

METHOD, APPARATUS, DEVICE AND STORAGE MEDIUM FOR TIMBRE CREATION AND USE

[0001] The present application claims priority to Chinese Patent Application No. 202410233361.4, filed on February 29, 2024, and entitled "METHOD AND APPARATUS, DEVICE AND STORAGE MEDIUM FOR TIMBRE CREATION AND USE ", which is hereby incorporated by reference in its entirety.

FIELD

[0002] Example embodiments of the present disclosure generally relate to the field of computers, and in particular, to a method, an apparatus, a device, and a computer-readable storage medium for timbre creation and use.

BACKGROUND

[0003] Currently, more and more applications are designed to provide various services for users. Many applications can support users to exchange messages. In the process of people exchanging information over the Internet, various types of audio have become an important medium for people to express socially and exchange information. Therefore, it is expected to generate audio with better timbres.

SUMMARY

[0004] In a first aspect of the present disclosure, a method for timbre creation and use is provided. The method includes: presenting a timbre creation page in response to a timbre creation request, a reference text and a recording control being displayed in the timbre creation page displaying; receiving an audio input by a user in response to detecting a trigger on the recording control; creating a target timbre associated with the user based on the received audio and the reference text; presenting a timbre confirmation page, the timbre confirmation page at least including a preview audio generated based on the target timbre; and storing in response to receiving a confirmation of the target timbre in the timbre confirmation page, the target timbre associated with the user for use in an audio.

[0005] In a second aspect of the present disclosure, an apparatus for timbre creation and use is provided. The apparatus includes: a creation page presenting module configured to present a timbre creation page in response to a timbre creation request, a reference text and a recording control being displayed in the timbre creation page; an audio receiving module configured to receive an audio input by a user in response to detecting a trigger on the recording control; a timbre creation module configured to create a target timbre associated with the user based on the received audio and the reference text; a confirmation page presenting module configured to present a timbre confirmation page, the timbre confirmation page at least including a preview audio generated based on the target timbre; and a timbre storing module configured to store, in response to receiving a confirmation of the target timbre in the timbre confirmation page, the target timbre associated with the user for use in an audio.

[0006] In a third aspect of the present disclosure, an electronic device is provided. The device includes at least one processing unit; and at least one memory, the at least one memory being coupled to the at least one processing

unit and storing instructions executable by the at least one processing unit, wherein the instructions, when executed by the at least one processing unit, cause the device to perform the method of the first aspect.

[0007] In a fourth aspect of the present disclosure, a computer-readable storage medium is provided. The computer-readable storage medium has a computer program stored thereon, wherein the computer program is executable by a processor to implement the method of the first aspect.

[0008] In a fifth aspect of the present disclosure, a computer program product is provided. The computer program product is tangibly stored in a computer storage medium and includes computer-executable instructions, wherein the computer-executable instructions, when executed by a device, cause the device to perform the method of the first aspect.

[0009] It should be understood that the content described in this Summary is not intended to identify key or essential features of the embodiments of the present disclosure, nor is it intended to limit the scope of the present disclosure. Other features of the present disclosure will be readily envisaged through the following description.

BRIEF DESCRIPTION OF THE DRAWINGS

[0010] 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:

[0011] FIG. 1 shows a schematic diagram of an example environment in which the embodiments of the present disclosure can be implemented;

[0012] FIG. 2 shows a schematic diagram of an example architecture for timbre creation according to some embodiments of the present disclosure;

[0013] FIGS. 3A to 3L show schematic diagrams of example interfaces for timbre creation according to some embodiments of the present disclosure;

[0014] FIGS. 4A to 4C show schematic diagrams of example interfaces for using a target timbre according to some embodiments of the present disclosure;

[0015] FIGS. 5A to 5J show schematic diagrams of example interfaces for a user to discover a timbre creation entry during content browsing according to some embodiments of the present disclosure;

[0016] FIG. 6 shows a flowchart of a process for timbre creation and use according to some embodiments of the present disclosure;

[0017] FIG. 7 shows a flowchart of a process for timbre creation and use according to some embodiments of the present disclosure;

[0018] FIG. 8 shows a block diagram of an apparatus for timbre creation and use according to some embodiments of the present disclosure; and

[0019] FIG. 9 shows a block diagram of an electronic device capable of implementing one or more embodiments of the present disclosure.

DETAILED DESCRIPTION OF EMBODIMENTS

[0020] The embodiments of the present disclosure will be described in more detail below with reference to the drawings. Although certain 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 for illustrative purposes, and are not intended to limit the protection scope of the present disclosure.

[0021] In the description of the embodiments of the present disclosure, the term "include/comprise" and its 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 term "some embodiments" should be understood as "at least some embodiments". The following may also include other explicit and implicit definitions.

[0022] Here, unless explicitly stated, "in response to A" to perform a step does not mean that the step is performed immediately after "A", but may include one or more intermediate steps.

[0023] 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) should comply with requirements of corresponding laws, regulations and related provisions.

[0024] 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, use scope, use scenarios, etc., of personal information involved in the present disclosure and the authorization of the user shall be obtained in an appropriate manner in accordance with relevant laws and regulations.

[0025] For example, in response to receiving an active request from a user, 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, so that 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.

[0026] 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.

[0027] It may be understood that the above process of notifying and obtaining user authorization is only illustrative, and does not limit the implementation of the present disclosure. Other manners that satisfy relevant laws and regulations may also be applied to the implementation of the present disclosure.

[0028] As used herein, the term "model" may learn a correlation between corresponding inputs and outputs from training data, so that a corresponding output may be generated for a given input after the training is

completed. The generation of the model may be based on machine learning techniques. Deep learning is a machine learning algorithm that uses multiple layers of processing units to process inputs and provide corresponding outputs. A neural network model is an example of a model based on deep learning. Here, a "model" may also be referred to as a "machine learning model", a "learning model", a "machine learning network" or a "learning network", which terms are used interchangeably herein.

[0029] As briefly described above, in the process of people exchanging information over the Internet, various types of audio have become an important medium for people to express socially and exchange information. At present, the cost of timbre creation is relatively high. For example, if a user wants to copy his/her own timbre, he/she often needs to record voice data for a whole day, and it takes a long time to extract the timbre to copy the voice, and the effect is average, and the pronunciation is not natural enough.

[0030] This method is time-consuming and poor in user experience, and due to the high cost, it is difficult to achieve that everyone can really generate his/her own voice based on timbres.

[0031] In view of this, the embodiments of the present disclosure provide an improved method for timbre creation and use. In this solution, a timbre creation page is presented in response to a timbre creation request, a reference text and a recording control being displayed the timbre creation page. If a trigger on the recording control is detected, an audio input by a user is received. Then, a target timbre associated with the user is created based on the received audio and the reference text. Correspondingly, a timbre confirmation page is presented, the timbre confirmation page at least including a preview audio generated based on the target timbre. Further, in response to receiving a confirmation of the target timbre in the timbre confirmation page, the target timbre associated with the user is stored for use in an audio. In this way, the user can conveniently and quickly create his/her own timbre for generating personalized audio, which enhances the user experience. In the timbre creation process, the user can better understand the timbre creation method. In addition, by requiring the user to extract timbres by reading the reference text instead of uploading free audio, the problem of timbre imitation caused by audio imitation may be largely avoided.

[0032] The term "work " in the present disclosure refers to any type of media content or media work, which includes one or more types of content, including but not limited to audio files, video files, picture files, text files, etc. Specifically, the work may be a short video, music, a picture, a picture collection, a multimedia clip, an audio-visual material, etc. The present disclosure is not limited in this regard.

[0033] FIG. 1 shows a schematic diagram of an example environment 100 in which the embodiments of the present disclosure can be implemented. The environment 100 includes one or more users 110-1, 110-2, 110-3, ..., 110-N, who can send and receive messages through their associated terminal devices 120-1, 120-2, 120-3, ..., 120-N. For the convenience of discussion, the users 110-1, 110-2, 110-3, ..., 110-N may be collectively or individually referred to as a user 110, and the terminal devices 120-1, 120-2, 120-3, ..., 120-N may be collectively or individually referred to as a terminal device 120. In some scenarios, the user 110 may post and comment on works in a target platform through the associated terminal device 120. In some scenarios, the user 110 is also referred to as a poster of the work.

[0034] The terminal device 120 may be installed with an application 125 that supports message exchange (that is, the terminal device 120-1 is installed with the application 125-1, the terminal device 120-2 is installed with the application 125-2, the terminal device 120-3 is installed with the application 125-3, ..., the terminal device 120-N is installed with the application 125-N). It should be noted that the applications 125 installed in different terminal devices 120 may be completely identical applications or different applications (for example, different versions). The application 125 may be any appropriate application with a message sending and receiving function, which may be, for example, a dedicated chat application, a social application, a content sharing application, an office support application, etc.

[0035] In the environment 100 of FIG. 1, if the application 125 is active, the terminal device 120 may present the user interface of the application 125. This user interface may include various interfaces that may be provided by the application 125, such as a user interface supporting message exchange, a user interface supporting content browsing, a message sending and receiving interface, etc. Via different user interfaces, the application 125 may provide different content to the user 110. Via an appropriate manner, such as clicking or selecting any appropriate element in the user interface, the application 125 may also provide the user 110 with a choice and switch of presentation manner of associated content.

[0036] In some embodiments, different terminal devices 120 may also communicate with a server 130 through a network 132 to provide message exchange services. The server 130 may provide functions such as management, configuration and maintenance of the application 125.

[0037] The terminal device 120 may be any type of mobile terminal, fixed terminal or portable terminal, including a mobile phone, a desktop computer, a laptop computer, a notebook computer, a netbook computer, a tablet computer, a media computer, a multimedia tablet, a personal communication system (PCS) device, a personal navigation device, a personal digital assistant (PDA), an audio/video player, a digital camera/video camera, a television receiver, a radio broadcast receiver, an e-book device, a game device, or any combination of the foregoing, including accessories and peripherals of these devices or any combination thereof. In some embodiments, the terminal device 120 may also support any type of user-specific interface (such as a "wearable" circuit, etc.). The server 130 may be various types of computing systems/servers that may provide computing power, including but not limited to mainframes, edge computing nodes, computing devices in cloud environments, etc.

[0038] It should be understood that the structure and function of each element in the environment 100 are described for illustrative purposes only, and do not imply any limitation on the scope of the present disclosure.

[0039] Various example implementations of the present disclosure will be described in detail below.

[0040] FIG. 2 shows a schematic diagram of an example architecture 200 for timbre creation according to some embodiments of the present disclosure. As shown in FIG. 2, the architecture 200 involves the terminal device 120, the server 130, and a timbre extraction model 210. The timbre extraction model 210 is configured to extract timbre feature information in an audio based on an input audio for reuse in a new audio. The timbre extraction model 210 may be implemented based on any appropriate model, and the embodiments of the present disclosure are not limited in this regard.

[0041] The timbre extraction model 210 may run locally on the terminal device 120 or be deployed remotely, for example, on a server (such as the server 130), in a cloud environment, etc. In the case of local operation, the terminal device 120 may directly provide the input of the model to the timbre extraction model 210 installed locally, and obtain the model output generated by the timbre extraction model 210. In the case of remote operation, the terminal device 120 provides data to other electronic devices through a communication connection with the other electronic devices. The other electronic devices determine the model input based on the obtained data and provide this model input to the timbre extraction model 210. After obtaining the model output of the timbre extraction model 210, the other electronic devices provide the model output to the terminal device 120 through the communication connection with the terminal device 120, so that the terminal device 120 may obtain the model output.

[0042] For the convenience of description, the following uses an example in which the timbre extraction model 210 runs on the server 130 for description, that is, the terminal device 120 may provide the model input to the server 130, so that the server 130 may call the timbre extraction model 210 to perform timbre extraction. It should be understood that, in the following, for the convenience of discussion, various operations involved in timbre creation and use are described from the perspective of the terminal device, but it should be understood that some or all of these operations may also be performed on the server side and the execution results thereof may be sent to the terminal device, or the terminal device may be controlled by the server to perform these operations.

[0043] In the case where the application 125 is active, the terminal device 120 may present the user interface (for example, a timbre creation page) provided by the application 125. In the embodiments of the present disclosure, the terminal device 120 presents the timbre creation page in response to a timbre creation request. A reference text and a recording control is displayed in the timbre creation page displays. In the application 125, an appropriate interface and a timbre creation entry may be configured for the user 110 to initiate the timbre creation request. Some examples of the user 110 initiating the timbre creation request will be given below.

[0044] In the case where the timbre creation page is presented to the user, if the user 110 clicks on the recording control displayed on the timbre creation page by the terminal device 120, the user 110 begins to record. It should be understood that before or after the user clicks on the recording control, the user is informed of the purpose of obtaining the recording of the user 110, and the user 110 will begin to record only after giving consent. In the timbre creation page, prompt information may be provided to prompt the user to record audio by reading the reference text displayed in the page. Further, the terminal device 120 receives an audio 202 input by the user. The terminal device 120 creates a target timbre 212 associated with the user 110 based on the audio 202 and the reference text.

[0045] In some embodiments, the target timbre 212 may be generated using the timbre extraction model 210 based on the audio 202 and the reference text. In some embodiments, the terminal device 120 may send the audio 202 and the reference text to the server 130, and the server 130 may generate the target timbre 212. The terminal device 120 may send the audio 202 and the reference text to the server 130, and the server 130 may use the timbre extraction model 210 to generate the target timbre 212.

[0046] The following continues to describe example embodiments of target timbre creation with reference to the example interfaces/example pages shown in FIGS. 3A to 3L.

[0047] Only for the purpose of better understanding the embodiments of the present disclosure, the example user interfaces 301 to 309 shown in FIGS. 3A to 3L are referred to in the following example embodiments. FIGS. 3A to 3I show schematic diagrams of example interfaces for timbre creation according to some embodiments of the present disclosure. It should be understood that the user interfaces shown in FIGS. 3A to 3L and other drawings below are only examples, and various designs may actually exist. For example, each graphic element and/or control in the user interface may have different arrangements and different visual representations, one or more of the elements and/or controls may be omitted or replaced, and there may also be one or more other elements and/or controls. In addition, the user interface may contain any appropriate content. The scope of the present disclosure is not limited in this regard.

[0048] In the example embodiments of the present disclosure, the example user interfaces 301 to 309 may be implemented at the terminal device 120. For the convenience of discussion, the example user interfaces 301 to 309 will be described with reference to the environment 100 of FIG. 1 and/or FIG. 2.

[0049] In some embodiments, the terminal device 120 presents the timbre creation page in response to the timbre creation request. the reference text and the recording control is displayed in the timbre creation page . In some examples, the reference text displayed on the timbre creation page may be the text for recording audio.

[0050] The examples shown in FIGS. 3A to 3L involve a user initiating a timbre creation request in the editing page of media content. It is assumed that the user is editing media content with the expectation of posting the content. As shown in the example interfaces 301 to 304 in FIGS. 3A to 3D, after the user 110 clicks on the "text" control 311 presented on the interface 301 by the terminal device 120, the terminal device 120 presents a text input box 321 on the interface 302. The user 110 may input text (for example, copy A) in the text input box 321 presented by the terminal device 120. Subsequently, the terminal device 120 presents the user input text 330 (for example, copy A) on the interface 303. In the media content editing process, assuming that the user expects to perform text-to-speech on the input text, the user may trigger the "text-to-speech" function 322. In the interface 303, a timbre selection panel 334 may be provided. If the user 110 selects the "first timbre" 331, the terminal device 120 will read the "copy A" in the "first timbre" 331. It may be understood that the timbres presented to the user 110 are existing timbres in the timbre library or timbres that are authorized to be used.

[0051] In some embodiments, the terminal device 120 may present an entry 332 for creating timbres on the interface 303 (for example, the interface may be a corresponding interface of the timbre selection panel 334). If the user 110 clicks on the entry 332 for creating timbres, the terminal device 120 will detect the timbre creation request and present the function introduction interface 340 about timbre creation on the interface 304. The function introduction interface 340 may provide some prompt information about timbre creation. For example, on the interface 140, the terminal device 120 will display the title: "Create your own timbre", relevant descriptions: "Record a few sentences out loud to create a timbre that may be used with reference text in XX", "You may set your timbre as private and delete it at any time", and "Recording only takes 10 seconds. Make

sure you are in a quiet environment and avoid using wireless headsets". Certainly, these information are only examples.

[0052] After the function introduction interface, the terminal device 120 may display the timbre creation page, the interface 305 shown in FIG. 3E. The timbre creation page may provide the recording control and the reference text. In some embodiments, the terminal device 120 receives the audio input by the user in response to detecting a trigger on the recording control by the user. The terminal device 120 creates the target timbre associated with the user based on the audio and the reference text.

[0053] As shown in the example interfaces 305 to 313 in FIGS. 3E to 3M, the terminal device 120 presents the reference text 351 (for example, copy B) and the recording control 350 on the interface 305. The terminal device 120 starts recording based on the user 110 clicking on the recording control 350 to receive the audio of the user 110. In some examples, the reference text may be pre-configured. In some embodiments, a reference text library may be set up, and each time a user requests to create a timbre, a reference text may be randomly selected from the reference text library and displayed to the user. In some embodiments, the length of the reference text may be set to allow the obtained audio to be sufficient for the timbre extraction model 210 to correctly extract timbres. Depending on the capability of the timbre extraction model 210, when requiring the user to record audio, only one sentence or several sentences need to be recorded, and the recording is performed with a countdown of a certain number of seconds. In this way, the user can create timbres easily and quickly.

[0054] In order to guide the user to use the timbre creation function, the timbre creation page may also provide prompt information to instruct the user to record by reading the reference text displayed on the page and guide the user to use the recording control, as shown in FIG. 3E. After the user 110 clicks on the recording control 350, the user 110 begins to read the copy B. At this time, the terminal device 120 starts to receive the corresponding audio of the user 110 reading the copy B. If the user 110 clicks on the "stop" control 360 in FIG. 3F, the terminal device 120 will stop receiving the audio input by the user 110. In some embodiments, the terminal device 120 creates the target timbre of the user 110 based on the received audio and the reference text. In some embodiments, considering that timbre creation takes a certain amount of time, a progress indication of timbre generation may be provided. As shown in FIG. 3G, the terminal device 120 presents a corresponding "generating" control 370 on the interface 307 to indicate the progress of timbre generation.

[0055] In some embodiments, after the target timbre is extracted based on the audio input by the user, the terminal device 120 also presents the timbre confirmation page. The timbre confirmation page at least includes the preview audio generated based on the target timbre, so that the user may determine whether the generated timbre is satisfactory and whether further adjustment or re-generation is required by listening to the preview audio. As shown in FIG. 3H, the terminal device 120 provides the preview control 381 for the preview audio on the timbre confirmation page 308 and displays the corresponding indication 380 of "Your voice is ready". The terminal device 120 plays the preview audio based on the user 110 (for example, user A) clicking on the preview control 381. At this time, the audio has the extracted timbre. In some embodiments, if timbre creation is performed during the media content creation process and there is user input text, the preview audio may be generated based on the user input text. For example, in the example shown in FIG. 3H, the preview audio is the

audio generated after performing the text-to-speech operation on the user input text 330 (that is, "copy A"), and the extracted timbre is applied to this audio.

[0056] In some embodiments, in the absence of the user input text associated with the timbre creation request, the terminal device 120 presents the timbre confirmation page 309 at least including the preview audio generated based on the target timbre. The preview audio includes the audio with the target timbre that is generated by performing the text-to-speech operation on the predetermined text.

[0057] As shown in the timbre confirmation page 309 in FIG. 3I, after the terminal device 120 finishes loading the target timbre, the timbre of the user 110 (for example, user A) is automatically applied to the media content currently being created. Accordingly, the terminal device 120 provides the preview control 381 for the preview audio on the timbre confirmation page 309 and displays the corresponding indication of "Your voice is ready". The terminal device 120 plays the corresponding audio of the user 110 (for example, user A) reading the "predetermined text" 390 based on the user 110 (for example, user A) clicking on the preview control 381.

[0058] The above describes the case where the terminal device 120 generates the audio with the target timbre in the absence of the reference text input by the user. The following will continue to describe re-recording the audio to generate the audio with the target timbre with reference to Fig. 3I.

[0059] In some embodiments, after the terminal device 120 finishes loading the target timbre, the timbre of the user 110 (for example, user A) is automatically applied to the media content currently being created. As shown in FIG. 3H, with the consent of the user 110, the generated target timbre is identified as the "timbre of user A" 388, which is applied to the media content that the user 110 is currently editing.

[0060] The terminal device 120 stores the target timbre associated with the user for use in the audio in response to receiving the confirmation of the target timbre in the timbre confirmation page. As shown in FIG 3H, if the user 110 confirms the timbre currently generated, the generated timbre may be confirmed by triggering the "use" control 382. In some examples, if the user 110 confirms the target timbre in the timbre confirmation page, the terminal device 120 stores the target timbre for use in the audio. For example, after the user 110 confirms the target timbre, the terminal device 120 may directly apply the timbre to the media content currently being created. The terminal device 120 may also store the timbre for the user to use directly next time. In some other examples, if the user 110 (for example, user A) actively sets the timbre as a public timbre, that is, the user 110 authorizes his timbre to be used, another user B may directly use the timbre.

[0061] In this way, the cost of timbre creation for users can be reduced, and each user may be supported to freely create and use his own timbre according to his needs.

[0062] In some embodiments, the terminal device 120 displays another reference text and the recording control in the timbre creation page in response to detecting a re-generation request in the timbre creation page. For example, if the user 110 is not satisfied with the audio with the target timbre that is currently generated, the user may click on the "restart" control 383 to re-record the audio.

[0063] In some other embodiments, the terminal device 120 prompts the user to re-record the audio in response to detecting that the received audio does not match the reference text. Subsequently, the terminal device 120 displays another reference text and the recording control in the timbre creation page.

[0064] In some examples, if the terminal device 120 detects that the audio input by the user 110 does not match the reference text, the user is prompted to re-record the audio. After entering the audio recording page, the terminal device 120 automatically switches the recording copy.

[0065] By requiring the user to create the timbre by reading a specific text, the present disclosure can ensure to avoid the situation of timbre imitation or forgery. The following describes re-recording the audio in the case where an error occurs in the process of generating the target timbre with reference to the example interface 3010 shown in FIG. 3J.

[0066] In some embodiments, the terminal device 120 prompts the user to re-record the audio in response to detecting that the background noise in the audio exceeds a noise threshold or the duration of the audio is below a preset threshold. Then, the terminal device 120 displays another reference text and the recording control in the timbre creation page. In some examples, when the target timbre is not successfully generated, or when the close button is clicked in the recording process, the terminal device 120 will exit the recording page without saving the recorded data and timbre, and a prompt message will be popped up.

[0067] As shown in the example interface 3010 in FIG. 3J, when the terminal device 120 detects an error in the process of generating the target timbre, the prompt message 3010-1 of "Error! Please try again" will be presented on the interface 3010. When the user re-enters the recording page, the terminal device 120 will present another reference text 3010-2 (for example, copy C) and the recording control 3010-3 on the interface 3010.

[0068] For example, when detecting that the background noise exceeds the default value in the process of generating the target timbre, the terminal device 120 will present the prompt message of "Noise environment detected, please try again" on the interface 3010.

[0069] For another example, when detecting that the recording duration is less than 3 seconds in the process of generating the timbre, the terminal device 120 will present the prompt message of "Please record at least 3 seconds to ensure the best results" on the interface 3010.

[0070] In some examples, in the process of generating the target timbre, the terminal device 120 detects an error in automatic speech recognition (ASR), and the prompt message "There is a pronunciation problem with this recording. Please try again" will be presented on the interface 3010.

[0071] In some other examples, in the process of generating the target timbre, the terminal device 120 detects a network error, and the prompt message "Please ensure the network connection. Please try again" will be presented on the interface 3010. In the process of generating the target timbre, the terminal device 120 detects that the service connection fails, and the prompt message "This service is currently unavailable. Please try again in a few minutes" will be presented on the interface 3010.

[0072] The above describes the case where the terminal device 120 re-records the audio in the case of an error occurring in the process of generating the target timbre. The following describes that the terminal device 120 applies the target timbre to the edited work after the target timbre is created with reference to FIG. 3I and FIGS. 3K to 3L.

[0073] In some embodiments, the terminal device 120 detects the timbre creation request from the user in the media content editing page. In some embodiments, the terminal device 120 jumps to the editing page in response

to receiving the confirmation of the target timbre in the timbre confirmation page. The target timbre is selected to be applied to the audio of the media item being edited.

[0074] In some examples, the terminal device 120 creates the target timbre after detecting the timbre creation request from the user in the media content editing page. Then, after the target timbre is created, the "use" control is directly clicked on to apply the created target timbre directly to the edited work.

[0075] As shown in the example interfaces 309 and 3011 in FIGS. 3I, 3H, and 3K, after the target timbre of user A is created, user A directly clicks on the "use" control 382 to apply the created target timbre of user A directly to the edited work. At this time, the terminal device 120 presents the target timbre of user A in the area 3011-1 above the timbres 3011-2 (for example, the first timbre, the second timbre, ..., timbre N) pre-provided by the terminal device 120 on the interface 3011. It should be understood that the pre-provided timbres presented on the terminal device 120 are all existing timbres in the timbre library or timbres that are authorized to be used.

[0076] In some embodiments, the terminal device 120 deletes the stored target timbre in response to receiving the deletion request for the stored target timbre. The terminal device 120 presents the timbre creation page in response to receiving the re-recording request for the stored target timbre. The timbre creation page displays another reference text and the recording control.

[0077] As shown in the example interface 3012 in FIG. 3L, the user 110 clicks on the "more" control 3012-1 in the area 3011-1 to manage the generated target timbre (for example, the target timbre of user A). Based on the user clicking on the control 3012-1, the terminal device 120 presents the function card 3012-2 on the interface 1032.

[0078] If the user selects to delete the target timbre of user A on the function card 3012-2, the terminal device 120 deletes the stored target timbre of user A. If the user selects to re-record the target timbre of user A on the function card 3012-2, the terminal device 120 presents the timbre creation page.

[0079] User A may select to set the permission of the target timbre of user A on the function card 3012-2. For example, if the permission of the target timbre of user A is set as a private timbre, the target timbre of user A may only be used by user A. If the permission of the target timbre of user A is set as a public timbre, the target timbre of user A may be used by other users.

[0080] The above describes the process of timbre creation. The following describes the user using the target timbre with reference to the interface 401 shown in FIG. 4A. FIGS. 4A to 4C show schematic diagrams of example interfaces for using the target timbre according to some embodiments of the present disclosure.

[0081] In some embodiments, the terminal device 120 receives the post request for the target media. The target media at least includes the target audio with the target timbre. For example, the terminal device 120 creates the timbre based on the reference text and the audio, and stores the target timbre of the user 110 (for example, user A) for use in the target audio. The user 110 (for example, user A) publishes the target media of the target audio with the target timbre.

[0082] In some embodiments, the terminal device 120 posts the target media content in association with the tag information related to the target timbre in response to the post request. In some examples, the target media

may be audio, text plus image plus audio, or video. The tag information may be a timbre anchor or a sticker bubble.

[0083] As shown in the example interface 401 in FIG. 4A, if the user 110 (for example, user A) posts the target media, the terminal device 120 presents the target media content on the interface 401. If the user 110 (for example, user A) posts the target media, the terminal device 120 also presents the tag information on the interface 401. For example, the control 411 corresponding to the timbre anchor (for example, the timbre of user A) and the control 410 corresponding to the sticker bubble.

[0084] In some embodiments, the tag information at least includes the entry of the detail page corresponding to the target timbre and/or the entry of the timbre usage page corresponding to the target timbre. In some examples, if the user sets the permission of his target timbre as a private permission, the tag information includes the entry of the timbre usage page corresponding to the target timbre. At this time, the terminal device 120 presents the recording page in response to the click operation of the user.

[0085] In some other examples, if the user sets the permission of his target timbre as a public permission, the control 411 corresponding to the timbre anchor may also be the entry of the detail page corresponding to the target timbre. At this time, the terminal device 120 presents the recording page in response to the user clicking on the control 411.

[0086] In some examples, if the user 110 (for example, user A) sets the permission of his target timbre as private, the terminal device 120 will not present the control 411 corresponding to the timbre anchor.

[0087] In some embodiments, the tag information also includes the indication information of the intelligent generation of the target timbre. As shown in FIG. 4A, the terminal device 120 presents the indication information 412 of the intelligent generation presented on the interface 401. This helps to let other users notice that the timbre in the media content is not a real person's recorded voice, but a voice copied based on the intelligently generated timbre after the media content is posted.

[0088] The following describes that the user 110 (for example, user B) discovers the timbre creation entry in the content browsing process with reference to the example interfaces 501 to 5010 shown in FIGS. 5A to 5J. FIGS. 5A to 5J show schematic diagrams of example interfaces for a user to discover a timbre creation entry in a content browsing process according to some embodiments of the present disclosure. For the convenience of understanding, the following will be described with reference to the example interfaces 401 to 403 shown in FIGS. 4A to 4C.

[0089] In some embodiments, if the user 110 requests to create the target timbre in a non-editing page, the terminal device 120 jumps to the editing page of the media content in response to receiving the confirmation of the target timbre in the timbre confirmation page, and the target timbre is selected in the editing page to create the media content.

[0090] For example, if the timbre creation process is entered by selecting the function of creating the target timbre from the timbre detail page, after the target timbre is confirmed, the "try" button is clicked on to enter the editing page, and the user may choose to publish the creation.

[0091] In some embodiments, the non-editing page includes a sound detail page or a timbre detail page of another timbre. In some examples, if the user 110 (for example, user A) publishes the target media including the target timbre, and user B clicks on the control 410 corresponding to the timbre anchor on the page 401, the terminal device 120 will present the sound detail page 403. In some examples, if user A publishes the target media including the target timbre, and the permission of the target timbre is set to private, the terminal device 120 will not present the control 410 corresponding to the timbre anchor on the page 401.

[0092] In some examples, if the user 110 (for example, user A) publishes the target media including the target timbre, and user B clicks on the control 410 corresponding to the sticker bubble on the page 401, the terminal device 120 will present the sound detail page 402. For example, for the target media published by user A, user B may choose to shoot the same style, that is, user B enters the Markov decision process (MDP) display control 410.

[0093] In some examples, the user 110 (for example, user B) may click on the "create timbre" control 420 presented by the terminal device 120 on the interface 402 and/or click on the "create timbre" control 430 presented by the terminal device 120 on the interface 403 to start creating the target timbre of user B.

[0094] As shown in the example interfaces 501 to 5010 in FIGS. 5A to 5J, the terminal device 120 presents the interface 501 based on user B clicking on the control 430 and/or the control 420. User B clicks on the "continue" control 510 on the interface 501 to send the request for creating the target timbre to the terminal device 120. The terminal device 120 receives the request from user B and presents the interface 502.

[0095] The terminal device 120 receives that user B clicks on the "start recording" control 520 to start recording audio. The terminal device 120 stops recording based on the reception that user B clicks on the "stop" control 530. The terminal device 120 generates the timbre of user B based on the reception of the audio of user B and the text (for example, copy D). In the generation process, the terminal device 120 presents the corresponding prompt control 540 of "processing" on the interface 504.

[0096] After creating the target timbre of user B, the terminal device 120 presents the timbre confirmation page 505 including the preview audio generated based on the target timbre of user B. The terminal device 120 plays the audio of reading the predetermined text based on user B clicking on the preview control 550. The terminal device 120 presents the editing page 506 based on user B clicking on the "try" control 551. User B clicks on the "next" control 560 on the editing interface to publish the creation.

[0097] In some examples, on the editing page 506, user B may tap on the area 561 to view the text or edit the text. On the editing page 507, the terminal device 120 receives that user B clicks on the "edit" control 570 and presents the page 508 for user B to edit the text 572 in the input box 580.

[0098] On the editing page 507, the terminal device 120 receives that user B clicks on the "text-to-speech" control 571 and enters the text-to-speech panel to convert the text 572 into audio.

[0099] After the timbre of user B is created, user B directly clicks on the "use" control to directly apply the created timbre of user B to the edited work. At this time, the terminal device 120 presents the target timbre of user B in the upper area 590 on the interface 509. User B may click on the "preview with increased volume" control 5010-1 on the interface 5010 presented by the terminal device 120 to preview the media content.

[00100] The following describes the process of timbre creation and use with reference to FIG. 6 for the convenience of understanding. FIG. 6 shows a flowchart of a process 600 for timbre creation and use according to some embodiments of the present disclosure. The following will be described with reference to FIG. 1 for the convenience of discussion.

[00101] At block 610, the user 110 inputs text in the input box provided by the terminal device 120 and enters the text-to-speech panel by clicking on the text-to-speech control provided by the terminal device 120. At block 611, the terminal device 120 adds a new timbre list to the text-to-speech panel. At block 612, the terminal device 120 detects the timbre creation request sent by the user 110 (for example, by detecting the timbre creation entry provided in the panel triggered by the user) and presents the timbre creation page.

[00102] At block 613, the terminal device 120 allows the user 110 to enter the recording process based on the user 110 clicking on the recording control on the timbre creation page. After creating the timbre based on the audio recorded by the user, at block 614, the terminal device 120 determines whether the user 110 has input text.

[00103] At block 616, if the user 110 has not input text, the terminal device 120 previews the reading effect with the preset copy. At block 617, the terminal device 120 determines whether to contribute. For example, the terminal device 120 determines whether the user 110 chooses to post the media content corresponding to the audio with the timbre of the user 110. At block 618, if the user 110 does not choose to contribute, the terminal device 120 presents the prompt message of "save timbre" on the interface. At block 619, if the user 110 chooses to contribute, the terminal device 120 contributes the media content in the form of copy + text-to-speech reading.

[00104] At block 615, if the user 110 has input text, the terminal device 120 previews the copy (for example, the text input by the user 110) with the recorded timbre of the user. At block 620, the terminal device 120 contributes the media content in the form of timbre copying. At block 627, the terminal device 120 contributes the media content corresponding to the ordinary text-to-speech.

[00105] At block 621, the terminal device 120 determines whether the permission of the timbre of the user 110 is public. At block 622, if the user 110 sets the permission of his timbre as non-public, the terminal device 120 does not display the control corresponding to the timbre anchor on the page and directly enters the recording process.

[00106] At block 623, if the user 110 sets the permission of his timbre as public, the terminal device 120 presents the control corresponding to the timbre anchor and the control corresponding to the sticker bubble on the page. At block 624, the user 110 clicks on the control corresponding to the timbre anchor presented by the terminal device 120 on the interface to enter the timbre detail page. At block 625, the terminal device 120 supports reusing the timbre of the user 110. At block 626, the terminal device 120 presents the function entry on the timbre detail page it presents. For example, the entry for recording audio.

[00107] At block 628, after the media content corresponding to the audio with the timbre is published, the terminal device 120 presents the page entry for shooting the same style on the interface. After the media content corresponding to the ordinary text-to-speech is published, the terminal device 120 presents the page entry for shooting the same style on the interface. At block 629, the terminal device 120 directly enters the recording

process based on the user 110 clicking on the entry of the page for shooting the same style and/or the user 110 clicking on the function entry it presents.

[00108] Through the present disclosure, the discovery of functions is improved, and at the same time, the user better understands the functions through the video on the consumption side (user side). In the audio recording section, by limiting uploading (for example, only allowing current recording) and at the same time strictly limiting the recording duration, the risk of impersonating other people's timbres can be reduced, thereby protecting the user's own voice copyright. Through the corresponding control of the timbre anchor presented by the terminal device 120, the voice can be reused by others. Furthermore, further interactive play methods of the voice are supported to build a voice community.

[00109] FIG. 7 shows a flowchart of a process 700 for timbre creation and use according to some embodiments of the present disclosure. The process 700 may be implemented at the terminal device 120. The process 700 is described below with reference to FIG. 1.

[00110] At block 710, the terminal device 120 presents a timbre creation page in response to a timbre creation request, a reference text and a recording control being displayed in the timbre creation page.

[00111] At block 720, the terminal device 120 receives an audio input by a user in response to detecting a trigger on the recording control.

[00112] At block 730, the terminal device 120 creates a target timbre associated with the user based on the received audio and the reference text.

[00113] At block 740, the terminal device 120 presents a timbre confirmation page, the timbre confirmation page at least including a preview audio generated based on the target timbre.

[00114] At block 750, the terminal device 120 stores, in response to receiving a confirmation of the target timbre in the timbre confirmation page, the target timbre associated with the user for use in an audio.

[00115] In some embodiments, in accordance with a presence of user input text associated with the timbre creation request, the preview audio includes an audio with the target timbre that is generated by performing a text-to-speech operation on the user input text, or in the absence of the user input text associated with the timbre creation request, the preview audio includes an audio with the target timbre that is generated by performing the text-to-speech operation on a predetermined text.

[00116] In some embodiments, the process 700 further includes: prompting the user to re-record the audio in response to detecting that the received audio does not match the reference text; and displaying another reference text and the recording control in the timbre creation page.

[00117] In some embodiments, the process 700 further includes: prompting the user to re-record the audio in response to detecting that background noise in the audio exceeds a noise threshold or a duration of the audio is below a preset threshold; and displaying another reference text and the recording control in the timbre creation page.

[00118] In some embodiments, the process 700 further includes: displaying another reference text and the recording control in the timbre creation page in response to detecting a re-generation request in the timbre creation page.

[00119] In some embodiments, the timbre creation request is detected in an editing page of media content, and the method further includes: jumping to the editing page in response to receiving the confirmation of the target timbre in the timbre confirmation page, where the target timbre is selected to be applied to an audio of a media item being edited.

[00120] In some embodiments, the timbre creation request is detected in a non-editing page, and the method further includes: jumping to an editing page of media content in response to receiving the confirmation of the target timbre in the timbre confirmation page, the target timbre being selected in the editing page to create the media content.

[00121] In some embodiments, the non-editing page includes a sound detail page, or a timbre detail page of another timbre.

[00122] In some embodiments, the process 700 further includes: receiving a post request for a target media, the target media at least including a target audio having the target timbre; and post, in response to the post request, target media content in association with tag information related to the target timbre.

[00123] In some embodiments, the tag information at least includes an entry to a detail page corresponding to the target timbre and/or an entry to a timbre usage page corresponding to the target timbre.

[00124] In some embodiments, the tag information further includes indication information of intelligent generation of the target timbre.

[00125] In some embodiments, the process 700 further includes: deleting the stored target timbre in response to receiving a deletion request for the stored target timbre; and presenting the timbre creation page in response to receiving a re-recording request for the stored target timbre, another reference text and the recording control being displayed in the timbre creation page.

[00126] FIG. 8 shows a schematic structural block diagram of an apparatus 800 for timbre creation and use according to some embodiments of the present disclosure. The apparatus 800 may be implemented as or included in the terminal device 120. Each module/component in the apparatus 800 may be implemented by hardware, software, firmware, or any combination thereof.

[00127] As shown in the figure, the apparatus 800 includes a creation page presenting module 810 configured to present a timbre creation page in response to a timbre creation request, a reference text and a recording control being displayed in the timbre creation page.

[00128] The apparatus 800 further includes an audio receiving module 820 configured to receive an audio input by a user in response to detecting a trigger on the recording control.

[00129] The apparatus 800 further includes a timbre creation module 830 configured to create a target timbre associated with the user based on the received audio and the reference text.

[00130] The apparatus 800 further includes a confirmation page presenting module 840 configured to present a timbre confirmation page, the timbre confirmation page including at least a preview audio generated based on the target timbre.

[00131] The apparatus 800 further includes a timbre storing module 850 configured to store, in response to receiving a confirmation of the target timbre in the timbre confirmation page, the target timbre associated with the user for use in an audio.

[00132] In some embodiments, in accordance with a presence of user input text associated with the timbre creation request, the preview audio includes an audio with the target timbre that is generated by performing a text-to-speech operation on the user input text, or in the absence of the user input text associated with the timbre creation request, the preview audio includes an audio with the target timbre that is generated by performing the text-to-speech operation on a predetermined text.

[00133] In some embodiments, the apparatus 800 further includes an audio prompting module configured to prompt the user to re-record the audio in response to detecting that the received audio does not match the reference text; and display another reference text and the recording control in the timbre creation page.

[00134] In some embodiments, the audio prompting module is configured to prompt the user to re-record the audio in response to detecting that background noise in the audio exceeds a noise threshold or a duration of the audio is below a preset threshold; and display another reference text and the recording control in the timbre creation page.

[00135] In some embodiments, the apparatus 800 further includes a generation request module configured to display another reference text and the recording control in the timbre creation page in response to detecting a re-generation request in the timbre creation page.

[00136] In some embodiments, the timbre creation request is detected in an editing page of media content, and the apparatus 800 further includes a jumping module configured to jump to the editing page in response to receiving the confirmation of the target timbre in the timbre confirmation page, where the target timbre is selected to be applied to an audio of a media item being edited.

[00137] In some embodiments, the timbre creation request is detected in a non-editing page, and the jumping module is further configured to jump to an editing page of media content in response to receiving the confirmation of the target timbre in the timbre confirmation page, the target timbre being selected in the editing page to create the media content.

[00138] In some embodiments, the non-editing page includes a sound detail page, or a timbre detail page of another timbre.

[00139] In some embodiments, the apparatus 800 further includes a request receiving module configured to receive a post request for a target media, the target media at least including a target audio having the target timbre; and post, in response to the post request, target media content in association with tag information related to the target timbre.

[00140] In some embodiments, the tag information at least includes an entry to a detail page corresponding to the target timbre and/or an entry to a timbre usage page corresponding to the target timbre.

[00141] In some embodiments, the tag information further includes indication information of intelligent generation of the target timbre.

[00142] In some embodiments, the apparatus 800 further includes a request deletion module configured to delete the stored target timbre in response to receiving a deletion request for the stored target timbre; and present the timbre creation page in response to receiving a re-recording request for the stored target timbre, another reference text and the recording control being displayed in the timbre creation page.

[00143] FIG. 9 shows a block diagram of an electronic device 900 in which one or more embodiments of the present disclosure may be implemented. It should be understood that the electronic device 900 shown in FIG. 9 is only illustrative, and should not constitute any limitation on the function and scope of the embodiments described herein. The electronic device 900 shown in Fig. 9 may be used to implement the terminal device 120 in FIG. 1 or the apparatus 800 in FIG. 8.

[00144] As shown in FIG. 9, the electronic device 900 is in the form of a general electronic device. The components of the electronic device 900 may include, but are not limited to, one or more processors or processing units 910, a memory 920, a storage device 930, one or more communication units 940, one or more input devices 950, and one or more output devices 960. The processing unit 910 may be an actual or virtual processor, and may perform various processing based on the programs stored in the memory 920. In a multi-processor system, multiple processing units execute computer-executable instructions in parallel to improve the parallel processing capability of the electronic device 900.

[00145] The electronic device 900 typically includes multiple computer storage medium. Such medium may be any available medium that is accessible to the electronic device 900, including, but not limited to, volatile and non-volatile medium, removable and non-removable medium. The memory 920 may be a volatile memory (for example, a register, cache, a random access memory (RAM)), a non-volatile memory (such as a read-only memory (ROM), an electrically erasable programmable read-only memory (EEPROM), a flash memory), or any combination thereof. The storage device 930 may be any removable or non-removable medium, and may include a machine-readable medium such as a flash drive, a disk, or any other medium, which may be used to store information and/or data (such as training data for training) and may be accessed within the electronic device 900.

[00146] The electronic device 900 may further include other removable/non-removable, volatile/non-volatile memory medium. Although not shown in FIG. 9, a disk drive for reading from or writing to a removable, non-volatile disk (such as a "floppy disk"), and an optical disk drive for reading from or writing to a removable, non-volatile optical disk may be provided. In these cases, each drive may be connected to the bus (not shown) by one or more data medium interfaces. The memory 920 may include a computer program product 925 having one or more program modules configured to perform various methods or acts of the various embodiments of the present disclosure.

[00147] The communication unit 940 enables communication with other electronic devices through the communication medium. Additionally, the functions of the components of the electronic device 900 may be implemented with a single computing cluster or multiple computing machines, which may communicate through communication connections. Therefore, the electronic device 900 may use a logical connection with one or more

other servers, a network personal computer (PC), or another network node to operate in a networked environment.

[00148] The input device 950 may be one or more input devices, such as a mouse, a keyboard, a tracking ball, etc. The output device 960 may be one or more output devices, such as a display, a speaker, a printer, etc. The electronic device 900 may also communicate with one or more external devices (not shown) as needed through the communication unit 940, such as a storage device, a display device, etc., communicate with one or more devices that enable the user to interact with the electronic device 900, or communicate with any devices (such as a network card, a modem, etc.) that enable the electronic device 900 to communicate with one or more other electronic devices. Such communication may be performed via input/output (I/O) interfaces (not shown).

[00149] According to an example implementation of the present disclosure, a computer-readable storage medium is provided, having computer-executable instructions stored thereon, where the computer-executable instructions are executed by a processor to implement the method described above. According to an example implementation of the present disclosure, there is further provided a computer program product, which is tangibly stored on a non-transitory computer-readable medium and includes computer-executable instructions, and the computer-executable instructions are executed by a processor to implement the method described above.

[00150] Various aspects of the present disclosure are described herein with reference to flowcharts and/or block diagrams of methods, apparatuses, devices, and computer program products implemented according to the present disclosure. It should be understood that each block of the flowchart and/or block diagrams, and combinations of blocks in the flowchart and/or block diagrams, may be implemented by computer-readable program instructions.

[00151] 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, such that the instructions, when executed by the processing unit of the computer or other programmable data processing apparatus, produce an apparatus for implementing the functions/acts specified in one or more blocks of the flowchart and/or block diagrams. These computer-readable program instructions may also be stored in a computer-readable storage medium, these instructions cause the computer, the programmable data processing apparatus, and/or other devices to work in a particular manner, and thus, the computer-readable medium storing the instructions includes a manufactured article, which includes instructions for implementing various aspects of the functions/acts specified in one or more blocks of the flowchart and/or block diagrams.

[00152] The computer-readable program instructions may be loaded onto a computer, another programmable data processing apparatus, or other devices, so that a series of operations and steps are performed on the computer, the other programmable data processing apparatus, or the other devices to produce a computer-implemented process, such that the instructions executed on the computer, the other programmable data processing apparatus, or the other devices implement the functions/acts specified in one or more blocks of the flowchart and/or block diagrams.

[00153] The flowchart and block diagrams in the drawings show the possibly implemented architectures, functions and operations of the systems, methods and computer program products according to multiple

implementations of the present disclosure. In this regard, each block in the flowchart or block diagram may represent a module, program segment or part of instructions, which includes 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 performed substantially in parallel, or they may sometimes be performed in the reverse order, depending on the functions involved. It should also be noted that each block in the block diagrams and/or flowcharts, and combinations of the blocks in the block diagrams and/or flowcharts may be implemented in special purpose hardware-based systems that perform the specified functions or acts, or may be implemented in a combination of special purpose hardware and computer instructions.

[00154] The implementations of the present disclosure have been described above, and the above description is illustrative, non-exhaustive, and not limited to the disclosed implementations. Without departing from the scope and spirit of the illustrated implementations, 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 implementations, the practical applications or improvements to the technology in the market, or to enable other ordinary skilled persons in the art to understand the implementations disclosed herein.

CLAIMS

1. A method for timbre creation and use, comprising:
 - presenting a timbre creation page in response to a timbre creation request, a reference text and a recording control being displayed in the timbre creation page;
 - receiving an audio input by a user in response to detecting a trigger on the recording control;
 - creating a target timbre associated with the user based on the received audio and the reference text;
 - presenting a timbre confirmation page, the timbre confirmation page at least comprising a preview audio generated based on the target timbre; and
 - storing, in response to receiving a confirmation of the target timbre in the timbre confirmation page, the target timbre associated with the user for use in an audio.
2. The method of claim 1, wherein in accordance with a presence of user input text associated with the timbre creation request, the preview audio comprises an audio with the target timbre that is generated by performing a text-to-speech operation on the user input text, or
 - in the absence of user input text associated with the timbre creation request, the preview audio comprises an audio with the target timbre that is generated by performing the text-to-speech operation on predetermined text.
3. The method of any one of claims 1 to 2, further comprising:
 - prompting the user to re-record the audio in response to detecting that the received audio does not match the reference text; and
 - displaying another reference text and the recording control in the timbre creation page.
4. The method of any one of claims 1 to 3, further comprising:
 - prompting the user to re-record the audio in response to detecting that background noise in the audio exceeds a noise threshold or a duration of the audio is below a preset threshold; and
 - displaying another reference text and the recording control in the timbre creation page.
5. The method of any one of claims 1 to 4, further comprising:
 - displaying another reference text and the recording control in the timbre creation page in response to detecting a re-generation request in the timbre creation page.
6. The method of any one of claims 1 to 5, wherein the timbre creation request is detected in an editing page of media content, and the method further comprises:
 - jumping to the editing page in response to receiving the confirmation of the target timbre in the timbre confirmation page, wherein the target timbre is selected to be applied to an audio of a media item being edited.

7. The method of any one of claims 1 to 6, wherein the timbre creation request is detected in a non-editing page, and the method further comprises:

jumping to an editing page of media content in response to receiving the confirmation of the target timbre in the timbre confirmation page, the target timbre being selected in the editing page to create the media content.

8. The method of claim 7, wherein the non-editing page comprises a sound detail page, or a timbre detail page of another timbre.

9. The method of any one of claims 1 to 8, further comprising:

receiving a post request for a target media, the target media at least comprising a target audio having the target timbre; and

posting, in response to the post request, target media content in association with tag information related to the target timbre.

10. The method of claim 9, wherein the tag information at least comprises an entry to a detail page corresponding to the target timbre and/or an entry to a timbre usage page corresponding to the target timbre.

11 The method of claim 9, wherein the tag information further comprises indication information of intelligent generation of the target timbre.

12. The method of any one of claims 1 to 11, further comprising:

deleting the stored target timbre in response to receiving a deletion request for the stored target timbre; and

presenting the timbre creation page in response to receiving a re-recording request for the stored target timbre, another reference text and the recording control being displayed in the timbre creation page.

13. An apparatus for timbre creation and use, comprising:

a creation page presenting module configured to present a timbre creation page in response to a timbre creation request, a reference text and a recording control being displayed in the timbre creation page;

an audio receiving module configured to receive an audio input by a user in response to detecting a trigger on the recording control;

a timbre creation module configured to create a target timbre associated with the user based on the received audio and the reference text;

a confirmation page presenting module configured to present a timbre confirmation page, the timbre confirmation page at least comprising a preview audio generated based on the target timbre; and

a timbre storing module configured to store, in response to receiving a confirmation of the target timbre in the timbre confirmation page, the target timbre associated with the user for use in an audio.

14. An electronic device, comprising:

at least one processing unit; and

at least one memory coupled to the at least one processing unit and storing instructions executable by the at least one processing unit, wherein the instructions, when executed by the at least one processing unit, cause the electronic device to perform the method of any one of claims 1 to 12.

15. A computer-readable storage medium having a computer program stored thereon, wherein the computer program is executable by a processor to implement the method of any one of claims 1 to 12.

16. A computer program product tangibly stored in a computer storage medium and comprising computer-executable instructions, wherein the computer-executable instructions, when executed by a device, cause the device to perform the method of any one of claims 1 to 12.