

CLAIMS:

What is claimed is:

1. A system for processing audio signals, comprising:

5 an authoring component configured to receive a plurality of audio signals and to generate a plurality of monophonic audio streams and one or more metadata sets associated with each of the audio streams and specifying a playback location of a respective audio stream, wherein the audio streams are identified as either channel-based audio or object-based audio, and wherein the playback location of the channel-based audio comprises speaker designations of speakers in a speaker array, and the playback location of the object-based audio comprises a location in three-dimensional space; and further wherein a first set of metadata is applied by default to one or more of the plurality of audio streams, and a second set of metadata is associated with a specific condition of a playback environment and is applied to the one or more of the plurality of audio streams instead of the first set if a condition of the playback environment matches the specific condition of the playback environment; and

15 a rendering system coupled to the authoring component and configured to receive a bitstream encapsulating the plurality of monophonic audio streams and the one or more data sets, and to render the audio streams to a plurality of speaker feeds corresponding to speakers in the playback environment in accordance with the one or more metadata sets based on the condition of the playback environment.

2. The system of claim 1 wherein each metadata set includes metadata elements associated with each object-based stream, the metadata elements for each object-based stream specifying spatial parameters controlling the playback of a corresponding object-based sound, and comprising one or more of: sound position, sound width, and sound velocity; and further wherein each metadata set includes metadata elements associated with each channel-based stream, and the speaker array comprises speakers arranged in a defined surround sound configuration, and wherein the metadata elements associated with each channel-based stream comprises designations of surround-sound channels of the speakers in the speaker array in accordance with a defined surround-sound standard.

3. The system of claim 1 wherein the speaker array includes additional speakers for playback of object-based streams that are positioned in the playback environment in accordance with set up instructions from a user based on the condition of the playback environment, and wherein the playback condition depends on variables comprising: size and shape of a room of the playback environment, occupancy, material composition, and ambient noise; and further wherein the system receives a set-up file from the user that includes at least a list of speaker designations and a mapping of channels to individual speakers of the speaker array, information regarding grouping of speakers, and a run-time mapping based on a relative position of speakers to the playback environment.

4. The system of claim 1 wherein the authoring component includes a mixing console having controls operable by the user to specify playback levels of the audio streams comprising the original audio content, and wherein the metadata elements associated with each respective object-based stream are automatically generated upon input to the mixing console controls by the user.

5. The system of claim 1 wherein the metadata sets include metadata to enable upmixing or downmixing of at least one of the channel-based audio streams and the object-based audio streams in accordance with a change from a first configuration of the speaker array to a second configuration of the speaker array.

6. The system of claim 3 wherein the content type is selected from the group consisting of: dialog, music, and effects, and each content type is embodied in a respective set of channel-based streams or object-based streams, and further wherein sound components of each content type are transmitted to defined speaker groups of one or more speaker groups designated within the speaker array.

7. The system of claim 6 wherein the speakers of the speaker array are placed at specific positions within the playback environment, and wherein metadata elements associated with each respective object-based stream specify that one or more sound components are rendered to a speaker feed for playback through a speaker nearest an intended playback location of the sound component, as indicated by the position metadata.

8. The system of claim 1 wherein the playback location comprises a spatial position relative to a screen within the playback environment, or a surface that encloses the playback environment, and wherein the surface comprises a front plane, a back plane, a left plane, right plane, an upper plane, and a lower plane.

5

9. The system of claim 1 further comprising a codec coupled to the authoring component and the rendering component and configured to receive the plurality of audio streams and metadata and to generate a single digital bitstream containing the plurality of audio streams in an ordered fashion.

10

10. The system of claim 9 wherein the rendering component further comprises means for selecting a rendering algorithm utilized by the rendering component, the rendering algorithm selected from the group consisting of: binaural, stereo dipole, Ambisonics, Wave Field Synthesis (WFS), multi-channel panning, raw stems with position metadata, dual balance, and vector-based amplitude panning.

15

11. The system of claim 1 wherein the playback location for each of the audio streams is independently specified with respect to either an egocentric frame of reference or an allocentric frame of reference, wherein an egocentric frame of reference is taken in relation to a listener in the playback environment, and wherein the allocentric frame of reference is taken with respect to a characteristic of the playback environment.

20

12. A system for processing audio signals, comprising:

an authoring component configured to receive a plurality of audio signals and to generate a plurality of monophonic audio streams and metadata associated with each of the audio streams and specifying a playback location of a respective audio stream, wherein the audio streams are identified as either channel-based audio or object-based audio, and wherein the playback location of the channel-based audio comprises speaker designations of speakers in a speaker array, and the playback location of the object-based audio comprises a location in three-dimensional space, and wherein each object-based audio stream is rendered in at least one specific speaker of the speaker array; and

25

30

a rendering system coupled to the authoring component and configured to receive a bitstream encapsulating the plurality of monophonic audio streams and metadata, and to render the audio streams to a plurality of speaker feeds corresponding to speakers in a

5 playback environment, wherein the speakers of the speaker array are placed at specific positions within the playback environment, and wherein metadata elements associated with each respective object-based stream specify that one or more sound components are rendered to a speaker feed for playback through a speaker nearest an intended playback location of the sound component, such that an object-based stream effectively snaps to the speaker nearest to the intended playback location.

10 13. The system of claim 12 wherein the metadata comprises two or more metadata sets, and the rendering system renders the audio streams in accordance with one of the two or more metadata sets based on a condition of the playback environment, wherein a first set of metadata is applied to one or more of the plurality of audio streams for a first condition of the playback environment, and a second set of metadata is applied to the one or more plurality of audio streams for a second condition of the playback environment; and wherein each metadata set includes metadata elements associated with each object-based stream, the
15 metadata elements for each object-based stream specifying spatial parameters controlling the playback of a corresponding object-based sound, and comprising one or more of: sound position, sound width, and sound velocity; and further wherein each metadata set includes metadata elements associated with each channel-based stream, and the speaker array comprises speakers arranged in a defined surround sound configuration, and wherein the
20 metadata elements associated with each channel-based stream comprise designations of surround-sound channels of the speakers in the speaker array in accordance with a defined surround-sound standard.

25 14. The system of claim 12 wherein the speaker array includes additional speakers for playback of object-based streams that are positioned in the playback environment in accordance with set up instructions from a user based on a condition of the playback environment, and wherein the playback condition depends on variables comprising: size and shape of a room of the playback environment, occupancy, material composition, and ambient noise; and further wherein the system receives a set-up file from the user that includes at least
30 a list of speaker designations and a mapping of channels to individual speakers of the speaker array, information regarding grouping of speakers, and a run-time mapping based on a relative position of speakers to the playback environment, and wherein an object stream rendered to a speaker feed for playback through a speaker nearest an intended playback location of the sound component snaps to a single speaker of the additional speakers.

15. The system of claim 14 wherein the intended playback location comprises a spatial position relative to a screen within the playback environment or a surface that encloses the playback environment, and wherein the surface comprises a front plane, a back plane, a left plane, a top plane, and a floor plane.

16. A system for processing audio signals, comprising:

an authoring component configured to receive a plurality of audio signals and to generate a plurality of monophonic audio streams and metadata associated with each of the audio streams and specifying a playback location of a respective audio stream, wherein the audio streams are identified as either channel-based audio or object-based audio, and wherein the playback location of the channel-based audio comprises speaker designations of speakers in a speaker array, and the playback location of the object-based audio comprises a location in three-dimensional space relative to a playback environment containing the speaker array, and wherein each object based audio stream is rendered in at least one specific speaker of the speaker array; and

a rendering system coupled to the authoring component and configured to receive a first map of speakers to audio channels comprising a list of speakers and their respective locations within the playback environment and a bitstream encapsulating the plurality of monophonic audio streams and metadata, and to render the audio streams to a plurality of speaker feeds corresponding to speakers in the playback environment in accordance with a run-time mapping based on a relative position of speakers to the playback environment and a condition of the playback environment.

17. The system of claim 16 wherein the condition of the playback environment depends on variables comprising: size and shape of a room of the playback environment, occupancy, material composition, and ambient noise.

18. The system of claim 17 wherein the first map is specified in a set-up file that includes at least a list of speaker designations and a mapping of channels to individual speakers of the speaker array, and information regarding grouping of speakers.

19. The system of claim 18 wherein the intended playback location comprises a spatial position relative to a screen within the playback environment or a surface of an enclosure

containing the playback environment, and wherein the surface comprises one of: a front plane, a back plane, a side plane, a top plane, and a floor plane of the enclosure.

20. The system of claim 19 wherein the speaker array comprises speakers arranged in a defined surround sound configuration, and wherein the metadata elements associated with each channel-based stream comprise designations of surround-sound channels of the speakers in the speaker array in accordance with a defined surround-sound standard, and further wherein certain object-based streams are played through additional speakers of the speaker array, and wherein the run-time mapping dynamically determines which individual speakers of the speaker array play back a corresponding object-based stream during a playback process.

21. A method of authoring audio signals for rendering, comprising:
receiving a plurality of audio signals;

generating a plurality of monophonic audio streams and one or more metadata sets associated with each of the audio streams and specifying a playback location of a respective audio stream, wherein the audio streams are identified as either channel-based audio or object-based audio, and wherein the playback location of the channel-based audio comprises speaker designations of speakers in a speaker array, and the playback location of the object-based audio comprises a location in three-dimensional space relative to a playback environment containing the speaker array; and further wherein a first set of metadata is applied to one or more of the plurality of audio streams for a first condition of the playback environment, and a second set of metadata is applied to the one or more of plurality of audio streams for a second condition of the playback environment; and

encapsulating the plurality of monophonic audio streams and the one or more metadata sets in a bitstream for transmission to a rendering system configured to render the audio streams to a plurality of speaker feeds corresponding to speakers in the playback environment in accordance with the one or more metadata sets based on a condition of the playback environment.

22. The method of claim 21 wherein each metadata set includes metadata elements associated with each object-based stream, the metadata elements for each object-based stream specifying spatial parameters controlling the playback of a corresponding object-based sound, and comprising one or more of: sound position, sound width, and sound velocity; and further

wherein each metadata set includes metadata elements associated with each channel-based stream, and the speaker array comprises speakers arranged in a defined surround sound configuration, and wherein the metadata elements associated with each channel-based stream comprises designations of surround-sound channels of the speakers in the speaker array in accordance with a defined surround-sound standard.

23. The method of claim 21 wherein the speaker array includes additional speakers for playback of object-based streams that are positioned in the playback environment, the method further comprising receiving set up instructions from a user based on the condition of the playback environment, and wherein the playback condition depends on variables comprising: size and shape of a room of the playback environment, occupancy, material composition, and ambient noise; the setup instructions further including at least a list of speaker designations and a mapping of channels to individual speakers of the speaker array, information regarding grouping of speakers, and a run-time mapping based on a relative position of speakers to the playback environment.

24. The method of claim 23 further comprising:
receiving, from a mixing console having controls operated by a user to specify playback levels of the audio streams comprising the original audio content; and
automatically generating the metadata elements associated with each respective object-based stream generated upon receipt of the user input.

25. A method of rendering audio signals, comprising:
receiving a bitstream encapsulating a plurality of monophonic audio streams and one or more metadata sets in a bitstream from an authoring component configured to receive a plurality of audio signals, and generate a plurality of monophonic audio streams and one or more metadata sets associated with each of the audio streams and specifying a playback location of a respective audio stream, wherein the audio streams are identified as either channel-based audio or object-based audio, and wherein the playback location of the channel-based audio comprises speaker designations of speakers in a speaker array, and the playback location of the object-based audio comprises a location in three-dimensional space relative to a playback environment containing the speaker array; and further wherein a first set of metadata is applied to one or more of the plurality of audio streams for a first condition of the

playback environment, and a second set of metadata is applied to the one or more plurality of audio streams for a second condition of the playback environment; and

rendering the plurality of audio streams to a plurality of speaker feeds corresponding to speakers in the playback environment in accordance with the one or more metadata sets
5 based on a condition of the playback environment.

26. A method of creating audio content comprising a plurality of monophonic audio streams processed in an authoring component, wherein the monophonic audio streams comprise at least one channel-based audio stream and at least one object-based audio stream,
10 the method comprising:

indicating whether each audio stream of the plurality of audio streams is a channel-based stream or an object-based stream;

associating with each channel-based stream a metadata element specifying a channel position for rendering the respective channel-based stream to one or more speakers within a
15 playback environment;

associating with each object-based stream one or metadata elements specifying an object-based position for rendering the respective object-based stream to one or more speakers within the playback environment with respect to an allocentric frame of reference defined with respect to size and dimensions of the playback environment; and

20 assembling the plurality of monophonic streams and associated metadata into a signal.

27. The method of claim 26 wherein the playback environment includes an array of speakers placed at defined locations and orientations relative to a reference point of an enclosure embodying the playback environment.
25

28. The method of claim 27 wherein a first set of speakers of the array of speakers comprises speakers arranged according to a defined surround sound system, and wherein a second set of speakers of the array of speakers comprises speakers arranged according to an adaptive audio scheme.
30

29. The method of claim 28 further comprising:

defining an audio type for sets of the plurality of monophonic audio streams, wherein the audio type is selected from the group consisting of dialog, music, and effects; and

transmitting the sets of audio streams to specific sets of speakers based on the audio type for a respective set of audio streams.

30. The method of claim 29 further comprising automatically generating the metadata elements through an authoring component implemented in a mixing console having controls operable by a user to specify playback levels of the monophonic audio streams.

31. The method of claim 30 further comprising packaging the plurality of monophonic audio streams and associated metadata elements into a single digital bitstream within an encoder.

32. A method of creating audio content comprising:

determining values of one or more metadata elements in a first metadata group associated with programming of the audio content for processing in a hybrid audio system handling both channel-based and object-based audio content;

determining values of one or more metadata elements in a second metadata group associated with storage and rendering characteristics of the audio content in the hybrid audio system; and

determining values of one or more metadata elements in a third metadata group associated with audio source position and control information for rendering the channel-based and object-based audio content.

33. The method of claim 32 wherein the audio source position for rendering the channel-based audio content comprises names associated with speakers in a surround sound speaker system, wherein the names define a location of respective speakers relative to one or more reference positions in the playback environment.

34. The method of claim 33 wherein the control information for rendering the channel-based audio content comprises upmix and downmix information for rendering audio content in different surround sound configurations, and wherein the metadata includes metadata for enabling or disabling an upmix and/or downmix function.

35. The method of claim 32 wherein the audio source position for rendering the object-based audio content comprises values associated with one or more mathematical functions

specifying an intended playback location for playback of a sound component of the object based audio content.

36. The method of claim 35 wherein the mathematical functions are selected from the group consisting of: three-dimensional coordinates specified as x, y, z coordinate values, a surface definition plus a set of two-dimensional coordinates, and a curve definition plus a one-dimensional linear position coordinate, and a scalar position on a screen in the playback environment.

37. The method of claim 36 wherein the control information for rendering the object-based audio content comprises values specifying individual speakers or speaker groups within the playback environment through which the sound component is played.

38. The method of claim 37 wherein the control information for rendering the object-based audio content further comprises a binary value specifying the sound source to be snapped to a nearest speaker or nearest speaker group within the playback environment.

39. A method of defining an audio transport protocol comprising:
defining values of one or more metadata elements in a first metadata group associated with programming of the audio content for processing in a hybrid audio system handling both channel-based and object-based audio content;
defining values of one or more metadata elements in a second metadata group associated with storage and rendering characteristics of the audio content in the hybrid audio system; and
defining values of one or more metadata elements in a third metadata group associated with audio source position and control information for rendering the channel-based and object-based audio content.

What is claimed is:

1. A system for processing audio signals, comprising an authoring component configured to:
receive a plurality of audio signals; generate an adaptive audio mix comprising a plurality of
5 monophonic audio streams and one or more metadata sets associated with each of the
plurality of monophonic audio streams and specifying a playback location of a respective
monophonic audio stream, wherein at least some of the plurality of monophonic audio
streams are identified as channel-based audio and wherein the others of the plurality of
monophonic audio streams are identified as object-based audio, and wherein the playback
10 location of the channel-based audio comprises speaker designations of speakers in a speaker
array, and the playback location of the object-based audio comprises a location in three-
dimensional space relative to a playback environment containing the speaker array; and
further wherein a first metadata set is applied to one or more of the plurality of monophonic
audio streams for a first condition of the playback environment, and a second metadata set is
15 applied to the one or more of the plurality of monophonic audio streams for a second
condition of the playback environment; and encapsulate the plurality of monophonic audio
streams and the at least two metadata sets in a bitstream for transmission to a rendering
system configured to render the plurality of monophonic audio streams to a plurality of
speaker feeds corresponding to speakers in the playback environment in accordance with the
20 at least two metadata sets based on a condition of the playback environment.

2. The system of claim 1 wherein the authoring component includes a mixing console
having controls operable by the user to specify playback levels of the plurality of
monophonic audio streams comprising the original audio content, and wherein the metadata
25 elements associated with each respective object-based stream are automatically generated
upon input to the mixing console controls by the user.

3. The system of claim 1 further comprising an encoder coupled to the authoring component
and configured to receive the plurality of monophonic audio streams and metadata and to
30 generate a single digital bitstream containing the plurality of monophonic audio streams in
an ordered fashion.

4. A system for processing audio signals, comprising a rendering system configured to:
receive a bitstream encapsulating a plurality of monophonic audio streams and at least two

metadata sets in a bitstream from an authoring component configured to receive a plurality of audio signals, and generate a plurality of monophonic audio streams and one or more metadata sets associated with each of the plurality of monophonic audio streams and specifying a playback location of a respective monophonic audio stream, wherein at least
5 some of the plurality of monophonic audio streams are identified as channel-based audio and wherein the others of the plurality of monophonic audio streams are identified as object-based audio, and wherein the playback location of the channel-based audio comprises speaker designations of speakers in a speaker array, and the playback location of the object-based audio comprises a location in three-dimensional space relative to a playback
10 environment containing the speaker array; and further wherein a first metadata set is applied to one or more of the plurality of monophonic audio streams for a first condition of the playback environment, and a second metadata set is applied to the one or more plurality of monophonic audio streams for a second condition of the playback environment; and render the plurality of monophonic audio streams to a plurality of speaker feeds corresponding to
15 speakers in the playback environment in accordance with the at least two metadata sets based on a condition of the playback environment.

5. The system of claim 4 wherein each metadata set includes metadata elements associated with each object-based stream, the metadata elements for each object-based stream
20 specifying spatial parameters controlling the playback of a corresponding object-based sound, and comprising one or more of: sound position, sound width, and sound velocity; and further wherein each metadata set includes metadata elements associated with each channel-based stream, and the speaker array comprises speakers arranged in a defined surround sound configuration, and wherein the metadata elements associated with each channel-based
25 stream comprises designations of surround-sound channels of the speakers in the speaker array in accordance with a defined surround-sound standard.

6. The system of claim 4 wherein the speaker array includes additional speakers for playback of object-based streams that are positioned in the playback environment in
30 accordance with set up instructions from a user based on the condition of the playback environment, and wherein the playback condition depends on variables comprising: size and shape of a room of the playback environment, occupancy, material composition, and ambient noise; and further wherein the system receives a set-up file from the user that includes at least a list of speaker designations and a mapping of channels to individual

speakers of the speaker array, information regarding grouping of speakers, and a mapping based on a relative position of speakers to the playback environment.

7. The system of claim 4 wherein the metadata sets include metadata to enable upmixing or downmixing of at least one of the channel-based monophonic audio streams and the object-based monophonic audio streams in accordance with a change from a first configuration of the speaker array to a second configuration of the speaker array.

8. The system of claim 6 wherein the metadata sets include metadata indicative of a content type of a monophonic audio stream; wherein the content type is selected from the group consisting of: dialog, music, and effects, and each content type is embodied in a respective set of channel-based streams or object-based streams, and further wherein sound components of each content type are transmitted to defined speaker groups of one or more speaker groups designated within the speaker array.

9. The system of claim 8 wherein the speakers of the speaker array are placed at specific positions within the playback environment, and wherein metadata elements associated with each respective object-based stream specify that one or more sound components are rendered to a speaker feed for playback through a speaker nearest an intended playback location of the sound component, as indicated by the position metadata.

10. The system of claim 4 wherein the playback location comprises a spatial position relative to a screen within the playback environment, or a surface that encloses the playback environment, and wherein the surface comprises a front plane, a back plane, a left plane, right plane, an upper plane, and a lower plane.

11. The system of claim 4 wherein the rendering system further comprises means for selecting a rendering algorithm utilized by the rendering system, the rendering algorithm selected from the group consisting of: binaural, stereo dipole, Ambisonics, Wave Field Synthesis (WFS), multi-channel panning, raw stems with position metadata, dual balance, and vector-based amplitude panning.

12. The system of claim 4 wherein the playback location for each of the plurality of monophonic audio streams is independently specified with respect to either an egocentric

frame of reference or an allocentric frame of reference, wherein an egocentric frame of reference is taken in relation to a listener in the playback environment, and wherein the allocentric frame of reference is taken with respect to a characteristic of the playback environment.

5

13. A method of authoring audio signals for rendering, comprising: receiving a plurality of audio signals; generating an adaptive audio mix comprising a plurality of monophonic audio streams and one or more metadata sets associated with each of the plurality of monophonic audio streams and specifying a playback location of a respective monophonic audio stream, wherein at least some of the plurality of monophonic audio streams are identified as channel-based audio and wherein the others of the plurality of monophonic audio streams are identified as object-based audio, and wherein the playback location of the channel-based audio comprises speaker designations of speakers in a speaker array, and the playback location of the object-based audio comprises a location in three-dimensional space relative to a playback environment containing the speaker array; and further wherein a first metadata set is applied to one or more of the plurality of monophonic audio streams for a first condition of the playback environment, and a second metadata set is applied to the one or more of the plurality of monophonic audio streams for a second condition of the playback environment; and encapsulating the plurality of monophonic audio streams and the one or more metadata sets in a bitstream for transmission to a rendering system configured to render the plurality of monophonic audio streams to a plurality of speaker feeds corresponding to speakers in the playback environment in accordance with the at least two metadata sets based on a condition of the playback environment.

14. The method of claim 13 further comprising: receiving, from a mixing console having controls operated by a user to specify playback levels of the plurality of monophonic audio streams comprising the original audio content; and automatically generating the metadata elements associated with each respective object-based stream generated upon receipt of the user input.

30

15. A method of rendering audio signals, comprising: receiving a bitstream encapsulating a plurality of monophonic audio streams and at least two metadata sets in a bitstream from an authoring component configured to receive a plurality of audio signals, and generate a plurality of monophonic audio streams and one or more metadata sets associated with each

of the plurality of monophonic audio streams and specifying a playback location of a respective monophonic audio stream, wherein at least some of the plurality of monophonic audio streams are identified as channel-based audio and wherein the others of the plurality of monophonic audio streams are identified as object-based audio, and wherein the playback location of the channel-based audio comprises speaker designations of speakers in a speaker array, and the playback location of the object-based audio comprises a location in three-dimensional space relative to a playback environment containing the speaker array; and further wherein a first metadata set is applied to one or more of the plurality of monophonic audio streams for a first condition of the playback environment, and a second metadata set is applied to the one or more plurality of monophonic audio streams for a second condition of the playback environment; and rendering the plurality of monophonic audio streams to a plurality of speaker feeds corresponding to speakers in the playback environment in accordance with the at least two metadata sets based on a condition of the playback environment.

16. The method of claim 15 wherein each metadata set includes metadata elements associated with each object-based stream, the metadata elements for each object-based stream specifying spatial parameters controlling the playback of a corresponding object-based sound, and comprising one or more of: sound position, sound width, and sound velocity; and further wherein each metadata set includes metadata elements associated with each channel-based stream, and the speaker array comprises speakers arranged in a defined surround sound configuration, and wherein the metadata elements associated with each channel-based stream comprises designations of surround-sound channels of the speakers in the speaker array in accordance with a defined surround-sound standard.

17. The method of claim 15 wherein the speaker array includes additional speakers for playback of object-based streams that are positioned in the playback environment, the method further comprising receiving set up instructions from a user based on the condition of the playback environment, and wherein the playback condition depends on variables comprising: size and shape of a room of the playback environment, occupancy, material composition, and ambient noise; the setup instructions further including at least a list of speaker designations and a mapping of channels to individual speakers of the speaker array, information regarding grouping of speakers, and a mapping based on a relative position of speakers to the playback environment.