CLAIMS

What is claimed is:

1. A voiceprint recognition method, comprising:

collecting, by a client, a test voice of a user, and sending a voice recognition request to a background server, the voice recognition request comprises a user ID and the test voice;

receiving, by the background server, the voice recognition request, and determining the voice recognition request to be processed with a message queue and an asynchronous mechanism;

acquiring, by the background server, a target voiceprint feature which corresponds to the user ID of the voice recognition request to be processed, and acquiring a test voiceprint feature which corresponds to the test voice of the voice recognition request to be processed;

judging, by the background server according to the target voiceprint feature and the test voiceprint feature, whether the target voiceprint feature and the test voiceprint feature correspond to the same user, and outputting the result of the judging to the client; and

receiving and displaying, by the client, the result of the judging;

wherein it is judged according to the target voiceprint feature and the test voiceprint feature whether the target voiceprint feature and the test voiceprint feature correspond to the same user, comprising:

applying dimensionality reduction to the target voiceprint feature and the test voiceprint feature respectively with a PLDA algorithm, to acquire the target dimensionality reduction value and the test dimensionality reduction value;

performing cosine measurement on the target dimensionality reduction value and the test dimensionality reduction value using a cosine measurement function to acquire a cosine measurement value; and

judging whether the cosine measurement value is greater than a similarity threshold; wherein if the cosine measurement value is greater than a similarity threshold, the target voiceprint feature and the test voiceprint feature correspond to the same user; and if the cosine measurement value is not greater than a similarity threshold, the target voiceprint feature and the test voiceprint feature do not correspond to the same user.

2. The voiceprint recognition method according to claim 1, wherein the background server acquires a target voiceprint feature which corresponds to the user ID of the voice recognition request to be processed and acquires a test voiceprint feature which corresponds to the test voice of the voice recognition request to be processed, comprising:

querying a voiceprint feature library according to the user ID of the voice recognition request to be processed to acquire the target voiceprint feature which corresponds to the user ID of the voice recognition request to be processed; and

processing the test voiceprint feature of the voice recognition request to be processed using a Gaussian mixture model-universal background model to acquire the test voiceprint feature which corresponds to the test voice of the voice recognition request to be processed.

3. The voiceprint recognition method according to claim 2, further comprising:

performing MFCC feature extraction on a training voice to acquire an MFCC acoustic feature;

performing voice activity detection on the MFCC acoustic feature to estimate a Gaussian mixture model parameter;

training the universal background model with the Gaussian mixture model parameter to acquire the Gaussian mixture model-universal background model;

receiving a voiceprint registration request, wherein the voiceprint registration request comprises a user ID and a target training voice;

training the target training voice with the Gaussian mixture model-universal background model to acquire the target voiceprint feature; and

storing the user ID and the target voiceprint feature in the voiceprint feature library.

4. The voiceprint recognition method according to claim 1, wherein the PLDA algorithm comprises:

initializing μ and W;

calculating w using $w \leftarrow \frac{W_i^{-\frac{1}{2}}(w-\mu_i)}{\left\|W_i^{-\frac{1}{2}}(w-\mu_i)\right\|}$; and

calculating W again using $W = \frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s}(w_i^s - y_s)(w_i^s - y_s)^t$, and returning to the step of calculating w using $w \leftarrow \frac{W_i^{-\frac{1}{2}}(w-\mu_i)}{\left\|W_i^{-\frac{1}{2}}(w-\mu_i)\right\|}$, until w is less than a specified threshold;

wherein μ is a mean voiceprint vector; W is the distance between classes; w is the voiceprint feature; and i is the number of iterations;

the cosine measurement function comprises:

$$score(w_{train}, w_{test}) = \frac{(w_{train})^t w_{test}}{\|w_{train}\|\|w_{test}\|}$$

wherein w_{train} is the target voiceprint feature, w_{test} is the test voiceprint feature, and t is the time.

5. A voiceprint recognition device, comprising:

a client configured to collect a test voice of a user and send a voice recognition request to a background server, the voice recognition request comprises a user ID and the test voice;

the background server configured to receive the voice recognition request and determine the voice recognition request to be processed with a message queue and an asynchronous mechanism;

the background server configured to acquire a target voiceprint feature which corresponds to the user ID of the voice recognition request to be processed, and acquire a test voiceprint feature which corresponds to the test voice of the voice recognition request to be processed;

the background server configured to judge, according to the target voiceprint feature and the test voiceprint feature, whether the target voiceprint feature and the test voiceprint feature correspond to the same user, and output the result of the judging to the client; and

the client configured to receive and display the result of the judging;

wherein the background server comprises:

a feature dimensionality reducing unit configured to reduce the dimensionality of the target voiceprint feature and the test voiceprint feature respectively with a PLDA algorithm, to acquire the target dimensionality reduction value and the test dimensionality reduction value;

a cosine measurement processing unit configured to perform cosine measurement on the target dimensionality reduction value and the test dimensionality reduction value using a cosine measurement function to acquire a cosine measurement value; and

a user recognition judging unit configured to judge whether the cosine measurement value is greater than a similarity threshold; wherein if the cosine measurement value is greater than a similarity threshold, the target voiceprint feature and the test voiceprint feature correspond to the same user; and if the cosine measurement value is not greater than a similarity threshold, the target voiceprint feature and the test voiceprint feature do not correspond to the same user.

6. The voiceprint recognition device according to claim 5, wherein the background server comprises:

a feature querying unit configured to query a voiceprint feature library according to the user ID of the voice recognition request to be processed to acquire the target voiceprint feature which corresponds to the user ID of the voice recognition request to be processed; and

a feature processing unit configured to process the test voiceprint feature of the voice recognition request to be processed using a Gaussian mixture model-universal background model to acquire the test voiceprint feature which corresponds to the test voice of the voice recognition request to be processed.

7. The voiceprint recognition device according to claim 6, wherein the background server further comprises:

an acoustic feature extracting unit configured to perform MFCC feature extraction on a training voice to acquire an MFCC acoustic feature;

a voice activity detecting unit configured to perform voice activity detection on the MFCC acoustic feature to estimate a Gaussian mixture model parameter;

a model training unit configured to train the universal background model with the Gaussian mixture model parameter to acquire the Gaussian mixture model-universal background model;

a registration voice receiving unit configured to receive a voiceprint registration request, wherein the voiceprint registration request comprises a user ID and a target training voice;

a target voiceprint feature acquiring unit configured to train the target training voice with the Gaussian mixture model-universal background model to acquire the target voiceprint feature; and

a target voiceprint feature storing unit configured to store the user ID and the target voiceprint feature in the voiceprint feature library.

8. The voiceprint recognition device according to claim 5, wherein the PLDA algorithm comprises:

initializing μ and W;

calculating w using $w \leftarrow \frac{W_i^{-\frac{1}{2}}(w-\mu_i)}{\left\|W_i^{-\frac{1}{2}}(w-\mu_i)\right\|}$; and

calculating W again using $W = \frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s}(w_i^s - y_s)(w_i^s - y_s)^t$, and returning to the step of calculating w using $w \leftarrow \frac{W_i^{-\frac{1}{2}}(w-\mu_i)}{\left\|W_i^{-\frac{1}{2}}(w-\mu_i)\right\|}$, until w is less than a specified threshold;

wherein μ is a mean voiceprint vector; W is the distance between classes; w is the voiceprint feature; and i is the number of iterations;

the cosine measurement function comprises:

$$score(w_{train}, w_{test}) = \frac{(w_{train})^t w_{test}}{\|w_{train}\|\|w_{test}\|}$$

wherein w_{train} is the target voiceprint feature, w_{test} is the test voiceprint feature, and t is the time.

9. A background server comprising a memory, a processor, and a computer program stored in the memory and executable on the processor, wherein the processor, when executing the computer program, implements the steps of:

collecting, by a client, a test voice of a user, and sending a voice recognition request to the background server, the voice recognition request comprises a user ID and the test voice;

receiving, by the background server, the voice recognition request, and determining the voice recognition request to be processed with a message queue and an asynchronous mechanism;

acquiring, by the background server, a target voiceprint feature which corresponds to the user ID of the voice recognition request to be processed, and acquiring a test voiceprint feature which corresponds to the test voice of the voice recognition request to be processed;

judging, by the background server according to the target voiceprint feature and the test voiceprint feature, whether the target voiceprint feature and the test voiceprint feature correspond to the same user, and outputting the result of the judging to the client; and

receiving and displaying, by the client, the result of the judging;

wherein it is judged according to the target voiceprint feature and the test voiceprint feature whether the target voiceprint feature and the test voiceprint feature correspond to the same user, comprising:

applying dimensionality reduction to the target voiceprint feature and the test voiceprint feature respectively with a PLDA algorithm, to acquire the target dimensionality reduction value and the test dimensionality reduction value;

performing cosine measurement on the target dimensionality reduction value and the test dimensionality reduction value using a cosine measurement function to acquire a cosine measurement value; and

judging whether the cosine measurement value is greater than a similarity threshold; wherein if the cosine measurement value is greater than a similarity threshold, the target voiceprint feature and the test voiceprint feature correspond to the same user; and if the cosine measurement value is not greater than a similarity threshold, the target voiceprint feature and the test voiceprint feature do not correspond to the same user.

10. The background server according to claim 9, wherein the background server acquires a target voiceprint feature which corresponds to the user ID of the voice recognition request to be processed and acquires a test voiceprint feature which corresponds to the test voice of the voice recognition request to be processed, comprising:

querying a voiceprint feature library according to the user ID of the voice recognition request to be processed to acquire the target voiceprint feature which corresponds to the user ID of the voice recognition request to be processed; and

processing the test voiceprint feature of the voice recognition request to be processed using a Gaussian mixture model-universal background model to acquire the test voiceprint feature which corresponds to the test voice of the voice recognition request to be processed.

11. The background server according to claim 10, further comprising:

performing MFCC feature extraction on a training voice to acquire an MFCC acoustic feature;

performing voice activity detection on the MFCC acoustic feature to estimate a Gaussian mixture model parameter;

training the universal background model with the Gaussian mixture model parameter to acquire the Gaussian mixture model-universal background model;

receiving a voiceprint registration request, wherein the voiceprint registration request comprises a user ID and a target training voice;

training the target training voice with the Gaussian mixture model-universal background model to acquire the target voiceprint feature; and

storing the user ID and the target voiceprint feature in the voiceprint feature library.

12. The background server according to claim 9, wherein the PLDA algorithm comprises:

initializing μ and W;

calculating w using $w \leftarrow \frac{W_i^{-\frac{1}{2}}(w-\mu_i)}{\left\|W_i^{-\frac{1}{2}}(w-\mu_i)\right\|}$; and

calculating W again using $W = \frac{1}{n}\sum_{s=1}^{s}\sum_{i=1}^{n_s}(w_i^s - y_s)(w_i^s - y_s)^t$, and returning to the step of calculating w using $w \leftarrow \frac{W_i^{-\frac{1}{2}}(w-\mu_i)}{\left\|W_i^{-\frac{1}{2}}(w-\mu_i)\right\|}$, until w is less than a specified threshold;

wherein μ is a mean voiceprint vector; W is the distance between classes; w is the voiceprint feature; and i is the number of iterations;

the cosine measurement function comprises:

$$score(w_{train}, w_{test}) = \frac{(w_{train})^t w_{test}}{\|w_{train}\|\|w_{test}\|}$$

wherein w_{train} is the target voiceprint feature, w_{test} is the test voiceprint feature, and t is the time.

13. A computer-readable storage medium in which a computer program is stored, wherein the computer program, when being executed by a processor, implements the steps of:

collecting, by a client, a test voice of a user, and sending a voice recognition request to a background server, wherein the voice recognition request comprises a user ID and the test voice;

receiving, by the background server, the voice recognition request, and determining the voice recognition request to be processed with a message queue and an asynchronous mechanism;

acquiring, by the background server, a target voiceprint feature which corresponds to the user ID of the voice recognition request to be processed, and acquiring a test

voiceprint feature which corresponds to the test voice of the voice recognition request to be processed;

judging, by the background server according to the target voiceprint feature and the test voiceprint feature, whether the target voiceprint feature and the test voiceprint feature correspond to the same user, and outputting the result of the judging to the client; and

receiving and displaying, by the client, the result of the judging;

wherein it is judged according to the target voiceprint feature and the test voiceprint feature whether the target voiceprint feature and the test voiceprint feature correspond to the same user, comprising:

applying dimensionality reduction to the target voiceprint feature and the test voiceprint feature respectively with a PLDA algorithm, to acquire the target dimensionality reduction value and the test dimensionality reduction value;

performing cosine measurement on the target dimensionality reduction value and the test dimensionality reduction value using a cosine measurement function to acquire a cosine measurement value; and

judging whether the cosine measurement value is greater than a similarity threshold; wherein if the cosine measurement value is greater than a similarity threshold, the target voiceprint feature and the test voiceprint feature correspond to the same user; and if the cosine measurement value is not greater than a similarity threshold, the target voiceprint feature and the test voiceprint feature do not correspond to the same user.

14. The computer-readable storage medium according to claim 13, wherein the background server acquires a target voiceprint feature which corresponds to the user ID of the voice recognition request to be processed and acquires a test voiceprint feature which corresponds to the test voice of the voice recognition request to be processed, comprising:

querying a voiceprint feature library according to the user ID of the voice recognition request to be processed to acquire the target voiceprint feature which corresponds to the user ID of the voice recognition request to be processed; and

processing the test voiceprint feature of the voice recognition request to be processed using a Gaussian mixture model-universal background model to acquire the test voiceprint feature which corresponds to the test voice of the voice recognition request to be processed.

15. The computer readable storage medium of claim 14, further comprising:

performing MFCC feature extraction on a training voice to acquire an MFCC acoustic feature;

performing voice activity detection on the MFCC acoustic feature to estimate a Gaussian mixture model parameter;

training the universal background model with the Gaussian mixture model parameter to acquire the Gaussian mixture model-universal background model;

receiving a voiceprint registration request, wherein the voiceprint registration request comprises a user ID and a target training voice;

training the target training voice with the Gaussian mixture model-universal background model to acquire the target voiceprint feature; and

storing the user ID and the target voiceprint feature in the voiceprint feature library.

16. The computer-readable storage medium of claim 13, wherein the PLDA algorithm comprises:

initializing μ and W;

calculating w using $w \leftarrow \frac{W_i^{-\frac{1}{2}}(w-\mu_i)}{\left\|W_i^{-\frac{1}{2}}(w-\mu_i)\right\|}$; and

calculating W again using $W = \frac{1}{n}\sum_{s=1}^{s}\sum_{i=1}^{n_s}(w_i^s - y_s)(w_i^s - y_s)^t$, and returning to

the step of calculating w using $w \leftarrow \frac{W_i^{-\frac{1}{2}}(w-\mu_i)}{\left\| W_i^{-\frac{1}{2}}(w-\mu_i) \right\|}$, until w is less than a specified threshold;

wherein μ is a mean voiceprint vector; W is the distance between classes; w is the voiceprint feature; and i is the number of iterations;

the cosine measurement function comprises:

$$score(w_{train}, w_{test}) = \frac{(w_{train})^t w_{test}}{\|w_{train}\| \|w_{test}\|}$$

wherein w_{train} is the target voiceprint feature, w_{test} is the test voiceprint feature, and t is the time.