

# **METHOD AND APPARATUS FOR GENERATING LYRICS, METHOD AND APPARATUS FOR DISPLAYING LYRICS, ELECTRONIC DEVICE, AND STORAGE MEDIUM**

[0001] The present application claims priority to Chinese Patent Application No. 201810513535.7, filed on May 25, 2018 and titled "METHOD AND APPARATUS FOR GENERATING LYRICS, METHOD AND APPARATUS FOR DISPLAYING LYRICS, ELECTRONIC DEVICE, AND STORAGE MEDIUM", which is incorporated herein by reference in its entirety.

## **TECHNICAL FIELD**

[0002] The present invention relates to the field of an Internet technology, and more particularly to a method and apparatus for generating lyrics, a method and apparatus for displaying lyrics, an electronic device, and a storage medium.

## **BACKGROUND**

[0003] With the development of Internet technology, many music players not only support the online playback of massive songs, but also provide users with a Karaoke song service. The Karaoke song service refers to playing a music accompaniment of a song with a music player and displaying lyrics on the current playing interface, so that users can sing along with the music accompaniment while watching the lyrics.

[0004] At present, on the current playing interface, a terminal is usually in a form of Karaoke, i.e., displaying a piece of lyrics corresponding to the current playing moment, and dynamically annotating which character in the lyrics is currently being played through a font color. During the actual singing of a song, in order to better express the feelings of the song, an original singer usually changes original pronunciations of some characters in lyrics into more artistic pronunciations of characters for singing. However, if a user is unfamiliar with original singer's pronunciations of the song, he/she will often sing original pronunciations of the characters, which is not consistent with the original singer's pronunciations of the song, resulting errors in the singing of the user. Therefore,

a method for displaying lyrics and a method for generating lyrics are urgently needed to ensure that users can accurately sing the original singer's pronunciation of each character.

## **SUMMARY**

**[0005]** Embodiments of the present application disclose a method and apparatus for generating lyrics, a method and apparatus for displaying lyrics, an electronic device, and a storage medium, which can solve the problem that users in the related art sing wrong characters in lyrics. The technical solutions are summarized as follows.

**[0006]** In a first aspect, a method for generating lyrics is provided. The method includes:

**[0007]** acquiring lyrics of a target song and an audio file of the target song;

**[0008]** determining, according to a character to be annotated among a plurality of characters in the lyrics and audio clips corresponding to the plurality of characters in the audio file, a candidate pronunciation of the character to be annotated;

**[0009]** if the character to be annotated has one candidate pronunciation, determining the candidate pronunciation as the corresponding pronunciation of the character to be annotated in the target song;

**[0010]** if the character to be annotated has at least two candidate pronunciations, determining the candidate pronunciation semantically matching the character to be annotated as the corresponding pronunciation of the character to be annotated in the target song; and

**[0011]** generating a first lyric file of the target song according to the plurality of characters and the corresponding pronunciation of the character to be annotated in the target song.

**[0012]** In a second aspect, a method for displaying lyrics is provided. The method is applied to a terminal and includes:

**[0013]** when a lyric display instruction is received, acquiring a first lyric file of a target song, the lyric display instruction being configured to display lyrics of the target song;

**[0014]** acquiring, from the first lyric file, the lyrics of the target song and a corresponding pronunciation of a character to be annotated among a plurality of characters of the lyrics in the target song;

**[0015]** displaying the plurality of characters of the lyrics; and

**[0016]** in displaying the character to be annotated, annotating, at a target position of the character to be annotated, the corresponding pronunciation of the character to be annotated in the target song, the target position being located above the character to be annotated.

**[0017]** In a third aspect, an apparatus for generating lyrics is provided. The apparatus includes:

**[0018]** an acquisition module configured to acquire lyrics of a target song and an audio file of the target song;

**[0019]** a determination module configured to determine, according to a character to be annotated among a plurality of characters in the lyrics and audio clips corresponding to the plurality of characters in the audio file, a candidate pronunciation of the character to be annotated;

**[0020]** the determination module being further configured to, if the character to be annotated has one candidate pronunciation, determine the candidate pronunciation as the corresponding pronunciation of the character to be annotated in the target song;

**[0021]** the determination module being further configured to, if the character to be annotated has at least two candidate pronunciations, determine the candidate pronunciation semantically matching the character as the corresponding pronunciation of the character to be annotated in the target song; and

**[0022]** a generation module configured to generate a first lyric file of the target song according to the corresponding pronunciations of the plurality of characters and the character to be annotated in the target song.

**[0023]** In a fourth aspect, an apparatus for displaying lyrics is provided. The apparatus is applied to a terminal and includes:

**[0024]** an acquisition module configured to, when a lyric display instruction is received, acquire a first lyric file of a target song, the lyric display instruction being configured to display lyrics of the target song;

**[0025]** the acquisition module being further configured to acquire, from the first lyric file, the lyrics of the target song and a corresponding pronunciation of a character to be annotated among a plurality of characters of the lyrics in the target song;

[0026] a display module configured to display the plurality of characters of the lyrics;  
and

[0027] an annotating module configured to, in displaying the character to be annotated, mark, at a target position of the character to be annotated, the corresponding pronunciation of the character to be annotated in the target song, the target position being located above the character to be annotated.

[0028] In a fifth aspect, an electronic device is provided. The device includes a processor and a memory, at least one instruction being stored in the memory, and loaded and executed by the processor to implement the operations executed by the method for generating the lyrics according to the first aspect, or the method for displaying the lyrics according to the second aspect.

[0029] In a sixth aspect, a computer readable storage medium is provided. At least one instruction is stored in the storage medium, and loaded and executed by a processor to implement the operations of the method for generating the lyrics according to the first aspect, or the method for displaying the lyrics according to the second aspect.

[0030] The technical solutions provided by the embodiments of the present invention may include the following benefits.

[0031] In the embodiments of the present invention, a server may determine, based on the lyrics of the target song and the audio file of the target song, the corresponding pronunciation of each character to be annotated in the lyrics in the target song, and may further generate the first lyric file of the target song according to the pronunciations of the plurality of characters and the character to be annotated, thereby binding a corresponding pronunciation for each character to be annotated. Therefore, in the subsequent display of the lyrics, the pronunciations can be displayed synchronously to ensure that the user can sing each character of the target song correctly. In addition, when the terminal displays the lyrics, the pronunciation may also be annotated above the corresponding character to be annotated, such that the pronunciation is clearly visible, and users can accurately and quickly find the corresponding pronunciations of the character to be annotated, thereby improving the accuracy of the displayed lyrics.

## **BRIEF DESCRIPTION OF THE DRAWINGS**

[0032] In order to describe the technical solutions in the embodiments of the present invention more clearly, the following briefly introduces the accompanying drawings required for describing the embodiments. Apparently, the accompanying drawings in the following description show merely some embodiments of the present invention, and a person of ordinary skill in the art may also derive other drawings from these accompanying drawings without creative efforts.

[0033] FIG. 1 is a schematic diagram of an implementation environment provided by an embodiment of the present disclosure;

[0034] FIG. 2 is a flowchart of a method for generating lyrics according to an embodiment of the present invention;

[0035] FIG. 3 is a flowchart of a method for displaying lyrics according to an embodiment of the present invention;

[0036] FIG. 4 is a schematic diagram of an interface for displaying lyrics according to an embodiment of the present invention;

[0037] FIG. 5 is a schematic diagram of an interface for displaying lyrics according to an embodiment of the present invention;

[0038] FIG. 6 is a schematic structural diagram of an apparatus for generating lyrics provided by an embodiment of the present invention;

[0039] FIG. 7 is a schematic structural diagram of an apparatus for displaying lyrics provided by an embodiment of the present invention;

[0040] FIG. 8 is a schematic structural diagram of a terminal provided by an embodiment of the present invention; and

[0041] FIG. 9 is a schematic structural diagram of a server provided by an embodiment of the present invention.

## **DETAILED DESCRIPTION**

[0042] To make the objects, technical solutions and advantages of the present application clearer, the embodiments of the present invention will be described below in detail in conjunction with the accompanying drawings. It is obvious that the described embodiments are part rather than all of the embodiments of the present invention. All other embodiments obtained by those of ordinary skill in the art based on the

embodiments of the present invention without creative work are within the scope of protection of the present invention.

**[0043]** FIG. 1 is a schematic diagram of an implementation environment provided by an embodiment of the present invention, the implementation environment including a terminal 101 and a server 102. The terminal 101 and the server 102 are connected via a network. An application is installed on the terminal 101. The terminal 101 may acquire a target song from the server 102, play the target song, display lyrics based on the application, and the like.

**[0044]** The server 102 may, based on a plurality of characters to be annotated in the lyrics of the target song and an audio file of the target song, determine pronunciations of the characters to be annotated among the plurality of characters in advance, so as to obtain the corresponding pronunciations of the characters to be annotated in the target song. The server may store the lyrics and the corresponding pronunciations of the characters to be annotated in the target song, and send the lyrics of the target song and the corresponding pronunciations of the characters to be annotated to the terminal 101. When the terminal 101 plays the target song, the lyrics of the target song are displayed synchronously, and the corresponding pronunciations of the characters to be annotated are annotated above the characters to be annotated.

**[0045]** The above-mentioned process of acquiring the pronunciations of the characters to be annotated among the plurality of characters may also be performed by the terminal 101. That is, after the terminal 101 acquires the lyrics of the target song from the server 102, the pronunciations are determined and annotated for the characters to be annotated in the lyrics.

**[0046]** It should be noted that in the actual singing of some songs, in order to better express the feelings expressed by a song, original pronunciations of some characters in lyrics are changed into pronunciations of more artistic characters. For example, in some Japanese songs, "未来" is often sung as "あ す", but a Chinese character corresponding to "あ す" is "明日", rather than the correct pronunciation of "未来". For the needs of the artistic expression of the songs, "未来" is sung as the pronunciation "あす" of "明日".

**[0047]** In the present invention, the pronunciations of the characters to be annotated of the target song as an original singer sings is written into a lyric file of the target song for

storage, thereby facilitating the subsequent display of the pronunciations of the characters to be annotated as the original singer sings based on the lyric file.

**[0048]** This application may be an application of a music player or an application in which a music playback plugin is installed. The terminal may be a mobile phone terminal, a PAD (Portable Android Device) terminal, a computer terminal or the like. The server 102 is a background server of the application. The server 102 may be a server, or a server cluster composed of several servers, or a cloud computing server center.

**[0049]** FIG. 2 is a flowchart of a method for generating lyrics according to an embodiment of the present invention. An executive main body of this embodiment of the present invention is a terminal. Referring to FIG. 2, the method includes the following steps.

**[0050]** In 201, a server acquires lyrics of a target song and an audio file of the target song.

**[0051]** The audio file includes audio clips corresponding to the plurality of characters in the lyrics. The target song is any song that needs to generate lyrics, and generally a song that includes Chinese characters in lyrics, the Chinese characters including, but not limited to, simplified Chinese characters, traditional Chinese characters, or the like. The plurality of characters may be text symbols in a target language, for example, Chinese characters in Japanese, furigana, or the like. In this embodiment of the present invention, the server may determine a pronunciation of each character to be annotated in the lyrics based on the pronunciations of the target song as an original singer sings, thereby accurately obtaining the pronunciations consistent with the original singer's pronunciations of the target song.

**[0052]** In this step, the server may acquire a second lyric file and an original audio file of the target song, then acquires the lyrics of the target song from the second lyric file, extract audio clips corresponding to the plurality of characters from the original audio file, and generate the audio file according to the audio clips corresponding to the plurality of characters.

**[0053]** The audio clips may be audio clips in a human voice band. In this step, the server may obtain the audio file by eliminating an accompaniment. The step for the server to acquire the audio file of the target song may be as follows: the server acquires the

original audio file of the target song, the original audio file including collected audio clips in various bands; the server extracts audio clips in the human voice band from the original audio file, and generates the audio file of the target song according to the audio clips in the human voice band, thereby eliminating the accompaniment of the target song. Generally, the server extracts audio clips of the target song directly in front of a space from the original audio file, and selects the audio clips in the human voice band. Generally, the human voice band includes an intermediate frequency and a high frequency.

**[0054]** In lyric files of some songs, display times of a plurality of characters are often included, the display times being used to indicate in displaying each character during the playing process of the song. In addition, in a lyric file, each character and the display time of this character are stored in a spliced manner. The display time often exists between the characters. For example, the lyrics "雨 の 空港" in the lyric file is actually stored in a form of "(39112) 雨 (39803) の (40356) 空 (40606) 港 (41176)", wherein the numerals in brackets indicate display times of characters, and the characters are cut off by the display times. The server may acquire a second lyric file of the target song, extract the lyrics of the target song and the display times of a plurality of characters in the lyrics from the second lyric file, and establish an index relationship between each character and the display time of this character, thereby ensuring that the characters can also be displayed based on the display times during subsequent display. The second lyric file is an original lyric file of the target song, and includes the lyrics of the target song and the display times of a plurality of characters in the lyrics.

**[0055]** The server may store the characters and the display times separately in a form of arrays, so that the characters may be in one-to-one correspondence with the display times. In a possible design, the step for the server to acquire the lyrics of the target song may be as follows: the server acquires the second lyric file of the target song, and establishes a first array and a third array. The first array is configured to store a plurality of characters in the lyrics. The third array is configured to store the display times of the plurality of characters. The server writes the plurality of characters into the first array, writes the display times of the plurality of characters into the third array, and determine the plurality of characters in the first array as the lyrics of the target song. In the third array, the server

may store the display time of each character in the third array according to a storage order of each character in the first array. Based on the storage order, each character in the first array is in one-to-one correspondence with the display time of this character in the third array, thereby improving the accuracy of the lyrics and the display times.

**[0056]** It should be noted that the server may determine pronunciations subsequently based on characters by extracting the characters. Meanwhile, the server extracts the lyrics by establishing the index relationship between each character and the display time of this character without destroying the original content of the second lyric file, thereby ensuring that an accurate lyric file can be finally obtained.

**[0057]** In 202, the server determines a Chinese character among the plurality of characters as the character to be annotated,

**[0058]** In this embodiment of the present invention, the server may recognize a Chinese character among the plurality of characters by using a preset recognition algorithm, and determine the Chinese character as the character to be annotated. The Chinese character includes, but is not limited to, a simplified Chinese character, a traditional Chinese character, or the like. The server generally stores encoded data of the plurality of characters. The preset recognition algorithm may be configured to recognize the plurality of characters based on an encoding table of encoding modes of the plurality of characters. Of course, this embodiment of the present invention does not limit the encoding modes of characters. For example, the encoding table may be an encoding table of a Unicode (Uniform Code) encoding mode.

**[0059]** It should be noted that the server recognizes the character to be annotated by recognizing the plurality of characters first, and thus directly determines the pronunciation directly based on the character to be annotated among the plurality of characters, thereby reducing the number of characters to be processed subsequently, and further improving the efficiency of determining the pronunciation. In addition, in this embodiment, the Chinese character is taken as the character to be annotated, and the pronunciation of the Chinese character in the lyrics is then annotated. For a user whose mother tongue is not Chinese or Japanese, it is convenient for the user to learn a foreign language by singing, which not only greatly satisfies the listening pleasure, but also

achieves the effect of learning foreign languages, and greatly enriches the user experience, thereby greatly improving the practical value of this method.

**[0060]** In 203, the server recognizes pronunciations of the plurality of characters according to the audio clips corresponding to the plurality of characters, and determines a candidate pronunciation of the character to be annotated according to the character to be annotated among the plurality of characters.

**[0061]** In this embodiment, the server may recognize the plurality of pronunciations by means of a preset speech recognition algorithm based on the audio clips corresponding to the plurality of characters, and determine the candidate pronunciation of the character to be annotated according to relative positions between the character to be annotated among the plurality of characters and characters other than the character to be annotated.

**[0062]** However, there may be noise in the audio clips corresponding to the plurality of characters in the audio file. Alternatively, during actual singing, a syllable of a character may be stretched, which may cause that the pronunciations of the plurality of characters possibly include some fuzzy pronunciations during voice recognition. The server may temporarily use an intermediate pronunciation to indicate the pronunciation that has not been uniquely determined. Therefore, this step includes the following two case.

**[0063]** In the first case, if the pronunciations of a plurality of characters recognized based on audio clips do not include fuzzy pronunciations, that is, when each pronunciation is uniquely determined, the server determines, according to a character to be annotated among the plurality of characters, a candidate pronunciation of the character to be annotated from the pronunciations of the plurality of characters.

**[0064]** The server determines the corresponding candidate pronunciation of the character to be annotated among the plurality of pronunciations according to the positions of the characters other than the character to be annotated among the plurality of characters in the plurality of pronunciations. Of course, in the case that a plurality of characters to be annotated are provided, the server may further determine corresponding candidate pronunciations of the characters to be annotated in the plurality of pronunciations according to the positions of other characters adjacent to the characters to be annotated in the plurality of pronunciations.

**[0065]** It should be noted that in this step, only one candidate pronunciation corresponding to the character to be annotated is determined by the server. Subsequently, the following step 204 may be used to determine the corresponding pronunciation of the character to be annotated in the target song.

**[0066]** In the second case, if the pronunciations of a plurality of characters recognized based on audio clips include fuzzy pronunciations, the server determines, according to a character to be annotated among the plurality of characters, an intermediate pronunciation of the character to be annotated from the pronunciations of the plurality of characters, and recognize at least two candidate pronunciations of the character to be annotated according to the intermediate pronunciation.

**[0067]** The intermediate pronunciation can be a Roman pronunciation. The server may match at least two candidate pronunciations corresponding to the character to be annotated from a dictionary engine based on the Roman pronunciation. The at least two candidate pronunciations are furigana used to represent pronunciations of Chinese characters in Japanese. For example, taking Japanese as an example, with respect to audio clips corresponding to "ペガサス幻想そうさ夢だけは" in lyrics, a plurality of pronunciations recognized by voice may be "ペガサス fantaji そうさゆめだけは", wherein "fantaji" is a Roman pronunciation. The server may determine, based on the Roman pronunciation "fantaji", three candidate pronunciations corresponding to the character to be annotated as: "ふぁんたじ", "ファンタジ", and "ファンタジー".

**[0068]** It should be noted that in this step, at least two candidate pronunciations corresponding to the character to be annotated are determined by the server. Subsequently, the following step 205 may be used to determine the corresponding pronunciation of the character to be annotated in the target song.

**[0069]** It should be noted that the above steps 202-203 are the specific implementation of "the server determines, according to a character to be annotated among the plurality of characters in the lyrics and audio clips corresponding to the plurality of characters in the audio file, the candidate pronunciation of the character to be annotated". In practice, the above steps are of the process of determining the character to be annotated in the lyrics, and then performing voice recognition based on the audio file. However, the server may also perform voice recognition based on the audio file to determine the pronunciations of

the plurality of characters, and then determine the character to be annotated in the lyrics, thereby determining the candidate pronunciations, which will not be specifically limited in this embodiment of the present invention.

**[0070]** In 204, if the character to be annotated has one candidate pronunciation, the server determines the candidate pronunciation as the corresponding pronunciation of the character to be annotated in the target song.

**[0071]** When the server recognizes that the candidate pronunciation corresponding to the character to be annotated is uniquely determined, with respect to each character to be annotated, the server may directly determine the candidate pronunciation corresponding to the character to be annotated as the corresponding pronunciation of the character to be annotated in the target song.

**[0072]** Taking Japanese as an example, the server can use furigana to represent the pronunciations of the plurality of characters. Based on a plurality of characters "誰でももし男(ひ)女(と)が人(ひ)間(と)で ありたいなら" in the lyrics and the plurality of pronunciations "だれでももしひとがひとでありたいなら" recognized according to the audio clips corresponding to the plurality of characters, the process of determining the pronunciations of "誰", "男女", and "人間" among the plurality of characters may be as follows.

**[0073]** a. the plurality of pronunciations are matched with the plurality of characters, and positions of "でももし", "が", "で ありたいなら" of "誰でももし男(ひ)女(と)が人(ひ)間(と)で ありたいなら" in "だれでももしひとがひとでありたいなら" are determined respectively.

**[0074]** b. respective pronunciations of the plurality of "誰", "男女" and "人間" in this "だれでももしひとがひとでありたいなら" are determined respectively according to the relative positions of the characters "誰", "男女" and "人間" and the recognized "でももし", "が", "で ありたいなら", wherein

**[0075]** "誰" is pronounced as "だ れ";

**[0076]** "男女" is correspondingly pronounced as "ひ と"; and

**[0077]** "人間" is correspondingly pronounced as "ひ と".

**[0078]** It should be noted that since only the pronunciations of the lyrics in Japanese songs need to be determined, it is only necessary to recognize pronunciations of respective furigana in Japanese when performing voice recognition based on audio clips, thereby reducing the range of pronunciations to be recognized. In terms of fault tolerance, the pronunciation recognition only needs to be performed based on the various existing furigana in the Japanese range, without considering other languages, thereby improving the accuracy of a plurality of recognized pronunciations. In addition, the server may further limit a voice interval corresponding to each of the characters to be annotated in the plurality of pronunciations based on each of the characters to be annotated and a plurality of furigana, thereby improving the accuracy of subsequent pronunciation determination of the characters to be annotated.

**[0079]** In 205, if the character to be annotated has at least two candidate pronunciations, the server determines the candidate pronunciation semantically matching the character to be annotated as the corresponding pronunciation of the character to be annotated in the target song.

**[0080]** In this step, the server may establish and store dictionary engines in advance, and looks up a candidate pronunciation matching with a voice based on the dictionary engines. With respect to each candidate pronunciation, the server looks up an intermediate character corresponding to each candidate pronunciation from the first dictionary engine; looks up a semantic meaning of the intermediate character from the second dictionary engine; select, from the at least two candidate pronunciations, the candidate pronunciation of which the semantic meaning of the intermediate character matches the semantic meaning of the character to be annotated according to the semantic meaning of the character to be annotated; and determine the selected candidate pronunciation as the corresponding pronunciation of the character to be annotated in the target song.

**[0081]** The first dictionary engine includes a plurality of candidate pronunciations and the intermediate characters corresponding to the plurality of candidate pronunciations. The second dictionary engine includes a plurality of intermediate characters and a plurality of semantic meanings corresponding to the plurality of intermediate characters. The corresponding pronunciation of the character to be annotated in the target song may be represented by furigana in Japanese. The intermediate character may be a text symbol

in a language, for example, a text symbol in English, Chinese, German, Japanese, French, or the like. The first dictionary engine may be a Japanese-Chinese dictionary, a Japanese-Japanese dictionary, a Japanese-English dictionary, a Japanese-German dictionary, a Japanese-French dictionary, or the like, which will not be specifically limited in this embodiment of the present invention.

**[0082]** Taking Japanese as an example, a plurality of pronunciations "ペガサス fantaji そうさゆめだけは" corresponding to "ペガサス幻想そうさ夢だけは" in the lyrics are illustrated with "幻想" as an example. The process of determining this pronunciation is as follows:

**[0083]** a. firstly matching a plurality of characters in the lyrics with the plurality of pronunciations, wherein it can be seen that the pronunciation corresponding to "幻想" is an intermediate pronunciation, having a Roman pronunciation of "fantaji";

**[0084]** b. determining three candidate pronunciations corresponding to the Roman pronunciation "fantaji" being: "ふぁんたじ", "ファンタジ" and "ファンタジー";

**[0085]** c. with respect to each candidate pronunciation, looking up an intermediate character corresponding to each candidate pronunciation in the first dictionary engine, and looking up a semantic meaning of the intermediate character from the second dictionary engine to determine the candidate pronunciation "ファンタジー", which corresponds to an English character "Fantasy" in the Japanese-English dictionary;

**[0086]** d. looking up the semantic meaning of each intermediate character from the second dictionary engine, semantically matching the semantic meaning of each intermediate character with the semantic meaning of "幻想", and determining an English character "Fantasy" corresponding to the candidate pronunciation "ファンタジー", wherein the corresponding semantic meaning of "Fantasy" in the second dictionary engine may be "幻想", which is the same as that of "幻想" in Japanese; and

**[0087]** e. determining the pronunciation corresponding to "幻想" is "ファンタジー".

**[0088]** Through the above steps a-e, the foreign language is used as an intermediate character, and furigana "ファンタジー" having the same semantic meaning as "幻想" is further determined from a plurality of furigana by looking up a dictionary engine established in advance.

**[0089]** It should be noted that the server, by using a foreign language as an intermediate character, first converts the pronunciation of the character to be annotated into an intermediate character based on a plurality of possible candidate pronunciations corresponding to ambiguous Roman pronunciations, and looks up the semantic meaning corresponding to the intermediate character and match a candidate pronunciation that is closest to the semantic meaning of the character to be annotated from a plurality of candidate pronunciations, thereby further analyzing the pronunciation of the character to be annotated after the artistic evolution during actual singing, and greatly improving the accuracy of determining the pronunciation of the character to be annotated.

**[0090]** In 206, the server generates a first lyric file of the target song according to the corresponding pronunciations of the plurality of characters and the character to be annotated in the target song.

**[0091]** In this step, the server creates the first lyric file, store the corresponding pronunciations of the plurality of characters and the character to be annotated in the target song into the first lyric file, and establishes an index relationship between the character to be annotated and the corresponding pronunciation of the character to be annotated in the target song. The server may establish an index relationship between the character and the pronunciation based on arrays. The server establishes a second array, and writes the corresponding pronunciation of the character to be annotated in the target song into the second array; and add the first array and the second array to the first lyric file. With respect to each character to be annotated, the server stores the pronunciation corresponding to the character to be annotated into a second storage position of the third array according to a first storage position of the character to be annotated in the first array, wherein the first storage position and the third storage position may both be associated bytes (for example, first bytes) in the first array and the second array respectively.

**[0092]** In this embodiment of the present invention, in step 201, the server may also acquire display times of the plurality of characters. In this step, the server may also add the display times to the first lyric file. Therefore, the step of generating the first lyric file of the target song by the server according to the corresponding pronunciations of the plurality of characters and the character to be annotated in the target song may also be as follows: the server creates the first lyric file, establishes the second array, and writes the

corresponding pronunciation of the character to be annotated in the target song into the second array; and the server adds the first array, the second array, and the third array to the first lyric file.

**[0093]** In a possible design, in some languages, there may be cases that the number of characters is not equal to the number of pronunciations. For example, two characters included in "欠片" in Japanese actually correspond to three pronunciations. Therefore, when a target word exists in the lyrics, the server may also determine a display time of the target word according to display times of at least two adjacent characters to be annotated included in the target word, and update the third array according to the display time of the target word. Finally, the server adds the first array, the second array, and the updated third array to the first lyric file. The target word includes at least two adjacent characters to be annotated, and the number of the included characters to be annotated is not equal to the number of pronunciations of the target word. The manner of for the server to determine the display time of the target word is as follows: merging the display times of at least two adjacent characters to be annotated included in the target word, and determining the merged display time as the display time of the target word.

**[0094]** In Japanese, the number of pronunciations of many words is not equal to the number of characters they include, for example, 欠片, 明後日, 一昨日, 黄昏, 彷徨, 躊躇, 煙草, 大人 and the like. Taking 欠片 as an example, before the third array is updated, the display times stored in the second lyric file is shown in Table 1 below.

Table 1

|             |       |       |       |       |
|-------------|-------|-------|-------|-------|
| Third array | 39112 | 39803 | 40356 | 41176 |
| First array | 欠     | 片     | の     | 色     |

**[0095]** Obviously, "欠" and "片" each correspond to a display time. However, in the actual display, since "欠" and "片" jointly correspond to three pronunciations, "欠" and "片" are displayed separately according to two display times, which will obviously cause a display error. By merging the two display times, the display time stored in the updated third array is obtained as shown in Table 2 below.

Table 2

|              |       |   |       |       |
|--------------|-------|---|-------|-------|
| Third array  | 39112 |   | 40356 | 41176 |
| First array  | 欠     | 片 | の     | 色     |
| Second array | かけら   |   |       | いろ    |

**[0096]** It should be noted that the server merges the display times of a plurality of adjacent characters in the target word, so that the plurality of adjacent characters can be used as one word, corresponding to one separate display time, thereby avoiding the problem that the display time of each character does not match the pronunciation in actual singing because a song is displayed based on each character corresponding to one display time in actual playing. Further, the display time of each character matches the actual pronunciation of this character, and the display time of each word matches the pronunciation of this word. Therefore, the display times of the lyrics in the first lyric file can accurately match the pronunciations of the lyrics, thereby further improving the accuracy of the finally obtained first lyric file.

**[0097]** In the embodiments of the present invention, a server may determine, based on the lyrics of the target song and the audio file of the target song, the corresponding pronunciation of each character to be annotated in the lyrics in the target song, and may further generate the first lyric file of the target song according to the pronunciations of the plurality of characters and the character to be annotated, thereby binding a corresponding pronunciation for each character to be annotated. Therefore, in the subsequent display of the lyrics, the pronunciations can be displayed synchronously to ensure that the user can sing each character of the target song correctly.

**[0098]** FIG. 3 is a flowchart of a method for displaying lyrics according to an embodiment of the present invention. An executive main body of this embodiment of the present invention is a terminal. Referring to FIG. 3, the method includes the following steps.

**[0099]** In 301, when a lyric display instruction is received, the terminal acquires a first lyric file of a target song.

**[00100]** In this embodiment of the present invention, the lyric display instruction is configured to display lyrics of the target song. When the terminal receives the lyric

display instruction, the terminal may acquire the first lyric file of the target song from a local side or a server according to an identifier of the target song. The lyric display instruction may be obtained when a user triggers the terminal to play the target song, or may be obtained when the user triggers the display of the lyric file.

**[00101]** It should be noted that the first lyric file is a lyric file generated based on lyrics and the target pronunciations of a plurality of characters in advance. The specific generation process is shown in the above steps 201-206. The first lyric file includes at least the lyric of the target song, and a corresponding pronunciation of a character to be annotated among the plurality of characters of the lyrics in the target song.

**[00102]** In step 302, the terminal acquires, from the first lyric file, the lyrics of the target song and a corresponding pronunciation of a character to be annotated among a plurality of characters of the lyrics in the target song.

**[00103]** In this embodiment of the present invention, in the first lyric file, the lyrics and the pronunciation of the character to be annotated may be stored in a form of arrays, respectively. The terminal acquires a first array and a second array from the first lyric file, reads a plurality of characters in the lyrics from the first array, and reads a pronunciation of the character to be annotated among the plurality of characters from the second array. The terminal may determine a second storage position associated with the second array based on a first storage position of each character to be annotated in the first array, and read the pronunciation of the character to be annotated from the second storage position.

**[00104]** In a possible design, the first lyric file may further include display times of the plurality of characters in the lyrics. The terminal may also acquire the display times of the plurality of characters from the first lyric file. This process may be as follows: the terminal acquires a third array from the first lyric file, and acquires the display times of the plurality of characters from the third array. The display times of the plurality of characters include a display time of a target word in the lyrics and a display time of each character other than the target word.

**[00105]** The target word includes at least two adjacent characters to be annotated, and the number of the included characters to be annotated is not equal to the number of pronunciations of the target word. The display time of the target word is a merged display

time of the display times of at least two adjacent characters to be annotated included in the target word.

**[00106]** In 303, the terminal displays the plurality of characters of the lyrics.

**[00107]** In this embodiment of the present invention, the terminal highlights a currently played target word in the plurality of characters according to the display time of the target word; and highlights a currently played character among the characters other than the target word according to the display time of each character other than the target word. The terminal may highlight the character or target word being played by a font color of the character. For example, the characters or target word that has been played are rendered as a first color, and the characters or target word that has not been played are rendered as a second color.

**[00108]** In 304, in displaying the character to be annotated, the terminal marks, at a target position of the character to be annotated, the corresponding pronunciation of the character to be annotated in the target song.

**[00109]** In this embodiment of the present invention, the target position is located above the character to be annotated. In displaying the character to be annotated, the terminal highlights the currently played character to be annotated and the corresponding pronunciation of the currently played character to be annotated in the target song according to the display time of the character to be annotated. Of course, the character to be annotated may include the target word. The terminal may highlight the currently played target word and the corresponding pronunciation of the currently played target word in the target song according to the display time of the target word.

**[00110]** It should be noted that, in some prior arts, in order to simplify the logic, unique pronunciations of the lyrics are annotated with brackets as shown in FIG. 4. For example, such annotating "誰でももし男女（ひと）が人間（ひと）でありたいなら" will cause some misunderstanding to users. For example, the users may think that "女" is pronounced as "ひと" alone, and the pronunciation of "男" is unknown. Especially for compound words, for example, 時間旅行（タイムトリップ）, the user cannot distinguish whether the pronunciation in the brackets is "旅行" or "時間旅行" after browsing. Meanwhile, in the case that the lyrics are highlighted, there is no guarantee that the characters and pronunciations will be highlighted at the same display pace. For

example, 男女 (ひと) have two syllables only, but becomes six syllables (including brackets) in the case of corresponding to a line of text. When "男" is highlighted based on the display time of "男", "男女" needs to be rendered as a first color; and when "女" is highlighted, the remaining portion "ひと" is correspondingly rendered as the first color, thereby resulting in a display error and bringing poor user experience to the user.

**[00111]** In this embodiment, as shown in FIG. 5, the terminal may display the pronunciation above the character to be annotated. When the user browses the character to be annotated, the pronunciation of the character to be annotated can be found clearly and accurately, thereby improving the efficiency of the user in knowing the pronunciation of the character to be annotated, and improving the accuracy of the lyric display. Meanwhile, the user sings based on the pronunciation to ensure that the pronunciation of each character can be sung accurately during singing.

**[00112]** In this embodiment of the present invention, when receiving the lyric display instruction, the terminal may display the lyrics of the target song and the corresponding pronunciation of the character to be annotated among the plurality of characters of the lyrics in the target song, thereby ensuring that a user can sing the pronunciation of each character accurately during singing. In addition, the terminal may also mark the pronunciation above the corresponding character to be annotated, so that the pronunciation is clearly visible, and the user can accurately and quickly find the corresponding pronunciation of the character to be annotated, thereby improving the accuracy of lyrics display.

**[00113]** FIG. 6 is a schematic structural diagram of an apparatus for generating lyrics provided by an embodiment of the present invention. Referring to FIG. 6, the apparatus includes: an acquisition module 601, a determination module 602, and a generation module 603.

**[00114]** The acquisition module 601 is configured to acquire lyrics of a target song and an audio file of the target song, the audio file including audio clips corresponding to a plurality of characters in the lyrics.

**[00115]** The determination module 602 is configured to determine, according to a character to be annotated among the plurality of characters and audio clips corresponding to the plurality of characters, a candidate pronunciation of the character to be annotated.

**[00116]** The determination module 602 is further configured to, if the character to be annotated has one candidate pronunciation, determine the candidate pronunciation as the corresponding pronunciation of the character to be annotated in the target song.

**[00117]** The determination module 602 is further configured to, if the character to be annotated has at least two candidate pronunciations, determine the candidate pronunciation semantically matching the character to be annotated as the corresponding pronunciation of the character to be annotated in the target song.

**[00118]** The generation module 603 is configured to generate a first lyric file of the target song according to the corresponding pronunciations of the plurality of characters and the character to be annotated in the target song.

**[00119]** Optionally, the determination module 602 includes:

**[00120]** a determination unit configured to determine a Chinese character among the plurality of characters as the character to be annotated;

**[00121]** an identification unit configured to identify the pronunciations of the plurality of characters according to the audio clips corresponding to the plurality of characters, wherein

**[00122]** the determination unit is further configured to determine the candidate pronunciation of the character to be annotated from the pronunciations of the plurality of characters according to the character to be annotated among the plurality of characters, or determine an intermediate pronunciation of the character to be annotated from the pronunciations of the plurality of characters according to the character to be annotated among the plurality of characters, and recognize at least two candidate pronunciations of the character to be annotated according to the intermediate pronunciation.

**[00123]** Optionally, the determination module 602 includes:

**[00124]** a lookup unit configured to, for each candidate pronunciation, look up an intermediate character corresponding to each candidate pronunciation from a first dictionary engine, and look up a semantic meaning of the intermediate character from a second dictionary engine, the first dictionary engine including a plurality of candidate

pronunciations and intermediate characters corresponding to the plurality of candidate pronunciations, and the second dictionary engine including a plurality of intermediate characters and a plurality of semantic meanings corresponding to the plurality of intermediate characters; and

**[00125]** a selection unit configured to select, from the at least two candidate pronunciations, the candidate pronunciation of which the semantic meaning of the intermediate character matches the semantic meaning of the character to be annotated according to the semantic meaning of the character to be annotated, and determine the selected candidate pronunciation as the corresponding pronunciation of the character to be annotated in the target song.

**[00126]** Optionally, the acquisition module 601 is configured to, acquire a second lyric file of the target song, the second lyric file including lyrics of the target song and display times of a plurality of characters of the lyrics; establish a first array and a third array, write the plurality of characters into the first array, write the display times of the plurality of characters into the third array, and determine the plurality of characters in the first array as the lyrics of the target song; and acquire an original audio file of the target song, extract audio clips in a human voice band in the original audio file, and generate the audio file.

**[00127]** Optionally, the generation module 603 is configured to establish a second array, and write the corresponding pronunciation of the character to be annotated in the target song into the second array; when a target word exists in the lyrics, determine a display time of the target word according to the display times of at least two adjacent characters to be annotated included in the target word, the target word including at least two adjacent characters to be annotated, and the number of the included characters to be annotated being not equal to the number of pronunciations of the target word; write the display time of the target word into the third array; and add the first array, the second array, and the third array into the first lyric file.

**[00128]** The manner of determining the display time of the target word is as follows: merging the display times of at least two adjacent characters to be annotated included in the target word, and determining the merged display time as the display time of the target word.

**[00129]** Optionally, the generation module 603 is configured to establish the second array, and write the corresponding pronunciation of the character to be annotated in the target song into the second array; and add the first array and the second array to the first lyric file, the first array being configured to store a plurality of characters of the lyrics.

**[00130]** In the embodiments of the present invention, a server may determine, based on the lyrics of the target song and the audio file of the target song, the corresponding pronunciation of each character to be annotated of the lyrics in the target song, and may further generate the first lyric file of the target song according to the pronunciations of the plurality of characters and the character to be annotated, thereby binding a corresponding pronunciation for each character to be annotated. Therefore, in the subsequent display of the lyrics, the pronunciations can be displayed synchronously to ensure that the user can sing each character of the target song correctly.

**[00131]** FIG. 7 is a schematic structural diagram of an apparatus for displaying lyrics provided by an embodiment of the present invention. The apparatus is applied to a terminal. Referring to FIG. 7, the apparatus includes: an acquisition module 701, a display module 702, and an annotating module 703.

**[00132]** The acquisition module 701 is configured to, when a lyric display instruction is received, acquire a first lyric file of a target song, the lyric display instruction being configured to display lyrics of the target song.

**[00133]** The acquisition module 701 is further configured to acquire, from the first lyric file, the lyrics of the target song and a corresponding pronunciation of a character to be annotated among a plurality of characters of the lyrics in the target song.

**[00134]** The display module 702 is configured to display the plurality of characters of the lyrics.

**[00135]** The annotating module 703 is configured to, in displaying the character to be annotated, mark, at a target position of the character to be annotated, the corresponding pronunciation of the character to be annotated in the target song, the target position being located above the character to be annotated.

**[00136]** Optionally, the display module 702 includes:

**[00137]** a first display unit configured to highlight a currently played target word in the plurality of characters according to the display time of the target word in the plurality of characters; and

**[00138]** a second display unit configured to highlight a currently played character among the characters other than the target word according to the display time of each character other than the target word among the plurality of characters.

**[00139]** The target word includes at least two adjacent characters to be annotated, and the number of the included characters to be annotated is not equal to the number of pronunciations of the target word. The display time of the target word is a merged display time of the display times of at least two adjacent characters to be annotated included in the target word.

**[00140]** Optionally, the second display unit is configured to, in displaying the character to be annotated, highlight the currently played character to be annotated, and the corresponding pronunciation of the currently played character to be annotated in the target song.

**[00141]** In this embodiment of the present invention, when receiving the lyric display instruction, the terminal may display the lyrics of the target song and the corresponding pronunciation of the character to be annotated among the plurality of characters of the lyrics in the target song, thereby ensuring that a user can sing the pronunciation of each character accurately during singing. In addition, the terminal may also mark the pronunciation above the corresponding character to be annotated, so that the pronunciation is clearly visible, and the user can accurately and quickly find the corresponding pronunciation of the character to be annotated, thereby improving the accuracy of displaying lyrics.

**[00142]** All the above-mentioned optional technical solutions may be used in any combination to form optional embodiments of the present invention, which will not be described in detail here.

**[00143]** It should be noted that the apparatus for generating lyrics or the apparatus for displaying lyrics provided by the present embodiment only takes division of all the functional modules as an example for explanation when generating or displaying lyrics.

In practice, the above functions can be finished by the different functional modules as required. That is, the internal structure of the apparatus is divided into different functional modules to finish all or part of the functions described above. In addition, the apparatus for generating lyrics and the apparatus for displaying lyrics provided by the present embodiment have the same concept as the method for generating lyrics and the method for displaying lyrics provided by the foregoing embodiment. Refer to the method embodiment for the specific implementation process of the apparatus, which will not be repeated herein.

**[00144]** FIG. 8 is a structural block diagram of a terminal in accordance with one embodiment of the present invention. The terminal 800 may be a smart phone, a tablet computer, an MP3 (Moving Picture Experts Group Audio Layer III) player, an MP4 (Moving Picture Experts Group Audio Layer IV) player, a laptop or desk computer. The terminal 800 may also be called a UE (User Equipment), a portable terminal, a laptop terminal, a desk terminal, etc.

**[00145]** Generally, the terminal 800 includes a processor 801 and a memory 802.

**[00146]** The processor 801 may include one or more processing cores, such as a 4-core processor and an 8-core processor. The processor 801 may be formed by at least one hardware of a DSP (Digital Signal Processing), an FPGA (Field-Programmable Gate Array), and a PLA (Programmable Logic Array). The processor 801 may also include a main processor and a coprocessor. The main processor is a processor for processing the data in an awake state, and is also called a CPU (Central Processing Unit). The coprocessor is a low-power-consumption processor for processing the data in a standby state. In some embodiments, the processor 801 may be integrated with a GPU (Graphics Processing Unit), which is configured to render and draw the content that needs to be displayed by a display screen. In some embodiments, the processor 801 may also include an AI (Artificial Intelligence) processor configured to process computational operations related to machine learning.

**[00147]** The memory 802 may include one or more computer-readable storage mediums, which can be non-transitory. The memory 802 may also include a high-speed random access memory, as well as a non-volatile memory, such as one or more disk

storage devices and flash storage devices. In some embodiments, the non-transitory computer-readable storage medium in the memory 802 is configured to store at least one instruction. The at least one instruction is configured to be executed by the processor 801 to implement the method for generating lyrics and the method for displaying lyrics provided by the method embodiments of the present invention.

**[00148]** In some embodiments, the terminal 800 also optionally includes a peripheral device interface 803 and at least one peripheral device. The processor 801, the memory 802, and the peripheral device interface 803 may be connected by a bus or a signal line. Each peripheral device may be connected to the peripheral device interface 803 by a bus, a signal line or a circuit board. For example, the peripheral device includes at least one of a radio frequency circuit 804, a touch display screen 805, a camera 806, an audio circuit 807, a positioning component 808 and a power source 809.

**[00149]** The peripheral device interface 803 may be configured to connect at least one peripheral device associated with an I/O (Input/Output) to the processor 801 and the memory 802. In some embodiments, the processor 801, the memory 802 and the peripheral device interface 803 are integrated on the same chip or circuit board. In some other embodiments, any one or two of the processor 801, the memory 802 and the peripheral device interface 803 may be implemented on a separate chip or circuit board, which is not limited in the present embodiment.

**[00150]** The radio frequency circuit 804 is configured to receive and transmit an RF (Radio Frequency) signal, which is also referred to as an electromagnetic signal. The radio frequency circuit 804 communicates with a communication network and other communication devices via the electromagnetic signal. The radio frequency circuit 804 converts the electrical signal into the electromagnetic signal for transmission, or converts the received electromagnetic signal into the electrical signal. Optionally, the radio frequency circuit 804 includes an antenna system, an RF transceiver, one or more amplifiers, a tuner, an oscillator, a digital signal processor, a codec chipset, a subscriber identity module card, and the like. The radio frequency circuit 804 can communicate with other terminals via at least one wireless communication protocol. The wireless communication protocol includes, but not limited to, the World Wide Web, a metropolitan area network, an intranet, various generations of mobile communication

networks (2G, 3G, 4G, and 5G), a wireless local area network, and/or a WiFi (Wireless Fidelity) network. In some embodiments, the RF circuit 804 may also include NFC (Near Field Communication) related circuits, which is not limited in the present invention.

**[00151]** The display screen 805 is configured to display a UI (User Interface). The UI may include graphics, text, icons, videos, and any combination thereof. When the display screen 805 is a touch display screen, the display screen 805 also has the capacity to acquire touch signals on or over the surface of the display screen 805. The touch signal may be input into the processor 801 as a control signal for processing. At this time, the display screen 805 may also be configured to provide virtual buttons and/or virtual keyboards, which are also referred to as soft buttons and/or soft keyboards. In some embodiments, one display screen 805 may be disposed on the front panel of the terminal 800. In some other embodiments, at least two display screens 805 may be disposed respectively on different surfaces of the terminal 800 or in a folded design. In further embodiments, the display screen 805 may be a flexible display screen disposed on the curved or folded surface of the terminal 800. Even the display screen 805 may have an irregular shape other than a rectangle; that is, the display screen 805 may be an irregular-shaped screen. The display screen 805 may be an LCD (liquid Crystal Display) screen, an OLED (Organic Light-Emitting Diode) screen or the like.

**[00152]** The camera component 806 is configured to capture images or videos. In some embodiments of the present invention, the camera component 806 includes a front camera and a rear camera. Usually, the front camera is placed on the front panel of the terminal, and the rear camera is placed on the back of the terminal. In some embodiments, at least two rear cameras are disposed, and are at least one of a main camera, a depth-of-field camera, a wide-angle camera, and a telephoto camera respectively, so as to realize a background blurring function achieved by fusion of the main camera and the depth-of-field camera, panoramic shooting and VR (Virtual Reality) shooting functions achieved by fusion of the main camera and the wide-angle camera or other fusion shooting functions. In some embodiments, the camera component 806 may also include a flashlight. The flashlight may be a mono-color temperature flashlight or a two-color temperature flashlight. The two-color temperature flash is a combination of a warm

flashlight and a cold flashlight and can be used for light compensation at different color temperatures.

**[00153]** The audio circuit 807 may include a microphone and a speaker. The microphone is configured to collect sound waves of users and environments, and convert the sound waves into electrical signals which are input into the processor 801 for processing, or input into the RF circuit 804 for voice communication. For the purpose of stereo acquisition or noise reduction, there may be a plurality of microphones respectively disposed at different locations of the terminal 800. The microphone may also be an array microphone or an omnidirectional acquisition microphone. The speaker is then configured to convert the electrical signals from the processor 801 or the radio frequency circuit 804 into the sound waves. The speaker may be a conventional film speaker or a piezoelectric ceramic speaker. When the speaker is the piezoelectric ceramic speaker, the electrical signal can be converted into not only human-audible sound waves but also the sound waves which are inaudible to humans for the purpose of ranging and the like. In some embodiments, the audio circuit 807 may also include a headphone jack.

**[00154]** The positioning component 808 is configured to locate the current geographic location of the terminal 800 to implement navigation or LBS (Location Based Service). The positioning component 808 may be a positioning component based on the American GPS (Global Positioning System), the Chinese Beidou system, the Russian Grenas System or the European Galileo system.

**[00155]** The power source 809 is configured to power up various components in the terminal 800. The power source 809 may be alternating current, direct current, a disposable battery, or a rechargeable battery. When the power source 809 includes the rechargeable battery, the rechargeable battery may be a wired rechargeable battery or a wireless rechargeable battery. The rechargeable battery may also support the fast charging technology.

**[00156]** In some embodiments, terminal 800 also includes one or more sensors 810. The one or more sensors 810 include, but not limited to, an acceleration sensor 811, a gyro sensor 812, a pressure sensor 813, a fingerprint sensor 814, an optical sensor 815 and a proximity sensor 816.

**[00157]** The acceleration sensor 811 may detect magnitudes of accelerations on three coordinate axes of a coordinate system established by the terminal 800. For example, the acceleration sensor 811 may be configured to detect components of a gravitational acceleration on the three coordinate axes. The processor 801 may control the touch display screen 805 to display a user interface in a landscape view or a portrait view according to a gravity acceleration signal collected by the acceleration sensor 811. The acceleration sensor 811 may also be configured to collect motion data of a game or a user.

**[00158]** The gyro sensor 812 can detect a body direction and a rotation angle of the terminal 800, and can cooperate with the acceleration sensor 811 to collect a 3D motion of the user on the terminal 800. Based on the data collected by the gyro sensor 812, the processor 801 can serve the following functions: motion sensing (such as changing the UI according to a user's tilt operation), image stabilization during shooting, game control and inertial navigation.

**[00159]** The pressure sensor 813 may be disposed on a side frame of the terminal 800 and/or a lower layer of the touch display screen 805. When the pressure sensor 813 is disposed on the side frame of the terminal 800, a user's holding signal to the terminal 800 can be detected. The processor 801 can perform left-right hand recognition or quick operation according to the holding signal collected by the pressure sensor 813. When the pressure sensor 813 is disposed on the lower layer of the touch display screen 805, the processor 801 controls an operable control on the UI according to a user's pressure operation on the touch display screen 805. The operable control includes at least one of a button control, a scroll bar control, an icon control and a menu control.

**[00160]** The fingerprint sensor 814 is configured to collect a user's fingerprint. The processor 801 identifies the user's identity based on the fingerprint collected by the fingerprint sensor 814, or the fingerprint sensor 814 identifies the user's identity based on the collected fingerprint. When the user's identity is identified as trusted, the processor 801 authorizes the user to perform related sensitive operations, such as unlocking the screen, viewing encrypted information, downloading software, paying, and changing settings. The fingerprint sensor 814 may be provided on the front, back, or side of the terminal 800. When the terminal 800 is provided with a physical button or a

manufacturer's Logo, the fingerprint sensor 814 may be integrated with the physical button or the manufacturer's Logo.

**[00161]** The optical sensor 815 is configured to collect ambient light intensity. In one embodiment, the processor 801 may control the display brightness of the touch display screen 805 according to the ambient light intensity collected by the optical sensor 815. For example, when the ambient light intensity is high, the display brightness of the touch display screen 805 is increased; and when the ambient light intensity is low, the display brightness of the touch display screen 805 is decreased. In another embodiment, the processor 801 may also dynamically adjust shooting parameters of the camera component 806 according to the ambient light intensity collected by the optical sensor 815.

**[00162]** The proximity sensor 816, also referred to as a distance sensor, is usually disposed on the front panel of the terminal 800. The proximity sensor 816 is configured to capture a distance between the user and a front surface of the terminal 800. In one embodiment, when the proximity sensor 816 detects that the distance between the user and the front surface of the terminal 800 becomes gradually smaller, the processor 801 controls the touch display screen 805 to switch from a screen-on state to a screen-off state. When it is detected that the distance between the user and the front surface of the terminal 800 gradually increases, the processor 801 controls the touch display screen 805 to switch from the screen-off state to the screen-on state.

**[00163]** It will be understood by those skilled in the art that the structure shown in FIG. 8 does not constitute a limitation to the terminal 800, and may include more or less components than those illustrated, or combine some components or adopt different component arrangements.

**[00164]** Fig. 9 is a block diagram showing a server according to an embodiment of the present invention. For example, the server 900 may vary significantly in configuration or capabilities, and may include one or more processors 901 (for example, one or more central processing units (CPUs)), and one or more memories 902. The one or more memories 902 store at least one instruction which may be loaded and executed by the processor 901 to realize the methods for generating lyrics provided by the above embodiments. Of course, the server may also include a wired or wireless interface, a

keyboard, an input/output interface and the like to facilitate input/output, and other components for realizing functions of the apparatus, which will not be described in detail herein.

**[00165]** In one exemplary embodiment, there is also provided a computer-readable storage medium, e.g., a memory including instructions. The instructions may be executed by the processor in the terminal to finish the method for generating lyrics or the method for displaying lyrics provided by the above embodiments. For example, computer-readable storage medium may be a read only memory (ROM), a random access memory (RAM), a CD-ROM, a magnetic tape, a floppy disk, an optical data storage device, or the like.

**[00166]** Persons of ordinary skill in the art can understand that all or part of the steps described in the above embodiments can be completed through hardware, or through relevant hardware instructed by applications stored in a non-transitory computer readable storage medium, such as a read-only memory, a disk or a CD, etc.

**[00167]** The foregoing descriptions are merely exemplary embodiments of the present invention, and are not intended to limit the present invention. Within the spirit and principles of the disclosure, any modifications, equivalent substitutions, improvements, etc., are within the protection scope of the present invention.

## CLAIMS

1. A method for generating lyrics, comprising:

acquiring lyrics of a target song and an audio file of the target song;

determining, according to a character to be annotated among a plurality of characters in the lyrics and audio clips corresponding to the plurality of characters in the audio file, a candidate pronunciation of the character to be annotated;

if the character to be annotated has one candidate pronunciation, determining the candidate pronunciation as a corresponding pronunciation of the character to be annotated in the target song;

if the character to be annotated has at least two candidate pronunciations, determining a candidate pronunciation semantically matching the character to be annotated as the corresponding pronunciation of the character to be annotated in the target song; and

generating a first lyric file of the target song according to the plurality of characters and the corresponding pronunciation of the character to be annotated in the target song.

2. The method according to claim 1, wherein determining, according to the character to be annotated among the plurality of characters in the lyrics and the audio clips corresponding to the plurality of characters in the audio file, the candidate pronunciation of the character to be annotated comprises:

determining a Chinese character among the plurality of characters as the character to be annotated;

identifying pronunciations of the plurality of characters according to the audio clips corresponding to the plurality of characters;

determining the candidate pronunciation of the character to be annotated from the pronunciations of the plurality of characters according to the character to be annotated among the plurality of characters, or determining an intermediate pronunciation of the character to be annotated from the pronunciation of the plurality of characters according to the character to be annotated among the plurality of characters, and identifying at least two candidate pronunciations of the character to be annotated according to the intermediate pronunciation.

3. The method according to claim 1, wherein if the character has at least two candidate pronunciations, determining the candidate pronunciation semantically matching the character as the corresponding pronunciation of the character to be annotated in the target song comprises:

for each candidate pronunciation, looking up an intermediate character corresponding to each candidate pronunciation from a first dictionary engine, and looking up a semantic meaning of the intermediate character from a second dictionary engine, the first dictionary engine comprising a plurality of candidate pronunciations and the intermediate characters corresponding to the plurality of candidate pronunciations, and the second dictionary engine comprising a plurality of intermediate characters and a plurality of semantic meanings corresponding to the plurality of intermediate characters; and

selecting, from the at least two candidate pronunciations, a candidate pronunciation of which the semantic meaning of the intermediate character matches the semantic meaning of the character to be annotated according to the semantic meaning of the character to be annotated, and determining the selected candidate pronunciation as the corresponding pronunciation of the character to be annotated in the target song.

4. The method according to claim 1, wherein acquiring the lyrics of the target song and the audio file of the target song comprises:

acquiring a second lyric file of the target song, the second lyric file comprising the lyrics of the target song and display times of the plurality of characters of the lyrics;

establishing a first array and a third array, writing the plurality of characters into the first array, writing the display times of the plurality of characters into the third array, and determining the plurality of characters in the first array as the lyrics of the target song; and

acquiring an original audio file of the target song, extracting audio clips in a human voice band in the original audio file, and generating the audio file.

5. The method according to claim 4, wherein generating the first lyric file of the target song according to the corresponding pronunciations of the plurality of characters and the character to be annotated in the target song comprises:

establishing a second array, and writing the corresponding pronunciation of the character to be annotated in the target song into the second array;

when a target word exists in the lyrics, determining a display time of the target word according to display times of at least two adjacent characters to be annotated comprised in the target word, the target word comprising at least two adjacent characters to be annotated, and the number of the comprised characters to be annotated being not equal to the number of pronunciations of the target word;

writing the display time of the target word into the third array; and

adding the first array, the second array, and the third array to the first lyric file, wherein

the manner of determining the display time of the target word is as follows: merging the display times of at least two adjacent characters to be annotated included in the target word, and determining the merged display time as the display time of the target word.

6. The method according to claim 1, wherein generating the first lyric file of the target song according to the corresponding pronunciations of the plurality of characters and the character to be annotated in the target song comprises:

establishing the second array, and writing the corresponding pronunciation of the character to be annotated in the target song into the second array; and

adding the first array and the second array to the first lyric file, the first array being configured to store the plurality of characters of the lyrics.

7. A method for displaying lyrics, wherein the method is applied to a terminal and comprises:

when a lyric display instruction is received, acquiring a first lyric file of a target song, the lyric display instruction being configured to display lyrics of the target song;

acquiring, from the first lyric file, the lyrics of the target song and a corresponding pronunciation of a character to be annotated among a plurality of characters of the lyrics in the target song;

displaying the plurality of characters of the lyrics; and

in displaying the character to be annotated, annotating, at a target position of the character to be annotated, the corresponding pronunciation of the character to be annotated in the target song, the target position being above the character to be annotated.

8. The method according to claim 7, wherein the first lyric file further comprises display times of the plurality of characters; and correspondingly, displaying the plurality of characters of the lyrics comprises:

highlighting a currently played target word in the plurality of characters according to a display time of the target word in the plurality of characters; and

highlighting a currently played character among characters other than the target word according to a display time of each character other than the target word among the plurality of characters, wherein

the target word comprises at least two adjacent characters to be annotated, and the number of the comprised characters to be annotated is not equal to the number of pronunciations of the target word; and the display time of the target word is a merged display time of display times of at least two adjacent characters to be annotated comprised in the target word.

9. The method according to claim 8, wherein in displaying the character to be annotated, annotating, at the target position of the character to be annotated, the corresponding pronunciation of the character to be annotated in the target song comprises:

in displaying the character to be annotated, highlighting the currently played character to be annotated and the corresponding pronunciation of the currently played character to be annotated in the target song.

10. An apparatus for generating lyrics, comprising:

an acquisition module configured to acquire lyrics of a target song and an audio file of the target song;

a determination module configured to determine, according to a character to be annotated among a plurality of characters in the lyrics and audio clips corresponding to the plurality of characters in the audio file, a candidate pronunciation of the character to be annotated;

the determination module being further configured to, if the character to be annotated has one candidate pronunciation, determine the candidate pronunciation as a corresponding pronunciation of the character to be annotated in the target song;

the determination module being further configured to, if the character to be annotated has at least two candidate pronunciations, determine a candidate pronunciation semantically matching the character as the corresponding pronunciation of the character to be annotated in the target song; and

a generation module configured to generate a first lyric file of the target song according to the corresponding pronunciations of the plurality of characters and the character to be annotated in the target song.

11. The apparatus according to claim 10, wherein the determination module comprises:

a determination unit configured to determine a Chinese character among the plurality of characters as the character to be annotated;

an identification unit configured to identify pronunciations of the plurality of characters according to the audio clips corresponding to the plurality of characters, wherein

the determination unit is further configured to determine the candidate pronunciation of the character to be annotated from the pronunciations of the plurality of characters according to the character to be annotated among the plurality of characters, or determine an intermediate pronunciation of the character to be annotated from the pronunciations of the plurality of characters according to the character to be annotated among the plurality of characters, and identify at least two candidate pronunciations of the character to be annotated according to the intermediate pronunciation.

12. The apparatus according to claim 10, wherein the determination module comprises:

a lookup unit configured to, for each candidate pronunciation, look up an intermediate character corresponding to each candidate pronunciation from a first dictionary engine, and look up a semantic meaning of the intermediate character from a second dictionary engine, the first dictionary engine comprising a plurality of candidate pronunciations and intermediate characters corresponding to the plurality of candidate pronunciations, and the second dictionary engine comprising a plurality of intermediate characters and a plurality of semantic meanings corresponding to the plurality of intermediate characters; and

a selection unit configured to select, from the at least two candidate pronunciations, the candidate pronunciation of which the semantic meaning of the intermediate character matches the semantic meaning of the character to be annotated according to the semantic meaning of the character to be annotated, and determine the selected candidate pronunciation as the corresponding pronunciation of the character to be annotated in the target song.

13. An apparatus for displaying lyrics, which is applied to a terminal and comprises:

an acquisition module configured to, when a lyric display instruction is received, acquire a first lyric file of a target song, the lyric display instruction being configured to display lyrics of the target song;

the acquisition module being further configured to acquire, from the first lyric file, the lyrics of the target song and a corresponding pronunciation of a character to be annotated among a plurality of characters of the lyrics in the target song;

a display module configured to display the plurality of characters of the lyrics; and

an annotating module configured to, in displaying the character to be annotated, mark, at a target position of the character to be annotated, the corresponding pronunciation of the character to be annotated in the target song, the target position being above the character to be annotated.

14. An electronic device, comprising a processor and a memory, at least one instruction being stored in the memory, and loaded and executed by the processor to implement the operations executed by the method for generating lyrics according to any one of claims 1 to 6, or the method for displaying lyrics according to any one of claims 7 to 9.

15. A computer readable storage medium, at least one instruction being stored in the storage medium, and loaded and executed by a processor to implement the operations of the method for generating lyrics according to any one of claims 1 to 6, or the method for displaying lyrics according to any one of claims 7 to 9.