METHOD FOR GENERATING ACCOMPANIMENT, DEVICE, AND STORAGE MEDIUM

TECHNICAL FIELD

[0001] The disclosure relates to the field of computer application technology, and in particular to a method for generating an accompaniment, a device, and a storage medium.

BACKGROUND

[0002] With the development of virtual reality technology, virtual three-dimensional (3D) audio technology is gradually optimized. Virtual 3D audio technology can create a stereo dynamic effect. Applying virtual 3D audio technology to a singing application can give users an immersive experience. At present, when virtual 3D audio technology is applied to multi-person chorus scenes, in the existing technical solution, weighted superimposing is directly performed on multiple voices. However, the sound under this processing method is not good enough in stereo, resulting in poor user experience.

SUMMARY

[0003] Embodiments of the disclosure provide a method for generating an accompaniment, a device, and a storage medium, which can achieve an all-round stereo audio effect and improve user experience.

[0004] In an aspect, embodiments of the disclosure provide a method for generating an accompaniment as follows. A dry signal set is obtained, where the dry signal set includes x dry signals corresponding to a target song, and x is an integer greater than 1. A virtual sound signal is generated based on a dry signal corresponding to each virtual three-dimensional spatial acoustic image position of N virtual three-dimensional spatial acoustic image positions, where the x dry signals correspond to the N virtual three-dimensional spatial acoustic image positions, N is an integer greater than 1, the N virtual three-dimensional spatial acoustic image positions are different from one another, and each virtual three-dimensional spatial acoustic image position is allowed to correspond to one or more dry signals of the x dry signals. Each virtual sound signal

in a virtual sound signal set is merged to obtain a chorus dry sound, where the virtual sound signal set includes the virtual sound signal at each virtual three-dimensional spatial acoustic image position of the *N* virtual three-dimensional spatial acoustic image positions. Audio effect synthesis is performed on the chorus dry sound and background music of the target song according to a sound effect optimizing rule to obtain an accompaniment of the target song.

[0005] In another aspect, embodiments of the disclosure provide a method for playing and processing an accompaniment as follows. A user interface is displayed for receiving a selection instruction for a target song. An accompaniment corresponding to the target song is obtained when the selection instruction received on the user interface indicates that an accompaniment mode for the target song is a chorus accompaniment mode. The accompaniment corresponding to the target song is played, where the accompaniment is generated according to a chorus dry sound and background music, where the chorus dry sound is generated according to multiple dry signals in a dry signal set, the multiple dry signals in the dry signal set correspond to multiple different virtual three-dimensional spatial acoustic image positions, and the dry signal set is obtained according to dry signals recorded by multiple users for the target song.

[0006] In another aspect, embodiments of the disclosure provide an apparatus for generating an accompaniment, and the apparatus includes an obtaining unit and a processing unit.

[0007] The obtaining unit is configured to obtain a dry signal set, where the dry signal set includes *x* dry signals corresponding to a target song, and *x* is an integer greater than 1. The obtaining unit is further configured to generate a virtual sound signal based on a dry signal corresponding to each virtual three-dimensional spatial acoustic image position of *N* virtual three-dimensional spatial acoustic image positions, where the *x* dry signals correspond to the *N* virtual three-dimensional spatial acoustic image positions, *N* is an integer greater than 1, the *N* virtual three-dimensional spatial acoustic image positions are different from one another, and each virtual three-dimensional spatial acoustic image position is allowed to correspond to one or more dry signals of the *x* dry signals. The processing unit is configured to merge each virtual sound signal in a virtual sound signal set to obtain a chorus dry sound, where the virtual sound signal set includes the virtual sound signal at each virtual three-dimensional spatial acoustic image position of the *N* virtual three-dimensional spatial acoustic image positions. The processing unit is further configured to perform audio effect synthesis on the chorus dry sound and background music of the target song according to a sound effect optimizing rule to obtain an

accompaniment of the target song.

[0008] In another aspect, embodiments of the disclosure provide an apparatus for playing and processing an accompaniment, and the apparatus includes an obtaining unit and a processing unit.

[0009] The obtaining unit is configured to display a user interface for receiving a selection instruction for a target song; and obtain an accompaniment corresponding to the target song when the selection instruction received on the user interface indicates that an accompaniment mode for the target song is a chorus accompaniment mode. The processing unit is configured to play the accompaniment corresponding to the target song, where the accompaniment is generated according to a chorus dry sound and background music, where the chorus dry sound is generated according to multiple dry signals in a dry signal set, the multiple dry signals in the dry signal set correspond to multiple different virtual three-dimensional spatial acoustic image positions, and the dry signal set is obtained according to dry signals recorded by multiple users for the target song.

[0010] Accordingly, embodiments of the disclosure provide an electronic device including a memory, a processor, and a network interface. The processor is connected to the memory and the network interface, where the network interface is configured to provide a network communication function, the memory is configured to store program codes, and the processor is configured to invoke the program codes to perform the method in embodiments of the disclosure.

[0011] Accordingly, embodiments of the disclosure provide a computer-readable storage medium storing a computer program operable with a processor to implement the method in the embodiments of the disclosure.

[0012] Accordingly, embodiments of the disclosure provide a computer program product or a computer program. The computer program product or a computer program includes computer instructions stored in a computer-readable storage medium. A processor of a computer device reads from the computer-readable storage medium and executes the computer instructions, to enable the computer device to perform the method in embodiments of the disclosure.

[0013] In embodiments of the disclosure, on the one hand, virtual sound signals for each dry signal at different virtual three-dimensional spatial acoustic image positions corresponding to a target song in a dry signal set can be obtained. The virtual sound signals corresponding to each dry signal are merged to obtain a chorus dry sound. At last, audio effect synthesis is performed

on the chorus dry sound and background music of the target song according to a sound effect optimizing rule to obtain an accompaniment of the target song. On the other hand, a selection instruction from a user for the target song can be received. When the received selection instruction indicates that an accompaniment mode for the target song is a chorus accompaniment mode, the accompaniment corresponding to the target song can be obtained and played. In this way, acoustic image positions of the dry signals in a virtual three-dimensional space can be simulated comprehensively to achieve a surrounded stereo audio effect, so that users will have an immersive feeling in listening and have an immersive experience when obtaining the corresponding accompaniment.

BRIEF DESCRIPTION OF THE DRAWINGS

[0014] In order to describe technical solutions in embodiments of the disclosure or in related art more clearly, the following will give a brief introduction to the accompanying drawings required for describing embodiments. Apparently, the accompanying drawings hereinafter described are merely some embodiments of the disclosure. Based on these drawings, those of ordinary skill in the art can also obtain other drawings without creative effort.

[0015] FIG. 1 is a schematic diagram illustrating an application scenario of a method for generating an accompaniment provided in embodiments of the disclosure.

[0016] FIG. 2 is a schematic flowchart of a method for generating an accompaniment provided in embodiments of the disclosure.

[0017] FIG. 3a is a schematic diagram illustrating a horizontal plane, an upper plane, and a lower plane in a method for generating an accompaniment provided in embodiments of the disclosure.

[0018] FIG. 3b is a schematic diagram illustrating a virtual three-dimensional acoustic image position in a method for generating an accompaniment provided in embodiments of the disclosure.

[0019] FIG. 3c is a schematic diagram illustrating partition of each plane at intervals of a preset angle in a method for generating an accompaniment provided in embodiments of the disclosure.

[0020] FIG. 4 is a schematic flowchart of another method for generating an accompaniment provided in embodiments of the disclosure.

[0021] FIG. 5 is a schematic flowchart illustrating obtaining of a binaural signal corresponding to a dry signal in a dry signal set in a method for generating an accompaniment provided in embodiments of the disclosure.

[0022] FIG. 6 is a schematic flowchart of a method for playing and processing an accompaniment provided in embodiments of the disclosure.

[0023] FIG. 7a is a schematic flowchart illustrating obtaining of an accompaniment corresponding to a target song in a method for playing and processing an accompaniment provided in embodiments of the disclosure.

[0024] FIG. 7b is a schematic diagram illustrating displaying of a first single-sentence interface in a method for playing and processing an accompaniment provided in embodiments of the disclosure.

[0025] FIG. 7c is a schematic diagram illustrating displaying of a second single-sentence interface in a method for playing and processing an accompaniment provided in embodiments of the disclosure.

[0026] FIG. 8a is a schematic structural diagram of an apparatus for generating an accompaniment provided in embodiments of the disclosure.

[0027] FIG. 8b is a schematic structural diagram of an apparatus for playing and processing an accompaniment provided in embodiments of the disclosure.

[0028] FIG. 9 is a schematic structural diagram of an electronic device provided in embodiments of the disclosure.

DETAILED DESCRIPTION

[0029] In order to make purposes, technical solutions, and advantages of embodiments of the disclosure clearer, technical solutions of embodiments of the disclosure will be described clearly and completely below with reference to the accompanying drawings in the embodiments of the disclosure. Apparently, embodiments described herein are some embodiments, rather than all embodiments, of the disclosure. Based on the embodiments of the disclosure, all other embodiments obtained by those of ordinary skill in the art without creative effort shall fall within the protection scope of the disclosure.

[0030] Before further detailed description of embodiments of the disclosure, the nouns and terms involved in embodiments of the disclosure are explained. The nouns and terms involved in

the embodiments of the disclosure are subject to the following interpretations.

[0031] 1) Dry signal: a dry signal in embodiments of the disclosure refers to a pure human voice signal without accompaniment music. The dry signal is a monophonic sound signal, that is, the dry signal does not contain any directional information.

[0032] 2) Binaural signal: “binaural” means two sound channels. The principle is that people can determine a specific position of a sound source according to a phase difference between the left ear and the right ear when hearing a sound. The binaural signal in embodiments of the disclosure refers to a left channel sound signal and a right channel sound signal.

[0033] 3) Head Related Transfer Functions (HRTF): HRTF can be referred to as a binaural transfer function, which describes transmission of sound waves from a sound source to ears. HRTF is a set of filters with the principle that time domain convolution is equivalent to frequency domain convolution, and can calculate a virtual sound signal transmitted to the ears according to HRTF data corresponding to sound source position information.

[0034] The embodiment of the disclosure provides a method for generating an accompaniment, a device, and a storage medium. In embodiments of the disclosure, a dry signal set including multiple dry signals corresponding to a same target song can be obtained. Virtual sound signals corresponding to each dry signal at different virtual three-dimensional spatial acoustic image positions are obtained, where the dry signals are included in the dry signal set. Then the virtual sound signals corresponding to each dry signal are merged to obtain a chorus dry sound. At last, audio effect synthesis is performed on the chorus dry sound and background music of the target song according to a sound effect optimizing rule to obtain an accompaniment of the target song. In this way, on the one hand, acoustic image positions of each dry signal included in the dry signal set in a virtual three-dimensional space can be simulated comprehensively, and then the virtual sound signals corresponding to each dry signal at different virtual three-dimensional spatial acoustic image positions are merged to obtain the chorus dry sound, thereby achieving the surrounded stereo audio effect. On the other hand, audio effect synthesis can be performed on the chorus dry sound and the background music according to the sound effect optimizing rule to obtain the accompaniment, so as to enhance an immersive audio effect. In general, compared with a processing manner of directly superimposing each dry signal, richer audio effects can be obtained and user experience can be improved in the disclosure.

[0035] Reference is made to FIG. 1 which is a schematic diagram illustrating an application

scenario of a method for generating an accompaniment provided in embodiments of the disclosure. As illustrated in FIG. 1, the application scenario may include a smart device 100 which communicates with a server 110 in a wired manner or in a wireless manner, and the server 110 is connected to a database 120.

[0036] The method for generating an accompaniment provided in embodiments of the disclosure can be implemented by an electronic device such as the smart device 100. For example, when the smart device 100 receives a selection instruction indicating that an accompaniment mode for the target song is a chorus accompaniment mode, a dry signal set corresponding to the target song can be obtained. A virtual sound signal is generated based on a dry signal corresponding to each virtual three-dimensional spatial acoustic image position of N virtual three-dimensional spatial acoustic image positions. For example, the virtual sound signal may be a binaural signal. The virtual sound signals corresponding to each dry signal are merged to obtain a chorus dry sound. Audio effect synthesis is performed on the chorus dry sound and background music of the target song according to a sound effect optimizing rule to obtain an accompaniment. As an example, an option of "chorus accompaniment" is illustrated in the smart device 100 in FIG. 1. A user may control by voice to generate a selection instruction for the chorus accompaniment mode or may trigger a selection control displayed on a user interface to generate a selection instruction for the chorus accompaniment mode. The dry signal set may be pre-stored locally by the smart device 100, or the dry signal set may be obtained by the smart device 100 from the server 110 or the database 120.

[0037] The method for generating an accompaniment provided in embodiments of the disclosure can further be implemented by an electronic device such as the server 110. For example, when the server 110 receives a selection instruction indicating that an accompaniment mode for a target song is a chorus accompaniment mode, a dry signal set corresponding to the target song can be obtained. A virtual sound signal is generated based on a dry signal corresponding to each virtual three-dimensional spatial acoustic image position of N virtual three-dimensional spatial acoustic image positions. For example, the virtual sound signal may be a binaural signal. The virtual sound signals corresponding to each dry signal are merged to obtain a chorus dry sound. Audio effect synthesis is performed on the chorus dry sound and background music of the target song according to a sound effect optimizing rule to obtain an accompaniment. The dry signal set may be pre-stored locally by the server 110, or the dry signal set may be

obtained by the server 110 from the database 120. The obtained accompaniment may be stored locally or stored in the database 120 to be invoked when needed. Of course, the server 110 may start generating the accompaniment without having to wait for receiving the selection instruction which indicates that the accompaniment mode for the target song is the chorus accompaniment mode. The server 110 may start performing relevant operations in the method for generating an accompaniment in the disclosure to generate the accompaniment at an appropriate time, for example, when the server 110 load is low, when the server 110 receives a new dry signal of the target song, or when the server receives a management operation for generating the accompaniment. Preferably, a chorus accompaniment may be generated in advance and stored in the server. After a large number of accompaniments of songs are generated, the user may send a selection instruction for the target song in a manner such as selecting “chorus accompaniment” on the user interface via the smart device 100. In this way, the server 110 can respond to the selection instruction, find the chorus accompaniment of the target song from the large number of generated accompaniments, and send the chorus accompaniment to the smart device 100.

[0038] The method for generating an accompaniment provided in embodiments of the disclosure can further be implemented collaboratively by electronic devices such as the smart device 100 and the server 110. For example, the server 110 may generate a virtual sound signal based on a dry signal corresponding to each virtual three-dimensional spatial acoustic image position of N virtual three-dimensional spatial acoustic image positions. For example, the virtual sound signal may be a binaural signal. The virtual sound signals corresponding to each dry signal are merged to obtain a chorus dry sound. Audio effect synthesis is performed on the chorus dry sound and background music of the target song according to a sound effect optimizing rule to obtain an accompaniment. The obtained accompaniment is sent to the smart device 100.

[0039] The method for generating an accompaniment provided in embodiments of the disclosure can further be implemented by electronic devices such as the smart device 100 and the server 110 via running a computer program. For example, the computer program may be a native program or software module in an operating system, a local application (APP), or a mini program. In short, the computer program can be an application, a module, or a plug-in in any form, which is not limited in embodiments of the disclosure.

[0040] The smart device involved in embodiments of the disclosure may be, but is not limited to, a personal computer, a laptop, a smart phone, a tablet computer, a smart watch, a

smart voice interaction device, a smart home appliance, an in-vehicle terminal, and a smart wearable device. The server may be an independent physical server, or a server cluster or a distributed system formed by multiple physical servers, or a cloud server that provides basic cloud computing services such as cloud services, cloud databases, cloud computing, cloud functions, cloud storage, network services, cloud communication, middleware services, domain name services, security services, content delivery networks (CDNs), big data, and artificial intelligence platforms. The smart device and the server may be directly or indirectly connected in a wired or wireless communication manner, which is not limited in embodiments of the disclosure.

[0041] It is understood that the number of dry signals and virtual three-dimensional spatial acoustic image positions illustrated in FIG. 1 is only an example. According to implementation requirements, a dry signal set may include any number of dry signals, and any number of virtual three-dimensional spatial acoustic image positions may exist in a virtual three-dimensional space.

[0042] Further, reference is made to FIG. 2 which is a schematic flowchart of a method for generating an accompaniment provided in embodiments of the disclosure. The method in embodiments of the disclosure can be applied to an electronic device, such as a smart phone, a tablet computer, a smart wearable device, a personal computer, or a server. The method may include but is not limited to the following operations.

[0043] At S201, a dry signal set is obtained.

[0044] In embodiments of the disclosure, an electronic device can obtain the dry signal set which includes multiple dry signals corresponding to a target song.

[0045] In an embodiment, the dry signal set can be obtained from an audio database. The audio database includes initial dry signals recorded by multiple users when singing a same song. It is noted that the initial dry signals in the audio database are recorded with authorization and consent of the users. An electronic device may select satisfactory dry signals according to sound parameters of the initial dry signals as the dry signal set.

[0046] In an embodiment, the electronic device may select the satisfactory dry signals from the initial dry signal set according to an intonation characteristic parameter and a timbre characteristic parameter. The intonation characteristic parameter may include any one or more of a pitch parameter, a rhythm parameter, or a prosodic parameter. Satisfactory dry signals selected

according to the intonation characteristic parameter have the characteristic of high consistency between the song and the accompaniment in pitch and rhythm. The timbre characteristic parameter may include any one or more of a noise parameter, an energy parameter, or a speed parameter. Satisfactory dry signals selected according to the timbre characteristic parameter have the characteristics of clear audio, appropriate audio energy, and uniform audio speed. A sequence for selecting the satisfactory dry signals is not limited in embodiments of the disclosure. For example, the electronic device may first select the satisfactory dry signals according to the intonation characteristic parameter, and then select satisfactory dry signals that meet a preset timbre characteristic parameter requirement from satisfactory dry signals that meet a preset intonation characteristic parameter requirement. Alternatively, the electronic device may first select satisfactory dry signals according to the timbre characteristic parameter, and then select satisfactory dry signals that meet the preset intonation characteristic parameter requirement from satisfactory dry signals that meet the preset timbre characteristic parameter requirement. The dry signal set including the dry signals selected from the initial dry signal set in this manner has excellent quality in intonation and timbre.

[0047] At S202, a virtual sound signal is generated based on a dry signal corresponding to each virtual three-dimensional spatial acoustic image position of *N* virtual three-dimensional spatial acoustic image positions.

[0048] In an embodiment of the disclosure, the electronic device can simulate different acoustic image positions of each dry signal in the dry signal set in a virtual three-dimensional space, and then generate a virtual sound signal based on the dry signal corresponding to each virtual three-dimensional spatial acoustic image position of the *N* virtual three-dimensional spatial acoustic image positions. The virtual sound signal may be, for example, a binaural signal. The *N* virtual three-dimensional spatial acoustic image positions are different from one another, and each virtual three-dimensional spatial acoustic image position may correspond to one or more dry signals.

[0049] In an embodiment, the *N* virtual three-dimensional spatial acoustic image positions may be obtained through simulation in the virtual three-dimensional space as follows. As illustrated in FIG. 3a, in the virtual three-dimensional space, the positive direction of the x-axis corresponds to the front side of the head, the positive direction of the y-axis corresponds to the left side of the head, and the positive direction of the z-axis corresponds to the upper side of the

head. The virtual three-dimensional space is partitioned by three planes: a horizontal plane 301, an upper plane 302, and a lower plane 303, where an angle between the upper plane 302 and the horizontal plane is a first angle threshold, and an angle between the lower plane 303 and the horizontal plane is a second angle threshold. As illustrated in FIG. 3b, each virtual three-dimensional spatial acoustic image position in the virtual three-dimensional space includes an azimuth angle and an elevation angle. Assuming that θ represents the azimuth angle of the virtual three-dimensional spatial acoustic image position, and φ represents the elevation angle of the virtual three-dimensional spatial acoustic image position, each virtual three-dimensional spatial acoustic image position can be represented as (θ, φ). Accordingly, the horizontal plane 301 is a plane corresponding to an elevation angle of 0°, the upper plane is a plane corresponding to an elevation angle of the first angle threshold, where the first angle threshold may be any angle above the horizontal plane, and the lower plane is a plane corresponding to an elevation angle of the second angle threshold, where the second angle threshold may be any angle below the horizontal plane. For example, the upper plane may be a plane corresponding to an elevation angle of 40°, and the lower plane may be a plane corresponding to an elevation angle of -40°. The azimuth angle θ may be used to describe an angle between the virtual three-dimensional spatial acoustic image position and a target direction line on a plane in a clockwise direction. Further, as illustrated in FIG. 3c, multiple virtual three-dimensional spatial acoustic image positions can be obtained by partitioning different planes corresponding to different elevation angles at intervals of respective preset angles. Specifically, n1 virtual three-dimensional spatial acoustic image positions on the horizontal plane can be obtained by partitioning the horizontal plane at intervals of the first preset angle. n2 virtual three-dimensional spatial acoustic image positions on the upper plane can be obtained by partitioning the upper plane at intervals of the second preset angle. n3 virtual three-dimensional spatial acoustic image positions on the lower plane can be obtained by partitioning the lower plane at intervals of the third preset angle. For example, assuming that the first preset angle is 10°, and the second preset angle and the third preset angle are both 15°, then 36 virtual three-dimensional spatial acoustic image positions can be obtained by partitioning the horizontal plane at intervals of 10°, 24 virtual three-dimensional spatial acoustic image positions can be obtained by partitioning the upper plane at intervals of 15°, and 24 virtual three-dimensional spatial acoustic image positions can also be obtained by partitioning the lower plane at intervals of 15°. In total, 84 different virtual three-dimensional

spatial acoustic image positions can be obtained in this way. It is noted that the first preset angle, the second preset angle, and the third preset angle in embodiments of the disclosure each can be any preset angle. Specific values of the above three preset angles are only used as examples and do not constitute a limitation on embodiments of the disclosure. In this way, multiple virtual three-dimensional spatial acoustic image positions can be virtually created on three different planes in the virtual three-dimensional space at intervals of different azimuth angles, thereby achieving an all-round immersive simulation of a sound source.

[0050] In an embodiment, each virtual three-dimensional spatial acoustic image position may correspond to one or more dry signals. The electronic device may obtain a virtual sound signal corresponding to the one or more dry signals at each virtual three-dimensional spatial acoustic image position. Specifically, the electronic device can obtain virtual sound signals corresponding to each dry signal in the virtual three-dimensional space as follows. An azimuth angle and an elevation angle of a virtual three-dimensional spatial acoustic image position corresponding to the dry signal are obtained. A head related transfer function (HRTF) corresponding to the virtual three-dimensional spatial acoustic image position is determined according to the azimuth angle and elevation angle of the virtual three-dimensional spatial acoustic image position. A virtual sound signal corresponding to the dry signal at the virtual three-dimensional spatial acoustic image position can be calculated according to the azimuth angle and the elevation angle of the virtual three-dimensional spatial acoustic image position and corresponding HRTF data. For example, the azimuth angle and the elevation angle of the virtual three-dimensional spatial acoustic image position corresponding to dry signal X are (θ, φ), and the HRTF data corresponding to the virtual three-dimensional spatial acoustic image position is expressed as follows.

$$\mathrm{HRTF}(\theta,\ \varphi)\begin{cases} Y_L = X \cdot H_L \\ Y_R = \mathrm{X} \cdot H_R \end{cases}$$

[0051] The virtual sound signal corresponding to the dry signal X at the virtual three-dimensional spatial acoustic image position is calculated to be a binaural signal Y_L and Y_R, where Y_L is a left channel signal and Y_R is a right channel signal.

[0052] In an embodiment, the electronic device can obtain part of the dry signals from the dry signal set. For example, the electronic device may randomly obtain or select dry signals with better intonation and timbre according to a new selecting rule, and perform delay processing on

each of the selected part of the dry signals to obtain a delayed binaural signal corresponding to each dry signal of the part of the dry signals. Specifically, in performing delay processing on the dry signal, 8 different pairs of time parameters can be selected. It is noted that the 8 pairs of time parameters represent 8 time parameters for obtaining delayed left channel signals and 8 time parameters for obtaining delayed right channel signals. A total of 16 time parameters will be selected. For example, 16 different parameters may be selected in the range of 21ms~79ms (based on 80ms of the reverberation time for the general room impulse response). Alternatively, 16 (or another number) different parameters may be randomly selected in other reasonable ranges as time parameters according to actual needs. In this way, a dry signal at the left ear or the right ear of the head can be simulated to enrich the audio effect. In an embodiment, a selection manner and setting of time parameters (delay duration parameters) for the delay processing can be adjusted via an interface, so as to facilitate flexible configuration for users who produce chorus audio. It is noted that the above operations of obtaining the virtual sound signal, obtaining the delayed left channel signal and the delayed right channel signal may be performed simultaneously or successively, which is not limited in embodiments of the disclosure.

[0053] At S203, each virtual sound signal in the virtual sound signal set is merged to obtain a chorus dry sound.

[0054] In an embodiment of the disclosure, the electronic device can merge each virtual sound signal in the virtual sound signal set to obtain the chorus dry sound.

[0055] In an embodiment, the virtual sound signals corresponding to each dry signal can be merged through normalization to adjust the loudness of a merged virtual sound signal to [-1dB, 1dB]. The virtual sound signals subjected to the merging process include the virtual sound signal corresponding to the dry signal at each virtual three-dimensional spatial acoustic image position of the *N* virtual three-dimensional spatial acoustic image positions and the delayed binaural signals obtained by delaying part of the dry signals in the dry signal set.

[0056] At S204, audio effect synthesis is performed on the chorus dry sound and background music of the target song according to a sound effect optimizing rule to obtain an accompaniment.

[0057] In an embodiment of the disclosure, the electronic device can obtain a final accompaniment by performing audio effect synthesis on the chorus dry sound and the background music of the target song according to the sound effect optimizing rule. For example, the sound effect optimizing rule may be adjusting a sound parameter of the background music of

the target song and a sound parameter of the above obtained virtual sound signals corresponding to the multiple dry signals, where for example, the sound parameter may be a common adjustable parameter such as loudness and timbre.

[0058] In an embodiment, after obtaining the chorus dry sound, the electronic device can obtain the background music of the target song. When an energy relationship between the obtained chorus dry sound and the background music of the target song does not satisfy an energy ratio condition, the electronic device can adjust the energy relationship between the chorus dry sound and the background music of the target song. Here, the energy ratio condition may be set as that a ratio between an energy value of the chorus dry sound and an energy value of the background music of the target song is less than a ratio threshold, or a loudness of the chorus dry sound is 3dB lower than a loudness of the background music of the target song. In this manner, the energy of the chorus dry sound can be avoided to be greater than the energy of the background music of the target song, thereby obtaining a more harmonious accompaniment.

[0059] In embodiments of the disclosure, the virtual sound signals corresponding to each dry signal at different virtual three-dimensional spatial acoustic image positions in the virtual three-dimensional space can be obtained. The virtual sound signals are merged to obtain the chorus dry sound. Audio effect synthesis is performed on the chorus dry sound and the background music of the target song according to the sound effect optimizing rule to obtain the accompaniment of the target song, thereby achieving a surrounded stereo effect in listening and enhancing an immersive audio effect to provide good user experience.

[0060] Further, reference is made to FIG. 4 which is a schematic flowchart of another method for generating an accompaniment provided in embodiments of the disclosure. The method in embodiments of the disclosure can be applied to an electronic device. For example, the electronic device may be a smart phone, a tablet computer, a smart wearable device, a personal computer, a server, etc. The method can include but is not limited to the following operations.

[0061] At S401, an initial dry signal set is obtained from an audio database.

[0062] In an embodiment of the disclosure, the electronic device can obtain the initial dry signal set from the audio database. It is noted that the initial dry signal set in the audio database is recorded with authorization and consent of users.

[0063] In an embodiment, the audio database may be a separate database or may be

integrated with the electronic device, that is, the audio database may be stored in the electronic device. Here, the initial dry signal set refers to, in the audio database, a set of original dry signals recorded with authorization and consent of the users when singing a same song.

[0064] At S402, dry signals are selected from the initial dry signal set as the dry signal set according to a sound parameter of each initial dry signal.

[0065] In embodiments of the disclosure, the electronic device can select satisfactory dry signals from the initial dry signal set according to the sound parameter of each initial dry signal, thereby narrowing the initial dry signal set to obtain the dry signal set.

[0066] In an embodiment, the sound parameter of the initial dry signal may include an intonation characteristic parameter and a timbre characteristic parameter of the initial dry signal. The intonation characteristic parameter may include any one or more of a pitch parameter, a rhythm parameter, or a prosodic parameter. The timbre characteristic parameter may include any one or more of a noise parameter, an energy parameter, or a speed parameter. In this way, initial dry signals with poor audio effects such as noisy, out-of-tune, too short audio time, low audio energy, and popping sounds can be removed from the initial dry signal set to obtain the dry signal set with excellent intonation and timbre.

[0067] At S403, an HRTF corresponding to each virtual three-dimensional spatial acoustic image position of N virtual three-dimensional spatial acoustic image positions is obtained.

[0068] In embodiments of the disclosure, the electronic device may obtain the N virtual three-dimensional spatial acoustic image positions in a virtual three-dimensional space and obtain the HRTF corresponding to each virtual three-dimensional spatial acoustic image position according to the N virtual three-dimensional spatial acoustic image positions.

[0069] In an embodiment, the HRTF corresponding to each virtual three-dimensional spatial acoustic image position in the virtual three-dimensional space can be pre-stored in an HRTF database, so that the electronic device may invoke a corresponding HRTF from the HRTF database according to the virtual three-dimensional spatial acoustic image position.

[0070] At S404, a dry signal corresponding to a target virtual three-dimensional spatial acoustic image position is processed with an HRTF corresponding to the target virtual three-dimensional spatial acoustic image position to obtain a virtual sound signal at the target virtual three-dimensional spatial acoustic image position.

[0071] In embodiments of the disclosure, the electronic device can process a target dry signal

according to the HRTF corresponding to the target virtual three-dimensional spatial acoustic image position to obtain a virtual sound signal corresponding to the target dry signal at the target virtual three-dimensional spatial acoustic image position. The target virtual three-dimensional spatial acoustic image position may be any one of the *N* virtual three-dimensional spatial acoustic image positions, and the target dry signal may be any one dry signal in the dry signal set.

[0072] In an embodiment, the HRTF corresponding to the target virtual three-dimensional spatial acoustic image position is HRTF data corresponding to the target virtual three-dimensional spatial acoustic image position. According to an azimuth angle and an elevation angle of the target virtual three-dimensional spatial acoustic image position, HRTF data corresponding to the target virtual three-dimensional spatial acoustic image position can be determined from known HRTF data. Then the electronic device may convolve the target dry signal with the HRTF data corresponding to the target virtual three-dimensional spatial position to obtain the virtual sound signal corresponding to the target dry signal at the target virtual three-dimensional spatial acoustic image position.

[0073] At S405, *p* dry signals are obtained from *x* dry signals included in the dry signal set.

[0074] In embodiments of the disclosure, the electronic device can randomly obtain the *p* dry signals from the *x* dry signals included in the dry signal set. It is noted that operations at S404 and S405 may be executed simultaneously or successively, which is not limited in the disclosure.

[0075] At S406, delay processing is performed on each dry signal of the *p* dry signals to obtain a delayed left channel signal and a delayed right channel signal corresponding to each dry signal of the *p* dry signals.

[0076] In embodiments of the disclosure, the electronic device can perform delay processing on each dry signal of the *p* dry signals for m1 time parameters to obtain m1 delayed dry signals corresponding to each dry signal of the *p* dry signals, and obtain a delayed left channel signal corresponding to each dry signal of the *p* dry signals by superimposing the m1 delayed dry signals corresponding to each dry signal, where m1 is a positive integer. The electronic device further performs delay processing on each dry signal of the *p* dry signals for m2 time parameters to obtain m2 delayed dry signals corresponding to each dry signal of the *p* dry signals, and obtain a delayed right channel signal corresponding to each dry signal of the *p* dry signals by superimposing the m2 delayed dry signals corresponding to each dry signal, where m2 is a

positive integer.

[0077] In an embodiment, the electronic device can process a dry signal via 16 delay devices with different time parameters to obtain 16 dry signals with different delays and attenuation degrees. The 16 dry signals with different delays and attenuation degrees are evenly classified into two groups. Dry signals with different delays and attenuation degrees in each group are superimposed separately, so as to finally obtain the delayed left channel signal and the delayed right channel signal corresponding to the dry signal.

[0078] In an embodiment, before the delayed binaural signal corresponding to each dry signal of the p dry signals is obtained, a sound field of the dry signal may be widened by adding a bass enhancement and reverberation simulation module to reduce a correlation between the delayed left channel signal and the delayed right channel signal of the binaural signal obtained by delay processing. It is noted that operations at S403 and S404 for obtaining the virtual sound signal and operations at S405 and S406 for obtaining the delayed left channel signal and the delayed right channel signal may be performed simultaneously or successively, which is not limited in embodiments of the disclosure, where the operation at S405 and the operation at S406 are two optional operations.

[0079] At S407, virtual sound signals in a virtual sound signal set are merged to obtain a chorus dry sound.

[0080] In embodiments of the disclosure, the electronic device can merge the virtual sound signals in the virtual sound signal set to obtain the chorus dry sound. Here, the virtual sound signal set includes virtual sound signals corresponding to each dry signal obtained by the electronic device by simulating the N virtual three-dimensional spatial positions and delayed binaural signals obtained by the electronic device by performing delaying processing on the p dry signals of the dry signal set.

[0081] In an embodiment, each virtual sound signal in the virtual sound signal set is a binaural signal. The binaural signal includes a left channel signal and a right channel signal. The left channel signal and the right channel signal may be processed separately in merging each virtual sound signal. The left channel signal and the right channel signal are applicable to a same processing rule. Here, the merging process may be implemented through normalization, so that a loudness of a merged binaural signal is [-1dB, 1dB]. For example, it is assumed that there are 1000 binaural signals including 1000 left channel signals and 1000 right channel signals. Each

left channel signal is normalized separately, and the 1000 normalized left channel signals are added together and then divided by 1000 to obtain a merged left channel signal. Similarly, after each right channel signal is normalized, the 1000 normalized right channel signals are added together and then divided by 1000 to obtain a merged right channel signal. In this way, the chorus dry sound can be obtained.

[0082] In an embodiment, an energy relationship between the obtained chorus dry sound and background music of a target song may satisfy an energy ratio condition or may not satisfy the energy ratio condition. If the energy relationship between the obtained chorus dry sound and the background music satisfies the energy ratio condition, the operation at S408 may be ignored. Correspondingly, if the energy relationship between the obtained chorus dry sound and the background music does not satisfy the energy ratio condition, the operation at S408 is to be performed.

[0083] At S408, the background music of the target song is obtained, and the energy relationship between the chorus dry sound and the background music is adjusted.

[0084] In embodiments of the disclosure, the electronic device can obtain the background music of the target song and adjust the energy relationship between the chorus dry sound and the corresponding background music. An energy relationship between an adjusted chorus dry sound and an adjusted background music satisfies the energy ratio condition.

[0085] In an embodiment, the energy of the chorus dry sound may be too huge and cover the energy of the background music. By adjusting the chorus dry sound and the background music, the energy relationship between the adjusted chorus dry sound and the adjusted background music can satisfy the energy ratio condition, so that the situation where the energy of the chorus dry sound is too huge can be dealt with in this way. The energy ratio condition can be set as that a ratio between an energy value of the chorus dry sound and an energy value of the background music is less than a ratio threshold, or a loudness of the chorus dry sound is 3dB lower than a loudness of the background music.

[0086] In an embodiment, after the background music of the target song is obtained, according to a detailed description of the operation at S202 illustrating that the virtual sound signal is generated based on the dry signal corresponding to each virtual three-dimensional spatial acoustic image position of the N virtual three-dimensional spatial acoustic image positions, the background music may be subjected to a same process to obtain a chorus dry

sound and background music with similar effects, thereby achieving more harmonious and unified listening experience.

[0087] At S409, spectrum coordination processing is performed on the chorus dry sound at a preset frequency band.

[0088] In embodiments of the disclosure, the electronic device can perform spectrum coordination processing on the chorus dry sound at the preset frequency band.

[0089] In an embodiment, the electronic device can achieve spectrum coordination by additionally performing spectrum notch processing at the preset frequency band. For example, the electronic device may additionally perform spectrum notch processing of about 6dB around 4kHz. In this way, high-frequency current sound caused by spectrum incoordination can be prevented, so that the chorus dry sound can sound more natural.

[0090] At S410, the loudness of the background music can be obtained.

[0091] In embodiments of the disclosure, the electronic device can obtain the loudness of the background music.

[0092] At S411, when the loudness is less than a loudness threshold, the loudness of the background music is increased to the loudness threshold.

[0093] In embodiments of the disclosure, when the loudness of the background music is less than the loudness threshold, the electronic device can increase the loudness of the background music to the loudness threshold. For example, the loudness threshold can be set as -14dB. When the loudness of the background music is less than -14dB, the electronic device can increase the loudness to -14dB.

[0094] At S412, an accompaniment is obtained.

[0095] In embodiments of the disclosure, the electronic device superimposes the chorus dry sound and the background music to obtain a final accompaniment. It is noted that the accompaniment may be obtained according to any one operation or a combination of multiple operations at S408 to S411. Moreover, in an embodiment, the operations at S408 to S411 may be selectively performed according to actual needs. For example, there may be a situation where the energy relationship between the chorus dry sound and the background music does not need to be adjusted, so that the operation at S408 will not be performed. Similarly, the operation of performing spectrum coordination processing on the chorus dry sound at the preset frequency band may also be optional. For another example, operations at S410 and S411 may not be

performed. In FIG. 4, only the technical solutions adopted to make the accompaniment more harmonious and natural and the timbre better are illustrated. A sequence of energy relationship adjustment at S408, spectrum coordination adjustment at S409, and loudness adjustment at S410 and S411 is not limited in the disclosure.

[0096] In an embodiment, an obtained final accompaniment can be stored in a database, so that the electronic device may directly obtain a corresponding accompaniment from the database when receiving a chorus request for a same song.

[0097] In embodiments of the disclosure, the virtual sound signals corresponding to each dry signal can be obtained by simulating the *N* virtual three-dimensional spatial positions, and the delayed binaural signal corresponding to each dry signal can be obtained by performing delay processing on each dry signal, thereby enriching the chorus dry sound. In addition, the energy relationship between the chorus dry sound and the background music is adjusted to make the final accompaniment sounds more harmonious and natural, so as to enable users to clearly feel a sense of space and immersion when chorusing.

[0098] Further, reference is made to FIG. 5 which is a schematic flowchart illustrating obtaining of a virtual sound signal in a method for generating an accompaniment provided in embodiments of the disclosure. The virtual sound signal is obtained as follows. A virtual sound signal corresponding to a dry signal at each virtual three-dimensional spatial acoustic image position of *N* virtual three-dimensional spatial acoustic image positions is obtained. Delay processing is performed on each dry signal of *p* dry signals to obtain a delayed binaural signal corresponding to each dry signal of the *p* dry signals.

[0099] In embodiments of the disclosure, after a dry signal set is obtained, the virtual sound signal corresponding to the dry signal at each virtual three-dimensional spatial acoustic image position of the *N* virtual three-dimensional spatial acoustic image positions can be obtained. Delay processing can be performed on each dry signal of the *p* dry signals in the dry signal set to obtain the delayed binaural signal corresponding to each dry signal of the *p* dry signals.

[0100] For example, as illustrated in FIG. 5, corresponding virtual sound signals are obtained in the above two manners respectively for dry signal X and dry signal W which are included in the dry signal set, where the dry signal X and the dry signal W each can be any dry signal in the dry signal set. Specifically, after obtaining the dry signal X in the dry signal set, the electronic device can describe position information, i.e., (θ, φ), of a virtual three-dimensional spatial

acoustic image position according to an azimuth angle and an elevation angle of the virtual three-dimensional spatial acoustic image position. Then, according to the position information (θ, φ) of the virtual three-dimensional spatial acoustic image position, HRTF (θ, φ) corresponding to the virtual three-dimensional spatial acoustic image position can be determined. After the dry signal X is convolved with the HRTF (θ, φ) corresponding to the virtual three-dimensional spatial acoustic image position, the virtual sound signal corresponding to the dry signal at the virtual three-dimensional spatial acoustic image position can be obtained. The virtual sound signal is a binaural signal including a left channel signal Y_L and a right channel signal Y_R. The virtual sound signal corresponding to the dry signal obtained in this way can enhance stereoscopic immersion for the user.

[0101] In addition, after obtaining the dry signal W in the dry signal set, the electronic device may perform delay processing on the dry signal W. For example, the electronic device may perform delay processing on the dry signal W via a total of 16 delay devices with different time parameters, where the 16 delay devices include $d_L(1)$, $d_L(2)$, …, $d_L(8)$, $d_R(1)$, $d_R(2)$, …, and $d_R(8)$. Dry signals obtained through delay processing via 8 delay devices, i.e., $d_L(1)$, $d_L(2)$, …, and $d_L(8)$ are superimposed to obtain a delayed left channel signal W_L. Dry signals obtained through delay processing via 8 delay devices, i.e., $d_R(1)$, $d_R(2)$, …, and $d_R(8)$ are superimposed to obtain a delayed right channel signal W_R. A delayed binaural signal corresponding to the dry signal obtained in this way can simulate a binaural signal at the left ear or right ear of the head, thereby enriching the listening effect for the user.

[0102] In an embodiment, an obtained final virtual sound signal set includes the above two cases, that is, the final virtual sound signal set is: Z={ZL, ZR}, $ZL=Y_L+W_L$; $ZR=Y_R+W_R$. It is noted that the above operations of obtaining the virtual sound signal and obtaining the delayed left channel signal and the delayed right channel signal may be performed simultaneously or successively, which is not limited in the disclosure. By obtaining the virtual sound signal corresponding to dry signal in the dry signal set in the above two different manners, a chorus scene may be comprehensively presented thereby enriching the audio effect.

[0103] Further, reference is made to FIG. 6 which illustrates a method for playing and processing an accompaniment provided in embodiments of the disclosure. The method in embodiments of the disclosure can be applied to an electronic device. The electronic device may

be, for example, a smart device such as a smart phone, a tablet computer, a smart wearable device, and a personal computer, or a server, etc. The method can include but is not limited to the following operations.

[0104] At S601, a user interface is displayed.

[0105] In embodiments of the disclosure, the electronic device may display the user interface for receiving a selection instruction from a user for a target song.

[0106] In an embodiment, the selection instruction may include a selection instruction for an accompaniment mode for the target song. The accompaniment mode for the target song may be, but is not limited to, a chorus accompaniment mode, an original sound accompaniment mode, or an artificial intelligence (AI) accompaniment mode.

[0107] In an embodiment, the selection instruction may be an instruction generated by the user by triggering a selection control displayed on the user interface, or may be a selection instruction generated by the user by controlling the electronic device by voice. For example, a voice control from the user for the electronic device may be “please play in the chorus accompaniment mode”. In this way, the electronic device may generate a selection instruction indicating that the accompaniment mode for the target song is the chorus accompaniment mode.

[0108] At S602, when a selection instruction received on the user interface indicates that the accompaniment mode for the target song is the chorus accompaniment mode, an accompaniment corresponding to the target song is obtained.

[0109] In embodiments of the disclosure, when the selection instruction received on the user interface indicates that the accompaniment mode for the target song is the chorus accompaniment mode, the electronic device may obtain an accompaniment corresponding to the chorus accompaniment mode for the target song.

[0110] In an embodiment, selection controls for the accompaniment mode for the target song may be displayed on the user interface. The selection controls may include a chorus accompaniment mode selection control and an original sound accompaniment mode selection control. Before obtaining the accompaniment corresponding to the target song, the electronic device may detect whether a selection operation for the chorus accompaniment mode selection control is obtained. When the selection operation for the chorus accompaniment mode selection control is obtained, it is determined that the selection instruction received on the user interface indicates that the accompaniment mode for the target song is the chorus accompaniment mode.

[0111] In an embodiment, the accompaniment corresponding to the chorus accompaniment mode is generated according to a chorus dry sound and background music. The chorus dry sound may be generated according to a virtual sound signal set. The virtual sound signal set includes a virtual sound signal generated, according to an obtained dry signal set, at each virtual three-dimensional spatial acoustic image position of N virtual three-dimensional spatial acoustic image positions. Multiple dry signals in the dry signal set may correspond to multiple different virtual three-dimensional spatial acoustic image positions. Each virtual three-dimensional spatial acoustic image position may correspond to one or more dry signals. The dry signal set is obtained according to dry signals recorded by multiple users for the target song. It is noted that the dry signals for the target song are recorded with authorization and consent of the users. Specifically, the method for generating the accompaniment corresponding to the chorus accompaniment mode may be referred to embodiments illustrated in FIG. 2 to FIG. 5 above, which will not be repeated herein.

[0112] At S603, the accompaniment corresponding to the target song is played.

[0113] In embodiments of the disclosure, after obtaining the accompaniment corresponding to the chorus accompaniment mode for the target song, the electronic device can play the accompaniment for the user.

[0114] In an embodiment, the accompaniment corresponding to the target song may be applied in a karaoke scene. The user may sing while the accompaniment is playing. With the authorization and consent of the user, the electronic device may merge captured singing of the user with the accompaniment corresponding to the target song and play the captured singing of the user and the accompaniment after merging, so as to provide the user with a unique experience like being at a concert.

[0115] In an embodiment, as illustrated in FIG. 7a, an electronic device can obtain an accompaniment corresponding to a target song as follows.

[0116] At S701, an accompaniment request is sent to a server.

[0117] In embodiments of the disclosure, the electronic device may send the accompaniment request to the server. The accompaniment request may include identification information of the target song.

[0118] In an embodiment, the identification information of the target song is the unique information indicating the target song. For example, the identification information may be the

name of the target song.

[0119] At S702, a chorus dry sound and background music returned by the server in response to the accompaniment request are received.

[0120] In embodiments of the disclosure, the electronic device may receive the chorus dry sound and the background music returned by the server in response to the accompaniment request for the target song.

[0121] In an embodiment, the server may return the chorus dry sound and the background music separately, or may merge the chorus dry sound and the background music before returning them. A specific return manner may be selected according to settings from the user.

[0122] At S703, a target chorus dry sound segment is determined from the chorus dry sound.

[0123] In embodiments of the disclosure, the electronic device may determine the target chorus dry sound segment according to a returned chorus dry sound.

[0124] In an embodiment, the electronic device may display a first single-sentence interface. As illustrated in FIG. 7b, each single sentence in text data corresponding to the chorus dry sound is displayed on the first single-sentence interface according to a time playing node sequence of the chorus dry sound. The user may select the target chorus dry sound segment according to each single sentence displayed on the first single-sentence interface.

[0125] In an embodiment, the target chorus dry sound segment may include some of or all of the single sentences in the chorus dry sound, which may be specifically determined by a selection operation of the user.

[0126] At S704, an accompaniment corresponding to the target song is obtained according to a chorus dry sound and background music corresponding to the target chorus dry sound segment.

[0127] In embodiments of the disclosure, the electronic device may obtain the accompaniment corresponding to the target song according to the chorus dry sound and the background music corresponding to the target chorus dry sound segment selected by the user.

[0128] In an embodiment, the electronic device may display a second single-sentence interface. As illustrated in FIG. 7c, the second single-sentence interface may be displayed in a playing process of the accompaniment corresponding to the target song. Each single sentence in the text data corresponding to the accompaniment may be displayed according to a time playing node sequence of the accompaniment.

[0129] In an embodiment, the electronic device may further detect whether a mute selection

operation for the chorus dry sound in the accompaniment is obtained in the playing process. When the mute selection operation for the chorus dry sound in the accompaniment is obtained in the playing process, playing of the chorus dry sound in the accompaniment can be terminated at a current time playing node, and only the background music in the accompaniment is played.

[0130] In embodiments of the disclosure, on the one hand, the selection instruction of the user for the target song can be received. When the selection instruction of the user indicates that the accompaniment mode for the target song is the chorus accompaniment mode, the accompaniment corresponding to the target song can be obtained and played. On the other hand, the accompaniment of the target song in the chorus accompaniment mode is generated according to the chorus dry sound and the background music. The target chorus segment can be determined from the chorus dry sound. The accompaniment corresponding to the target song can be generated according to the chorus dry sound and the background music corresponding to the target chorus dry sound segment. In this way, the user can have the experience of being at a concert and have an immersive listening experience when the accompaniment corresponding to the target song is played. In addition, the user can flexibly select the chorus dry sound in the accompaniment, so that the accompaniment can be more interesting, thereby improving user experience.

[0131] Further, reference is made to FIG. 8a which is a schematic structural diagram of an apparatus for generating an accompaniment provided in embodiments of the disclosure. The apparatus in embodiments of the disclosure can be applied to an electronic device. The electronic device may be, for example, a smart phone, a tablet computer, a smart wearable device, a personal computer, a server, etc. In an embodiment, as illustrated in FIG. 8a, an accompaniment generating apparatus 80 can include an obtaining unit 801 and a processing unit 802.

[0132] The obtaining unit 801 is configured to obtain a dry signal set. The dry signal set includes x dry signals corresponding to a target song, where x is an integer greater than 1. A virtual sound signal is generated based on a dry signal corresponding to each virtual three-dimensional spatial acoustic image position of N virtual three-dimensional spatial acoustic image positions, where the x dry signals correspond to the N virtual three-dimensional spatial acoustic image positions, N is an integer greater than 1, the N virtual three-dimensional spatial acoustic image positions are different from one another, and each virtual three-dimensional spatial acoustic image position is allowed to correspond to one or more dry signals of the x dry signals.

The processing unit 802 is configured to merge each virtual sound signal in a virtual sound signal set to obtain a chorus dry sound. The virtual sound signal set includes a virtual sound signal at each virtual three-dimensional spatial acoustic image position of the *N* virtual three-dimensional spatial acoustic image positions. The processing unit 802 is further configured to perform audio effect synthesis on the chorus dry sound and background music of the target song according to a sound effect optimizing rule to obtain an accompaniment of the target song.

[0133] In an embodiment, the obtaining unit 801 can further be configured to obtain an initial dry signal set from an audio database. The audio database includes initial dry signals recorded by multiple users when singing a same song. The processing unit 802 can further be configured to select dry signals from the initial dry signal set as the dry signal set according to a sound parameter of each initial dry signal.

[0134] In an embodiment, the dry signal set includes a dry signal selected from the initial dry signal set according to an intonation characteristic parameter and a timbre characteristic parameter. The intonation characteristic parameter includes any one or more of a pitch parameter, a rhythm parameter, or a prosodic parameter. The timbre characteristic parameter includes any one or more of a noise parameter, an energy parameter, or a speed parameter.

[0135] In an embodiment, the *N* virtual three-dimensional spatial acoustic image positions include n1 virtual three-dimensional spatial acoustic image positions on a horizontal plane obtained by partitioning the horizontal plane at intervals of a first preset angle; n2 virtual three-dimensional spatial acoustic image positions on an upper plane obtained by partitioning the upper plane at intervals of a second preset angle, where an angle between the upper plane and the horizontal plane is a first angle threshold; and n3 virtual three-dimensional spatial acoustic image positions on a lower plane obtained by partitioning the lower plane at intervals of a third preset angle, where an angle between the lower plane and the horizontal plane is a second angle threshold, where the angle between the lower plane and the horizontal plane is the second angle threshold; where n1, n2, and n3 are positive integers, and a sum of the n1, n2, and n3 is equal to *N*.

[0136] In an embodiment, the obtaining unit 801 can further be configured to obtain an HRTF corresponding to each virtual three-dimensional spatial acoustic image position of the *N* virtual three-dimensional spatial acoustic image positions. The processing unit 802 can further be configured to process a dry signal corresponding to a target virtual three-dimensional spatial

acoustic image position with an HRTF corresponding to the target virtual three-dimensional spatial acoustic image position to obtain a virtual sound signal at the target virtual three-dimensional spatial acoustic image position. The virtual sound signal at the target virtual three-dimensional spatial acoustic image position is a binaural signal. The target virtual three-dimensional spatial acoustic image position is any one of the *N* virtual three-dimensional spatial acoustic image positions.

[0137] In an embodiment, the virtual sound signal set further includes a delayed left channel signal and a delayed right channel signal corresponding to each dry signal of *p* dry signals. The obtaining unit 801 can further be configured to obtain the *p* dry signals from the *x* dry signals included in the dry signal set, where *p* is a positive integer and is less than or equal to *x*. The processing unit 802 can further be configured to perform delay processing on each dry signal of the *p* dry signals for m1 time parameters to obtain m1 delayed dry signals corresponding to each dry signal of the *p* dry signals, and obtain a delayed left channel signal corresponding to each dry signal of the *p* dry signals by superimposing the m1 delayed dry signals corresponding to each dry signal, where m1 is a positive integer. The processing unit 802 can further be configured to perform delay processing on each dry signal of the *p* dry signals for m2 time parameters to obtain m2 delayed dry signals corresponding to each dry signal of the *p* dry signals and obtain delayed right channel signal corresponding to each dry signal of the *p* dry signals by superimposing the m2 delayed dry signals corresponding to each dry signal, where m2 is a positive integer.

[0138] In an embodiment, the obtaining unit 801 can further be configured to obtain the background music of the target song. The processing unit 802 can further be configured to adjust an energy relationship between the chorus dry sound and the background music. An energy relationship between an adjusted chorus dry sound and an adjusted background music satisfies an energy ratio condition. The accompaniment is obtained according to the adjusted chorus dry sound and the adjusted background music.

[0139] In an embodiment, the processing unit 802 can further be configured to perform spectrum coordination processing on the chorus dry sound at a preset frequency band. The obtaining unit 801 can further be configured to obtain a loudness of the background music. The processing unit 802 can further be configured to increase the loudness of the background music to a loudness threshold when the loudness of the background music is less than the loudness

threshold. The accompaniment is obtained according to a spectrum-coordination-processed chorus dry sound and loudness-processed background music.

[0140] It is noted that a specific implementation manner of each operation and contents not mentioned in an embodiment corresponding to FIG. 8a can be referred to embodiments illustrated in FIG. 2 to FIG. 5 and the aforementioned contents, which will not be repeated here.

[0141] Further, reference is made to FIG. 8b which is a schematic structural diagram of an apparatus for playing and processing an accompaniment provided in embodiments of the disclosure. The apparatus in embodiments of the disclosure can be applied to an electronic device. The electronic device may be, for example, a smart phone, a tablet computer, a smart wearable device, a personal computer, a server, etc. In an embodiment, as illustrated in FIG. 8b, an accompaniment playing and processing apparatus 81 may include an obtaining unit 811 and a processing unit 812.

[0142] The obtaining unit 811 is configured to display a user interface for receiving a selection instruction for a target song. When the selection instruction received on the user interface indicates that an accompaniment mode for the target song is a chorus accompaniment mode, an accompaniment corresponding to the target song is obtained. The processing unit 812 is configured to play the accompaniment corresponding to the target song. The accompaniment is generated according to a chorus dry sound and background music, where the chorus dry sound is generated according to multiple dry signals in a dry signal set, the multiple dry signals in the dry signal set correspond to multiple different virtual three-dimensional spatial acoustic image positions, and the dry signal set is obtained according to dry signals recorded by multiple users for the target song.

[0143] In an embodiment, the chorus dry sound is generated according to a virtual sound signal set. The virtual sound signal set includes a virtual sound signal at each virtual three-dimensional spatial acoustic image position of *N* virtual three-dimensional spatial acoustic image positions, and the virtual sound signal is generated according to an obtained dry signal set. Multiple dry signals in the dry signal set correspond to the *N* virtual three-dimensional spatial acoustic image positions. N is an integer greater than 1. The *N* virtual three-dimensional spatial acoustic image positions are different from one another. Each virtual three-dimensional spatial acoustic image position is allowed to correspond to one or more dry signals.

[0144] In an embodiment, selection controls for the accompaniment mode for the target song

are displayed on the user interface. The selection control for the accompaniment mode includes a chorus accompaniment mode selection control and an original sound accompaniment mode selection control. The processing unit 812 can further be configured to detect whether a selection operation for the chorus accompaniment mode selection control is obtained, before the accompaniment corresponding to the target song is obtained. The processing unit 812 can further be configured to determine that the selection instruction received on the user interface indicates that the accompaniment mode for the target song is the chorus accompaniment mode when the selection operation for the chorus accompaniment mode selection control is obtained.

[0145] In an embodiment, the processing unit 812 can further be configured to send an accompaniment request to a server, where the accompaniment request contains identification information of the target song. The obtaining unit 811 can further be configured to receive the chorus dry sound and the background music returned by the server in response to the accompaniment request. The processing unit 812 can further be configured to determine a target chorus dry sound segment from the chorus dry sound, and obtain the accompaniment corresponding to the target song according to a chorus dry sound and background music corresponding to the target chorus dry sound segment.

[0146] In an embodiment, the processing unit 812 can further be configured to display a first single-sentence interface, and display each single sentence in text data corresponding to the chorus dry sound according to a time playing node sequence of the chorus dry sound. The target chorus dry sound segment is determined based on a single-sentence selection operation on the first single-sentence interface.

[0147] In an embodiment, the processing unit 812 can further be configured to display a second single-sentence interface, display each single sentence in text data corresponding to the accompaniment according to a time playing node sequence of the accompaniment, and detect whether a mute selection operation for the chorus dry sound in the accompaniment is obtained in a playing process. The processing unit 812 can further be configured to terminate playing of the chorus dry sound at a current time playing node when the mute selection operation for the chorus dry sound in the accompaniment is obtained in the playing process.

[0148] It is noted that a specific implementation manner of each operation and contents not mentioned in an embodiment corresponding to FIG. 8b can be referred to embodiments illustrated in FIG. 2 to FIG. 7c and aforementioned contents, which will not be repeated here.

[0149] Further, reference is made to FIG. 9 which is a schematic structural diagram of an electronic device provided in embodiments of the disclosure. The electronic device may include a network interface 901, a memory 902, and a processor 903. The network interface 901, the memory 902, and the processor 903 are connected via one or more communication buses. The communication buses are configured to implement connection and communication between these components. The network interface 901 may include a standard wired interface and a standard wireless interface (for example, a WIFI interface). The memory 902 may include a volatile memory such as a random access memory (RAM). The memory 902 may further include a non-volatile memory such as a flash memory, a solid-state drive (SSD), etc. The memory 902 may further include a combination of the abovementioned types of memory. The processor 903 may be a central processing unit (CPU). The processor 903 may further include a hardware chip. The abovementioned hardware chip may be an application-specific integrated circuit (ASIC), a programmable logic device (PLD), etc. The PLD may be a field-programmable gate array (FPGA), a generic array logic (GAL), or the like.

[0150] Optionally, the memory 902 is further configured to store program instructions. The processor 903 can further invoke the program instructions to implement the following.

[0151] A dry signal set is obtained, where the dry signal set includes x dry signals corresponding to a target song, and x is an integer greater than 1. A virtual sound signal is generated based on a dry signal corresponding to each virtual three-dimensional spatial acoustic image position of N virtual three-dimensional spatial acoustic image positions, where the x dry signals correspond to the N virtual three-dimensional spatial acoustic image positions, N is an integer greater than 1, the N virtual three-dimensional spatial acoustic image positions are different from one another, and each virtual three-dimensional spatial acoustic image position is allowed to correspond to one or more dry signals of the x dry signals. Each virtual sound signal in a virtual sound signal set is merged to obtain a chorus dry sound, where the virtual sound signal set includes the virtual sound signal at each virtual three-dimensional spatial acoustic image position of the N virtual three-dimensional spatial acoustic image positions. Audio effect synthesis is performed on the chorus dry sound and background music of the target song according to a sound effect optimizing rule to obtain an accompaniment of the target song.

[0152] In an embodiment, the processor 903 can further invoke the program instructions to implement the following. An initial dry signal set is obtained from an audio database. The audio

database includes initial dry signals recorded by multiple users when singing the target song. *x* dry signals from the initial dry signal set are selected as the dry signal set according to a sound parameter of each initial dry signal.

[0153] In an embodiment, the dry signal set includes a dry signal selected from the initial dry signal set according to an intonation characteristic parameter and a timbre characteristic parameter. The intonation characteristic parameter includes any one or more of a pitch parameter, a rhythm parameter, or a prosodic parameter. The timbre characteristic parameter includes any one or more of a noise parameter, an energy parameter, or a speed parameter.

[0154] In an embodiment, the *N* virtual three-dimensional spatial acoustic image positions include n1 virtual three-dimensional spatial acoustic image positions on a horizontal plane obtained by partitioning the horizontal plane at intervals of a first preset angle; n2 virtual three-dimensional spatial acoustic image positions on an upper plane obtained by partitioning the upper plane at intervals of a second preset angle, where an angle between the upper plane and the horizontal plane is a first angle threshold; and n3 virtual three-dimensional spatial acoustic image positions on a lower plane obtained by partitioning the lower plane at intervals of a third preset angle, where an angle between the lower plane and the horizontal plane is a second angle threshold, n1, n2, and n3 are positive integers, and a sum of the n1, n2, and n3 is equal to *N*.

[0155] In an embodiment, the processor 903 may further invoke the program instructions to implement the following. An HRTF corresponding to each virtual three-dimensional spatial acoustic image position of the *N* virtual three-dimensional spatial acoustic image positions is obtained. A dry signal corresponding to a target virtual three-dimensional spatial acoustic image position is processed with an HRTF corresponding to the target virtual three-dimensional spatial acoustic image position to obtain a virtual sound signal at the target virtual three-dimensional spatial acoustic image position. The virtual sound signal at the target virtual three-dimensional spatial acoustic image position is a binaural signal. The target virtual three-dimensional spatial acoustic image position is any virtual three-dimensional spatial acoustic image position of the *N* virtual three-dimensional spatial acoustic image positions.

[0156] In an embodiment, the virtual sound signal set further includes a delayed left channel signal and a delayed right channel signal corresponding to each dry signal of *p* dry signals. The processor 903 can further invoke the program instructions to implement the following. *p* dry signals are obtained from the *x* dry signals included in the dry signal set, where *p* is a positive

integer and is less than or equal to *x*. Delay processing is performed on each dry signal of the *p* dry signals for m1 time parameters to obtain m1 delayed dry signals corresponding to each dry signal of the *p* dry signals. A delayed left channel signal corresponding to each dry signal of the *p* dry signals is obtained by superimposing the m1 delayed dry signals corresponding to each dry signal, where m1 is a positive integer. Delay processing is performed on each dry signal of the *p* dry signals for m2 time parameters to obtain m2 delayed dry signals corresponding to each dry signal of the *p* dry signals. A delayed right channel signal corresponding to each dry signal of the *p* dry signals is obtained by superimposing the m2 delayed dry signals corresponding to each dry signal, where m2 is a positive integer.

[0157] In an embodiment, the processor 903 may further invoke the program instructions to implement the following. The background music of the target song is obtained. An energy relationship between the chorus dry sound and the background music is adjusted. An energy relationship between an adjusted chorus dry sound and an adjusted background music satisfies an energy ratio condition. The accompaniment is obtained according to the adjusted chorus dry sound and the adjusted background music.

[0158] In an embodiment, the processor 903 may further invoke the program instructions to implement the following. Spectrum coordination processing is performed on the chorus dry sound at a preset frequency band. A loudness of the background music is obtained. When the loudness of the background music is less than a loudness threshold, the loudness of the background music is increased to the loudness threshold. The accompaniment is obtained according to a spectrum-coordination-processed chorus dry sound and loudness-processed background music.

[0159] Optionally, the memory 902 is further configured to store the program instructions. The processor 903 may further invoke the program instructions to implement the following.

[0160] A user interface is displayed for receiving a selection instruction for a target song. When the selection instruction received on the user interface indicates that the accompaniment mode for the target song is a chorus accompaniment mode, an accompaniment corresponding to the target song is obtained. The accompaniment corresponding to the target song is played. The accompaniment is generated according to a chorus dry sound and background music. The chorus dry sound is generated according to multiple dry signals in a dry signal set. The multiple dry signals in the dry signal set correspond to multiple different virtual three-dimensional spatial

acoustic image positions. The dry signal set is obtained according to dry signals recorded by multiple users for the target song.

[0161] In an embodiment, the chorus dry sound is generated according to a virtual sound signal set. The virtual sound signal set includes a virtual sound signal at each virtual three-dimensional spatial acoustic image position of *N* virtual three-dimensional spatial acoustic image positions, where the virtual sound signal is generated according to an obtained dry signal set. Multiple dry signals in the dry signal set correspond to the *N* virtual three-dimensional spatial acoustic image positions. *N* is an integer greater than 1. The *N* virtual three-dimensional spatial acoustic image positions are different from one another. Each virtual three-dimensional spatial acoustic image position is allowed to correspond to one or more dry signals.

[0162] In an embodiment, selection controls for the accompaniment mode for the target song are displayed on the user interface. The selection controls for the accompaniment mode include a chorus accompaniment mode selection control and an original sound accompaniment mode selection control. Before obtaining the accompaniment corresponding to the target song, the processor 903 may further invoke the program instructions to implement the following. Whether a selection operation for the chorus accompaniment mode selection control is obtained is detected. When the selection operation for the chorus accompaniment mode selection control is obtained, it is determined that the selection instruction received on the user interface indicates that the accompaniment mode for the target song is the chorus accompaniment mode.

[0163] In an embodiment, the processor 903 may further invoke the program instructions to implement the following. An accompaniment request to a server is sent. The accompaniment request contains identification information of the target song. The chorus dry sound and the background music returned by the server in response to the accompaniment request are received. A target chorus dry sound segment is determined from the chorus dry sound. The accompaniment corresponding to the target song is obtained according to a chorus dry sound corresponding to the target chorus dry sound segment and background music corresponding to the target chorus dry sound segment.

[0164] In an embodiment, the processor 903 can further invoke the program instructions to implement the following. A first single-sentence interface is displayed. Each single sentence in text data corresponding to the chorus dry sound is displayed according to a time playing node sequence of the chorus dry sound. The target chorus dry sound segment is determined based on a

single-sentence selection operation on the first single-sentence interface.

[0165] In an embodiment, the processor 903 can further invoke the program instructions to implement the following. A second single-sentence interface is displayed. Each single sentence in text data corresponding to the accompaniment is displayed according to a time playing node sequence of the accompaniment. Whether a mute selection operation for the chorus dry sound in the accompaniment is obtained is detected in a playing process. When the mute selection operation for the chorus dry sound in the accompaniment is obtained in the playing process, playing of the chorus dry sound is terminated at a current time playing node.

[0166] It is understood that problem-solving principles and beneficial effects of the electronic device 90 described in embodiments of the disclosure are similar to problem-solving principles and beneficial effects of embodiments illustrated in FIG. 2 to FIG. 7c and the aforementioned contents in the disclosure, which are not repeated here for the sake of brevity.

[0167] In addition, the disclosure further provides a computer-readable storage medium storing a computer program. The computer program is operable with a processor to implement the method provided in the aforementioned embodiments.

[0168] Embodiments of the disclosure further provide a computer program product or a computer program which contains computer instructions stored in a computer-readable storage medium. A processor of a computer device reads the computer instructions from the computer-readable storage medium and executes the computer instructions, so that the computer device performs the method in embodiments of the disclosure.

[0169] Operations in methods in embodiments of the disclosure can be adjusted in sequence, combined, and deleted according to actual needs.

[0170] Units in apparatus in embodiments of the disclosure can be merged, divided, and deleted according to actual needs.

[0171] Those of ordinary skill in the art can understand that all or some operations of methods in the foregoing embodiments can be implemented by instructing related hardware by a computer program. The program can be stored in a computer-readable storage medium. The program is operable to perform operations including the embodiments of the above-mentioned methods. The computer-readable storage medium can be a magnetic disk, an optical disk, a read-only memory (ROM), a random access memory (RAM), etc.

[0172] The above embodiments are only part of embodiments of the disclosure, which

certainly cannot be used to limit the scope of the disclosure. The ordinary skill in the art can understand all or a part of the process to realize the above embodiments of the disclosure, and the equivalent changes made in accordance with the claims of this disclosure, still belong to the scope of the disclosure covered.

What is claimed is:

1. A method for generating an accompaniment, comprising:

obtaining a dry signal set, wherein the dry signal set comprises *x* dry signals corresponding to a target song, and *x* is an integer greater than 1;

generating a virtual sound signal based on a dry signal corresponding to each virtual three-dimensional spatial acoustic image position of *N* virtual three-dimensional spatial acoustic image positions, wherein the *x* dry signals correspond to the *N* virtual three-dimensional spatial acoustic image positions, *N* is an integer greater than 1, the *N* virtual three-dimensional spatial acoustic image positions are different from one another, and each virtual three-dimensional spatial acoustic image position is allowed to correspond to one or more dry signals of the *x* dry signals;

merging each virtual sound signal in a virtual sound signal set to obtain a chorus dry sound, wherein the virtual sound signal set comprises the virtual sound signal at each virtual three-dimensional spatial acoustic image position of the *N* virtual three-dimensional spatial acoustic image positions; and

performing audio effect synthesis on the chorus dry sound and background music of the target song according to a sound effect optimizing rule to obtain an accompaniment of the target song.

2. The method of claim 1, wherein obtaining the dry signal set comprises:

obtaining an initial dry signal set from an audio database, wherein the audio database comprises initial dry signals recorded by a plurality of users when singing the target song; and

selecting, according to a sound parameter of each initial dry signal, *x* dry signals from the initial dry signal set as the dry signal set.

3. The method of claim 2, wherein the sound parameter comprises an intonation characteristic parameter and a timbre characteristic parameter, wherein the intonation characteristic parameter comprises any one or more of a pitch parameter, a rhythm parameter, or a prosodic parameter; and the timbre characteristic parameter comprises any one or more of a noise parameter, an energy parameter, or a speed parameter.

4. The method of claim 1, wherein the *N* virtual three-dimensional spatial acoustic image positions comprise:

n1 virtual three-dimensional spatial acoustic image positions on a horizontal plane obtained by partitioning the horizontal plane at intervals of a first preset angle;

n2 virtual three-dimensional spatial acoustic image positions on an upper plane obtained by partitioning the upper plane at intervals of a second preset angle, wherein an angle between the upper plane and the horizontal plane is a first angle threshold; and

n3 virtual three-dimensional spatial acoustic image positions on a lower plane obtained by partitioning the lower plane at intervals of a third preset angle, wherein an angle between the lower plane and the horizontal plane is a second angle threshold,

wherein, n1, n2, and n3 are positive integers, and a sum of the n1, n2, and n3 is equal to *N*.

5. The method of any one of claims 1 to 4, wherein generating the virtual sound signal based on the dry signal corresponding to each virtual three-dimensional spatial acoustic image position of the *N* virtual three-dimensional spatial acoustic image positions comprises:

obtaining a head related transfer function (HRTF) corresponding to each virtual three-dimensional spatial acoustic image position of the *N* virtual three-dimensional spatial acoustic image positions; and

processing a dry signal corresponding to a target virtual three-dimensional spatial acoustic image position with an HRTF corresponding to the target virtual three-dimensional spatial acoustic image position to obtain a virtual sound signal at the target virtual three-dimensional spatial acoustic image position, wherein

the virtual sound signal at the target virtual three-dimensional spatial acoustic image position is a binaural signal; and

the target virtual three-dimensional spatial acoustic image position is any virtual three-dimensional spatial acoustic image position of the *N* virtual three-dimensional spatial acoustic image positions.

6. The method of claim 5, wherein obtaining the HRTF corresponding to each virtual three-dimensional spatial acoustic image position of the *N* virtual three-dimensional spatial acoustic

image positions comprises:

obtaining an azimuth angle and an elevation angle of each virtual three-dimensional spatial acoustic image position of the *N* virtual three-dimensional spatial acoustic image positions; and

determining the HRTF corresponding to each virtual three-dimensional spatial acoustic image position according to the azimuth angle and elevation angle of each virtual three-dimensional spatial acoustic image position.

7. The method of any one of claims 1 to 4, wherein the virtual sound signal set further comprises a delayed left channel signal and a delayed right channel signal corresponding to each dry signal of *p* dry signals; and

before merging each virtual sound signal in the virtual sound signal set to obtain the chorus dry sound, the method further comprises:

obtaining the *p* dry signals from the *x* dry signals comprised in the dry signal set, wherein *p* is a positive integer and is less than or equal to *x*;

performing delay processing on each dry signal of the *p* dry signals for m1 time parameters to obtain m1 delayed dry signals corresponding to each dry signal of the *p* dry signals, and obtaining a delayed left channel signal corresponding to each dry signal of the *p* dry signals by superimposing the m1 delayed dry signals corresponding to each dry signal, wherein m1 is a positive integer; and

performing delay processing on each dry signal of the *p* dry signals for m2 time parameters to obtain m2 delayed dry signals corresponding to each dry signal of the *p* dry signals, and obtaining a delayed right channel signal corresponding to each dry signal of the *p* dry signals by superimposing the m2 delayed dry signals corresponding to each dry signal, wherein m2 is a positive integer;

8. The method of any one of claims 1 to 4, wherein performing audio effect synthesis on the chorus dry sound and the background music of the target song according to the sound effect optimizing rule to obtain the accompaniment of the target song comprises:

obtaining the background music of the target song, and adjusting an energy relationship between the chorus dry sound and the background music, wherein an energy relationship

between an adjusted chorus dry sound and adjusted background music satisfies an energy ratio condition,

wherein the accompaniment is obtained according to the adjusted chorus dry sound and the adjusted background music.

9. The method of claim 8, wherein the energy ratio condition is set as a ratio between an energy value of the chorus dry sound and an energy value of the background music of the target song is less than a ratio threshold.

10. The method of claim 8, wherein the energy ratio condition is set as a loudness of the chorus dry sound is 3dB lower than a loudness of the background music of the target song.

11. The method of any one of claims 1 to 4, wherein performing audio effect synthesis on the chorus dry sound and the background music of the target song according to the sound effect optimizing rule to obtain the accompaniment of the target song comprises:

performing spectrum coordination processing on the chorus dry sound at a preset frequency band;

obtaining a loudness of the background music; and

increasing the loudness of the background music to a loudness threshold when the loudness of the background music is less than the loudness threshold,

wherein the accompaniment is obtained according to a spectrum-coordination-processed chorus dry sound and loudness-processed background music.

12. The method of any of claims 1 to 4, wherein the virtual sound signal at the target virtual three-dimensional spatial acoustic image position is a binaural signal, and merging each virtual sound signal in the virtual sound signal set to obtain a chorus dry sound comprises:

merging each left channel signal in the virtual sound signal set; and

merging each right channel signal in the virtual sound signal set;

13. A method for playing and processing an accompaniment, comprising:

displaying a user interface for receiving a selection instruction for a target song;

obtaining an accompaniment corresponding to the target song when the selection instruction received on the user interface indicates that an accompaniment mode for the target song is a chorus accompaniment mode; and

playing the accompaniment corresponding to the target song,

wherein the accompaniment is generated according to a chorus dry sound and background music, wherein the chorus dry sound is generated according to a plurality of dry signals in a dry signal set, the plurality of dry signals in the dry signal set correspond to a plurality of different virtual three-dimensional spatial acoustic image positions, and the dry signal set is obtained according to dry signals recorded by a plurality of users for the target song.

14. The method of claim 13, wherein the chorus dry sound is generated according to a virtual sound signal set, wherein the virtual sound signal set comprises a virtual sound signal at each virtual three-dimensional spatial acoustic image position of N virtual three-dimensional acoustic image positions, and the virtual sound signal is generated according to an obtained dry signal set, wherein

the plurality of dry signals in the dry signal set correspond to the N virtual three-dimensional spatial acoustic image positions, N is an integer greater than 1, the N virtual three-dimensional spatial acoustic image positions are different from one another, and each virtual three-dimensional spatial acoustic image position is allowed to correspond to one or more dry signals.

15. The method of claim 13 or 14, wherein selection controls for the accompaniment mode for the target song are displayed on the user interface, and the selection controls for the accompaniment mode comprise a chorus accompaniment mode selection control and an original sound accompaniment mode selection control; and before obtaining the accompaniment corresponding to the target song, the method further comprises:

detecting whether a selection operation for the chorus accompaniment mode selection control is obtained;

when the selection operation for the chorus accompaniment mode selection control is obtained, determining that the selection instruction received on the user interface indicates that the accompaniment mode for the target song is the chorus accompaniment mode.

16. The method of claim 13, wherein obtaining the accompaniment corresponding to the target song comprises:

sending an accompaniment request to a server, wherein the accompaniment request contains identification information of the target song;

receiving the chorus dry sound and the background music returned by the server in response to the accompaniment request;

determining a target chorus dry sound segment from the chorus dry sound; and

obtaining the accompaniment corresponding to the target song according to a chorus dry sound corresponding to the target chorus dry sound segment and background music corresponding to the target chorus dry sound segment.

17. The method of claim 16, wherein before determining the target chorus dry sound segment from the chorus dry sound, the method further comprises:

displaying a first single-sentence interface, and displaying each single sentence in text data corresponding to the chorus dry sound according to a time playing node sequence of the chorus dry sound, wherein

the target chorus dry sound segment is determined based on a single-sentence selection operation on the first single-sentence interface.

18. The method of claim 13 or 16, wherein after playing the accompaniment corresponding to the target song, the method comprises:

displaying a second single-sentence interface, and displaying each single sentence in text data corresponding to the accompaniment according to a time playing node sequence of the accompaniment;

detecting whether a mute selection operation for the chorus dry sound in the accompaniment is obtained in a playing process; and

when the mute selection operation for the chorus dry sound in the accompaniment is obtained in the playing process, terminating playing of the chorus dry sound at a current time playing node.

19. An electronic device comprising a memory, a processor, and a network interface, wherein the processor is connected to the memory and the network interface, wherein the network interface is configured to provide a network communication function, the memory is configured to store program codes, and the processor is configured to invoke the program codes to perform the method of any one of claims 1 to 18.

20. A computer-readable storage medium storing a computer program operable with a processor to implement the method of any one of claims 1 to 18.