What is claimed is:

1. A real-time speech translation method, comprising the following steps of:

performing speech recognition of a first speech information input by a first user to obtain first text data;

translating the first speech information to obtain a second synthesized speech corresponding to a second language category;

obtaining a first synthesized speech corresponding to a first language category on the basis of the first text data;

respectively collecting environmental voice data from the first user and a second user;

performing voiceprint detection on each of the environmental voice data;

sending a first preset alert message to the first user when detecting other people's voiceprints in the environmental voice data from the first user, the first preset alert message being used to remind the first user that there are other people nearby;

pausing the translation of the first speech information when detecting other people's voiceprints in the environmental voice data from the second user, and sending a second preset alert message to the second user;

resuming the translation of the first speech information when the other people's voiceprints are no longer detected in the environmental voice data from the second user; and

outputting the first synthesized speech to the first user and simultaneously outputting the second synthesized speech to the second user.

2. The real-time speech translation method of claim 1, wherein the step of translating the first speech information to obtain a second synthesized speech corresponding to a second language category further comprises the following steps of:

translating the first text data obtained from the speech recognition of the first speech information to obtain a second text data corresponding to the second language category; and

obtaining the second synthesized speech corresponding to the second language category on the basis of the second text data.

3. The real-time speech translation method of claim 2, wherein the step of obtaining a first synthesized speech according to the first text data, corresponding to a first language category further comprises the following steps of:

calculating an estimated duration for completing the playback of the second synthesized speech to obtain a first estimated duration;

obtaining a second timestamp information related to the second synthesized speech on the basis of the first estimated duration;

performing speech synthesis on the basis of the first text data to obtain the first synthesized speech; and

synchronizing a first timestamp corresponding of the first synthesized speech with the second timestamp information.

4. The real-time speech translation method of claim 1, further comprising the following steps of:

continuing collecting a next segment of the first speech information input by the first user when receiving a positive feedback from the first user regarding the first synthesized speech; and

collecting the first speech information re-input by the first user and outputting a notification message of the re-input of the first speech information to the second user when receiving a negative feedback from the first user regarding the first synthesized speech.

5. The real-time speech translation method of claim 2, further comprising

the following steps of:

respectively extracting interest tags and the corresponding timestamp information from the first text data and the second text data; and

respectively marking the interest tags on the first text data and the second text data according to the first timestamp information and the second timestamp information.

6. The real-time speech translation method of claim 5, before the step of outputting the first synthesized speech to the first user and simultaneously outputting the second synthesized speech to the second user, further comprising the following steps of:

respectively collecting voiceprint information from the first user and the second user;

respectively identifying the first user and the second user to obtain identity information on the basis of each of the voiceprint information;

determining a relationship type between the first user and the second user on the basis of each of the identity information; and

filtering the interest tags on the basis of the relationship type between the first user and the second user and a preset identity relationship interest database, and designating the remaining interest tags as target interest tags; and

wherein the step of outputting the first synthesized speech to the first user and simultaneously outputting the second synthesized speech to the second user further comprises the following step of:

providing a voice notification for the voice associated with the target interest tags in the first synthesized speech and second synthesized speech.

7. The real-time speech translation method of claim 1, wherein the step of performing speech recognition on a first speech information from a first user

to obtain first text data further comprises the following steps of:

obtaining preceding context data of the first speech information;

performing speech recognition of the first speech information to obtain initial text data; and

correcting the initial text data on the basis of the preceding context data to obtain the first text data.

8. The real-time speech translation method of claim 1, before the step of outputting the first synthesized speech to the first user and simultaneously outputting the second synthesized speech to the second user, further comprising the following step of:

when detecting a pair of the earphones for speech output worn by different users, controlling the earphones to operate in a first state; and

when detecting two pairs of earphones worn by the first user and the second user respectively, controlling the pairs of earphones to operate in a second state;

wherein, in the first state, each of the earphones is configured as a speech output channel of the user wearing it and outputs different speech from each other; in the second state, both earphones in one pair of earphones are configured as a speech output channel of the user wearing the and output the same speech.

9. The real-time speech translation method of claim 8, wherein the pair of earphones comprises an UWB communication module; before the step of outputting the first synthesized speech to the first user and simultaneously outputting the second synthesized speech to the second user, the real-time speech translation method further comprises the following steps of:

the UWB communication module detecting the distance between the two earphones in the pair of earphones in real-time;

when the distance between the earphones is larger than a first preset threshold, controlling the earphones to operate in the first state; and

when the distance between the earphones is less than the first preset threshold, controlling the earphones to operate in the second state.

10. The real-time speech translation method of claim 1, further comprising the following steps of:

forming an audio adjustment database on the basis of un-played audio frames of the first synthesized speech and audio frames of the collected first speech information that have not been converted into the first synthesized speech; and

adjusting a playing speed of the first synthesized speech on the basis of the number of audio frames in the audio adjustment database;

wherein, the playing speed of the first synthesized speech increases with an increase in the number of audio frames in the audio adjustment database and decreases with a decrease in the number of audio frames in the audio adjustment database.

11. A real-time speech translation system for implementing the method of claim 1, comprising:

a first text data generation module, configured to performing speech recognition of a first speech information input by a first user to obtain first text data;

a second synthesized speech generation module, configured to translate the first speech information to obtain a second synthesized speech corresponding to a second language category;

a first synthesized speech generation module, configured to obtain a first synthesized speech corresponding to a first language category on the basis of the first text data; and

a synthesized speech playback module, configured to the first synthesized speech to the first user and simultaneously outputting the second synthesized speech to a second user.

12. A real-time speech translation device, comprising:

a processor; and

a memory, configured to store programs for the processor to execute;

wherein, the processor is configured to execute the programs to implemented the steps of the real-time speech translation method according to any claim in claim 1 to claim 10.

13. A storage medium for storing programs, wherein the programs are implemented the steps of the real-time speech translation method according to any claim in claim 1 to claim 10 by a processor.