CLAIMS

1. A voice processing method, characterized in that applied to a voice processing system comprising a client and a first server, wherein the client and the first server can be communicatively connected, and the client comprises a voice receiving unit, a voice processing unit, a voice playback unit, and a display unit; the method comprises:

the voice receiving unit collecting voice information in at least one language and sending non-decoded voice information to the voice processing unit;

the voice processing unit receiving the voice information, obtaining text information corresponding to the voice information, and sending the text information to the first server;

the first server determining a category corresponding to the text information, executing processing logic corresponding to the category by filtering the text information even when the text information is vague and difficult to recognize, obtaining a first response information of the text information corresponding to the category based on content of the text information, and sending the first response information to the voice processing unit;

the voice processing unit receiving the first response information , and obtaining a corresponding second response information based on the first response information, wherein the second response information includes any combination of the following: the first response information, a voice response information corresponding to the first response information; controlling the voice playback unit to play the voice response information and/or displaying the first response information on the display unit.

2. The method according to claim 1, characterized in that the first response information comprises any combination of the following: text response information, image response information, action response instruction; the first server determining the category corresponding to the text information, executing the processing logic corresponding to the category, obtaining the first response information of the text information corresponding to the category, comprises:

in the case where the category is instruction, determining the action response instruction that matches the text information, or any combination of the action

response instruction and the following content: text response information, image response information;

in the case where the category is Q&A, inputting the text information into the language model to obtain the text response information corresponding to the text information.

3. A voice processing method characterized in that applied to a client, wherein the client comprises a voice receiving unit, a voice processing unit, a voice playback unit, and a display unit, wherein the client and a first server can be communicatively connected; the method comprises:

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

the voice processing unit receiving the voice information in at least one language, obtaining the text information corresponding to non-decoded voice information, and sending the text information to the first server; wherein, the first server is used to determine a category corresponding to the text information, executing the processing logic corresponding to the category by filtering the text information even when the text information is vague and difficult to recognize, obtaining a first response information of the text information corresponding to the category based on content of the text information, and sending the first response information to the voice processing unit;

the voice processing unit receiving the first response information , and obtaining a corresponding second response information based on the first response information, wherein the second response information includes any combination of the following: the first response information, a voice response information corresponding to the first

response information; controlling the voice playback unit to play the voice response information and/or displaying the first response information on the display unit.

4. A voice processing method characterized in that applied to a first server, wherein the first server and the client can be communicatively connected, the client comprises a voice receiving unit, a voice processing unit, a voice playback unit, and a display unit; the method comprises:

the first server receiving text information sent by the voice processing unit, and the text information is generated based on non-decoded voice information sent by the voice receiving unit;

the first server determining a category corresponding to the text information, executes processing logic corresponding to the category by filtering the text information even when the text information is vague and difficult to recognize, obtaining a first response information of the text information corresponding to the category based on content of the text information, and sending the first response information to the voice processing unit to generate a corresponding second response information based on the first response information, wherein the second response information includes any combination of the following: the first response information, a voice response information corresponding to the first response information; controlling the voice playback unit to play the voice response information and/or displaying the first response information on the display unit.

5. The method according to claim 4, characterized in that the first response information comprises any combination of the following: text response information, image response information, action response instruction; the first server determining the category corresponding to the text information, executing the processing logic corresponding to the category, obtains the first response information of the text information corresponding to the category, comprises:

in the case where the category is instruction, determining the action response instruction that matches the text information, or any combination of the action response instruction and the following content: text response information, image response information;

in the case where the category is Q&A, inputting the text information into the language model to obtain the text response information corresponding to the text information.

6. A voice processing system, characterized in that the voice processing system comprises a client and a first server, wherein the client and the first server can be communicatively connected, the client comprises a voice receiving unit, a voice processing unit, a voice playback unit, and a display unit, wherein:

the voice receiving unit is configured to collect voice information in at least one language and send the voice information to the voice processing unit;

the voice processing unit is configured to receive non-decoded voice information, obtain the text information corresponding to the voice information, and send the text information to the first server;

the first server is configured to determine a category corresponding to the text information, execute the processing logic corresponding to the category by filtering the text information even when the text information is vague and difficult to recognize, obtain a first response information of the text information corresponding to the category based on content of the text information, and send the first response information to the voice processing unit, the first response information at least includes the text response information corresponding to the text information;

the voice processing unit is further configured to receive the first response information, and obtain a corresponding second response information based on the first response information, wherein the second response information includes any combination of the following: the first response information, a voice response information corresponding to the first response information; control the voice playback unit to play the voice response information and/or display the first response information on the display unit.

7. A first server, characterized in that the first server and the client can be communicatively connected, the client comprises a voice receiving unit, a voice processing unit, a voice playback unit, and a display unit, wherein the first server is configured as:

receives text information sent by the voice processing unit, and the text information is generated based on the voice information sent by the voice receiving unit;

determines a category corresponding to the text information, executes processing logic corresponding to the category by filtering the text information even when the text information is vague and difficult to recognize, obtains a first response information of the text information corresponding to the category based on content of the text information, and sends the first response information to the voice processing unit to generate a corresponding second response information based on the first response information, wherein the second response information includes any combination of the following: the first response information, a voice response information corresponding to the first response information; control the voice playback unit to play the voice response information and/or display the first response information on the display unit.