target variable referencing a name of a particular first node representing a particular Set, the particular Set consisting of a first playback device and a second playback device in a bonded zone, and wherein causing the one or more particular playback devices to play back audio content according to the one or more playback commands comprises causing the first playback device and the second playback device to play back respective channels of the audio content in synchrony.

15. The NMD of claim **12**,

wherein the data structure further comprises one or more third nodes in a third level as parents to one or more respective child second nodes to represent Rooms in respective Areas of the hierarchy,

wherein determining that the one or more first keywords within the voice input represent respective target variables comprises determining that at least one first voice keyword within the voice input represents a target variable referencing a name of a particular third node representing an Area including a first Room and a second Room, the first Room including a first Set that consists of a first playback device and the second Room including a second Set that consists of a second playback device, and

wherein causing the one or more particular playback devices to play back audio content according to the one or more playback commands comprises causing the first playback device and the second playback device to play back the audio content in synchrony.

16. The NMD of claim **15**, wherein causing the one or more particular playback devices to play back audio content according to the one or more playback commands comprises causing the first playback device and the second playback device to form a synchrony group.

17. The NMD of claim **12**, wherein determining that one or more first keywords within the voice input represent respective target variables comprises determining that at least one first voice command within the voice input represents a target variable referencing a name of the root node, and wherein causing the one or more particular playback devices to play back audio content according to the one or more playback commands comprises causing all playback devices represented by nodes within the data structure to play back the audio content in synchrony.

18. The NMD of claim **12**, wherein determining that one or more first voice commands within the voice input represent respective target variables comprises:

searching the nodes of the data structure for nodes having assigned names that match target variables represented by the one or more first keywords; and

determining that the one or more first keywords match names assigned to the one or more particular nodes.

19. The NMD of claim **12**, wherein the functions further comprise:

when another keyword is detected in the recorded audio, processing an additional voice input within a portion of the audio recording comprising the another keyword, wherein processing the additional voice input comprises:

determining, based on the data structure representing the media playback system, that one or more third keywords within the additional voice input represent respective additional target variables indicating one or more additional nodes of the data structure, each additional target variable referencing a name of a respective node of the data structure; and

determining that one or more fourth keywords within the voice input correspond to one or more smart device commands; and

causing, via the network interface, one or more particular smart devices to carry out the one or more smart device commands, wherein the one or more particular smart devices include (a) all smart devices represented by the one or more additional nodes of the data structure and (b) all smart illumination devices represented by child nodes of the one or more additional nodes of the data structure.

20. A tangible, non-transitory computer-readable medium having instructions encoded therein that, when executed by at least one processor of a network microphone device (NMD), cause the NMD to perform functions comprising:

storing a data structure comprising nodes in a hierarchy representing a media playback system, wherein the data structure comprises (i) a root node representing the media playback system as a Home of the hierarchy, (ii) one or more first nodes in a first level, the first nodes representing respective devices of the media playback system as Sets of the hierarchy, and (iii) one or more second nodes in a second level as parents to one or more respective child first nodes to represent Sets in respective Rooms of the hierarchy, wherein the nodes in the hierarchy are assigned respective names;

recording, via a microphone array, audio into a buffer;

monitoring the recorded audio for keywords;

when a keyword is detected in the recorded audio, processing a voice input within a portion of the audio recording comprising the keyword, wherein processing the voice input comprises:

(i) determining, based on the data structure representing the media playback system, that one or more first keywords within the voice input represent respective target variables indicating one or more particular nodes of the data structure, each target variable referencing a name of a respective node of the data structure; and

(ii) determining that one or more second keywords within the voice input correspond to one or more playback commands; and

causing, via a network interface, one or more particular playback devices to play back audio content according to the one or more playback commands, wherein the one or more particular playback devices include (a) all playback devices represented by the one or more particular nodes of the data structure and (b) all playback devices represented by child nodes of the one or more particular nodes of the data structure.

* * * * *