

- 1 -

Description**Audio Dynamics Processing using a Reset**

5

Technical Field

The invention relates to audio signal processing. In particular, the invention relates to an audio dynamics processor or processing method that uses a reset mechanism or process in order to adapt quickly to content changes in the audio signal. A reset signal
10 may be generated by analyzing the audio signal itself or the reset may be triggered from an external event such as a channel change on a television set or an input selection change on an audio/visual receiver. In the case of an external trigger, one or more indicators of the state of the dynamics processor for a current audio source may be saved and associated with that audio source before switching to a new audio source. Then, if the
15 system switches back to the first audio source, the dynamics processor may be reset to the state previously stored or an approximation thereof. The invention also relates to computer programs for practicing such methods or controlling such apparatus.

Background Art

An object of audio dynamics processing is to alter the level or dynamics of an
20 audio signal to be within some desired limits. This is generally achieved by creating a time-varying measure of an audio signal's level (rms level or peak level, for example) and then computing and applying a time-varying signal modification (a gain change, for example) that is a function of the level estimate. Dynamics processors employing such a mode of operation include automatic gain controls (AGCs), dynamic range controls
25 (DRCs), expanders, limiters, noise gates, etc. Various types of signal dynamics processing are set forth in International Patent Application PCT/US 2005/038579 of Alan Jeffrey Seefeldt, published as WO 2006/047600 on May 4, 2006. The application designates the United States among other entities. The application is hereby incorporated by reference in its entirety.

30 FIG. 1 depicts a high level block diagram of a generic audio dynamics processor. The processor may be considered to have two paths, an upper "signal" path 2 and a lower "control" path 4. On the lower path, a dynamics control process or controller ("Dynamics Control") 6 measures the level of the audio signal and generates one or more time-varying modification parameters as a function of the level measure. As shown, the modification
35 parameters are derived from the input audio signal. Alternatively, the modification

- 2 -

parameters may be derived from the processed (output) audio or from a combination of the input and output audio signals. On the upper audio path 2, the modification parameters generated by the Dynamics Control 6 are applied to the audio to generate the processed audio. The application of modification parameters to an audio signal may be accomplished in many known ways and is shown generically by the multiplier symbol 8. For example, in the case of a simple automatic gain control device or process, there may be a single wideband gain modification parameter that controls the gain of a variable gain/loss device or process in the main path. In practice, the audio may also be delayed prior to the application of the modification parameters in order to compensate for any delay associated with the computation of the modification parameters in the dynamics control process. For simplicity in presentation, a delay is not shown in FIG. 1 or other figures herein.

In a dynamics control process, it is typical that both the signal level measure and the resulting modification parameters are computed continuously over time. In addition, either or both the signal level measure and the modification parameters are usually smoothed across time to minimize perceptible artifacts from being introduced into the processed audio. The smoothing is most often performed using a "fast attack" and a "slow release", meaning that the modification parameters change relatively quickly in response to an increase in the signal level and respond more slowly as the signal level decreases. Such smoothing is in accordance with the dynamics of natural sounds and the way in which humans perceive changes in loudness over time. Consequently, such time smoothing is nearly universal in audio dynamics processors.

For some dynamics processing applications, the time constants associated with such smoothing may be quite large; on the order of one or more seconds. An AGC, for instance, may compute an estimate of the long-term average level of a signal using large time constants and then use the resulting estimate to generate slowly varying modification parameters that move the average level of the audio closer to a desired target level. In this case, large time constants may be desirable in order to preserve the short-term dynamics of the audio signal. Suppose that such an AGC is operating on the audio of a television set with the intent of maintaining a consistent average level across programming and across various channels. In such a situation, the content of the audio signal being processed by the AGC may abruptly change or have a discontinuity, when a channel is changed for example, and the associated average level of the audio signal may therefore also abruptly change or have a discontinuity. With its large time constants,

- 3 -

however, the AGC takes a considerable amount of time to converge to a new level and bring the modified level of the processed audio in line with the desired target level. During such adaptation time, a viewer of the television may perceive the level of the audio to be too loud or too soft. As a result, the viewer may quickly reach for the remote control to adjust the volume—only to find himself or herself fighting the AGC as it converges.

A typical prior art solution to the problem just described involves using time constants that adapt based on the dynamics of the signal. For example, if the short-term level of the signal is significantly greater or less than the smoothed level as defined by some threshold boundaries around the smoothed level, then the smoothing operation switches to faster attack and/or release time constants, respectively, until the short-term level falls back within the threshold boundaries around the smoothed level. Subsequently, the system switches back to the original slower time constants. Such a system may reduce the adaptation time of the AGC, but the thresholds and shorter time constants must be chosen carefully. In general, for any reasonable thresholds, signals may exist in which the original desired signal dynamics fluctuate outside of the threshold boundaries around the average level, thus causing the smoothing process to falsely switch into the fast attack or release mode. Due to the possibly frequent occurrence of such false switching, the fast attack and release mode time constants must not be chosen to be too short in order to avoid instability of the AGC during normal program material. As a result, the convergence of the AGC during abrupt transitions or discontinuities in the audio content may still not be as fast as desired.

It is therefore the object of the present invention to provide a better solution to the problem of dynamics processing adaptation time during audio content changes.

Disclosure Of The Invention.

According to an aspect of the invention, a method for processing an audio signal (or apparatus practicing the method) includes altering the dynamics of the audio signal in accordance with a dynamics control process, detecting a change in the content of the audio signal, and resetting the dynamics control process in response to detecting such a change. Detecting a change in the content of the audio signal may comprise one or both of (1) detecting an occurrence in the temporal evolution of the audio signal in which the level of the audio signal remains below a threshold, L_{silence} , throughout a time interval no shorter than a first time threshold, t_{silence} , and (2) detecting an occurrence in the temporal evolution of the audio signal in which the level of the audio signal decreases by an

- 4 -

amount greater than a difference threshold, L_{drop} , within a time interval no greater than a second time threshold, t_{drop} .

According to another aspect of the invention, a method for processing an audio signal (or apparatus practicing the method) includes altering the dynamics of the audio signal in accordance with a dynamics control process, receiving an externally-supplied reset trigger signal indicating a change in the audio signal, and resetting the dynamics control process in response to the reset trigger signal. The reset trigger signal may indicate one or more of (1) the act of a user changing a channel, (2) the act of a user changing input sources, (3) the act of a user selecting a play, rewind or forward function, (4) the switching of one file to another, (5) a change in program, (6) the switching of one audio coding format to another, and (7) a change in coding parameters.

According to either of the above aspects, resetting the dynamics control process may speed up the rate at which the dynamics processing adapts to changes in the input signal. Resetting the dynamics control process may set the value of one or more process-controlling parameters or a signal of which such one or more process-controlling parameters are a function to a stored or default value. Resetting the dynamics control process may shorten the value of one or more time constants employed in determining one or more process-controlling parameters or a signal of which such one or more process-controlling parameters are a function.

Although the adaptive-time-constants approach described above may be somewhat helpful in reducing the adaptation time of an audio dynamics processor or process and may be used along with aspects of the present invention, such a processor or process may be improved significantly by augmenting it with an explicit reset mechanism or process that is activated upon the occurrence of certain abrupt changes or discontinuities in the content of the audio signal being processed. FIG. 2 depicts the present invention in a general form in which a further control path, a reset control path 10, is added to the generic dynamics processor of FIG. 1. A reset control mechanism or process ("Reset Control") 12 may respond to one or both of the audio signal itself (the input audio signal or a combination of the input and output audio signals) and an external trigger signal. Thus, there are three possible configurations: (1) the audio signal itself (which, in turn may comprise the input audio signal or a combination of the input and output audio signals), (2) an external trigger signal, and (3) both the audio signal itself and an external trigger signal. If a decision to reset is made by Reset Control 12, the reset control mechanism or process sends a reset message to the Dynamics Control 6,

- 5 -

instructing it to reset certain aspects of its dynamics control state in a way that allows the audio dynamics processing to rapidly adapt to the new level of the incoming audio signal. The resetting of certain aspects of the dynamics control state may, for example, (1) set one or more time constants to a shorter value for a short period of time (thereby speeding up the adaptation rate of the audio processing), and/or (2) set a smoothed measure of the audio signal level or one or more modification parameters to a stored value or to an average or default value.

Analyzing the Audio Signal

to Determine the Occurrence of an Abrupt Change or Discontinuity in Signal Content

A reset decision may be initiated by Reset Control 12 in a number of ways, for example, by analyzing the incoming audio signal itself or by receiving an external reset trigger signal. In analyzing the audio signal, the Reset Control 12 attempts to detect conditions in the signal indicative of an abrupt change or discontinuity in content. Upon detecting such a condition with a certain degree of confidence, Reset Control 12 makes a reset decision and subsequently instructs Dynamics Control 6 to reset one or more aspects of its dynamics control state. Although other techniques may be utilized to detect abrupt changes or discontinuities in signal content, two practical and effective examples for doing so are described below.

Using an External Trigger

to Determine the Occurrence of an Abrupt Change or Discontinuity in Signal Content

In many applications, highly reliable external information may be available to trigger a reset. For example, in a television set or cable "set-top box", the act of a user changing a channel may serve as an external trigger of a reset. Also, in a television or audio/video receiver, the act of a user changing input sources, for example from "Video 1" to "Video 2", may serve to trigger a reset. Other examples include a file-based media player, for example, a portable audio device or a digital video recorder. In such cases, a reset may be triggered externally when the system finishes playing one file and switches to another. A reset might also be triggered when a user explicitly changes files himself or herself, presses play, or fast forwards or rewinds to a new location in a piece of content. In the case in which the audio being processed has been decoded from a digital data compressed stream, Dolby Digital audio, for example, other sources of an external reset trigger signal may exist. "Dolby" and "Dolby Digital" are trademarks of Dolby Laboratories Licensing Corporation. For example, the Dolby Digital audio coding system is the mandatory audio format for High Definition Television in the United States. On a

- 6 -

given channel of programming, the audio is sent continuously, but the Dolby Digital format may change at program boundaries. For example, the main program may be encoded in 5.1 channel format, but a commercial may be encoded in stereo format. A Dolby Digital decoder necessarily detects such changes and may pass the information to the audio dynamics processor of the present invention as a reset trigger. Other changes in coding parameters may also be employed, for example changes in the Dolby Digital dialog normalization parameter, "DIALNORM", that generally is held constant for a single piece of content. A similar external reset trigger may be generated from other audio coding formats as well, such a MP3 and AAC. In addition to the sources of external reset triggers just listed, others are possible, and the invention is not meant to be limited by those listed.

In the case where the invention is used in a device in which the audio may originate from a set of known sources, the invention may be further enhanced as shown in FIG. 3. Along the bottom of the figure are depicted N audio sources ("Audio Source 1" through "Audio Source N ") 14-1 through 14- N , where it is assumed that one of the N audio sources is selected for playback through the dynamics processing system by way of a source selection device or process ("Source Select") 16. Such audio sources could represent, but are not limited to, various channels on a television or set-top box, the various inputs on an audio/visual receiver, or various files on a portable media player. Stored with each audio source are indicators of the state of one or more aspects of the dynamics control process as it existed the last time that audio particular source was played. Such aspects of the state of the dynamics control may include, for example, one or more of (1) the smoothed measure of the audio signal level, (2) one or more modification parameters resulting from the smoothed measure of the audio signal level, and (3) one or more of the time constants employed in obtaining a smoothed measure of the audio signal level. In an exemplary embodiment described below, the smoothed measure of the audio signal level is employed, which indirectly affects the gain modification parameter (in the example, gain is a function of the smoothed audio signal level). When a different audio source is selected through the source select process, prior to actually switching to the new audio source, the Source Select 16 sends an external reset trigger to the Reset Control 12'. This in turn causes the Reset Control 12' to capture indicators of the current state of the dynamics control process for the currently playing audio source. The Reset Control 12' then saves indicators of the dynamics control state in the N dynamics control state storage locations associated with the current audio

- 7 -

selection (Dynamic Control State Storage 1 through N) 18-1 through 18-N. Next, the Source Select 16 switches to the new audio source and sends the associated stored indicators of the dynamics control state to the reset control process that in turn resets the dynamics control process using the stored state indicators.

5 By using aspects of the last dynamics control state of an audio source to reset the dynamics control process when switching to that source, better performance may be achieved than in the absence of employing such an initial state. Without such information, the dynamics control process relies on the incoming audio signal after the reset as it adapts to the new audio source. On a short-term basis, the level of the audio
10 may fluctuate rapidly, and therefore the modification parameters may also fluctuate rapidly as the dynamics control process attempts to adapt rapidly to the new audio. By using aspects of the last state of the dynamics processor associated with the audio source to reset the dynamics processor, it is likely that the dynamics processor will begin in a state that is closer to the state to which it will eventually converge for that audio source.

15 As a result, the fast adaptation time constants after the reset need not be decreased as much as may otherwise be required, thereby providing improved stability without sacrificing overall adaptation time. As an example, consider switching between two television stations, the first with audio at an average level of -15 dB relative to full-scale digital and the second with audio at an average level of -30 dB, both with a range of plus
20 or minus 5 dB around their respective average. Suppose that just before switching to the second station, the average level of the first station is at -13 dB. Suppose further that the average level of the second station is at -33 dB after the switch. This is a difference of -20 dB. If one desires the dynamics processing system to adapt in half a second, for example, an adaptation rate of -40 dB per second is required in the absence of any other
25 information. Such a rate is quite fast and may introduce instability in the processed audio. On the other hand, suppose that the last dynamics processing state stored with the second station corresponds to a level of -28 dB. Then, after the reset, the dynamics control process may be reset with this state and a difference of only $-33 \text{ dB} - (-28 \text{ dB}) = -5 \text{ dB}$ exists. Thus, an adaptation rate of only -10 dB per second (requiring less of a shortening
30 of the time constant) is required to converge to the desired level within half a second. This example illustrates the advantages in storing the last state of the dynamics control process with the audio source that resulted in such state.

- 8 -

The Effect When a Reset is Triggered

Triggering a reset speeds up the audio dynamics processing adaptation rate (the rate at which the dynamics processing adapts to changes in the input signal). For example, when a reset is triggered, one or more time constants associated with the adaptation rate may quickly switch to significantly smaller values and then smoothly return to their original large values over a specified period of time. Alternatively, the one or more time constants may switch to significantly smaller values and remain at such values for a specified period of time. In an AGC arrangement, a time constant control signal $c[t]$ may momentarily decrease the time constants used in computing the smoothed average level. For example, the smoothing time constants may be decreased over a duration of approximately one second following a reset. In an example embodiment, the time constant control signal $c[t]$ may start at a value of "1" (indicating a maximum change effect on one or more time constants) upon occurrence of a reset trigger; when it is "1" the attack and release coefficients are set to values significantly smaller than their nominal values. As $c[t]$ decays to zero over a short period of time such as one second (the length of the time period is not critical), the coefficients are interpolated back to their nominal normal (non-reset) values. Alternatively or in addition, upon occurrence of a reset trigger, the value of a smoothed average level computed by the dynamics control process may be reset to a state stored with a particular audio source or to a default value.

Description Of The Drawings

FIG. 1 is a schematic functional block diagram showing, at a high level, a generic audio dynamics processor.

FIG. 2 is a schematic functional block diagram showing, in accordance with aspects of the present invention, an example of an embodiment of the present invention, in which, in a general form, a further control path, a reset control path, is added to the generic dynamics processor of FIG. 1.

FIG. 3 is a schematic functional block diagram showing, in accordance with aspects of the present invention, an example of an embodiment of the present invention, in which an audio dynamics processor is further enhanced.

FIG. 4 shows an exemplary input/output function for a typical automatic gain control (AGC).

FIG. 5 shows a set of bandpass frequency responses chosen to mimic the critical band filtering observed along the basilar membrane in the human ear.

- 9 -

FIG. 6 is a schematic functional block diagram showing, in accordance with aspects of the present invention, an example of an AGC embodiment of the present invention.

Best Mode For Carrying Out The Invention

5 The described invention is applicable to a wide variety of dynamics processing applications, such as dynamic range control, compressors, limiters, expanders, etc. It is particularly useful, however, when applied to an AGC in which, typically, large time constants result in slow adaptation in the absence of the present invention. Thus, the preferred embodiment describes, as an example, the invention's application to an AGC.

Basic AGC Implementation

10 A basic implementation of an AGC computes a time-varying estimate of a signal's average level and computes a desired modified output level that is a function of the input level and a desired target level. A time varying signal modification may then be computed as a function of the difference between the input level and the desired output
15 level. The function mapping input to output is designed to bring the level of the modified audio closer to the desired target level, and FIG. 4 depicts one such input/output function. For an input above the target level, the AGC calls for signal attenuation, and for an input below the target level, the AGC calls for signal boost. The degree of attenuation or boost may be controlled by varying the slope of the line in FIG. 4.

20 Ideally, the measure used to compute the average level of the audio should correlate with the human perception of loudness. This may be achieved in many ways; for example, by a weighted mean-square power measure or a psychoacoustic loudness measure. A simple unweighted mean-square power measure is slightly less accurate than the two methods just mentioned, but still demonstrates a high degree of correlation with
25 human loudness perception for most real-world audio signals. Due to its computational simplicity, an unweighted mean-square power measure is used in a preferred embodiment, but should not be seen as limiting the invention.

Although, in principle, aspects of the invention may be implemented in analog and/or digital domains, practical implementations are likely to be implemented in the
30 digital domain in which each of the audio signals are represented by individual samples or samples within blocks of data. The input level estimate and the corresponding desired output level and signal modification parameters may be computed continuously for an analog audio signal or on a sample by sample basis from a digital signal, but for this example embodiment it is instead desirable to compute such quantities for consecutive

- 10 -

overlapping blocks of samples of a digital audio signal. This is due mainly to the fact that digital block processing is useful for detecting reset conditions from the signal itself, as is described later. Letting the digital audio signal be represented by $x[n]$, the overlapping blocks of the audio signal may be computed as:

$$x[n, t] = w[n]x[n + tN/2] \quad \text{for} \quad 0 < n < N-1 \quad (1)$$

where N is the block length, $N/2$ is the degree of overlap between consecutive blocks, t is the block index, and $w[n]$ is a window function, such as a sine window. For signals sampled at 44100 Hz, setting $N=512$ or $N=1024$ works well. Further details of a digital audio processing arrangement employing overlapping blocks of samples are set forth in U.S. Patent 5,899,969 of Fielder et al ("Frame-based audio with gain-control words"), which patent is hereby incorporated by reference in its entirety. As is described below, the AGC computes a time varying gain $G[t]$ that is then multiplied with each block of the signal. Each of these modified blocks is then overlap-added to produce the final modified audio signal $y[n + tN/2]$:

$$y[n + tN/2] = G[t]x[n, t] + G[t-1]x[n + N/2, t-1] \quad \text{for} \quad 0 < n < N/2 \quad (2)$$

As a first step in computing the gain $G[t]$, the time-varying instantaneous level of the signal $L[t]$ is computed as the mean square power of each block $x[n, t]$. In decibels, the level is computed as:

$$L[t] = 10 \log_{10} \left(\frac{1}{N} \sum_{n=0}^{N-1} x^2[n, t] \right) \quad (3)$$

where 0 dB corresponds to the level of a full scale digital square wave.

Next, the instantaneous level may be smoothed with a fast attack and a slow release to generate a smoothed average level of the audio signal $\bar{L}[t]$:

- 11 -

$$\bar{L}[t] = \begin{cases} \alpha[t] \cdot \bar{L}[t-1] + (1 - \alpha[t])L[t] & L[t] \geq L_{\min} \\ \bar{L}[t-1] & L[t] < L_{\min} \end{cases} \quad (4a)$$

5 where

$$\alpha[t] = \begin{cases} \alpha_{\text{attack}} & L[t] \geq \bar{L}[t-1] \\ \alpha_{\text{release}} & L[t] < \bar{L}[t-1] \end{cases} \quad (4b)$$

10 and

$$\alpha_{\text{release}} > \alpha_{\text{attack}} \quad (4c)$$

The smoothing coefficients α_{attack} and α_{release} are chosen to give the desired attack
 15 and release times. One way to specify this is the half-decay time of the smoothing filter,
i.e., the time it takes the time smoothing filter's impulse response to decay to one half its
 original value. Although choosing α_{attack} corresponding to a half decay time of 1 second
 and α_{release} corresponding to a half decay time of 4 seconds works well for an AGC, the
 values are not critical. The value $L_{\min}$ in Equation 4a represents a minimum level above
 20 which the instantaneous signal level $L[t]$ must be in order for the smoothed level to be
 updated. This prevents the smoothed average level from falling too low, and the
 corresponding gain from rising too high, when the signal falls into relative silence.
 Assuming 0 dB represents the level of a full scale digital square wave, setting $L_{\min} = -60$
 dB is a reasonable choice, although the level is not critical.

25 From the smoothed average level $\bar{L}[t]$ and the AGC input/output function F_{AGC} ,
 as shown in FIG. 4, the desired output level $\bar{L}_{\text{out}}[t]$ is computed:

$$\bar{L}_{\text{out}}[t] = F_{\text{AGC}}\{\bar{L}[t]\} \quad (5)$$

- 12 -

Finally, the gain $G[t]$ is computed from the difference between the output level $\bar{L}_{out}[t]$ and the input level $\bar{L}[t]$:

$$G[t] = 10^{(\bar{L}_{out}[t] - \bar{L}[t])/20}$$

(6)

When $\bar{L}_{out}[t] > \bar{L}[t]$, the gain is greater than one, meaning the signal is boosted, and when $\bar{L}_{out}[t] < \bar{L}[t]$, the gain is less than one, meaning the signal is attenuated.

10 *AGC with Reset*

As mentioned above, the reset mechanism or function may cause one of two or both of two actions: (1) setting one or more time constants to a shorter value for a short period of time (thereby speeding up the adaptation rate of the audio processing), and (2) setting a smoothed measure of the audio signal level or one or more modification parameters to a stored value or to an average or default value, which may be expressed as $\bar{L}_{init}$ (see Eqn. 8a below and the descriptions before and after that equation). First, resetting of one or more time constants is described.

Because the modification parameter, gain $G[t]$, is a function of the smoothed average level $\bar{L}[t]$, the rate at which the described AGC adapts to the audio is governed by the attack and release times used in computing $\bar{L}[t]$. As stated above, such smoothing time constants are chosen to be relatively large, on the order of several seconds, so that the AGC does not change the level of the audio too rapidly during the temporal evolution of normal audio content. However, if the audio content abruptly changes or has a discontinuity, the AGC may take too long to adapt, especially if the new content is at a significantly different level than the previous content. To speed up the adaptation, the reset signal described earlier may be used to trigger a modification of one or more of the time constants associated with the level smoothing. In this case, when a reset is triggered, the time constants quickly switch to significantly smaller values and then smoothly return to their original large values (or approximations thereof) over a specified period of time. Other methods for speeding up the adaptation rate after a reset is triggered are possible. For example, instead of smoothly returning to their original values or approximations thereof the time constants may be kept at their smaller values over a specified period of

- 13 -

time and then returned directly to their original values or approximations thereof. Another possibility is to provide a step-wise return of the time constants over a specified period of time. The invention is not meant to be limited manner in which the time constants return to their original values or an approximation thereof.

5 To trigger a reset there may exist a binary reset signal $R[t]$ such the $R[t]=0$ during normal operation and $R[t]=1$ when a reset is desired at time block t . The reset may be triggered by analyzing the audio signal or by an external source, as described earlier. Details of examples of triggering a reset through signal analysis are described below. In the case of an external trigger, there may exist a stored state with which the
10 dynamics process may be reset, as shown in FIG. 3. For the described preferred embodiment, such a stored state may consist of the value of the smoothed average level $\bar{L}[t]$ at the last time block at which a particular audio source was processed. The initial state, retrieved from storage during an externally-triggered reset (or, alternatively, an initial state may be a default value whether or not an externally-triggered reset is
15 employed), may be represented by the value $\bar{L}_{init}$.

From the binary reset signal $R[t]$, a time constant control signal may be generated and subsequently used to shorten momentarily the time constants used in computing the smoothed average level $\bar{L}[t]$, thereby speeding up the processing adaptation rate. A time constant control signal $c[t]$ may be computed so that it equals "1" at the moment a reset
20 occurs ($R[t]=1$), and then decays exponentially to "0" at a rate controlled by a decay coefficient $\lambda < 1$:

$$c[t] = 1 \text{ if } R[t] = 1$$

(7a)

otherwise,

$$c[t] = \begin{cases} \lambda c[t-1] & L[t] \geq L_{min} \\ c[t-1] & L[t] < L_{min} \end{cases}$$

(7b)

- 14 -

The decay coefficient λ may be chosen to give a half decay time of one second, for example, which means that the smoothing time constants are shortened over a duration of approximately one second following a reset. Note also that the control signal may be updated only if the signal level $L[t] \geq L_{\min}$ in order to prevent fast adaptation from occurring during relative silence.

Rather than using fixed attack and release coefficients to smooth the signal level, as shown in Equation 4b, the coefficients may now vary over time as a function of the time constant control signal $c[t]$. When $c[t]=1$, meaning that a reset has just occurred, the attack and release coefficients may be set equal to values significantly smaller than the values α_{attack} and $\alpha_{release}$ from Equation 4b (about ten percent of the original values or less, for example). As $c[t]$ then decays to "0", the coefficients may be interpolated back to their nominal values of α_{attack} and $\alpha_{release}$. Optionally, in addition, the small coefficient values used immediately after a reset may also vary depending on whether an initial state value $\bar{L}_{init}$ exists. If no stored initial state exists, then very small coefficient values may be used, and if a stored initial state does exist, then slightly larger coefficient values may be used. In the latter case it is assumed that initializing $\bar{L}[t]$ with $\bar{L}_{init}$ starts $\bar{L}[t]$ at a value closer to the value to which it will eventually converge. Therefore, a slightly slower adaptation time, corresponding to larger coefficients, may be utilized in comparison to the case where no initial state is available. The adaptive smoothing process may be expressed as:

$$\bar{L}[t] = \bar{L}_{init} \text{ if } R[t] = 1 \text{ and } \bar{L}_{init} \text{ exists} \quad (8a)$$

otherwise,

$$\bar{L}[t] = \begin{cases} \gamma[t] \cdot \bar{L}[t-1] + (1-\gamma[t])L[t] & L[t] \geq L_{\min} \\ \bar{L}[t-1] & L[t] < L_{\min} \end{cases} \quad (8b)$$

where

- 15 -

$$\gamma[t] = \begin{cases} \gamma_{attack}[t] & L[t] \geq \bar{L}[t-1] \\ \gamma_{release}[t] & L[t] < \bar{L}[t-1] \end{cases},$$

(8c)

$$\gamma_{attack}[t] = c[t]\beta_{attack}[t] + (1 - c[t])\alpha_{attack},$$

(8d)

$$\gamma_{release}[t] = c[t]\beta_{release}[t] + (1 - c[t])\alpha_{release},$$

(8e)

$$\beta_{attack}[t] = \begin{cases} \beta_{afast} & R[t] = 1 \text{ and } \bar{L}_{init} \text{ exists} \\ \beta_{aslow} & R[t] = 1 \text{ and } \bar{L}_{init} \text{ does not exist,} \\ \beta_{attack}[t-1] & \text{otherwise} \end{cases},$$

(8f)

$$\beta_{release}[t] = \begin{cases} \beta_{rfast} & R[t] = 1 \text{ and } \bar{L}_{init} \text{ exists} \\ \beta_{rslow} & R[t] = 1 \text{ and } \bar{L}_{init} \text{ does not exist,} \\ \beta_{release}[t-1] & \text{otherwise} \end{cases},$$

(8g)

$$\beta_{afast} < \beta_{aslow} \ll \alpha_{attack},$$

(8h)

and

$$\beta_{rfast} < \beta_{rslow} \ll \alpha_{release}$$

(8i)

The values β_{afast} and β_{rfast} represent the initial attack and release smoothing coefficients immediately after a reset when no initial state, $\bar{L}_{init}$, is provided. Values corresponding to half decays times of 50 milliseconds and 200 milliseconds, respectively,

- 16 -

have been found to be usable. The particular values are not critical. The values β_{aslow} and β_{rslow} represent the larger initial attack and release smoothing coefficients immediately after a reset when initial state, $\bar{L}_{init}$, is provided. In this case, values corresponding to half decays times equal to twice those of their faster counterparts; 100 milliseconds and 400 milliseconds, respectively, have been found to be usable. Also in this case, the particular values are not critical.

Reset Detection Through Signal Analysis

The most reliable method for triggering a reset in the disclosed invention is to receive a trigger externally from a mechanism or process that is known to be directly tied to the change of audio content. Several such mechanisms were discussed earlier; a channel change on a television set or an input selection change on an audio/visual receiver, for example. In many cases, however, no such external mechanisms may be available, and therefore the system must rely on analyzing the audio signal itself to determine if a reset should be triggered. Suppose, for example, that the disclosed invention resides in a television set to which a user has connected an external cable set-top box. The cable set-top box is responsible for tuning-in and changing channels, and the decoded audio is merely passed on to the television as a continuous audio stream. The television therefore receives no explicit information about when a new channel is selected. The only information available is that which may be inferred from the audio stream it receives.

A relatively simple but effective way to trigger a reset through signal analysis is to detect an occurrence in the temporal evolution of the audio signal in which the level of the audio signal remains below a threshold, $L_{silence}$, throughout a minimum time period (a time interval no shorter than a time threshold $t_{silence}$). In other words, to detect a period of relative silence having a duration at least as long as a specified threshold time interval. Such an occurrence is likely to indicate an abrupt change or discontinuity in the content of the audio. Many devices, cable set-top boxes in particular, mute the audio signal for a short period of time when the audio source changes. The act of muting reduces the audio signal to a level far lower than is generally encountered in normal audio content. In the case of digital audio, for example, it may reduce the samples of the signal to zero. Such conditions in the temporal evolution of the signal may be detected through analysis of the short-term level $L[t]$ discussed earlier. If $L[t]$ remains below than a threshold $L_{silence}$

- 17 -

throughout a time interval of at least t_{silence} , then a reset may be triggered. This approach may be expressed as:

$$R[t] = \begin{cases} 0 & s[t] < t_{\text{silence}} \\ 1 & s[t] \geq t_{\text{silence}} \end{cases}$$

5 (9a)

where a silence count signal $s[t]$ is updated according to:

$$s[t] = \begin{cases} 0 & L[t] > L_{\text{silence}} \\ s[t-1] + 1 & L[t] \leq L_{\text{silence}} \end{cases}$$

10 (9b)

Assuming that 0 dB corresponds to the level of a full scale digital square wave, then setting L_{silence} equal to -90 dB, for example, works well in practice, although the level is not critical.. Setting t_{silence} to a time of 0.25 seconds, for example, is a reasonable
15 choice for many applications, although the time period is not critical.

In some situations, a gap of relative silence may not be introduced when the audio content abruptly changes or has a discontinuity, and therefore the above technique may not be effective in triggering a reset. An alternative reset triggering occurrence to deal with additional cases is therefore desirable. In any situation in which the audio content
20 abruptly changes or has a discontinuity, the level of the new audio will be either greater than, equal to, or less than the level of the previous audio content. Because most dynamics processors employ an attack that is significantly faster than its release, the case in which the level of the new audio content is greater than or equal to the old is less problematic than when the new audio content is significantly lower than the old. In the
25 latter case, the slow release time may result in an objectionably long adaptation time. Therefore, triggering a reset when the new audio content is significantly lower than the old may be considered particularly desirable. To do so, one may exploit a feature common to most audio signals composed of natural sounds. Natural sounds may exhibit abrupt increases in level, but they generally exhibit a more gradual decay. Sounds that
30 are cut off abruptly are not very common in most audio recordings. However, if audio

- 18 -

content switches abruptly, through an input selection change for example, and the level of the new audio input is significantly lower than the old, then the resulting audio stream exhibits a drastic, unnatural, instantaneous drop in level. Such a drop in level may be detected from the short-term level $L[t]$ and used to trigger a reset.

Thus, another effective way to trigger a reset through signal analysis is to detect an occurrence in the temporal evolution of the audio signal in which the level of the audio signal decreases by an amount greater than a difference threshold L_{drop} within a time interval no greater than a time threshold t_{drop} . More specifically, if the difference between $L[t]$ and $L[t-1]$ is more than a specified threshold L_{drop} within a time interval t_{drop} , then a reset is triggered:

$$R[t] = \begin{cases} 1 & L[t] - L[t-1] < L_{drop} \\ 0 & \text{otherwise} \end{cases}$$

(10)

A suitable time difference t_{drop} is the time of one digital processing block. For example, with a block overlap of one half, one block time corresponds to $N/2$ samples. If $N=512$ and the sampling rate is 48000 Hz, t_{drop} is about 5 milliseconds. If $N=1024$, t_{drop} is about 10 milliseconds. A level difference L_{drop} of about -10 to -20 dB has been found suitable, although the level difference is not critical.

The above technique is effective in triggering a reset when the level of new audio content is significantly lower than the old, but it may sometimes be susceptible to falsely triggering a reset during normal audio content. The described technique looks for a drop in the total level of the signal, and sometimes a dominant portion of the audio signal, a kick drum for example, may exhibit such behavior, while the remaining portion of the signal does not. The dominant signal component, however, contributes most significantly to the overall level and therefore may cause a reset to be triggered. To deal with such unintended triggering, the detection scheme may be improved by detecting a reset when the level of the signal exhibits a significant level drop simultaneously across a plurality of frequency bands (for example, all frequency bands) within the threshold time period. This way, all such frequency bands contribute equally to the reset detection process, regardless of the absolute level in any given band. To implement such a technique, the instantaneous level within a plurality of frequency bands is first computed. This level is represented as $L[b, t]$, where b indicates the band index. It may be generated efficiently

- 19 -

by computing the Discrete Fourier Transform (DFT) of each audio block $x[n, t]$ and multiplying the DFT with a bandpass frequency response $C_b[k]$ for each band b :

$$L[b, t] = 10 \log_{10} \left(\sum_{k=0}^{N-1} C_b^2[k] |X[k, t]|^2 \right)$$

5 (11a)

where $X[k, t]$, the DFT of $x[n, t]$, is computed as:

$$X[k, t] = \sum_{n=0}^{N-1} x[n, t] e^{-j \frac{2\pi kn}{N}}$$

10 (11b)

The set of bandpass frequency responses $C_b[k]$ may be advantageously chosen to mimic the critical band filtering observed along the basilar membrane in the human ear. As suggested by Moore and Glasberg("A Model for the Prediction of Thresholds, Loudness, and Partial Loudness," Brian Moore, Brian Glasberg, and Thomas Baer, J. Audio Eng. Soc., Vol. 45, No. 4, 1997 April), such filters exhibit an approximately rounded exponential shape and are spaced uniformly on the Equivalent Rectangular Bandwidth (ERB) scale. FIG. 5 depicts a suitable set of filters with a spacing of 1 ERB, resulting in a total of 40 bands.

20 From the instantaneous level per band, $L[b, t]$, the difference between successive time blocks, averaged across all bands may be computed:

$$D[t] = \frac{1}{B} \sum_{b=1}^B L[b, t] - L[b, t-1]$$

(12)

25

For real world audio signals, $D[t]$ is extremely small only if the level of the signal drops significantly in the majority of bands b . If the difference is less than some threshold, D_{drop} , then a reset is detected:

- 20 -

$$R[t] = \begin{cases} 1 & D[t] < D_{drop} \\ 0 & otherwise \end{cases}$$

(13)

In practice, setting D_{drop} between -10 and -20 dB works well, although the setting is not critical. The resulting multiband technique is less susceptible to falsely triggering a reset than the simpler technique that looks only at the difference in total signal level.

FIG. 6 shows a block diagram of an AGC embodying aspects of the present invention. A digital audio input signal $x[n, t]$ (see Eqn. 1) is applied to two paths of an AGC device or process. In the control path, a "Compute Level" device or process 20 computes a measure of the audio, which may be a time-varying instantaneous level of the signal $L[t]$, computed as the mean square power of each block of the input signal $x[n, t]$ (see Eqn. 3). Next, the instantaneous level $L[t]$ is time smoothed in a "Smooth Level" device or process 22 to generate a smoothed average level of the audio signal $\bar{L}[t]$ (see Eqn. 4a). In the absence of a reset, the time smoothing may have a fast attack and a slow release characteristic. The Smooth Level 22 may be controlled by a "Time Constant Control Signal $c[t]$ " input and/or an "Initialize $\bar{L}[t]$ " input. The Time Constant Control Signal $c[t]$ input may cause one or more of the time constants employed in the Smooth Level 22 to be altered upon the occurrence of a reset, as described above. The Initialize $\bar{L}[t]$ input may cause the output of the Smooth Level 22 to assume a stored or default value $\bar{L}_{init}$ upon occurrence of a reset, as described above. The Smooth Level 22 output $\bar{L}[t]$ is applied to an "Apply Input/Output Function F_{AGC} " device or function 24 and to a "Compute Gain" device or function 26. Device or function 24 applies the function described in connection with FIG. 4 to provide a desired output level $\bar{L}_{out}[t]$ (see Eqn. 5). The Compute Gain 26 computes the gain $G[t]$ from the difference between the output level $\bar{L}_{out}[t]$ and the input level $\bar{L}[t]$ (see Eqn. 6). When $\bar{L}_{out}[t] > \bar{L}[t]$, the gain is greater than one, meaning the signal is boosted, and when $\bar{L}_{out}[t] < \bar{L}[t]$, the gain is less than one, meaning the signal is attenuated. The time varying gain $G[t]$, a modification parameter, is then applied to "Multiply" device or process 28 where it is then multiplied with each block of the signal $x[n, t]$ to produce a modified audio signal $y[n + tN/2]$ (see Eqn. 2). Although as described, the gain modification parameter $G[t]$ is frequency independent, a

- 21 -

frequency-dependent gain modification parameter $G[b, t]$ may be computed in which b is a band index. In this case Compute Level 20 may compute a frequency-dependent instantaneous signal level $\bar{L}[b, t]$, Smooth Level 22 may provide a frequency-dependent output $\bar{L}[b, t]$ (its control inputs Time Constant Control Signal $c[t]$ and Initialize $\bar{L}[t]$ may also be frequency dependent), Apply Function F_{AGC} may apply a frequency-dependent function, and Compute Gain 26 provides a frequency-dependent time-varying gain $G[b, t]$.

Implementation

The invention may be implemented in hardware or software, or a combination of both (e.g., programmable logic arrays). Unless otherwise specified, the algorithms included as part of the invention are not inherently related to any particular computer or other apparatus. In particular, various general-purpose machines may be used with programs written in accordance with the teachings herein, or it may be more convenient to construct more specialized apparatus (e.g., integrated circuits) to perform the required method steps. Thus, the invention may be implemented in one or more computer programs executing on one or more programmable computer systems each comprising at least one processor, at least one data storage system (including volatile and non-volatile memory and/or storage elements), at least one input device or port, and at least one output device or port. Program code is applied to input data to perform the functions described herein and generate output information. The output information is applied to one or more output devices, in known fashion.

Each such program may be implemented in any desired computer language (including machine, assembly, or high level procedural, logical, or object oriented programming languages) to communicate with a computer system. In any case, the language may be a compiled or interpreted language.

Each such computer program is preferably stored on or downloaded to a storage media or device (e.g., solid state memory or media, or magnetic or optical media) readable by a general or special purpose programmable computer, for configuring and operating the computer when the storage media or device is read by the computer system to perform the procedures described herein. The inventive system may also be considered to be implemented as a computer-readable storage medium, configured with a computer program, where the storage medium so configured causes a computer system to operate in a specific and predefined manner to perform the functions described herein.

- 22 -

A number of embodiments of the invention have been described. Nevertheless, it will be understood that various modifications may be made without departing from the spirit and scope of the invention. For example, some of the steps described herein may be order independent, and thus can be performed in an order different from that described.

Claims as further amended under Article 34 PCT

- 5 1. A method for processing an audio signal, comprising
altering the dynamics of the audio signal in accordance with a dynamics control
process,
detecting an occurrence in the temporal evolution of the audio signal in which the
level of the audio signal decreases by an amount greater than a difference threshold, L_{drop} ,
10 within a time interval no greater than a second time threshold, t_{drop} , wherein said detecting
detects the decrease in level of the audio signal in a plurality of frequency bands, and
resetting the dynamics control process in response to said detecting.
- 15 2. A method according to claim 1 further comprising:
detecting an occurrence in the temporal evolution of the audio signal in
which the level of the audio signal remains below a threshold, $L_{silence}$,
throughout a time interval no shorter than a first time threshold, $t_{silence}$, and
resetting the dynamics control process in response to said detecting.
- 20 3. A method according to claim 1 or claim 2 wherein resetting the dynamics control
process speeds up the rate at which the dynamics processing adapts to changes in the input
signal.
- 25 4. A method according to any one of claims 1-3 wherein resetting the dynamics
control process sets the value of one or more process-controlling parameters or a signal of
which such one or more process-controlling parameters are a function to a stored or default
value.
- 30 5. A method according to any one of claim 1-4 wherein resetting the dynamics control
process shortens the value of one or more time constants employed in determining one or
more process-controlling parameters or a signal of which such one or more process-
controlling parameters are a function.

PCT/US2007/020747

- 24 -

A8/26622 WO EP

6. A method according to any one of claims 1-5, further comprising
receiving an externally-supplied reset trigger signal indicating a change in the audio
signal, and

5 resetting the dynamics control process in response to the reset trigger signal.

7. A method according to claim 6 wherein the reset trigger signal indicates one or
more of

10 the act of a user changing a channel,
the act of a user changing input sources,
the act of a user selecting a play, rewind or forward function,
the switching of one file to another,
a change in program,
the switching of one audio coding format to another, and
15 a change in coding parameters.

8. Apparatus comprising means adapted to carry out the steps of the method of any
one of claims 1 through 7.

20 9. A computer program, stored on a computer-readable medium causing a computer
to carry out the steps of the method of any one of claims 1 through 7.