CLAIMS

1. A voice interaction method, characterized in that the voice interaction method is applied to a voice interaction system, wherein the voice interaction system comprises a client and a first server, the client comprises a voice receiving unit, a voice playback unit, and a terminal processing unit, the first server comprises or connects to a language model, the language model comprises one or more language sub-model, the language sub-model correspond to one or more languages, and the method comprises:

the voice receiving unit collecting voice information and sending the voice information to the terminal processing unit;

the terminal processing unit obtaining text information corresponding to the voice information based on the voice information, and sending the text information to the first server;

the first server inputting the text information to the target language sub-model in the language model, which the target language sub-model corresponds to a language category of the text information; obtaining text response information corresponding to the text information returned by the target language sub-model, which the text response information has the same language category as the text information; sending the text response information to the terminal processing unit;

the terminal processing unit obtaining voice response information corresponding to the text response information based on the text response information, and controlling the voice playback unit to play the voice response information in the language category of the text information,

characterized in that the client further comprises a camera unit;

the camera unit is configured for capturing image information of voice objects;

the terminal processing unit obtaining text information corresponding to the voice information based on the voice information collected by the voice receiving unit, comprising:

the terminal processing unit intercepting the voice information collected by the voice receiving unit based on the image information of the voice object to obtain key voice information;

the terminal processing unit obtaining text information corresponding to the key voice information based on the key voice information.

2. The method according to claim 1, characterized in that the text information comprises a language identifier, which identifies the language category of the voice information;

the first server inputting the text information to the target language sub-model in the language model; obtaining the text response information corresponding to the text information returned by the target language sub-model, comprising:

the first server inputting the text information to the target language sub-model corresponding to the language identifier in the language model, and obtaining the text response information corresponding to the language identifier.

3. The method according to claim 2, characterized in that the first server comprises a language response module, wherein the language response module comprises or connects to the language model, the first server inputting the text information to the target language sub-model corresponding to the language identifier in the language model, and obtaining the text response information corresponding to the language identifier, comprising:

the language response module inputting the text information to the target language sub-model corresponding to the language identifier in the language model, and obtaining the text response information corresponding to the language identifier.

4. The method according to claim 3, characterized in that the language response module inputting the text information to the target language sub-model corresponding to the language identifier in the language model, and obtaining the text response information corresponding to the language identifier, comprising:

the language response module using the target language sub-model corresponding to the language identifier in the language model to semantically understand the text information and obtaining an understanding result;

if the understanding result indicates that the text information is Q&A type information, the language response module generating text response information corresponding to the text information in the language category characterized by the language identifier by using the target language sub-model.

5. The method according to claim 4, characterized in that the first server further comprises a text processing module, and the method further comprises:

if the understanding result indicates that the text information is instruction type information, the language response module sending a response instruction corresponding to the text information to the terminal processing unit through the text processing module.

6. The method according to claim 2, characterized in that the client further comprises a display unit, and the method further comprises:

the terminal processing unit controlling the display unit to display the text corresponding to the text response information in the language category identified by the language identifier, and controlling an interaction object displayed on the display unit to perform an interaction action corresponding to the text response information.

7. The method according to claim 6, characterized in that the text response information comprises language identifier, and the method further comprises:

the terminal processing unit controlling the display unit to display elements corresponding to the language identifier based on the text response information.

8. The method according to claim 5, characterized in that before the language response module inputting the text information to the target language sub-model corresponding to the language identifier in the language model, the method further comprises:

the text processing module receiving the text information sent by the terminal processing unit and classifying the text information to obtain classification result;

if the classification result indicates that the text information is instruction type information, the text processing module sending the response instruction corresponding to the text information to the terminal processing unit;

if the classification result indicates that the text information is Q&A type information, the text processing module sending the text information to the language response module.

9. The method according to claim 1, characterized in that the terminal processing unit obtains text information corresponding to the voice information based on the voice information, comprising:

the terminal processing unit decoding the voice information to obtain the text information corresponding to the voice information;

alternatively, the terminal processing unit uploading the voice information to a second server for decoding and obtaining the text information returned by the second server.

10. A voice interaction method, characterized in that the voice interaction method is applied to a client, which the client and the first server can be communicatively connected; the client comprises a voice receiving unit, a voice playback unit, and a terminal processing unit; the first server comprises or connects a language model, the language model comprises one or more language sub-model, and the language sub-model correspond to one or more languages; the method comprises:

the voice receiving unit collecting voice information and sending the voice information to the terminal processing unit;

the terminal processing unit obtaining the text information corresponding to the voice information based on the voice information, sending text information to the first server, so that the first server inputting the text information to the target language sub-model in the language model, and obtaining text response information corresponding to the text information returned by the target language sub-model; the target language sub-model corresponds to a language category of the text information, which the text response information has the same language category as the text information;

the terminal processing unit obtaining voice response information corresponding to the text response information based on the text response information, and controlling the voice playback unit to play the voice response information in the language category of the text information,

characterized in that the client further comprises a camera unit;

the camera unit is configured for capturing image information of voice objects;

the terminal processing unit obtaining text information corresponding to the voice information based on the voice information collected by the voice receiving unit, comprising:

the terminal processing unit intercepting the voice information collected by the voice receiving unit based on the image information of the voice object to obtain key voice information;

the terminal processing unit obtaining text information corresponding to the key voice information based on the key voice information.

11. The method according to claim 10, characterized in that the client further comprises a display unit, and the method further comprises:

the terminal processing unit controlling the display unit to display the text corresponding to the text response information in the language category identified by the language identifier, and controlling an interaction object displayed on the display unit to perform an interaction action corresponding to the text response information.

12. The method according to claim 11, characterized in that the text response information comprises language identifier, and the method further comprises:

the terminal processing unit controlling the display unit to display elements corresponding to the language identifier based on the text response information.

13. The method according to claim 10, characterized in that the terminal processing unit obtaining text information corresponding to the voice information based on the voice information, comprising:

the terminal processing unit decoding the voice information to obtain the text information corresponding to the voice information;

alternatively, the terminal processing unit uploading the voice information to a second server for decoding and obtaining the text information returned by the second server.

14. A voice interaction method characterized in that the voice interaction method is applied to a first server, which the client and the first server can be communicatively connected; the client comprises a voice receiving unit, a voice playback unit, and a terminal processing unit; the first server comprises or connects a language model, the language model comprises one or more language sub-model, and the language sub-model correspond to one or more languages; the method comprises:

the first server receiving text information sent by the terminal processing unit; the terminal processing unit is used for obtaining text information corresponding to voice information collected by the voice receiving unit;

the first server inputting the text information to a target language sub-model in the language model, which the target language sub-model corresponds to a language category of the text information; obtaining the text response information corresponding to the text information returned by the target language sub-model, which the text response information has the same language category as the text information;

the first server sending the text response information to the terminal processing unit, so that the terminal processing unit obtaining voice response information corresponding to the text response information based on the text response information, and controlling the voice playback unit to play the voice response information.

15. The method according to claim 14, characterized in that the text information comprises a language identifier, which identifies the language category of the voice information;

the first server inputting the text information to the target language sub-model in the language model; obtaining the text response information corresponding to the text information returned by the target language sub-model, comprising:

the first server inputting the text information to the target language sub-model corresponding to the language identifier in the language model, and obtaining the text response information corresponding to the language identifier.

16. The method according to claim 15, characterized in that the first server comprises a language response module, wherein the language response module comprises or connecting to the language model, the first server inputting the text information to the target language sub-model corresponding to the language identifier in the language model, and obtaining the text response information corresponding to the language identifier, comprising:

the language response module inputting the text information to the target language sub-model corresponding to the language identifier in the language model, and obtaining the text response information corresponding to the language identifier.

17. The method according to claim 16, characterized in that the language response module inputting the text information to the target language sub-model corresponding to the language identifier in the language model, comprising:
the language response module using the target language sub-model corresponding to the language identifier in the language model to semantically understand the text information and obtaining an understanding result;
if the understanding result indicates that the text information is Q&A type information, the language response module generating text response information corresponding to the text information in the language category characterized by the language identifier by using the target language sub-model.

18. The method according to claim 17, characterized in that the first server further comprises a text processing module, and the method further comprises:
if the understanding result indicates that the text information is instruction type information, the language response module sending a response instruction corresponding to the text information to the terminal processing unit through the text processing module.

19. The method according to claim 18, characterized in that before the language response module inputting the text information to the target language sub-model corresponding to the language identifier in the language model, the method further comprises:
the text processing module receiving the text information sent by the terminal processing unit and classifying the text information to obtain classification result;
if the classification result indicates that the text information is instruction type information, the text processing module sending the response instruction corresponding to the text information to the terminal processing unit;
if the classification result indicates that the text information is Q&A type information, the text processing module sending the text information to the language response module.

20. A client, characterized in that the client and the first server can be communicatively connected, wherein the client comprises a voice receiving unit, a voice playback unit, and a terminal processing unit; the first server comprises or connects a language

model, the language model comprises one or more language sub-model, and the language sub-model correspond to one or more languages; wherein:

the voice receiving unit is configured to collect voice information and send the voice information to the terminal processing unit;

the terminal processing unit is configured to obtain text information corresponding to the voice information based on the voice information, send text information to the first server, so that the first server inputs the text information to the target language sub-model in the language model, and obtain text response information corresponding to the text information returned by the target language sub-model; the target language sub-model corresponds to the language category of the text information, which the text response information has the same language category as the text information;

the terminal processing unit is further configured to obtain voice response information corresponding to the text response information based on the text response information, and control the voice playback unit to play the voice response information in the language category of the text information,

characterized in that the client further comprises a camera unit;

the camera unit is configured for capturing image information of voice objects;

the terminal processing unit obtaining text information corresponding to the voice information based on the voice information collected by the voice receiving unit, comprising:

the terminal processing unit intercepting the voice information collected by the voice receiving unit based on the image information of the voice object to obtain key voice information;

the terminal processing unit obtaining text information corresponding to the key voice information based on the key voice information.

21. A first server, characterized in that the client and the first server can be communicatively connected, wherein the client comprises a voice receiving unit, a voice playback unit, and a terminal processing unit; the first server comprises or connects a language model, the language model comprises one or more language sub-model, and the language sub-model correspond to one or more languages; wherein:

the first server is configured to receive text information sent by the terminal processing unit; the terminal processing unit is used to obtain the text information corresponding to voice information collected by the voice receiving unit;

the first server is configured to input the text information to a target language sub-model in the language model, which the target language sub-model corresponds to the language category of the text information; obtain text response information corresponding to the text information returned by the target language sub-model, which the text response information has the same language category as the text information;

the first server is configured to send the text response information to the terminal processing unit, so that the terminal processing unit obtains voice response information corresponding to the text response information based on the text response information, and control the voice playback unit to play the voice response information, characterized in that the client further comprises a camera unit;

the camera unit is configured for capturing image information of voice objects;

the terminal processing unit obtaining text information corresponding to the voice information based on the voice information collected by the voice receiving unit, comprising:

the terminal processing unit intercepting the voice information collected by the voice receiving unit based on the image information of the voice object to obtain key voice information;

the terminal processing unit obtaining text information corresponding to the key voice information based on the key voice information.