用于确定音频的卡点的方法、装置、设备和产品

本申请要求于 2024 年 06 月 14 日提交的，申请号为 202410773416.0、发明名称为“用于确定音频的卡点的方法、装置、设备和产品”的中国专利申请的优先权，该申请的全部内容通过引用结合在本申请中。

技术领域

本公开总体涉及计算机领域，并且更具体地，涉及用于确定音频的卡点的方法、装置、电子设备和程序产品。

背景技术

视频剪辑是对一个视频源进行非线性编辑的过程，这涉及将图片、背景音乐、特效、场景等素材与原始视频进行重混合。在这个过程中，背景音乐的音乐卡点在视频剪辑中发挥着重要作用，它能够使得视频作品更具观赏性。

音乐卡点是指在制作视频时，选择一段特定的音乐节奏，使得视频的画面与这段音乐的节奏相匹配，以达到增强画面感染力和观看体验的效果。例如，在旅行游记、视频生活日志、影视剧解说等类型的视频中，通过音乐卡点可以迅速将观众带入剧情，增强观看体验。

发明内容

本公开的实施例提供了一种用于确定音频的卡点的方法、装置、电子设备和程序产品。

根据本公开的第一方面，提供了一种用于确定音频的卡点的方法。该方法包括获取用于制作视频的音频。该方法包括确定音频的多个候选卡点以及多个候选卡点的多个能量。此外，该方法还包括基于多个候选卡点以及多个能量，确定音频的多个目标卡点，其中多个目标卡点用于匹配视频的画面。

在本公开的第二方面中，提供了一种的用于确定音频的卡点的装置。该装置包括音频获取模块，被配置为获取用于制作视频的音频。该装置还包括候选卡点确定模块，被配置为确定音频的多个候选卡点以及多个候选卡点的多个能量。此外，该装置还包括目标卡点确定模块，被配置为基于多个候选卡点以及多个能量，确定音频的多个目标卡点，其中多个目标卡点用于匹配视频的画面。

在本公开的第三方面中，提供了一种电子设备。该电子设备包括处理器以及与处理器耦合的存储器，存储器具有存储于其中的指令，指令在被处理器执行时，使得

电子设备执行根据第一方面的方法。

在本公开的第四个方面，提供了一种计算机程序产品，其上存储有包括计算机可执行指令，其中所述计算机可执行指令被处理器执行以实现第一方面的方法。

发明内容部分是为了以简化的形式来介绍对概念的选择，它们在下文的具体实施方式中将被进一步描述。发明内容部分无意标识要求保护的主题的关键特征或主要特征，也无意限制要求保护的主题的范围。

附图说明

结合附图并参考以下详细说明，本公开各实施例的上述和其它特征、优点及方面将变得更加明显。在附图中，相同或相似的附图标记表示相同或相似的元素，其中：

图 1 示出了本公开的一些实施例可以在其中实现的示例环境的示意图；

图 2 示出了本公开的一些实施例的用于确定音频的卡点的方法的流程图；

图 3 示出了本公开的一些实施例的音频的卡点的示意图；

图 4 示出了本公开的一些实施例的用于在多个候选卡点中确定音频的目标卡点的示意图；

图 5 示出了本公开的一些实施例的通过滑动窗口来确定音频的卡点的示意图；

图 6A 示出了本公开的一些实施例的用于导入待剪辑的视频的示意图；

图 6B 示出了本公开的一些实施例的用于导入音频的示意图；

图 7 示出了本公开的一些实施例的用于确定音频的卡点的装置的框图；以及

图 8 示出了本公开的一些实施例的电子设备的框图。

在所有附图中，相同或相似参考数字表示相同或相似元素。

具体实施方式

可以理解的是，本技术方案所涉及的数据（包括但不限于数据本身、数据的获取或使用）应当遵循相应法律法规及相关规定的要求。

可以理解的是，在使用本公开各实施例公开的技术方案之前，均应当依据相关法律法规通过恰当的方式对本公开所涉及个人信息的类型、使用范围、使用场景等告知用户并获得用户的授权。

例如，在接收到用户的主动请求时，向用户发送提示信息，以明确地提示用户，其请求执行的操作将需要获取和使用到用户的个人信息。从而，使得用户可以根据提示信息来自主地选择是否向执行本公开技术方案的操作的电子设备、应用程序、服务器或存储介质等软件或硬件提供个人信息。

作为一种可选的但非限定性的实现方式，响应于接收到用户的主动请求，向用户发送提示信息的方式例如可以是弹窗的方式，弹窗中可以以文字的方式呈现提示信

息。此外，弹窗中还可以承载供用户选择“同意”或者“不同意”向电子设备提供个人信息的选择控件。

可以理解的是，上述通知和获取用户授权过程仅是示意性的，不对本公开的实现方式构成限定，其它满足相关法律法规的方式也可应用于本公开的实现方式中。

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一些实施例，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。

在本公开的实施例的描述中，术语“包括”及其类似用语应当理解为开放性包括，即“包括但不限于”。术语“基于”应当理解为“至少部分地基于”。术语“一个实施例”或“该实施例”应当理解为“至少一个实施例”。术语“第一”、“第二”等等可以指代不同的或相同的对象，除非明确说明。下文还可能包括其它明确的和隐含的定义。

如前所述，将视频的画面节奏与音乐的节奏相匹配在视频剪辑中是一种重要的技巧。在相关技术中，通常是将音乐的节拍点（beat）作为音乐的节奏点，这种方法不能满足用户对视频卡点音乐的需求。因为在音乐的每个小节（measure）没有发生速度（tempo）变化的情况下，节拍点作为音乐中的基本时间单位，其在音乐的时间维度上是均匀分布的。此时，仅将节拍点作为音乐的节奏点，并不能契合视频的创作者的卡点需求。此外，仅使用音乐的节拍点（beat）作为视频的节奏点，还可能过滤掉一个节拍之内的明显的节奏变化的点，比如一个节拍之内的连续的几个重低音。

本公开的实施例中，通过检测音频中的多个更大范围的潜在的节奏点（候选卡点）以及这些候选卡点的能量，能够精确地确定出音频中关键的节奏点（目标卡点），这些目标卡点可以被用于与视频画面的精准匹配。这种根据音频的能量分析来确定音频的节奏的方法，能够确定出包括但不限于节拍点在内的更多的在音频中具有影响力的多个候选卡点，进而再根据这些候选卡点的能量确定在音频中具有影响力和重要性的目标卡点，不仅能够确保关键卡点的有效保留，还能够满足用户对根据卡点来应用视频转场的需求，提升了用户体验。

图 1 示出了本公开的一些实施例可以在其中实现的示例环境 100 的示意图。如图 1 所示，显示界面 102 是为待编辑的视频 104 编辑音频 110 的卡点的界面。在显示界面 102 上，显示有待编辑的视频 104，通过播放工具栏 106，用户可以对待编辑的视频 104 暂停，前进或者后退，还可以通过全屏播放按钮来实现待编辑视频 104 的全屏展示，播放工具栏 106 上还显示有待编辑视频 104 的总时间长度以及当前的定位时间戳。

继续参考图 1，音频 110 是用户选择的为待编辑的视频 104 进行配乐的音频。为了使得待编辑视频 104 的节奏与音频 110 的节奏相匹配，可以为用户呈现其使用的音频 110 的卡点，这样可以方便用户在编辑视频时，能够在音乐发生明显变化的时候应用转场，使得音视频和谐一致。例如，当用户打开自动卡点 108 的按钮，可以在音频 110 的下方显示当前音频 110 的卡点 112，这些卡点 112 不局限于音频 110 的每个小节的均匀分布的节拍，还可以包括有音频 110 的鼓点或者是低音乐器(base)如吉他开始时的点位。

图 1 所示的卡点 112 中的点位 112-1 可以是一节拍之内的连续 3 个鼓点。在音乐中，小节是一段固定时值的节奏单位，通常以小节线分隔。每个小节通常包含一个或多个节拍。节拍点是音乐中的基本时间单位，节拍点在乐曲中是均匀分布的，除非有特定的速度变化标记（如渐快、渐慢等）。

参考图 1，为了实现友好的用户交互和提升用户的体验，除了为用户自动识别卡点外，用户也可以通过添加卡点 114 按钮来实现手动对已经生成的卡点 112 的编辑，例如可以自行添加卡点或者删除已生成的至少包括节拍点在内的多个候选卡点。当用户操作后，可以通过工具栏 116 来实现对音视频的保存或者取消。通过这种为用户选定的音频确定大量的多样化的不局限于节拍点的多个候选卡点的方法，能够满足用户希望在音乐发生明显变化时来同步实现画面变化的需求，提升了用户体验感。

下面将结合图 2 至图 8 详细描述根据本公开实施例的过程。为了便于理解，在下文描述中提及的具体数据均是示例性的，并不用于限定本公开的保护范围。可以理解，以下描述的实施例还可以包括未示出的附加动作和/或可以省略所示出的动作，本公开的范围在此方面不受限制。

图 2 示出了本公开的一些实施例的用于确定音频的卡点的方法 200 的流程图。方法 200 可以由用于确定音频的卡点的装置执行。该装置例如可以是服务器，也可以是配置于服务器的系统，也可以是独立的装置或系统。该装置可以通过软件和/或硬件的方式实现。接下来以用于确定音频的卡点的装置为执行主体为例，对方法 200 进行示意性地说明。参考图 2，方法 200 可以包括框 202、框 204 以及框 206。

在框 202，获取用于制作视频的音频。在一些实施例中，这些视频可以是短视频也可以是长视频，这些视频可以从云端获取或者从本地获取。用于制作视频的音频可以是纯音乐音频或者是包含人声的音频等，这些音频可以从本地获取或者从云端获取。

在框 204，确定音频的多个候选卡点以及多个候选卡点的多个能量。这些候选点可以通过起始点检测（onset detection）的方法而得到，这些候选卡点是多种形式的卡点，比如可以是音频的节拍点，还可以是音频的某个乐器如鼓的鼓点，又或者是某个低音乐器如吉他的开始点位。例如还可以是一个节拍之内有好几个连续的重低音的鼓

点。每个点位都有其各自的能量，在一些实施例中，这些能量代表音频中音符和/或变化的音乐事件的强度。

在框 206，基于多个候选卡点以及多个能量，确定音频的多个目标卡点，其中多个目标卡点用于匹配视频的画面。例如，当确定了多个候选卡点后，考虑到确定的候选卡点的数量可能过于密集从而影响了用户卡点的转场效果，可以在根据多个候选卡点的能量来从多个候选卡点中确定出适用于用户进行视频画面踩点转场的目标卡点，这样用户即可以根据确定出来的目标卡点来实现视频画面与音频节奏点的同步变化。

本公开的实施例中，检测音频的多个候选卡点以及多个候选卡点的能量，并根据这些候选卡点以及候选卡点的能量来确定音频的目标卡点，从而可以根据目标卡点来匹配视频画面。这种根据音频的能量来确定音频的节奏的方法，能够确定出更多的包括但不限于节拍点在内的在音频中具有影响力的多个候选卡点，进而再根据这些候选卡点的能量有效保留在音频中具有影响力和重要性的目标卡点，这样不仅能够确保关键卡点的保留，还能够满足用户对根据卡点来应用视频转场的需求，提升了用户体验。

图 3 示出了本公开的一些实施例的音频的卡点 300 的示意图。为了清晰地说明本公开的一些实施例的卡点的点位，下面将结合图 3 来说明。当一个声音从开始到结束的过程可以大致分为起始（onset）、瞬态（attack）、衰减（decay）这三个过程。在声音的开始阶段，声音的起始点 310 标志着声音能量的突然增加和声音事件的开始，声音的振幅和能量可以从零或非常低的水平开始显著上升，这是声音可被识别的起点，也是后续阶段的开端。图 3 所示的起始点 310 即为音频中的卡点。

继续参考图 3，在声音的起始阶段后紧跟着瞬态阶段 320，瞬态阶段 320 描述了声音从起始到最大音量的快速上升过程，即声音在极短的时间内迅速增强。在瞬态阶段 320，声音的振幅包络急剧增加，音量快速上升并达到峰值，这个阶段通常很短，但它决定了声音的清晰度和冲击力。爆发性声音 330 通常与音符开始时的瞬态变化重叠，其可描述为在乐音或声音事件开始时出现的短持续时间和高振幅的类噪声声音分量。爆发性声音 330 区域信号变化快速且不可预测。例如，在钢琴演奏中，爆发性声音 330 对应于琴键被敲击、琴弦开始振动的初始阶段。

如图 3 所示，当声音达到峰值后，音量逐渐减小直至声音消失的过程是衰减阶段 340。在这个阶段，声音的振幅逐渐降低，能量逐渐消散。在一些乐器中，衰减可能是迅速的，而在其它情形下，衰减可能是较为缓慢的。

图 4 示出了本公开的一些实施例的用于在多个候选卡点中确定音频的目标卡点 400 的示意图。参考图 4，在 410，检测音频的候选卡点。在一些实施例中，可以采用 onset 检测的方法来确定出音频的候选卡点。在一些实施例中，这种 onset 检测的方法可以是基于音频能量的，还可以是基于频谱的，或者是可以基于频谱和能量的结合的

方法的。在一些实施例中，这些被检测出来的候选卡点可以是音频的节拍点或者某个声音或者音乐事件的起始点。通过 onset 检测不仅能得到数量更多，类型更全的候选卡点，还能得到候选卡点的时间戳以及候选卡点的能量值。

如图 4 所示，在 420，将候选卡点按照能量值降序排序。假设检测到有以下候选卡点的表 1。（时间单位：秒）

时间戳	2.5	3.3	3.4	8.0	……	8.1
能量值	100	90	80	60	……	70

经过根据能量降序排序后，可以得到表 2。

能量值	100	90	80	70	……	60
时间戳	2.5	3.3	3.4	8.1	……	8.0

见图 4，在 430，初始化数量数组。该数组用于记录时间滑窗中已放入的候选卡点的数量，并且使用时间滑窗相对于时间起点的步长偏移量作为下标。例如，假设时间序列的数据是从 T0 时刻开始的，使用 10 秒的时间滑窗，从 T0 时刻开始，第一个 10 秒的窗口的偏移量是 0（因为它从起点开始），第二个 10 秒的窗口的偏移量是 1（因为它从 T0+1 开始），依此类推。这样，能够简化时间滑窗的管理和访问方式。

继续参考图 4，在 440，通过滑动窗口来遍历排序过的候选卡点。下面将结合图 5 来说明本公开的实施例的一些通过滑动窗口来确定音频的卡点 500 的示意图。假设设定的每个时间窗口内的候选卡点的上限为 20 个，时间窗口大小为 5 个步长，每次移动的步长为 1。

如图 4 所示，在 442，根据目标卡点的时间戳计算所有包含候选卡点的时间窗口。在图 5 中，区域 510 示出了候选卡点的时间戳。假设有一个候选卡点在[n,n+1]之间，那么在[n,n+1]之间的这个候选卡点可以影响区域 520 中的时间窗口[n-4,n+1]、时间窗口[n-3,n+2]、时间窗口[n-2,n+3]、时间窗口[n-1,n+4]以及时间窗口[n,n+5]。可以理解的是，在这里不考虑负的时间点，所以起始时间至少为 0。

继续参考图 4，由于候选卡点可能不是恰好在时间窗口的起始或结束时发生，一个候选卡点可能会影响多个相邻的时间窗口。为了确保在任何 5 个步长时间窗口内候选卡点的数量小于 20 个，需要检查并更新所有受影响的时间窗口的计数值。在 444，判断时间窗口中已有的候选卡点的数量是否大于等于上限即 20 个，如果不是的话，则在 448，将该候选卡点加入结果列表中，并更新用于记录每个窗口内已放入的候选卡点的数量。结合图 5，在区域 530 中的数组中的计数器记录的值都小于上限 20，表示可以将在[n,n+1]之间的这个候选卡点加入结果列表之中，并且更新数组的值即对数组中的对应的窗口的计数值均加 1。

继续参考图 4，在 444，判断时间窗口中已有的候选卡点的数量是否大于等于上

限即 20，如果是的话，则在 446，跳过这个候选卡点并重新判断下一个候选卡点。结合图 5，例如时间窗口[n+1,n+6]中的值等于 20，那么与时间窗口[n+1,n+6]对应的候选卡点将被跳过，并判断这个候选卡点后的候选卡点。如此，针对时间戳上的每个候选卡点都进行如上操作，可以将满足条件的候选卡点加入至结果列表。进而，在 450，对结果列表中的候选卡点按照时间戳的顺序升序排序，从而在 460，将升序排序后的结果列表中的候选卡点作为音频的目标卡点。

为了更清晰地说明本公开的实施例的通过滑动窗口来确定音频的卡点的过程，下面将举例说明，假设在窗口大小为 10 秒，步长为 1 秒的任意时间窗口内，设定窗口的候选卡点的数量小于 20 个，并优先保留能量最高的候选卡点。假设降序排序后的多个候选卡点的时间戳及其能量值如表 3 所示：

能量值	100	90	80	70	60	……
时间戳	2.5	3.3	3.5	8.0	8.5	……

接下来，可以定义的数量数组来记录每个时间窗口内的候选卡点的数量，时间窗口的大小是 10 秒，步长是 1 秒，可以为每个可能的起始时间点（从 0 秒开始，每 1 秒 1 个点）来初始化一个计数器。假设初始化数组为[0,0,0,0,0,0,0,0,0,0]，这里的每个元素代表一个时间窗口的候选卡点的计数器，下标代表时间窗口的起始时间（以秒为单位）。

接着，将通过滑动窗口来遍历排序过的候选卡点。针对 2.5 秒的候选卡点，其影响的时间窗口有[0-10],[1-11]以及[2-12]。判断这些时间窗口的计数器，均为 0，未达到上限，将 2.5 秒的候选卡点加入结果列表，并更新数量数组为[1,1,1,0,0,0,0,0,0,0]。

针对 3.3 秒的候选卡点，其影响的时间窗口有[0-10],[1-11],[2-12]以及[3-13]。判断这些时间窗口的计数器，前三个时间窗口计数为 1，未达到上限，[3-13]的时间窗口的计数器为 0，将 3.3 秒的候选卡点加入结果列表，并更新数量数组为[2,2,2,1,0,0,0,0,0,0]。

针对 3.5 秒的候选卡点，其影响的时间窗口有[0-10],[1-11],[2-12]以及[3-13]。判断这些时间窗口的计数器，未达到上限，将 3.5 秒的候选卡点加入结果列表，并更新数量数组为[3,3,3,2,0,0,0,0,0,0]。

针对 8.0 秒的候选卡点，其影响的时间窗口有[0-10],[1-11]……[8-18]。判断这些时间窗口的计数器，未达到上限，将 8.0 秒的候选卡点加入结果列表，并更新数量数组为[4,4,4,3,1,1,1,1,1,1]。

针对 8.5 秒的候选卡点，其影响的时间窗口有[0-10]，[1-11]……[8-18]。判断这些时间窗口的计数器，未达到上限，将 8.5 秒的候选卡点加入结果列表，并更新数量数组为[5,5,5,4,2,2,2,2,2,2]。

接着对结果列表中的候选卡点，按照时间戳的顺序升序排序，可以得到表 4，如下：

时间戳	2.5	3.3	3.5	8.0	8.5	……
能量值	100	90	80	70	60	……

表 4 所示的候选卡点即为音频的目标卡点。

又例如，假设有一系列的经过能量降序排序后候选卡点，时间从 0 秒开始，每隔 0.1 秒就有一个候选卡点，且能量值相同。当处理到 1.0 秒的候选卡点时，时间窗口[0-10]的计数就会增加到 10，当处理到 1.9 秒的候选卡点时，时间窗口[0-10]的计数就会增加到 20，此时该时间窗口已经达到了设定的上限，该 1.9 秒的候选卡点将被跳过，并判断后面的 2.0 秒的候选卡点。接下来的候选卡点（如 2.0 秒、2.1 秒等）也会影响到[0-10]的时间窗口，所以这些候选卡点也不会被加入至结果列表中。可以理解的是，实际应用中候选卡点可能不是规律地出现，但以上有助于理解根据滑动窗口来筛选候选卡点的过程。

图 6A 示出了本公开的一些实施例的用于导入待剪辑的视频的 600A 的示意图。如图 6A 所示，在剪辑功能 610A 的显示界面 602A 的显眼位置显示有用于导入待剪辑视频的“开始创作”按钮 604A，用户点击后即可以进度导入视频或者照片的界面。在按钮 604A 的下方，显示有用户可以使用的模板区域 606A，这样能够方便用户快速进入创作状态。在显示页面 602A 的中下部区域 608A，展示了用户之前导入的但未完成编辑的视频草稿，这样能够方便用户随时继续编辑。

图 6B 示出了本公开的一些实施例的用于导入音频 600B 的示意图。结合图 6A，当用户点击“开始创作”按钮 604A，并导入了待编辑的视频 604B 后，可以跳转到显示界面 602B。在显示界面 602B 中，用户可以为待编辑的视频 604B 配乐，并根据该配乐来调整视频 604B 的画面。具体而言，用户可以通过点击添加音频按钮 608B 来跳转到添加音频的页面，从而可以添加其需要的音频，这些音频可以保存于本地也可以从云端获取。当导入了音频后，可以通过区域 606B 的关闭原声按钮来关闭视频的原声，这样能够方便用户聚焦于导入的音频与视频 604B 的匹配度上。当用户完成对视频已经音频的编辑后，可以点击导出按钮 610B 来保存和分享他们的作品。用户还可以看到他们作品的分辨率，还可以调整他们的作品的分辨率。

图 7 示出了本公开的一些实施例的用于确定音频的卡点的装置 700 的框图。如图 7 所示，装置 700 包括音频获取模块 702，被配置为获取用于制作视频的音频。装置 700 还包括候选卡点确定模块 704，被配置为确定音频的多个候选卡点以及多个候选卡点的多个能量。此外，装置 700 还包括目标卡点确定模块 706，被配置基于多个候选卡点以及多个能量，确定音频的多个目标卡点，其中多个目标卡点用于匹配视频

的画面。

图 8 示出了本公开的一些实施例的电子设备 800 的框图，设备 800 可以是本公开的实施例所描述的设备或装置。如图 8 所示，设备 800 包括中央处理单元（CPU）和/或图形处理单元（GPU）801，其可以根据存储在只读存储器（ROM）802 中的计算机程序指令或者从存储单元 808 加载到随机访问存储器（RAM）803 中的计算机程序指令，来执行各种适当的动作和处理。在 RAM 803 中，还可以存储设备 800 操作所需的各种程序和数据。CPU/GPU 801、ROM 802 以及 RAM 803 通过总线 804 彼此相连。输入/输出（I/O）接口 805 也连接至总线 804。虽然未在图 8 中示出，设备 800 还可以包括协处理器。

设备 800 中的多个部件连接至 I/O 接口 805，包括：输入单元 806，例如键盘、鼠标等；输出单元 807，例如各种类型的显示器、扬声器等；存储单元 808，例如磁盘、光盘等；以及通信单元 809，例如网卡、调制解调器、无线通信收发机等。通信单元 809 允许设备 800 通过诸如互联网的计算机网络和/或各种电信网络与其他设备交换信息/数据。

上文所描述的各个方法或过程可以由 CPU/GPU 801 来执行。例如，在一些实施例中，方法可被实现为计算机软件程序，其被有形地包含于机器可读介质，例如存储单元 808。在一些实施例中，计算机程序的部分或者全部可以经由 ROM 802 和/或通信单元 809 而被载入和/或安装到设备 800 上。当计算机程序被加载到 RAM 803 并由 CPU/GPU 801 执行时，可以执行上文描述的方法或过程中的一个或多个步骤或动作。

在一些实施例中，以上所描述的方法和过程可以被实现为计算机程序产品。计算机程序产品可以包括计算机可读存储介质，其上载有用于执行本公开的各个方面的计算机可读程序指令。

计算机可读存储介质可以是可以保持和存储由指令执行设备使用的指令的有形设备。计算机可读存储介质例如可以是但不限于电存储设备、磁存储设备、光存储设备、电磁存储设备、半导体存储设备或者上述的任意合适的组合。计算机可读存储介质的更具体的例子（非穷举的列表）包括：便携式计算机盘、硬盘、随机存取存储器（RAM）、只读存储器（ROM）、可擦式可编程只读存储器（EPROM 或闪存）、静态随机存取存储器（SRAM）、便携式压缩盘只读存储器（CD-ROM）、数字多功能盘（DVD）、记忆棒、软盘、机械编码设备、例如其上存储有指令的打孔卡或凹槽内凸起结构、以及上述的任意合适的组合。这里所使用的计算机可读存储介质不被解释为瞬时信号本身，诸如无线电波或者其他自由传播的电磁波、通过波导或其他传输媒介传播的电磁波（例如，通过光纤电缆的光脉冲）、或者通过电线传输的电信号。

本文所描述的计算机可读程序指令可以从计算机可读存储介质下载到各个计算/处理设备，或者通过网络、例如互联网、局域网、广域网和/或无线网下载到外部计

算机或外部存储设备。网络可以包括铜传输电缆、光纤传输、无线传输、路由器、防火墙、交换机、网关计算机和/或边缘服务器。每个计算/处理设备中的网络适配卡或者网络接口从网络接收计算机可读程序指令，并转发该计算机可读程序指令，以供存储在各个计算/处理设备中的计算机可读存储介质中。

用于执行本公开操作的计算机程序指令可以是汇编指令、指令集架构（ISA）指令、机器指令、机器相关指令、微代码、固件指令、状态设置数据、或者以一种或多种编程语言的任意组合编写的源代码或目标代码，编程语言包括面向对象的编程语言，以及常规的过程式编程语言。计算机可读程序指令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（例如利用互联网服务提供商来通过互联网连接）。在一些实施例中，通过利用计算机可读程序指令的状态信息来个性化定制电子电路，例如可编程逻辑电路、现场可编程门阵列（FPGA）或可编程逻辑阵列（PLA），该电子电路可以执行计算机可读程序指令，从而实现本公开的各个方面。

这些计算机可读程序指令可以提供给通用计算机、专用计算机或其它可编程数据处理装置的处理单元，从而生产出一种机器，使得这些指令在通过计算机或其它可编程数据处理装置的处理单元执行时，产生了实现流程图和/或框图中的一个或多个框中规定的功能/动作的装置。也可以把这些计算机可读程序指令存储在计算机可读存储介质中，这些指令使得计算机、可编程数据处理装置和/或其他设备以特定方式工作，从而，存储有指令的计算机可读介质则包括一个制造品，其包括实现流程图和/或框图中的一个或多个框中规定的功能/动作的各个方面的指令。

也可以把计算机可读程序指令加载到计算机、其它可编程数据处理装置、或其它设备上，使得在计算机、其它可编程数据处理装置或其它设备上执行一系列操作步骤，以产生计算机实现的过程，从而使得在计算机、其它可编程数据处理装置、或其它设备上执行的指令实现流程图和/或框图中的一个或多个框中规定的功能/动作。

附图中的流程图和框图显示了根据本公开的多个实施例的设备、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框可以代表一个模块、程序段或指令的一部分，所述模块、程序段或指令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。在有些作为替换的实现中，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，两个连续的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，这取决于所涉及的功能。也要注意的是，框图和/或流程图中的每个框、以及框图和/或流程图中的框的组合，

可以用执行规定的功能或动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。

以上已经描述了本公开的各实施例，上述说明是示例性的，并非穷尽性的，并且也不限于所公开的各实施例。在不偏离所说明的各实施例的范围和精神的情况下，对于本技术领域的普通技术人员来说许多修改和变更都是显而易见的。本文中所用术语的选择，旨在最好地解释各实施例的原理、实际应用或对市场中技术的技术改进，或者使得本技术领域的其它普通技术人员能理解本文公开的各实施例。

下列出了本公开的一些示例实现。

示例 1. 一种用于确定音频的卡点的方法，包括：

获取用于制作视频的音频；

确定所述音频的多个候选卡点以及所述多个候选卡点的多个能量；以及

基于所述多个候选卡点以及所述多个能量，确定所述音频的多个目标卡点，所述多个目标卡点用于匹配视频的画面。

示例 2. 根据示例 1 所述的方法，其中确定所述音频的多个候选卡点以及所述多个候选卡点的多个能量包括：

确定所述多个候选卡点中的每个候选卡点在所述音频中的时间戳以及能量值，所述能量值指示所述音频的音符和/或音乐事件的强度；以及

基于所述能量值，对所述多个候选卡点中的每个候选卡点降序排序。

示例 3. 根据示例 1-2 中任一项所述的方法，其中基于所述多个候选卡点以及所述多个能量，确定所述音频的多个目标卡点包括：

基于预定单位时间内的卡点的上限，通过遍历经降序排列的所述多个候选卡点，来确定所述音频的所述多个目标卡点。

示例 4. 根据示例 1-3 中任一项所述的方法，其中基于预定单位时间内的上限，通过遍历经降序排列的所述多个候选卡点，来确定所述音频的所述多个目标卡点包括：

由时间滑动窗口基于滑动时间步长在经降序排列的所述多个候选卡点上确定多个时间窗口，所述时间滑动窗口指示所述预定单位时间；

基于经降序排列的每个候选卡点以及每个候选卡点的时间戳，确定每个时间窗口内的所述候选卡点的数量；以及

基于所述数量，确定所述多个候选卡点中的每个候选卡点是否为所述目标卡点。

示例 5. 根据示例 1-4 中任一项所述的方法，其中基于经降序排列的每个候选卡点以及每个候选卡点的时间戳，确定每个时间窗口内的所述候选卡点的数量包括：

基于所述滑动时间步长，定义数量数组，所述数量数组用于记录每个时间窗口中的候选卡点的数量。

示例 6. 根据示例 1-5 中任一项所述的方法，其中基于所述数量，确定所述多个候选卡点中的每个候选卡点是否为所述目标卡点包括：

响应于每个时间窗口内的所述候选卡点的所述数量大于所述上限，跳过给定候选卡点；或者

响应于每个时间窗口内的所述候选卡点的所述数量小于等于所述上限，将给定候选卡点加入结果列表，并更新所述数量数组；以及

将所述结果列表中的所加入的给定候选卡点确定为所述音频的所述目标卡点。

示例 7. 根据示例 1-6 中任一项所述的方法，其中将所述结果列表中的所加入的给定候选卡点确定为所述音频的所述目标卡点包括：

基于所述结果列表中的所加入的多个候选卡点的时间戳，对所述多个候选卡点升序排序，经升序排序的所述结果列表中的所述候选卡点为所述音频的所述目标卡点。

示例 8. 根据示例 1-7 中任一项所述的方法，获取所述音频包括：

响应于检测到用户对第一按钮的点击，显示第一页面，所述第一页面用于显示本地视频素材以及云端视频素材中的至少一项；以及

响应于所述视频已被导入，显示第二页面，所述第二页面用于显示第二按钮，所述第二按钮用于导入所述音频。

示例 9. 根据示例 1-8 中任一项所述的方法，还包括：

响应于检测到所述用户对所述第二按钮的点击，显示第三页面，所述第三页面用于显示本地音频素材以及云端音频素材中的至少一项；以及

响应于检测到所述用户对所述第二页面上的第三按钮的点击，导出所编辑的视频。

示例 10. 一种用于确定音频的卡点的装置，包括：

音频获取模块，被配置为获取用于制作视频的音频；

候选卡点确定模块，被配置为确定所述音频的多个候选卡点以及所述多个候选卡点的多个能量；以及

目标卡点确定模块，被配置为基于所述多个候选卡点以及所述多个能量，确定所述音频的多个目标卡点，所述多个目标卡点用于匹配视频的画面。

示例 11. 根据示例 10 所述的装置，其中候选卡点确定模块包括：

能量值确定模块，被配置为确定所述多个候选卡点中的每个候选卡点在所述音频中的时间戳以及能量值，所述能量值指示所述音频的音符和/或音乐事件的强度；以及

降序模块，被配置为基于所述能量值，对所述多个候选卡点中的每个候选卡点降序排序。

示例 12. 根据示例 10-11 中任一项所述的装置，其中目标卡点确定模块包括：

遍历模块，被配置为基于预定单位时间内的卡点的上限，通过遍历经降序排列的所述多个候选卡点，来确定所述音频的所述多个目标卡点。

示例 13. 根据示例 10-12 中任一项所述的装置，其中遍历模块包括：

时间窗口确定模块，被配置为由时间滑动窗口基于滑动时间步长在经降序排列的所述多个候选卡点上确定多个时间窗口，所述时间滑动窗口指示所述预定单位时间；

数量确定模块，被配置为基于经降序排列的每个候选卡点以及每个候选卡点的时间戳，确定每个时间窗口内的所述候选卡点的数量；以及

判断模块，被配置为基于所述数量，确定所述多个候选卡点中的每个候选卡点是否为所述目标卡点。

示例 14. 根据示例 10-13 中任一项所述的装置，其中数量确定模块包括：

数组定义模块，被配置为基于所述滑动时间步长，定义数量数组，所述数量数组用于记录每个时间窗口中的候选卡点的数量。

示例 15. 根据示例 10-14 中任一项所述的装置，其中判断模块包括：

跳过模块，被配置为响应于每个时间窗口内的所述候选卡点的所述数量大于等于所述上限，跳过给定候选卡点；或者

加入模块，被配置为响应于每个时间窗口内的所述候选卡点的所述数量小于所述上限，将给定候选卡点加入结果列表，并更新所述数量数组；以及

确定模块，被配置为将所述结果列表中的所加入的给定候选卡点确定为所述音频的所述目标卡点。

示例 16. 根据示例 10-15 中任一项所述的装置，其中确定模块包括：

升序模块，被配置为基于所述结果列表中的所加入的多个候选卡点的时间戳，对所述多个候选卡点升序排序，经升序排序的所述结果列表中的所述候选卡点为所述音频的所述目标卡点。

示例 17. 根据示例 10-16 中任一项所述的装置，其中音频获取模块包括：

第一页面实现模块，被配置为响应于检测到用户对第一按钮的点击，显示第一页面，所述第一页面用于显示本地视频素材以及云端视频素材中的至少一项；以及

第二页面显示模块，被配置为响应于所述视频已被导入，显示第二页面，所述第二页面用于显示第二按钮，所述第二按钮用于导入所述音频。

示例 18. 根据示例 10-17 中任一项所述的装置，还包括：

第三页面显示模块，被配置为响应于检测到所述用户对所述第二按钮的点击，显示第三页面，所述第三页面用于显示本地音频素材以及云端音频素材中的至少一项；以及

导出模块，被配置为响应于检测到所述用户对所述第二页面上的第三按钮的点击，导出所编辑的视频。

示例 19. 一种电子设备，包括：

处理器；以及

与所述处理器耦合的存储器，所述存储器具有存储于其中的指令，所述指令在被处理器执行时，使得所述电子设备执行动作，所述动作包括：

获取用于制作视频的音频；

确定所述音频的多个候选卡点以及所述多个候选卡点的多个能量；以及

基于所述多个候选卡点以及所述多个能量，确定所述音频的多个目标卡点，所述多个目标卡点用于匹配视频的画面。

示例 20. 根据示例 19 所述的电子设备，其中确定所述音频的多个候选卡点以及所述多个候选卡点的多个能量包括：

确定所述多个候选卡点中的每个候选卡点在所述音频中的时间戳以及能量值，所述能量值指示所述音频的音符和/或音乐事件的强度；以及

基于所述能量值，对所述多个候选卡点中的每个候选卡点降序排序。

示例 21. 根据示例 19-20 中任一项所述的电子设备，其中基于所述多个候选卡点以及所述多个能量，确定所述音频的多个目标卡点包括：

基于预定单位时间内的卡点的上限，通过遍历经降序排列的所述多个候选卡点，来确定所述音频的所述多个目标卡点。

示例 22. 根据示例 19-21 中任一项所述的电子设备，其中基于预定单位时间内的上限，通过遍历经降序排列的所述多个候选卡点，来确定所述音频的所述多个目标卡点包括：

由时间滑动窗口基于滑动时间步长在经降序排列的所述多个候选卡点上确定多个时间窗口，所述时间滑动窗口指示所述预定单位时间；

基于经降序排列的每个候选卡点以及每个候选卡点的时间戳，确定每个时间窗口内的所述候选卡点的数量；以及

基于所述数量，确定所述多个候选卡点中的每个候选卡点是否为所述目标卡点。

示例 23. 根据示例 19-22 中任一项所述的电子设备，其中基于经降序排列的每个候选卡点以及每个候选卡点的时间戳，确定每个时间窗口内的所述候选卡点的数量包括：

基于所述滑动时间步长，定义数量数组，所述数量数组用于记录每个时间窗口中的候选卡点的数量。

示例 24. 根据示例 19-23 中任一项所述的电子设备，其中基于所述数量，确定所述多个候选卡点中的每个候选卡点是否为所述目标卡点包括：

响应于每个时间窗口内的所述候选卡点的所述数量大于等于所述上限，跳过给定候选卡点；或者

响应于每个时间窗口内的所述候选卡点的所述数量小于所述上限，将给定候选卡点加入结果列表，并更新所述数量数组；以及

将所述结果列表中的所加入的给定候选卡点确定为所述音频的所述目标卡点。

示例 25. 根据示例 19-24 中任一项所述的电子设备，其中将所述结果列表中的所加入的给定候选卡点确定为所述音频的所述目标卡点包括：

基于所述结果列表中的所加入的多个候选卡点的时间戳，对所述多个候选卡点升序排序，经升序排序的所述结果列表中的所述候选卡点为所述音频的所述目标卡点。

示例 26. 根据示例 19-25 中任一项所述的电子设备，其中获取所述音频包括：

响应于检测到用户对第一按钮的点击，显示第一页面，所述第一页面用于显示本地视频素材以及云端视频素材中的至少一项；以及

响应于所述视频已被导入，显示第二页面，所述第二页面用于显示第二按钮，所述第二按钮用于导入所述音频。

示例 27. 根据示例 19-26 中任一项所述的电子设备，还包括：

响应于检测到所述用户对所述第二按钮的点击，显示第三页面，所述第三页面用于显示本地音频素材以及云端音频素材中的至少一项；以及

响应于检测到所述用户对所述第二页面上的第三按钮的点击，导出所编辑的视频。

示例 28. 一种计算机可读存储介质，其上存储有计算机可执行指令，其中所述计算机可执行指令被处理器执行以实现根据示例 1 至 9 中任一项所述的方法。

示例 29. 一种计算机程序产品，所述计算机程序产品被有形地存储在计算机可读介质上并且包括计算机可执行指令，所述计算机可执行指令在由设备执行时使所述设备执行根据示例 1 至 9 中任一项所述的方法。

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公开，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. 一种用于确定音频的卡点的方法，包括：

获取用于制作视频的音频；

确定所述音频的多个候选卡点以及所述多个候选卡点的多个能量；以及

基于所述多个候选卡点以及所述多个能量，确定所述音频的多个目标卡点，所述多个目标卡点用于匹配视频的画面。

2. 根据权利要求 1 所述的方法，其中确定所述音频的多个候选卡点以及所述多个候选卡点的多个能量包括：

确定所述多个候选卡点中的每个候选卡点在所述音频中的时间戳以及能量值，所述能量值指示所述音频的音符和/或音乐事件的强度；以及

基于所述能量值，对所述多个候选卡点中的每个候选卡点降序排序。

3. 根据权利要求 2 所述的方法，其中基于所述多个候选卡点以及所述多个能量，确定所述音频的多个目标卡点包括：

基于预定单位时间内的卡点的上限，通过遍历经降序排列的所述多个候选卡点，来确定所述音频的所述多个目标卡点。

4. 根据权利要求 3 所述的方法，其中基于预定单位时间内的上限，通过遍历经降序排列的所述多个候选卡点，来确定所述音频的所述多个目标卡点包括：

由时间滑动窗口基于滑动时间步长在经降序排列的所述多个候选卡点上确定多个时间窗口，所述时间滑动窗口指示所述预定单位时间；

基于经降序排列的每个候选卡点以及每个候选卡点的时间戳，确定每个时间窗口内的所述候选卡点的数量；以及

基于所述数量，确定所述多个候选卡点中的每个候选卡点是否为所述目标卡点。

5. 根据权利要求 4 所述的方法，其中基于经降序排列的每个候选卡点以及每个候选卡点的时间戳，确定每个时间窗口内的所述候选卡点的数量包括：

基于所述滑动时间步长，定义数量数组，所述数量数组用于记录每个时间窗口中的候选卡点的数量。

6. 根据权利要求 5 所述的方法，其中基于所述数量，确定所述多个候选卡点中的每个候选卡点是否为所述目标卡点包括：

响应于每个时间窗口内的所述候选卡点的所述数量大于等于所述上限，跳过给定候选卡点；或者

响应于每个时间窗口内的所述候选卡点的所述数量小于所述上限，将给定候选卡

点加入结果列表，并更新所述数量数组；以及

将所述结果列表中的所加入的给定候选卡点确定为所述音频的所述目标卡点。

7. 根据权利要求 6 所述的方法，其中将所述结果列表中的所加入的给定候选卡点确定为所述音频的所述目标卡点包括：

基于所述结果列表中的所加入的多个候选卡点的时间戳，对所述多个候选卡点升序排序，经升序排序的所述结果列表中的所述候选卡点为所述音频的所述目标卡点。

8. 根据权利要求 1 所述的方法，其中获取所述音频包括：

响应于检测到用户对第一按钮的点击，显示第一页面，所述第一页面用于显示本地视频素材以及云端视频素材中的至少一项；以及

响应于所述视频已被导入，显示第二页面，所述第二页面用于显示第二按钮，所述第二按钮用于导入所述音频。

9. 根据权利要求 8 所述的方法，还包括：

响应于检测到所述用户对所述第二按钮的点击，显示第三页面，所述第三页面用于显示本地音频素材以及云端音频素材中的至少一项；以及

响应于检测到所述用户对所述第二页面上的第三按钮的点击，导出所编辑的视频。

10. 一种的用于确定音频的卡点的装置，包括：

音频获取模块，被配置为获取用于制作视频的音频；

候选卡点确定模块，被配置为确定所述音频的多个候选卡点以及所述多个候选卡点的多个能量；以及

目标卡点确定模块，被配置为基于所述多个候选卡点以及所述多个能量，确定所述音频的多个目标卡点，所述多个目标卡点用于匹配视频的画面。

11. 一种电子设备，包括：

处理器；以及

与所述处理器耦合的存储器，所述存储器具有存储于其中的指令，所述指令在被处理器执行时，使得所述电子设备执行根据权利要求 1 至 9 中任一项所述的方法。

12. 一种计算机程序产品，其包括计算机可执行指令，其中所述计算机可执行指令被处理器执行以实现根据权利要求 1 至 9 中任一项所述的方法。