CLAIMS

1. A device configured to play one or more of a plurality of audio streams, the device comprising: a memory configured to store the plurality of audio streams, each of the audio streams representative of a soundfield; and one or more processors coupled to the memory, and configured to: present a user interface to a user; obtain a first indication from the user via the user interface representing a desired listening position; select, based on the first indication, a first at least one audio stream of the plurality of audio streams; obtain a second indication from the user via the user interface representing an audio source distance threshold, the audio source distance threshold being representative of a distance from an audio source associated with the desired listening position; determine that the user is moving from one location to another location; determine an audio source distance as a distance between the audio source and a virtual location of the user; compare the audio source distance to the audio source distance threshold; based on the audio source distance being greater than the audio source distance threshold, select, a second at least one audio stream of the plurality of audio streams, wherein the second at least one audio stream is not identical to the first at least one audio stream.

2. The device of claim 1, wherein the memory is further configured to store location information associated with coordinates of an acoustical space in which a corresponding one of the plurality of audio streams was captured or synthesized.

3. The device of claim 2, wherein the user interface comprises one or more of a graphical user interface, a gesture-based user interface, a voice command-based user interface, a touch-based user interface.

4. The device of claim 3, wherein the user interface is configured to obtain user input in at least one of single touch, multi-touch, gesture, voice command, or tap.

5. The device of claim 3, wherein the user interface comprises the graphical user interface, and the graphical user interface comprises representations associated with coordinates of the acoustical space in which the plurality of audio streams were captured or synthesized.

6. The device of claim 5, wherein the representations are arranged in the graphical user interface with a spatial relationship representing relative positions of the coordinates of the

acoustical space in which the plurality of audio streams were captured or synthesized.

7. The device of claim 6, wherein the one or more processors is further configured to, in response to obtaining the first indication representing the desired listening position, output, for display, a graphical user interface element suggesting an alternative listening position.

8. The device of claim 1, wherein the one or more processors is further configured to combine at least two audio streams based on the first indication by at least one of mixing the at least two audio streams or interpolating a third audio stream based on the at least two audio streams.

9. The device of claim 1, wherein the one or more processors are further configured to obtain via the user interface an importance indication representing an importance assigned to an audio stream, and wherein the importance indicates a relative gain to be applied to the audio stream.

10. The device of claim 1, wherein the one or more processors are further configured to set, based on the second indication, the audio source distance threshold.

11. The device of claim 1, wherein the one or more processors are further configured to obtain a third indication, via the user interface, from the user that the user desires to activate a snapping mode, the snapping mode being configured to transform spatial coordinates for audio rendering based on a present location of the user and at least one audio source distance threshold.

12. The device of claim 11, wherein the snapping mode is a hard snapping mode including a single audio source distance threshold or a soft snapping mode including two audio source distance thresholds.

13. The device of claim 12, wherein the audio source distance threshold is a first audio source distance threshold and wherein the one or more processors are further configured to: determine a second audio source distance threshold, and wherein the one or more processors are configured to select the second at least one audio stream of the plurality of audio streams, further based on the second audio source distance threshold.

14. The device of claim 13, wherein the one or more processors are configured to combine the two audio streams by applying a function F(x) to the two audio streams.

15. The device of claim 1, wherein the device comprises an extended reality headset, and wherein a displayed world comprises a visual scene represented by video data captured by a camera.

16. The device of claim 1, wherein the device comprises an extended reality headset, and wherein a displayed world comprises a virtual world.

17. The device of claim 1, wherein the device comprises a mobile handset.

18. The device of claim 1, wherein the device further comprises a wireless transceiver, the wireless transceiver being coupled to the one or more processors and being configured to receiver a wireless signal, wherein the wireless signal comprises at least one of Bluetooth, or Wi-Fi, or conforms to a fifth generation (5G) cellular protocol.

19. A method comprising: storing, by a memory, a plurality of audio streams, each of the audio streams representative of a soundfield, the memory being communicatively coupled to one or more processors; presenting, by the one or more processors, a user interface to a user; obtaining, by the one or more processors, from the user via the user interface, a first indication representing a desired listening position; selecting, based on the first indication, a first at least one audio stream of the plurality of audio streams; obtaining, by the one or more processors, from the user via the user interface, a second indication representing an audio source distance threshold, the audio source distance threshold being representative of a distance from an audio source associated with the desired listening position; determining that the user is moving from one location to another location; determining an audio source distance as a distance between the audio source and a virtual location of the user; comparing the audio source distance to the audio source distance threshold; and based on the audio source distance being greater than the audio source distance threshold, selecting, by the one or more processors, a second at least one audio stream of the plurality of audio streams, wherein the second at least one audio stream is not identical to the first at least one audio stream.

20. The method of claim 19, further comprising storing location information associated with coordinates of an acoustical space in which a corresponding one of the plurality of audio streams was captured or synthesized.

21. The method of claim 20, wherein the user interface comprises one or more of a graphical user interface, a gesture-based user interface, a voice command-based user interface, a touch-based user interface.

22. The method of claim 21, further comprising, obtaining, by the one or more processors, via the user interface, user input, the user input comprising at least one of single touch, multi-touch, gesture, voice command, or tap.

23. The method of claim 21, wherein the user interface comprises the graphical user interface, and the graphical user interface comprises representations associated with coordinates of the acoustical space in which the plurality of audio streams were captured or synthesized.

24. The method of claim 23, wherein the representations are arranged in the graphical user interface with a spatial relationship representing relative positions of the coordinates of the acoustical space in which the plurality of audio streams were captured or synthesized.

25. The method of claim 24, further comprising outputting, in response to obtaining the first indication representing the desired listening position, and for display, a graphical user interface element suggesting an alternative listening position.

26. The method claim 19, further comprising combining, by the one or more processors, at least two audio streams based on the first indication by at least one of mixing the at least two audio streams or interpolating a third audio stream based on the at least two audio streams.

27. The method of claim 19, further comprising, obtaining, by the one or more processors via the user interface an importance indication representing an importance assigned to an audio stream, and wherein the importance indicates a relative gain to be applied to the audio stream.

28. The method of claim 19, further comprising setting, by the one or more processors, and based on the second indication, the audio source distance threshold.

29. The method of claim 28, further comprising obtaining, by the one or more processors via the user interface, a third indication from the user that the user desires to activate a snapping mode, the snapping mode being configured to transform spatial coordinates for audio rendering based on a present location of the user and at least one audio source distance threshold.

30. The method of claim 29, wherein the snapping mode is a hard snapping mode including a single audio source distance threshold or a soft snapping mode including two audio source distance thresholds.

31. The method of claim 30, wherein the audio source distance threshold is a first audio source distance threshold, the method further comprising: determining, by the one or more processors via the user interface, a second audio source distance threshold, wherein the selecting the second at least one audio stream of the plurality of audio streams is further based on the second audio source distance threshold.

32. The method of claim 31, wherein the one or more processors are configured to combine the two audio streams by applying a function F(x) to the two audio streams.

33. The method of claim 19, wherein the method is performed by an extended reality headset, and wherein a displayed world comprises a visual scene represented by video data captured by a camera.

34. The method of claim 19, wherein the method is performed on an extended reality headset, and wherein a displayed world comprises a virtual world.

35. The method of claim 19, wherein the method is performed on a mobile handset.

36. The method of claim 19, further comprising receiving, by a wireless transceiver, a wireless signal, wherein the wireless signal comprises at least one of a Bluetooth signal, a Wi-Fi signal, or a signal in accordance with a fifth generation (5G) cellular standard.

37. A device configured to play one or more of a plurality of audio streams, the device comprising: means for storing the plurality of audio streams, each of the audio streams representative of a soundfield; means for presenting a user interface to a user; means for

obtaining, from a user via the user interface, a first indication representing a desired listening position; means for selecting, based on the first indication, a first at least one audio stream of the plurality of audio streams; means for obtaining, from the user via the user interface, a second indication representing an audio source distance threshold, the audio source distance threshold being representative of a distance from an audio source associated with the desired listening position; and means for determining that the user is moving from one location to another location; means for determining an audio source distance as a distance between the audio source and a virtual location of the user; means for comparing the audio source distance to the audio source distance threshold; and means for selecting, based on the audio source distance being greater than the audio source distance threshold, a second at least one audio stream of the plurality of audio streams, wherein the second at least one audio stream is not identical to the first at least one audio stream.

38. A non-transitory computer-readable storage medium having stored thereon instructions that, when executed, cause one or more processors to: store a plurality of audio streams, each of the audio streams representative of a soundfield; present a user interface to a user; obtain, via the user interface from the user, a first indication representing a desired listening position; select, based on the first indication, a first at least one audio stream of the plurality of audio streams; obtain, via the user interface from the user, a second indication representing an audio source distance threshold, the audio source distance threshold being representative of a distance from an audio source associated with the desired listening position; determine that the user is moving from one location to another location; determine an audio source distance as a distance between the audio source and a virtual location of the user; compare the audio source distance to the audio source distance threshold; and based on the audio source distance being greater than the audio source distance threshold, select, based on the first indication and the second indication, a second at least one audio stream of the plurality of audio streams, wherein the second at least one audio stream is not identical to the first at least one audio stream.