METHOD AND APPARATUS FOR DETERMINING BEAT SYNCHRONIZATION POINT OF AUDIO, DEVICE, AND PRODUCT

[0001] The present application claims priority to Chinese Patent Application No. 202410773416.0, filed on June 14, 2024, and entitled "METHOD AND APPARATUS FOR DETERMINING BEAT SYNCHRONIZATION POINT OF AUDIO, DEVICE, AND PRODUCT", which is hereby incorporated by reference in its entirety.

TECHNICAL FIELD

[0002] The present disclosure generally relates to the field of computers, and more specifically, to a method, an apparatus, an electronic device, and a program product for determining a point of an audio.

BACKGROUND

[0003] Video editing is a process of nonlinear editing of a video source, which involves remixing images, background music, effects, scenes, and other assets with an original video. In this process, the music point of the background music plays an important role in video editing, which may make the video work more ornamental.

[0004] The music point refers to selecting a specific music rhythm when generating a video, so that the images of the video match the rhythm of the music, thereby achieving the effect of enhancing the appeal of the images and the viewing experience. For example, in videos of travel notes, video life logs, movie and TV drama reviews, and other types, the audience may be quickly brought into the plot through the music point, enhancing the viewing experience.

SUMMARY

[0005] Embodiments of the present disclosure provide a method, an apparatus, an electronic device, and a program product for determining a point of an audio.

[0006] According to a first aspect of the present disclosure, there is provided a method for determining a point of an audio. The method includes obtaining audio for generating a video. The method includes determining a plurality of candidate points of the audio and a plurality of energies of the plurality of candidate points. In addition, the method further includes determining a plurality of target points of the audio based on the plurality of candidate points and the plurality of energies, the plurality of target points being configured to match images of the video.

[0007] In a second aspect of the present disclosure, there is provided an apparatus for determining a point of an audio. The apparatus includes an audio obtaining module configured to obtain audio for generating a video. The apparatus further includes a candidate point determination module configured to determine a plurality of candidate points of the audio and a plurality of energies of the plurality of candidate points. In

addition, the apparatus further includes a target point determination module configured to determine a plurality of target points of the audio based on the plurality of candidate points and the plurality of energies, the plurality of target points being configured to match images of the video.

[0008] In a third aspect of the present disclosure, there is provided an electronic device. The electronic device includes a processor and a memory coupled with the processor, the memory having instructions stored therein, the instructions, when executed by the processor, causing the electronic device to perform the method according to the first aspect.

[0009] In a fourth aspect of the present disclosure, there is provided a computer program product having stored thereon computer-executable instructions, the computer-executable instructions being executed by a processor to implement the method of the first aspect.

[0010] The Summary is provided to introduce a selection of concepts in a simplified form, which will be further described in detail in the following Detailed Description. The Summary is not intended to identify key features or essential features of the claimed subject matter, nor is it intended to be configured to limit the scope of the claimed subject matter.

BRIEF DESCRIPTION OF THE DRAWINGS

[0011] The above and other features, advantages, and aspects of the embodiments of the present disclosure will become more apparent in combination with the drawings and with reference to the following detailed description. In the drawings, the same or similar reference symbols refer to the same or similar elements, where:

[0012] Fig. 1 shows a schematic diagram of an example environment in which some embodiments of the present disclosure may be implemented;

[0013] Fig. 2 shows a flowchart of a method for determining a point of an audio according to some embodiments of the present disclosure;

[0014] Fig. 3 shows a schematic diagram of audio point according to some embodiments of the present disclosure;

[0015] Fig. 4 shows a schematic diagram of determining target point of audio among a plurality of candidate points according to some embodiments of the present disclosure;

[0016] Fig. 5 shows a schematic diagram of determining a point of an audio through a sliding window according to some embodiments of the present disclosure;

[0017] Fig. 6A shows a schematic diagram of importing a video to be edited according to some embodiments of the present disclosure;

[0018] Fig. 6B shows a schematic diagram of importing an audio according to some embodiments of the present disclosure;

[0019] Fig. 7 shows a block diagram of an apparatus for determining a point of an audio according to some embodiments of the present disclosure; and

[0020] Fig. 8 shows a block diagram of an electronic device according to some embodiments of the present

disclosure.

[0021] Throughout the drawings, the same or similar reference numbers refer to the same or similar elements.

DETAILED DESCRIPTION OF EMBODIMENTS

[0022] It may be understood that the data involved in the technical solution (including but not limited to the data itself, acquisition, or use of the data) shall comply with the requirements of corresponding laws, regulations, and related provisions.

[0023] It may be understood that before using the technical solution disclosed in the embodiments of the present disclosure, the user shall be informed of the type, range of use, use scenarios, etc., of personal information involved in the present disclosure in an appropriate manner, and the authorization of the user shall be obtained in accordance with relevant laws and regulations.

[0024] For example, when an active request of the user is received, prompt information is sent to the user to clearly prompt the user that the requested operation will require access to and use of the user's personal information. In this way, the user may independently choose, based on the prompt information, whether to provide the personal information to software or hardware, such as an electronic device, an application, a server or a storage medium, that performs the operations of the technical solution of the present disclosure.

[0025] As an optional but non-limiting implementation, in response to receiving the active request from the user, the prompt information may be sent to the user in the form of, for example, a pop-up window, in which the prompt information may be presented in text. In addition, the pop-up window may also include a selection control for the user to choose whether to "agree" or "disagree" to provide the personal information to the electronic device.

[0026] It may be understood that the above process of notifying and obtaining user authorization is only illustrative and does not limit the implementations of the present disclosure, and other manners that satisfy the relevant laws and regulations may also be applied in the implementations of the present disclosure.

[0027] The embodiments of the present disclosure will be described in more detail below with reference to the drawings. Although some embodiments of the present disclosure are shown in the drawings, it should be understood that the present disclosure may be implemented in various forms and should not be construed as limited to the embodiments set forth herein; on the contrary, these embodiments are provided for a more thorough and complete understanding of the present disclosure. It should be understood that the drawings and embodiments of the present disclosure are only used for illustrative purposes, and are not configured to limit the protection scope of the present disclosure.

[0028] In the description of the embodiments of the present disclosure, the term "include/comprise" and similar terms should be understood as open-ended inclusions, that is, "include/comprise but not limited to". The term "based on" should be understood as "at least partially based on". The term "one embodiment" or "the embodiment" should be understood as "at least one embodiment". The terms "first", "second", etc. may refer to different or the same objects, unless explicitly stated. Other definitions, both explicit and implicit, may also be

included below.

[0029] As mentioned above, matching the image rhythm of a video with the rhythm of music is an important technique in video editing. In the related art, the beat of music is usually regarded as the rhythm point of the music, which cannot meet the user's needs for the music point in a video. This is because, in the case where there is no change in tempo in each measure of music, the point, as the basic unit of time in music, is evenly distributed in the time dimension of the music. At this time, only taking the point as the rhythm point of the music cannot meet the needs of video creators for the point. In addition, only using the beat of music as the rhythm point of a video may also filter out an obvious rhythm change point within a beat, such as several consecutive heavy bass notes within a beat.

[0030] In the embodiments of the present disclosure, by detecting a plurality of a large range of potential rhythm points (candidate points) in an audio and energies of these candidate points, key rhythm points (target points) in the audio may be accurately determined, and these target points may be used for precise matching with video images. This method of determining the rhythm of an audio according to energy analysis of the audio may determine more influential candidate points in the audio, including but not limited to the beats, and then determine influential and important target points in the audio according to the energies of these candidate points. This may not only ensure effective reservation of the key points, but also meet the user's needs for applying video transitions according to the points, thus improving the user experience.

[0031] Fig. 1 shows a schematic diagram of an example environment 100 in which some embodiments of the present disclosure may be implemented. As shown in Fig. 1, a display interface 102 is an interface for editing the point of the audio 110 for a video to be edited 104. The video to be edited 104 is displayed on the display interface 102, and through a playing toolbar 106, a user may pause, fast forward, or rewind the video to be edited 104, and may also achieve full-screen display of the video to be edited 104 through a full-screen playing button. The total duration and the current positioning time stamp of the video to be edited 104 are also displayed on the playing toolbar 106.

[0032] Continuing to refer to Fig. 1, an audio 110 is an audio selected by the user for the soundtrack of the video to be edited 104. In order to match the rhythm of the video to be edited 104 with the rhythm of the audio 110, the point of the audio 110 used by the user may be presented for the user, which may facilitate the user to apply transitions when the music changes significantly during video editing, making the audio and video harmonious. For example, when the user turns on a button of automatic point 108, the point 112 of the current audio 110 may be displayed under the audio 110, and the point 112 is not limited to the evenly distributed beats of each measure of the audio 110, but may also include the drum beat of the audio 110 or the point at which a bass instrument (bass) such as a guitar starts.

[0033] A point 112-1 in the point 112 shown in Fig. 1 may be three consecutive drum beats within a beat. In music, a measure is a rhythmic unit of fixed duration, usually separated by bar lines. Each measure usually contains one or more beats. The point is the basic unit of time in music, and the points are evenly distributed in a piece of music, unless there are specific tempo change marks (such as accelerando, ritardando, etc.).

[0034] Referring to Fig. 1, in order to achieve friendly user interaction and improve user experience, in

addition to automatically identifying the point for the user, the user may also manually edit the generated point 112 through an add point 114 button. For example, the user may add a point on his or her own or delete a plurality of generated candidate points, including at least the beats. After the user's operation, the audio and video may be saved or cancelled through a toolbar 116. Through this method of determining a large number of diverse candidate points not limited to the beats for the audio selected by the user, the user's need to synchronously achieve image changes when the music changes significantly may be satisfied, thus improving the user experience.

[0035] The process according to the embodiments of the present disclosure will be described in detail below with reference to Fig. 2 to Fig. 8. For ease of understanding, the specific data mentioned in the following description is exemplary and does not limit the protection scope of the present disclosure. It may be understood that the embodiments described below may further include additional actions not shown and/or may omit the actions shown, and the scope of the present disclosure is not limited in this regard.

[0036] Fig. 2 shows a flowchart of a method 200 for determining a point of an audio according to some embodiments of the present disclosure. The method 200 may be performed by an apparatus for determining a point of an audio. The apparatus may be, for example, a server, a system configured in the server, or an independent apparatus or system. The apparatus may be implemented by software and/or hardware. Next, method 200 will be schematically described by taking the apparatus for determining a point of an audio as the execution body as an example. Referring to Fig. 2, the method 200 may include a block 202, a block 204, and a block 206.

[0037] At block 202, an audio for generating a video is obtained. In some embodiments, these videos may be short videos or long videos, and these videos may be obtained from the cloud or locally. The audio for generating the video may be pure music audio or audio including human voice, etc., and this audio may be obtained locally or from the cloud.

[0038] At block 204, a plurality of candidate points of the audio and a plurality of energies of the plurality of candidate points are determined. These candidate points may be obtained by onset detection, and these candidate points are various forms of points, such as the beat of the audio, the drum beat of a certain instrument (such as a drum) of the audio, or the starting point of a certain bass instrument (such as a guitar). For example, it may also be that there are several consecutive heavy bass drum points within a point. Each point has its own energy, and in some embodiments, these energies represent the intensity of notes and/or changing music events in the audio.

[0039] At block 206, a plurality of target points of the audio are determined based on the plurality of candidate points and the plurality of energies, wherein the plurality of target points are configured to match images of the video. For example, after the plurality of candidate points are determined, considering that the number of the determined candidate points may be too dense and thus affect the transition effect of the user's point, the target points suitable for the user to perform image transition may be determined from the plurality of candidate points according to the energies of the plurality of candidate points, so that the user may achieve synchronous change of the video images and the audio rhythm point according to the determined target points.

[0040] In the embodiments of the present disclosure, the plurality of candidate points of the audio and the energies of the plurality of candidate points are detected, and the target points of the audio are determined according to these candidate points and the energies of the candidate points, so that the video images may be matched according to the target points. This method of determining the rhythm of the audio according to the energy of the audio may determine more influential candidate points in the audio, including but not limited to the beats, and then effectively reserve the influential and important target points in the audio according to the energies of these candidate points, which may not only ensure the reservation of the key points, but also meet the user's need for applying video transitions according to the points, thus improving the user experience.

[0041] Fig. 3 shows a schematic diagram of audio point 300 according to some embodiments of the present disclosure. To clearly illustrate the points of the point in some embodiments of the present disclosure, the following will illustrate in conjunction with Fig. 3. A process of a sound from start to end may be roughly divided into three processes: onset, attack, and decay. In the starting stage of the sound, an onset 310 of the sound marks a sudden increase of sound energy and the start of a sound event, and the amplitude and energy of the sound may rise significantly from zero or a very low level, which is the starting point at which the sound may be recognized and the beginning of the subsequent stage. The onset 310 shown in Fig. 3 is the point in the audio.

[0042] Continuing to refer to Fig. 3, the onset stage of the sound is followed by an attack stage 320, which describes a rapid rise of the sound from the onset to the maximum volume, that is, the sound is rapidly enhanced in a very short time. In the attack stage 320, the amplitude envelope of the sound increases sharply, and the volume rises quickly and reaches a peak. This stage is usually very short, but it determines the clarity and impact of the sound. A transient 330 usually overlaps with the transient change at the beginning of a note, and it may be described as a noise-like sound component with short duration and high amplitude that appears at the beginning of a tone or sound event. The signal in the transient 330 region changes fast and unpredictably. For example, in piano playing, the transient 330 corresponds to the initial stage when a key is struck, and a string starts to vibrate.

[0043] As shown in Fig. 3, after the sound reaches its peak, the process in which the volume gradually decreases until the sound disappears is a decay stage 340. In this stage, the amplitude of the sound gradually decreases, and the energy gradually dissipates. In some instruments, the decay may be rapid, while in other cases, the decay may be relatively slow.

[0044] Fig. 4 shows a schematic diagram of determining target points 400 of audio among a plurality of candidate points according to some embodiments of the present disclosure. Referring to Fig. 4, at 410, the candidate points of the audio are detected. In some embodiments, an onset detection method may be configured to determine the candidate points of the audio. In some embodiments, the onset detection method may be based on audio energy, frequency spectrum, or a combination of frequency spectrum and energy. In some embodiments, these detected candidate points may be the points of the audio or the onset of a certain sound or music event. Through the onset detection, not only can more candidate points of more complete types be obtained, but also the time stamps of the candidate points and the energy values of the candidate points may

be obtained.

[0045] As shown in Fig. 4, at 420, the candidate points are sorted in descending order of energy values. It is assumed that the following Table 1 of candidate points is detected (the time unit is seconds).

time stamp	2.5	3.3	3.4	8.0	……	8.1
energy value	100	90	80	60	……	70

[0046] After sorting according to energy in descending order, Table 2 may be obtained.

energy value	100	90	80	70	……	60
time stamp	2.5	3.3	3.4	8.1	……	8.0

[0047] Referring to Fig. 4, at 430, a number array is initialized. The array is configured to record the number of candidate points that have been put into the sliding time window, and the step offset of the sliding time window relative to the start of the time is used as the subscript. For example, it is assumed that the data of the time series starts from time T0, and a 10-second sliding time window is used. Starting from time T0, the offset of the first 10-second window is 0 (because it starts from the start), the offset of the second 10-second window is 1 (because it starts from T0+1), and so on. In this way, the management and access of the sliding time window may be simplified.

[0048] Continuing to refer to Fig. 4, at 440, the sorted candidate points are traversed through a sliding window. The following will illustrate some schematic diagrams of determining a point of an audio 500 through a sliding window according to the embodiments of the present disclosure in conjunction with Fig. 5. It is assumed that the upper limit of the number of candidate points in each time window is set to 20, the size of the time window is 5 steps, and the step size of each movement is 1.

[0049] As shown in Fig. 4, at 442, all time windows including a candidate point are calculated according to the time stamp of the target point. In Fig. 5, region 510 shows the time stamp of the candidate point. It is assumed that there is a candidate point between [n, n+1], then this candidate point between [n, n+1] may affect a time window [n-4, n+1], a time window [n-3, n+2], a time window [n-2, n+3], a time window [n-1, n+4] and a time window [n, n+5] in a region 520. It may be understood that negative time points are not considered here, so the start time is at least 0.

[0050] Continuing to refer to Fig. 4, because the candidate point may not occur exactly at the start or end of the time window, one candidate point may affect a plurality of adjacent time windows. In order to ensure that the number of candidate points in any 5-step time window is less than 20, it is necessary to check and update the count values of all affected time windows. At 444, whether the number of existing candidate points in the time window is greater than or equal to the upper limit, that is, 20, is determined, and if not, at 448, the candidate point is added to the result list, and the number of candidate points that have been put into each window is updated. With reference to Fig. 5, the values recorded by the counters in the array in a region 530 are all less than the upper limit 20, indicating that this candidate point between [n, n+1] may be added to the result list, and the values of the array are updated, that is, the count values of the corresponding windows in the array are all increased by 1.

[0051] Continuing to refer to Fig. 4, at 444, whether the number of existing candidate points in the time

window is greater than or equal to the upper limit, that is, 20, is determined, and if yes, at 446, the candidate point is skipped, and the next candidate point is re-determined. With reference to Fig. 5, for example, if the value in the time window [n+1, n+6] is equal to 20, the candidate point corresponding to the time window [n+1, n+6] will be skipped, and the candidate point after this candidate point is determined. In this way, the above operations are performed on each candidate point on the timestamp, and the candidate points that meet the conditions may be added to the result list. Then, at 450, the candidate points in the result list are sorted in ascending order of time stamp, so that at 460, the candidate points in the result list sorted in ascending order are used as the target points of the audio.

[0052] In order to more clearly illustrate the process of determining the audio point through the sliding window in the embodiments of the present disclosure, the following is an example. It is assumed that in any time window with a window size of 10 seconds and a step size of 1 second, the number of candidate points in the window is set to be less than 20, and the candidate point with the highest energy is preferentially reserved. It is assumed that the time stamps and energy values of a plurality of candidate points sorted in descending order are shown in Table 3 below.

energy value	100	90	80	70	60	……
time stamp	2.5	3.3	3.5	8.0	8.5	……

[0053] Next, a number array may be defined to record the number of candidate points in each time window. The size of the time window is 10 seconds, and the step size is 1 second, and a counter may be initialized for each possible start time point (starting from 0 second, one point every 1 second). It is assumed that the initialization array is [0,0,0,0,0,0,0,0,0,0], and each element here represents a counter of candidate points in a time window, and the subscript represents the start time of the time window (in seconds).

[0054] Next, the sorted candidate points will be traversed through the sliding window. For the 2.5-second candidate point, the time windows it affects are [0-10], [1-11], and [2-12]. The counters of these time windows are all 0, which have not reached the upper limit, so the 2.5-second candidate point is added to the result list, and the number array is updated to [1,1,1,0,0,0,0,0,0,0].

[0055] For the 3.3-second candidate point, the time windows it affects are [0-10], [1-11], [2-12], and [3-13]. The counters of these time windows are determined, and the count of the first three time windows is 1, which has not reached the upper limit, and the counter of the time window [3-13] is 0. The 3.3-second candidate point is added to the result list, and the number array is updated to [2,2,2,1,0,0,0,0,0,0].

[0056] For the 3.5-second candidate point, the time windows it affects are [0-10], [1-11], [2-12], and [3-13]. The counters of these time windows are determined, and they have not reached the upper limit. The 3.5-second candidate point is added to the result list, and the number array is updated to [3,3,3,2,0,0,0,0,0,0].

[0057] For the 8.0-second candidate point, the time windows it affects are [0-10], [1-11]...[8-18]. The counters of these time windows are determined, and they have not reached the upper limit. The 8.0-second candidate point is added to the result list, and the number array is updated to [4,4,4,3,1,1,1,1,1,1].

[0058] For the 8.5-second candidate point, the time windows it affects are [0-10], [1-11]...[8-18]. The counters of these time windows are determined, and they have not reached the upper limit. The 8.5-second

candidate point is added to the result list, and the number array is updated to [5,5,5,4,2,2,2,2,2].

[0059] Then, the candidate points in the result list are sorted in ascending order of time stamp, and Table 4 may be obtained as follows.

time stamp	2.5	3.3	3.5	8.0	8.5	……
energy value	100	90	80	70	60	……

[0060] The candidate points shown in Table 4 are the target points of the audio.

[0061] For another example, it is assumed that there is a series of candidate points sorted in descending order of energy, the time starts from 0 second, there is a candidate point every 0.1 second, and the energy values are the same. When a 1.0-second candidate point is processed, the count of the time window [0-10] will increase to 10, and when a 1.9-second candidate point is processed, the count of the time window [0-10] will increase to 20. At this time, the time window has reached the set upper limit, the 1.9-second candidate point will be skipped, and a subsequent 2.0-second candidate point is determined. The following candidate points (such as 2.0 seconds, 2.1 seconds, etc.) will also affect the time window [0-10], so these candidate points will not be added to the result list. It may be understood that in practical applications, the candidate points may not appear regularly, but the above is helpful to understand the process of selecting candidate points according to the sliding window.

[0062] Fig. 6A shows a schematic diagram of importing a video to be edited 600A according to some embodiments of the present disclosure. As shown in Fig. 6A, a "Start creating" button 604A for importing the video to be edited is displayed in a prominent position on a display interface 602A of an editing function 610A, and after the user clicks the button, the user may proceed to an interface for importing the video or a photo. Below the button 604A, a template region 606A that may be used by the user is displayed, which may facilitate the user to quickly enter a creating state. In a middle lower region 608A of the display interface 602A, a video draft that has been imported by the user but has not been edited is displayed, which may facilitate the user to continue editing at any time.

[0063] Fig. 6B shows a schematic diagram of importing an audio 600B according to some embodiments of the present disclosure. With reference to Fig. 6A, when the user clicks the "Start creating" button 604A and imports a video to be edited 604B, the display interface 602B may be jumped to. In the display interface 602B, the user may add a soundtrack to the video to be edited 604B and adjust the image of the video 604B according to the soundtrack. Specifically, the user may jump to an interface for adding audio by clicking an add audio button 608B, so that the required audio may be added, and this audio may be saved locally or obtained from the cloud. After the audio is imported, the original sound of the video may be turned off through a mute original sound button in a region 606B, which may facilitate the user to focus on the matching degree between the imported audio and the video 604B. After the user finishes editing the video and the audio, the user may click an export button 610B to save and share their work. The user may also see the resolution of their work and may adjust the resolution of their work.

[0064] Fig. 7 shows a block diagram of an apparatus 700 for determining a point of an audio according to some embodiments of the present disclosure. As shown in Fig. 7, the apparatus 700 includes an audio

obtaining module 702 configured to obtain an audio for generating a video. The apparatus 700 also includes a candidate point determination module 704 configured to determine a plurality of candidate points of the audio and a plurality of energies of the plurality of candidate points. In addition, the apparatus 700 also includes a target point determination module 706 configured to determine a plurality of target points of the audio based on the plurality of candidate points and the plurality of energies, the plurality of target points being configured to match images of the video.

[0065] Fig. 8 shows a block diagram of an electronic device 800 according to some embodiments of the present disclosure, and the device 800 may be the device or apparatus described in the embodiments of the present disclosure. As shown in Fig. 8, the device 800 includes a central processing unit (CPU) and/or a graphics processing unit (GPU) 801, which may perform various appropriate actions and processing operations according to computer program instructions stored in a read-only memory (ROM) 802 or computer program instructions loaded from a storage unit 808 into a random access memory (RAM) 803. In the RAM 803, various programs and data required for the operation of the device 800 may also be stored. The CPU/GPU 801, the ROM 802, and the RAM 803 are connected to each other through a bus 804. An input/output (I/O) interface 805 is also connected to the bus 804. Although not shown in Fig. 8, the device 800 may also include a coprocessor.

[0066] A plurality of components in the device 800 are connected to the I/O interface 805, and the components include: an input unit 806, such as a keyboard, a mouse, etc.; an output unit 807, such as various types of displays, speakers, etc.; a storage unit 808, such as a magnetic disk, an optical disc, etc.; and a communication unit 809, such as a network card, a modem, a wireless communication transceiver, etc. The communication unit 809 allows the device 800 to exchange information/data with other devices through computer network, such as the Internet and/or various telecommunication networks.

[0067] The various methods or processes described above may be executed by the CPU/GPU 801. For example, in some embodiments, the method may be implemented as a computer software program, which is tangibly contained in a machine-readable medium, such as the storage unit 808. In some embodiments, part or all of the computer program may be loaded and/or installed on the device 800 via the ROM 802 and/or the communication unit 809. When the computer program is loaded into the RAM 803 and executed by the CPU/GPU 801, one or more steps or actions in the methods or processes described above may be executed.

[0068] In some embodiments, the methods and processes described above may be implemented as a computer program product. The computer program product may include a computer-readable storage medium having computer-readable program instructions for performing various aspects of the present disclosure stored thereon.

[0069] The computer-readable storage medium may be a tangible device that holds and stores instructions used by an instruction execution device. The computer-readable storage medium may be, for example, but not limited to, an electrical storage device, a magnetic storage device, an optical storage device, an electromagnetic storage device, a semiconductor storage device, or any suitable combination of the above. More specific examples (a non-exhaustive list) of the computer-readable storage medium include: a portable

computer disk, a hard disk, a random access memory (RAM), a read-only memory (ROM), an erasable programmable read-only memory (EPROM or flash memory), a static random access memory (SRAM), a portable compact disc read-only memory (CD-ROM), a digital versatile disc (DVD), a memory stick, a floppy disk, a mechanically encoded device, such as a punch card or a raised structure in a groove on which instructions are stored, and any suitable combination of the above. The computer-readable storage medium used here is not interpreted as a transitory signal itself, such as a radio wave or other electromagnetic waves, an electromagnetic wave propagating through a waveguide or other transmission medium (for example, a light pulse through an optical fiber cable), or an electrical signal transmitted through a wire.

[0070] The computer-readable program instructions described herein may be downloaded from the computer-readable storage medium to various computing/processing devices, or downloaded to an external computer or an external storage device through a network, such as the Internet, a local area network, a wide area network and/or a wireless network. The network may include a copper transmission cable, an optical fiber transmission, a wireless transmission, a router, a firewall, a switch, a gateway computer and/or an edge server. A network adapter card or a network interface in each computing/processing device receives the computer-readable program instructions from the network and forwards the computer-readable program instructions for storage in the computer-readable storage medium in each computing/processing device.

[0071] The computer program instructions for performing the operations of the present disclosure may be assembly instructions, instruction set architecture (ISA) instructions, machine instructions, machine-related instructions, microcode, firmware instructions, state setting data, or source code or object code written in any combination of one or more programming languages, including object-oriented programming languages and conventional procedural programming languages. The computer-readable program instructions may be executed entirely on a user's computer, partly executed on the user's computer, executed as an independent software package, partly executed on the user's computer and partly executed on a remote computer, or entirely executed on the remote computer or server. In the case of involving the remote computer, the remote computer may be connected to the user's computer through any kind of network, including a local area network (LAN) or a wide area network (WAN), or it may be connected to an external computer (for example, connected through the Internet with the aid of an Internet service provider). In some embodiments, by customizing an electronic circuit, such as a programmable logic circuit, a field programmable gate array (FPGA), or a programmable logic array (PLA), with the state information of the computer-readable program instructions, the electronic circuit may execute the computer-readable program instructions, thereby implementing various aspects of the present disclosure.

[0072] These computer-readable program instructions may be provided to a processing unit of a general-purpose computer, a special-purpose computer or other programmable data processing apparatus to produce a machine, so that when these instructions are executed by the processing unit of the computer or other programmable data processing apparatus, an apparatus for implementing the functions/actions specified in one or more blocks in the flowchart and/or block diagram is produced. These computer-readable program instructions may also be stored in the computer-readable storage medium, and these instructions make the

computer, the programmable data processing apparatus and/or other devices work in a specific manner, so that the computer-readable medium storing the instructions includes a manufactured product, which includes instructions for implementing various aspects of the functions/actions specified in one or more blocks in the flowchart and/or block diagram.

[0073] The computer-readable program instructions may also be loaded onto a computer, other programmable data processing apparatus or other devices, so that a series of operation steps are executed on the computer, other programmable data processing apparatus or other devices to produce a computer-implemented process, so that the instructions executed on the computer, other programmable data processing apparatus or other devices implement the functions/actions specified in one or more blocks in the flowchart and/or block diagram.

[0074] The flowchart and block diagram in the drawings show the possibly implemented architectures, functions, and operations of the device, the method, and the computer program product according to the embodiments of the present disclosure. In this regard, each block in the flowchart or block diagram may represent a module, program segment, or part of an instruction, and the module, program segment, or part of an instruction contains one or more executable instructions for implementing the specified logical functions. In some alternative implementations, the functions marked in the blocks may also occur in an order different from that marked in the drawings. For example, two consecutive blocks may actually be executed substantially in parallel, or they may sometimes be executed in the reverse order, depending on the functions involved. It should also be noted that each block in the block diagram and/or the flowchart, and the combination of the blocks in the block diagram and/or the flowchart, may be implemented by a special-purpose hardware-based system that executes specified functions or actions, or may be implemented by a combination of special-purpose hardware and computer instructions.

[0075] The embodiments of the present disclosure have been described above, and the above description is exemplary, non-exhaustive, and not limited to the disclosed embodiments. Without departing from the scope and spirit of the illustrated embodiments, many modifications and changes will be apparent to those of ordinary skill in the art. The terms used herein are chosen to best explain the principles of the embodiments, the actual application, or the technical improvement to the technology in the market, or to enable other ordinary skilled persons in the art to understand the embodiments disclosed herein.

[0076] The following are some example implementations of the present disclosure.

[0077] Example 1. A method for determining a point of an audio, including:

[0078] obtaining an audio for generating a video;

[0079] determining a plurality of candidate points of the audio and a plurality of energies of the plurality of candidate points; and

[0080] determining a plurality of target points of the audio based on the plurality of candidate points and the plurality of energies, the plurality of target points being configured to match images of the video.

[0081] Example 2. The method of Example 1, where determining the plurality of candidate points of the audio and the plurality of energies of the plurality of candidate points includes:

[0082] determining a time stamp and an energy value, in the audio, of each of the plurality of candidate points, the energy value indicating an intensity of a note and/or a music event of the audio; and

[0083] sorting each of the plurality of candidate points in descending order based on the energy value.

[0084] Example 3. The method of any one of Examples 1-2, where determining the plurality of target points of the audio based on the plurality of candidate points and the plurality of energies includes:

[0085] determining the plurality of target points of the audio by traversing the plurality of candidate points sorted in descending order and based on an upper limit of points within a predetermined unit time.

[0086] Example 4. The method of any one of Examples 1-3, where determining the plurality of target points of the audio by traversing the plurality of candidate points sorted in descending order and based on the upper limit of the points within the predetermined unit time includes:

[0087] determining a plurality of time windows on the plurality of candidate points sorted in descending order using a sliding time window and based on a sliding time step, the sliding time window indicating the predetermined unit time;

[0088] determining a number of the candidate points within each of the plurality of time windows based on each of the plurality of candidate points sorted in descending order and the time stamp of each of the plurality of candidate points; and

[0089] determining whether each of the plurality of candidate points is the target point based on the number.

[0090] Example 5. The method of any one of Examples 1-4, where determining the number of the candidate points within each of the plurality of time windows based on each of the plurality of candidate points sorted in descending order and the time stamp of each of the plurality of candidate points includes:

[0091] defining a number array based on the sliding time step, the number array being configured to record a number of candidate points in each time window.

[0092] Example 6. The method of any one of Examples 1-5, where determining whether each of the plurality of candidate points is the target point based on the number includes:

[0093] skipping a given candidate point in response to the number of the candidate points within each time window being greater than or equal to the upper limit; or

[0094] adding a given candidate point to a result list and updating the number array in response to the number of the candidate points within each time window being less than the upper limit; and

[0095] determining the given candidate point added to the result list as the target points of the audio.

[0096] Example 7. The method of any one of Examples 1-6, where determining the given candidate point added to the result list as the target point of the audio includes:

[0097] sorting a plurality of candidate points added to the result list in ascending order based on time stamps of the plurality of candidate points added to the result list, the plurality of candidate points in the result list sorted in ascending order being the plurality of target points of the audio.

[0098] Example 8. The method of any one of Examples 1-7, where obtaining the audio includes:

[0099] displaying a first interface in response to detecting a click by a user on a first button, the first

interface being configured to display at least one of a local video asset or a cloud video asset; and

[00100] displaying a second interface in response to the video having been imported, the second interface being configured to display a second button, the second button being configured to import the audio.

[00101] Example 9. The method of any one of Examples 1-8, further including:

[00102] displaying a third interface in response to detecting the click by the user on the second button, the third interface being configured to display at least one of a local audio asset or a cloud audio asset; and

[00103] exporting an edited video in response to detecting a click of the user on a third button on the second interface.

[00104] Example 10. An apparatus for determining a point of an audio, including:

[00105] an audio obtaining module configured to obtain an audio for generating a video;

[00106] a candidate point determination module configured to determine a plurality of candidate points of the audio and a plurality of energies of the plurality of candidate points; and

[00107] a target point determination module configured to determine a plurality of target points of the audio based on the plurality of candidate points and the plurality of energies, the plurality of target points being configured to match images of the video.

[00108] Example 11. The apparatus of Example 10, where the candidate point determination module includes:

[00109] an energy value determination module configured to determine a time stamp and an energy value, in the audio, of each of the plurality of candidate points, the energy value indicating an intensity of a note and/or a music event of the audio; and

[00110] a descending order module configured to sort each of the plurality of candidate points in descending order based on the energy value.

[00111] Example 12. The apparatus of any one of Examples 10-11, where the target point determination module includes:

[00112] a traversing module configured to determine the plurality of target points of the audio by traversing the plurality of candidate points sorted in descending order and based on an upper limit of points within a predetermined unit time.

[00113] Example 13. The apparatus of any one of Examples 10-12, where the traversing module includes:

[00114] a time window determination module configured to determine a plurality of time windows on the plurality of candidate points sorted in descending order using a sliding time window and based on a sliding time step, the sliding time window indicating the predetermined unit time;

[00115] a number determination module configured to determine a number of the candidate points within each of the plurality of time windows based on each of the plurality of candidate points sorted in descending order and the time stamp of each of the plurality of candidate points; and

[00116] a judging module configured to determine whether each of the plurality of candidate points is the target point based on the number.

[00117] Example 14. The apparatus of any one of Examples 10-13, where the number determination module includes:

[00118] an array definition module configured to define a number array based on the sliding time step, the number array being configured to record a number of candidate points in each time window.

[00119] Example 15. The apparatus of any one of Examples 10-14, where the judging module includes:

[00120] a skipping module configured to skip a given candidate point in response to the number of the candidate points within each time window being greater than or equal to the upper limit; or

[00121] an adding module configured to add a given candidate point to a result list and update the number array in response to the number of the candidate points within each time window being less than the upper limit; and

[00122] a determination module configured to determine the given candidate point added to the result list as the target point of the audio.

[00123] Example 16. The apparatus of any one of Examples 10-15, where the determination module includes:

[00124] an ascending order module configured to sort a plurality of candidate points added to the result list in ascending order based on time stamps of the plurality of candidate points added to the result list, the plurality of candidate points in the result list sorted in ascending order being the plurality of target points of the audio.

[00125] Example 17. The apparatus of any one of Examples 10-16, where the audio obtaining module includes:

[00126] a first interface implementation module configured to display a first interface in response to detecting a click by a user on a first button, the first interface being configured to display at least one of a local video asset or a cloud video asset; and

[00127] a second interface display module configured to display a second interface in response to the video having been imported, the second interface being configured to display a second button, the second button being configured to import the audio.

[00128] Example 18. The apparatus of any one of Examples 10-17, further including:

[00129] a third interface display module configured to display a third interface in response to detecting the click by the user on the second button, the third interface being configured to display at least one of a local audio asset or a cloud audio asset; and

[00130] an exporting module configured to export an edited video in response to detecting a click by the user on a third button on the second interface.

[00131] Example 19. An electronic device, including:

[00132] a processor; and

[00133] a memory coupled with the processor, the memory having instructions stored therein, the instructions, when executed by the processor, causing the electronic device to perform acts including:

[00134] obtaining an audio for generating a video;

[00135] determining a plurality of candidate points of the audio and a plurality of energies of the plurality of candidate points; and

[00136] determining a plurality of target points of the audio based on the plurality of candidate points and the plurality of energies, the plurality of target points being configured to match images of the video.

[00137] Example 20. The electronic device of Example 19, where determining the plurality of candidate points of the audio and the plurality of energies of the plurality of candidate points includes:

[00138] determining a time stamp and an energy value, in the audio, of each of the plurality of candidate points, the energy value indicating an intensity of a note and/or a music event of the audio; and

[00139] sorting each of the plurality of candidate points in descending order based on the energy value.

[00140] Example 21. The electronic device of any one of Examples 19-20, where determining the plurality of target points of the audio based on the plurality of candidate points and the plurality of energies includes:

[00141] determining the plurality of target points of the audio by traversing the plurality of candidate points sorted in descending order and based on an upper limit of points within a predetermined unit time.

[00142] Example 22. The electronic device of any one of Examples 19-21, where determining the plurality of target points of the audio by traversing the plurality of candidate points sorted in descending order and based on the upper limit of the points within the predetermined unit time includes:

[00143] determining a plurality of time windows on the plurality of candidate points sorted in descending order using a sliding time window and based on a sliding time step, the sliding time window indicating the predetermined unit time;

[00144] determining a number of the candidate points within each of the plurality of time windows based on each of the plurality of candidate points sorted in descending order and the time stamp of each of the plurality of candidate points; and

[00145] determining whether each of the plurality of candidate points is the target point based on the number.

[00146] Example 23. The electronic device of any one of Examples 19-22, where determining the number of the candidate points within each of the plurality of time windows based on each of the plurality of candidate points sorted in descending order and the time stamp of each of the plurality of candidate points includes:

[00147] defining a number array based on the sliding time step, the number array being configured to record a number of candidate points in each time window.

[00148] Example 24. The electronic device of any one of Examples 19-23, where determining whether each of the plurality of candidate points is the target point based on the number includes:

[00149] skipping a given candidate point in response to the number of the candidate points within each time window being greater than or equal to the upper limit; or

[00150] adding a given candidate point to a result list and updating the number array in response to the number of the candidate points within each time window being less than the upper limit; and

[00151] determining the given candidate point added to the result list as the target point of the audio.

[00152] Example 25. The electronic device of any one of Examples 19-24, where determining the given candidate point added to the result list as the target point of the audio includes:

[00153] sorting a plurality of candidate points added to the result list in ascending order based on time stamps of the plurality of candidate points added to the result list, the plurality of candidate points in the result list sorted in ascending order being the plurality of target points of the audio.

[00154] Example 26. The electronic device of any one of Examples 19-25, where obtaining the audio includes:

[00155] displaying a first interface in response to detecting a click by a user on a first button, the first interface being configured to display at least one of a local video asset or a cloud video asset; and

[00156] displaying a second interface in response to the video having been imported, the second interface being configured to display a second button, the second button being configured to import the audio.

[00157] Example 27. The electronic device of any one of Examples 19-26, further including:

[00158] displaying a third interface in response to detecting the click by the user on the second button, the third interface being configured to display at least one of a local audio asset or a cloud audio asset; and

[00159] exporting an edited video in response to detecting a click by the user on a third button on the second interface.

[00160] Example 28. A computer-readable storage medium having stored thereon computer-executable instructions, the computer-executable instructions being executed by a processor to implement the method of any one of Examples 1-9.

[00161] Example 29. A computer program product tangibly stored on a computer-readable medium and including computer-executable instructions, the computer-executable instructions, when executed by a device, causing the device to perform the method of any one of Examples 1-9.

[00162] Although the present disclosure has been described in language specific to structural features and/or logical actions of the method, it should be understood that the subject matter defined in the appended claims is not necessarily limited to the specific features or actions described above. Rather, the specific features and actions described above are merely example forms for implementing the claims.

CLAIMS

1. A method for determining a point of an audio, comprising:

obtaining an audio for generating a video;

determining a plurality of candidate points of the audio and a plurality of energies of the plurality of candidate points; and

determining a plurality of target points of the audio based on the plurality of candidate points and the plurality of energies, the plurality of target points being configured to match images of the video.

2. The method of claim 1, wherein determining the plurality of candidate points of the audio and the plurality of energies of the plurality of candidate points comprises:

determining a time stamp and an energy value, in the audio, of each of the plurality of candidate points, the energy value indicating an intensity of a note and/or a music event of the audio; and

sorting each of the plurality of candidate points in descending order based on the energy value.

3. The method of claim 2, wherein determining the plurality of target points of the audio based on the plurality of candidate points and the plurality of energies comprises:

determining the plurality of target points of the audio by traversing the plurality of candidate points sorted in descending order and based on an upper limit of points within a predetermined unit time.

4. The method of claim 3, wherein determining the plurality of target points of the audio by traversing the plurality of candidate points sorted in descending order and based on the upper limit of the points within the predetermined unit time comprises:

determining a plurality of time windows on the plurality of candidate points sorted in descending order using a sliding time window and based on a sliding time step, the sliding time window indicating the predetermined unit time;

determining a number of the candidate points within each of the plurality of time windows based on each of the plurality of candidate points sorted in descending order and the time stamp of each of the plurality of candidate points; and

determining whether each of the plurality of candidate points is the target point based on the number.

5. The method of claim 4, wherein determining the number of the candidate points within each of the plurality of time windows based on each of the plurality of candidate points sorted in descending order and the time stamp of each of the plurality of candidate points comprises:

defining a number array based on the sliding time step, the number array being configured to record a number of candidate points in each time window.

6. The method of claim 5, wherein determining whether each of the plurality of candidate points is the target point based on the number comprises:

skipping a given candidate point in response to the number of the candidate points within each time window being greater than or equal to the upper limit; or

adding a given candidate point to a result list and updating the number array in response to the number of the candidate points within each time window being less than the upper limit; and

determining the given candidate point added to the result list as the target point of the audio.

7. The method of claim 6, wherein determining the given candidate point added to the result list as the target point of the audio comprises:

sorting a plurality of candidate points added to the result list in ascending order based on time stamps of the plurality of candidate points added to the result list, the plurality of candidate points in the result list sorted in ascending order being the plurality of target points of the audio.

8. The method of claim 1, wherein obtaining the audio comprises:

displaying a first interface in response to detecting a click by a user on a first button, the first interface being configured to display at least one of a local video asset or a cloud video asset; and

displaying a second interface in response to the video having been imported, the second interface being configured to display a second button, the second button being configured to import the audio.

9. The method of claim 8, further comprising:

displaying a third interface in response to detecting the click by the user on the second button, the third interface being configured to display at least one of a local audio asset or a cloud audio asset; and

exporting an edited video in response to detecting a click by the user on a third button on the second interface.

10. An apparatus for determining a point of an audio, comprising:

an audio obtaining module configured to obtain an audio for generating a video;

a candidate point determination module configured to determine a plurality of candidate points of the audio and a plurality of energies of the plurality of candidate points; and

a target point determination module configured to determine a plurality of target points of the audio based on the plurality of candidate points and the plurality of energies, the plurality of target points being configured to match images of the video.

11. An electronic device, comprising:

a processor; and

a memory coupled with the processor, the memory having instructions stored therein, the instructions, when executed by the processor, causing the electronic device to perform the method of any of claims 1 to 9.

12. A computer program product comprising computer-executable instructions, wherein the computer-

executable instructions are executed by a processor to implement the method of any of claims 1 to 9.