

INTERACTION METHOD AND APPARATUS, ELECTRONIC DEVICE AND STORAGE MEDIUM

TECHNICAL FIELD

[0001] The present disclosure relates to the field of computer technology, and in particular, to an interaction method and apparatus, electronic device and storage medium.

BACKGROUND

[0002] With the rapid development of the Internet, a live stream has become an important way of information dissemination. Due to different time slots for different viewers to watch a live stream, it is impossible for a streamer to do live 24 hours a day in order to meet the needs of the different viewers. The problem may be solved by using virtual human to do live, however, the interaction technology between the virtual human as a streamer and the viewers needs to be researched and developed.

SUMMARY

[0003] According to one aspect of the present disclosure, provided is an interaction method. The method includes: receiving a first message from a client; based on indication content included in the first message, obtaining driving data matching the indication content; and controlling a display interface of client to play a response animation of an interactive object by using the driving data.

[0004] In combination with any of the embodiments provided by the present disclosure, based on the indication content included in the first message, obtaining the driving data matching the indication content includes: obtaining reply content in response to the indication content, wherein the reply content includes a reply text; and based on at least one target text included in the reply text, obtaining control parameters for a set movement of the interactive object matching the target text.

[0005] In combination with any of the embodiments provided by the present disclosure, based on the indication content included in the first message, obtaining the driving data matching the indication content includes: obtaining reply content in response to the indication content, wherein the reply content includes a phoneme sequence; and obtaining control parameters for the interactive object matching the phoneme sequence.

[0006] In combination with any of the embodiments provided by the present disclosure, the control parameters for the interactive object include a posture control vector of at least one local region, and obtaining the control parameters for the interactive object matching

the phoneme sequence includes: obtaining a first encoding sequence corresponding to the phoneme sequence by performing feature encoding on the phoneme sequence; obtaining a feature encoding corresponding to at least one phoneme based on the first encoding sequence; and obtain the posture control vector of at least one local region of the interactive object corresponding to the feature encoding.

[0007] In combination with any of the embodiments provided by the present disclosure, the method further includes: sending to the client an indication message including the reply content to cause the client to display the reply content based on the indication message.

[0008] In combination with any of the embodiments provided by the present disclosure, controlling the display interface of client to play a response animation of the interactive object by using the driving data includes: sending the driving data of the interactive object to the client to cause the client to generate a response animation based on the driving data; and controlling the client to play the response animation on the display interface; or, adjusting parameters of a virtual model for the interactive object based on the driving data; and generating the response animation of the interactive object by using a rendering engine based on the adjusted parameters of the virtual model, and sending the response animation to the client.

[0009] According to one aspect of the present disclosure, provided is an interaction method. The method includes: sending a first message including indication content to a server in response to a user input operation from a client; playing a response animation of an interactive object on a display interface of the client, based on a second message from the server in response to the first message.

[0010] In combination with any of the embodiments provided by the present disclosure, the indication content includes text content; the method further includes: displaying the text content on the client, and/or, playing an audio file corresponding to the text content.

[0011] In combination with any of the embodiments provided by the present disclosure, displaying the text content on the client includes: generating a bullet screen comment message of the text content; and displaying the bullet screen comment message on the display interface of the client.

[0012] In combination with any of the embodiments provided by the present disclosure, the second message includes a reply text in response to the indication content; the method further includes: displaying the reply text on the display interface of the client, and/or determining and playing the audio file corresponding to the reply text.

[0013] In combination with any of the embodiments provided by the present disclosure, the second message includes driving data of the interactive object; playing the response animation of the interactive object on a display interface of the client, based on the second message from the server in response to the first message includes: adjusting parameters of a virtual model for the interactive object based on the driving data; and generating the response animation of the interactive object by using a rendering engine based on the adjusted parameters of the virtual model, and displaying the response animation on the display interface of the client; wherein the driving data includes: control parameters for the interactive object which match a phoneme sequence corresponding to the reply text; and/or control parameters for a set movement of the interactive object which match at least one target text included in the reply text.

[0014] In combination with any of the embodiments provided by the present disclosure, the second message includes a response animation of the interactive object in response to the indication content.

[0015] In combination with any of the embodiments provided by the present disclosure, the user input operation includes a body posture correspondingly made by the user following a body movement image displayed on the display interface; and in response to the user input operation from the client, obtaining a user behavior image including the body posture; identifying body posture information in the user behavior image; and driving the interactive object displayed on the display interface to make a response based on the body posture information.

[0016] In combination with any of the embodiments provided by the present disclosure, driving the interactive object displayed on the display interface to make a response based on the body posture information includes: determining a matching degree between the body posture information and the body posture in the body movement image; and driving the interactive object displayed on the display interface to make a response based on the matching degree.

[0017] In combination with any of the embodiments provided by the present disclosure, driving the interactive object displayed on the display interface to make a response based on the matching degree includes: in response to determining that the matching degree meets a set condition, instructing the interactive object displayed on the display interface to make a first response, wherein the first response includes a body movement and/or a voice prompt which indicates that a displayed posture is qualified; and displaying a next body movement

image; and in response to determining that the matching degree does not meet the set condition, instructing the interactive object displayed on the display interface to make a second response, wherein the second response includes a body movement and/or a voice prompt which indicates that the displayed posture is unqualified; and keeping displaying a current body movement image.

[0018] According to one aspect of the present disclosure, provided is an interaction apparatus. The apparatus includes: a receiving unit, configured to receive a first message from a client; an obtaining unit, configured to, based on indication content included in the first message, obtain driving data matching the indication content; and a driving unit, configured to control a display interface of client to play a response animation of an interactive object by using the driving data.

[0019] In combination with any of the embodiments provided by the present disclosure, the obtaining unit is specifically configured to: obtain reply content in response to the indication content, and the reply content includes a reply text; and based on at least one target text included in the reply text, obtain control parameters for a set movement of the interactive object matching the target text.

[0020] In combination with any of the embodiments provided by the present disclosure, the obtaining unit is configured to: obtain reply content in response to the indication content, and the reply content includes a phoneme sequence; and obtain control parameters for the interactive object matching the phoneme sequence.

[0021] In combination with any of the embodiments provided by the present disclosure, the control parameters for the interactive object include a posture control vector of at least one local region, and the obtaining unit, when obtaining the control parameters for the interactive object matching the phoneme sequence, is configured to: obtain a first encoding sequence corresponding to the phoneme sequence by performing feature encoding on the phoneme sequence; obtain a feature encoding corresponding to at least one phoneme based on the first encoding sequence; and obtain the posture control vector of at least one local region of the interactive object corresponding to the feature encoding.

[0022] In combination with any of the embodiments provided by the present disclosure, the apparatus further includes a sending unit, which is configured to send to the client an indication message including the reply content to cause the client to display the reply content based on the indication message.

[0023] In combination with any of the embodiments provided by the present disclosure, the driving unit is configured to: send the driving data of the interactive object to the client to cause the client to generate a response animation based on the driving data; and control the client to play the response animation on the display interface; or, adjust parameters of a 2D or 3D virtual model for the interactive object based on the driving data; and generate the response animation of the interactive object by using a rendering engine based on the adjusted parameters of the 2D or 3D virtual model, and send the response animation to the client.

[0024] According to one aspect of the present disclosure, provided is an interaction apparatus. The apparatus includes: a sending unit, configured to send a first message including indication content to a server in response to a user input operation from a client; and a playing unit, configured to play a response animation of an interactive object on a display interface of the client, based on a second message from the server in response to the first message.

[0025] In combination with any of the embodiments provided by the present disclosure, the indication content includes text content; the apparatus further includes a first displaying unit, which is configured to display the text content on the client, and/or, determine and play an audio file corresponding to the text content.

[0026] In combination with any of the embodiments provided by the present disclosure, the first displaying unit, when displaying the text content on the client, is configured to: generate a bullet screen comment message of the text content; and display the bullet screen comment message on the display interface of the client.

[0027] In combination with any of the embodiments provided by the present disclosure, the second message includes a reply text in response to the indication content; the apparatus further includes a second displaying unit, which is configured to display the reply text on the display interface of the client, and/or determine and playing the audio file corresponding to the reply text.

[0028] In combination with any of the embodiments provided by the present disclosure, the second message includes driving data of the interactive object; and the playing unit is configured to: adjust parameters of a virtual model for the interactive object based on the driving data; and generate the response animation of the interactive object by using a rendering engine based on the adjusted parameters of the virtual model, and display the response animation on the display interface of the client; wherein the driving data includes:

control parameters for the interactive object which match a phoneme sequence corresponding to the reply text; and/or control parameters for a set movement of the interactive object which match at least one target text included in the reply text.

[0029] In combination with any of the embodiments provided by the present disclosure, the second message includes a response animation of the interactive object in response to the indication content.

[0030] In combination with any of the embodiments provided by the present disclosure, the user input operation includes a body posture correspondingly made by the user following a body movement image displayed on the display interface; and the generating unit is configured to: in response to the user input operation from the client, obtain a user behavior image including the body posture; identify body posture information in the user behavior image; and drive the interactive object displayed on the display interface to make a response based on the body posture information.

[0031] In combination with any of the embodiments provided by the present disclosure, the generating unit is specifically configured to: determine a matching degree between the body posture information and the body posture in the body movement image; and drive the interactive object displayed on the display interface to make a response based on the matching degree.

[0032] In combination with any of the embodiments provided by the present disclosure, the generating unit is configured to: in response to determining that the matching degree meets a set condition, instructing the interactive object displayed on the display interface to make a first response, wherein the first response includes a body movement and/or a voice prompt which indicates that a displayed posture is qualified; and displaying a next body movement image; and in response to determining that the matching degree does not meet the set condition, instructing the interactive object displayed on the display interface to make a second response, wherein the second response includes a body movement and/or a voice prompt which indicates that the displayed posture is unqualified; and keeping displaying a current body movement image.

[0033] According to one aspect of the present disclosure, provided is an electronic device. The electronic device includes a processor and a memory for storing computer instructions executable by the processor, wherein the processor is configured to execute the computer instructions to implement the interaction method proposed in any of the embodiments of the present disclosure.

[0034] According to one aspect of the present disclosure, provided is a computer-readable storage medium. The computer-readable storage medium stores a computer program, when being executed by a processor, the computer program causes the processor to implement the interaction method proposed in any of the embodiments of the present disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

[0035] In order to more clearly illustrate one or more embodiments of the present specification, the following is a brief description of the drawings required for the embodiments, and it will be obvious that the drawings in the following description are only some of the embodiments recorded in the present specification, and that other drawings may be obtained on the basis of these drawings without inventive work to those skilled in the art.

[0036] FIG. 1 illustrates a schematic flowchart of an interaction method in accordance with at least one embodiment of the present disclosure.

[0037] FIG. 2 illustrates a schematic diagram of a live streaming process applying an interaction method in accordance with at least one embodiment of the present disclosure.

[0038] FIG. 3 illustrates a schematic flowchart of a method for obtaining a posture control vector in accordance with at least one embodiment of the present disclosure.

[0039] FIG. 4 illustrates a schematic flowchart of another interaction method in accordance with at least one embodiment of the present disclosure.

[0040] FIG. 5 illustrates a structural schematic diagram of an interaction apparatus in accordance with at least one embodiment of the present disclosure.

[0041] FIG. 6 illustrates a structural schematic diagram of another interaction apparatus in accordance with at least one embodiment of the present disclosure.

[0042] FIG. 7 illustrates a structural schematic diagram of an electronic device in accordance with at least one embodiment of the present disclosure.

[0043] FIG. 8 illustrates a structural schematic diagram of another electronic device in accordance with at least one embodiment of the present disclosure.

DETAILED DESCRIPTION OF THE EMBODIMENTS

[0044] Examples will be described in detail herein, with the illustrations thereof represented in the drawings. When the following descriptions involve the drawings, like numerals in different drawings refer to like or similar elements unless otherwise indicated. The embodiments described in the following examples do not represent all embodiments consistent with the present disclosure. Rather, they are merely examples of apparatuses and

methods consistent with some aspects of the present disclosure as detailed in the appended claims.

[0045] The term “and/or” herein is merely an association relationship for describing associated objects, and indicates that there may be three relationships, for example, A and/or B may indicate that there are three cases: A alone, both A and B, and B alone. In addition, the term “at least one” herein means any one of a plurality or any combination of at least two of a plurality, for example, including at least one of A, B and C, and may mean including any one or more elements selected from the set of A, B, and C.

[0046] A live stream may take place at any time by using virtual human as a streamer, and a 24-hour uninterrupted live stream may be achieved, to meet the different needs of different viewers for the time to watch the live stream. In the case of virtual human as an interactive object for a user in the live stream, how to provide timely feedback to questions raised by the user, and how to interact with the user in a vivid and natural way are problems that need to be solved urgently.

[0047] In view of this, the present disclosure proposes to provide an interaction solution, the interaction solution may be applied to any scenario involving interaction with a virtual interactive object such as a live stream and the like.

[0048] The interaction method provided in embodiments of the present disclosure may be applied to a terminal device or a server, and the terminal device may be, for example, an electronic device with a client, such as a cell phone, a tablet computer and the like, and the present disclosure is not limited to the form of the terminal device. The client is, for example, a client for a live video stream, and includes a client for a live video stream, a client for physical interaction, etc. The server may be any server capable of providing processing capabilities for interactive objects.

[0049] The interactive object may be any kind of interactive object that interacts with the user, which may be a virtual character, and may also be other virtual images, such as a virtual animal, a virtual article, a cartoon image, etc., that can achieve interaction functions. The interactive object may be created based on 2D virtual models or 3D virtual models, and the interactive object is obtained by rendering 2D or 3D virtual models. The user may be a physical person, or a robot, or other intelligent device. The interaction between the interactive object and the user may be in an active or passive manner.

[0050] Exemplarily, in the live video stream scenario, the animation including the interactive object may be displayed on the display interface of the client, and the user may

perform an input operation, such as inputting text, inputting voice, movement triggering, key triggering, etc., on the client of the terminal device to interact with the interactive object.

[0051] FIG. 1 illustrates a schematic flowchart of an interaction method in accordance with at least one embodiment of the present disclosure, the interaction method may be applied on a server. As shown in FIG. 1, the method includes the following steps 101 to 103.

[0052] At step 101, a first message from a client is received.

[0053] Exemplarily, instructions carried in the first message may include information input by the user to perform an input operation via the client, and the input operation of the user includes an input text operation, an input voice operation, a movement-triggered operation, a key-triggered operation, etc. The input information may be sent from the client to the server; optionally, the input information may be displayed directly at the client when the client sends the input information to the server. The form of the indication content carried in the first message includes, but is not limited to, a text, a voice, an image (e.g., an expression and movement image), video, and the like. The specific form of the first message is related to the application scenario. For example, in the live video stream scenario, the client may be a client supporting the function of watching live video stream. The first message may be sent out after the client captures that the user inputs text content on the display interface, the indication content carried in the first message is, for example, the input text content, and the indication content may be displayed on the display interface in the form of bullet screen; another example, in the somatosensory interaction scenario, the first message may be sent out after the client captures a behavior image of the user, and the indication content carried in the first message is, for example, the captured behavior image of the user. Of course, embodiments of the present disclosure do not limit the sending mechanism of the first message and the form of the indicated content carried in the first message.

[0054] At step 102, based on indication content included in the first message, driving data matching the indication content is obtained.

[0055] Exemplarily, the driving data includes one or more of sound driving data, expression driving data, and movement driving data. In one embodiment, the driving data may be prestored on a server or other associated business server, and upon receipt of the first message from the client, a search may be performed on the server or other associated business server based on the indication content to obtain driving data matching the indication content. In another embodiment, the driving data may be generated based on the

indication content, for example, by inputting the indication content into a pre-trained deep learning model, to predict the driving data corresponding to the indication content.

[0056] At step 103, the display interface of the client is controlled to play a response animation of the interactive object, by using the driving data,

[0057] In embodiments of the present disclosure, the interactive object is obtained by rendering a virtual model such as a 2D or 3D virtual model. The virtual model may be generated as a custom model or may be obtained by converting an image or video of a character. Embodiments of the present disclosure are not limited as to how the virtual model is generated.

[0058] The response animation may be generated based on the driving data, by controlling the display interface of the client, such as a live video stream interface, to play the response animation of the interactive object, and capable of showing a response of the interactive object to the first message from the client, which includes outputting a piece of language, and/or making a movement, an expression, etc.

[0059] In embodiments of the present disclosure, the server receives the first message from the client, and obtains driving data matching the indication content based on the indication content included in the first message, and the display interface of the client is controlled by using the driving data, to play the response animation of the interactive object, so as to display the response of the interactive object, such that the interactive object may provide timely feedback to the indication content of the user, and the timely interaction with the user is realized.

[0060] FIG. 2 is an exemplary illustration of an interaction method in accordance with at least one embodiment of the present disclosure applied to a live stream. As shown in Figure 2, the interactive object is a 3D virtual character with the image of a doctor. On the display interface of the client, the process of live stream of the 3D virtual character as a streamer may be displayed, and the user of the client may send the first message carrying the indication content by the input of the indication content on the display interface, accordingly, the server, after receiving the first message from the client, may identify the indication content, such as “how to wash hands”, and thereby the driving data matching the indication content is obtained based on the indication content, and according to the driving data, the client is controlled to display the response of the 3D virtual character to the indication content on “how to wash hands”. For example, the 3D virtual character is controlled to

output a voice corresponding to “how to wash hands”, and at the same time make a movement and/or expression matching the output voice.

[0061] In some embodiments, the indication content includes text content. The response content for the indication content may be obtained according to the following: identifying the linguistic intent expressed by the text content based on a Natural Language Processing (NLP) algorithm; and obtaining the response content matching the linguistic intent.

[0062] In some embodiments, the text content may be processed by using a pre-trained neural network model for natural language processing, such as Convolutional Neural Networks (CNN), Recurrent Neural Network (RNN), Long Short-Term Memory network (LSTM), etc. The category of linguistic intent expressed by the text content is determined, by inputting the text content included in the first message to the neural network model, and classifying the linguistic intent expressed by the text content.

[0063] Since the text content included in the first message may include multiple meanings, the natural language processing algorithm is used to identify the actual intent of the user, such that the content that the user really wants to get can be fed directly, and the interaction experience of the user is improved.

[0064] In some embodiments, a reply content matching the linguistic intent is found in a predetermined database based on the linguistic intent, and further, the server may generate driving data for enabling the interactive object to express the reply content based on the reply content. wherein the database may be deployed on a server or a cloud server, which is not limited in the present disclosure.

[0065] In the case where the linguistic intent is identified, the server may extract parameters, i.e., entities, related to the linguistic intent, from the text content. For example, the entities may be determined through word segmentation and information extraction. In the data corresponding to the linguistic intent classification, the reply text matching the linguistic intent may be further determined by the entities. It should be understood by those skilled in the art that the above means are only exemplary, and other means are be used to obtain the reply text matching the linguistic intent, which is not limited in the present disclosure.

[0066] In some embodiments, the server may generate voice driving data based on the reply content, and the voice driving data includes, for example, a phoneme sequence corresponding to the reply text included in the reply content. a voice corresponding to the

phoneme sequence is generated, and the client is controlled to output the voice, such that the interactive object outputs the voice expressing the content of the reply text.

[0067] In some embodiments, the server may generate movement driving data based on the reply content, such that the interactive object performs a movement that expresses the reply content.

[0068] In an example, where the reply content includes a reply text, movement driving data may be generated based on the reply content by the following: based on at least one target text included in the reply text, control parameters for a set movement matching the target text of the interactive object are obtained.

[0069] The target text may be set keywords, key phrases, key sentences, etc. Illustrated by the example of the keywords “washing hands”, and in the case where the reply text includes “washing hands”, it can be determined that the reply text includes the target text. a set movement matching each target text may be set in advance, and each set movement may be realized by a set of control parameter sequences, for example, the displacement of multiple skeletal points forms a set of control parameters, and the control parameter sequences formed by multiple sets of control parameters are used to adjust model parameters for the interactive object, such that the interactive object makes the set movement.

[0070] In embodiments of the present disclosure, the interactive object replies to the first message in the form of a movement, such that the user is able to obtain an intuitive and vivid response to the first message, and the interaction experience of the user is improved.

[0071] In some embodiments, the voice information corresponding to the target text may be determined; time information for outputting the voice information is obtained; execution time of the set movement corresponding to the target text is determined based on the time information; and the interactive object is controlled to execute the set movement with the control parameters corresponding to the target text in accordance with the execution time.

[0072] In the case where the client is controlled to output the voice based on the phoneme sequence corresponding to the reply text, the time information of the voice corresponding to the target text may be determined, such as time of starting to output the voice corresponding to the target text, time of ending the output, and duration. The execution time of the set movement corresponding to the target text is determined based on the time information, and within the execution time, or within a certain range of the execution time, the interactive object is controlled to execute the set movement with the control parameters corresponding to the target text.

[0073] In embodiments of the present disclosure, for each target text, the duration of outputting the corresponding voice and the duration of controlling the movement according to the corresponding control parameters are consistent or similar, so that the time of outputting the voice corresponding to the target text and the time of performing the movement of the interactive object is matched, and thereby the voice and movement of the interactive object are synchronized and coordinated, and the user generates the feeling that the interactive object responds during the live stream, and the user's experience of interacting with the streamer during the live stream is improved.

[0074] In some embodiments, posture driving data may be generated based on the reply text, to enable the client to display a posture of the interactive object that matches the voice corresponding to the reply text, such as making corresponding an expression and a movement.

[0075] In an example, the reply content may also include a phoneme sequence, or, in the case where the reply content includes a response text, a phoneme sequence corresponding to the reply text may also be extracted, and after obtaining the reply content including the phoneme sequence, control parameters for the interactive object matching the phoneme sequence may be obtained. wherein the control parameters for the interactive object include a posture control vector of at least one local region, and obtaining the control parameters for the interactive object matching the phoneme sequence includes: obtaining a first encoding sequence corresponding to the phoneme sequence by performing feature encoding on the phoneme sequence; obtaining a feature encoding corresponding to at least one phoneme based on the first encoding sequence; and obtaining a posture control vector of at least one local region of the interactive object corresponding to the feature encoding.

[0076] In some embodiments, the client is controlled to play the voice corresponding to the reply text and display a response animation of the posture of the interactive object matching the voice, such that the response of the interactive object is made more anthropomorphic, and more vivid and natural, and the interaction experience of the user is improved.

[0077] In an embodiment where the control parameters for the interactive object include a posture control vector of at least one local region, the posture control vector may be obtained by the following ways.

[0078] First, an encoding sequence corresponding to the phoneme sequence is obtained by performing feature encoding on the phoneme sequence corresponding to the reply text. Here,

the encoding sequence corresponding to the phoneme sequence of the text data is referred to as a first encoding sequence in order to differentiate from the subsequently mentioned encoding sequence.

[0079] For multiple phonemes included in the phoneme sequence, a sub-encoding sequence corresponding to each phoneme is generated.

[0080] In an example, it is detected whether there is a first phoneme at various time points, and the first phoneme is any one of the multiple phonemes; an encoding value at a time point with the first phoneme is set to a first value, and an encoding value at a time point without the first phoneme is set to a second value. The sub-encoding sequence corresponding to the first phoneme is obtained after assigning the encoding value at various time points. For example, an encoding value at a time point with the first phoneme may be set to 1, and an encoding value at a time point without the first phoneme may be set to 0. It will be understood for those skilled in the art that the above setting of the encoding values is only an example, and the encoding values may also be set to other values, but the present disclosure is not limited thereto.

[0081] Then, the first encoding sequence corresponding to the phoneme sequence is obtained based on the sub-encoding sequences respectively corresponding to the multiple phonemes.

[0082] In an example, for the sub-encoding sequence corresponding to the first phoneme, a Gaussian filter may be used to perform a Gaussian convolution operation on successive values of the first phoneme in time, to filter a matrix corresponding to the feature encoding, and to smooth the movement transition of the mouth region during phoneme conversion.

[0083] FIG. 3 illustrates a schematic flowchart of a method for obtaining a posture control vector as proposed in at least one embodiment of the present disclosure. As shown in FIG. 3, a phoneme sequence 310 includes phonemes j, i1, j and ie4 (for brevity, only some of the phonemes are shown), and for each of the phonemes j, i1 and ie4, sub-encoding sequences 321, 322 and 323 respectively corresponding to the above phonemes are obtained. In each of the sub-coding sequences, an encoding value at time (time unit in seconds(s) in Figure 3) with the phoneme is a first value (e.g., 1), and an encoding value at time (time unit in seconds(s) in Figure 3) without the phoneme is a second value (e.g., 0). Illustrated by the example of sub-encoding sequence 321, a value of the sub-encoding sequence 321 at time with the phoneme j in the phoneme sequence 310 is the first value, and a value of sub-

encoding sequence 321 at time without the phoneme j is the second value. All sub-encoding sequences form a first encoding sequence 320.

[0084] Then, a feature encoding corresponding to at least one phoneme is obtained based on the first encoding sequence.

[0085] Based on the encoding values of the sub-encoding sequences 321, 322 and 323 respectively corresponding to the phonemes j, i1 and ie4, and the durations of the corresponding phonemes in the three sub-encoding sequences, i.e., the duration of the phoneme j in the sub-encoding sequence 321, the duration of the phoneme i1 the in sub-encoding sequence 322, and the duration of the phoneme ie4 in the sub-encoding sequence 323, feature information of the sub-encoding sequences 321, 322 and 323 is obtained.

[0086] In an example, a Gaussian filter may be used to respectively perform a Gaussian convolution operation on successive values in time of the phonemes j, i1 and ie4 in the sub-encoding sequences 321, 322 and 323, to smooth the feature encodings to obtain a smoothed first encoding sequence 330. That is, the Gaussian filter is used to perform the Gaussian convolution operation on successive values from 0 to 1 in time of the phonemes, so that stages of change of the encoding values from the second value to the first value or from the first value to the second value in the respective encoding sequence become smooth. For example, the values of the encoding sequence besides 0 and 1 also present intermediate state values, such as 0.2, 0.3, etc., and the posture control vectors obtained based on these intermediate state values make the movement transitions and expression changes of the interactive character more smooth and more natural, and the interaction experience of a target object is improved.

[0087] In some embodiments, a feature encoding corresponding to at least one phoneme may be obtained by performing a sliding window operation on the first encoding sequence. wherein the first encoding sequence may be an encoding sequence that has undergone a Gaussian convolution operation.

[0088] A sliding window operation is performed on the encoding sequence with a set length of time window and a set step size, and a feature encoding in the time window is used as a feature encoding of the corresponding at least one phoneme, after the sliding window operation is completed, a second encoding sequence is obtained based on multiple feature encodings. As shown in FIG. 3, a time window with a set length is slid over the first encoding sequence 320 or the smoothed first encoding sequence 330, to obtain feature encoding 1, feature encoding 2, feature encoding 3, and so on, respectively, i.e., the feature

encodings 1, 2, 3, ..., and M are obtained after traversing the first encoding sequence, so as to obtain a second encoding sequence 340. Where M is a positive integer, and a value thereof is determined based on a length of the first encoding sequence, a length of the time window and a sliding step size of the time window.

[0089] Based on the feature encoding 1, 2, 3, ..., and M, the corresponding posture control vectors 1, 2, 3, ..., and M are obtained, respectively, such that a posture control vector sequence 350 is obtained.

[0090] The posture control vector sequence 350 and the second encoding sequence 340 are aligned in time. Since each feature encoding in the second encoding sequence is obtained based on at least one phoneme in the phoneme sequence, each feature vector in the posture control vector sequence 350 is also obtained based on at least one phoneme in the phoneme sequence. While the phoneme sequence corresponding to the text data is played, the interactive object is driven to make a movement according to the posture control vector sequence, i.e., the interactive object can be driven to make voices according to the text content and make a movement synchronized with the voices at the same time, such that the target object feels that the interactive object is speaking, and the interactive experience of the target object is improved.

[0091] Assuming that the feature encoding starts to be output at a set time of a first time window, the posture control vector before the set time may be set to a default value, that is, when the phoneme sequence is just started to be played, the interactive object is driven to make a default movement, and after the set time, the interactive object is driven to make a movement by using the posture control vector sequence obtained according to the first encoding sequence. Illustrated by the example of FIG. 3, the feature coding 1 starts to be output at time t_0 , the default posture control vector corresponds to the time before time t_0 .

[0092] In some embodiments, when the time interval between phonemes in the phoneme sequence is greater than a set threshold, the interactive object is driven to make a movement according to the set posture control vector of the local region. That is, when the interactive character pauses for a long time, the interactive object is driven to make a set movement. For example, when there is a long pause in the output voice, the interactive character may be driven to make a smiling expression or make a slight body sway to avoid the interactive character standing upright with no expression when there is a long pause, so that the interactive object speaks naturally and smoothly, and the interactive feeling of the target object is improved.

[0093] In some embodiments, for at least one target text included in the reply text, control parameters for a set movement of the interactive object matching the at least one target text are obtained, to drive the interactive object to execute the set movement; for the reply content other than the at least one target text, control parameters for the interactive object may be obtained according to the phoneme corresponding to the reply content, so as to drive the interactive object to make a posture matching the pronunciation of the reply content, such as an expression and a movement.

[0094] Illustrated by the example of the live stream shown in FIG. 2, in the case where the received first message includes the text content “how to wash hands”, through a natural language processing algorithm, the user's linguistic intent may be identified as “advice on how to wash hands”. By searching in the preset database, the content in response to how to wash hands may be obtained and used as the reply text. By generating movement driving data, voice driving data and posture driving data according to the reply text, the interactive object may answer the question “how to wash hands” by voice, and at the same time make an expression and a movement matching the pronunciation, and demonstrate how to wash hands by body movements.

[0095] In some embodiments, an indication message including the reply content may be send to the client to cause the client to display the reply text based on the indication message.

[0096] For example, for the reply text in response to the question of “how to wash hands”, the instruction message including the reply text may be sent to the client to display the instruction message in the form of text on the client, so that the user may more accurately receive the information conveyed by the interactive object.

[0097] In some embodiments, the virtual model corresponding to the interactive object (the virtual model may be a 2D or 3D virtual model) may be stored on the client. In the case, the driving data of the interactive object may be sent to the client to cause the client to generate a response animation based on the driving data and control the client to play the response animation. For example, the client is controlled to adjust parameters of a virtual model for the interactive object based on the control parameters included in the driving data and generate the response animation of the interactive object by using a rendering engine based on the adjusted parameters of the virtual model and play the response animation to response to the first message. When the virtual model is a 2D virtual model, the parameters of the virtual model are parameters of a 2D virtual model, and when the virtual model is a 3D virtual model, the parameters of the virtual model are parameters of a 3D virtual model.

For another example, the server may determine a control instruction for controlling the response mode of the interactive object based on the driving data, and send the control instruction to the client, so that the client displays a screen for response of the interactive object based on the control instruction.

[0098] In the case where the amount of data of the virtual model of the interactive object is small and the performance usage for the client is not high, the driving data may be sent to the client, such that the client may generate the response animation based on the driving data, to conveniently and flexibly display the screen for response of the interactive object.

[0099] In some embodiments, the virtual model corresponding to the interactive object is stored on a server or a cloud server. In the case, the parameters of the virtual model for the interactive object may be adjusted based on the driving data; and the response animation of the interactive object may be generated by using a rendering engine based on the adjusted parameters of the virtual model, and the response animation is sent to the client, and the response animation shows the movement or expression of the interactive object. By sending the response animation to the client to realize the response of the interactive object, the lag caused by rendering operations of the client may be avoided, and a high-quality response animation may be displayed on the client, and the interactive experience of the user is improved.

[0100] FIG. 4 illustrates a schematic flowchart of another interaction method in accordance with at least one embodiment of the present disclosure. The interaction method may be applied to the client. The method includes the following steps 401 to 402.

[0101] At step 401, a first message including indication content is sent to a server in response to a user input operation from a client.

[0102] Exemplarily, the user input operation includes an input text operation, an input voice operation, a movement trigger operation, a key trigger operation, etc. The first message is sent to the server in response to the user input operation, and the indication content carried in the first message includes, but is not limited to, one or more of text, voice, image (e.g., an expression image, a movement image), video, etc. For example, in the live video stream scenario, the client may be a client supporting the function of watching live video stream. The first message may be sent out after the client captures that the user inputs text content on the display interface, the indication content carried in the first message is, for example, the input text content, and the indication content may be displayed on the display interface in the form of bullet screen. For another example, in a somatosensory

interaction scenario, the first message may be sent after the client collects a user behavior image, and the indication content carried in the first message is, for example, the user behavior image. Of course, embodiments of the present disclosure do not limit the sending mechanism of the first message and the form of the indicated content carried in the first message.

[0103] At step 402, a response animation of an interactive object is played on a display interface of the client, based on a second message from the server in response to the first message.

[0104] The second message is generated by the server in response to the indication content included in the first message and is used to make the client display an interactive object that responds to the indication content.

[0105] In embodiments of the present disclosure, the interactive object is obtained by rendering a virtual model such as a 2D or 3D virtual model. The virtual model may be generated as a custom model or may be obtained by converting an image or video of a character. Embodiments of the present disclosure are not limited as to how the virtual model is generated.

[0106] In embodiments of the present disclosure, the first message including the indication content is sent to the server based on the user input operation, and the interactive object is displayed on the client, based on the second message of the server in response to the first message, such that the interactive object may provide timely feedback to the indication content of the user, and realize timely interaction with the user.

[0107] In some embodiments, the indication content includes text content; the method further includes: displaying the text content on the display interface of the client, and/or, determining and playing an audio file corresponding to the text content. That is, the text content input by the user may be displayed on the client; and the audio file corresponding to the text content may also be played on the client, and the voice corresponding to the text content may be output.

[0108] In some embodiments, displaying the text content on the client includes: generating a bullet screen comment message of the text content; and displaying the bullet screen comment message on the display interface of the client.

[0109] In a live video stream scenario, for the text content input by the user, the bullet screen comment message may be correspondingly generated, and the bullet screen comment message may be displayed on the display interface of the client. Illustrated by the example

of FIG. 2, in the case where the user inputs “how to wash hands” on the live interactive interface of the client, the bullet screen comment message “how to wash hands” corresponding to the text content may be displayed on the display interface.

[0110] In some embodiments, the second message includes a reply text in response to the indication content; the method further includes: displaying the reply text on the display interface of the client, and/or determine and playing the audio file corresponding to the reply text.

[0111] The reply text of the indication content may be obtained in the following manner: identifying the linguistic intent expressed by the text content and searching for a reply text matching the linguistic intent from a preset database. The details may be found in the above-mentioned embodiments, and no more tautology here.

[0112] Illustrated by the example of the live video stream scenario, the reply text in response to the bullet screen comment message of the user may be displayed on the display interface in the same as the form of bullet screen comment message; and the audio file corresponding to the reply text may be played on the display interface, i.e., the voice corresponding to the reply text may be output, the bullet screen comment message of the user may be replied accurately and intuitively, and the interactive experience of the user is improved.

[0113] In some embodiments, the second message includes: control parameters for the interactive object which match a phoneme sequence corresponding to the reply text; and/or control parameters for a set movement of the interactive object which match at least one target text included in the reply text. Playing the response animation of the interactive object on a display interface of the client, based on the second message from the server in response to the first message includes: adjusting parameters of a virtual model for the interactive object based on the driving data; and generating the response animation of the interactive object by using a rendering engine based on the adjusted parameters of the virtual model, and displaying the response animation on the display interface of the client. wherein the method for generating the control parameters for the interactive object which match a phoneme sequence corresponding to the reply text, and the control parameters for a set movement of the interactive object which match at least one target text included in the reply text, may be found in the above-mentioned embodiments, and no more tautology here.

[0114] In the case where the amount of data of the virtual model of the interactive object is small and the performance usage for the client is not high, the client obtains the driving

data, and the response animation is generated based on the driving data, such that the screen for response of the interactive object may be conveniently and flexibly displayed.

[0115] In some embodiments, the second message further includes a response animation of the interactive object in response to the indication content; and playing the response animation of the interactive object on a display interface of the client, based on the second message from the server in response to the first message includes: displaying the response animation on the display interface of the client.

[0116] In some embodiments, the virtual model corresponding to the interactive object is stored on a server or a cloud server. in the case, the response animation may be generated on a server or on a cloud server. The specific method for generating the response animation may be found in the above-mentioned embodiments, and no more tautology here.

[0117] By sending the response animation to the client to realize the response of the interactive object, the lag caused by rendering operations of the client may be avoided, and a high-quality response animation may be displayed on the client, and the interactive experience of the user is improved.

[0118] In some embodiments, the user input operation includes a body posture correspondingly made by the user following a body movement image displayed on the display interface; and in the case, in response to the user input operation from the client, the method further includes: obtaining a user behavior image including the body posture; identifying body posture information in the user behavior image; and driving the interactive object displayed on the display interface to make a response based on the body posture information.

[0119] In some embodiments, driving the interactive object displayed on the display interface to make a response based on the body posture information includes: determining a matching degree between the body posture information and the body posture in the body movement image; and driving the interactive object displayed on the display interface to make a response based on the matching degree.

[0120] In some embodiments, driving the interactive object displayed on the display interface to make a response based on the matching degree includes: in response to determining that the matching degree meets a set condition, instructing the interactive object displayed on the display interface to make a first response, wherein the first response includes a body movement and/or a voice prompt which indicates that a displayed posture is qualified; and displaying a next body movement image; and in response to determining

that the matching degree does not meet the set condition, instructing the interactive object displayed on the display interface to make a second response, wherein the second response includes a body movement and/or a voice prompt which indicates that the displayed posture is unqualified; and keeping displaying a current body movement image.

[0121] Exemplarily, the following are some embodiments in which the embodiments of the present disclosure are applied in the live video scenario.

[0122] In some embodiments, a first message received from the client is a user bullet screen text transmitted by a live stream platform.

[0123] In some embodiments, after analyzing the intention of the bullet screen by a natural language processing algorithm, a corresponding response is obtained, and then the content of the response is broadcast by the interactive object. In addition, the movement corresponding to the content of the response may be further displayed by the interactive object.

[0124] In some embodiments, the natural language processing capability of the client is directly integrated to perform natural language processing on the indication content included in the first message, so as to obtain a reply text matching and conforming the linguistic intent of the indication content, and the output text corresponding to the reply text is directly provided to the interactive object for playing.

[0125] In some embodiments, the interactive object may imitate the speech content of the user. For example, the voice is converted into text, and a voice feature is obtained based on the voice, and the voice corresponding to the text is output based on the voice feature, to realize the interactive object to imitate the speech content of the user.

[0126] In some embodiments, the interactive object may also make a page display of the content returned by the natural language processing, which can be displayed in accordance with the pre-designed content to be displayed, as well as the interactive way to display UI content, thus making the display of the reply content more eye-catching and attracting the user's attention.

[0127] In the above embodiment, live streaming real-time interaction is allowed, during the live stream, the user may interact with the interactive object in real time and get feedback. The present disclosure provides a new method of live TV, which may be carried out continuously and may also automatically produce video content.

[0128] The interactive object may be presented as virtual human in 3D form. The combination of Artificial Intelligence (AI) simulation animation generation capability and

natural language understanding capability enables the virtual human to communicate with the user in a vocal and physical manner like a real person. The virtual human may generate corresponding mouth shape, expression, eyes and whole-body movement according to the reply content, and finally output high-quality, audio-video synchronized voice and multi-dimensional animation content, the complete virtual human image is presented to the user naturally.

[0129] In some implementations, the virtual human may be quickly linked to content service libraries of different knowledge fields and efficiently applied to more industries. Furthermore, virtual human with various styles such as hyper realism and cartoon may be provided according to the needs of different scenes and support intelligent interaction with the user through AI technologies such as face recognition and posture recognition. For example, the virtual human with a hyper-realistic style may be used as an intelligent reception clerk of a bank, a business hall, or a service hall to communicate with customers authentically and effectively, and service quality and customer satisfaction are improved.

[0130] In some embodiments, the virtual human with a cartoon style may be applied to fun and interactive-oriented scenarios, such as an intelligent guide in an offline superstore, or an intelligent coach, a virtual teacher, etc., to achieve the purpose of customer attraction, stimulating interest, strengthening the teaching effect, etc.

[0131] At least one embodiment of the present disclosure further provides an interaction apparatus that may be applied to a server. As shown in FIG. 5, the interaction apparatus 50 includes: a receiving unit 501, configured to receive a first message from a client; an obtaining unit 502, configured to, based on indication content included in the first message, obtain driving data matching the indication content; and a driving unit 503, configured to control a display interface of client to play a response animation of an interactive object by using the driving data.

[0132] In some embodiments, the obtaining unit 502 is specifically configured to: obtain reply content in response to the indication content, and the reply content includes a reply text; and based on at least one target text included in the reply text, obtain control parameters for a set movement of the interactive object matching the target text.

[0133] In some embodiments, the obtaining unit 502 is configured to: obtain reply content in response to the indication content, and the reply content includes a phoneme sequence; and obtain control parameters for the interactive object matching the phoneme sequence.

[0134] In some embodiments provided by the present disclosure, the control parameters for the interactive object include a posture control vector of at least one local region, and the obtaining unit 502, when obtaining the control parameters for the interactive object matching the phoneme sequence, is configured to: obtain a first encoding sequence corresponding to the phoneme sequence by performing feature encoding on the phoneme sequence; obtain a feature encoding corresponding to at least one phoneme based on the first encoding sequence; and obtain the posture control vector of at least one local region of the interactive object corresponding to the feature encoding.

[0135] In some embodiments, the apparatus further includes a sending unit, which is configured to send to the client an indication message including the reply content to cause the client to display the reply content based on the indication message.

[0136] In some embodiments, the driving unit 503 is configured to: send the driving data of the interactive object to the client to cause the client to generate a response animation based on the driving data; and control the client to play the response animation on the display interface; or, adjust parameters of a 2D or 3D virtual model for the interactive object based on the driving data; and generate the response animation of the interactive object by using a rendering engine based on the adjusted parameters of the 2D or 3D virtual model, and send the response animation to the client.

[0137] At least one embodiment of the present disclosure further provides another interaction apparatus that may be applied to a client. As shown in FIG. 6, the interaction apparatus 60 includes: a sending unit 601, configured to send a first message including indication content to a server in response to a user input operation from a client; and a playing unit 602, configured to play a response animation of an interactive object on a display interface of the client, based on a second message from the server in response to the first message. The interactive object is obtained by rendering a virtual model such as a 2D or 3D virtual model.

[0138] In some embodiments, the indication content includes text content; the apparatus further includes a first displaying unit, which is configured to display the text content on the display interface of the client, and/or, determine and play an audio file corresponding to the text content.

[0139] In some embodiments, the first displaying unit, when displaying the text content on the client, is configured to: generate a bullet screen comment message of the text content; and display the bullet screen comment message on the display interface of the client.

[0140] In some embodiments, the second message includes a reply text in response to the indication content; the apparatus further includes a second displaying unit, which is configured to display the reply text on the display interface of the client, and/or determine and play the audio file corresponding to the reply text.

[0141] In some embodiments, the second message includes driving data of the interactive object; and the playing unit 602 is configured to: adjust parameters of a virtual model for the interactive object based on the driving data; and generate the response animation of the interactive object by using a rendering engine based on the adjusted parameters of the virtual model, and display the response animation on the display interface of the client; wherein the driving data includes: control parameters for the interactive object which match a phoneme sequence corresponding to the reply text; and/or control parameters for a set movement of the interactive object which match at least one target text included in the reply text.

[0142] In some embodiments, the second message includes a response animation of the interactive object in response to the indication content.

[0143] In some embodiments, the user input operation includes a body posture correspondingly made by the user following a body movement image displayed on the display interface; and the generating unit is configured to: in response to the user input operation from the client, obtain a user behavior image including the body posture; identify body posture information in the user behavior image; and drive the interactive object displayed on the display interface to make a response based on the body posture information.

[0144] In some embodiments, the generating unit is specifically configured to: determine a matching degree between the body posture information and the body posture in the body movement image; and drive the interactive object displayed on the display interface to make a response based on the matching degree.

[0145] In some embodiments, the generating unit is configured to: in response to determining that the matching degree meets a set condition, instructing the interactive object displayed on the display interface to make a first response, wherein the first response includes a body movement and/or a voice prompt which indicates that a displayed posture is qualified; and displaying a next body movement image; and in response to determining that the matching degree does not meet the set condition, instructing the interactive object displayed on the display interface to make a second response, wherein the second response includes a body movement and/or a voice prompt which indicates that the displayed posture is unqualified; and keeping displaying a current body movement image.

[0146] At least one embodiment of the present disclosure provides an electronic device. As shown in FIG. 7, the electronic device includes a processor 702 and a memory 701 for storing computer instructions executable by the processor 702, wherein the processor 702 is configured to execute the computer instructions to implement the interaction method proposed in the server embodiments of the present disclosure.

[0147] At least one embodiment of the present disclosure further provides a computer-readable storage medium. The computer-readable storage medium stores a computer program, when being executed by the processor 702, the computer program causes the processor 702 to implement the interaction method proposed in the server embodiments of the present disclosure.

[0148] At least one embodiment of the present disclosure provides an electronic device. As shown in FIG. 8, the electronic device includes a processor 802 and a memory 801 for storing computer instructions executable by the processor 802, wherein the processor 802 is configured to execute the computer instructions to implement the interaction method proposed in the client embodiments of the present disclosure.

[0149] At least one embodiment of the present disclosure further provides a computer-readable storage medium. The computer-readable storage medium stores a computer program, when being executed by the processor 802, the computer program causes the processor 802 to implement the interaction method proposed in the client embodiments of the present disclosure.

[0150] It will be understood for those skilled in the art that one or more embodiments of the present specification may be provide as a method, a system, or a computer program product. Therefore, one or more embodiments of the present specification may take a form of full hardware embodiment, full software embodiment, or embodiment combining software and hardware. Further, one or more embodiments of the present specification may take a form of computer program product implemented on one or more computer-readable storage mediums (including but not limited to disk memory, CD-ROM, optical memory, etc.) including computer available program codes.

[0151] Each embodiment in the present specification is described in a progressive manner, and the same and similar parts between each embodiment may be referred to each other, and each embodiment focuses on the differences from other embodiments. In particular, for the data processing device embodiment, since it substantially corresponds to the method

embodiment, the description thereof is relatively simple, and relevant parts can be referred to the description of the method embodiment.

[0152] Specific embodiments of the present specification are described above. Other embodiments are within the scope of the claims. In some cases, the acts or steps in the claims may be performed in a different order than in the embodiments and still achieve the desired result. In addition, the processes depicted in the drawings do not necessarily require the particular order or the successive order shown in order to achieve the desired results. In some embodiments, multitasking and parallel processing are also possible or may be advantageous.

[0153] The embodiments of the subject matter and functional operations described in the present specification may be implemented in the following: a digital electronic circuit, tangibly embodied computer software or firmware, computer hardware including the structure disclosed in the present specification and structural equivalent thereof, or a combination of one or more of them. The embodiments of the subject matter described in the present specification may be implemented as one or more computer programs, i.e., one or more modules of computer program instructions encoded on a tangible non-transitory program carrier to be executed by a data processing device or to control the operation of a data processing device. Alternatively or additionally, the program instructions may be encoded on an artificially generated propagation signal, such as a machine-generated electrical, optical, or electromagnetic signal, which is generated to encode and transmit the information to a suitable receiver device for execution by the data processing device. The computer storage medium may be a machine-readable storage device, a machine-readable storage substrate, a random or serial access memory device, or a combination of one or more of them.

[0154] The processing and logic flow described in the present specification may be performed by one or more programmable computers executing one or more computer programs to perform corresponding functions by operating on input data and generating outputs. The processing and logic flow may also be performed by a dedicated logic circuit - such as FPGA (Field Programmable Gate Array) or ASIC (Application Specific Integrated Circuit) - and the device may also be implemented as a dedicated logic circuit.

[0155] A computer suitable for executing a computer program include, for example, a general-purpose and/or specialized microprocessor, or any other type of central processing unit. Typically, the central processing unit will receive instructions and data from read-only

memory and/or random-access memory. The basic components of a computer include a central processing unit for implementing or executing instructions and one or more memory devices for storing instructions and data. Typically, the computer will also include one or more mass storage devices for storing data, such as a disk, a magnetic disk, or an optical disk, or the computer will be operable to couple to the mass storage device to receive data from or transmit data to it, or both. However, it is not necessary for a computer to have such a device. In addition, a computer can be embedded in another device, such as a cell phone, a personal digital assistant (PDA), a mobile audio or video player, a game console, a global positioning system (GPS) receiver, or a portable storage device such as a universal serial bus (USB) flash drive, just to name a few.

[0156] A computer-readable storage medium suitable for storing computer program instructions and data includes all forms of non-transitory memory, media, and memory devices, including, for example, a semiconductor memory device (e.g., EPROM, EEPROM, and a flash memory device), a disk (e.g., an internal hard disk or a removable disk), a magnetic disk, and a CD ROM and DVD-ROM disk. The processor and memory may be supplemented by or incorporated into a dedicated logic circuit.

[0157] While the specification includes many specific embodiment details, these should not be interpreted as limiting the scope of any invention or the scope of protection claimed but are primarily intended to describe the features of specific embodiments of a particular invention. Certain features described in multiple embodiments in the present specification may also be implemented in combination in a single embodiment. On the other hand, the various features described in a single embodiment may also be implemented separately in multiple embodiments or implemented in any suitable sub-combination. Further, although the features may function in some combinations as described above and even initially claimed protection, one or more features from the combination claimed protection may be removed from the combination in some cases, and the combination claimed protection may refer to a sub-combination or a variant of a sub-combination.

[0158] Similarly, although the operations are depicted in a particular order in the drawings, this should not be construed as requiring that the operations be performed in the particular order shown or sequentially, or that all of the illustrated operations be performed to achieve the desired result. In some embodiments, multitasking and parallel processing are also possible or may be advantageous. Furthermore, the separation of the various system modules and components in the above embodiments should not be understood to require

such separation in all embodiments, and it should be understood that the program components and systems described may often be integrated together in a single software product or packaged into multiple software products.

[0159] Therefore, particular embodiments of the subject matter have been described. Other embodiments are within the scope of the claims. In some cases, the acts or steps in the claims may be performed in a different order than in the embodiments and still achieve the desired result. In addition, the processes depicted in the drawings do not necessarily require the particular order or successive order shown in order to achieve the desired results. In some embodiments, multitasking and parallel processing are also possible or may be advantageous.

[0160] The above are only some embodiments of the present disclosure and are not intended to limit the present disclosure, any modifications, equivalent substitutions, improvements, etc. made within the spirit and principle of the present disclosure shall be covered by the scope of the present disclosure.

CLAIMS

1. An interaction method, comprising:

receiving a first message from a client;

based on indication content comprised in the first message, obtaining driving data matching the indication content; and

controlling a display interface of client to play a response animation of an interactive object by using the driving data.

2. The method according to claim 1, wherein based on the indication content comprised in the first message, obtaining the driving data matching the indication content comprises:

obtaining reply content in response to the indication content, wherein the reply content comprises a reply text; and

based on at least one target text comprised in the reply text, obtaining control parameters for a set movement of the interactive object matching the target text.

3. The method according to claim 1 or 2, wherein based on the indication content comprised in the first message, obtaining the driving data matching the indication content comprises:

obtaining reply content in response to the indication content, wherein the reply content comprises a phoneme sequence; and

obtaining control parameters for the interactive object matching the phoneme sequence.

4. The method according to claim 3, wherein

the control parameters for the interactive object comprise a posture control vector of at least one local region, and obtaining the control parameters for the interactive object matching the phoneme sequence comprises:

obtaining a first encoding sequence corresponding to the phoneme sequence by performing feature encoding on the phoneme sequence;

obtaining feature encoding corresponding to at least one phoneme based on the first encoding sequence; and

obtain the posture control vector of at least one local region of the interactive object corresponding to the feature encoding.

5. The method according to any one of claims 1 to 4, further comprising:

sending to the client an indication message comprising the reply content to cause the client to display the reply content based on the indication message.

6. The method according to any one of claims 1 to 5, wherein controlling the display interface of client to play a response animation of the interactive object by using the driving data comprises:

sending the driving data of the interactive object to the client to cause the client to generate a response animation based on the driving data; and controlling the client to play the response animation on the display interface; or

adjusting parameters of a virtual model for the interactive object based on the driving data; and generating the response animation of the interactive object by using a rendering engine based on the adjusted parameters of the virtual model; and sending the response animation to the client.

7. An interaction method, comprising:

sending a first message comprising indication content to a server in response to a user input operation from a client;

playing a response animation of an interactive object on a display interface of the client, based on a second message from the server in response to the first message.

8. The method according to claim 7, wherein the indication content comprises text content; the method further comprises: displaying the text content on the client, and/or, playing an audio file corresponding to the text content.

9. The method according to claim 8, wherein displaying the text content on the client comprises:

generating a bullet-screen comment message of the text content; and

displaying the bullet-screen comment message on the display interface of the client.

10. The method according to any one of claims 7 to 9, wherein the second message comprises a reply text in response to the indication content;

the method further comprises:

displaying the reply text on the display interface of the client, and/or

determining and playing the audio file corresponding to the reply text.

11. The method according to any one of claims 8 to 10, wherein the second message comprises driving data of the interactive object;

playing the response animation of the interactive object on a display interface of the client, based on the second message from the server in response to the first message comprises:

adjusting parameters of a virtual model for the interactive object based on the driving data;

generating the response animation of the interactive object by using a rendering engine based on the adjusted parameters of the virtual model; and

displaying the response animation on the display interface of the client;

wherein the driving data comprises: control parameters for the interactive object which match a phoneme sequence corresponding to the reply text; and/or control parameters for a set movement of the interactive object which match at least one target text comprised in the reply text.

12. The method according to any one of claims 8 to 11, wherein the second message comprises a response animation of the interactive object in response to the indication content.

13. The method according to claim 7, wherein

the user input operation comprises a body posture correspondingly made by the user following a body movement image displayed on the display interface; and

in response to the user input operation from the client, the method further comprises:

obtaining a user behavior image comprising the body posture;

identifying body posture information in the user behavior image; and

driving the interactive object displayed on the display interface to make a response based on the body posture information.

14. The method according to claim 13, wherein driving the interactive object displayed on the display interface to make a response based on the body posture information comprises:

determining a matching degree between the body posture information and a body posture in the body movement image; and

driving the interactive object displayed on the display interface to make a response based on the matching degree.

15. The method according to claim 14, wherein driving the interactive object displayed on the display interface to make a response based on the matching degree comprises:

in response to determining that the matching degree meets a set condition, instructing the interactive object displayed on the display interface to make a first response, wherein the first response comprises a body movement and/or a voice prompt which indicates that a displayed posture is qualified; and displaying a next body movement image; and

in response to determining that the matching degree does not meet the set condition, instructing the interactive object displayed on the display interface to make a second response, wherein the second response comprises a body movement and/or a voice prompt which indicates that the displayed posture is unqualified; and keeping displaying a current body movement image.

16. An interaction apparatus, comprising:

a receiving unit, configured to receive a first message from a client;

an obtaining unit, configured to, based on indication content comprised in the first message, obtain driving data matching the indication content; and

a driving unit, configured to control a display interface of client to play a response animation of an interactive object by using the driving data.

17. An interaction apparatus, comprising:

a sending unit, configured to send a first message comprising indication content to a server in response to a user input operation from a client; and

a playing unit, configured to play a response animation of an interactive object on a display interface of the client, based on a second message from the server in response to the first message.

18. An electronic device, comprising a processor and a memory for storing computer instructions executable by the processor, wherein the processor is configured to execute the computer instructions to implement the method according to any one of claims 1 to 6, or the method according to any one of claims 7 to 15.

19. A computer-readable storage medium storing a computer program, when being executed by a processor, the computer program causes the processor to implement the method according to any one of claims 1 to 6, or the method according to any one of claims 7 to 15.