

CLAIMS

What is claimed is:

1. A method, comprising:

5 acquiring sound source angle data corresponding to a speech data frame of a to-be-separated role collected by a speech collection device, the speech collection device comprises a microphone array, the acquiring the sound source angle data corresponding to the speech data frame of the to-be-separated role collected by the speech collection device including:

10 acquiring a covariance matrix of the speech data frame received by at least some microphones in the microphone array;

performing eigenvalue decomposition on the covariance matrix to obtain multiple eigenvalues;

15 selecting a first quantity of largest eigenvalues from the multiple eigenvalues, and forming a speech signal sub-space based on eigenvectors corresponding to the selected eigenvalues; and

determining the sound source angle data based on the speech signal sub-space;

20 performing identification on the to-be-separated role based on the sound source angle data to obtain a first identification result of the to-be-separated role; and

separating a role based on the first identification result of the to-be-separated role.

2. The method according to claim 1, wherein after the acquiring the sound source angle data corresponding to the speech data frame of the to-be-separated role collected by the speech collection device, the method further comprises:

25 performing voice activity detection on the speech data frame of the to-be-separated role to obtain a speech data frame with a voice activity;

filtering and smoothing the speech data frame with the voice activity based on an energy spectrum of the speech data frame of the to-be-separated role to obtain a filtered and smoothed speech data frame; and

30 updating the sound source angle data based on the filtered and smoothed speech data frame to obtain updated sound source angle data.

3. The method according to claim 2, wherein the filtering and smoothing the speech data frame with the voice activity based on the energy spectrum of the speech data frame of the to-be-separated role to obtain the filtered and smoothed speech data frame comprises:

5 filtering and smoothing the speech data frame with the voice activity through a median filter based on spectral flatness of the energy spectrum of the speech data frame of the to-be-separated role to obtain the filtered and smoothed speech data frame.

10 4. The method according to claim 1, wherein the performing the identification on the to-be-separated role based on the sound source angle data to obtain the first identification result of the to-be-separated role comprises:

performing sequence clustering on the sound source angle data to obtain a sequence clustering result of the sound source angle data; and

15 determining that a role identifier corresponding to the sequence clustering result of the sound source angle data is the first identification result of the to-be-separated role.

5. The method according to claim 4, wherein the performing the sequence clustering on the sound source angle data to obtain the sequence clustering result of the sound source angle data comprises:

20 determining a distance between the sound source angle data and a sound source angle sequence clustering center; and

determining the sequence clustering result of the sound source angle data based on the distance between the sound source angle data and the sound source angle sequence clustering center.

25

6. The method according to claim 1, wherein after the obtaining the first identification result of the to-be-separated role, the method further comprises:

30 performing voiceprint identification on a speech data frame of the to-be-separated role within a preset time period to obtain a second identification result of the to-be-separated role; and

in response to determining that the first identification result is different from the second identification result, using the second identification result to correct the first identification result to obtain a final identification result of the to-be-separated role.

7. The method according to claim 1, wherein the first quantity is equivalent to an estimated quantity of sound sources.

8. The method according to claim 6, wherein after the obtaining the final
5 identification result of the to-be-separated role, the method further comprises:

acquiring face image data of the to-be-separated role collected by an image collection device;

performing face recognition on the face image data to obtain a third identification result of the to-be-separated role; and

10 in response to determining that the third identification result is different from the second identification result, using the third identification result to correct the second identification result to obtain the final identification result of the to-be-separated role.

9. An apparatus comprising:

15 one or more processors; and

one or more memories storing thereon computer-readable instructions that, executable by the one or more processors, cause the one or more processors to perform acts comprising:

20 acquiring sound source angle data corresponding to a speech data frame of a role collected by a speech collection device, the speech collection device comprises a microphone array, the acquiring the sound source angle data corresponding to the speech data frame of the to-be-separated role collected by the speech collection device including:

25 acquiring a covariance matrix of the speech data frame received by at least some microphones in the microphone array;

performing eigenvalue decomposition on the covariance matrix to obtain multiple eigenvalues;

30 selecting a first quantity of largest eigenvalues from the multiple eigenvalues, and forming a speech signal sub-space based on eigenvectors corresponding to the selected eigenvalues;

determining the sound source angle data based on the speech signal sub-space;

performing identification on the role based on the sound source angle data

to obtain an identification result of the role; and
displaying identity data of the role on an interactive interface of the speech collection device based on the identification result of the role.

5 10. The apparatus according to claim 9, wherein after the acquiring the sound source angle data corresponding to the speech data frame of the role collected by the speech collection device, the acts further comprise:

switching on a lamp of the speech collection device in a sound source direction indicated by the sound source angle data.

10

11. The apparatus according to claim 9, wherein the acts further comprise:
displaying a speaking action image or a speech waveform image of the role on the interactive interface of the speech collection device.

15 12. One or more memories storing thereon computer-readable instructions that, executable by one or more processors, cause the one or more processors to perform acts comprising:

acquiring sound source angle data corresponding to a speech data frame of a to-be-separated role collected by a speech collection device, the speech collection device
20 comprises a microphone array, the acquiring the sound source angle data corresponding to the speech data frame of the to-be-separated role collected by the speech collection device including:

acquiring a covariance matrix of the speech data frame received by at least some microphones in the microphone array;

25 performing eigenvalue decomposition on the covariance matrix to obtain multiple eigenvalues;

selecting a first quantity of largest eigenvalues from the multiple eigenvalues, and forming a speech signal sub-space based on eigenvectors corresponding to the selected eigenvalues; determining the sound source angle data based on the speech signal sub-space;
30

performing identification on the to-be-separated role based on the sound source angle data to obtain a first identification result of the to-be-separated role; and

separating a role based on the first identification result of the to-be-separated role.

13. The one or more memories according to claim 12, wherein after the acquiring the sound source angle data corresponding to the speech data frame of the to-be-separated role collected by the speech collection device, the acts further comprise:

performing voice activity detection on the speech data frame of the to-be-separated role to obtain a speech data frame with a voice activity;

filtering and smoothing the speech data frame with the voice activity based on an energy spectrum of the speech data frame of the to-be-separated role to obtain a filtered and smoothed speech data frame; and

updating the sound source angle based on the filtered and smoothed speech data frame to obtain updated sound source angle data.

14. The one or more memories according to claim 13, wherein the filtering and smoothing the speech data frame with the voice activity based on the energy spectrum of the speech data frame of the to-be-separated role to obtain the filtered and smoothed speech data frame comprises:

filtering and smoothing the speech data frame with the voice activity through a median filter based on a spectral flatness of the energy spectrum of the speech data frame of the to-be-separated role to obtain the filtered and smoothed speech data frame.

15. The one or more memories according to claim 12, wherein the performing the identification on the to-be-separated role based on the sound source angle data to obtain the first identification result of the to-be-separated role comprises:

performing sequence clustering on the sound source angle data to obtain a sequence clustering result of the sound source angle data; and

determining that a role identifier corresponding to the sequence clustering result of the sound source angle data is the first identification result of the to-be-separated role.

16. The one or more memories according to claim 15, wherein the performing the sequence clustering on the sound source angle data to obtain the sequence clustering result of the sound source angle data comprises:

determining a distance between the sound source angle data and a sound source angle sequence clustering center; and

determining the sequence clustering result of the sound source angle data based

on the distance between the sound source angle data and the sound source angle sequence clustering center.

17. The one or more memories according to claim 12, wherein after the
5 obtaining the first identification result of the to-be-separated role, the acts further comprise:

performing voiceprint identification on a speech data frame of the to-be-separated role within a preset time period to obtain a second identification result of the to-be-separated role; and

10 in response to determining that the first identification result is different from the second identification result, using the second identification result to correct the first identification result to obtain a final identification result of the to-be-separated role.