CLAIMS

What is claimed is:

1. An audio encoding method, comprising:

selecting a first target virtual speaker from a preset virtual speaker set based on a current scene audio signal;

generating a first virtual speaker signal based on the current scene audio signal and attribute information of the first target virtual speaker;

encoding the first virtual speaker signal to obtain a bitstream; and

obtaining a main sound field component from the current scene audio signal based on the preset virtual speaker set;

wherein:

selecting the first target virtual speaker from the preset virtual speaker set comprises:

selecting the first target virtual speaker from the preset virtual speaker set based on the main sound field component; and

selecting the first target virtual speaker from the preset virtual speaker set based on the main sound field component comprises:

selecting a higher order ambisonics (HOA) coefficient for the main sound field component from an HOA coefficient set based on the main sound field component, wherein HOA coefficients in the HOA coefficient set are in a one-to-one correspondence with virtual speakers in the preset virtual speaker set; and

determining, as the first target virtual speaker, a virtual speaker that corresponds to the HOA coefficient for the main sound field component and is in the preset virtual speaker set.

2. The method according to claim 1, wherein the selecting the first target virtual speaker from the virtual speaker set based on the main sound field component comprises:

obtaining a configuration parameter of the first target virtual speaker based on the main sound field component;

generating, based on the configuration parameter of the first target virtual speaker, an HOA coefficient for the first target virtual speaker; and

determining, as the target virtual speaker, a virtual speaker that corresponds to the HOA coefficient for the first target virtual speaker and that is in the virtual speaker set.

3. The method according to claim 2, wherein the obtaining a configuration parameter of the first target virtual speaker based on the main sound field component comprises:

determining configuration parameters of a plurality of virtual speakers in the virtual speaker set based on configuration information of an audio encoder; and

selecting the configuration parameter of the first target virtual speaker from the configuration parameters of the plurality of virtual speakers based on the main sound field component.

4. The method according to claim 2 or 3, wherein the configuration parameter of the first target virtual speaker comprises location information and HOA order information of the first target virtual speaker; and

the generating, based on the configuration parameter of the first target virtual speaker, an HOA coefficient for the first target virtual speaker comprises:

determining, based on the location information and the HOA order information of the first target virtual speaker, the HOA coefficient for the first target virtual speaker.

5. The method according to any one of claims 1 to 4, wherein the method further comprises:

encoding the attribute information of the first target virtual speaker, and writing encoded attribute information into the bitstream.

6. The method according to any one of claims 1 to 5, wherein the current scene audio signal comprises a to-be-encoded higher order ambisonics HOA signal, and the attribute information of the first target virtual speaker comprises the HOA coefficient of the first target virtual speaker; and

the generating a first virtual speaker signal based on the current scene audio signal and attribute information of the first target virtual speaker comprises:

performing linear combination on the to-be-encoded HOA signal and the HOA coefficient to obtain the first virtual speaker signal.

7. The method according to any one of claims 1 to 6, wherein the current scene audio signal comprises a to-be-encoded higher order ambisonics HOA signal, and the attribute information of the first target virtual speaker comprises the location information of the first target virtual speaker; and

the generating a first virtual speaker signal based on the current scene audio signal and attribute information of the first target virtual speaker comprises:

obtaining, based on the location information of the first target virtual speaker, the HOA coefficient for the first target virtual speaker; and

performing linear combination on the to-be-encoded HOA signal and the HOA coefficient to obtain the first virtual speaker signal.

8. The method according to any one of claims 1 to 7, wherein the method further comprises:

selecting a second target virtual speaker from the virtual speaker set based on the current scene audio signal;

generating a second virtual speaker signal based on the current scene audio signal and attribute information of the second target virtual speaker; and

encoding the second virtual speaker signal, and writing an encoded second virtual speaker signal into the bitstream.

9. The method according to claim 8, wherein the method further comprises:

performing alignment processing on the first virtual speaker signal and the second virtual speaker signal to obtain an aligned first virtual speaker signal and an aligned second virtual speaker signal;

correspondingly, the encoding the second virtual speaker signal comprises:

encoding the aligned second virtual speaker signal; and

correspondingly, the encoding the first virtual speaker signal comprises:

encoding the aligned first virtual speaker signal.

10. The method according to any one of claims 1 to 7, wherein the method further comprises:

selecting a second target virtual speaker from the virtual speaker set based on the current scene audio signal; and

generating a second virtual speaker signal based on the current scene audio signal and attribute information of the second target virtual speaker; and

correspondingly, the encoding the first virtual speaker signal comprises:

obtaining a downmixed signal and side information based on the first virtual speaker signal and the second virtual speaker signal, wherein the side information indicates a relationship between the first virtual speaker signal and the second virtual speaker signal; and

encoding the downmixed signal and the side information.

11. The method according to claim 10, wherein the method further comprises:

performing alignment processing on the first virtual speaker signal and the second virtual speaker signal to obtain an aligned first virtual speaker signal and an aligned second virtual speaker signal;

correspondingly, the obtaining a downmixed signal and side information based on the first virtual speaker signal and the second virtual speaker signal comprises:

obtaining the downmixed signal and the side information based on the aligned first virtual speaker signal and the aligned second virtual speaker signal; and

correspondingly, the side information indicates a relationship between the aligned first virtual speaker signal and the aligned second virtual speaker signal.

12. The method according to any one of claims 8 to 11, wherein before the selecting a second

target virtual speaker from the virtual speaker set based on the current scene audio signal, the method further comprises:

determining, based on an encoding rate and/or signal type information of the current scene audio signal, whether a target virtual speaker other than the first target virtual speaker needs to be obtained; and

selecting the second target virtual speaker from the virtual speaker set based on the current scene audio signal if the target virtual speaker other than the first target virtual speaker needs to be obtained.

13. An audio decoding method, comprising:

receiving a bitstream;

decoding the bitstream to obtain a virtual speaker signal;

decoding the bitstream to obtain attribute information of a target virtual speaker; and

obtaining a reconstructed scene audio signal based on attribute information of a target virtual speaker and the virtual speaker signal;

wherein:

the attribute information of the target virtual speaker comprises location information of the target virtual speaker; and

obtaining the reconstructed scene audio signal comprises:

determining a higher order ambisonics (HOA) coefficient of the target virtual speaker based on the location information of the target virtual speaker; and

performing synthesis processing on the virtual speaker signal and the HOA coefficient of the target virtual speaker to obtain the reconstructed scene audio signal.

14. The method according to claim 13, wherein the virtual speaker signal is a downmixed signal obtained by downmixing a first virtual speaker signal and a second virtual speaker signal, and the method further comprises:

decoding the bitstream to obtain side information, wherein the side information indicates a relationship between the first virtual speaker signal and the second virtual speaker signal; and

obtaining the first virtual speaker signal and the second virtual speaker signal based on the side information and the downmixed signal; and

correspondingly, the obtaining a reconstructed scene audio signal based on attribute information of a target virtual speaker and the virtual speaker signal comprises:

obtaining the reconstructed scene audio signal based on the attribute information of the target virtual speaker, the first virtual speaker signal, and the second virtual speaker signal.

15. An audio encoding apparatus, comprising:

an obtaining module, configured to obtain a main sound field component from a current scene audio signal based on a preset virtual speaker set, and select a first target virtual speaker from the preset virtual speaker set based on the main sound field component of the current scene audio signal;

a signal generation module, configured to generate a first virtual speaker signal based on the current scene audio signal and attribute information of the first target virtual speaker; and

an encoding module, configured to encode the first virtual speaker signal to obtain a bitstream;

wherein the obtaining module is configured to: select an HOA coefficient for the main sound field component from a higher order ambisonics HOA coefficient set based on the main sound field component, wherein HOA coefficients in the HOA coefficient set are in a one-to-one correspondence with virtual speakers in the virtual speaker set; and determine, as the first target virtual speaker, a virtual speaker that corresponds to the HOA coefficient for the main sound field component and that is in the virtual speaker set.

16. The apparatus according to claim 15, wherein the obtaining module is configured to: obtain a configuration parameter of the first target virtual speaker based on the main sound field component; generate, based on the configuration parameter of the first target virtual speaker, an HOA coefficient for the first target virtual speaker; and determine, as the target virtual speaker, a virtual speaker that corresponds to the HOA coefficient for the first target virtual speaker and that is in the virtual speaker set.

17. The apparatus according to claim 16, wherein the obtaining module is configured to: determine configuration parameters of a plurality of virtual speakers in the virtual speaker set based on configuration information of an audio encoder; and select the configuration parameter of the first target virtual speaker from the configuration parameters of the plurality of virtual speakers based on the main sound field component.

18. The apparatus according to claim 16 or 17, wherein the configuration parameter of the first target virtual speaker comprises location information and HOA order information of the first target virtual speaker; and

the obtaining module is configured to determine, based on the location information and the HOA order information of the first target virtual speaker, the HOA coefficient for the first target virtual speaker.

19. The apparatus according to any one of claims 15 to 18, wherein the encoding module is further configured to encode the attribute information of the first target virtual speaker, and write encoded attribute information into the bitstream.

20. The apparatus according to any one of claims 15 to 19, wherein the current scene audio signal

comprises a to-be-encoded HOA signal, and the attribute information of the first target virtual speaker comprises the HOA coefficient of the first target virtual speaker; and

the signal generation module is configured to perform linear combination on the to-be-encoded HOA signal and the HOA coefficient to obtain the first virtual speaker signal.

21. The apparatus according to any one of claims 15 to 19, wherein the current scene audio signal comprises a to-be-encoded higher order ambisonics HOA signal, and the attribute information of the first target virtual speaker comprises the location information of the first target virtual speaker; and

the signal generation module is configured to: obtain, based on the location information of the first target virtual speaker, the HOA coefficient for the first target virtual speaker; and perform linear combination on the to-be-encoded HOA signal and the HOA coefficient to obtain the first virtual speaker signal.

22. The apparatus according to any one of claims 15 to 21, wherein

the obtaining module is configured to select a second target virtual speaker from the virtual speaker set based on the current scene audio signal;

the signal generation module is configured to generate a second virtual speaker signal based on the current scene audio signal and attribute information of the second target virtual speaker; and

the encoding module is configured to encode the second virtual speaker signal, and write an encoded second virtual speaker signal into the bitstream.

23. The apparatus according to claim 22, wherein

the signal generation module is configured to perform alignment processing on the first virtual speaker signal and the second virtual speaker signal to obtain an aligned first virtual speaker signal and an aligned second virtual speaker signal;

correspondingly, the encoding module is configured to encode the aligned second virtual speaker signal; and

correspondingly, the encoding module is configured to encode the aligned first virtual speaker signal.

24. The apparatus according to any one of claims 15 to 21, wherein

the obtaining module is configured to select a second target virtual speaker from the virtual speaker set based on the current scene audio signal;

the signal generation module is configured to generate a second virtual speaker signal based on the current scene audio signal and attribute information of the second target virtual speaker; and

correspondingly, the encoding module is configured to obtain a downmixed signal and side information based on the first virtual speaker signal and the second virtual speaker signal, wherein

the side information indicates a relationship between the first virtual speaker signal and the second virtual speaker signal; and encode the downmixed signal and the side information.

25. The apparatus according to claim 24, wherein

the signal generation module is configured to perform alignment processing on the first virtual speaker signal and the second virtual speaker signal to obtain an aligned first virtual speaker signal and an aligned second virtual speaker signal;

correspondingly, the encoding module is configured to obtain the downmixed signal and the side information based on the aligned first virtual speaker signal and the aligned second virtual speaker signal; and

correspondingly, the side information indicates a relationship between the aligned first virtual speaker signal and the aligned second virtual speaker signal.

26. The apparatus according to any one of claims 15 to 25, wherein the obtaining module is configured to: before the selecting a second target virtual speaker from the virtual speaker set based on the current scene audio signal, determine, based on an encoding rate and/or signal type information of the current scene audio signal, whether a target virtual speaker other than the first target virtual speaker needs to be obtained; and select the second target virtual speaker from the virtual speaker set based on the current scene audio signal if the target virtual speaker other than the first target virtual speaker needs to be obtained.

27. An audio decoding apparatus, comprising:

a receiving module, configured to receive a bitstream;

a decoding module, configured to decode the bitstream to obtain a virtual speaker signal and decode the bitstream to obtain attribute information of a target virtual speaker; and

a reconstruction module, configured to obtain a reconstructed scene audio signal based on the attribute information of the target virtual speaker and the virtual speaker signal;

wherein the attribute information of the target virtual speaker comprises location information of the target virtual speaker; and

the reconstruction module is configured to determine an HOA coefficient of the target virtual speaker based on the location information of the target virtual speaker; and perform synthesis processing on the virtual speaker signal and the HOA coefficient of the target virtual speaker to obtain the reconstructed scene audio signal.

28. The apparatus according to claim 27, wherein the virtual speaker signal is a downmixed signal obtained by downmixing a first virtual speaker signal and a second virtual speaker signal, and the apparatus further comprises a signal compensation module, wherein

the decoding module is configured to decode the bitstream to obtain side information, wherein the side information indicates a relationship between the first virtual speaker signal and the second virtual speaker signal;

the signal compensation module is configured to obtain the first virtual speaker signal and the second virtual speaker signal based on the side information and the downmixed signal; and

correspondingly, the reconstruction module is configured to obtain the reconstructed scene audio signal based on the attribute information of the target virtual speaker, the first virtual speaker signal, and the second virtual speaker signal.

29. An audio encoding apparatus, wherein the audio encoding apparatus comprises at least one processor, and the at least one processor is configured to be coupled to a memory, and read and execute instructions in the memory, to implement the method according to any one of claims 1 to 12.

30. The audio encoding apparatus according to claim 29, wherein the audio encoding apparatus further comprises the memory.

31. An audio decoding apparatus, wherein the audio decoding apparatus comprises at least one processor, and the at least one processor is configured to be coupled to a memory, and read and execute instructions in the memory, to implement the method according to any one of claims 13 to 14.

32. The audio decoding apparatus according to claim 31, wherein the audio decoding apparatus further comprises the memory.

33. A computer-readable storage medium, comprising instructions, wherein when the instructions are run on a computer, the computer is enabled to perform the method according to any one of claims 1 to 12 or claims 13 to 14.

34. A computer-readable storage medium, comprising a bitstream generated by using the method according to any one of claims 1 to 12.