

METHOD AND SYSTEM FOR ACQUIRING ONLINE AUDIO-VISUAL CONTENT, SEPARATING TRACKS, INTEGRATING ARTIFICIAL INTELLIGENCE AND OUTPUTTING AI-GENERATED MEDIUM

FIELD OF THE INVENTION

[0001] The present invention relates to a method and a system for separating tracks of an online audio-visual content, integrating an AI-generated content generated by artificial intelligence, and outputting an AI-generated medium.

BACKGROUND OF THE INVENTION

[0002] Chinese Patent Publication No. CN116939323A discloses a music matching method, a device, an electronic device and a computer-readable storage medium, which mainly separates the tracks of the music to be matched and obtains target tracks of the music to be matched. The target tracks include a target vocal track, a target accompaniment track, a target bass track, and a target drumbeat track. According to the target drumbeat track, a first target drumbeat sequence and a second target drumbeat sequence are obtained. The terminal selects at least one target music from the candidate music and designates the video to be matched, which contains the target music, as the target video.

[0003] The above-mentioned patent has the following limitations:

[0004] 1. The matching between the target music and the target video is performed based on the target drumbeat track.

[0005] 2. It is unable to automatically generate a generated audio track that matches the target audio track.

[0006] 3. It can only separate multiple target audio tracks from the music to be matched and cannot perform operations such as object recognition, edge detection, or main character extraction for audio-visual sources containing images.

[0007] 4. It is unable to automatically generate a new matching image, matching animation or matching movie based on the selected target audio track.

SUMMARY OF THE INVENTION

[0008] According to one aspect of the present invention, a method for acquiring an online audio-visual content, separating tracks, integrating artificial intelligence and outputting an AI-generated medium comprises the following steps: a device server of an audio-visual playback device or a generative artificial intelligence server that is in signal communication with the device server acquiring an online audio-visual content from a streaming platform; the device server or the generative artificial intelligence server separating a plurality track contents from the online audio-visual content through a sampling module; the generative artificial intelligence server selecting part or all of the track contents, the selected track contents being aligned and positioned on a timeline, the generative artificial intelligence server generating at least one AI-generated content according to the selected track contents, the at least one AI-generated content being aligned with the track contents and positioned on the timeline; the generative artificial intelligence server mixing the selected track contents with the at least one AI-generated content, outputting an AI-generated medium and transmitting it back to the device server; the device server playing the AI-generated medium on a playback interface.

[0009] According to another aspect of the present invention, a system for acquiring an online audio-visual content, separating tracks, integrating artificial intelligence and outputting an AI-generated medium, being in signal communication with a streaming platform, comprises an audio-visual playback device, a sampling module, and a generative artificial intelligence server. The audio-visual playback device has a device server and a playback interface that is in signal communication with the device server. The sampling module is in signal communication with the device server. The generative artificial intelligence server is in signal communication with the device server and the sampling module. The device server or the generative artificial intelligence server acquires an online audio-visual content from the streaming platform, and then the device server or the generative artificial intelligence server separates a plurality track contents from the online audio-visual content through the sampling module. The generative artificial intelligence server selects part or all of the track contents. The selected track contents are aligned and positioned on a timeline. The generative artificial intelligence server generates at least one AI-generated content according to the selected track contents. The at least one AI-generated content is aligned with the track contents and positioned on the timeline. The generative artificial intelligence server mixes the selected track contents with the at least one AI-generated content, outputs an AI-generated medium and transmits it back to the device server. The device server plays the AI-generated medium on a playback interface.

[0010] According to the above technical features, the present invention can achieve the following effects:

[0011] 1. The use of Application Programming Interface (API) facilitates collaboration between different systems or application programs. Moreover, APIs can be modularized and configured separately according to different capture targets, making the operation more flexible and easier to maintain.

[0012] 2. The generative artificial intelligence server automatically generates one or more AI-generated contents matching the timeline according to the input track contents through artificial intelligence. All the AI-generated contents generated by the generative artificial intelligence server are selectable independent objects. The generative artificial intelligence server can superimpose the selected AI-generated contents on the input track contents, thereby quickly generating a new AI-generated medium that can be widely used in different fields, such as entertainment, education, creative production, and the like.

BRIEF DESCRIPTION OF THE DRAWINGS

[0013] FIG. 1 is a block diagram of a system of one embodiment of the present invention;

[0014] FIG. 2 is a flow chart of a method of the embodiment of the present invention;

[0015] FIG. 3 is a first schematic diagram of the implementation of the embodiment of the present invention, illustrating that a searching and sorting interface has a single search mode and an integrated search mode;

[0016] FIG. 4 is a second schematic diagram of the implementation of the embodiment of the present invention, illustrating that the searching and sorting interface displays a search result;

[0017] FIG. 5 is a third schematic diagram of the implementation of the embodiment of the present invention, illustrating the track separation of an online audio-visual content that contains only audio;

[0018] FIG. 6 is a fourth schematic diagram of the implementation of the embodiment of the present invention, illustrating that the online audio-visual content that contains only audio is mixed with an AI-generated content to output an AI-generated medium;

[0019] FIG. 7 is a fifth schematic diagram of the implementation of the embodiment of the present invention, illustrating the track separation of an online audio-visual content that contains video and audio;

[0020] FIG. 8 is a sixth schematic diagram of the implementation of the embodiment of the present invention, illustrating that the online audio-visual content that contains video and audio is mixed with an AI-generated content to output an AI-generated medium; and

[0021] FIG. 9 is a seventh schematic diagram of the implementation of the embodiment of the present invention, illustrating that the online audio-visual content that contains video and audio is mixed with an AI-generated content to output another AI-generated medium.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0022] Embodiments of the present invention will now be described, by way of example only, with reference to the accompanying drawings.

[0023] The present invention discloses a system (hereinafter referred to as the system) and a method (hereinafter referred to as the method) for acquiring an online audio-visual content, separating tracks, integrating artificial intelligence and outputting an AI-generated medium. FIG. 1 is a block diagram of the system.

[0024] The system is in signal communication with a plurality of streaming platforms A, such as video streaming platforms like YouTube and Netflix, as well as music streaming platforms like Spotify and KKBOX. Each streaming platform A is equipped with an application programming interface server A1.

[0025] The system comprises an audio-visual playback device 1, a sampling module 2, and a generative artificial intelligence server 3.

[0026] The audio-visual playback device 1 includes a device server 11, a playback interface 12, a searching and sorting interface 13, an activation interface 14, and a database 15. The playback interface 12, the searching and sorting interface 13, the activation interface 14 and the database 15 are in signal communication with the device server 11, respectively.

[0027] The device server 11 is in signal communication with the application programming interface server A1 of each streaming platform A. According to the number of the streaming platforms A, a plurality of activation interfaces 14 are provided, corresponding to the streaming platforms A one by one. The database 15 may be a physical hard drive, a storage area network, a cloud hard drive, etc.

[0028] The sampling module 2 is in signal communication with the device server 11. Preferably, the sampling module 2 includes a front-end neural network, an object recognition model,

and an edge detection algorithm. The front-end neural network is one of the following: a multi-source separation model (UMX), an audio separation network (TasNet), a speech separation model (Looking to listen), a sound separation model (Demucs), and a dual-path transformation network (DPTNet). The object recognition model is one of the following: YOLO8, Vision Master, Vision AI, and Amazon Rekognition. The edge detection algorithm is one of the following: Canny edge detection algorithm, Laplacian edge detection algorithm, and Sobel edge detection algorithm.

[0029] The generative artificial intelligence server 3 is in signal communication with the device server 11 via the sampling module 2.

[0030] Referring to FIG. 1 and FIG. 2, the method comprises the following steps:

[0031] First, setting up the system, for example, purchasing the video-visual playback device 1 and connecting the video-visual playback device 1 to the sampling module 2 and the generative artificial intelligence server 3 in the cloud. Alternatively, the application program configured by the system is loaded into the user's existing playback device or computer, so as to use the playback device or the computing unit of the computer as the device server 11 to execute the method.

[0032] Referring to FIG. 3 and FIG. 4, in conjunction with FIG. 1 and FIG. 2, the user can search for a desired online audio-visual content B (referring to FIG. 5) such as a video, audio, music, etc. through a single search mode or an integrated search mode of the searching and sorting interface 13.

[0033] In the single search mode, the user can select different search options corresponding to the streaming platforms A shown in FIG. 3 according to whether he/she wants to play audio-only music or audio-visual content.

[0034] After selecting the single search mode for one of the streaming platforms A, the user enters a keyword that he/she wants to search for. The device server 11 sends an audio-visual search request to the selected application programming interface server A1 based on the keyword using an application programming interface (API).

[0035] The application programming interface server A1 conducts a search according to the audio-visual search request and obtains a search result corresponding to the keyword. The device server 11 obtains a plurality of link codes of the search result from the application programming interface server A1 using the API. Each link code corresponds to a different online audio-visual content B. The searching and sorting interface 13 sorts the link codes and displays the search result

after first excluding the link code corresponding to the online audio-visual content B that is in an invalid state.

[0036] Taking YouTube as an example, after YouTube receives the audio-visual search request, the application programming interface server A1 of YouTube will conduct a search using YouTube's own search engine, and there will be multiple videos or audios in the resulting search result.

[0037] The present invention integrates the URLs, names, creators, and other information of the videos and audios into the link codes, and excludes the link codes that are invalid, such as those with copyright issues, playback permission issues, and violations of the policy of the streaming platform A. Without the need to open the YouTube webpage or the YouTube app interface, the searching and sorting interface 13 can fully display YouTube's search result, and the playback will not be interrupted due to the invalid condition, so that the user's audio and video viewing experience will not be affected.

[0038] In the integrated search mode, for example, the integrated search option shown in FIG. 3, the user enters the keyword to be searched, and the device server 11 sends the audio-visual search request to the application programming interface servers A1 of all the streaming platforms A. The application programming interface servers A1 each conduct the search independently to obtain their own search result.

[0039] The device server 11 can integrate multiple online audio-visual contents B from different streaming platforms A into one data entry if they have the same name and creator, or it can also integrate multiple online audio-visual contents B into one data entry if they have the same creator and the names that are not identical but all contain the specified keyword. In the searching and sorting interface 13, it is indicated that the data has sources from different streaming platforms A. The search results of all the streaming platforms A are displayed in the single searching and sorting interface 13 for the user to select, as shown in FIG. 4.

[0040] The user can click on the corresponding source icon to add the link code of a source to a playback queue. The user can delete, reorder the link code through the option of the playback queue shown in FIG. 3. This is a familiar and intuitive way for operation of the user.

[0041] The device server 11 or the generative artificial intelligence server 3 sequentially obtains the link codes from the playback queue, parses the link codes via the application programming interface (API), and acquires the online audio-visual content B from the application programming

interface server A1, without needing to redirect to the webpage or app of the other streaming platformS A, and without downloading the online audio-visual content B.

[0042] Directly downloading the online audio-visual content B from the streaming platform A, without using the API, is also one of the feasible implementations of the present invention.

[0043] Preferably, the device server 11 or the generative artificial intelligence server 3 first verifies whether the streaming platform A is activated, such as whether the user is logged in or whether a paid subscription mode has been enabled.

[0044] If the streaming platform A is not activated, the device server 11 displays the activation interface 14 corresponding to the streaming platform A in the audio-visual playback device 1 for the user to activate it and then continue to obtain the online audio-visual content B to ensure that the source of the online audio-visual content B is normal.

[0045] Next, the front-end neural network of the sampling module 2 samples and separates a plurality of track contents C1, C2, C3 from the online audio-visual content B according to a frequency feature. As shown in FIG. 5, the track contents C1, C2, C3 include an audio, video, main character image, lyrics text, dialogue text, narration text, etc. The main character image is preferably captured by the object recognition model and the edge detection algorithm.

[0046] The generative artificial intelligence server 3 randomly selects or selects part or all of the track contents C1, C2, C3 according to the user's instructions. The selected track contents C1, C2, C3 are aligned and positioned on a timeline. For example, all the track contents C1, C2, C3 share the same playback start time at 0 seconds in the sampling process. The playback start time is marked as the starting point of each track content C1, C2, C3 on the timeline. Then, the starting points of the selected track contents C1, C2, C3 are aligned to complete the positioning of the track contents C1, C2, C3.

[0047] The generative artificial intelligence server 3 generates at least one AI-generated content D1, D2, D3 according to the selected track contents C1, C2, C3. As shown in FIG. 6, the AI-generated contents D1, D2, D3 are aligned with the track contents C1, C2, C3 and positioned on the timeline. The AI-generated contents D1, D2, D3 include AI-generated audios, AI-generated static images, AI-generated dynamic animations, AI-generated texts, etc. The generative artificial intelligence server 3 can perform an additional change, such as a change in style, tempo, melody, volume and pitch, to part of the selected track contents C1, C2, C3 automatically or according to a

user's instructions, in order to meet the needs of different users for a wide range of purposes, such as entertainment, education, and creativity.

[0048] Finally, the generative artificial intelligence server 3 mixes the selected track contents C1, C2, C3 with the AI-generated contents D1, D2, D3, outputs an AI-generated medium E and transmits it back to the device server 11. As shown in FIG. 6, the device server 11 plays the AI-generated medium E on the playback interface 12 for user viewing.

[0049] Referring to FIG. 5 and FIG. 6, in conjunction with FIG. 1 and FIG. 2, in one implementation, the online audio-visual content B contains only audio, and the track contents C1, C2, C3 after separation also contain only audio, such as drum sounds, violin sounds, and piano sounds. The generative artificial intelligence server 3 can generate AI-generated images, AI-generated animations, AI-generated texts, etc. to form the AI-generated contents D1, D2, D3, such as static bay buildings, continuously changing sailboats, etc. The generative artificial intelligence server 3 mixes the track contents C1, C2, C3 with the AI-generated contents D1, D2, D3 to output the AI-generated medium E.

[0050] Referring to FIG. 7 through FIG. 9, in conjunction with FIG. 1 and FIG. 2, in another implementation, the online audio-visual content B' contains both audio and video, and the track contents C1', C2', C3', C4' after separation contain an image, such as a main character image, in addition to the audio. The generative artificial intelligence server 3 can generate the AI-generated contents D1', D2', and mix the track contents C1', C2', C3', C4' with the AI-generated contents D1', D2' to output the AI-generated medium E1'.

[0051] Or, only the selected track contents C1', C2', C3' are mixed with the AI-generated contents D3', D4', D5', such as AI-generated images, AI-generated animations, AI-generated texts, etc., to output a different AI-generated medium E2', as shown in FIG. 9.

[0052] Or, only part of the audio and part of the video of the track contents are selected and combined with an AI-generated text. Or, only part of the images of the track contents is selected and combined with an AI-generated video and AI-generated animation. Or, instead of selecting the sailboat from the main character image, the waves from the images of the track contents are selected and combined with the AI-generated audio and other content. The methods of selection and combination are not restricted.

[0053] Referring to FIG. 6, in conjunction with FIG. 1 and FIG. 2, more application scenarios of the present invention are as follow.

[0054] The user, being a violinist, can separate the orchestra music from Spotify into individual tracks such as violin track, piano track, and trumpet track. After excluding the violin track, the remaining separated track contents C1, C2, C3 are combined with music score and preferred AI-generated animations to produce the generated medium E. This allows the user to practice playing the violin while the generated medium E is played, helping to better master the techniques required for the piece.

[0055] Or, the user may record a company meeting and upload it to YouTube to separate the voices of different colleagues into individual tracks. The voice of a colleague who speaks relatively quietly is amplified, and then combined with the conversation content and translation by the generative artificial intelligence server 3, the AI-generated medium E is output as a multi-language meeting record.

[0056] For the link codes that have been acquired a threshold number of times, the device server 11 may store the link codes in the database 15 for quick acquisition by the device server 11 or the generative artificial intelligence server 3 the next time the user wants to play them again. The device server 11 can count the link codes commonly used by users of different regions, different ages, and different genders, and then make customized recommendations for a particular user based on region, age, gender, etc.

[0057] Although particular embodiments of the present invention have been described in detail for purposes of illustration, various modifications and enhancements may be made without departing from the spirit and scope of the present invention. Accordingly, the present invention is not to be limited except as by the appended claims.

WHAT IS CLAIMED IS:

1. A method for acquiring an online audio-visual content, separating tracks, integrating artificial intelligence and outputting an AI-generated medium, comprising the following steps:

a device server (11) of an audio-visual playback device (1) or a generative artificial intelligence server (3) that is in signal communication with the device server (11) acquiring an online audio-visual content (B, B') from a streaming platform (A);

the device server (11) or the generative artificial intelligence server (3) separating a plurality track contents (C1, C2, C3, C1', C2', C3', C4') from the online audio-visual content (B, B') through a sampling module (2);

the generative artificial intelligence server (3) selecting part or all of the track contents (C1, C2, C3, C1', C2', C3', C4'), the selected track contents (C1, C2, C3, C1', C2', C3', C4') being aligned and positioned on a timeline, the generative artificial intelligence server (3) generating at least one AI-generated content (D1, D2, D3, D1', D2', D3', D4', D5') according to the selected track contents (C1, C2, C3, C1', C2', C3', C4'), the at least one AI-generated content (D1, D2, D3, D1', D2', D3', D4', D5') being aligned with the track contents (C1, C2, C3, C1', C2', C3', C4') and positioned on the timeline;

the generative artificial intelligence server (3) mixing the selected track contents (C1, C2, C3, C1', C2', C3', C4') with the at least one AI-generated content (D1, D2, D3, D1', D2', D3', D4', D5'), outputting an AI-generated medium (E, E', E2') and transmitting it back to the device server (11);

the device server (11) playing the AI-generated medium (E, E', E2') on a playback interface (12).

2. The method as claimed in claim 1, wherein the audio-visual playback device (1) includes a searching and sorting interface (13) that is in signal communication with the device server (11); the device server (11) sends an audio-visual search request to an application programming interface server (A1) of the streaming platform (A) according to a keyword using an application programming interface (API); after the application programming interface server (A1) conducts a search according to the audio-visual search request and obtains a search result corresponding to the keyword, the device server (11) obtains a plurality of link codes of the search result from the application programming interface server (A1) using the application programming interface (API), each of the link codes corresponds to the respective online audio-visual content (B, B'); the device server (11) obtains a corresponding one of the link codes, the device server (11) or the generative artificial intelligence server (3) parses the corresponding link code and acquires the corresponding online audio-visual

content (B, B') from the application programming interface server (A1); the searching and sorting interface (13) sorts the link codes in the audio-visual playback device (1) and displays the search result.

3. The method as claimed in claim 2, wherein the audio-visual playback device (1) further includes a database (15) that is in signal communication with the device server (11); after the searching and sorting interface (13) displays the search result, a user selects at least one of the link codes, the searching and sorting interface (13) adds the selected link code to a playback queue, the device server (11) or the generative artificial intelligence server (3) sequentially obtains the link codes from the playback queue; when the number of times that one of the link codes is obtained reaches a threshold, the device server (11) stores the one of the link codes in the database (15).

4. The method as claimed in claim 2, wherein the streaming platform (A) includes a plurality of streaming platforms (A) each having the application programming interface server (A1) that is in signal communication with the device server (11), the searching and sorting interface (13) has a single search mode and an integrated search mode; when the searching and sorting interface (13) is operated in the single search mode, the device server (11) sends the audio-visual search request to the application programming interface server (A1) of one of the streaming platforms (A), the audio-visual playback device (1) displays the search result of the one of the streaming platforms (A); when the searching and sorting interface (13) is operated in the integrated search mode, the device server (11) sends the audio-visual search request to the application programming interface servers (A1) of all the streaming platforms (A), and the device server (11) integrates and displays the link codes from the streaming platforms that correspond to the same online audio-visual content (B, B').

5. The method as claimed in claim 2, wherein before the searching and sorting interface (13) displays the search result, the device server (11) first excludes the link code corresponding to the online audio-visual content (B, B') that is in an invalid state, and the device server (11) or the generative artificial intelligence server (3) first verifies whether the streaming platform (A) is activated before the streaming platform (A) acquires the online audio-visual content (B, B').

6. The method as claimed in claim 1, wherein before outputting the AI-generated medium, the generative artificial intelligence server (3) first performs an additional change to at least one part of the selected track contents (C1, C2, C3, C1', C2', C3', C4'), and the additional change includes one of style, tempo, melody, volume and pitch.

7. The method as claimed in claim 1, wherein the sampling module (2) includes a front-end neural network, an object recognition model and an edge detection algorithm, the front-end neural

network samples the track contents (C1, C2, C3, C1', C2', C3', C4') from the online audio-visual content (B, B') according to a frequency feature, and a main character image is captured by the object recognition model and the edge detection algorithm.

8. A system for acquiring an online audio-visual content, separating tracks, integrating artificial intelligence and outputting an AI-generated medium, being in signal communication with a streaming platform, comprising:

an audio-visual playback device (1), having a device server (11) and a playback interface (12) that is in signal communication with the device server (11);

a sampling module (2), being in signal communication with the device server (11); and

a generative artificial intelligence server (3), being in signal communication with the device server (11) and the sampling module (2);

wherein the device server (11) or the generative artificial intelligence server (3) acquires an online audio-visual content (B, B') from the streaming platform (A), and then the device server (11) or the generative artificial intelligence server (3) separates a plurality track contents (C1, C2, C3, C1', C2', C3', C4') from the online audio-visual content (B, B') through the sampling module (2);

wherein the generative artificial intelligence server (3) selects part or all of the track contents (C1, C2, C3, C1', C2', C3', C4'), the selected track contents (C1, C2, C3, C1', C2', C3', C4') are aligned and positioned on a timeline, the generative artificial intelligence server (3) generates at least one AI-generated content (D1, D2, D3, D1', D2', D3', D4', D5') according to the selected track contents (C1, C2, C3, C1', C2', C3', C4'), and the at least one AI-generated content (D1, D2, D3, D1', D2', D3', D4', D5') is aligned with the track contents (C1, C2, C3, C1', C2', C3', C4') and positioned on the timeline;

wherein the generative artificial intelligence server (3) mixes the selected track contents (C1, C2, C3, C1', C2', C3', C4') with the at least one AI-generated content (D1, D2, D3, D1', D2', D3', D4', D5'), outputs an AI-generated medium (E, E', E2') and transmits it back to the device server (11);

wherein the device server (11) plays the AI-generated medium (E, E', E2') on a playback interface (12).

9. The system as claimed in claim 8, wherein the streaming platform (A) has an application programming interface server (A1), the device server (11) is in signal communication with the application programming interface server (A1), the audio-visual playback device (1) includes a searching and sorting interface (13) that is in signal communication with the device server (11); the

device server (11) sends an audio-visual search request to the application programming interface server (A1) according to a keyword using an application programming interface (API); after the application programming interface server (A1) conducts a search and obtains a search result corresponding to the keyword, the device server (11) obtains a plurality of link codes corresponding to the search result from the application programming interface server (A1) using the application programming interface (API), each of the link codes corresponds to the respective online audio-visual content (B, B'); the device server (11) obtains a corresponding one of the link codes, the device server (11) or the generative artificial intelligence server (3) parses the corresponding link code using the application programming interface (API) and then acquires the corresponding online audio-visual content (B, B') from the application programming interface server (A1) using the application programming interface (API); the searching and sorting interface (13) sorts the link codes in the audio-visual playback device (1) and displays the search result.

10. The system as claimed in claim 9, wherein the audio-visual playback device (1) further includes a database (15) that is in signal communication with the device server (11); after the searching and sorting interface (13) displays the search result, a user selects at least one of the link codes, the searching and sorting interface (13) adds the selected link code to a playback queue, the device server (11) or the generative artificial intelligence server (3) sequentially obtains the link codes from the playback queue; when the number of times that one of the link codes is obtained reaches a threshold, the device server (11) stores the one of the link codes in the database (15).

11. The system as claimed in claim 9, wherein the streaming platform (A) includes a plurality of streaming platforms (A) each having the application programming interface server (A1) that is in signal communication with the device server (11), the searching and sorting interface (13) has a single search mode and an integrated search mode; when the searching and sorting interface (13) is operated in the single search mode, the device server (11) sends the audio-visual search request to the application programming interface server (A1) of one of the streaming platforms (A), the audio-visual playback device (1) displays the search result of the one of the streaming platforms (A); when the searching and sorting interface (13) is operated in the integrated search mode, the device server (11) sends the audio-visual search request to the application programming interface servers (A1) of all the streaming platforms (A), and the device server (11) integrates and displays the link codes from the streaming platforms that correspond to the same online audio-visual content (B, B').

12. The method as claimed in claim 9, wherein before the searching and sorting interface (13) displays the search result, the device server (11) first excludes the link code corresponding to the online audio-visual content (B, B') that is in an invalid state, and the device server (11) or the

generative artificial intelligence server (3) first verifies whether the streaming platform (A) is activated before the streaming platform (A) acquires the online audio-visual content (B, B').

13. The method as claimed in claim 8, wherein before outputting the AI-generated medium, the generative artificial intelligence server (3) first performs an additional change to at least one part of the selected track contents (C1, C2, C3, C1', C2', C3', C4'), and the additional change includes one of style, tempo, melody, volume and pitch.

14. The system as claimed in claim 8, wherein the sampling module (2) includes a front-end neural network, an object recognition model and an edge detection algorithm, the front-end neural network samples the track contents (C1, C2, C3, C1', C2', C3', C4') from the online audio-visual content (B, B') according to a frequency feature, and a main character image is captured by the object recognition model and the edge detection algorithm.