CLAIMS:

1. A three-way calling terminal for a mobile human-machine coordination calling robot, comprising:

a first speech interface, configured to transfer call audio between a call object and a back-end processing module, the call object representing a user for establishing call connection to the three-way calling terminal by using a user terminal;

a CODEC1 module, configured to encode and/or decode the call audio between the call object and the back-end processing module;

a second speech interface, configured to transfer call audio between the human agent and the call object, wherein during a call, the call audio between the call object and the back-end processing module is received through the second speech interface by the human agent to determine whether to perform human intervention; and when the human intervention is required, a speech of the human agent is transmitted to the three-way calling terminal through the second speech interface;

a CODEC2 module, configured to encode and/or decode the call audio between the human agent and the call object;

a call control module, configured to process a control signal, and automatically make, answer, and hang up a call;

a data processing submodule, configured to process speech data and perform data transfer between the data processing submodule and the back-end processing module; and

a networking submodule, configured to perform network connection to the back-end processing module.

2. The three-way calling terminal for a mobile human-machine coordination calling robot according to claim 1, wherein the three-way calling terminal further comprises a display module configured to display a call record or call-related information between the three-way calling terminal and the call object.

3. The three-way calling terminal for a mobile human-machine coordination calling robot according to claim 1 or 2, wherein the three-way calling terminal further comprises a keypad

submodule, through which a control instruction is inputted.

4. A calling system comprising the three-way calling terminal according to any one of claims 1 to 3, a back-end processing module, and a human agent, and at least one communications terminal, wherein the human agent is connected to the three-way calling terminal through the communications terminal.

5. The calling system according to claim 4, wherein the back-end processing module is configured to process speech data transmitted by the three-way calling terminal, and generate response speech and text to be transmitted back to the three-way calling terminal.

6. The calling system according to claim 4 or 5, wherein the back-end processing module comprises a dialogue management submodule, a speech recognition submodule, an intention recognition submodule, and a speech synthesis submodule;

the dialogue management submodule is configured to control procedure and logic of a dialogue, and generate response text;

the speech recognition submodule is configured to recognize and convert received speech of the call object into text;

the intention recognition submodule is configured to recognize an intention of the call object based on the text of the recognized speech; and

the speech synthesis submodule is configured to synthesize and transmit speech from the response text to the three-way calling terminal.