SPATIAL AUDIO RENDERING ADAPTIVE TO SIGNAL LEVEL AND LOUDSPEAKER PLAYBACK LIMIT THRESHOLDS

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of priority to U.S. Provisional Patent Application No. 63/392,794 filed July 27, 2022, U.S. Provisional Patent Application No. 63/413,923 filed October 6, 2022 and U.S. Provisional Patent Application No. 63/505,652 filed June 1, 2023.

TECHNICAL FIELD

The disclosure pertains to systems and methods for rendering audio for playback by a set of speakers.

BACKGROUND

Audio devices, including but not limited to smart audio devices, have been widely deployed and are becoming common features of many homes. Although existing systems and methods for controlling audio devices provide benefits, improved systems and methods would be desirable. International Patent Publication No. WO 2022/064905 describes an information processing device, information processing method, and program.

NOTATION AND NOMENCLATURE

Throughout this disclosure, including in the claims, "speaker" and "loudspeaker" are used synonymously to denote any sound-emitting transducer (or set of transducers) driven by a single speaker feed. A typical set of headphones includes two speakers.

Throughout this disclosure, including in the claims, the expression performing an operation "on" a signal or data (e.g., filtering, scaling, transforming, or applying gain to, the signal or data) is used in a broad sense to denote performing the operation directly on the signal or data, or on a processed version of the signal or data (e.g., on a version of the signal that has undergone preliminary filtering or pre-processing prior to performance of the operation thereon).

Throughout this disclosure including in the claims, the expression "system" is used in a broad sense to denote a device, system, or subsystem. For example, a subsystem that implements a decoder may be referred to as a decoder system, and a system including such a

CLAIMS

1. An audio processing method, comprising:

receiving, by a control system and via an interface system, audio data, the audio data including one or more audio signals and associated spatial data, the spatial data indicating an intended perceived spatial position corresponding to an audio signal (1205);

rendering, by the control system, the audio data for reproduction via a set of two or more loudspeakers of an environment, to produce loudspeaker signals, wherein:

rendering each of the one or more audio signals included in the audio data involves a mapping for each audio signal to the loudspeaker signals, the mapping being a time- and frequency-varying mapping;

the mapping for each audio signal is computed as a function of an audio signal's intended perceived spatial position, physical positions associated with the loudspeakers and a time- and frequency-varying representation of loudspeaker signal level relative to a maximum playback limit of each loudspeaker;

each mapping is computed to approximately achieve the intended perceived spatial position of an audio signal when the loudspeaker signals are played back over the two or more corresponding loudspeakers located at associated loudspeaker positions;

a representation of loudspeaker signal level relative to a maximum playback limit is computed for each audio signal as a function of one or more of the audio signals and their perceived spatial positions; and

the mapping of an audio signal into a particular loudspeaker signal is reduced as the representation of loudspeaker signal level relative to a maximum playback limit increases above a threshold, while the mapping is increased into one or more other loudspeakers for which the representations of signal level relative to the maximum playback limits of one or more other loudspeakers are less than a threshold (1210); and

providing, via the interface system, the loudspeaker signals to at least two loudspeakers of the set of loudspeakers of the environment (1215);

wherein the mapping from audio signal to loudspeaker signals is determined by:

querying a data structure indexed by the intended perceived spatial position and a level estimate for each audio signal, or

interpolating from a set of pre-computed speaker mappings, the set being indexed by the intended perceived spatial position and a level estimate for each audio signal or by an intended level estimate for each audio signal.

2. The method of claim 1, wherein the mapping is computed over an entire audible frequency range or over a subset of an audible frequency range.

3. The method of claim 1 or 2, wherein computing the mapping involves optimizing a cost function including a first term that models how closely the intended perceived spatial position is achieved as a function of mapping an audio signal into loudspeaker signals, and a second term that assigns a cost to activating each of the loudspeakers, and optionally wherein the cost of activating each loudspeaker is based, at least in part, on a function of the representation of loudspeaker signal level relative to the maximum playback limit.

4. The method of any one of claims 1–3, wherein the representation of loudspeaker signal level relative to the maximum playback limit corresponds to one or more of a digital signal level, a limiter gain, or an acoustic signal level or wherein the representation of loudspeaker signal level relative to the maximum playback limit is computed as a difference between the level estimate for each audio signal and playback limit thresholds for each loudspeaker.

5. The method of any one of claims 1-4, wherein the level estimate for each audio signal is based, at least in part, on a zone-based rendering of all the audio signals.

6. The method of any one of claims 1-5, wherein the level estimate for each audio signal is based, at least in part, on previously-computed loudspeaker signals, and optionally wherein the level estimate for each audio signal is further dependent upon a participation of each loudspeaker in a plurality of spatial zones.

7. The method of any one of claims 1–6, further comprising smoothing the level estimate for each audio signal across time, across frequency, or across both time and frequency.

8. The method of any one of claims 1-7, wherein the level estimate for each audio signal is represented as a broadband gain multiplied with a spectral shape, and optionally wherein the spectral shape is selected from a plurality of spectral shapes, each spectral shape of the plurality of spectral shapes corresponding to a content type.

9. The method of any one of claims 1–8, wherein reducing a mapping into one loudspeaker and increasing a mapping into another loudspeaker occurs gradually as the representation of signal level relative to a maximum playback level increases above a threshold.

10. The method of any one of claims 1–9, further comprising controlling a degree of reduction of mapping into one loudspeaker and an increase of mapping into another loudspeaker according to one or more of an audio format, a codec, or metadata or according to a knee parameter.

11. The method of any one of claims 1–10, wherein the intended perceived spatial position corresponds with a channel of a channel-based audio format, corresponds with metadata, or corresponds with both the channel and the metadata.

12. The method of any one of claims 1–11, wherein approximately achieving the intended perceived spatial position of an associated audio signal involves minimizing a difference between a perceived spatial position and the intended perceived spatial position, given available loudspeakers and associated loudspeaker positions.

13. The method of any one of claims 1–12, wherein approximately achieving the intended perceived spatial position of an associated audio signal involves minimizing a cost function.

14. A system configured to perform the method of any one of claims 1–13.

15. One or more non-transitory media having instructions stored thereon for controlling one or more devices to perform the method of any one of claims 1–13.