本元素合并，共用第一个文本元素对应的音频片段的时间戳。

示例性地，假设插入文本之前，字幕文本为：今（00:00-00:01）天（00:01-00:02）真（00:02-00:04）开（00:04-00:05）心（00:05-00:06）啊（00:06-00:07），
5 括号内表示文本元素对应的音频片段的时间戳信息。

情形一、在文本元素“真”的后面插入文本元素“的”之后，更新后的字幕文本为：今（00:00-00:01）天（00:01-00:02）真的（00:02-00:04）开（00:04-00:05）心（00:05-00:06）啊（00:06-00:07），括号内表示文本元素对应的音频片段的时间戳信息。
10

对比可知，执行插入新增文本元素之后，“真的”共用原先“真”所对应的音频片段的时间戳信息（00:02-00:04）。

情形二、在文本元素“今”的前面插入新增文本元素“哈哈”之后，更新后的字幕文本为：哈哈今（00:00-00:01）天（00:01-00:02）真（00:02-00:04）
15 开（00:04-00:05）心（00:05-00:06）啊（00:06-00:07），括号内表示文本元素对应的音频片段的时间戳信息。

对比可知，执行插入新增文本元素之后，“哈哈我”共用原先“我”所对应的音频片段的时间戳（00:00-00:01）。

S106、响应文本替换指令，采用替换文本替换所述字幕文本中的一个或多个文本元素得到更新后的字幕文本。
20

替换时，替换文本对应的时间戳信息等同于被替换文本元素所对应的音频片段的时间戳信息。在一次替换中，替换文本中可以包括一个或者多个文本元素，替换文本可以理解为一个整体，被替换文本元素的数量也可以为一个或多个位置连续的文本元素。

示例性地，假设插入文本之前，字幕文本为：今（00:00-00:01）天（00:01-00:02）真（00:02-00:04）开（00:04-00:05）心（00:05-00:06）啊（00:06-00:07），
25 括号内表示文本元素对应的音频片段的时间戳信息。

假设，采用替换文本“难过”替换“开心”，采用替换文本“呀”替换“啊”，替换之后得到的更新后的字幕文本为：今（00:00-00:01）天（00:01-00:02）真（00:02-00:04）难过（00:04-00:06）呀（00:06-00:07），括号内表
30

示文本元素对应的音频片段的时间戳信息。

对比可知，替换之后，“难过”采用原先“开心”分别对应的音频片段的时间戳之和（00:04-00:06）；“呀”采用原先“啊”对应的音频片段的时间戳（00:06-00:07）。

5 对字幕文本进行编辑可以根据需求选择上述一种或多种编辑方式。

S107、根据更新后的字幕文本中各文本元素分别对应音频片段的时间戳信息，确定所述多媒体素材中与各所述文本元素分别匹配的素材片段。

S108、将各所述文本元素分别与相匹配的时间内的素材片段合成，以为所述多媒体素材重新添加字幕。

10 步骤 S107 至步骤 S108 分别与前述图 1 所示实施例中步骤 S102 和步骤 S103 的实现方式类似，可参照前述图 1 所示实施例的详细描述。

若通过调用动态字幕资源包自动实现，将更新后的字幕文本以及更新后的字幕文本包括的各文本元素的时间戳信息重新传入动态字幕资源包，动态字幕资源包为更新后的字幕文本包括的各文本元素重新批量应用预先设定的字幕动画样式，并将带有预设的字幕动画样式的文本元素叠加在相匹配的素材片段中，从而为多媒体素材重新添加字幕。

15

本实施例的方法，能够满足用户在为多媒体素材添加字幕时对字幕内容进行调整的需求，针对更新后的字幕文本，能够自动生成带有指定字幕效果的字幕，方便用户使用，有利于提高用户体验。

20 通过图 1 所示实施例的方法为多媒体素材添加字幕后，用户还可以对字幕当前所采用的字幕动画样式进行调整，以得到符合用户预期的字幕效果。图 3 为本公开另一实施例提供的字幕处理方法的流程示意图。请参阅图 3 所示，本实施例的方法包括：

S301、在对多媒体素材进行编辑的过程中，对多媒体素材对应的音频进行语音识别得到音频对应的字幕文本以及字幕文本包括的各文本元素分别对应音频片段的时间戳信息。

25

S302、根据各文本元素分别对应音频片段的时间戳信息与多媒体素材中各素材单元进行匹配，确定多媒体素材中与各文本元素分别匹配的素材片段；其中，与文本元素相匹配的素材片段在编辑时间线上的时间与文本元素对应的音频片段在编辑时间线上的时间一致。

30

S303、为各文本元素批量应用指定的第一字幕动画样式，将具有第一字幕动画样式的文本元素与相匹配的时间内的素材片段合成得到有字幕文本采用所述第一字幕动画样式逐字跳出动画效果的目标多媒体素材。

本实施例中步骤 S301 至步骤 S302 分别与图 1 所示实施例中步骤 S101 至步骤 S103 类似，可参照图 1 所示实施例的详细描述，此处不再赘述。需要说明的是，第一字幕动画样式可以理解为应用程序默认的字幕动画样式。

S304、响应字幕动画样式切换指令，为各文本元素批量应用所述第二字幕动画样式，将具有第二字幕动画样式的文本元素与相匹配的时间内的素材片段合成得到有字幕文本采用第二字幕动画样式逐字跳出动画效果的目标多媒体素材。

其中，应用程序可以通过电子设备向用户提供的字幕动画样式编辑的页面，页面中可以展示一个或多个可供用户选择的字幕动画样式对应的区域或者控件，用户可以通过操作字幕动画样式对应的区域或者控件输入字幕动画样式切换指令。

若通过调用动态字幕资源包自动实现，将更新后的字幕文本以及更新后的字幕文本包括的各文本元素的时间戳信息重新传入动态字幕资源包，动态字幕资源包为更新后的字幕文本包括的各文本元素重新批量应用用户指定的第二字幕动画样式，并将带有第二字幕动画样式的文本元素叠加在相匹配的素材片段中，从而为多媒体素材重新添加字幕。

本实施例的方法能够满足用户对后期调整字幕效果的需求，且支持字幕动画样式批量编辑，字幕处理效率较高。

基于前述描述，本公开以实施例将以电子设备为例，结合附图和应用场景，对本公开提供的字幕处理方法进行举例说明。为了便于说明，图 4A-图 4K 中，以电子设备为手机，手机中安装有视频编辑类应用程序（简称应用 1）为例进行示意。此外，用户向应用 1 中导入的多媒体素材为视频素材。

请参阅图 4A-图 4I，图 4A-图 4I 为本公开实施例提供的人机交互界面示意图。

应用 1 可以在手机上示例性地显示如图 4A 所示的用户界面 11，用户界面 11 用于显示多媒体素材编辑页面（以下简称为编辑页面），应用 1 在编辑页面中执行某个功能集合，如预览播放多媒体素材的剪辑效果、为多媒体素

材添加背景音乐、为多媒体素材添加滤镜、贴纸、文字等等。

参照图 4A 所示，用户界面 11 包括区域 a1，为多媒体素材的剪辑效果预览区域；用户界面 11 还包括区域 a2，在区域 a2 中可以按照时间轴展示多媒体素材以及在编辑过程中添加的其他剪辑素材。用户界面 11 还包括区域 a3，
5 区域 a3 中可以向用户提供多种剪辑功能入口。例如，区域 a3 中包括控件 101，控件 101 用于进入到应用 1 的文本功能集合页，该文本功能集合页中包括多个控件，多个控件分别对应不同的文本功能。

示例性地，在应用 1 接收到用户在图 4A 所示的用户界面 11 中执行如点击控件 101 的操作后，应用 1 可以在手机上示例性地显示如图 4B 所示的用户界面 12，用户界面 12 中展示了应用 1 提供的文本功能集合页，文本功能
10 集合页中可以向用户提供各种不同文本功能的入口，用户通过入口进入相应的文本功能操作页面向多媒体素材中添加文本内容。

用户界面 12 包括：区域 a4，区域 a4 中包括：新建文本功能、文字模板功能、字幕识别功能、歌词识别功能、贴纸功能以及涂鸦笔功能分别对应的
15 入口。其中，用户界面 12 中所示的控件 102 即为字幕识别功能对应的入口。

应用 1 接收到用户在图 4B 所示的用户界面 12 中执行如点击控件 102 的操作后，应用 1 可以在手机上示例性地显示如图 4C 所示的用户界面 13，用户界面 13 用于展示了应用 1 提供的字幕识别面板，字幕识别面板中可以向用户提供识别类型的选项、语言种类选择入口、标记无效片段的开关、动态
20 字幕的开关以及同时清空已有字幕的开关。

其中，动态字幕表示为多媒体素材添加具有字幕文本逐字跳出动画效果的字幕的功能。具体地，当动态字幕的开关为关闭状态时，添加的字幕所呈现的字幕效果是以语句片段为单条字幕的形式出现；动态字幕的开关为打开状态时，添加的字幕所呈现的字幕效果是字幕文本逐字跳出的效果，即字幕
25 文本中的文本元素逐个出现且文本元素在相应的音频片段开始时显示。

在一些实施例，可以记忆用户的选择，打开字幕识别面板时显示用户上次退出字幕识别时动态字幕的开关状态，更加符合用户的使用习惯。在应用 1 首次更新了动态字幕功能时，动态字幕的开关可以为关闭状态，即如用户界面 13 所示。

30 应用 1 接收到用户在如图 4C 所示的用户界面 13 中执行如点击动态字幕

的开关按钮的操作后，显示如图 4D 所示的用户界面 14，用户界面 14 中，动态字幕的开关状态为打开状态。

用户界面 14 中还包括控件 103，控件 103 用于指示开始语音识别并添加具有逐字跳出动画效果的字幕。应用 1 响应于用户在用户界面 14 中执行如
5 点击控件 103 的操作后，应用 1 在手机上示例性显示如图 4E 所示的用户界面 15，用户界面 15 中字幕识别面板关闭，且在区域 a4 中显示提示内容，例如动画和提示文字，以提示用户当前正在创建动态字幕动画。为了减小提示动画以及提示文字对区域 a1 中展示的预览画面的遮挡，区域 a4 可以位于区域 a1 的上方，应理解区域 a4 也可以位于其他位置，本公开对此不做限定。

10 其中，结合前文所述，用户对动态字幕的开关以及对控件 103 的操作触发应用 1 对多媒体素材对应的音频进行语音识别并自动添加具有逐字跳出动画效果的动态字幕。

当动态字幕动画创建完成后，应用 1 可以在手机上示例性地显示如图 4F 所示的用户界面 16，用户界面 16 中，区域 a4 中可以显示提示内容，例如提示
15 提示文字“识别成功，已自动生成字幕”。

之后，用户可以点击预览播放按钮，在区域 a1 中预览字幕效果，如果符合用户预期，则可以将编辑好的多媒体素材导出为目标视频进行发布或者保存。

结合图 4A 至图 4F 所示的交互过程，本公开在字幕识别的前置阶段为用户
20 提供了动态字幕开关，更便于用户使用。且记忆用户上一次退出字幕识别面板时，动态字幕开关的开关状态，用户再次使用时无需进行操作，从而再使用上再次提效，用户也无需执行过多的操作。

为了更好地满足用户需求，应用 1 还为用户提供了在后置的环节添加动态字幕或者修改已有字幕动画样式的功能。

25 示例性地，在图 4F 所示的用户界面 16 的基础上，区域 a2 中按照时间轴展示了多媒体素材以及字幕文本分别对应的标识，通过对区域 a2 中所展示的字幕文本的标识进行操作（如点击），可以触发对字幕进行再次编辑。应用 1 接收到用户对用户界面 16 的区域 a2 中任一字幕包含的文本片段的点击操作，应用 1 可以在手机上示例性地显示如图 4G 所示的用户界面 17。

30 用户界面 17 中，区域 a1 中显示字幕文本对应的文本框 104，文本框 104

中包含当前预览位置对应的文本内容，可以是一个或者多个语句（即文本片段），且区域 a1 中还可以显示针对文本框进行操作的控件，例如，旋转、复制、用户还可以通过双指触发放大或者缩小文本框的尺寸，文本框中文本元素的尺寸也会随着文本框的大小变化而变化。用户界面 17 中还包括区域 a5，

5 区域 a5 用于显示字幕编辑功能集合页的区域，字幕编辑功能集合页提供了对当前添加的字幕进行编辑的多种编辑功能的入口，例如，可以包括：批量编辑字幕、字幕分割、复制字幕、编辑字幕、删除字幕、花字以及字幕动画样式等等功能分别对应的入口。其中，用户界面 17 中包含控件 105，控件 105 用于进入字幕动画面板，以为当前字幕添加字幕效果（其中包含动态字幕效果）

10 或者对当前字幕所采用的字幕动画样式进行修改。

应用 1 接收用户在用户界面 17 中执行如点击控件 105 的操作后，显示如图 4H 所示的用户界面 18，用户界面 18 包括区域 a6。

其中，区域 a6 用户展示字幕动画面板，字幕动画面板包括设置动画样式的标签 106 以及字体标签、样式标签、花字标签、文字模板标签等。一些实施例中，可以如图 4H 所示，进入字幕动画样式面板可以默认定位至标签 106，并显示标签 106 的相关内容。另一些实施例中，可以定位至其他标签，应用 1 在接收到用户对标签 106 的点击操作之后，显示标签 106 的相关内容。

15

参照图 4H 所示，区域 a6 中还包括：动态字幕开关 107，通过操作动态字幕开关 107 可以为当前字幕添加具有文本元素逐个显示的字幕效果。

20 在一些实施例中，若在前置阶段，用户已添加过动态字幕，则此处可以显示为开启状态；若在前置阶段，用户未使用动态字幕，则此处可以显示为关闭状态，用户可以切换用户界面 18 中显示的动态字幕开关 107 的开关状态为打开状态。其中，图 4H 所示实施例中，动态字幕开关 107 为关闭状态。

此外，区域 a6 中还包括：设置字幕入场样式的标签 108、设置字幕出场样式的标签 109、设置字幕循环样式的标签 110、设置动态字幕动画样式的标签 111 以及区域 a7，其中，区域 a7 用于根据当前定位的标签，显示相应标签的内容。一些情况下，动态字幕开关 107 为关闭状态时，可以默认显示任一标签相关的内容，例如图 4H 所示用户界面 17，默认显示标签 108 对应的

25 相关内容。

30 其中，当应用 1 接收到用户针对用户界面 18 中动态字幕开关 107 执行

的操作（如点击操作），动态字幕开关 107 由关闭状态切换为打开状态时，应用 1 可在手机上示例性地显示如图 4I 所示的用户界面 19，参照图 4I 所示，用户界面 19 中，动态字幕开关 107 为打开状态，且标签 111 为选中的状态，在区域 a7 用于显示与动态字幕相关的一种或多种可供用户选择的动态字幕动画样式，多种动态字幕动画样式对应的显示标识可以按照由左向右的方式依次排列，用户通过左右滑动屏幕可以来回进行查看。其中，应用 1 中默认的动态字幕动画样式可以显示由左向右的第一个位置，以使用户清楚了解应用 1 默认使用的动态字幕动画样式是哪个。

其中，区域 a7 中还可以包括禁用按钮 112，禁用按钮可以设置在区域 a7 的最左侧，当然也可以设置在其他位置，本公开对此不做限定。当用户点击禁用按钮 112 对应关闭动态字幕效果，动态字幕开关 107 会切换为关闭状态。

假设用户点击了区域 a7 中由左向右的第二个动态字幕动画样式，相当于向应用 1 输入字幕动画样式切换指令，应用 1 响应字幕动画样式切换指令，为字幕文本中包括的各文本元素应用第二个动态字幕样式。用户可以多次切换动态字幕动画样式，直至得到符合用户预期的字幕效果。

在图 4H 所示的用户界面 18 以及图 4I 所示的用户界面 19 的基础上，区域 a5 中还包括：区域 a8，区域 a8 用于展示文本编辑框，用户可以通过文本编辑框删除字幕文本中的文本元素、插入新增文本或者替换原先的文本元素，用户针对文本编辑框的操作相当于向应用 1 输入了删除指令、插入指令以及替换指令。在对区域 a8 中文本编辑框中的文本内容进行编辑时，编辑后的文本内容同步地展示在区域 a1 中所示的文本框 104 中，有利于用户预览编辑后的字幕内容以及字幕内容在多媒体素材片段的视频帧图像中的显示效果。

通过如上所示的图 4F 至图 4I 所示实施例，在后置阶段通过在字幕动画样式面板中设置动态字幕开关以及动态字幕动画样式标签，满足用户在后置环节添加动态字幕的需求以及调整动态字幕所采用的字幕动态样式的需求。

需要说明的是，上述图 4A 至图 4I 所示的交互界面示意图并不是对本公开提供的字幕处理方法的限制，应理解，一些控件、面板、标签的样式、触发方式等等均可以根据需求灵活调整。

图 5 为本公开一实施例提供的字幕处理装置的结构示意图。请参阅图 5 所示，本实施例提供的装置 500 包括：

语音识别模块 501，用于在对多媒体素材进行编辑的过程中，对所述多媒体素材对应的音频进行语音识别得到所述音频对应的字幕文本以及所述字幕文本包括的各文本元素分别对应音频片段的时间戳信息。

5 匹配模块 502，用于根据各所述文本元素分别对应音频片段的时间戳信息与所述多媒体素材中各素材单元进行匹配，确定所述多媒体素材中与各所述文本元素分别匹配的素材片段；其中，与所述文本元素相匹配的素材片段在编辑时间线上的时间与所述文本元素对应的音频片段在编辑时间线上的时间一致。

10 字幕合成模块 503，用于将各所述文本元素分别与相匹配的时间范围内的素材片段合成得到有字幕文本逐字跳出动画效果的目标多媒体素材。

在一些实施例中，在所述编辑时间线上，与所述文本元素相匹配的素材片段所属时间的起始时刻与所述文本元素对应的音频片段的起始时刻一致；且在所述编辑时间线上，与所述文本元素相匹配的素材片段所属时间的结束时刻与所述文本元素对应的音频片段的结束时刻一致，或者，与所述文本元素相匹配的素材片段所属时间的结束时刻晚于所述文本元素对应的音频片段的结束时刻。

20 在一些实施例中，字幕合成模块 503，具体用于为各所述文本元素批量应用指定的第一字幕动画样式，将具有第一字幕动画样式的文本元素与相匹配的时间内的素材片段合成得到有字幕文本采用所述第一字幕动画样式逐字跳出的动画效果的目标多媒体素材。

可选地，装置 500 还包括：字幕文本更新模块 504。

在一些实施例中，字幕文本更新模块 504，用于响应文本删除指令，从所述字幕文本中删除相应的文本元素得到更新后的字幕文本。

25 相应地，匹配模块 502，还用于根据所述更新后的字幕文本中各文本元素分别对应音频片段的时间戳信息，确定所述多媒体素材中与各所述文本元素分别匹配的素材片段。

字幕合成模块 503，还用于将所述更新后的字幕文本包括的各所述文本元素分别与相匹配的时间内的素材片段合成，以为所述多媒体素材重新添加具有字幕文本逐字跳出动画效果的字幕。

30 在一些实施例中，字幕文本更新模块 504，还用于响应文本插入指令，

在所述字幕文本中插入新增文本元素得到更新后的字幕文本。

- 相应地，匹配模块 502，还用于根据所述更新后的字幕文本中各文本元素分别对应音频片段的时间戳信息与所述多媒体素材中各素材单元进行匹配，确定所述多媒体素材中与各所述文本元素分别匹配的素材片段；其中，
- 5 新增文本元素与相邻的文本元素合并共用所述相邻的文本元素对应的音频片段的时间戳信息。

字幕合成模块 503，还用于将所述更新后的字幕文本包括的各所述文本元素分别与相匹配的时间范围内的素材片段进行结合，以为所述多媒体素材重新添加具有字幕文本逐字跳出动画效果的字幕。

- 10 在一些实施例中，若所述新增文本元素的插入位置为所述字幕文本的最前端，则将所述新增文本元素与所述字幕文本中的第一个文本元素合并，共用所述第一个文本元素对应的音频片段的时间戳；若所述新增文本元素的插入位置为所述字幕文本的中间或者最末端位置，所述新增文本元素与相邻的前一个文本元素合并，共用前一个文本元素对应的音频片段的时间戳。

- 15 在一些实施例中，字幕文本更新模块 504，还用于响应文本替换指令，采用替换文本替换所述字幕文本中的一个或多个文本元素得到更新后的字幕文本。

- 相应地，匹配模块 502，还用于根据所述更新后的字幕文本中各文本元素分别对应音频片段的时间戳信息，确定所述多媒体素材中与各所述文本元素分别匹配的素材片段；其中，所述替换文本对应被替换文本元素所对应的
- 20 音频片段的时间戳信息。

字幕合成模块 503，还用于将所述更新后的字幕文本包括的各所述文本元素分别与相匹配的时间内的素材片段合成，以为所述多媒体素材重新添加具有字幕文本逐字跳出动画效果的字幕。

- 25 在一些实施例中，字幕合成模块 503，还用于响应字幕动画样式切换指令，为各所述文本元素批量应用第二字幕动画样式，将具有所述第二字幕动画样式的文本元素与相匹配的时间范围内的素材片段合成得到有字幕文本采用第二字幕动画样式逐字跳出动画效果的目标多媒体素材。

- 在一些实施例中，所述多媒体素材对应的音频为所述多媒体素材包括的
- 30 原始音频或者为所述多媒体素材添加的背景音乐。

本实施例提供的字幕处理装置可以用于执行前述任一方法实施例的技术方案，其实现原理以及技术效果类似，可参照前述方法实施例的详细描述，简明起见，此处不再赘述。

5 示例性地，本公开提供一种电子设备，包括：一个或多个处理器；存储器；以及一个或多个计算机程序；其中一个或多个计算机程序被存储在存储器中；一个或多个处理器在执行一个或多个计算机程序时，使得电子设备实现前文实施例的字幕处理方法。

10 示例性地，本公开提供一种芯片系统，芯片系统应用于包括存储器和传感器的电子设备；芯片系统包括：处理器；当处理器执行前文实施例的字幕处理方法。

示例性地，本公开提供一种计算机可读存储介质，其上存储有计算机程序，计算机程序被处理器使得电子设备执行时实现前文实施例的字幕处理方法。

15 示例性地，本公开提供一种计算机程序产品，当计算机程序产品在计算机上运行时，使得计算机执行前文实施例的字幕处理方法。

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
20 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。

25 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、一种字幕处理方法，包括：

在对多媒体素材进行编辑的过程中，对所述多媒体素材对应的音频进行
5 语音识别得到所述音频对应的字幕文本以及所述字幕文本包括的各文本元素
分别对应音频片段的时间戳信息；

根据各所述文本元素分别对应音频片段的时间戳信息与所述多媒体素材
中各素材单元进行匹配，确定所述多媒体素材中与各所述文本元素分别匹
配的素材片段；其中，与所述文本元素相匹配的素材片段在编辑时间线上的
10 时间与所述文本元素对应的音频片段在编辑时间线上的时间一致；

将各所述文本元素分别与相匹配的时间内的素材片段合成得到有字幕
文本逐字跳出动画效果的目标多媒体素材。

2、根据权利要求1所述的方法，其中，在所述编辑时间线上，与所述文
本元素相匹配的素材片段所属时间的起始时刻与所述文本元素对应的音频
15 片段的起始时刻一致；且在所述编辑时间线上，与所述文本元素相匹配的素
材片段所属时间的结束时刻与所述文本元素对应的音频片段的结束时刻一
致，或者，与所述文本元素相匹配的素材片段所属时间的结束时刻晚于所述
文本元素对应的音频片段的结束时刻。

3、根据权利要求1或2所述的方法，其中，所述将各所述文本元素分别
20 与相匹配的时间内的素材片段合成得到有字幕文本逐字跳出动画效果的目
标多媒体素材，包括：

为各所述文本元素批量应用指定的第一字幕动画样式，将具有所述第一
字幕动画样式的文本元素与相匹配的时间内的素材片段合成得到有字幕文
本采用所述第一字幕动画样式逐字跳出动画效果的所述目标多媒体素材。

25 4、根据权利要求1-3任一项所述的方法，还包括：

响应文本删除指令，从所述字幕文本中删除相应的文本元素得到更新后
的字幕文本；

根据所述更新后的字幕文本中各文本元素分别对应音频片段的时间戳
信息，确定所述多媒体素材中与各所述文本元素分别匹配的素材片段；

30 将所述更新后的字幕文本包括的各所述文本元素分别与相匹配的时间

内的素材片段合成，以为所述多媒体素材重新添加具有字幕文本逐字跳出动画效果的字幕。

5、根据权利要求 1-4 任一项所述的方法，还包括：

响应文本插入指令，在所述字幕文本中插入新增文本元素得到更新后的
5 字幕文本；

根据所述更新后的字幕文本中各文本元素分别对应音频片段的时间戳信息与所述多媒体素材中各素材单元进行匹配，确定所述多媒体素材中与各所述文本元素分别匹配的素材片段；其中，新增文本元素与相邻的文本元素合并共用所述相邻的文本元素对应的音频片段的时间戳信息；

10 将所述更新后的字幕文本包括的各所述文本元素分别与相匹配的时间范围内的素材片段进行结合，以为所述多媒体素材重新添加具有字幕文本逐字跳出动画效果的字幕。

6、根据权利要求 5 所述的方法，其中，若所述新增文本元素的插入位置为所述字幕文本的最前端，则将所述新增文本元素与所述字幕文本中的第一个文本元素合并，共用所述第一个文本元素对应的音频片段的时间戳；
15

若所述新增文本元素的插入位置为所述字幕文本的中间或者最末端位置，所述新增文本元素与相邻的前一个文本元素合并，共用前一个文本元素对应的音频片段的时间戳。

7、根据权利要求 1-6 任一项所述的方法，还包括：

20 响应文本替换指令，采用替换文本替换所述字幕文本中的一个或多个文本元素得到更新后的字幕文本；

根据所述更新后的字幕文本中各文本元素分别对应音频片段的时间戳信息，确定所述多媒体素材中与各所述文本元素分别匹配的素材片段；其中，所述替换文本对应被替换文本元素所对应的音频片段的时间戳信息；

25 将所述更新后的字幕文本包括的各所述文本元素分别与相匹配的时间内的素材片段合成，以为所述多媒体素材重新添加具有字幕文本逐字跳出动画效果的字幕。

8、根据权利要求 3 所述的方法，还包括：

30 响应字幕动画样式切换指令，为各所述文本元素批量应用第二字幕动画样式，将具有所述第二字幕动画样式的文本元素与相匹配的时间范围内的素

材片段合成得到有字幕文本采用第二字幕动画样式逐字跳出动画效果的目标多媒体素材。

- 9、根据权利要求1至8任一项所述的方法，其中，所述多媒体素材对应的音频为所述多媒体素材包括的原始音频或者为所述多媒体素材添加的背景音乐。

10、一种字幕处理装置，包括：

语音识别模块，用于在对多媒体素材进行编辑的过程中，对所述多媒体素材对应的音频进行语音识别得到所述音频对应的字幕文本以及所述字幕文本包括的各文本元素分别对应音频片段的时间戳信息；

- 10 匹配模块，用于根据各所述文本元素分别对应音频片段的时间戳信息与所述多媒体素材中各素材单元进行匹配，确定所述多媒体素材中与各所述文本元素分别匹配的素材片段；其中，与所述文本元素相匹配的素材片段在编辑时间线上的时间与所述文本元素对应的音频片段在编辑时间线上的时间一致；

- 15 字幕合成模块，用于将各所述文本元素分别与相匹配的时间范围内的素材片段合成得到有字幕文本逐字跳出动画效果的目标多媒体素材。

11、一种电子设备，包括：存储器和处理器，其中，

所述存储器被配置为存储计算机程序指令；

- 20 所述处理器被配置为执行所述计算机程序指令，使得所述电子设备实现如权利要求1至9任一项所述的字幕处理方法。

12、一种可读存储介质，包括：计算机程序指令，其中，

电子设备的至少一个处理器执行所述计算机程序指令，使得所述电子设备实现如权利要求1至9任一项所述的字幕处理方法。

- 25 13、一种计算机程序产品，其中，电子设备执行所述计算机程序产品，使得所述电子设备实现如权利要求1至9任一项所述的字幕处理方法。