

JOINT STEREO CODING IN COMPLEX-VALUED FILTER BANK DOMAIN**CROSS-REFERENCE TO RELATED APPLICATIONS**

[0001] This application claims priority to U.S. Provisional Application No. 63/559,764 filed February 29, 2024, and U.S. Provisional Application No. 63/491,840, filed March 23, 2023, the entire contents of which is hereby incorporated by reference.

TECHNICAL FIELD

[0002] This application relates generally to audio and speech coding, and more specifically to mid-side stereo coding.

BACKGROUND

[0003] Unless otherwise indicated herein, the materials described in this section are not prior art to the claims in this application and are not admitted as prior art by inclusion in this section.

[0004] Joint coding of left (L) and right (R) channels of a stereo signal enables more efficient coding compared to independent coding of L and R channels. One approach for joint stereo coding is so-called mid/side (M/S) coding. In mid-side coding, a mid (M) signal is formed as a combination of the L and R channel signals, while a side (S) signal may be formed as a difference between the L and R channel signals. For example, the M signal may be of the form $M = (L + R)/2$; while the S signal may be of the form $S = (L - R)/2$. In the case of M/S coding, the M and S signals are coded instead of the L and R signals.

[0005] In some implementations, M/S stereo coding can be implemented in a time and frequency-variant manner. For example, a stereo encoder can apply L/R for encoding some frequency bands of the stereo signal, whereas M/S coding can be used for encoding other frequency bands of the stereo signal (frequency variant). Moreover, some encoders can switch over time between L/R and M/S coding (time-variant) methods.

[0006] It is with respect to these and other considerations that the disclosure made herein is presented.

BRIEF SUMMARY OF THE DISCLOSURE

[0007] Techniques are described for processing audio signals. Various embodiments described herein provide systems, methods and/or devices with extensions to mid-side stereo coding applied in a complex-valued filter domain. For example, embodiments provide an adjusted

phase alignment between a left audio channel and a right audio channel prior to the mid-side conversion in combination with real-valued prediction of the side signal from the mid signal in the encoder. Additionally, embodiments described herein provide a novel method of phase alignment in the complex domain and transmitting of the complex audio data.

[0008] The inter-channel phase difference which may be transmitted in a bitstream as metadata is used to reconstruct the original inter-channel phase relation at the decoder. As the signal is operated in the complex domain, the phase alignment operation can be applied without loss of information or risk of degradation, which is in general not the case when only real-valued data is encoded. Further, different processing blocks (e.g., phase alignment, mid-side conversion, and side signal prediction blocks) are enabled or disabled based on a level-dependent psychoacoustic model and, in some instances, based on parameter side rate cost.

[0009] Embodiments described herein improve spatial noise shaping (by preventing spatial unmasking) and improve energy compaction compared to known mid-side coding, particularly for signals with systematic inter-channel level differences and inter-channel phase shifts or time delays. Such characteristics are common for binaural signals which have been generated by filtering audio objects or channels with head-related transfer functions. Using mid-side coding extensions described herein, audio quality is improved at a given bitrate, or the bitrate may be lowered at a certain quality level.

[0010] In some aspects, embodiments described herein include an encoder receiving a left-channel (e.g., a left input signal) and a right-channel (e.g., a right input signal) that are binaural channels. Complex filter bank analysis is performed on the left-channel and the right-channel to convert the left-channel and the right-channel to a complex-valued filter bank domain. Converting the left-channel and the right-channel to the complex-valued filter bank domain prepares the left-channel and the right-channel for rendering by a head-tracking device, such as being processed by a head related transfer function (HRTF). The complex-valued filter bank domain signals may be, for example, one or more frequency bands.

[0011] In some aspects, stereo analysis is performed on the complex-valued filter bank domain left-channel and the complex-valued filter bank domain right-channel. The stereo analysis may identify energies of the complex-valued filter bank domain left-channel and energies of the complex-valued filter bank domain right-channel. Additionally, the stereo analysis may identify energies of a potential Mid signal and a potential Side signal. The Mid signal represents the sum of the left-channel and the right-channel. The Side signal represents the difference between the

left-channel and the right-channel. In some instances, the energy of a potential residual signal is determined.

[0012] In some aspects, the stereo analysis also generates stereo metadata based on the left-channel and the right-channel. For example, a covariance of the complex-valued filter bank domain left-channel and the complex-valued filter bank domain right-channel may be calculated. An inter-channel phase difference of the left-channel and the right-channel is determined based on the covariance. Additionally, a real-valued prediction coefficient (e.g., a side prediction coefficient) is calculated based on the energies of the complex-valued filter bank domain left-channel and the complex-valued filter bank domain right-channel, as well as the energies of the potential Mid signal and the potential Side signal.

[0013] In some aspects, a single stereo coding mode of a plurality of stereo coding modes is selected for signaling and encoding the left-channel and the right-channel. For example, for each possible stereo coding mode, a bit cost of signaling the left-channel and the right-channel in that stereo coding mode is determined. The bit cost is based on the energies of the complex-valued filter bank domain left-channel, the complex-valued filter bank domain right-channel, the possible Mid signal, the possible Side signal, and the possible residual signal. In some instances ratios of the energies of signals involved in each stereo coding mode are determined. The stereo coding mode is selected based on the bit cost. In some instances, the stereo coding mode with the lowest bit cost is selected.

[0014] In some aspects, the left-channel and the right-channel are processed according to the selected stereo coding mode. As one example, when the stereo coding mode is a separate coding mode, a stereo processor performs signaling of the left-channel and the right-channel (or the complex-valued filter bank domain left-channel and the complex-valued filter bank domain right-channel) directly without converting the signals to the Mid signal and the Side signal.

[0015] In another example, the selected stereo mode is a basic Mid/Side stereo mode. In this instance, the left-channel and the right-channel are converted to the Mid signal and the Side signal. The stereo processor performs signaling of the Mid signal and the Side signal.

[0016] In another example, the selected stereo mode is a Mid/Side stereo mode with adjusted phase alignment. In this instance, the left-channel and the right-channel are phase-aligned based on the inter-channel phase difference. After aligning, the left-channel and the right-channel are converted to the Mid signal and the Side signal. The stereo processor performs signaling of the

Mid signal and the Side signal, and the phase difference is encoded alongside the Mid signal and the side signal.

[0017] In another example, the selected stereo mode is a Mid/Side stereo mode with side prediction. In this instance, the left-channel and the right-channel are converted to the Mid signal and the Side signal. A residual signal is generated based on the Side signal and a prediction coefficient. The stereo processor performs signaling of the Mid signal and the residual signal, and the prediction coefficient is encoded alongside the Mid signal and the residual signal.

[0018] In another example, the selected stereo mode is a Mid/Side stereo mode with both adjusted phase alignment and side prediction. In this instance, the left-channel and the right-channel are phase-aligned based on the inter-channel phase difference. After aligning, the left-channel and the right-channel are converted to the Mid signal and the Side signal. A residual signal is generated based on the Side signal and a prediction coefficient. The stereo processor performs signaling of the Mid signal and the residual signal, and the prediction coefficient and phase difference are encoded alongside the Mid signal and the residual signal.

[0019] One example provides a method for encoding a stereo audio signal in a bitstream. The method includes passing a left-channel and a right-channel of the stereo audio signal to a complex-valued filter bank analysis block to responsively generate one or more frequency bands and calculating, for each of the one or more frequency bands, an energy of the left-channel, an energy of the right-channel, and a covariance of the left-channel and the right-channel. The method includes selecting, based on the calculated energy of the left-channel, the calculated energy of the right-channel, and the calculated covariance of the left-channel and the right-channel, a stereo coding mode in which to encode the left-channel and the right-channel. When the stereo coding mode is an extended Mid/Side coding mode, the method includes computing a phase difference between the left-channel and the right-channel, adjusting phase alignment between the left-channel and the right-channel based on the computed phase difference to generate an aligned left-channel and an aligned right-channel, transforming the aligned left-channel and the aligned right-channel to a Mid signal and a Side signal, generating a residual signal based on side prediction data and the Side signal; encoding the Mid signal, the residual signal, the phase difference, and the side prediction data in the bitstream, and outputting the bitstream for the selected stereo coding mode.

[0020] Another example provides an apparatus for encoding a stereo audio signal in a bitstream, the apparatus including an electronic processor. The electronic processor is configured to pass a

left-channel and a right-channel of the stereo audio signal to a complex-valued filter bank analysis block to responsively generate one or more frequency bands, calculate, for each of the one or more frequency bands, an energy of the left-channel, an energy of the right-channel, and a covariance of the left-channel and the right-channel and select, based on the calculated energy of the left-channel, the calculated energy of the right-channel, and the calculated covariance of the left-channel and the right-channel, a stereo coding mode in which to encode the left-channel and the right-channel. When the stereo coding mode is an extended Mid/Side coding mode, the electronic processor is configured to compute a phase difference between the left-channel and the right-channel, adjust phase alignment between the left-channel and the right-channel based on the computed phase difference to generate an aligned left-channel and an aligned right-channel, transform the aligned left-channel and the aligned right-channel to a Mid signal and a Side signal, generate a residual signal based on side prediction data and the Side signal, encode the Mid signal, the residual signal, the phase difference, and the side prediction data in the bitstream, and output the bitstream for the selected stereo coding mode.

[0021] Another example provides a method for decoding a stereo audio signal. The method includes receiving an encoded bitstream, decoding, from the bitstream, a replicated Mid signal, a replicated residual signal, and replicated stereo metadata, wherein the replicated stereo metadata includes a phase difference and side prediction data, converting the replicated Mid signal and the replicated residual signal to a replicated left channel and a replicated right channel using the stereo metadata, and passing the replicated left channel and the replicated right channel to a filter bank analysis block to recreate an original left channel and an original right channel.

[0022] Another example provides an apparatus for decoding a stereo audio signal, the apparatus including an electronic processor. The electronic processor is configured to receive an encoded bitstream, decode, from the bitstream, a replicated Mid signal, a replicated residual signal, and replicated stereo metadata, wherein the replicated stereo metadata includes a phase difference and side prediction data, convert the replicated Mid signal and the replicated residual signal to a replicated left channel and a replicated right channel using the replicated stereo metadata, and pass the replicated left channel and the replicated right channel to a filter bank analysis block to recreate an original left channel and an original right channel.

[0023] In this manner, various aspects of the present disclosure provide for processing of stereo audio signals, and effect improvements in at least the technical fields of audio encoding, audio decoding, virtual reality, and the like.

[0024] The embodiments described herein may be generally described as techniques, where the term “technique” may refer to system(s), device(s), method(s), computer-readable instruction(s), module(s), component(s), hardware logic, and/or operation(s) as suggested by the context as applied herein.

[0025] Features and technical benefits other than those explicitly described above will be apparent from a reading of the following Detailed Description and a review of the associated drawings. This Summary is provided to introduce a selection of techniques in a simplified form, and not intended to identify key or essential features of the claimed subject matter, which are defined by the appended claims.

DESCRIPTION OF THE DRAWINGS

[0026] These and other more detailed and specific features of various embodiments are more fully disclosed in the following description, reference being had to the accompanying drawings, in which:

[0027] FIG. 1 illustrates a block diagram of an example audio coding system in which various aspects of the present invention may be incorporated.

[0028] FIG. 2 illustrates a block diagram of an example encoder.

[0029] FIG. 3 illustrates a block diagram of an example stereo processing process.

[0030] FIG. 4A illustrates an example joint stereo process conversion.

[0031] FIG. 4B illustrates a table that provides each of the joint stereo coding types and their associated bitstream syntax elements.

[0032] FIG. 5A illustrates a block diagram of an example mode-based stereo processing unit, such as the mode-based stereo processing unit of FIG. 3.

[0033] FIG. 5B illustrates a table that provides an example bitstream syntax.

[0034] FIG. 6A illustrates a graph illustrating a stereo metadata rate for an extended Mid/Side coding mode per audio frame, in accordance with various aspects of the present disclosure.

[0035] FIGS. 6B-6D illustrate examples of pseudocode.

[0036] FIGS. 7A-7B illustrate a block diagram of various example methods for encoding stereo signals, which may be performed by the encoder of FIG. 2, in accordance with various aspects of the present disclosure.

[0037] FIG. 8A illustrates a block diagram of an example decoder.

[0038] FIG. 8B illustrates an example of pseudocode.

[0039] FIG. 9 illustrates a block diagram of various example methods for decoding stereo signals, which may be performed by the decoder of FIG. 8A, in accordance with various aspects of the present disclosure.

[0040] FIG. 10 illustrates a graph of a PEAQ evaluation for twelve audio items.

[0041] FIG. 11A illustrates a schematic block diagram of an example device architecture that may be used to implement various aspects of the present disclosure.

[0042] FIG. 11B illustrates a schematic block diagram of an example CPU implemented in the device architecture of FIG. 11A that may be used to implement various aspects of the present disclosure.

DETAILED DESCRIPTION

[0043] In the following description, numerous details are set forth, such as audio device configurations, timings, operations, and the like, in order to provide an understanding of one or more aspects of the present disclosure. It will be readily apparent to one skilled in the art that these specific details are merely examples and not intended to limit the scope of this application.

[0044] As used herein, the term “includes” and its variants are to be read as open-ended terms that mean “includes, but is not limited to.” The term “or” is to be read as “and/or” unless the context clearly indicates otherwise. The term “based on” is to be read as “based at least in part on.” The term “one example implementation” and “an example implementation” are to be read as “at least one example implementation.” The term “another implementation” is to be read as “at least one other implementation.” The terms “determined,” “determines,” or “determining” are to be read as obtaining, receiving, computing, calculating, estimating, predicting, or deriving. In addition, in the following description and claims, unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skills in the art to which this disclosure belongs.

[0045] Various Acronyms may appear throughout this disclosure and in the associated claims and/or drawings are listed below. Other commonly used acronyms and terms of art may be excluded from this list in the interest of brevity. Thus, a short list of acronyms is provided below as an easy reference for the reader.

ERB – Equivalent Rectangular Bandwidth
PEAQ – Perceptual Evaluation of Audio Quality
ODG – Objective Difference Grade
IVAS – Immersive Voice and Audio Services
MDFT – Modified Discrete Fourier Transform
MDCT – Modified Discrete Cosine Transform
HRTF – Head Related Transfer Function

[0046] FIG. 1 illustrates a block diagram of an audio coding system in which various aspects of the present invention may be incorporated. The example audio coding system 100 includes an encoder 110 and a decoder 120. The input of the encoder 110 corresponds to a first signal path 105, while an output of the encoder 100 corresponds to a second signal path 115. The input of the decoder 120 corresponds to the second signal path 115, while an output of the decoder corresponds to a third signal path 125.

[0047] The encoder 110 is configured to receive, from the first signal path 105, one or more streams of audio information representing one or more channels of audio signals. The encoder 110 is further configured to process the received streams of audio information and generate an encoded signal, which may be output to the second signal path 115. At the second signal path 115, the encoded signal may be stored (e.g., captured, buffered and/or recorded), or transmitted (e.g., via a wired or wireless communication medium). The decoder 120 is configured to receive the encoded signal from the second signal path 115. The decoder 120 is further configured to process the received encoded signal and generate a decoded signal, which may be output to the third signal path 125. The decoded signal that is generated by the decoder 120 corresponds to a replica of the audio information previously received by the encoder 110 from the first signal path 105. At the third signal path, the decoded signal may be stored (e.g., captured and/or recorded), transmitted (e.g., via a wireless or wired electronic communication medium), or output to a listening device (e.g., an audio processing device such as a receiver, speaker, soundbar, etc.). The audio coding system 100 may be an audio system capable of implementing an audio codec standard, such as the Immersive Voice and Audio Services (IVAS) standard. In such an instance, the encoded signal at the second signal path 115 may correspond to an IVAS bitstream.

[0048] In various examples described herein, the terms “replica” and “replica signal” are not intended to mean that the streams of audio information are “identical”. Instead, the term “replica” may indicate that the streams of audio information are approximately the same as the original audio information. For example, when the encoder 110 uses a lossless encoding

technique to generate the encoded signal, the decoder 120 can in principle recover a lossless version that is approximately the same as the original audio information from the streams. However, in examples where the encoder 110 uses a lossy encoding technique, such as for perceptual coding, the content of the recovered replica signal is generally not identical to the content of the original stream but it may be perceptually indistinguishable from the original content. Thus, the terms “replica” and “replica signal” are intended to cover both lossless and lossy encoding techniques as used herein.

[0049] FIG. 2 illustrates an example of the encoder 110. The encoder 110 includes a complex filter bank analysis block 205, a stereo processing block 210, an encoding block 215, and a bitstream writing block 220. The complex filter bank analysis block 205 is configured to receive left and right audio channels (for example, from one or more microphones). The left and right audio channels may be binaural channels that have a small delay between the channels and a correlation between the channels. The left and right audio channels may be processed by a head related transfer function (HRTF). In some instances, the left and right audio channels are Ambisonics signals. The complex filter bank analysis 205 is further configured to process the left and right audio channels to generate complex-value filter bank domain left signals and complex-valued filter bank domain right signals. The complex filter bank analysis block 205 is configured to output the complex-valued filter bank domain left signals and the complex-valued filter bank domain right signals to the stereo processing block 210. The complex-valued filter bank domain signals may include additional metadata for processing by a head tracking device. In some instances, the complex-valued filter bank domain signals are one or more frequency bands associated with the left-channel and the right-channel. In some examples, the complex-filter bank analysis may be a perfect reconstruction such as for a Modified Discrete Fourier Transform (MDFT), a Modified Discrete Cosine/Sine Transform (MDCT), or a near perfect reconstruction such as a complex modulated filter bank.

[0050] The stereo processing block 210 is configured to receive the complex-valued filter bank domain left signals and the complex-valued filter bank domain right signals from the complex filter bank analysis 205. The stereo processing block 210 is configured to perform stereo processing analysis by assembling filter bank frequency bins that are associated to a frame to frequency bands according to an auditory frequency scale, such as an Equivalent Rectangular Bandwidth (ERB) scale or a Bark scale. For every frequency band, the energy of the left and right channel and the covariance from both channels are calculated. Additionally, the energies of the Mid and Side signals are computed per frequency band. A correlation coefficient is computed per band from the covariance. When the correlation coefficient is higher than a

threshold, a phase difference between the left and right channel may be computed and used to adjust the phase alignment. The covariance may be updated based on the quantized phase difference.

[0051] Additionally, the stereo processing block 210 may compute a real-valued prediction coefficient, which can be used to remove redundancy in the Side signal with respect to the Mid signal. The adjusted phase alignment may lead to an increase of the Mid signal energy and a decrease of the Side signal energy. Additionally, the adjusted phase alignment may ensure that the Side signal energy is always lower than the Mid signal energy, which may lead to situations where the left and right signal parts cancel out in the Mid signal for certain frequencies. The side signal residual energy may be computed based on the quantized prediction coefficient. Further details regarding operation of the stereo processing block 210 are provided below.

[0052] The computed signal energies, as well as computed stereo processing metadata, are output by the stereo processing block 210 to encoding block 215. The encoding block 215 is configured to receive the computed signal energies and the computed stereo processing metadata from the stereo processing block 210. The encoding block 215 is further configured to encode the computed signal energies and the stereo processing metadata as an encoded signal. The encoded computed signal energies and the encoded stereo processing metadata are output by the encoding block 215 to the bitstream writing block 220. The bitstream writing block 220 is configured to receive the encoded computed signal energies and the encoded stereo processing metadata. The bitstream writing block 220 is configured to convert the encoded computed signal energies and the encoded stereo processing metadata to a bitstream. The bitstream may be output by the bitstream writing block 220.

[0053] Stereo processing is based on a determined stereo coding mode. Some example stereo coding modes, or methods, include at least: a Left/Right mode, a Mid/Side mode, and an extended Mid/Side mode. The selected stereo coding mode for a certain frequency band may be determined based on the estimated or computed inter-channel energy difference of the signal pairs for each possible mode. For example, the coding mode decision may be based on an estimated number of bits that would be required to encode each signal pair for the corresponding coding mode. The estimation of the required number of bits may be based on a level-dependent psychoacoustic model that assigns more bits to louder signal parts than to quieter signal parts. Accordingly, a signal pair with the highest energy difference may require the least number of bits at a certain quality level. In some instances, the bits needed to encode the stereo metadata may also be considered in determining the most efficient coding method. Additionally, in some

instances, the same level dependent psycho-acoustical model may also be used to select the encoding method for the audio data. For the Mid/Side modes, a bit savings estimate relative to the Left/Right coding may be computed and added up to determine the total bit savings on a per frame basis. After deduction of respective metadata bits, the stereo coding mode may be selected, which is most efficient (e.g., has the lowest cost) in terms of required bits.

[0054] FIG. 3 illustrates an example stereo processing process or method, such as that performed by stereo processing block 210. The example stereo processing block 210 includes a stereo analysis unit 305, a psychoacoustic model 310, a bit cost estimator unit 315, a mode decision unit 320, and a mode-based stereo processing unit 325. The stereo analysis unit 305 is configured to receive the complex-valued filter bank domain left and right signals from the complex filter bank analysis block 205. The stereo analysis unit 305 determines, by computation or estimation, the energies of the complex-valued filter bank domain left and right signals, the energies of the Mid signal and the Side signal, and residual energies. The stereo analysis unit 305 is further configured to provide the energies to the bit cost estimator unit 315. The stereo analysis unit 305 also determines, by computation or estimation, stereo metadata based on the complex-valued filter bank domain left and right signals. The stereo metadata may include, for example, one or more of a phase difference, a prediction coefficient, and/or a covariance of the complex-valued filter bank domain left and right signals, the Mid signal, and the Side signal.

[0055] The stereo metadata is provided by the stereo analysis unit 305 to the mode-based stereo processing unit 325 and the bit cost estimator unit 315. The bit cost estimator unit 315 estimates the required number of bits needed to encode the signals based on the determined energies provided by the stereo analysis unit 305 and based on the psychoacoustic model 310. For example, the bit cost estimator unit 315 estimates the required number of bits needed to encode the signals for each candidate (or possible) stereo coding mode. The bit cost estimator unit 315 is configured to transmit the estimated required number of bits to the mode decision unit 320. The mode decision unit 320 is configured to receive the estimated required number of bits, and responsively determines the stereo coding mode (for example, selects one of the candidate stereo coding modes) based on the estimated required number of bits. The mode decision unit 320 is further configured to provide the selected (or determined) stereo coding mode to the mode-based stereo processing unit 325. The mode-based stereo processing unit 325 is configured to receive the selected stereo coding mode from the mode decision unit 320. The mode-based stereo processing unit 325 signals the processed left and right signals using, for example, two bits per frame (e.g., signal 0 and signal 1). The processed left and right signals may correspond to the

left and right signals, Mid and Side signals, or Mid and residual signals, as described below in more detail.

[0056] FIG. 4A illustrates an example joint stereo process 400 performed by the stereo analysis unit 305. By performing the process 400, the stereo analysis unit 305 analyzes the complex-valued filter bank domain left and right signals received from the complex filter bank analysis block 205. For example, the stereo analysis unit 305 receives the complex-valued filter bank domain left signal L and the complex-valued filter bank domain right signal R . The left input signal L is provided according to Equation 1:

$$L(l, n) = L_r(l, n) + jL_i(l, n) \quad [\text{Equation 1}]$$

where l is the filter bank frequency bin, n is the filter bank time index, L_r is the real part of the left input signal L , and L_i is the imaginary (e.g., complex) part of the left input signal L .

[0057] The right input signal R is provided according to Equation 2:

$$R(l, n) = R_r(l, n) + jR_i(l, n) \quad [\text{Equation 2}]$$

where R_r is the real part of the right input signal R , and R_i is the imaginary part of the right input signal R .

[0058] In some instances, the right input signal R is phase rotated prior to computing Mid signal M and Side signal S . The phase-rotated right signal R' is provided according to Equation 3:

$$R'(l, n) = e^{j\varphi} R(l, n) \quad [\text{Equation 3}]$$

where φ is the phase rotation angle.

[0059] Next, the Mid signal M is determined according to Equation 4:

$$M(l, n) = 0.5(L(l, n) + R'(l, n)) \quad [\text{Equation 4}]$$

and the Side signal S is determined according to Equation 5:

$$S(l, n) = 0.5(L(l, n) - R'(l, n)) \quad [\text{Equation 5}]$$

[0060] A residual signal S' is also determined, as provided by Equation 6:

$$S'(l, n) = S(l, n) - pM(l, n) \quad [\text{Equation 6}]$$

where p is a real-valued prediction coefficient (and is provided below in Equation 15).

[0061] Stereo processing parameters (that are provided as the stereo metadata) are also generated by the stereo analysis unit 305. The stereo processing parameters are based on energy and covariance measures for each perceptual frequency band k in an audio frame f (which is omitted in the below computations for clarity). Therefore, energy compaction estimates for multiple stereo coding types referred to by the mode decision unit 320 are computed.

[0062] The energy for the left input signal L is provided by Equation 7:

$$E_k^L = \sum_{l=l_k^0}^{l_k^1} \sum_{n=n_f}^{n_f+N-1} L(l, n) L^*(l, n) \quad [\text{Equation 7}]$$

where $*$ denotes complex conjugate, N is the number of filter bank time slots per frame, n_f is the first filter bank time slot index in the audio frame f , and l_k^0 and l_k^1 correspond to the first and last filter bank frequency bins associated with the frequency band k . In some instances, the total number of frequency bands (K) is approximately 23, and the total number of filter bank frequency bins is greater than K (for example, 60, 64, 256, or more filter bank frequency bins).

[0063] The energy for the right input signal R is provided by Equation 8:

$$E_k^R = \sum_{l=l_k^0}^{l_k^1} \sum_{n=n_f}^{n_f+N-1} R(l, n) R^*(l, n) \quad [\text{Equation 8}]$$

[0064] The covariance with respect to the left input signal L and the right input signal R is provided by Equation 9:

$$C_k = \sum_{l=l_k^0}^{l_k^1} \sum_{n=n_f}^{n_f+N-1} L^*(l, n) R(l, n) \quad [\text{Equation 9}]$$

[0065] The inter-channel phase difference, which may be used to maximize the Mid signal energy, is provided by Equation 10:

$$\varphi_k = \begin{cases} \text{atan2}(\text{real}(C_k), \text{imag}(C_k)) & \text{if } \|C_k\|^2 > (Thr \times E_k^L E_k^R) \\ 0 & \text{otherwise} \end{cases} \quad [\text{Equation 10}]$$

where the threshold Thr may be greater than or equal to 0.5, but less than 1.0.

[0066] The real part of the covariance with phase-rotated right signal R' according to the quantized phase may be derived from the initial covariance using Equation 11:

$$C_{k,Q} = \text{real}(C_k e^{-j\varphi_{k,Q}}) \quad [\text{Equation 11}]$$

where $\varphi_{k,Q}$ is a quantized version of φ_k defined by Equation 12:

$$\varphi_{k,Q} = \text{round}(\varphi_k sf) \quad [\text{Equation 12}]$$

where $sf = \frac{x}{\pi}$ and x is an integer number > 0 .

[0067] The energy of the Mid signal M is computed according to Equation 13:

$$E_k^M = 0.25 (E_k^L + E_k^R + 2.0 C_{k,Q}) \quad [\text{Equation 13}]$$

[0068] The energy of the Side signal S is computed according to Equation 14:

$$E_k^S = 0.25 (E_k^L + E_k^R - 2.0 C_{k,Q}) \quad [\text{Equation 14}]$$

[0069] The prediction coefficient p , which minimizes the energy of the residual signal S' , is computed according to Equation 15:

$$p_k = 0.25 \frac{(E_k^L - E_k^R)}{E_k^M} \quad [\text{Equation 15}]$$

[0070] The residual signal energy is computed according to Equation 16:

$$E_k^{S'} = E_k^S + (p_{k,Q})^2 E_k^M - 0.5 p_{k,Q} (E_k^L - E_k^R) \quad [\text{Equation 16}]$$

where $p_{k,Q}$ is a quantized version of p_k .

[0071] Referring now to operations performed by the mode decision unit 320 of FIG. 3, four different types of joint stereo coding (e.g., advanced stereo coding modes) are considered for every frequency band k . Table 1, shown in FIG. 4B, provides each of the joint stereo coding types and their associated bitstream syntax elements. In Table 1, the type “0” corresponds to separate coding of the left and right signals, and is not a joint stereo coding type. The joint coding types are associated with a corresponding stereo mode $MSMode$. The related bitstream elements are provided to the mode-based stereo processing unit 325.

[0072] The stereo coding type $t = 0$ corresponds to a separate coding of the left and right signals. The stereo coding type $t = 1$ corresponds to a standard Mid/Side coding. The stereo coding type $t = 2$ corresponds to extended Mid/Side coding with adjusted phase alignment only. The stereo coding type $t = 3$ corresponds to extended Mid/Side coding with side prediction only. The stereo coding type $t = 4$ corresponds to extended Mid/Side coding with adjusted phase alignment and side prediction. The stereo coding types for any of $t = 1, 2, 3$, or 4 corresponds to joint coding types.

[0073] To determine the joint stereo mode, the bit cost estimator unit 315 determines an energy ratio quantity for each potential joint coding type (e.g., $t=1,2,3$, or 4) and the separate coding type (e.g., $t = 0$) as defined below by Equations 17 and 18:

$$R_k^0 = \left\lfloor 10 \log_{10} \left(\frac{E_k^L}{E_k^R} \right) \right\rfloor \quad [\text{Equation 17}]$$

$$R_k^t = \left\lfloor 10 \log_{10} \left(\frac{E_{k,t}^M}{E_{k,t}^R} \right) \right\rfloor \text{ for } t = 1, 2, 3, 4 \quad [\text{Equation 18}]$$

[0074] When the ratio for a certain potential joint coding type t ($=1,2,3,4$) is greater than the ratio for the separate coding type $t = 0$, then the certain joint coding type is marked for the respective frequency band k (e.g., a flag is set), as shown by Equation 19:

$$\text{MSFlags}_k^t = \begin{cases} 1 & \text{if } R_k^t > R_k^0 \\ 0 & \text{otherwise} \end{cases} \quad [\text{Equation 19}]$$

[0075] The reduction of required bits due to the energy compaction from joint coding is assessed according to Equation 20:

$$B_k^t = \max\{(R_k^t - R_k^0)(l_k^1 - l_k^0 + 1)N \text{ 2.0 } \textit{smrSlope}_k \textit{bitsPerdB}, 0\} \quad [\text{Equation 20}]$$

where $\textit{smrSlope}_k$ is a factor which relates the Level (in dB) above the threshold in quiet to the required signal-to-mask ratio (in dB) for just noticeable noise, as provided by the psychoacoustic model 310 (and is typically between the values of 0.25 and 0.4). In some instances, $\textit{smrSlope}_k$ is a band-dependent factor. An example of the psychoacoustic model 310 can be found in U.S. Patent Publication No. 2022/0415334, “A Psychoacoustic Model for Audio Processing,” incorporated herein by reference in its entirety. The constant $\textit{bitsPerdB}$ is an estimate of the number of bits needed to quantize a value to achieve a certain signal-to-noise ratio (SNR). The SNR is dependent on the utilized codec, and may be, for example, approximately 0.3.

[0076] The total reduction of required bits for a certain joint coding type is computed by summing the associated reduction in all bands, as provided by Equation 21:

$$B^t = \max\{-B_{\text{cost}}^t + \sum_{k=0}^{K-1} B_k^t, 0\} \quad [\text{Equation 21}]$$

considering the signaling cost B_{cost}^t according to the bitstream syntax, described below with respect to Table 2. The most-efficient (e.g., lowest cost) joint stereo coding type t is selected by the mode decision unit 320 as provided by Equation 22:

$$t_{\text{best}} = \textit{argmax}(B^t) \quad [\text{Equation 22}]$$

In some embodiments, the joint coding type may be selected such as to exceed bit reduction of another joint coding type by a certain percentage.

[0077] The mode-based stereo processing unit 325 is configured to receive the stereo metadata from the stereo analysis unit 305 (which includes the quantized phase difference and the

quantized prediction coefficient) and receive the selected joint stereo coding type t from the mode decision unit 320. Processing of the left input signal L and the right input signal R is performed by the mode-based stereo processing unit 325 based on the selected joint stereo coding type t (e.g., the selected stereo mode).

[0078] In instances where the separate coding type $t = 0$ is selected, the mode-based stereo processing unit 325 may simply pass the left input signal L and the right input signal R to the encoding block 215, without changing or altering the signals. In other instances, where the coding type $t = 1, 2, 3$, or 4 is selected, a Mid/Side mode, the mode-based stereo processing unit 325 processes the left input signal L and the right input signal R according to the selected joint stereo coding type t and the stereo metadata.

[0079] FIG. 5A illustrates a block diagram of an example mode-based stereo processing unit, such as mode-based stereo processing unit 325 of FIG. 3. The illustrated example may be used when a Mid/Side mode is selected. The example mode-based stereo processing unit 325 includes a phase-alignment block 505, a Mid/Side conversion block 510, and a side prediction block 515. The phase-alignment block 505 is configured to receive the left input signal L , the right input signal R , and the quantized phase difference. The phase-alignment block 505 is only needed for the stereo coding types $t = 2$ and $t = 4$. The phase-alignment block 505 performs phase alignment adjustment between the left input signal L and the right input signal R using the quantized phase difference $\varphi_{k,Q}$. In some examples, the adjustment for phase alignment may be applied to one of the L or R signals, while in other examples the adjustment for phase alignment may be applied to both the L and R signals. The phase-aligned left input signal L and the phase-aligned right input signal R are output by the phase-alignment block 505 to the side conversion block 510. In other joint stereo coding types, such as $t = 1$ and $t = 3$, the phase-alignment block 505 may be disabled or bypassed (e.g., $\varphi_{k,Q}$ may be set to 0).

[0080] The Mid/Side conversion block 510 is configured to receive the left input signal L and the right input signal R from the phase-alignment block 505, which may be phase-aligned as discussed above. The Mid/Side conversion block 510 converts the left input signal L and the right input signal R to Mid signal M and side signal S (and, in some instances, residual signal S'), as previously described with respect to FIG. 4A. The Mid signal M and the side signal S are output by the side conversion block 510 to the side prediction block 515.

[0081] The side prediction block 515 may be implemented for the stereo coding types $t = 3$ and $t = 4$. The side prediction block 515 receives the Mid signal M and the side signal S from the side conversion block 510 and receives the quantized prediction coefficient $p_{k,Q}$ from the mode

decision unit 320. The side prediction block 515 applies side prediction to the Mid signal M and the side signal S . In other joint stereo coding types, such as $t = 1$ and $t = 2$, the side prediction block 515 is disabled or bypassed (e.g., $p_{k,Q}$ is set to 0).

[0082] The stereo modes may be signaled, for example, by 2 bits per frame from the mode-based stereo processing unit 325. In this example, Mid/Side coding versus Left/Right coding may be signaled by one bit per frequency band. Alternatively, Mid/Side coding active for all frequency bands may be indicated by a single bit per frame. For the extended Mid/Side mode, the presence of phase difference data may be signaled by one bit per frame and the side prediction data may be signaled by one bit per frame. In this manner, the metadata size may be reduced for different types of stereo signals. If the side prediction data is present, the side prediction data may be Huffman entropy coded per differences across relevant frequency bands. If the phase difference data is present, then the phase difference data may be entropy coded. Table 2, shown in FIG. 5B, provides an example bitstream syntax.

[0083] During encoding (at the encoding block 215), the inter-channel phase difference data is linearly quantized. The scale factor used for quantization is selected such that the values π and $-\pi$ are precisely represented as shown in Equation 12. In some implementations, the second order differences of the phase symbols $\varphi_{k,Q}$ (e.g., the quantized inter-channel phase differences per band) along the relevant frequency bands may be computed and wrapped into a $[-\pi, +\pi]$ range such that no jumps larger than π are required. The second order differences may be computed by first computing the difference between the phase symbols in bands 1 to N and the phase symbols in bands 0 to N-1, where N is the number of bands where M/S processing with phase adjustment is active (for example, see variable “numMSBands” in Table 2 of FIG. 5B and in Pseudocode 1 of FIG. 6B). In a second step, the second order differences are computed as the differences between the new values in bands 2 to N and the new values in bands 1 to N, as shown in Pseudocode 2 of FIG. 6C. In this manner, the quantized number in the first frequency band may be unmodified, the quantized number in the second frequency band may represent a difference, and the quantized numbers in the rest of the frequency bands may correspond to differences from the first difference. The wrapped phase symbols may be entropy coded, where the symbol zero may be encoded using the least number of bits. The signal model has inter-channel time delay or a constant phase shift. In some instances, for linear frequency banding, the second order frequency differentials are substantially equal to zero and thus need only a small number of bits to be encoded. In other instances, the second order frequency differentials are substantially non-zero. Additionally, in some instances, only the first order difference is used for encoding. Regardless, the described entropy coding scheme for inter-channel phase difference using second

order frequency difference coding is more efficient (in the sense that it uses less bits for the same information) compared to plain frequency differential coding or non-differential coding for stereo coding content, such as binaural stereo.

[0084] FIG. 6A illustrates the stereo metadata rate for the extended Mid/Side coding mode per audio frame for an item with five samples (at 48 kHz) inter-channel time delay over a plurality of frames when encoding with either the first order differential or the second order differential. In FIG. 6A, the x-axis provides the plurality of frames and the y-axis provides the bitrate (in kB/s) of encoding for each frame. As shown in FIG. 6A, when encoding the quantized phase angles for jointly coded bands, as indicated by $MSFlags_k^t$, the quantized phase angles may be first arranged adjacently. As one particular example of the operations, refer to Pseudocode 1 as shown in FIG. 6B. Next, first and second order differences are computed for the arranged quantized phase angles. Lastly, the data is wrapped into the quantized 2π range before Huffman encoding. Pseudocode 2, as shown in FIG. 6C provides one example of computing the first and second order differences, and presented using a C-like format. Pseudocode 3, as shown in FIG. 6D, provides an example of wrapping of the phase symbols (variable $phaseQ$) into 2π range.

[0085] The phase symbols either represent absolute phase (first frequency band), the first order difference (second frequency band), or the second order difference across frequency bands (other frequency bands). The same phase wrapping previously described may apply if the phase data is coded using first order differentials. In Pseudocode 3, shown in FIG. 6D, a value of PHASE_MAX_VAL represents a value for $+\pi$ and a value of PHASE_MIN_VAL represents a value for $-\pi$, where the phase symbols are represented by integers in the range [PHASE_MIN_VAL, PHASE_MAX_VAL], where PHASE_MIN_VAL is defined as the negative PHASE_MAX_VAL and PHASE_MAX_VAL is the integer number x used for quantization in Equation 12 (for example, $x = 12$).

[0086] FIGS. 7A-7B illustrate a block diagram of various example methods 700 for encoding stereo signals, which may be performed by the encoder 110 of FIG. 2. The methods 700 may be performed by a processor, which may be configured to perform methods 700 via machine-executable instructions. The methods 700 may be broken into various blocks or partitions, such as blocks 705, 710, 715, 720, 725, 730, 735, 740, 745, 750, 755, 760, 765, and 770. The various process blocks illustrated in FIG. 7 provide examples of various methods disclosed herein, and it is understood that some blocks may be removed, added, combined, or modified without departing from the spirit of the present disclosure. For some examples, processing of the various

blocks, which may be described as processes, methods, steps, blocks, operations, or functions, may commence at block 705.

[0087] At block 705, “Passing Left-Channel and Right-Channel of Stereo Audio Signal To Filter Bank Analysis,” an example method 700 may include passing a left-channel and a right-channel of a stereo audio signal to a filter bank analysis. The filter bank analysis responsively generates one or more frequency bands. Processing may continue from block 705 to block 710.

[0088] At block 710, “Calculating Energy of Left-Channel, Energy of Right-Channel, and Covariance,” an example method 700 may include calculating the energy of the left-channel, calculating the energy of the right-channel, and calculating the covariance of the left-channel and the right channel, as previously described with respect to Equations 7-16. Processing may continue from block 710 to block 715.

[0089] At block 715, “Determining Bit Cost of a Plurality of Stereo Coding Modes,” an example method 700 may include determining a bit cost of a plurality of stereo coding modes, as previously described with respect to Equations 17-22 and Table 1. Processing may continue from block 715 to block 720. At block 720, the method 700 includes selecting a stereo coding mode, as previously described with respect to Table 1.

[0090] When a separated coding mode is selected at block 720, processing continues from block 720 to block 725. An example method 700 may include, at block 725, “Encoding Left-Channel and Right-Channel”, encoding the left-channel and the right-channel separately. For example, the stereo processing block 210 passes the left input signal and the right input signal to the encoding block 215.

[0091] When the Mid/Side coding mode is selected at block 720, processing continues from block 720 to block 730. An example method 700 may include, at block 730, “Converting To Mid Signal and Side Signal”, converting (for example, transforming) the left-channel and the right-channel to a Mid signal and a Side signal (for example, with the side conversion block 510). Processing may continue from block 730 to block 735. At block 735, “Encoding Mid Signal and Side Signal,” an example method 700 may include encoding the Mid signal and the Side signal.

[0092] When the extended Mid/Side coding mode is selected at block 720, processing continues from block 720 to block 740. An example method 700 may include, at block 740, “Adjusting Phase Alignment Based On Phase Difference,” adjusting phase alignment to the left-channel

and/or the right-channel based on a calculated phase difference (for example, with the phase-alignment block 505). Processing may continue from block 740 to block 745.

[0093] At block 745, “Converting to Mid Signal and Side Signal,” the method 700 includes converting the phase-aligned left-channel and right-channel to a Mid signal and a Side signal (for example, with the side conversion block 510). Processing may continue from block 745 to block 750.

[0094] At block 750, “Generating Residual Signal Using Side Prediction,” an example method 700 may include generating a residual signal using a side prediction coefficient (for example, with the side prediction block 515). Processing may continue from block 750 to block 755.

[0095] At block 755, “Encoding Mid Signal and Residual Signal,” an example method 700 may include encoding the Mid signal and the residual signal. In some instances, the phase difference and the side prediction coefficient are encoded alongside the Mid signal and the residual signal. Additionally, in some instances, only phase alignment adjustment or use of the side prediction coefficient is performed.

[0096] FIG. 8A illustrates a block diagram of an example decoder 120. The example decoder 120 reverses the encoding performed by the encoder 110. The example decoder 120 includes a bitstream reading block 805, a decoding block 810, an inverse stereo processing block 815, and a filter bank synthesis block 820. The bitstream reading block 805 receives the bitstream from the encoder 110. The bitstream reading block 805 processes the bitstream and provides the processed bitstream to the decoding block 810.

[0097] The decoding block 810 is configured to receive the processed bitstream from the bitstream reading block 805. The decoding block 810 processes (e.g., decodes) the processed bitstream to substantially replicate the Mid signal, the Side signal, and the stereo metadata. The replicated Mid signal, the replicated Side signal, and the replicated stereo metadata are provided by the decoding block 810 to the inverse stereo processing block 815.

[0098] The inverse stereo processing block 815 is configured to receive the replicated Mid signal, the replicated Side signal, and the replicated stereo metadata from the decoding block 810. The inverse stereo processing block 815 is configured to process the replicated Mid signal, the replicated Side signal, and the replicated stereo metadata to generate a replicated complex-valued filter bank domain left signal and a replicated complex-valued filter bank domain right signal. For example, to reverse the stereo processing, in instances where the residual signal is encoded, the inverse stereo processing block 815 first reconstructs the Side signal using side

prediction information included in the stereo metadata. Then, the Mid/Side transform is inverted. Finally, the original phase relation for the left signal and the right signal is reinstated based on the transmitted phase data. In some instances, phase adjustment (or alignment) is applied only to one channel. In other instances, phase adjustment (or alignment) is applied to both the left channel and the right channel. Equation 23 provides an example of decoding the jointly coded signals in matrix notation:

$$\begin{bmatrix} L \\ R \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & e^{-j\varphi_Q} \end{bmatrix} \begin{bmatrix} 1 + p_Q & 1 \\ 1 - p_Q & -1 \end{bmatrix} \begin{bmatrix} M \\ S' \end{bmatrix} \quad [\text{Equation 23}]$$

[0099] Decoding of the second order frequency differential coding of the transmitted phase signals is provided by Pseudocode 4 shown in FIG. 8B. The inverse stereo processing block 815 provides the replicated complex-valued filter bank domain left signal and the replicated complex-valued filter bank domain right signal to the filter bank synthesis block 820. The filter bank synthesis block 820 is configured to receive the replicated complex-valued filter bank domain left signal and the replicated complex-valued filter bank domain right signal. The filter bank synthesis block 820 converts the replicated complex-valued filter bank domain left signal and the replicated complex-valued filter bank domain right signal to a replicated left signal and a replicated right signal. The filter bank synthesis block 820 outputs the replicated original left signal and the replicated original right signal.

[0100] FIG. 9 illustrates a block diagram of various example methods 900 for decoding stereo signals, which may be performed by the decoder 120 of FIG. 8A. The example methods 900 may be performed by a processor, which may be configured to perform methods 900 via machine-executable instructions. The methods 900 may be broken into various blocks or partitions, such as blocks 905, 910, 915, and 920. The various process blocks illustrated in FIG. 9 provide examples of various methods disclosed herein, and it is understood that some blocks may be removed, added, combined, or modified without departing from the spirit of the present disclosure. In some examples of methods 900, stereo signals may have been previously encoded using an extended Side/Mid mode. For some examples, the processing of the various blocks, which may be described as processes, methods, steps, blocks, operations, or functions, may commence at block 905.

[0101] At block 905, “Receiving Encoded Bitstream,” an example method 900 may include receiving an encoded bitstream. For example, the decoder 120 receives a bitstream from the encoder 110. Processing may continue from block 905 to block 910. At block 910, “Decoding, From the Bitstream, the Mid Signal, the Residual Signal, and Stereo Metadata,” an example

method 900 may include decoding, from the bitstream, the replicated Mid signal, the replicated residual signal, and the replicated stereo metadata. Processing may continue from block 910 to block 915.

[0102] At block 915, “Converting the Mid Signal and the Residual Signal to a Left Channel and a Right Channel Using the Stereo Metadata,” an example method 900 may include converting the replicated Mid signal and the replicated residual signal to a replicated left channel and a replicated right channel using the replicated stereo metadata. For example, the Side signal is replicated using side prediction information included in the replicated stereo metadata. The replicated Mid signal and the replicated Side signal are converted to a replicated left channel and a replicated right channel. In instances where the left channel and the right channel are misaligned, phase adjustment is performed using a phase difference included in the stereo metadata. Processing may continue from block 915 to block 920.

[0103] At block 920, “Passing the Left Channel and the Right Channel to Filter Bank Analysis to Recreate Original Signals,” an example method 900 may include passing the replicated left channel and the replicated right channel to a filter bank analysis to replicate the original left channel signal and the original right channel signal.

[0104] FIG. 10 illustrates a graph of a Perceptual Evaluation of Audio Quality (PEAQ) for twelve audio items (provided along the x-axis). An audio item is an audio event captured by a microphone for encoding by the encoder 110. An Objective Difference Grade (ODG) value is provided (along the y-axis) for each audio item. The graph illustrates the benefits of joint stereo encoding. For example, audio quality was improved at least in audio item 5 (which had a small inter channel time delay) as indicated by the by ODG value for encoding item 5 with the extended Mid/Side mode being closer to zero than encoding item 5 with the Mid/Side mode. Audio item 9, which is a panned speech item, shows a similar improvement in audio quality, as indicated by the ODG for encoding item 9 with the extended Mid/Side mode being closer to zero than encoding item 9 with the Mid/Side mode. In the PEAQ evaluation, the codec was operated at 256 kb/s.

[0105] FIG. 11A illustrates a schematic block diagram of an example device architecture 1100 (e.g., an apparatus 1100) that may be used to implement various aspects of the present disclosure. Architecture 1100 includes but is not limited to servers and client devices, systems, and methods as described in reference to FIGS. 1-10. As shown, the architecture 1100 includes central processing unit (CPU) 1101 which is capable of performing various processes in accordance with a program stored in, for example, read only memory (ROM) 1102 or a program

loaded from, for example, storage unit 1108 to random access memory (RAM) 1103. The CPU 1101 may be, for example, an electronic processor 1101. In RAM 1103, the data required when CPU 1101 performs the various processes is also stored, as required. CPU 1101, ROM 1102, and RAM 1103 are connected to one another via bus 1104. Input/output interface 1105 is also connected to bus 1104.

[0106] The following components are connected to I/O interface 1105: input unit 1106, that may include a keyboard, a mouse, or the like; output unit 1107 that may include a display such as a liquid crystal display (LCD) and one or more speakers; storage unit 1108 including a hard disk, or another suitable storage device; and communication unit 1109 including a network interface card such as a network card (e.g., wired or wireless).

[0107] In some implementations, input unit 1106 includes one or more microphones in different positions (depending on the host device) enabling capture of audio signals in various formats (e.g., mono, stereo, spatial, immersive, and other suitable formats).

[0108] In some implementations, output unit 1107 include systems with various number of speakers. Output unit 1107 (depending on the capabilities of the host device) can render audio signals in various formats (e.g., mono, stereo, immersive, binaural, and other suitable formats).

[0109] In some embodiments, communication unit 1109 is configured to communicate with other devices (e.g., via a network). Drive 1110 is also connected to I/O interface 1105, as required. Removable medium 1111, such as a magnetic disk, an optical disk, a magneto-optical disk, a flash drive or another suitable removable medium is mounted on drive 1110, so that a computer program read therefrom is installed into storage unit 1108, as required. A person skilled in the art would understand that although apparatus 1100 is described as including the above-described components, in real applications, it is possible to add, remove, and/or replace some of these components and all these modifications or alteration all fall within the scope of the present disclosure.

[0110] In accordance with example embodiments of the present disclosure, the processes described above may be implemented as computer software programs or on a computer-readable storage medium. For example, embodiments of the present disclosure include a computer program product including a computer program tangibly embodied on a machine readable medium, the computer program including program code for performing methods. In such embodiments, the computer program may be downloaded and mounted from the network via the

communication unit 1109, and/or installed from the removable medium 1111, as shown in FIG. 11A.

[0111] FIG. 11B illustrates a schematic block diagram of an example CPU 1101 implemented in the device architecture 1100 of FIG. 11A that may be used to implement various aspects of the present disclosure. The CPU 1101 includes an electronic processor 1120 and a memory 1121. The electronic processor 1120 is electrically and/or communicatively connected to the memory 1121 for bidirectional communication. The memory 1121 stores encoding software 1122 and/or decoding software 1123. In some examples, memory 1121 may be located internal to the electronic processor 1120, such as for an internal cache memory or some other internally located ROM, RAM, or flash memory. In other examples, memory 1121 may be located external to the electronic processor 1120, such as in a ROM 1102, a RAM 1103, flash memory or a removable medium 1111, or another non-transitory computer readable medium that is contemplated for device architecture 1100. In some instances, the electronic processor 1120 may implement the encoding software 1122 stored in the memory 1121 to perform, among other things, any of the methods 700 of FIGS. 7A-7B. Additionally, the electronic processor 1120 may implement the decoding software 1123 stored in the memory 1121 to perform, among other things, any of the methods 900 of FIG. 9.

[0112] Generally, various example embodiments of the present disclosure may be implemented in hardware or special purpose circuits (e.g., control circuitry), software, logic or any combination thereof. For example, the units discussed above can be executed by control circuitry (e.g., CPU 1101 in combination with other components of FIG. 11A), thus, the control circuitry may be performing the actions described in this disclosure. Some aspects may be implemented in hardware, while other aspects may be implemented in firmware or software which may be executed by a controller, microprocessor or other computing device (e.g., control circuitry). While various aspects of the example embodiments of the present disclosure are illustrated and described as block diagrams, flowcharts, or using some other pictorial representation, it will be appreciated that the blocks, apparatus, systems, techniques or methods described herein may be implemented in, as non-limiting examples, hardware, software, firmware, special purpose circuits or logic, general purpose hardware or controller or other computing devices, or some combination thereof.

[0113] Additionally, various blocks shown in the flowcharts may be viewed as method steps, and/or as operations that result from operation of computer program code, and/or as a plurality of coupled logic circuit elements constructed to carry out the associated function(s). For example,

embodiments of the present disclosure include a computer program product including a computer program tangibly embodied on a machine readable medium, the computer program containing program codes configured to carry out the methods as described above.

[0114] In the context of the disclosure, a machine-readable medium may be any tangible medium that may contain or store a program for use by or in connection with an instruction execution system, apparatus, or device. The machine-readable medium may be a machine-readable signal medium or a machine-readable storage medium. A machine-readable medium may be non-transitory and may include but not limited to an electronic, magnetic, optical, electromagnetic, infrared, or semiconductor system, apparatus, or device, or any suitable combination of the foregoing. More specific examples of the machine-readable storage medium would include an electrical connection having one or more wires, a portable computer diskette, a hard disk, a random-access memory (RAM), a read-only memory (ROM), an erasable programmable read-only memory (EPROM or Flash memory), an optical fiber, a portable compact disc read-only memory (CD-ROM), an optical storage device, a magnetic storage device, or any suitable combination of the foregoing.

[0115] Computer program code for carrying out methods of the present disclosure may be written in any combination of one or more programming languages. These computer program codes may be provided to a processor of a general-purpose computer, special purpose computer, or other programmable data processing apparatus that has control circuitry, such that the program codes, when executed by the processor of the computer or other programmable data processing apparatus, cause the functions/operations specified in the flowcharts and/or block diagrams to be implemented. The program code may execute entirely on a computer, partly on the computer, as a stand-alone software package, partly on the computer and partly on a remote computer or entirely on the remote computer or server or distributed over one or more remote computers and/or servers.

[0116] A person skilled in the art realizes that the present invention by no means is limited to the embodiments described above. On the contrary, many modifications and variations are possible and considered within the scope of the appended claims. Various aspects and implementations of the present disclosure may also be appreciated from the following enumerated example embodiments (EEEs), which are not claims, and which may represent systems, methods, and devices, all arranged in accordance with aspects of the present disclosure.

[0117] EEE1. A method for encoding a stereo audio signal in a bitstream, the method comprising: passing a left-channel and a right-channel of the stereo audio signal to a complex-

valued filter bank analysis block to generate one or more frequency bands; calculating, for each of the one or more frequency bands, an energy of the left-channel, an energy of the right-channel, and a covariance of the left-channel and the right-channel; selecting, based on the calculated energy of the left-channel, the calculated energy of the right-channel, and the calculated covariance of the left-channel and the right-channel, a stereo coding mode in which to encode the left-channel and the right-channel; and when the stereo coding mode is an extended Mid/Side coding mode: computing a phase difference between the left-channel and the right-channel; adjusting phase alignment between the left-channel and the right-channel based on the computed phase difference to generate an aligned left-channel and an aligned right-channel; transforming the aligned left-channel and the aligned right-channel to a Mid signal and a Side signal; generating a residual signal based on side prediction data and the Side signal; encoding the Mid signal, the residual signal, the phase difference, and the side prediction data in a bitstream; and outputting the bitstream for the selected stereo coding mode.

[0118] EEE2. The method according to EEE1, wherein the method includes: when the stereo coding mode is a separated coding mode: encoding the left-channel and the right-channel in the bitstream.

[0119] EEE3. The method according to any one of EEE1 to EEE2, wherein the method includes: when the stereo coding mode is a Mid/Side coding mode: transforming the aligned left-channel and the aligned right-channel to the Mid signal and the Side signal; and encoding the Mid signal and the Side signal in the bitstream.

[0120] EEE4. The method according to any one of EEE1 to EEE3, wherein the Mid signal in the extended Mid/Side coding mode represents a sum of the left-channel and the right-channel, and wherein the Side signal in the extended Mid/Side coding mode represents a difference between the left-channel and the right-channel.

[0121] EEE5. The method according to any one of EEE1 to EEE4, wherein computing the phase difference includes: comparing the calculated covariance of the left-channel and the right-channel to an energy threshold, and setting the computed phase difference to zero when the calculated covariance of the left-channel and the right-channel is less than or equal to the energy threshold.

[0122] EEE6. The method according to any one of EEE1 to EEE5, wherein selecting the stereo coding mode includes: determining a bit cost associated with each of a plurality of stereo coding modes based on the calculated energy of the left-channel, the calculated energy of the right-

channel, the calculated covariance of the left-channel and the right-channel, and a cost of transmitting the stereo audio signal; and selecting the stereo coding mode based on the bit cost.

[0123] EEE7. The method according to EEE6, wherein determining the bit cost associated with each of the plurality of stereo coding modes includes: determining an energy ratio of signals included in each of the plurality of stereo coding modes; and comparing the energy ratio to a threshold, wherein the bit cost indicates a bit reduction between each of the plurality of stereo coding modes compared to coding the left-channel and the right-channel, and wherein the threshold is based on the calculated energy of the left-channel and the calculated energy of the right-channel.

[0124] EEE8. The method according to any one of EEE1 to EEE7, further comprising signaling the stereo coding mode that is selected using two bits per frame.

[0125] EEE9. The method according to EEE8, further comprising signaling the presence of the phase difference or the presence of the side prediction data using one bit per frame each.

[0126] EEE10. The method according to any one of EEE1 to EEE9, wherein the phase difference and the side prediction data are quantized.

[0127] EEE11. An apparatus for encoding a stereo audio signal in a bitstream, the apparatus comprising: an electronic processor configured to: pass a left-channel and a right-channel of the stereo audio signal to a complex-valued filter bank analysis block to generate one or more frequency bands; calculate, for each of the one or more frequency bands, an energy of the left-channel, an energy of the right-channel, and a covariance of the left-channel and the right-channel; select, based on the calculated energy of the left-channel, the calculated energy of the right-channel, and the calculated covariance of the left-channel and the right-channel, a stereo coding mode in which to encode the left-channel and the right-channel; and when the stereo coding mode is an extended Mid/Side coding mode: compute a phase difference between the left-channel and the right-channel; adjust phase alignment between the left-channel and the right-channel based on the computed phase difference to generate an aligned left-channel and an aligned right-channel; transform the aligned left-channel and the aligned right-channel to a Mid signal and a Side signal; generate a residual signal based on side prediction data and the Side signal; encode the Mid signal, the residual signal, the phase difference, and the side prediction data in the bitstream; and output the bitstream for the selected stereo coding mode.

[0128] EEE12. The apparatus according to EEE11, wherein the electronic processor is configured to: when the stereo coding mode is a separated coding mode: encode the left-channel and the right-channel in the bitstream.

[0129] EEE13. The apparatus according to any one of EEE11 to EEE12, wherein the electronic processor is configured to: when the stereo coding mode is a Mid/Side coding mode: transform the aligned left-channel and the aligned right-channel to the Mid signal and the Side signal; and encode the Mid signal and the Side signal in the bitstream.

[0130] EEE14. The apparatus according to any one of EEE11 to EEE13, wherein adjusting phase alignment between the left-channel and the right-channel includes adjusting a phase of the right-channel to align the left-channel and the right-channel.

[0131] EEE15. The apparatus according to any one of EEE11 to EEE14, wherein, to compute the phase difference, the electronic processor is configured to: compare the calculated covariance of the left-channel and the right-channel to an energy threshold, and set the computed phase difference to zero when the calculated covariance of the left-channel and the right-channel is less than or equal to the energy threshold.

[0132] EEE16. The apparatus according to any one of EEE11 to EEE15, wherein, to select the stereo coding mode, the electronic processor is configured to: determine a bit cost associated with each of a plurality of stereo coding modes based on the calculated energy of the left-channel, the calculated energy of the right-channel, the calculated covariance of the left-channel and the right-channel, and a cost of transmitting the stereo audio signal; and select the stereo coding mode based on the bit cost.

[0133] EEE17. The apparatus according to EEE16, wherein, to determine the bit cost associated with each of the plurality of stereo coding modes, the electronic processor is configured to: determine an energy ratio of signals included in each of the plurality of stereo coding modes; and compare the energy ratio to a threshold, wherein the bit cost indicates a bit reduction between each of the plurality of stereo coding modes compared to coding the left-channel and the right-channel, and wherein the threshold is based on the calculated energy of the left-channel and the calculated energy of the right-channel.

[0134] EEE18. The apparatus according to any one of EEE11 to EEE17, wherein the electronic processor is configured to signal the stereo coding mode that is selected using two bits per frame.

[0135] EEE19. The apparatus according to EEE18, wherein the electronic processor is configured to signal the presence of the phase difference or the side prediction data using one bit per frame each.

[0136] EEE20. The apparatus according to any one of EEE11 to EEE19, wherein the phase difference and the side prediction data are quantized.

[0137] EEE21. The apparatus according to any one of EEE11 to EEE20, wherein the phase difference is a linearly quantized inter-channel phase difference, and wherein to encode the phase difference, the electronic processor is configured to: compute, per frequency band, second order differences of the linearly quantized inter-channel phase difference, wrap the second order differences into a 2π range, and encode the wrapped second order differences.

[0138] EEE22. The method according to any one of EEE1 to EEE10, wherein the bitstream corresponds to an IVAS bitstream.

[0139] EEE23. The apparatus according to any one of EEE11 to EEE21, wherein the bitstream corresponds to an IVAS bitstream.

[0140] EEE24. A method for decoding a stereo audio signal, the method comprising: receiving an encoded bitstream; decoding, from the bitstream, a replicated Mid signal, a replicated residual signal, and replicated stereo metadata, wherein the replicated stereo metadata includes a phase difference and side prediction data; converting the replicated Mid signal and the replicated residual signal to a replicated left channel and a replicated right channel using the replicated stereo metadata; and passing the replicated left channel and the replicated right channel to a filter bank analysis block to recreate an original left channel and an original right channel.

[0141] EEE25. The method according to EEE24, wherein converting the replicated Mid signal and the replicated residual signal includes generating a Side signal from the replicated residual signal based on the side prediction data.

[0142] EEE26. The method according to any one of EEE24 to EEE25, further comprising aligning the replicated left channel and the replicated right channel using the phase difference.

[0143] EEE27. An apparatus for decoding a stereo audio signal, the apparatus comprising: an electronic processor configured to: receive an encoded bitstream; decode, from the bitstream, a replicated Mid signal, a replicated residual signal, and replicated stereo metadata, wherein the replicated stereo metadata includes a phase difference and side prediction data; convert the replicated Mid signal and the replicated residual signal to a replicated left channel and a

replicated right channel using the replicated stereo metadata; and pass the replicated left channel and the replicated right channel to a filter bank analysis block to recreate an original left channel and an original right channel.

[0144] EEE28. The apparatus according to EEE27, wherein, to convert the replicated Mid signal and the replicated residual signal, the electronic processor is configured to generate a Side signal from the replicated residual signal based on the side prediction data.

[0145] EEE29. The apparatus according to any one of EEE27 to EEE28, wherein the electronic processor is configured to align the replicated left channel and the replicated right channel using the phase difference.

[0146] EEE30. A method for encoding a stereo audio signal, comprising: determining an advanced stereo coding mode for encoding the stereo audio signal, wherein the advanced stereo coding mode is one selected from the group consisting of a left/right coding mode, a mid/side coding mode, and an extended mid/side coding mode, signaling the advanced stereo coding mode using two bits per frame, signaling, in response to the advanced stereo coding mode being the extended mid/side coding mode, phase difference data using one bit per frame, and signaling, in response to the advanced stereo coding mode being the extended mid/side coding mode, prediction data using one bit per frame.

[0147] EEE31. A non-transitory computer-readable storage medium recording a program of instructions that is executable by a device to perform the method according to any one of EEE1 to EEE10, EEE22, EEE24 to EEE26, or EEE30.

[0148] With regard to the processes, systems, methods, heuristics, etc. described herein, it should be understood that, although the steps of such processes, etc. have been described as occurring according to a certain ordered sequence, such processes could be practiced with the described steps performed in an order other than the order described herein. It further should be understood that certain steps could be performed simultaneously, that other steps could be added, or that certain steps described herein could be replaced, amended, or omitted. In other words, the descriptions of processes herein are provided for the purpose of illustrating certain embodiments and should in no way be construed so as to limit the claims.

[0149] Accordingly, it is to be understood that the above description is intended to be illustrative and not restrictive. Many embodiments and applications other than the examples provided would be apparent upon reading the above description. The scope should be determined, not with reference to the above description, but should instead be determined with reference to the

appended claims, along with the full scope of equivalents to which such claims are entitled. It is anticipated and intended that future developments will occur in the technologies discussed herein, and that the disclosed systems and methods will be incorporated into such future embodiments. In sum, it should be understood that the application is capable of modification and variation.

[0150] All terms used in the claims are intended to be given their broadest reasonable constructions and their ordinary meanings as understood by those knowledgeable in the technologies described herein unless an explicit indication to the contrary is made herein. In particular, use of the singular articles such as “a,” “the,” “said,” etc. should be read to recite one or more of the indicated elements unless a claim recites an explicit limitation to the contrary.

[0151] The Abstract of the Disclosure is provided to allow the reader to quickly ascertain the nature of the technical disclosure. It is submitted with the understanding that it will not be used to interpret or limit the scope or meaning of the claims. In addition, in the foregoing Detailed Description, it can be seen that various features are grouped together in various embodiments for the purpose of streamlining the disclosure. This method of disclosure is not to be interpreted as reflecting an intention that the claimed embodiments incorporate more features than are expressly recited in each claim. Rather, as the following claims reflect, inventive subject matter lies in less than all features of a single disclosed embodiment. Thus, the following claims are hereby incorporated into the Detailed Description, with each claim standing on its own as a separately claimed subject matter.

CLAIMS

What is claimed is:

1. A method (700) for encoding a stereo audio signal in a bitstream, the method comprising:
 - passing (705) a left-channel and a right-channel of the stereo audio signal to a complex-valued filter bank analysis block (205) to responsively generate one or more frequency bands;
 - calculating (710), for each of the one or more frequency bands, an energy of the left-channel, an energy of the right-channel, and a covariance of the left-channel and the right-channel;
 - selecting (720), based on the calculated energy of the left-channel, the calculated energy of the right-channel, and the calculated covariance of the left-channel and the right-channel, a stereo coding mode in which to encode the left-channel and the right-channel; and
 - when the stereo coding mode is an extended Mid/Side coding mode:
 - computing a phase difference between the left-channel and the right-channel;
 - adjusting (740) phase alignment between the left-channel and the right-channel based on the computed phase difference to generate an aligned left-channel and an aligned right-channel;
 - transforming (745) the aligned left-channel and the aligned right-channel to a Mid signal and a Side signal;
 - generating (750) a residual signal based on side prediction data and the Side signal; and
 - encoding (755) the Mid signal, the residual signal, and the side prediction data in the bitstream; and
 - outputting the bitstream for the selected stereo coding mode.
2. The method (700) of claim 1, wherein the bitstream corresponds to an IVAS bitstream.
3. The method (700) of claims 1 or 2, wherein the method (700) includes:
 - when the stereo coding mode is a separated coding mode:
 - encoding (725) the left-channel and the right-channel in the bitstream.
4. The method (700) of any of claim 1 to 3, wherein the method (700) includes:
 - when the stereo coding mode is a Mid/Side coding mode:
 - transforming (735) the aligned left-channel and the aligned right-channel to the Mid signal and the Side signal; and
 - encoding (740) the Mid signal and the Side signal in the bitstream.

5. The method (700) of any of claims 1 to 4, wherein the Mid signal in the extended Mid/Side coding mode represents a sum of the left-channel and the right-channel, and wherein the Side signal in the Mid/Side coding mode represents a difference between the left-channel and the right-channel.
6. The method (700) of any of claims 1 to 5, wherein computing the phase difference includes:
 - comparing the calculated covariance of the left-channel and the right-channel to an energy threshold, and
 - setting the computed phase difference to zero when the calculated covariance of the left-channel and the right-channel is less than or equal to the energy threshold.
7. The method (700) of any of claims 1 to 6, wherein selecting the stereo coding mode includes:
 - determining (715) a bit cost associated with each of a plurality of stereo coding modes based on the calculated energy of the left-channel, the calculated energy of the right-channel, the calculated covariance of the left-channel and the right-channel, and a cost of transmitting the stereo audio signal; and
 - selecting the stereo coding mode based on the bit cost.
8. The method (700) of claim 7, wherein determining (715) the bit cost associated with each of the plurality of stereo coding modes includes:
 - determining an energy ratio of signals included in each of the plurality of stereo coding modes; and
 - comparing the energy ratio to a threshold,
 - wherein the bit cost indicates a bit reduction between each of the plurality of stereo coding modes compared to coding the left-channel and the right-channel, and
 - wherein the threshold is based on the calculated energy of the left-channel and the calculated energy of the right-channel.
9. The method (700) of any of claims 1 to 8, further comprising signaling the stereo coding mode that is selected using two bits per frame.
10. The method (700) of claim 9, further comprising signaling the presence of the phase difference or the presence of the side prediction data using one bit per frame each.

11. The method (700) of any of claims 1 to 10, wherein the phase difference and the side prediction data are quantized.
12. A non-transitory computer-readable storage medium recording a program of instructions that is executable by a device to perform the method of any of claims 1 to 11.
13. An apparatus (1100) for encoding a stereo audio signal in a bitstream, the apparatus (1100) comprising:
- an electronic processor (1101) configured to:
 - pass (705) a left-channel and a right-channel of the stereo audio signal to a complex-valued filter bank analysis block (205) to responsively generate one or more frequency bands;
 - calculate (710), for each of the one or more frequency bands, an energy of the left-channel, an energy of the right-channel, and a covariance of the left-channel and the right-channel;
 - select (720), based on the calculated energy of the left-channel, the calculated energy of the right-channel, and the calculated covariance of the left-channel and the right-channel, a stereo coding mode in which to encode the left-channel and the right-channel; and
 - when the stereo coding mode is an extended Mid/Side coding mode (750):
 - compute a phase difference between the left-channel and the right-channel;
 - adjust (740) phase alignment between the left-channel and the right-channel based on the phase difference to generate an aligned left-channel and an aligned right-channel;
 - transform (745) the aligned left-channel and the aligned right-channel to a Mid signal and a Side signal;
 - generate (750) a residual signal based on side prediction data and the Side signal; and
 - encode (755) the Mid signal, the residual signal, and the side prediction data in the bitstream; and
 - output the bitstream for the selected stereo coding mode.
14. The apparatus (1100) of claim 13, wherein the electronic processor (1101) is configured to:
- when the stereo coding mode is a separated coding mode:
 - encode (725) the left-channel and the right-channel in the bitstream.

15. The apparatus (1100) of any of claims 13 to 14, wherein the electronic processor (1101) is configured to:
- when the stereo coding mode is a Mid/Side coding mode:
 - transform (730) the aligned left-channel and the aligned right-channel to the Mid signal and the Side signal; and
 - encode (735) the Mid signal and the Side signal in a bitstream.
16. The apparatus (1100) of any of claims 13 to 15, wherein adjusting the phase alignment between the left-channel and the right-channel includes adjusting a phase of the right-channel to align the left-channel and the right-channel.
17. The apparatus (1100) of any of claims 13 to 16, wherein, to compute the phase difference, the electronic processor (1101) is configured to:
- compare the calculated covariance of the left-channel and the right-channel to an energy threshold, and
 - set the computed phase difference to zero when the calculated covariance of the left-channel and the right-channel is less than or equal to the energy threshold.
18. The apparatus (1100) of any of claims 13 to 17, wherein, to select the stereo coding mode, the electronic processor (1101) is configured to:
- determine (715) a bit cost associated with each of a plurality of stereo coding modes based on the calculated energy of the left-channel, the calculated energy of the right-channel, the calculated covariance of the left-channel and the right-channel, and a cost of transmitting the stereo audio signal; and
 - select the stereo coding mode based on the bit cost.
19. The apparatus (1100) of claim 18, wherein, to determine (715) the bit cost associated with each of the plurality of stereo coding modes, the electronic processor (1101) is configured to:
- determine an energy ratio of signals included in each of the plurality of stereo coding modes; and
 - compare the energy ratio to a threshold,
 - wherein the bit cost indicates a bit reduction between each of the plurality of stereo coding modes compared to coding the left-channel and the right-channel, and
 - wherein the threshold is based on the calculated energy of the left-channel and the calculated energy of the right-channel.

20. The apparatus (1100) of any of claims 13 to 19, wherein the electronic processor (1101) is configured to signal the stereo coding mode that is selected using two bits per frame.

21. The apparatus (1100) of claim 20, wherein the electronic processor (1101) is configured to signal the presence of the phase difference or the side prediction data using one bit per frame each.