CLAIMS

1. A speech communication system with human-machine coordination, comprising a communications module, and further comprising:

an answering and calling-out module, allocating a human agent after a call is connected;

a speech intention understanding module, comprising a speech intention understanding model, for directly understanding an intention in a call process from speech;

a voice cloning module, comprising a voice cloning model, for synthesizing voice of a corresponding attendant from reply text;

a human-machine interaction module, invoking the speech intention understanding module through a speech robot to directly obtain an intention of a client from speech and further invoking, according to a pre-edited speechcraft, the voice cloning module to communicate with the client by the voice of the corresponding attendant;

an intervention prompting module, comprising a quantization analysis submodule, monitoring in real-time call content during a human-machine interaction to obtain an intervention willing probability and prompting the human agent according to a predefined rule; and

a manual intervention module, replying the client by speech of the corresponding attendant and/or by processing reply content of a random attendant into voice of the corresponding attendant through invoking the voice cloning module, and further comprising a retrieving and prompting submodule that can invoke the speech intention understanding module, to display a call intention and give the attendant a prompt for conversation.

2. The speech communication system with human-machine coordination according to claim 1, wherein the manual intervention module further comprises a quality inspection module for inspecting a speaking speed and content of the random attendant, and feeding back an inspection result to the human agent in real-time.

3. A speech communication method with human-machine coordination, comprising steps of:

allocating, by an answering and calling-out module, a human agent after a call is connected;

invoking a speech intention understanding module, through a speech robot of a human-machine interaction module, to directly obtain an intention of a client from speech and further invoking, according to a pre-edited speechcraft, a voice cloning module to communicate with the client by a voice of the corresponding attendant;

monitoring in real-time, by a quantization analysis submodule of an intervention prompting module, call content during a human-machine interaction to obtain an intervention willing probability and prompting the human agent according to a predefined rule; and

when the human agent receives an intervention prompt, through a manual intervention module, intervening in the call, displaying by a retrieving and prompting submodule an intention during a call process through human-machine interaction, replying the client by speech of the corresponding attendant and/or by processing reply content of a random attendant into voice of the corresponding attendant through invoking the voice cloning module, and giving the attendant a prompt for conversation during the call process by the retrieving and prompting submodule.

4. The speech communication method with human-machine coordination according to claim 3, wherein the quantization analysis submodule is configured to obtain the intervention willing probability for a current client needing manual intervention based on comparative analyses of big data of historical calls and a current human-machine interaction process.

5. The speech communication method with human-machine coordination according to claim 3, wherein when the random attendant selects to perform manual intervention, the retrieving and prompting submodule directly invokes the speech intention understanding module to display the intention of the call through human-machine interaction; and the retrieving and prompting submodule invokes the speech intention understanding module during a process in which the call is manually intervened, to obtain a newest call intention, and retrieves reference information from a database to generate and display a reply text.

6. The speech communication method with human-machine coordination according to claim 3, wherein manners by which the human agent performs an intervention include:

directly replying the client by speech by the corresponding attendant;

selecting, by a random attendant, prompting text generated by the retrieving and prompting submodule, which is generated into speech by the voice cloning module to reply the client;

manually inputting a text by a random attendant, which is generated into speech by the voice cloning module to reply the client;

speech of a random attendant, which is converted into text by a speech recognition submodule in the voice cloning module, and is then generated into speech by the voice cloning module to reply the client.

7. The speech communication method with human-machine coordination according to claim 3, wherein generation and applications of a speech intention understanding model in the speech intention understanding module include steps of:

S1. training the speech intention understanding model based on deep learning by using speech data with intention annotations;

S2. deploying the trained speech intention understanding model to a speech intention understanding server; and

S3. receiving the speech data during a conversation and inputting the speech data into the speech intention understanding server to invoke a speech intention understanding service, and returning a recognized intention.

8. The speech communication method with human-machine coordination according to claim 3, wherein generation and applications of a voice cloning model include steps of:

A1. collecting voice corpus of an attendant corresponding to the human agent, to train an initial voice cloning model;

A2. deploying the trained voice cloning model to a voice cloning server; and

A3. synthesizing, from a manually intervened reply text, voice of the corresponding attendant to play to the client