May 2023

DYNAMIC TORSO REFLECTION FILTERING FOR INTERACTIVE BINAURAL SPATIAL AUDIO BASED ON BIOLOGICALLY CONSTRAINED IMU DRIFT COMPENSATION

Philip Robinson
Meta Platforms Technologies, LLC

William Owen Brimijoin II
Meta Platforms Technologies, LLC

Antje Ihlefeld
Meta Platforms Technologies, LLC

Attorney Docket No. 31718-55106/US

DYNAMIC TORSO REFLECTION FILTERING FOR INTERACTIVE BINAURAL SPATIAL AUDIO BASED ON BIOLOGICALLY CONSTRAINED IMU DRIFT COMPENSATION

Inventors:
Philip Robinson
William Owen Brimijoin II
Antje Ihlefeld

FIELD OF THE INVENTION

[0001] The present disclosure generally relates to spatializing audio content, and specifically relates to dynamic torso reflection filtering for interactive binaural spatial audio based on biologically constrained IMU drift compensation.

BACKGROUND

[0002] Conventional audio systems use frequency-domain multiplication to process head-related transfer functions (HRTFs) for the generation of spatialized audio content. However, conventional HRTFs typically fail to account for changes in orientation between the head and the torso. Moreover, time-domain convolution of HRTFs require significant computational resources, power, and memory. This makes these devices not ideal for use in resource-constrained devices, such as a headset, with limited compute resources, limited memory, limited power, and small form factors.

DETAILED DESCRIPTION

[0003] Described herein is an audio system that uses dynamic torso reflection filtering for interactive binaural spatial audio based on biologically constrained IMU drift compensation. The audio system may be integrated into a device that includes and/or is communicatively coupled to a device (e.g., headset) that includes a drift compensation system for an inertial measurement unit (IMU). In some embodiments, the device may be a wearable device (e.g., a headset). The head and torso information are provided to an audio time and level difference renderer (TLDR) for generating spatialized audio content.

[0004] The drift compensation system comprises the IMU integrated into a headset worn by a user. The IMU of the headset tracks the position of the headset. The IMU may track the position of the headset in terms of yaw, roll, and pitch measurements. The measurements may be in terms of an angle measurement (e.g., in degrees or in radians) and/or in terms of a rate (e.g., in degrees per second, in radians per second, etc.). When the user is wearing the headset, the yaw measurement relates to the amount the user turns their head to the left or to the right (i.e., a rotation about the yaw axis). The roll measurement relates to the amount the user tilts their head to the left and to the right (i.e., a rotation about the roll axis). The pitch measurement relates to the amount the user tilts their head up and down (i.e., a rotation about the pitch axis). The IMU tracks the current yaw, roll, and pitch measurement relative to the last yaw, roll, and pitch measurement (i.e., the IMU is measuring changes in yaw, roll, and pitch measurements relative to the last yaw, roll, and pitch measurements). The IMU suffers from drift (i.e., an increasing difference in the actual position of the headset and the measured position of the headset).

[0005] The head of the user is biologically-constrained in its positioning relative to a torso of the user. A biological constraint of the user is a limitation or restriction that is based in part on the user's biology (e.g., how the head moves via the neck relative to the torso). For instance, the head of the user may only turn from the far left to the far right a maximum of 180 degrees (i.e., maximum yaw measurement). The head of the user may only tilt a maximum of 90 degrees (i.e., maximum roll measurement or maximum pitch measurement). Over time, the head of the user generally remains at some nominal position (e.g., squared to the shoulders looking straight ahead) relative to a torso of the user. A nominal position of the user may be established based on the positioning of the head of the user as the head of the user remains still over a predetermined time period. The nominal position may differ from user to user. For example, a nominal position of one user may have the head facing towards the front (i.e., squared with shoulders) with little to no tilt, and a nominal position for a different user may also have the head at a different position. A nominal position may be associated with a yaw measurement, a roll measurement, and a pitch measurement of zero degrees. After the neck of the user turns the head of the user to the left or to the right (i.e. introduces some yaw), the neck also tilts the head of the user (i.e., introduces some roll).

[0006] Utilizing the biological constraints and a set of position parameters, a drift correction component may be determined and may be used to compensate for drift present in the IMU over time. The set of position parameters may include, e.g., a yaw measurement, a roll measurement, a pitch measurement for the headset. In some embodiments, the set of position parameters may be measured relative to the torso. The set of position parameters may be determined by the IMU of the headset. In some embodiments, other secondary devices (e.g., other headsets, other mobile devices, etc.) may be used to determine some or all of the position parameters. The position parameters may include pose information about the user. Pose information describes the positioning of the head of the user. The pose information may include additional position measurements (i.e., a second yaw, roll, and pitch measurement), an indication the head positioning of the user (e.g., relative to the torso) has undergone a change, a notification of the type of change (i.e., a change in yaw, a change in roll, a change in pitch, or some combination thereof) the head positioning has undergone, or some combination thereof.

[0007] In one embodiment, a secondary device may determine pose information based on changes in sound detected by an audio system of the secondary device. In another embodiment, the headset may determine pose information based on changes in sound detected by an audio system of the headset. In one embodiment, the secondary device may determine pose information based on an analysis of one or more images of a head (including the headset) and a torso of the user captured by the secondary device. In another embodiment, the headset may determine pose information based on an analysis of one or more images of a head (including the headset) and a torso of the user provided by the secondary device, in accordance with one or more privacy settings of the user. The secondary device may determine pose information by utilizing an IMU integrated into the secondary device. The secondary device may provide, in accordance with privacy settings of the user, the pose information to the IMU to be included in the set of position parameters.

[0008] The set of position parameters and the biological constraints of the positioning of the head of the user are utilized (e.g., by the controller) to determine the drift correction component. The drift correction component describes a rate of correction. The drift correction component is applied to subsequent measurements (e.g., yaw measurements) of the IMU at the rate of correction. The drift correction component forces an estimated nominal position vector (i.e., a vector based on the nominal positioning of the head of the user) to a pointing vector (i.e., a vector based on the current measured positioning of the head of the user) during the subsequent measurements of the IMU. The drift correction component may include a yaw drift rate measured in degrees over time. The yaw drift rate may be inversely proportional to the amount of roll measured. In some embodiments, the drift correction component may further include a roll drift rate measured in degrees over time. In some embodiments, the drift correction component effectively adjusts for any drift error present in the subsequent yaw measurements, thereby compensating for drift present in the IMU over time.

[0009] The controller uses a head orientation and a torso orientation received from the drift compensation system to determine filters for presenting spatial audio. The controller is configured to use an audio TLDR and the received head orientation and torso orientation to determine the filters. The audio TLDR includes a cascaded series of infinite impulse response (IIR) filters and a pair of delays for the head orientation, and a separate set of cascaded IRR filters for the torso orientation. The controller applies the audio TLDR to an audio signal received at a single channel to generate spatialized audio content corresponding to multiple channels (e.g., left and right channel audio signals) for a given head orientation and torso orientation. The audio TLDR may have a set of configured monaural static filters (with 0, 1,2,…number of monaural static filters in the set) and a set of configured monaural dynamic filters (with 0, 1, 2,… number of monaural dynamic filters in the set) connected to the set of monaural static filters. The monaural static and dynamic filters are connected (i.e., receive input audio signal and generate an output audio signal) through the single channel. In some embodiments, there may also be static binaural filters that may perform individualized left/right speaker equalization. The audio TLDR also has a set of configured binaural dynamic filters (with 1, 2, …, number of pairs of binaural dynamic filters in the set) that are connected (i.e., receive an input audio signal and generate an output audio signal) through each channel of multiple audio channels (such as a connected left channel and a connected right channel). In addition, the audio TLDR may have a configured delay between the multiple audio channels. The head orientation and the torso orientation

[0010] The controller is configured to apply the head orientation information to a first set of dynamic binaural filters to generate intermediate audio content. Intermediate audio content is refers to the audio signals for the left channel, right channel, or both, that have been processed for head orientation, but have not yet been processed using torso orientation. Figure 1 is an example, audio TLDR that determines filters for spatialized audio using both a head orientation and torso orientation. As shown in Figure 1, a input mono channel is split into a left channel and right channel. The audio TLDR uses the head orientation as inputs to delay and a series of filters to form intermediate audio content that is ultimately input into the peak/notch filters with the torso orientation. The controller applies the torso orientation information and the intermediate audio content to a second set of dynamic binaural filters of the TLDR to generate spatialized audio content for the left and right channels. Note Figure 1 illustrates a specific set of filters and delays. Not in other embodiments, the number and/or type of filters may vary for generating intermediate audio content and/or for generating spatialized audio content. The audio controller instructs a speaker array of the headset to present the generated spatialized audio content using the left channel and right channel information to the user.

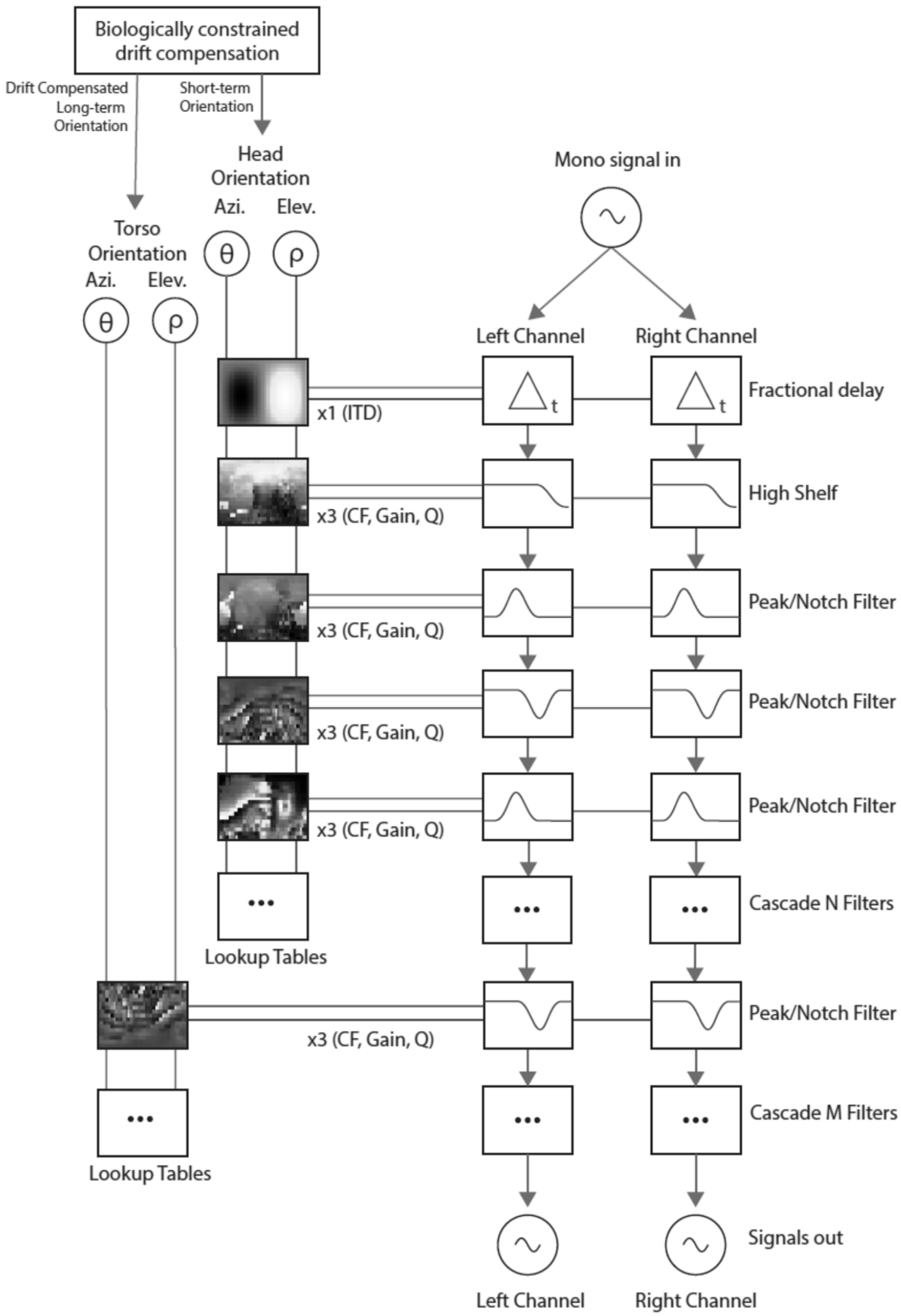

Figure 1

[0011] Embodiments of the invention may include or be implemented in conjunction with an artificial reality system. Artificial reality is a form of reality that has been adjusted in some manner before presentation to a user, which may include, e.g., a virtual reality (VR), an augmented reality (AR), a mixed reality (MR), a hybrid reality, or some combination and/or derivatives thereof. Artificial reality content may include completely generated content or generated content combined with captured (e.g., real-world) content. The artificial reality content may include video, audio, haptic feedback, or some combination thereof, any of which may be presented in a single channel or in multiple channels (such as stereo video that produces a three-dimensional effect to the viewer). Additionally, in some embodiments, artificial reality may also be associated with applications, products, accessories, services, or some combination thereof, that are used to create content in an artificial reality and/or are otherwise used in an artificial reality. The artificial reality system that provides the artificial reality content may be implemented on various platforms, including a wearable device (e.g., headset) connected to a host computer system, a standalone wearable device (e.g., headset), a mobile device or computing system, or any other hardware platform capable of providing artificial reality content to one or more viewers.

Additional Configuration Information

[0012] The foregoing description of the embodiments has been presented for illustration; it is not intended to be exhaustive or to limit the patent rights to the precise forms disclosed. Persons skilled in the relevant art can appreciate that many modifications and variations are possible considering the above disclosure.

[0013] Some portions of this description describe the embodiments in terms of algorithms and symbolic representations of operations on information. These algorithmic descriptions and representations are commonly used by those skilled in the data processing arts to convey the substance of their work effectively to others skilled in the art. These operations, while described functionally, computationally, or logically, are understood to be implemented by computer programs or equivalent electrical circuits, microcode, or the like. Furthermore, it has also proven convenient at times, to refer to these arrangements of operations as modules, without loss of generality. The described operations and their associated modules may be embodied in software, firmware, hardware, or any combinations thereof.

[0014] Any of the steps, operations, or processes described herein may be performed or implemented with one or more hardware or software modules, alone or in combination with other devices. In one embodiment, a software module is implemented with a computer program product comprising a computer-readable medium containing computer program code, which can be executed by a computer processor for performing any or all the steps, operations, or processes described.

[0015] Embodiments may also relate to an apparatus for performing the operations herein. This apparatus may be specially constructed for the required purposes, and/or it may comprise a general-purpose computing device selectively activated or reconfigured by a computer program stored in the computer. Such a computer program may be stored in a non-transitory, tangible computer readable storage medium, or any type of media suitable for storing electronic instructions, which may be coupled to a computer system bus. Furthermore, any computing systems referred to in the specification may include a single processor or may be architectures employing multiple processor designs for increased computing capability.

[0016] Embodiments may also relate to a product that is produced by a computing process described herein. Such a product may comprise information resulting from a computing process, where the information is stored on a non-transitory, tangible computer readable storage medium and may include any embodiment of a computer program product or other data combination described herein.

[0017] Finally, the language used in the specification has been principally selected for readability and instructional purposes, and it may not have been selected to delineate or circumscribe the patent rights. It is therefore intended that the scope of the patent rights be limited not by this detailed description, but rather by any claims that issue on an application based hereon. Accordingly, the disclosure of the embodiments is intended to be illustrative, but not limiting, of the scope of the patent rights, which is set forth in the following claims.

What is claimed is:

1. A method comprising:

receiving a head orientation and a torso orientation from an inertial measurement unit with biologically constrained drift compensation;

applying the head orientation information to a first set of dynamic binaural filters to generate intermediate audio content, wherein the first set of binaural dynamic filters are coupled via multiple channels for receiving input audio signals that are spilt from a single channel, and the multiple channels comprise a left channel and a right channel; and

applying the torso orientation information and the intermediate audio content to a second set of dynamic binaural filters to generate spatialized audio content for each channel of the multiple channels; and

presenting the generated spatialized audio content at multiple channels to a user.