



- (51) **International Patent Classification:**
H04R 3/00 (2006.01)
- (21) **International Application Number:**
PCT/US2013/033913
- (22) **International Filing Date:**
26 March 2013 (26.03.2013)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
13/445,769 12 April 2012 (12.04.2012) US
13/452,494 20 April 2012 (20.04.2012) US
- (71) **Applicant:** DTS LLC [US/US]; 5220 Las Virgenes Road, Calabasas, CA 91302 (US).
- (72) **Inventor:** TRACEY, James; 29562 Teracina Blvd., Laguna Niguel, CA 92677 (US).
- (74) **Agent:** KING, John, R.; Knobbe, Martens, Olson & Bear, LLP, 2040 Main Street, 14th Floor, Irvine, CA 92614 (US).
- (81) **Designated States** (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM,

AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

- (84) **Designated States** (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

(54) **Title:** SYSTEM FOR ADJUSTING LOUDNESS OF AUDIO SIGNALS IN REAL TIME

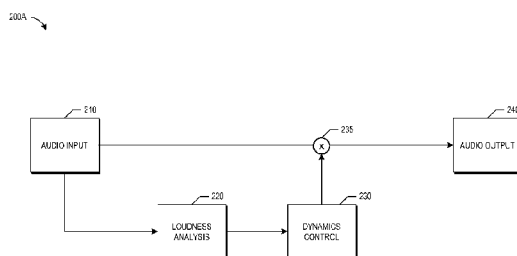


FIG. 2A

(57) **Abstract:** A method of adjusting a loudness of an audio signal in real time may include receiving an electronic audio signal and dividing the audio signal into a plurality of frames. Processing of the frames may be performed in real time. The processing may include measuring initial loudness values for blocks of samples in a frame to produce a plurality of initial loudness values, and computing a weighted average of at least some of the initial loudness values. The weights may be selected based on one or more of the recency of the initial loudness values, variation of the initial loudness values, and estimated information content of the audio signal. The processing may further include selectively discarding at least some of the loudness values that reach an adaptive loudness threshold. Weights can be adjusted based on the variation of the loudness values of the audio signal.

SYSTEM FOR ADJUSTING LOUDNESS OF AUDIO SIGNALS IN REAL TIME

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to U.S. Patent Application No. 13/445,769, filed on April 12, 2012, and entitled “System for Adjusting Loudness of Audio Signals in Real Time,” and U.S. Patent Application No. 13/452,494, filed on April 20, 2012, and entitled “System for Adjusting Loudness of Audio Signals in Real Time,” the disclosures of which are hereby incorporated by reference in their entirety.

BACKGROUND

[0002] One complaint voiced by many television viewers is the changes in volume viewers endure during commercial breaks and when switching between different channels. Similar volume extremes may also occur with other devices, such as portable audio players, A/V receivers, personal computers, and vehicle audio systems. One solution for this problem is automatic gain control (AGC). A typical automatic gain control (AGC) works by reacting to volume changes by cutting an audio signal at high amplitude and then boosting it at low amplitude—no matter where in the frequency range the loudness spike occurs.

[0003] When the AGC kicks in, unwanted changes and unnatural artifacts can often be heard in the form of pumping and breathing fluctuations. Pumping fluctuations can be the result of bass tones disappearing when the loudness suddenly increases, like during a loud action sequence. Breathing fluctuations can happen when low level hiss is boosted during quiet passages. Unfortunately, this brute force method of handling volume changes does not take into account how humans actually perceive change in volume.

SUMMARY

[0004] In certain embodiments, a method for adjusting loudness of an audio signal can include receiving an audio signal and dividing the audio signal into a plurality of frames. The method can include for a frame of the plurality of frames, by one or more processors, measuring initial loudness values for blocks of samples in the frame to produce a plurality of initial loudness values. The

method can also include computing a weighted average of at least some of the initial loudness values in the frame to produce an adjusted loudness value for the frame. The weights in the weighted average can be selected based at least partly on one or both of recency of the initial loudness values and variation of the initial loudness values. The method can further include adjusting loudness of the audio signal based at least partly on the adjusted loudness value.

[0005] In various embodiments, a system for adjusting loudness of an audio signal can include a loudness analysis module comprising one or more processors. The loudness analysis module can access a frame of an audio signal, the frame comprising a plurality of samples of the audio signal and measure loudness values for windows of the samples in the frame to produce a plurality of loudness values. The loudness analysis module can further compute a weighted average of at least some of the loudness values to produce an adjusted loudness value. The weights in the weighted average can be selected based at least partly on one or more of the following: timing of the loudness values and deviation of the loudness values. The system for adjusting loudness can also include a dynamics control module configured to adjust loudness of the audio signal based at least partly on the adjusted loudness value.

[0006] In certain implementations, a method of adjusting loudness of an audio signal can include sampling an audio signal to produce blocks of samples, measuring initial loudness values for the blocks of samples in the audio signal to produce a plurality of initial loudness values, and computing gain values to be applied to the audio signal based at least partly on the initial loudness values. The method can also include computing a weighted average of the initial loudness values or of the gain values. The weights in the weighted average can be selected based at least partly on one or both of recency and variation of the initial loudness values or the gain values. The method can further include adjusting loudness of the audio signal based at least partly on the computed gain values and the weighted average.

[0007] For purposes of summarizing the disclosure, certain aspects, advantages and novel features of the inventions have been described herein. It is to be understood that not necessarily all such advantages may be achieved in accordance with any particular embodiment of the inventions disclosed herein.

Thus, the inventions disclosed herein may be embodied or carried out in a manner that achieves or optimizes one advantage or group of advantages as taught herein without necessarily achieving other advantages as may be taught or suggested herein.

BRIEF DESCRIPTION OF THE DRAWINGS

[0008] Throughout the drawings, reference numbers may be re-used to indicate correspondence between referenced elements. The drawings are provided to illustrate embodiments of the inventions described herein and not to limit the scope thereof.

[0009] FIGURE 1 illustrates an embodiment of a broadcast system;

[0010] FIGURE 2A illustrates an embodiment of a system for adjusting the loudness of audio signals in real time;

[0011] FIGURE 2B illustrates another embodiment of a system for adjusting the loudness of audio signals in real time;

[0012] FIGURE 3A illustrates an embodiment of a system for determining loudness of audio signals in real time;

[0013] FIGURE 3B illustrates an embodiment of a frequency weighting filter;

[0014] FIGURES 3C-3D illustrate example frequency responses of filters of FIGURE 3B;

[0015] FIGURE 4A illustrates an embodiment of a process for determining loudness of audio signals in real time;

[0016] FIGURE 4B illustrates a chart showing example of overlapping segments of audio signals;

[0017] FIGURE 4C illustrates a chart showing example of weighting audio signals;

[0018] FIGURE 5 illustrates a chart showing an example weighting of audio signals;

[0019] FIGURE 6 illustrates a chart showing an example determination of loudness of audio signals in real time;

[0020] FIGURE 7A illustrates an embodiment of a system for determining information content of audio signals in real time;

[0021] FIGURE 7B illustrates an example frequency response of a bank of auditory filters similar to the filters used by the human ear.

DETAILED DESCRIPTION

Overview

[0022] Existing regulations, such as the Commercial Advertisement Loudness Mitigation Act (CALM Act), require broadcast and cable television stations to adopt industry technology standards that ensure that commercials are not louder than regular programming. Some volume control systems attempt to take loudness of an audio signal into account in determining how to vary gain. Loudness can be an attribute of the auditory system that can allow for classification of sounds on a scale from quiet to loud. Loudness can be measured in decibels or other units (such as the phon).

[0023] Audio and audiovisual programming, such as television programming, Internet streaming, and the like can be provided or streamed by content providers to a broadcasting center (which can be referred to as a broadcast head end). Some programming may be live, and such programming is broadcast to viewers live. Other programming, although not live programming, is streamed to the broadcasting center, and can be processed there in real time. Typically, compressed programming is provided, although programming may be provided in any form, such as uncompressed or a mix of compressed and uncompressed programming. The programming can include commercials, and loudness of the commercials may need to be adjusted in order to comply with the regulations. It can be advantageous to perform the adjustment in real time without having access to the entire program (e.g., entire episode of a TV show, entire news broadcast, entire movie, etc.)

[0024] This disclosure describes certain systems and methods for adjusting loudness of audio signals in real time. Techniques used by these systems and methods can include performing weighting of loudness measurements based on various criteria to more accurately track transient loudness spikes. Additional techniques can include increasing the size of an adaptive analysis window of the audio signal when feasible to reduce the impact of loudness on dynamic range. Further, these techniques can include adaptively

discarding selected loudness measurements to reduce the impact of noise and low-level signals that may represent silence on loudness adjustments. Moreover, techniques for detecting information content of an audio signal can be employed to enable loudness adjustment of the information-bearing portions.

[0025] Advantageously, in certain embodiments, applying one or more of these techniques can preserve or attempt to preserve the transparency of loudness adjustments to the audio signal while preserving the ability to react quickly to changes in the loudness. Transparency can relate to a degree or measure of perceptual difference between the original audio signal and an audio signal with adjusted loudness. Higher transparency can reflect the fact that the audio signal is substantially perceptually indistinguishable from the original audio signal. In some embodiments, applying one or more of the above-described features can also reduce the impact of loudness adjustments on the dynamic range of the audio signal.

[0026] In addition to having its ordinary meaning, the terms “real time,” “real time processing,” and the like can refer to any form of processing other than file-based processing. File-based processing, in the loudness adjustment context, can refer to loudness processing performed on an entire audio file or audio program at once, for example, by calculating an overall loudness of an entire program and adjusting the loudness accordingly. As another example, file-based processing can refer to loudness processing performed offline, such as in the context of video on demand (VOD) processing, audio and video on demand (AVOD) processing, etc. The terms “real time,” “real time processing,” and the like can, but need not, necessarily refer to processing performed without a perceptible delay from the point of view of a listener. Further, these terms can be need not refer to programming that is broadcast live. Rather, any programming can be broadcast and therefore adjusted for loudness in real time, including programming that has been stored as a file.

Example Loudness Adjustment System

[0027] Referring to **FIGURE 1**, an embodiment of a broadcast system 100 is shown. As is shown, one or more content providers 110, 111, and 112 deliver television programming to a broadcast head end 120. In some

embodiments, the broadcast head end 120 can be a facility for receiving television signals for processing and distribution over a television system. Processing may include recoding encoded and/or compressed content received from a content provider, aggregating or combining content for distribution downstream, inserting commercials, and so on. Broadcast head end 120 can operate automatically without user intervention or operate with user intervention. Content processed by the broadcast head end 120 can be distributed to one or more regional head ends 130, 131, and 132. A regional head end can perform further processing of the content, including insertion of commercials targeted for a specific region or locale. As is illustrated, the regional head end distributes final programming to one or more customers 140-145. The broadcast head end 120 and/or regional head ends 130-132 may include one or more processing devices, such as receivers and/or servers.

[0028] In some embodiments, the broadcast head end 120 broadcasts programming in real time, near-real time, or the like. For example, the broadcast head end 120 may not have access to the entire program before operating on the program, instead broadcasting the program as it is received from a content provider. As is explained above, storing or buffering the entire program (e.g., as a digital file) before processing the program can be considered a file-based operation. Processing performed by the regional head ends may be file-based, real time, or a combination of file-based and real time processing.

[0029] Referring to **FIGURE 2A**, an embodiment of an audio processing system 200A for adjusting the loudness of audio signals in real time is shown. The audio processing system 200A may be implemented in any machine that processes and/or reproduces audio, such as a head end device (e.g., a receiver, server, computer, and the like) or a user device (e.g., a computer, set top box, television, portable audio player, headphone, A/V receiver, vehicle audio system, and the like). Although in some embodiments the audio processing system 200A is implemented by a broadcast head end 120, the audio processing system 200A can be implemented by a regional head end 130-132 or any other machine that processes and/or reproduces audio. Advantageously, in certain embodiments, the audio processing system 200A adjusts a loudness of an audio input signal 210 to attempt to maintain the loudness at a certain level. For

example, the audio processing system 200A may attempt to maintain a certain loudness of television audio when a commercial begins broadcasting. The audio processing system 200A can perform these functions in real time.

[0030] As shown in FIGURE 2A, the audio processing system 200A receives an audio input signal 210. The audio input signal 210 can include one or more channels. For instance, two channels, such as stereo channels, 5.1, 6.1, or 7.1 surround sound channels, or matrix encoded channels such as Circle Surround encoded channels or the like may be provided. The audio input signal 210 can be an electrical signal or the like that represents a real, physical sound, such as music, voice, effects, combinations of the same, and the like.

[0031] The loudness analysis module 220 can include hardware and/or software for determining or estimating loudness of the audio input signal 210 in real time. In certain embodiments, the loudness analysis module 220 implements one or more loudness estimating techniques described in the following standards: Advanced Television Systems Committee, Inc.'s ATSC A/85, International Telecommunications Union's ITU BS.1770-1, and International Telecommunications Union's ITU BS.1770-1-2. Each of the ATSC A/85, ITU-R BS.1770-1, and ITU-R BS.1770-1-2 standards are hereby incorporated by reference in their entirety. For example, ITU-R BS.1770-1 and BS.1770-2 standards disclose determining loudness by integrating (or summing in the discrete time domain) weighted power of the audio signal channels over the duration of the audio signal, among other techniques that can be used by the loudness analysis module 220.

[0032] The dynamics control module 230 can include hardware and/or software for comparing the estimated loudness provided by the loudness analysis module 220 with a reference or target loudness level. The target loudness level can be a reference that is internal to the audio processing system 200A. For example, the target level can be full scale loudness (e.g., 0 dB), so that adjusting the loudness to this level preserves the dynamic range of the audio output signal 240. In another embodiment, the target level can be any level chosen by an operator or user at the head end. If the estimated loudness differs from the target loudness level, the dynamics control module 230 can determine the level difference between the estimated loudness and the target loudness

level. The dynamics control module 230 can further use this difference to adjust a gain applied to the audio input signal 210. The audio processing system 200A outputs an audio output signal 240, which has its loudness adjusted. The audio output signal 240 may be distributed to regional head ends 130-132 (in case the audio processing system 200A is implemented by the broadcast head end 120), or the audio output signal 240 may be distributed directly to customers (in case the audio processing system 200A is implemented by a regional head end 130-132).

[0033] The dynamics control module 230 can apply the loudness level difference (e.g., reflected as the gain) to the audio signal input 210 on a sample by sample basis via a mixer 235. For example, the dynamics control module 230 can compute the multiplicative inverse of the loudness level to determine the gain and apply the gain to the audio signal input 210. In certain embodiments, when more than one channel of audio input is provided, the dynamics control module can utilize more than one mixer 235 in order to apply the level difference to more than one channel. In certain embodiments, the dynamics control module 230 smoothes transitions between samples or blocks of samples of the audio input signal 210 to prevent jarring loudness transitions. In other embodiments, such smoothing is performed by weighing the loudness values as is described below. As a result, the mixer 235 may output or attempt to output an audio signal 240 that has a constant average loudness level or substantially constant average loudness level. In other embodiments, the loudness level of the output audio signal 240 may not be constant, but may still have less variability than the audio signal input 210. Thus, in certain embodiments, the audio processing system 200A can transform the audio input signal 210 into an audio output signal 240 that has a less variable loudness level than the audio input signal.

[0034] In some embodiments, a pre-processing module (not shown) may be included by the audio processing system 200A. The pre-processing module can be configured to receive the audio input signal 210. The pre-processing module may include hardware and/or software for gathering energy information from a channel (or each channel) of the audio input signal 210. The pre-processing module can be further configured to examine noise characteristics of the channel(s). Using the energy information and/or noise

characteristics in one embodiment, the pre-processing module can determine at least one dominant channel to be analyzed for loudness by a loudness analysis module 220. More generally, the pre-processing module may select a subset of the channels of the audio input signal 210 for loudness analysis. In certain embodiments, using fewer than all of the channels to determine loudness can reduce computing resources used to determine loudness.

[0035] FIGURE 2B illustrates another embodiment of an audio processing system 200B for adjusting the loudness of audio signals in real time. The audio processing system 200B may be implemented in any machine that processes and/or reproduces audio, such as a head end device (e.g., a receiver, server, computer, and the like) or a user device (e.g., a computer, set top box, television, portable audio player, headphone, A/V receiver, vehicle audio system, and the like). Although in some embodiments the audio processing system 200B is implemented by a broadcast head end 120, the audio processing system 200B can be implemented by a regional head end 130-132 or any other machine that processes and/or reproduces audio. Advantageously, in certain embodiments, the audio processing system 200B adjusts a loudness of an audio input signal 210 to maintain or attempt to maintain the loudness at a certain level. For example, the audio processing system 200B may attempt to maintain a certain loudness of television audio when a commercial begins broadcasting. The audio processing system 200B can perform these functions in real time.

[0036] As is described above, the audio processing system 200A receives an audio input signal 210. Further, as is described above, a pre-processing module (not shown) may be included in some embodiments. The dynamics control module 230 can include hardware and/or software for comparing a gated measured loudness 365 provided by the loudness analysis module 220 with a reference loudness level. The reference loudness level can be a reference that is internal to the audio processing system 200B. For example, the reference level can be set by a head end operator, such as to -24 dB, -20 dB, 0 dB (or full scale loudness, so that adjusting the loudness to this level preserves the dynamic range of the audio output signal 240), or another suitable value. If the estimated loudness differs from the reference loudness level, the dynamics control module 230 can determine the level

difference between the estimated loudness and the reference level. The dynamics control module 230 can further use this difference to adjust a gain to be applied to the audio input signal 210.

[0037] The dynamics control module 230 can provide the determined gain or the level difference to a limiter module 238. The limiter module 238 prevents or reduces clipping of the audio input signal 210 by reducing the computed gain to be applied to the signal 210 so that the peak amplitude of the gain-adjusted signal may be below a clipping threshold. The audio processing system 200B outputs the audio output signal 240, which has its loudness adjusted. The audio output signal 240 may be distributed to regional head ends 130-132 (in case the audio processing system 200B is implemented by the broadcast head end 120), or the audio output signal 240 may be distributed directly to customers (in case the audio processing system 200B is implemented by a regional head end 130-132). The limiter module 238 can apply the loudness level difference (e.g., reflected as the gain) to the audio signal input 210 on a sample by sample basis via the mixer 235. In certain embodiments, e.g., when more than one channel of audio input is provided, the limiter module 238 can utilize more than one mixer 235 in order to apply the level difference to more than one channel.

[0038] In certain embodiments, the dynamics control module 230 and/or the limiter module 238 smoothes transitions between samples or blocks of samples of the audio input signal 210 to prevent jarring loudness transitions. In other embodiments, such smoothing is performed by weighing the loudness values as is described below. As a result, the mixer 235 may output or attempt to output an audio signal 240 that has a constant average loudness level or substantially constant average loudness level. In other embodiments, the loudness level of the output audio signal 240 may not be constant, but may still have less variability than the loudness of the audio signal input 210. Thus, in certain embodiments, the audio processing system 200B can transform the audio input signal 210 into an audio output signal 240 that has a less variable loudness level than the audio input signal.

[0039] The framing module 215 can include hardware and/or software for buffering a segment of the audio input signal 210 for real time processing. In

some embodiments, the audio input signal 210 is a stream of a television program, an Internet stream, etc. and the framing module 215 can buffer a segment of the television program. The segment may contain a plurality of samples. In certain embodiments, the length of the frame is 12 seconds or approximately 12 seconds, although different frame lengths can be utilized, such as less than 12 seconds or more than 12 seconds, e.g., 20 seconds, 30 seconds, 45 seconds, 60 seconds, etc. In some embodiments, the length of the frame can be adjusted or varied as is described below.

[0040] The loudness analysis module 220 can include hardware and/or software for determining or estimating loudness of the audio input signal 210 in real time. In certain embodiments, the loudness analysis module 220 implements one or more loudness estimating techniques described in the ITU-R BS.1770-2 standard, incorporated above. With reference to **FIGURE 3A**, in some embodiments, the loudness analysis module 220 can include a system 300 for determining loudness of audio signals in real time. As is illustrated, the loudness determination system 300 can operate on multiple channels 305 of the framed audio input signal 210, such as left channel (L), right channel (R), center channel (C), left surround (Ls), and right surround (Rs). In other embodiments, the loudness determination system 300 can operate on fewer or greater number of channels and/or on channels different than those illustrated in **FIGURE 3A**.

[0041] The loudness determination system 300 applies filtering 310 of the audio input signal 305. As is illustrated, filtering 310 can be applied to each channel 305 of the framed audio input signal 210. In certain embodiments, filtering 310 performs two-stage frequency weighting of the audio input signal 305, as is illustrated in **FIGURE 3B**. The first filtering stage 311 can account for the acoustic effects of a human head modeled as a rigid sphere. The impulse response 311' of the first stage filter 311 is shown in **FIGURE 3C**. As is illustrated, in one embodiment, first stage filter 311 is a shelving filter or an equalizer configured to boost higher frequencies (e.g., provide treble boost). The coefficients of the first stage filter 311 are provided on page 4 of ITU-R BS.1770-1-2 standard (03/2011 edition). With reference to **FIGURE 3B**, the second filtering stage 312 applies a revised low-frequency B-weighting weighting curve (RLB weighting curve). The impulse response 312' of the second stage RLB

weighting filter 312 is shown in **FIGURE 3D**. As is illustrated, in one embodiment, second stage filter 312 is a high pass filter. The coefficients of the second stage filter 312 are provided on page 5 of ITU-R BS.1770-1-2 standard (03/2011 edition).

[0042] With reference to FIGURE 3A, filtered channel(s) 315 can be provided to the power calculator module 320, which computes the mean square value(s) 325 or average power of the filtered input signal. In certain embodiments, the power calculator module 320 computes the mean square value according to the following equation or some variation thereof:

$$z_i = \frac{1}{N} \sum_0^{N-1} y_i^2 \quad (1)$$

[0043] where N is the duration or interval (e.g., number of samples in one frame) of the audio input signal 210. The loudness determination system 300 can calculate initial measured loudness 345 over the length of the frame (which can be measured by the interval N) according to the following equation or some variation thereof (including a logarithmic version of at least a portion thereof):

$$Loudness, L = -0.691 + \sum_0^{N-1} G_i \cdot z_i \quad (2)$$

[0044] where G_i are weighting coefficient(s) 330 for the channels 325. Thus, initial measured loudness 345 can be calculated by summing 340 the mean square value(s) 325 and (optionally) scaling the sum by a constant (-0.691). In one embodiment, weighting coefficients 330 can be selected as follows: $G_L = 1.0$ (0 dB), $G_R = 1.0$ (0 dB), $G_C = 1.0$ (0 dB), $G_{LS} = 1.41$ (~ 1.5 dB), and $G_{RS} = 1.41$ (~ 1.5 dB). Other suitable values of weighting coefficients 330 may be used. In one embodiment, initial measured loudness 345 can be expressed in units called "LKFS," which refer to loudness, K-weighted, relative to full scale. If a logarithm is taken of the summation in equation (2), the loudness can be expressed in decibels. For example, if a 0 dB full-scale sine wave with frequency of 1 kHz is used as the audio input 210, its loudness will equal -3.01 LKFS. However, the loudness need not be expressed in decibels in some embodiments.

[0045] To calculate the gated measured loudness 365 in some embodiments, the loudness determination system 300 can divide or partition each frame of the audio input signal 210 into a set of overlapping windows or blocks. In one embodiment, the windows can be of equal length (although this is not required), and the overlap can be 75%. This is illustrated in **FIGURE 4B**, which shows a chart 420 of overlapping windows of audio signals according to one embodiment. The windows 423, 424, 425, and 426 overlap by 75% of their length or approximately 75% of their length. In another embodiment, a different overlap may be used, such as 10%, 25%, 50%, or the like. In an embodiment, the duration of windows 423, 424, 425, and 426 is 400 msec or approximately 400 msec to the nearest sample. In another embodiment, different duration can be selected, such as 100 msec, 150 msec, 200 msec, and so on. The selection of window duration and/or the overlap may depend on the sampling rate of the audio input signal 210. For example, 400 msec may be a suitable window length for 48 kHz sampling rate of the audio input signal 210. In certain embodiments, as is described below, filtering 310 and power calculation 320 operations can be performed on the overlapping windows. In other words, the loudness determination system 300 can determine initial measured loudness values 345 for each of the overlapping windows. In certain embodiments, frames of the audio input signal 210 can also be overlapped as described above, such that sliding frames of the audio input signal 210 are analyzed.

[0046] The power calculator module 320 can compute the mean square values for each overlapping window according to the following equation:

$$z_{ij} = \frac{1}{N} \sum_{Ng \cdot j \cdot step}^{Ng \cdot (j \cdot step + 1)} y_i^2 \quad (3)$$

[0047] where z_{ij} is the mean square value of the j th interval (or frame) of the i th channel of the audio input signal 210, step equals to $(1 - \text{overlap})$, N_g is duration (e.g., number of samples) in a sub-frame, and j is selected from the set of values $\left[0, 1, 2, \dots, \frac{N - N_g}{N_g \cdot step}\right]$. The initial measured loudness 345 of j th sub-frame can be calculated according to the following equation or the like:

$$L_j = -0.691 + \sum_0^{N-1} G_i \cdot z_{ij} \quad (4)$$

Adaptive Gating

[0048] The adaptive or dynamic gating module 360 of FIGURE 3A performs adaptive or dynamic gating of initial measured loudness 345. In certain embodiments, adaptive gating 360 employs a two-stage process. **FIGURE 4A** illustrates a process 400 for determining loudness in real time which can be performed by the adaptive gating module 360. With reference to FIGURE 4A, a frame of the audio input signal 210 is partitioned or divided into windows in block 402. In block 404, the process 400 measures or calculates the instantaneous loudness L_{wk} for each window in a frame. In one embodiment, this calculation can be performed utilizing equation (2) and/or equation (4). In an embodiment using 400 msec windows that overlap by 75% are used, a new loudness value L_{wk} may be computed every 100 msec. A first stage of adaptive gating 360 is performed in block 406, where loudness of each window is compared to a loudness threshold. The loudness threshold can be selected to reflect to discard loudness values that are likely to be periods of silence and/or low-level noise. In one embodiment, the loudness threshold can be selected as -70 dB (or -70 LKFS) or approximately -70 dB. In other embodiments, different values of the loudness threshold may be used, such as -80 dB or smaller, -65 dB or smaller or larger, and so on. This first stage of adaptive gating 360 can be referred to as absolute thresholding because the loudness threshold may be the same for most or all windows.

[0049] In block 406, the process 400 discards (e.g., excludes from further processing) windows for which loudness falls below the loudness threshold. This discarding can be performed because such windows may contain audio information that cannot be discerned by a listener, and such audio periods should not be taken into account when computing loudness of the audio input signal 210. In block 408, the process 400 determines mean loudness of the frame of the audio input signal 210. In one embodiment, the process 400 utilizes the following equation or the like for determining the mean loudness:

$$\bar{L} = \frac{1}{M} \sum_{k=0}^{M-1} L_{wk} \quad (5)$$

[0050] where M is the number of overlapping windows in the frame and L_{Wk} is the measured loudness of a k th window.

[0051] In certain embodiments, a second stage of adaptive gating 360 involves performing an adaptive or dynamic relative gating operation (which can be signal dependent). Advantageously, adaptive relative gating can account at least in part for transient events, which can reflect sudden, large changes in the loudness of the audio input signal 210. For example, in the audio signal, a loud portion (e.g., explosion) or vice versa may be immediately followed by a relatively quiet portion (e.g., whisper). As another example, a commercial may be inserted into an audio stream, and the commercial may have relative loudness that is larger than that of the content of the audio stream. It is desirable for the audio processing system 200A and/or 200B to track such transient events (e.g., high loudness to low loudness transitions) and take the differences in loudness caused by such events into account when adjusting loudness of the audio input signal. One reason is that doing so can provide greater transparency and preserve the dynamic range of the signal.

[0052] Another reason for using adaptive relating gating is that during real time processing of the audio signal, the audio processing system 200A and/or 200B may not have access to the entire audio program (e.g., entire television program, entire commercial, etc.). Instead, processing is performed on segments of the audio stream, and such segments may include audio content with different loudness values, such as part of a television program and part of a commercial. If the entire audio stream were available, determining loudness of such combined or aggregated content using static loudness correction techniques (e.g., file-based adjustment) could be performed separately for each different component. For example, as is described in ITU BS.1770-2, gated loudness L_{KG} can be determined by using a threshold obtained from subtracting a constant (e.g., 10 LKFS) from loudness measurements that are above an absolute threshold (e.g., -70 LKFS). However, when such static correction techniques are applied in real time, inaccurate loudness values may be determined. Basing loudness correction on such inaccurate values could adversely affect the loudness correction of the audio input signal 210.

[0053] Yet another reason for using adaptive relative gating can be maintenance of suitable responsiveness of the audio processing system 200A and/or 200B. While it may be generally desirable to perform loudness adjustment so that periods of silence do not substantially impact the adjustment, it can be advantageous for the audio processing system to quickly react to changes in the loudness (e.g., transitions from high loudness to low loudness) and adjust the loudness accordingly. For example, when beginning of a relatively quiet portion is played immediately after a loud commercial, the audio processing system 200A and/or 200B can advantageously recognize the abrupt change in loudness and quickly react by adjusting (e.g., increasing) the loudness of the audio input signal 210.

[0054] The process 400 performs adaptive gating 360 in blocks 410 and 412. In block 410 in one embodiment, the process can dynamically determine a deviation of the mean loudness of a current frame (e.g., computed in block 408) from the last determined gated measured loudness 365. A large deviation may indicate an onset of a transient event, which the process 400 may need to quickly react to. The deviation can be determined by using the following equations or the like:

$$\begin{aligned} S &= \bar{L} - C_1 - \Delta \\ \Delta &= \sum (L_{Wk} - L_{GMR}) \end{aligned} \quad (6)$$

[0055] where L_{Wk} is the loudness of windows in frames (as is measured in block 404) is the mean loudness value of the current frame, C_1 is a constant (e.g., 10 dB or another suitable value), and Δ is an accumulated difference between the instantaneous loudness value (e.g., calculated in block 404) and previously determined gated measured loudness L_{GMR} (e.g., determined for an immediately preceding frame). In some embodiments, the accumulated difference Δ can be bounded by lower bound and/or upper bound. For example, the lower bound can be 0 (thereby setting the upper range of adaptive gating to the constant C_1) or another suitable value smaller or greater than 0 (e.g., -10, -5, -1, 1, 2, 5, 10, etc.) As another example, the upper bound can be -40 dB or another suitable value smaller or greater than -40 dB (e.g., -20 dB, -25 dB, -30 dB, -35 dB, -45 dB, -50 dB, -55 dB, etc.). Thus, in one embodiment, the difference between the loudness output and the adaptive relative gate can range

from about -10 dB (e.g., per the BS 1770-2 specification) to about -40 dB (unlike the BS 1770-2 specification). Other ranges are also possible. In certain embodiments, the deviation can be determined by using the following equations or the like:

$$\begin{aligned} S &= \bar{L} - C_1 - \Delta \\ \Delta &= \sum (\bar{L} - L_{GMR}) \end{aligned} \quad (7)$$

[0056] where $\bar{L}$ is the mean loudness value of the current frame, C_1 is a constant (e.g., 10 dB or another suitable value), and Δ is an accumulated difference between the mean loudness value and previously determined gated measured loudness L_{GMR} (e.g., determined for an immediately preceding frame). Deviation Δ of equation (7) can be bounded as is described above. In one embodiment, deviation Δ of equations (6) and/or (7) can represent the difference between the instantaneous loudness or mean loudness value respectively and previously determined gated measured loudness 365, without accumulating the difference. At initialization, the previously determined gated measured loudness 365 can be set to a suitable initial value, such as the mean loudness (e.g., computed in block 408).

[0057] In some embodiments, deviation Δ can be an accumulated difference between the instantaneous loudness L_{wk} (e.g., determined in block 404) and mean loudness value of the current frame (or preceding frame, such as immediately preceding frame) $\bar{L}$. Deviation Δ can be bounded as is described above. In one embodiment, deviation Δ can represent the difference without accumulation.

[0058] In block 412, the process 400 discards from calculation windows of a frame for which loudness (e.g., calculated in block 404) is below the adaptive relative gating loudness of the frame determined in block 410. Such windows may contain audio information that cannot be discerned by a listener (e.g., periods of silence and/or noise), and such audio periods should not be taken into account when computing loudness of the audio input signal 210.

[0059] In some embodiments, the second stage of adaptive gating 360 is not performed. Instead, loudness values that have not been discarded in block 406 are utilized for computing the weighted average in block 414. In certain embodiments, the first and second stages of adaptive gating 360 are not

performed. Instead, loudness values measured for each window in block 404 are utilized in computing the weighted average in block 414.

Weighting Loudness Values

[0060] In block 414, the process 400 determines gated measured loudness 365 of the frame (or a window, part of a window, etc.). In some embodiments, the process 400 computes a weighted average of loudness values for windows that were not excluded from calculation in blocks 406 or 412. In one embodiment, the process 400 performs temporal weighting (e.g., weighting based on recency of the samples within a window and/or windows). Temporal weighting may be performed in cases when windows of the audio signal are overlapped, as is described above and illustrated in FIGURE 4B. Various types of weighting can be utilized. **FIGURE 4C** shows a chart 430 with two exemplary weighting schemes 432 and 434. As is illustrated, the weighting scheme 432 is non-linear and assigns progressively greater weights (computed non-linearly) to more recent samples and/or windows. The weighting scheme 432 can, for example, utilize the following equation or the like for computing the weights:

$$W_{Tk} = \alpha^k \quad (8)$$

[0061] where α is a constant and k is a sample index (within a window) and/or window index. The value of α can be selected so that a suitable emphasis is placed on more recent samples and/or windows. In one embodiment, α can be selected as a value greater than 1 (which can emphasize more recent samples and/or windows in relation to less recent samples and/or windows). In another embodiment, α can be selected as a number less than 1 (which can deemphasize more recent samples and/or windows in relation to less recent samples and/or windows).

[0062] As is illustrated, the weighting scheme 434 is linear, and assigns progressively greater weights (computed linearly) to more recent samples and/or windows. The weighting scheme 434 can, for example, utilize the following equation or the like for computing the weights:

$$W_{Tk} = C_2 k + C_3 \quad (9)$$

[0063] where C_2 and C_3 are constants and k is a sample (within a window) and/or window index. In one embodiment, the slope of the line 434 can be varied (e.g., by selecting a suitable value of C_2) in order to place a suitable emphasis on more recent samples and/or windows. A combination of linear and non-linear weights can be used in some embodiments.

[0064] In certain embodiments, the process 400 performs the weighting based on deviation of a loudness value (e.g., computed in block 404) from the mean loudness value of the frame (e.g., computed in block 408). This may be advantageous for quickly reacting to transient events. For example, the process 400 can utilize the following equation or the like for computing the weights:

$$W_{Mk} = \beta^{|L_k - \bar{L}|} \quad (10)$$

[0065] where β is a constant, L_k is measured loudness of a k th window (and/or sample within a window), and $\bar{L}$ is the (non-weighted) mean loudness of the frame. In one embodiment, β can be selected as a value greater than 1 (which can emphasize more recent windows and/or samples in relation to less recent windows and/or samples). In another embodiment, β can be selected as a number less than 1 (which can deemphasize more recent windows and/or samples in relation to less recent windows and/or samples). When there is no deviation from the mean loudness, the value of a weight computed according to equation (10) is 1. However, the value of the computed weight increases as there is greater deviation from the mean loudness value. Such windows (or samples) may likely correspond to transient events, and are thus emphasized by equation (10) in the calculation of the gated measured loudness 365.

[0066] In some embodiments, the values of α of equation (8) and/or β of equation (10) can be adjusted based on the statistics of the audio input signal 210. Statistics, for example, can be reflected by the variation (e.g., as measured by the standard deviation or a coefficient of variation) in loudness of the audio input signal 210. Variation of loudness of the audio input signal 210 can be measured using the determined loudness of windows in the frame (e.g., measured in block 404). For example, the following formula for the standard deviation or the like can be utilized for estimating the variation:

$$\sigma_L = \sqrt{\frac{1}{M} \sum_{k=0}^{M-1} (L_{wk} - \bar{L})^2} \quad (11)$$

[0067] where M is the number of overlapping windows in a frame, L_{wk} is loudness of a window, and $\bar{L}$ is the mean loudness of the frame (e.g., computed in block 408).

[0068] In some embodiments, if a portion of the signal is determined to have relatively constant or static loudness (e.g., variation of loudness of the frame is determined to be below a threshold) the values of α and/or β can be decreased to lower or dampen the sensitivity of the loudness adjustment and thereby attain a slower reaction to the changes in the loudness. On the other hand, if the portion of the signal is determined to have relatively dynamic loudness (e.g., variation of loudness of the frame is determined to be above the threshold), the values of α and/or β can be increased to increase the sensitivity of the loudness adjustment and thereby attain faster reaction to changes in the loudness.

[0069] In some embodiments, the values of α and/or β can be varied from a default value of about 5000 (to provide quick reaction to dynamic changes) to about 217 (to provide a more static adjustment). In other embodiments, other suitable upper and lower bounds can be used. For example, if the values of α and β were set to 1, then the weighted average of the loudness values would be an arithmetic mean. The values of α and β may also be lower than 1 in some embodiments. In certain embodiments, different upper and/or lower bounds can be used for the values α and β . In some embodiments, the values of α and β can be varied in a different manner (e.g., with different rates of decrease and/or increase). In some embodiments, the values of α and/or β are reset to the default value(s) (e.g., 5000) when variation of the loudness of the frame is determined to be above the threshold. Advantageously, such resetting may ensure or attempt to ensure that transient events are accounted for by the audio processing system.

[0070] Decreasing the values of α and/or β can result in the audio processing system 200A and/or 200B performing analysis that mimics using static loudness correction techniques or that otherwise applies less aggressive loudness leveling. This may be feasible for audio input signal 210 portions that

have substantially constant loudness, such as certain portions of music. Decreasing the values of α and/or β can preserve transparency of the audio input signal 210 in such cases.

[0071] **FIGURE 5** illustrates a chart 500 showing an example weighting of audio signals according to an embodiment. The x-axis reflects time in seconds, and the y-axis reflects values of the weights (e.g., W_{Mk}). Region 510 (which corresponds in part to region 610 of FIGURE 6) illustrates weights that may be derived for a transient event, such as a sudden, rapid decrease in loudness of the audio input signal 210. As is illustrated in chart 500, values of the weights are accordingly increased to account for the transient event.

[0072] With reference to FIGURE 4, in some embodiments, the process 400 utilizes both weighting schemes described above and reflected in equations (9) and (10).

[0073] **FIGURE 6** illustrates a chart 600 showing determination of loudness of audio signals in real time according to an embodiment. The x-axis reflects time in seconds, and the y-axis reflects loudness intensity in decibels. The chart 600 depicts processing of numerous frames (e.g., having duration of 12 seconds) of the audio input signal 210. The curve 602 represents instantaneous loudness L_{Wk} of windows in frames (as is measured in block 404). The curve 604 represents gated measured loudness of the frames (as is measured in block 414). The curve 606 represents mean loudness $\bar{L}$ of frames (as measured in block 408). The curve 608 represents the adaptive relative gating function (as determined in block 410).

[0074] As is shown in FIGURE 6, the curve 604 (which reflects gated measured loudness of the frames) can closely track the actual loudness of the audio input signal 210, which is reflected by the curve 602. In contrast, the curve 606 (which may, in one embodiment, reflect existing loudness correction techniques), does not track the actual loudness of the audio input signal 210 as closely in regions where the actual loudness undergoes rapid changes. As is shown in region 610, for example, the curve 606 deviates from the actual loudness by a significantly large amount, while the curve 604 more closely tracks the actual loudness in this example.

[0075] In some embodiments, the curve 604 closely tracks the actual loudness of the audio input signal 210 because the curve 608 (which reflects the adaptive relative gating function determined in block 410) accounts for rapid changes in the actual loudness. This is illustrated, for example, in region 610 which captures a transient event, such as a sudden, rapid decrease in loudness of the audio input signal 210. The transient event is reflected by the curve 602 decreasing from loudness above -10 dB at point 612 to loudness below -50 dB at point 614. This drop in loudness occurs over a short period of time, and may indicate beginning of a quiet portion of a television program. As is shown in the example scenario of FIGURE 6, loudness increases following the drop in point 614, which may indicate the end of the transient event. Because the adaptive relative gating function represented by the curve 608 is generally below the curve 602, the transient event is accounted for by the process 400. However, existing loudness determination techniques, which gate based on subtracting a constant (e.g., -10 dB) from the curve 606, would likely miss this transient event. In other words, such static loudness determination techniques (e.g., file-based adjustment) would likely discard this transient event from the loudness calculation. Hence, the transparency of loudness-adjusted audio input signal 210 would be adversely affected.

Adaptive Frame Length

[0076] In some embodiments, frame duration or length can be varied by the audio processing system 200A and/or 200B (e.g., instead of adjusting the values of α of equation (8) and β of equation (10)). For instance, duration of the frame can be adjusted based on the variation (e.g., as measured by the standard deviation or a coefficient of variation) in loudness of the audio input signal 210. Variation of loudness of the audio input signal 210 can be measured using determined loudness of windows in the frame (e.g., measured in block 404). For example, equation (11) can be utilized for estimating the variation.

[0077] In some embodiments, if the variation of loudness of the frame is determined to be below a threshold, the length of the frame is increased. In other words, relatively small variations of loudness may reflect that loudness of the portion of the audio input signal 210 covered by the frame is relatively static.

A limit on the length of the frame can be set, such as 48 seconds or another suitable value.

[0078] Increasing the frame duration (or widening the frame) may improve processing performance and save storage space. In addition, increasing the frame duration results in the audio processing system 200A and/or 200B performing analysis that mimics using static loudness correction techniques. In other words, loudness adjustment applies a substantially static loudness correction when the frame is widened. This may be possible for audio input signal 210 portions that have substantially constant loudness, such as certain portions of music. Widening the frame can preserve transparency of the audio input signal 210 in such cases.

[0079] In certain embodiments, when variation of loudness of the frame is determined to be above the threshold, the length of the frame is decreased. The rate of decrease of frame length (e.g., decrease step size) may be different than the rate of increase of frame length (e.g., increase step size). For example, the decrease step size may be larger than the increase step size, such as 50 times greater or another suitable value. In some embodiments, the frame length is reset to the initial frame length (e.g., 12 seconds) when variation of loudness of the frame is determined to be above the loudness threshold. Advantageously, such resetting may ensure or attempt to ensure that transient events are accounted for by the audio processing system.

Perceptual Information Content Estimation

[0080] With reference to FIGURE 2B, the perceptual information content estimation module 225 can determine or estimate the amount of perceptual information contained in a portion of the audio input signal 210 (e.g., a frame, a window, partial frame, partial window, a plurality of frames, a plurality of windows, etc.). The perceptual information content estimation module 225 can estimate an information content of the portion of the audio input signal 210. Based on the estimated information content, the perceptual information content estimation module 225 can determine weighting associated with adjusting the loudness of the portion of the audio input signal 210, such as W_{Pk} used in equation (12) described below. In one embodiment, the audio processing system 200B

and/or 200A may not process the portion if its informational content is determined to be below an information content threshold. Advantageously, estimating informational content can improve system performance in some embodiments.

[0081] For example, the portion of the audio input signal 210 (e.g., frame) may contain noise, such as random noise, background noise, or the like, and the perceptual information content estimation module 225 may determine that little or no information content is contained in the portion of the audio input signal 210. In such case, the audio processing system 200B and/or 200A may perform little or no loudness adjustment (e.g., set W_{pk} to zero for the window or entire frame) on the portion as it is unlikely to be perceived by the listener. As another example, the perceptual information content estimation module 225 may determine that a portion of the audio input signal 210 contains a substantial amount of information (e.g., the portion may be music or speech). In such case, the weighting may be set so that the portion is suitably emphasized for the purposes of loudness adjustment (e.g., set W_{pk} to a large value for the window or entire frame). For instance, weighting can be proportional to the difference between determined information content and the information content threshold. For example, audio signal portions with higher information content can be assigned greater weights.

[0082] **FIGURE 7A** illustrates an embodiment of perceptual information content estimation module 225. A Fourier transform module 710 transforms a portion of the audio input signal 210 to the frequency domain. For example, the Fourier transform module 710 can perform a fast Fourier transform (FFT) of the portion of the audio input signal 210. Any suitable duration of the portion of the audio input signal 210 can be used, such as 100 msec, 150 msec, 200 msec, etc. Frequency transformed audio input portion is subjected to band partitioning by the module 720. In one embodiment, band partitioning module 720 performs critical band partitioning of transformed audio input by utilizing a bank of filters that mimics the auditory response of a human ear. For example, the band partitioning module 720 can use overlapping, non-linearly spaced bandpass filter, such as logarithmically-spaced bandpass filters. An example of such filter bank 800 is illustrated in **FIGURE 7B**, which illustrates a bank of overlapping, non-

linearly spaced band-pass filters 802. As is shown in FIGURE 7B, the center frequencies of band-pass filters are spaced non-linearly, and this spacing increases as frequency increases. Other suitable filter banks can be used, such as the gammatone approximation filters described in U.S. Patent Publication No. 2009/0161883, titled "SYSTEM FOR ADJUSTING PERCEIVED LOUDNESS OF AUDIO SIGNALS," filed on December 19, 2008, which is hereby incorporated by reference herein in its entirety.

[0083] With reference to FIGURE 7A, the band partitioning module 720 outputs critical bands information (e.g., magnitude) 721, 722, 723, and 724 of the portion of the audio input signal 210. Critical bands can be band-pass filtered components of frequency-transformed portion of the audio input signal 210. Critical bands 721 through 724 are depicted in increasing order (e.g., critical band 724 has higher order or covers a greater bandwidth than critical band 721), and information contained in the critical bands can be ordered based on time (e.g., second sample in critical band 721 may correspond to audio signal sample that follows the first sample). A suitable number of critical bands can be used by the band partitioning module 720, such as between 20 and 25 bands, between 15 and 30 bands, or some other suitable number of bands. The frequency range covered by the band partitioning module 720 can be, for example, between 20 Hz and 20 kHz or some subset thereof covering at least a portion of the audible range of the human auditory system.

[0084] Critical band information is input to a modulation spectrum correlation module 730, which performs correlation of critical band information. In one embodiment, the modulation spectrum correlation module 730 computes cross-correlation for some or all critical bands (e.g., each critical band) against some or all other critical bands. For example, when 25 critical bands are used, the modulation spectrum correlation module 730 can compute 25x25 (or 625) cross-correlation values. In another embodiment, the modulation spectrum correlation module 730 can compute different number of cross-correlation values, such as not compute the autocorrelation for each critical band (e.g., compute 600 cross-correlation values in case when 25 critical bands are used). The modulation spectrum correlation module 730 can analyze cross-correlation values for similarity. In one embodiment, a substantially large degree of similarity

may indicate that the portion of the auditory input signal 210 has at least some, and potentially high, information content. For example, if a portion of the auditory input signal 210 is determined to have substantially large degree of similarity, the portion may have information in some or all frequency bands covered by band partitioning module 720. This may indicate that the portion of the signal carries audible information (e.g., music, speech, etc.). On the other hand, a substantially small or low degree of similarity may indicate that the portion of the auditory input signal 210 has low information content. For example, if a portion of the auditory input signal 210 is determined to have substantially small degree of similarity, it may indicate that the portion is noise (e.g., random noise, such as white noise, which tends to have a constant spectral density).

[0085] Based on correlation analysis, the modulation spectrum correlation module 730 can determine the information content of the portion of the audio input signal 210. In one embodiment, the modulation spectrum correlation module 730 can apply a non-linear function to determined correlations in order to determine or compute an information coefficient (e.g., weighting value). In another embodiment, a linear function can be utilized. The information coefficient is preferably a measure of information content (e.g., a value between 0 and 1). In another embodiment, the information coefficient may be a binary value that indicates presence or absence of information content (e.g., 0 may indicate little or no information content, and 1 may indicate presence of information content).

[0086] In one embodiment, the information coefficient can correspond to an information content weighting W_{Pk} . With reference to FIGURE 4A, the process 400 can utilize weighting W_{Pk} determined by a perceptual information content estimation block 225, as is explained below. With reference to FIGURE 3A, gated measured loudness 365 of a frame can be determined by process 400 in block 414 using the following equation or the like (assuming that the weights are normalized to 1):

$$L_{GMR} = \frac{\sum_{i=0}^{N-1} L_{wi} \cdot \alpha^i \cdot \beta^{|L_i - \bar{L}_i|} \cdot W_{Pi}}{\sum_{i=0}^{N-1} \alpha^i \cdot \beta^{|L_i - \bar{L}_i|} \cdot W_{Pi}} \quad (12)$$

where, N is the number of windows or blocks in a frame, α is a coefficient of temporal weighting (described above), β is a coefficient based on the deviation of the loudness value (described above), W_{Pi} is weighting based on perceptual content, L_i represents loudness for i th window or block (e.g., L_{Wi} as computed in block 404), and $\bar{L}_i$ represents mean loudness of the frame (e.g., as computed in block 408). In an embodiment that utilizes 400 msec windows that overlap by 75%, a new loudness value can be computed every 100 msec. In certain embodiments, N can represent number of overlapping windows in frame (e.g., 4 windows), number of window overlaps (e.g., 4 for each window when 400 msec windows overlap by 75%), and so on. In some embodiments, index k (as is used in some equations described above) can be used in place of index i . In various embodiments, $\bar{L}_i$ can correspond to mean loudness of a window.

Additional Embodiments

[0087] Although described herein primarily with respect to television broadcasting, the features of the systems and methods described herein can be implemented in other broadcasting and streaming scenarios. For instance, the systems and methods can be used to adjust the loudness of audio signals in radio broadcasts, which often have programming interspersed with commercials. In addition, the systems and methods described herein can be used to adjust the loudness of audio signals streamed or broadcast over a network, such as the Internet.

[0088] Although described herein primarily with respect to real-time processing, the features of the systems and methods described herein can be used in file-based processing. For instance, the systems and methods can be used to adjust the loudness of an entire audio file or audio program at once. As another example, the systems and methods can be used to adjust the loudness during VOD processing, AVOD processing, or the like.

[0089] Although described herein primarily with respect to performing the loudness adjustment directly using loudness values, systems and methods described herein can be used to perform loudness adjustment in different domains. For instance, instead of directly weighting the computed initial

loudness values, gain values can be determined (e.g., based on the loudness values, such as computing the multiplicative inverse of the loudness values) and weighted in a weighted average using any of the weighting techniques described above.

Conclusion

[0090] Depending on the embodiment, certain acts, events, or functions of any of the algorithms described herein can be performed in a different sequence, may be added, merged, or left out all together (e.g., not all described acts or events are necessary for the practice of the algorithm). Moreover, in certain embodiments, acts or events may be performed concurrently, e.g., through multi-threaded processing, interrupt processing, or multiple processors or processor cores, rather than sequentially.

[0091] The various illustrative logical blocks, modules, and algorithm steps described in connection with the embodiments disclosed herein may be implemented as electronic hardware, computer software, or combinations of both. To clearly illustrate this interchangeability of hardware and software, various illustrative components, blocks, modules, and steps have been described above generally in terms of their functionality. Whether such functionality is implemented as hardware or software depends upon the particular application and design constraints imposed on the overall system. The described functionality may be implemented in varying ways for each particular application, but such implementation decisions should not be interpreted as causing a departure from the scope of the disclosure.

[0092] The various illustrative logical blocks and modules described in connection with the embodiments disclosed herein may be implemented or performed by a machine, such as a general purpose processor, a digital signal processor (DSP), an application specific integrated circuit (ASIC), a field programmable gate array (FPGA) or other programmable logic device, discrete gate or transistor logic, discrete hardware components, or any combination thereof designed to perform the functions described herein. A general purpose processor may be a microprocessor, but in the alternative, the processor may be a processor, controller, microcontroller, or state machine, combinations of the

same, or the like. A processor may also be implemented as a combination of computing devices, e.g., a combination of a DSP and a microprocessor, a plurality of microprocessors, one or more microprocessors in conjunction with a DSP core, or any other such configuration.

[0093] The steps of a method or algorithm described in connection with the embodiments disclosed herein may be embodied directly in hardware, in a software module executed by a processor, or in a combination of the two. A software module may reside in RAM memory, flash memory, ROM memory, EPROM memory, EEPROM memory, registers, hard disk, a removable disk, a CD-ROM, or any other form of storage medium known in the art. An exemplary storage medium can be coupled to the processor such that the processor can read information from, and write information to, the storage medium. In the alternative, the storage medium may be integral to the processor. The processor and the storage medium may reside in an ASIC. The ASIC may reside in a user terminal. In the alternative, the processor and the storage medium may reside as discrete components in a user terminal.

[0094] Conditional language used herein, such as, among others, "can," "could," "might," "may," "e.g.," and the like, unless specifically stated otherwise, or otherwise understood within the context as used, is generally intended to convey that certain embodiments include, while other embodiments do not include, certain features, elements and/or states. Thus, such conditional language is not generally intended to imply that features, elements and/or states are in any way required for one or more embodiments or that one or more embodiments necessarily include logic for deciding, with or without author input or prompting, whether these features, elements and/or states are included or are to be performed in any particular embodiment.

[0095] While the above detailed description has shown, described, and pointed out novel features as applied to various embodiments, it will be understood that various omissions, substitutions, and changes in the form and details of the devices or algorithms illustrated may be made without departing from the spirit of the disclosure. As will be recognized, certain embodiments of the inventions described herein may be embodied within a form that does not provide all of the features and benefits set forth herein, as some features may be

used or practiced separately from others. The scope of certain inventions disclosed herein is indicated by the appended claims rather than by the foregoing description. All changes which come within the meaning and range of equivalency of the claims are to be embraced within their scope.

WHAT IS CLAIMED IS:

1. A method of adjusting loudness of an audio signal, the method comprising:

receiving an audio signal;

dividing the audio signal into a plurality of frames;

for a frame of the plurality of frames, by one or more processors:

measuring initial loudness values for blocks of samples in the frame to produce a plurality of initial loudness values, and

computing a weighted average of at least some of the initial loudness values in the frame to produce an adjusted loudness value for the frame, wherein weights in the weighted average are selected based at least partly on one or both of recency of the initial loudness values and variation of the initial loudness values; and

adjusting loudness of the audio signal based at least partly on the adjusted loudness value.

2. The method of claim 1, wherein the weightings enable said adjusting the loudness to track changes in the loudness of the audio signal more accurately than an arithmetic mean of the initial loudness values.

3. The method of any of the preceding claims, wherein the weights are selected based at least partly on both of the recency of the initial loudness values and variation of the initial loudness values.

4. The method of any of the preceding claims, wherein said computing the weights based on the variation of the initial loudness values comprises computing the weights based on deviation of the initial loudness values from a mean of the initial loudness values.

5. The method of any of the preceding claims, wherein at least some of the weights are nonlinear.

6. The method of claim 5, wherein at least some of the weights are selected using an exponential function.

7. The method of claim 6, further comprising damping the weights by reducing a coefficient of the exponential function based at least partly on a measure of the variation of the initial loudness values, wherein said damping the

weights reduces an impact on dynamic range of the audio signal by said adjusting the loudness of the audio signal.

8. The method of claim 7, wherein the measure of the variation is a standard deviation of the initial loudness values.

9. The method of any of the preceding claims, further comprising determining a length of the frame based at least in part on a variation of loudness of the plurality of initial loudness values.

10. The method of any of the preceding claims, wherein the weights are further selected based on a modulation spectrum analysis of the audio signal.

11. The method of claim 10, wherein the modulation spectrum analysis comprises computing a coefficient of variation among time-frequency representations of the audio signal.

12. The method of claim 11, wherein the computed coefficient of variation reflects a degree to which the audio signal comprises either information, noise, or both information and noise.

13. The method of any of claims 11 to 12, wherein larger weights are assigned in response to the coefficient being a first value and wherein lower weights are assigned in response to the coefficient being a second value lower than the first value.

14. A system for adjusting loudness of an audio signal, the system comprising:

a loudness analysis module comprising one or more processors,
the loudness analysis module configured to:

access a frame of an audio signal, the frame comprising a plurality of samples of the audio signal,

measure loudness values for windows of the samples in the frame to produce a plurality of loudness values, and

compute a weighted average of at least some of the loudness values to produce an adjusted loudness value, wherein weights in the weighted average are selected based at least partly on one or more of the following: timing of the loudness values and deviation of the loudness values; and

a dynamics control module configured to adjust loudness of the audio signal based at least partly on the adjusted loudness value.

15. The system of claim 14, wherein the dynamics control module is configured to compute one or more gains to apply to the audio signal based on the adjusted loudness value.

16. The system of claim of any of claims 14 to 15, further comprising a limiter configured to at least partially limit the one or more gains based on available headroom in the audio signal, thereby reducing clipping.

17. The system of any of claims 14 to 16, wherein at least some of the weights are linear.

18. The system of claim 17, wherein the loudness analysis module is further configured to damp the weights by reducing a slope of a linear function used to select the weights, said reducing being based at least partly on a measure of the deviation of the loudness values.

19. The system of claim 18, wherein the measure of the deviation is a standard deviation of the loudness values.

20. The system of any of claims 14 to 19, wherein the loudness analysis module is further configured to damp the weights by adjusting a coefficient of an exponential function of the variation, said adjusting comprising reducing the coefficient based on the variation.

21. A method of adjusting loudness of an audio signal, the method comprising:

- sampling an audio signal to produce blocks of samples;
- measuring initial loudness values for the blocks of samples in the audio signal to produce a plurality of initial loudness values;
- computing gain values to be applied to the audio signal based at least partly on the initial loudness values;
- computing a weighted average of the initial loudness values or of the gain values, wherein weights in the weighted average are selected based at least partly on one or both of recency and variation of the initial loudness values or the gain values; and
- adjusting loudness of the audio signal based at least partly on the computed gain values and the weighted average.

22. The method of claim 21, further comprising computing the weighted average of the gain values.

23. The method of any of claims 21 to 22, wherein said computing the gain values comprises obtaining a multiplicative inverse of the initial loudness values.

24. The method of any of claims 21 to 23, wherein said adjusting the loudness comprises first applying the weighted average of the gain values and then applying the weighted average of the gain values to the audio signal.

25. The method of any of claims 21 to 24, further comprising selectively discarding at least some of the loudness values that reach an adaptive loudness threshold, wherein a difference between the adaptive loudness threshold and the loudness values changes over time based on the loudness values.

26. The method of claim 25, wherein said selectively discarding at least some of the loudness values comprises discarding those of the loudness values that are below the adaptive loudness threshold.

27. The method of any of claims 25 to 26, further comprising computing the adaptive loudness threshold based at least in part on a previous loudness value of a frame that immediately precedes the frame.

28. The method of claim 27, wherein said computing the adaptive loudness threshold comprises computing a difference between the previous loudness value and a mean loudness value.

29. A method of adjusting loudness of an audio signal, the method comprising:

- receiving an audio signal;

- dividing the audio signal into a plurality of frames;

- for a frame of the plurality of frames, by one or more processors:

- measuring initial loudness values for blocks of samples in the frame to produce a plurality of initial loudness values,

- dynamically computing an adaptive loudness threshold, such that a difference between the adaptive loudness threshold and the initial loudness values changes over time based on the initial loudness values,

selectively discarding at least some of the initial loudness values that reach the adaptive loudness threshold; and
adjusting loudness of the audio signal based at least partly on the initial loudness values that have not been discarded.

30. The method of claim 29, wherein said selectively discarding at least some of the initial loudness values comprises discarding those of the initial loudness values that are below the adaptive loudness threshold.

31. The method of any of claims 28 to 29, further comprising computing the adaptive loudness threshold based at least in part on a previous loudness value of a second frame that precedes the frame.

32. The method of any of claims 28 to 31, wherein said computing the adaptive loudness threshold comprises computing a difference between the previous loudness value and at least one of: a mean loudness value of the frame and an initial loudness value.

33. The method of claim 32, wherein said computing the difference comprises accumulating the difference.

34. The method of any of claims 32 to 33, wherein the difference is bounded by a lower bound and an upper bound.

35. The method of any of claims 29 to 34, further comprising computing the adaptive loudness threshold based at least in part on a difference between an initial loudness value and the mean loudness value of the frame.

36. The method of any of claims 29 to 35, further comprising computing a weighted average of at least some of the initial loudness values to produce an adjusted loudness value, wherein said adjusting the loudness of the audio signal comprises adjusting the audio signal based at least partly on the adjusted loudness value.

37. The method of claim 36, wherein weights in the weighted average are computed based at least partly on one or both of recency of the initial loudness values and variation of the initial loudness values.

38. The method of any of claims 36 to 37, wherein the weights are further computed based at least partly on a modulation spectrum domain analysis of the audio signal, wherein the modulation spectrum domain analysis

comprises computing a coefficient of variation among time-frequency representations of the audio signal.

39. A system for adjusting loudness of an audio signal, the system comprising:

a loudness analysis module comprising one or more processors, the loudness analysis module configured to:

access a frame of an audio signal, the frame comprising a plurality of samples of the audio signal,

measure initial loudness values for blocks of the samples in the frame to produce a plurality of initial loudness values,

dynamically compute an adaptive loudness threshold, such that a difference between the adaptive loudness threshold and the initial loudness values changes over time based on the initial loudness values, and

selectively discard at least some of the initial loudness values that reach the adaptive loudness threshold; and

a dynamics control module configured to adjust loudness of the audio signal based at least partly on the initial loudness values that have not been discarded.

40. The system of claim 39, wherein the loudness analysis module is configured to selectively discard those of the initial loudness values that are below the adaptive loudness threshold.

41. The system of any of claims 39 to 40, wherein the loudness analysis module is further configured to compute the adaptive loudness threshold based at least in part on a previous loudness value of a second frame that precedes the frame.

42. The system of any of claims 39 to 41, wherein the loudness analysis module is configured to compute the adaptive loudness threshold based at least in part on a difference between the previous loudness value and at least one of: a mean loudness value of the frame and the initial loudness value.

43. The system of claim 42, wherein the loudness analysis module is configured to accumulate the difference.

44. The system of any of claims 42 to 43, wherein the difference is bounded by a lower bound and an upper bound.

45. The system of any of claims 39 to 44, wherein the loudness analysis module is further configured to compute the adaptive loudness threshold based at least in part on a difference between the initial loudness value and the mean loudness value of the frame.

46. The system of any of claims 39 to 45, wherein:

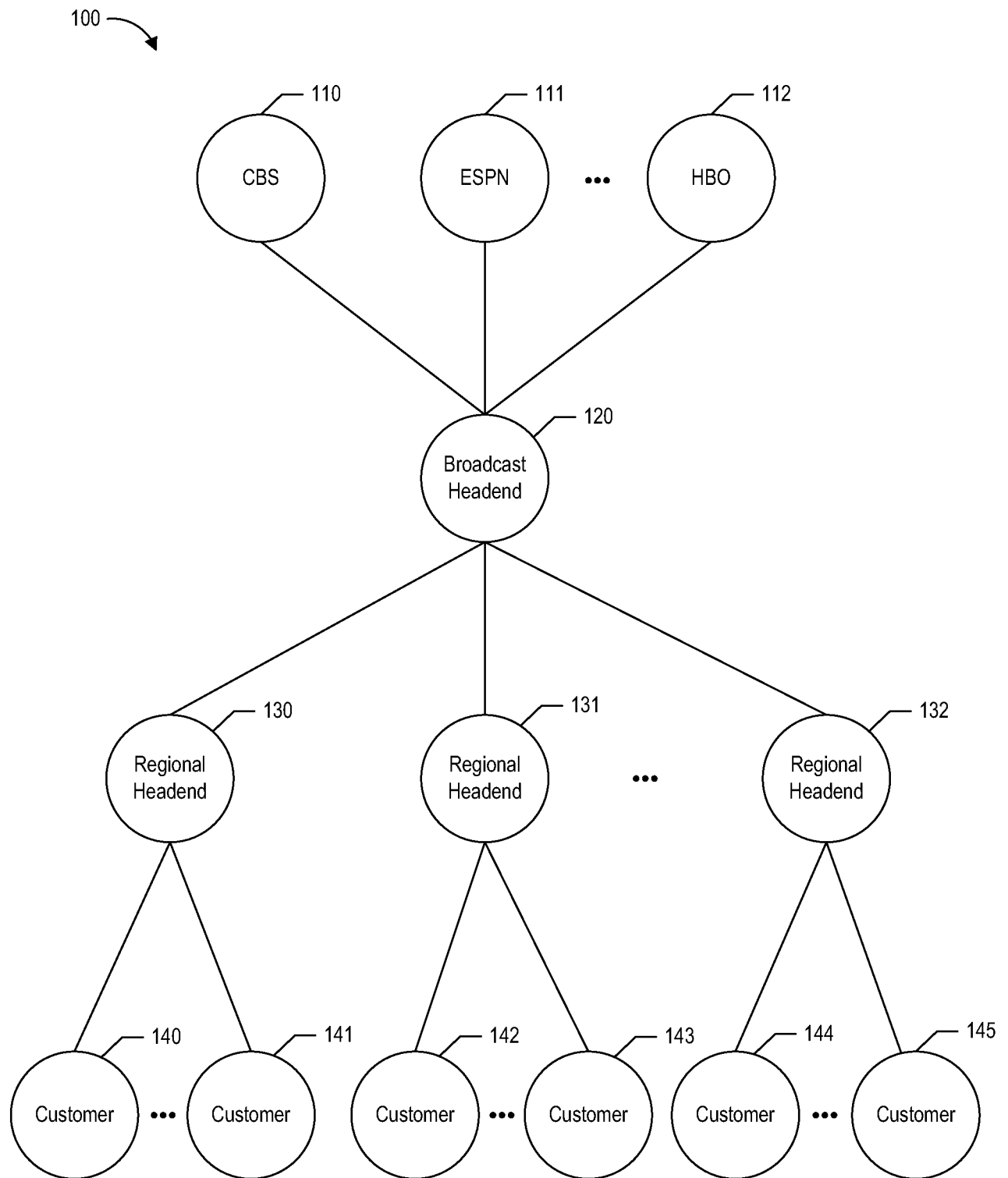
the loudness analysis module is further configured to compute a weighted average of at least some of the initial loudness values to produce an adjusted loudness value; and

the dynamics control module is further configured to adjust the loudness of the audio signal based at least partly on the adjusted loudness value.

47. The system of claim 46, wherein the loudness analysis module is configured to compute weights in the weighted average based at least partly on one or both of recency of the initial loudness values and variation of the initial loudness values.

48. The system of any of claims 46 to 47, further comprising a perceptual information content estimation module configured to perform a modulation spectrum domain analysis of the audio signal, wherein the modulation spectrum domain analysis comprises computing a coefficient of variation among time-frequency representations of the audio signal, and the loudness analysis module is further configured to compute the weights based at least partly on the modulation spectrum domain analysis of the audio signal.

1/14

**FIG. 1**

200A

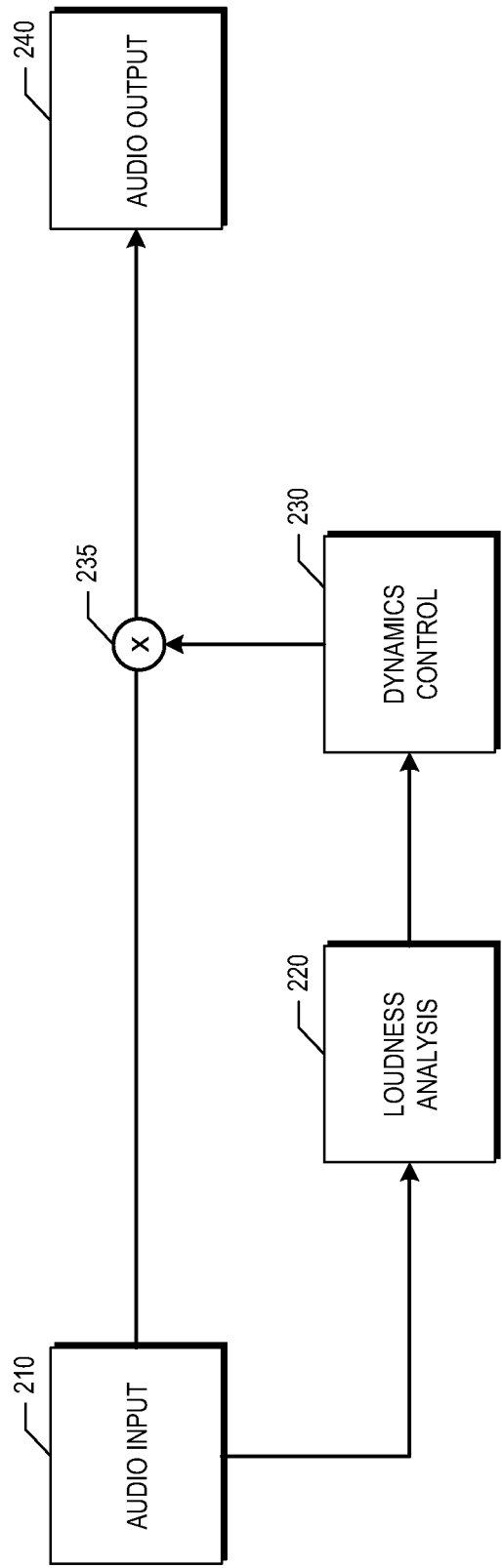


FIG. 2A

200B

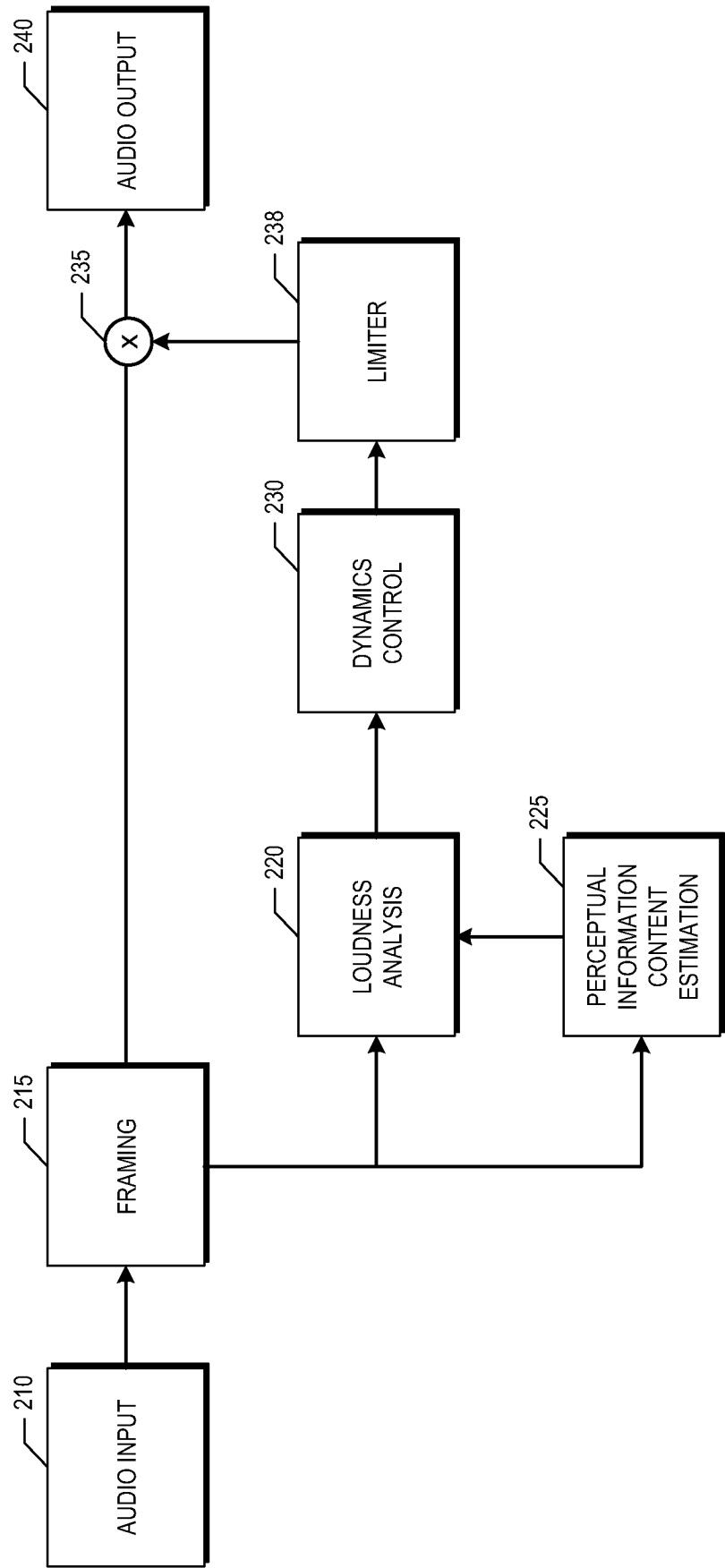


FIG. 2B

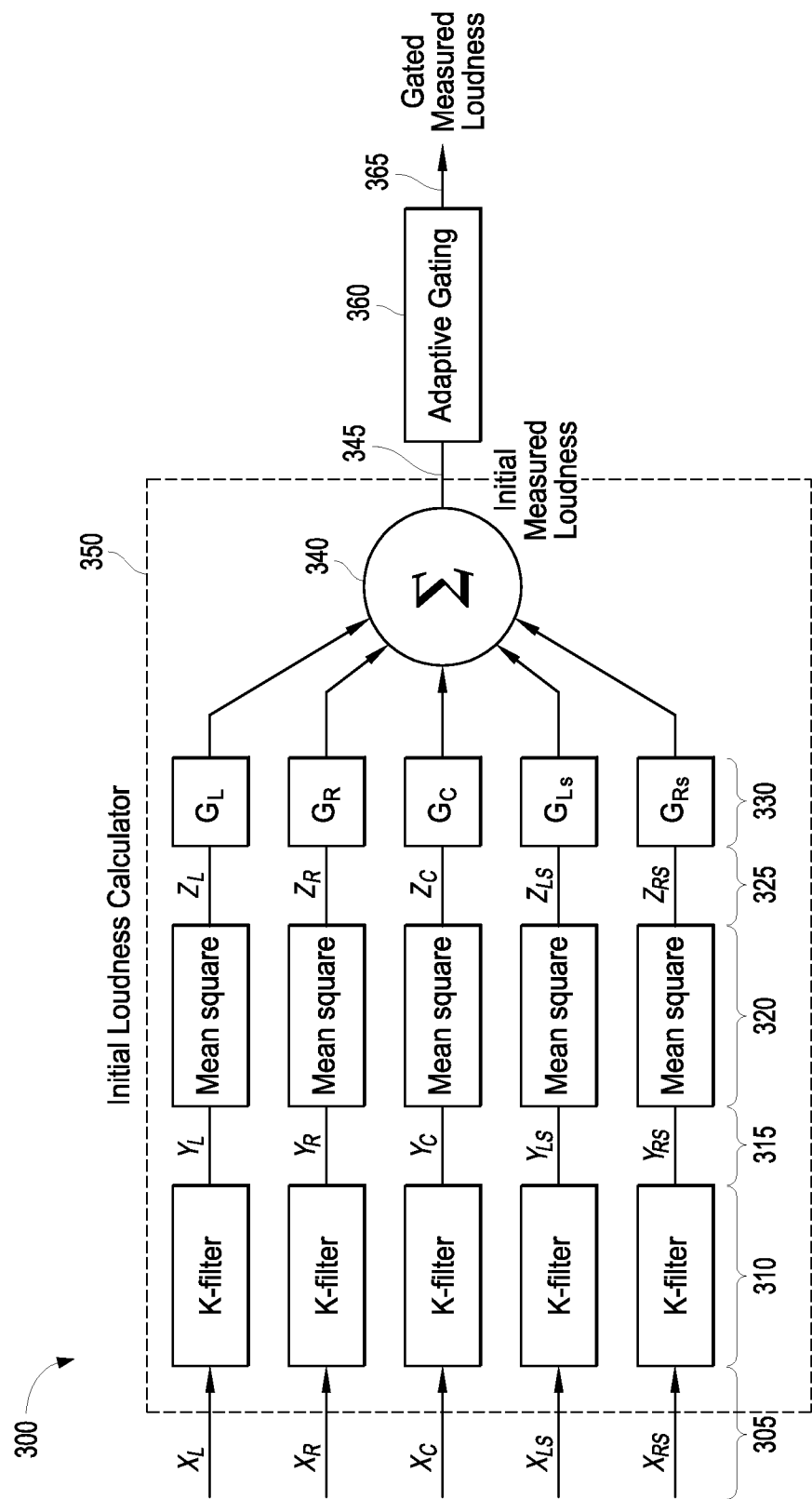


FIG. 3A

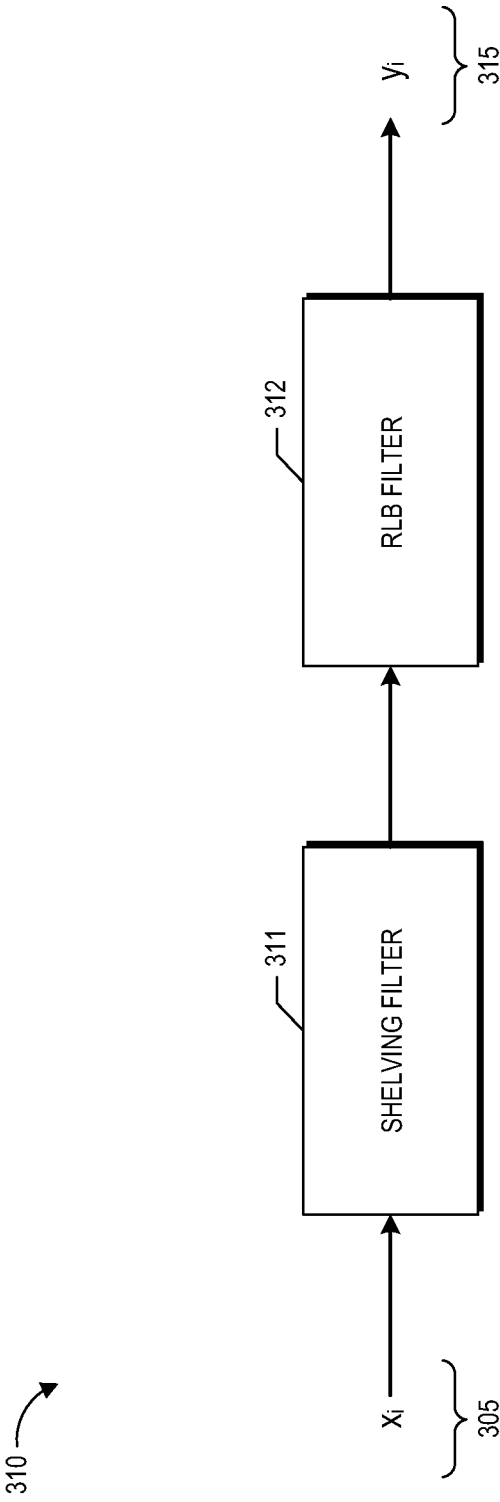
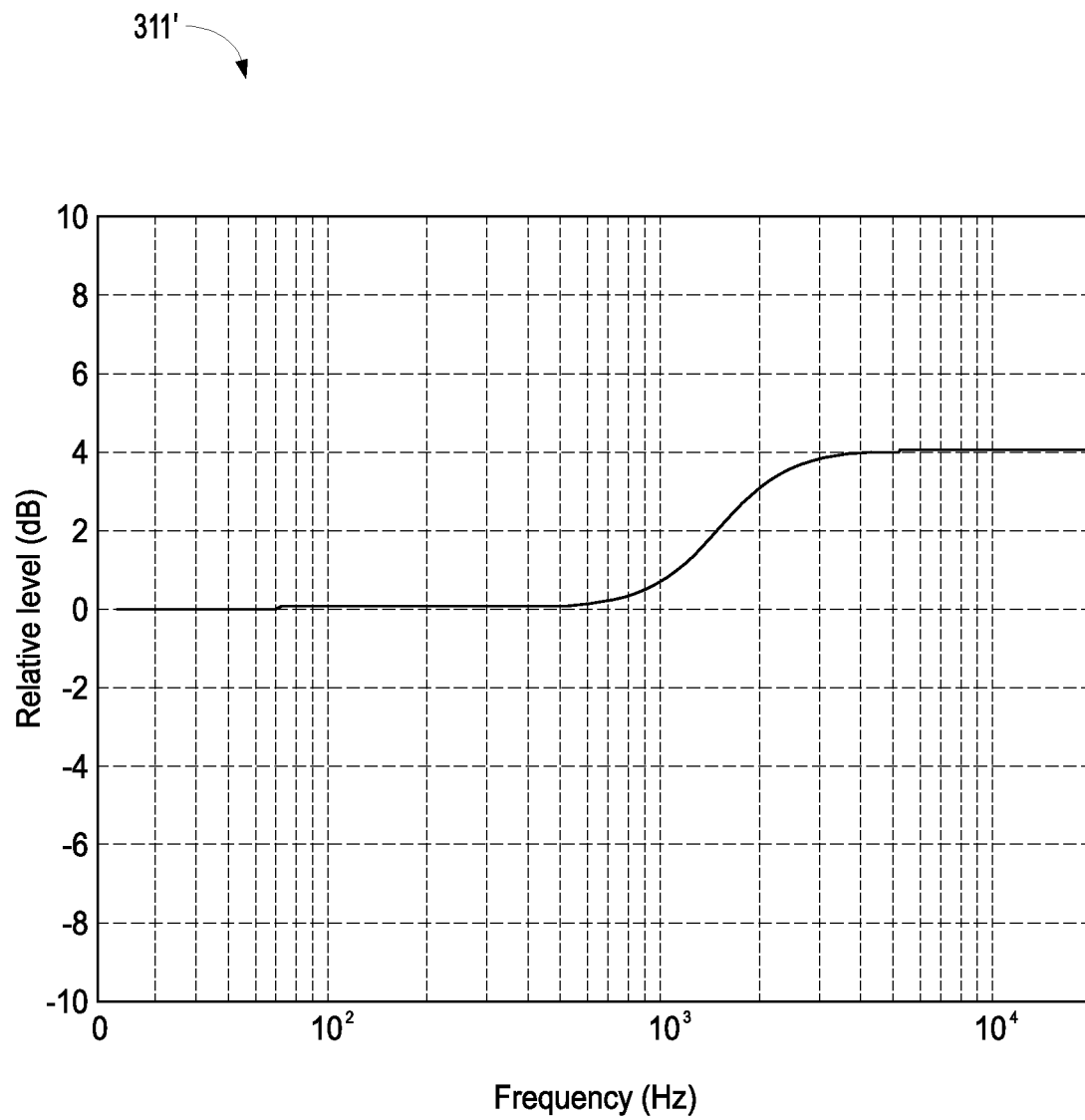
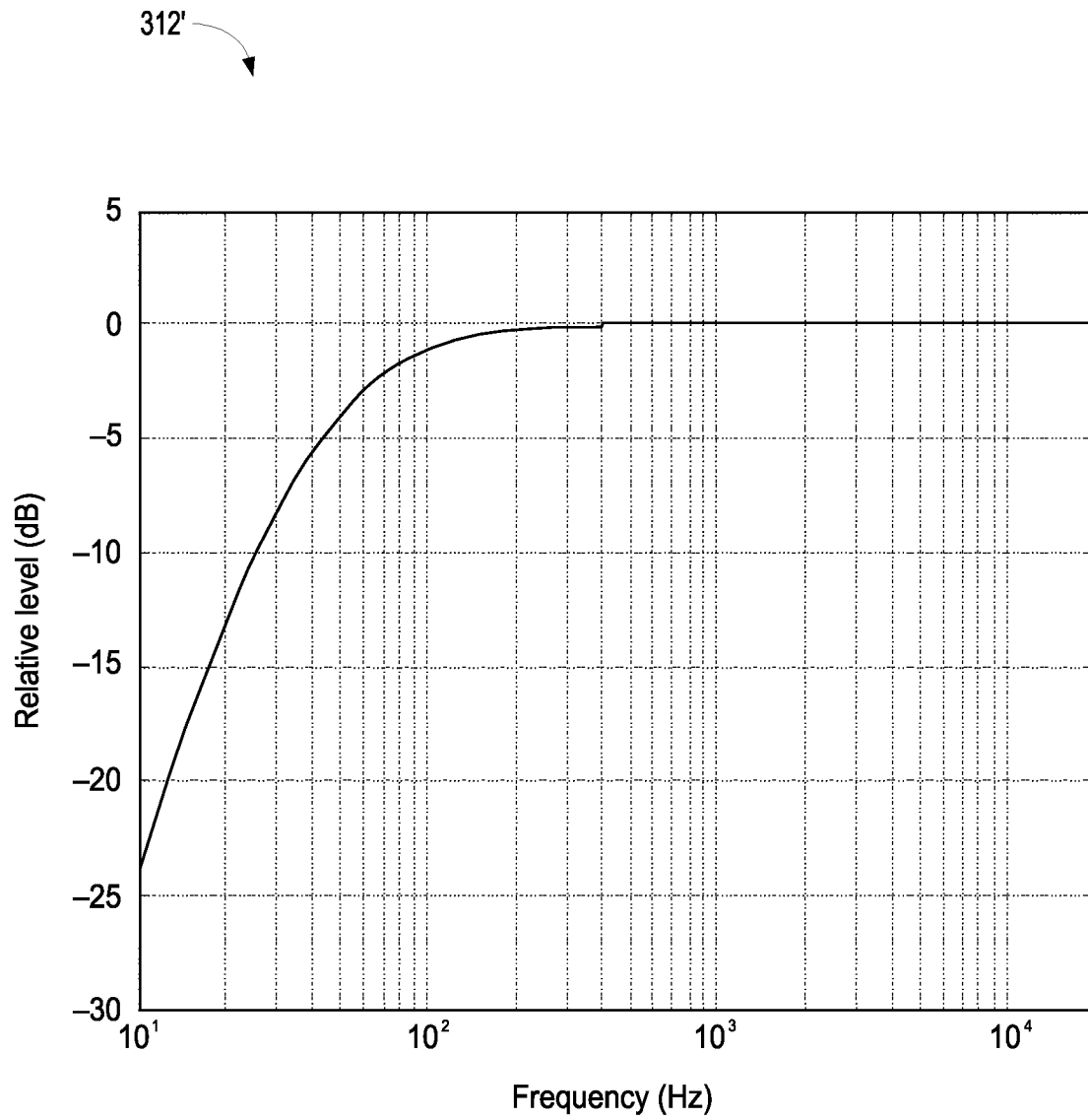
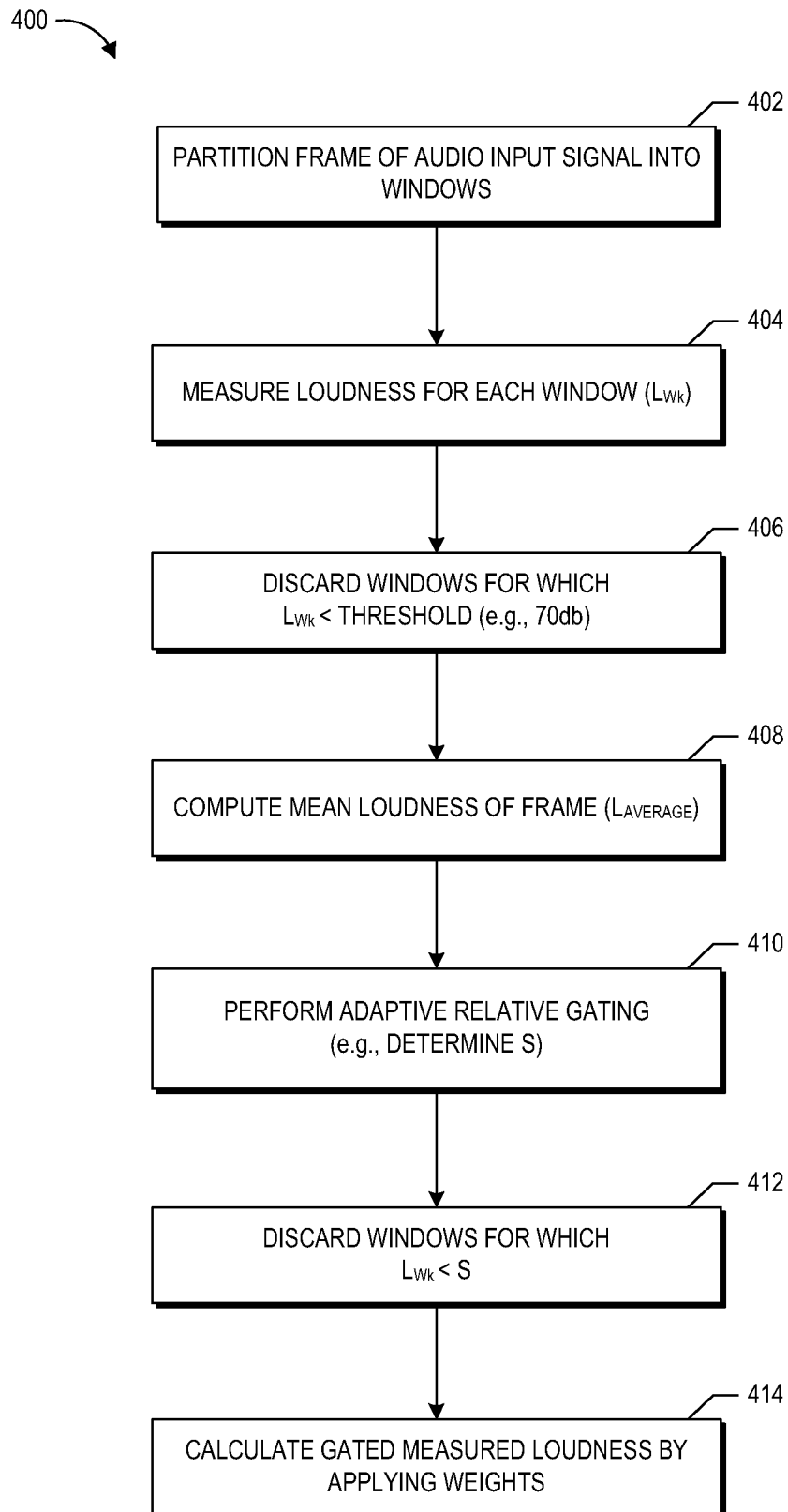


FIG. 3B

**FIG. 3C**

**FIG. 3D**

8/14

**FIG. 4A**

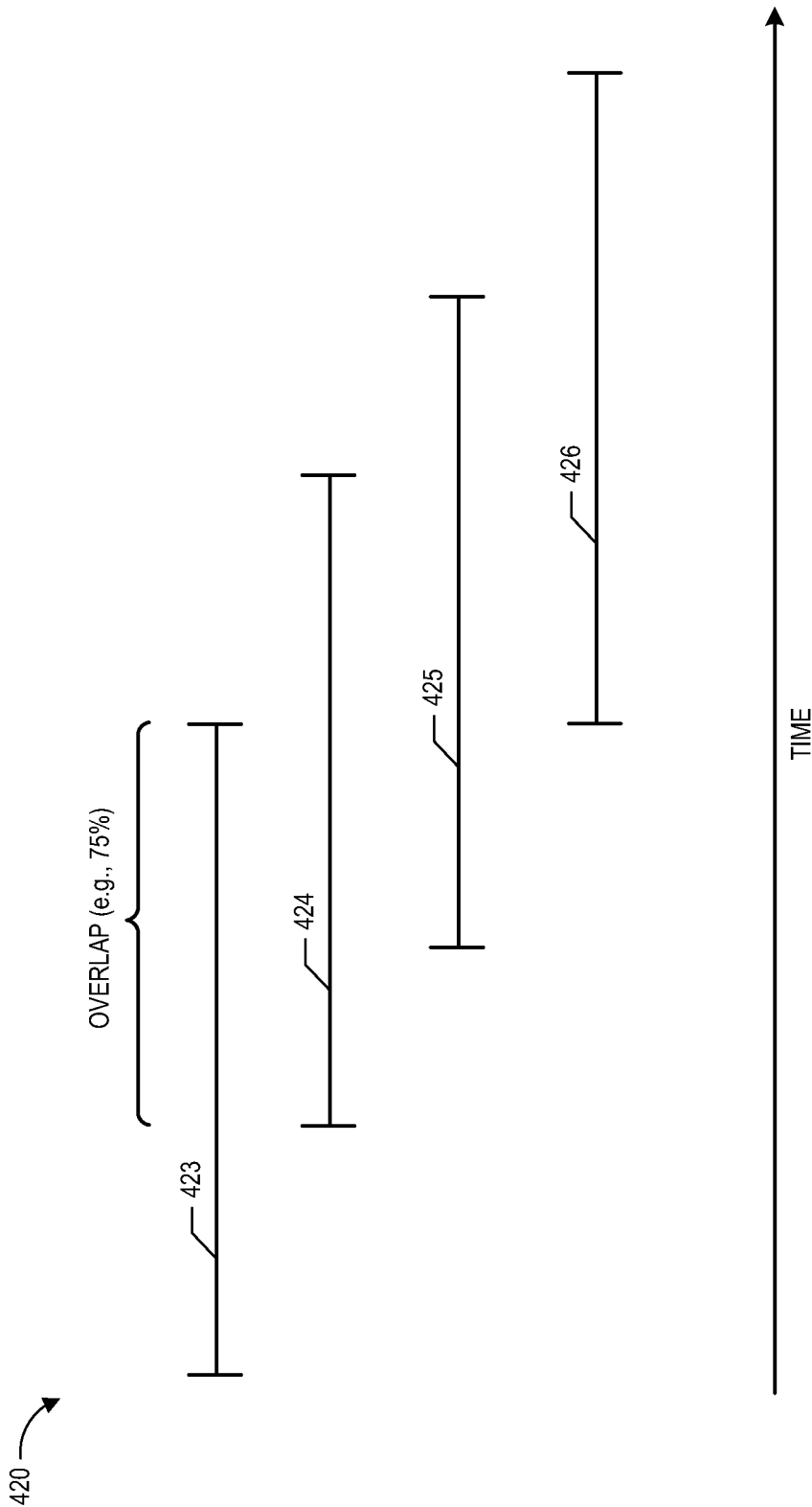


FIG. 4B

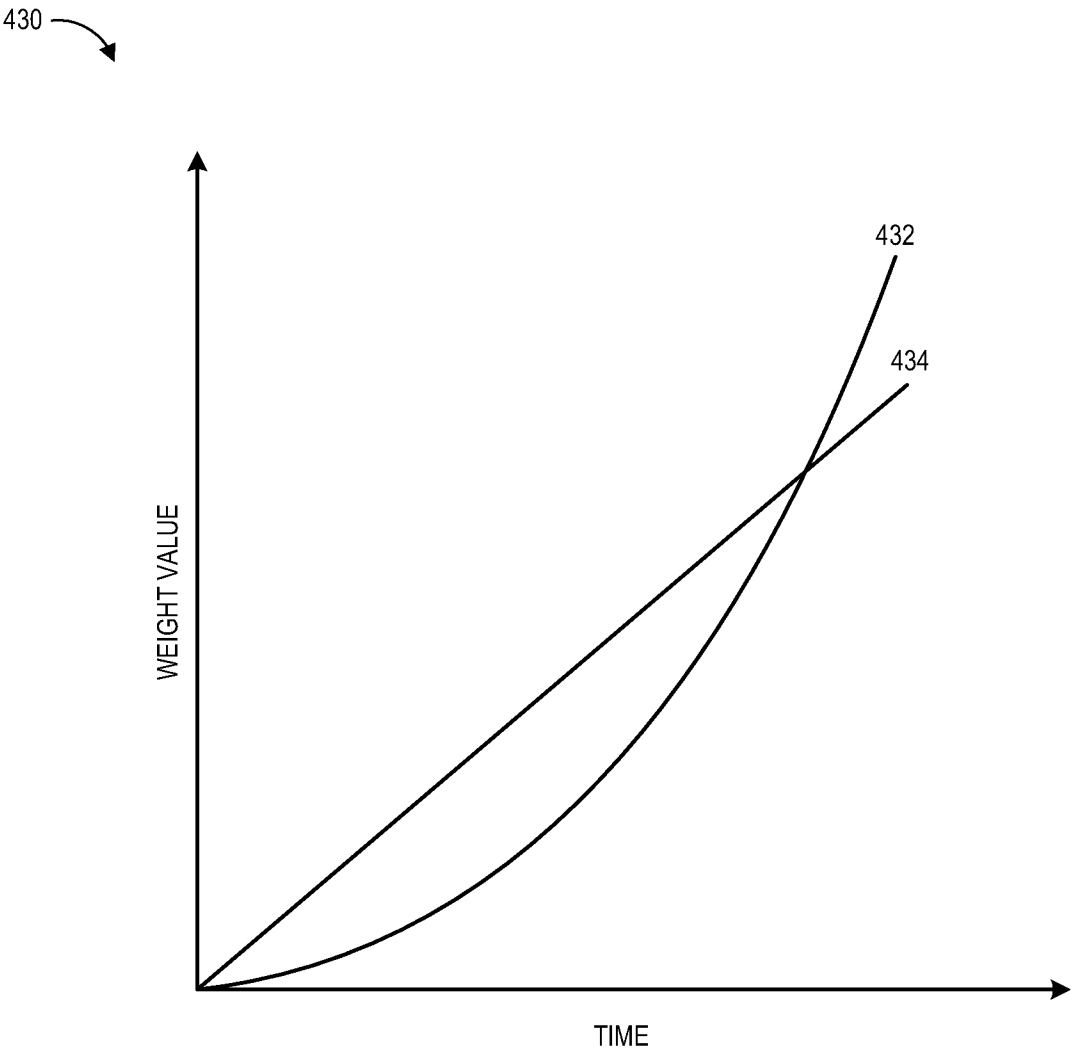


FIG. 4C

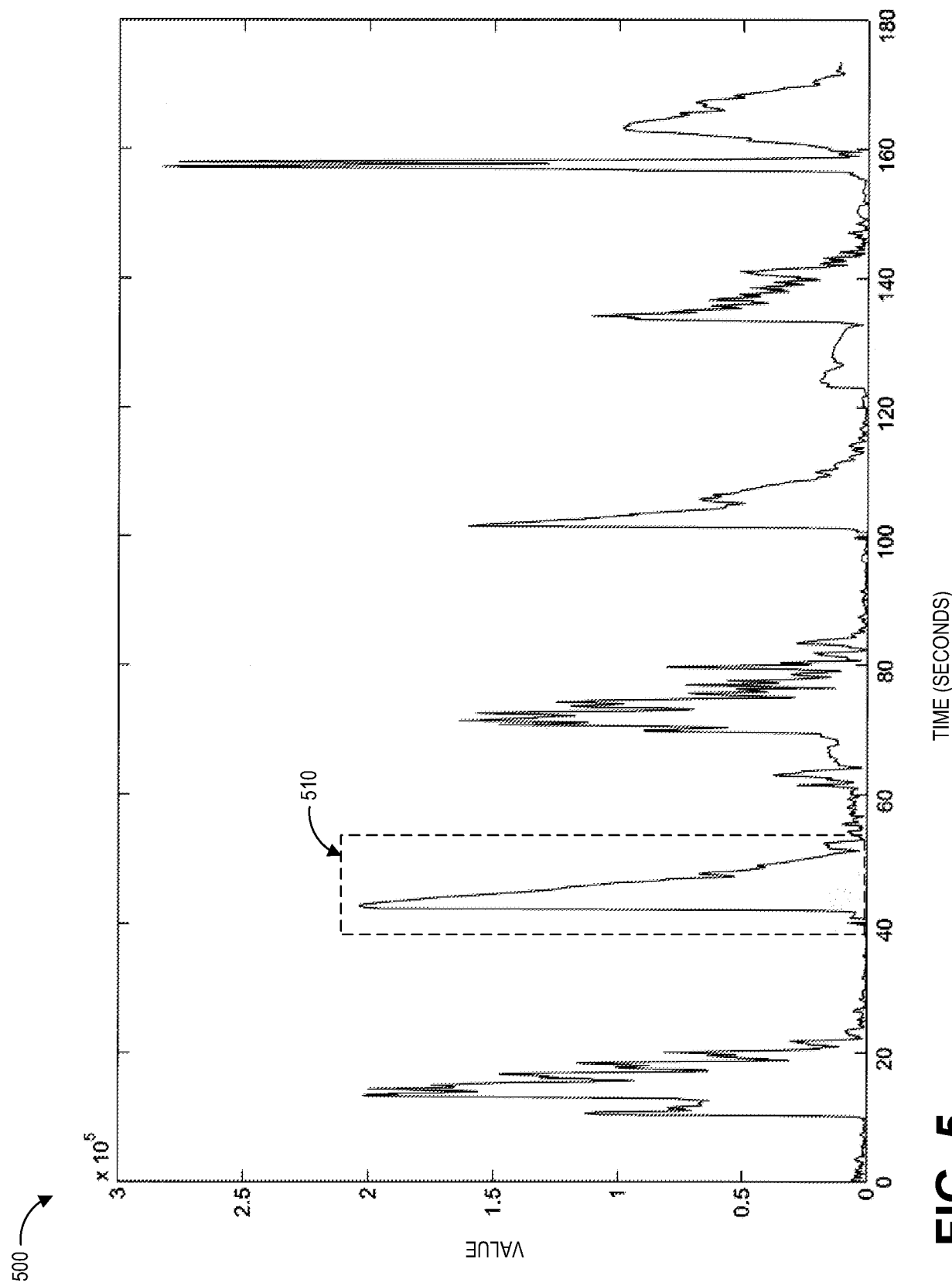


FIG. 5

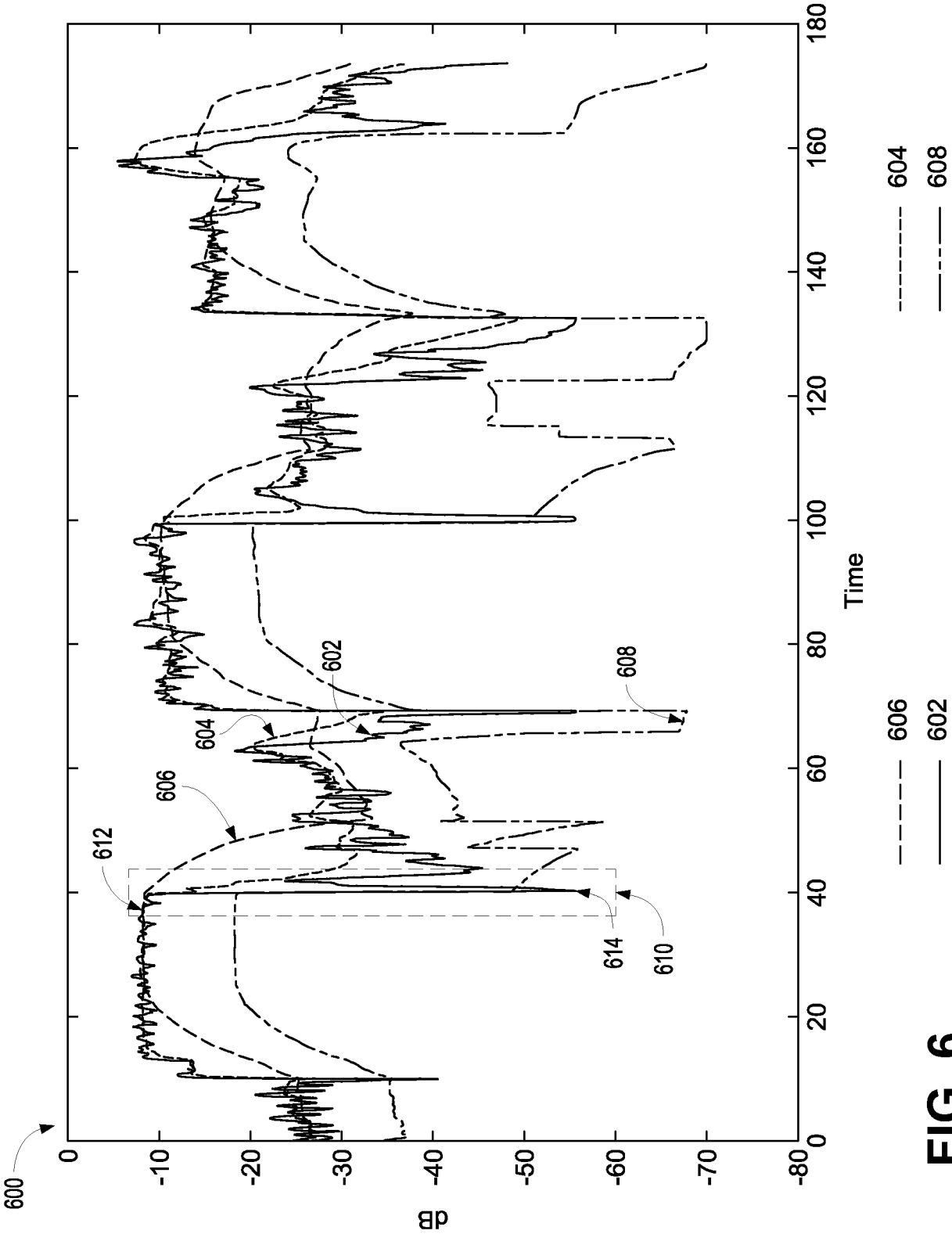


FIG. 6

225 →

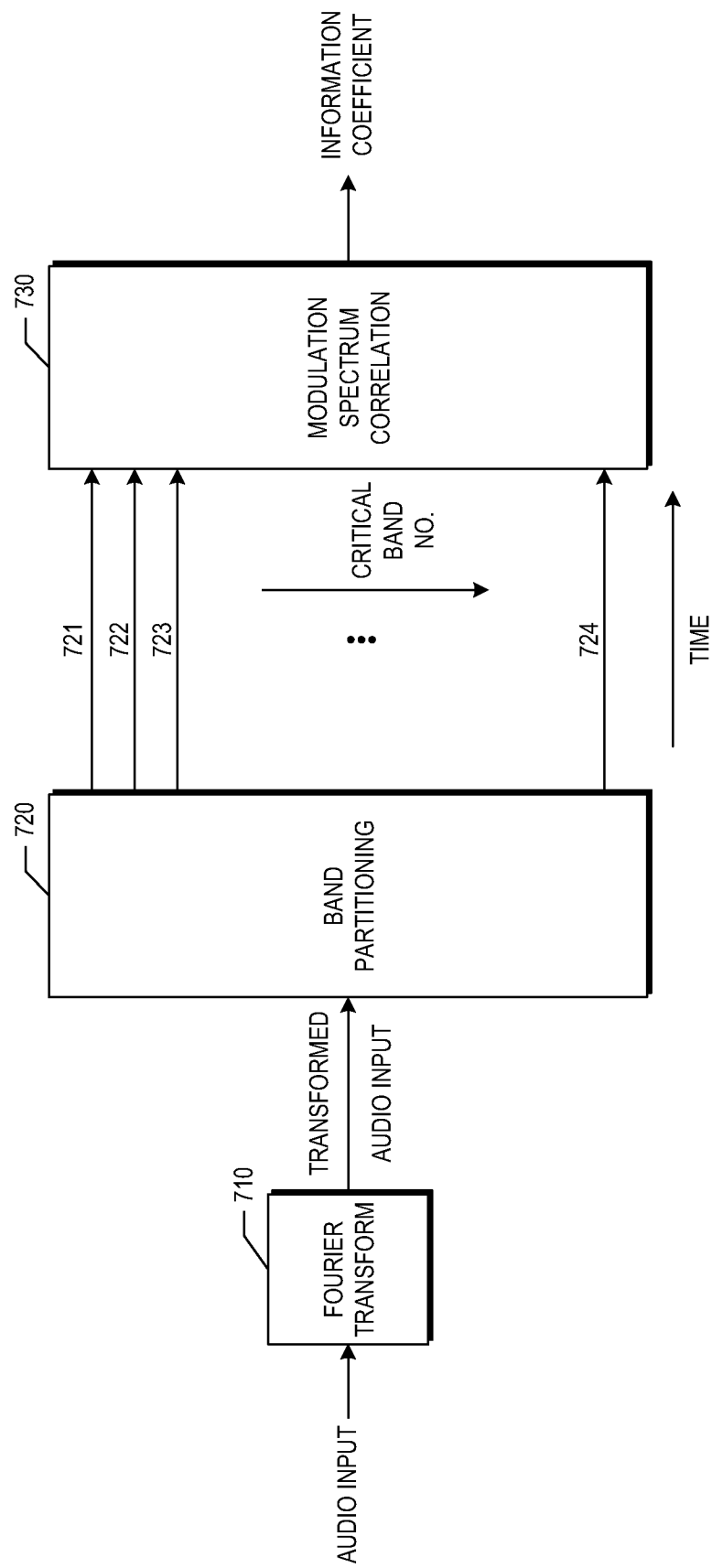


FIG. 7A

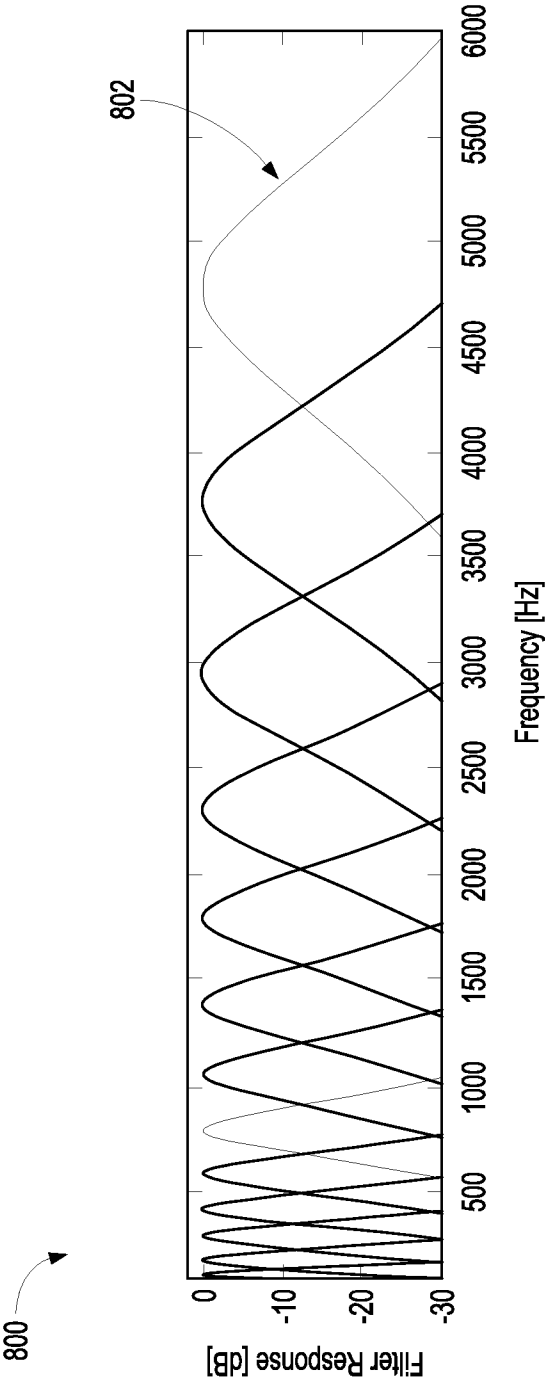


FIG. 7B