CLAIMS

1. An audio session management method, comprising:

receiving output signals from each microphone of a plurality of microphones in an audio environment, each microphone of the plurality of microphones residing in a microphone location of the audio environment, the output signals including signals corresponding to a current utterance of a person, wherein the output signals include speech and/or non-speech audio data;

determining, based on the output signals, one or more aspects of context information relating to the person, the context information including at least one of an estimated current location of the person or an estimated current proximity of the person to one or more microphone locations;

determining a closest loudspeaker-equipped audio device that is closest to the microphone location closest to the estimated current location of the person;

selecting two or more audio devices of the audio environment based, at least in part, on the one or more aspects of the context information, the two or more audio devices each including at least one loudspeaker and wherein the two or more audio devices include the closest loudspeaker-equipped audio device;

determining one or more types of audio processing changes to apply to audio data being rendered to loudspeaker feed signals for the two or more audio devices, the audio processing changes having an effect of increasing a speech to echo ratio at the microphone closest to the estimated current location of the person, wherein the echo comprises at least some of audio outputted by the two or more audio devices, and wherein at least one of the audio processing changes for the closest audio device is different from an audio processing change for a second audio device of said at least two audio devices, and wherein the one or more types of audio processing changes cause a reduction in loudspeaker reproduction level for the closest audio device; and

causing the one or more types of audio processing changes to be applied.

2. The method of claim 1, wherein the one or more types of audio processing changes involve spectral modification.

3. The method of any one of the preceding claims, wherein the one or more types of audio processing changes cause a reduction in loudspeaker reproduction level for the loudspeakers of the second audio device of said at least two audio devices.

4. The method of any one of claims 1–3, wherein selecting two or more audio devices of the audio environment comprises selecting *N* loudspeaker-equipped audio devices of the audio environment, *N* being an integer greater than 2.

5. The method of any one of claims 1–4, wherein selecting the two or more audio devices of the audio environment is based, at least in part, on an estimated current location of the person relative to at least one of a microphone location or a loudspeaker-equipped audio device location.

6. The method of any one of claims 1–5, wherein the one or more types of audio processing changes involve changing a rendering process to warp a rendering of audio signals away from the estimated current location of the person of interest.

7. The method of any one of claims 2-6, wherein the spectral modification involves reducing a level of audio data in a frequency band between 500 Hz and 3 KHz.

8. The method of any one of claims 1–7, wherein the one or more types of audio processing changes involve inserting at least one gap into at least one selected frequency band of an audio playback signal.

9. The method of any one of claims 1–8, wherein the one or more types of audio processing changes involve dynamic range compression.

10. The method of any one of claims 1–9, wherein selecting the two or more audio devices is based, at least in part, on a signal-to-echo ratio estimation for one or more microphone locations.

11. The method of claim 10, wherein selecting the two or more audio devices is based, at least in part, on determining whether the signal-to-echo ratio estimation is less than or equal to a signal-to-echo ratio threshold.

12. The method of claim 10, wherein determining the one or more types of audio processing changes is based on an optimization of a cost function that is based, at least in part, on the signal-to-echo ratio estimation.

13. The method of any one of claims 1–12, wherein selecting the two or more audio devices is based, at least in part, on a proximity estimation.

14. The method of any one of claims 1–13, further comprising:

determining multiple current acoustic features from the output signals of each microphone;

applying a classifier to the multiple current acoustic features, wherein applying the classifier involves applying a model trained on previously-determined acoustic features derived from a plurality of previous utterances made by the person in a plurality of user zones in the environment; and

wherein determining one or more aspects of context information relating to the person involves determining, based at least in part on output from the classifier, an estimate of a user zone in which the person is currently located.

15. The method of claim 14, wherein the estimate of the user zone is determined without reference to geometric locations of the plurality of microphones.

16. The method of claim 14 or claim 15, wherein the current utterance and the previous utterances comprise wakeword utterances.

17. The method of any one of claims 1–16, further comprising selecting at least one microphone according to the one or more aspects of the context information.

18. The method of any one of claims 1–17, wherein the one or more microphones reside in multiple audio devices of the audio environment.

19. The method of any one of claims 1–18, wherein the one or more microphones reside in a single audio device of the audio environment.

20. The method of any one of claims 1–19, wherein at least one of the one or more microphone locations corresponds to multiple microphones of a single audio device.

21. An apparatus configured to perform the method of any one of claims 1–20.

22. A system configured to perform the method of any one of claims 1–20.

23. One or more non-transitory media having software stored thereon, the software including instructions for controlling one or more devices to perform the method of any one of claims 1–20.