(12) UK Patent Application (19) GB (11) 2503867 (13) A

(43) Date of A Publication 15.01.2014

(21) Application No: 1208012.3

(22) Date of Filing: 08.05.2012

(71) Applicant(s):
Queen Mary & Westfield College, University of London
(Incorporated in the United Kingdom)
Mile End Road, LONDON, E1 4NS, United Kingdom

(72) Inventor(s):
Joshua Daniel Reiss
Stuart P Mansbidge
Alice Clifford
Zheng Ma
Sina Hafezi
Nicholas Jillings

(74) Agent and/or Address for Service:
Kilburn & Strode LLP
20 Red Lion Street, LONDON, WC1R 4PJ,
United Kingdom

(51) INT CL:
G10H 1/02 (2006.01) ***H04R 5/04*** (2006.01)
H04S 7/00 (2006.01)

(56) Documents Cited:
GB 2500790 A **WO 2011/034520 A1**
US 7825322 B1 **US 20120130516 A1**
US 20090049979 A1

(58) Field of Search:
INT CL **G10H, H04S**
Other: **EPODOC, WPI**

(54) Title of the Invention: **Audio processing**
Abstract Title: **Mixing and processing audio signals in accordance with audio features extracted from the audio signals**

(57) A method, apparatus, and computer readable medium for mixing a plurality of audio signals 101a, 101b, 101c, comprises: receiving a plurality or audio signals and extracting, preferably in parallel and in real-time, at least one audio feature from each of the plurality of audio signals. For each of the audio signals, an associated processing control function such as stereo pan position is determined, preferably from a predetermined processing rule, in accordance with a plurality of the extracted audio features and applied to each audio signal. The processed audio signals are then outputted 106. The audio features may be extracted 102a, 102b, 102c, on a frame-by-frame basis using a moving average filter. The audio features may include: loudness, loudness range, spectral masking, spectral centroid, spatial masking and a spectral balance. A maximum pan range of the audio signals may be determined and the processing control function may adjust the stereo pan position to optimise the panning width. The method provides autonomous multi-track audio production for use by sound engineers in the fields of both sound recording in a studio and live sound production at a venue.

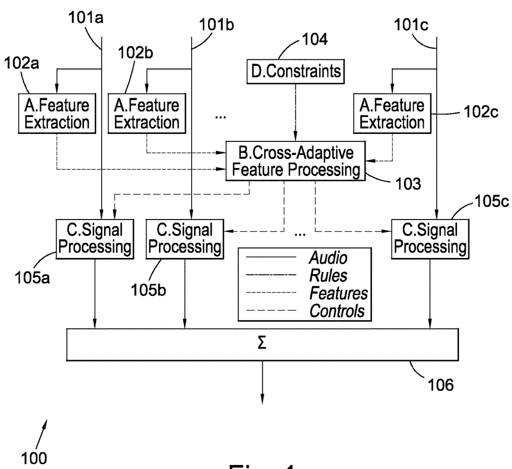


Fig. 1

This print incorporates corrections made under Section 117(1) of the Patents Act 1977.

At least one drawing originally filed was informal and the print reproduced here is taken from a later filed formal copy.

GB 2503867 A

19 09 13

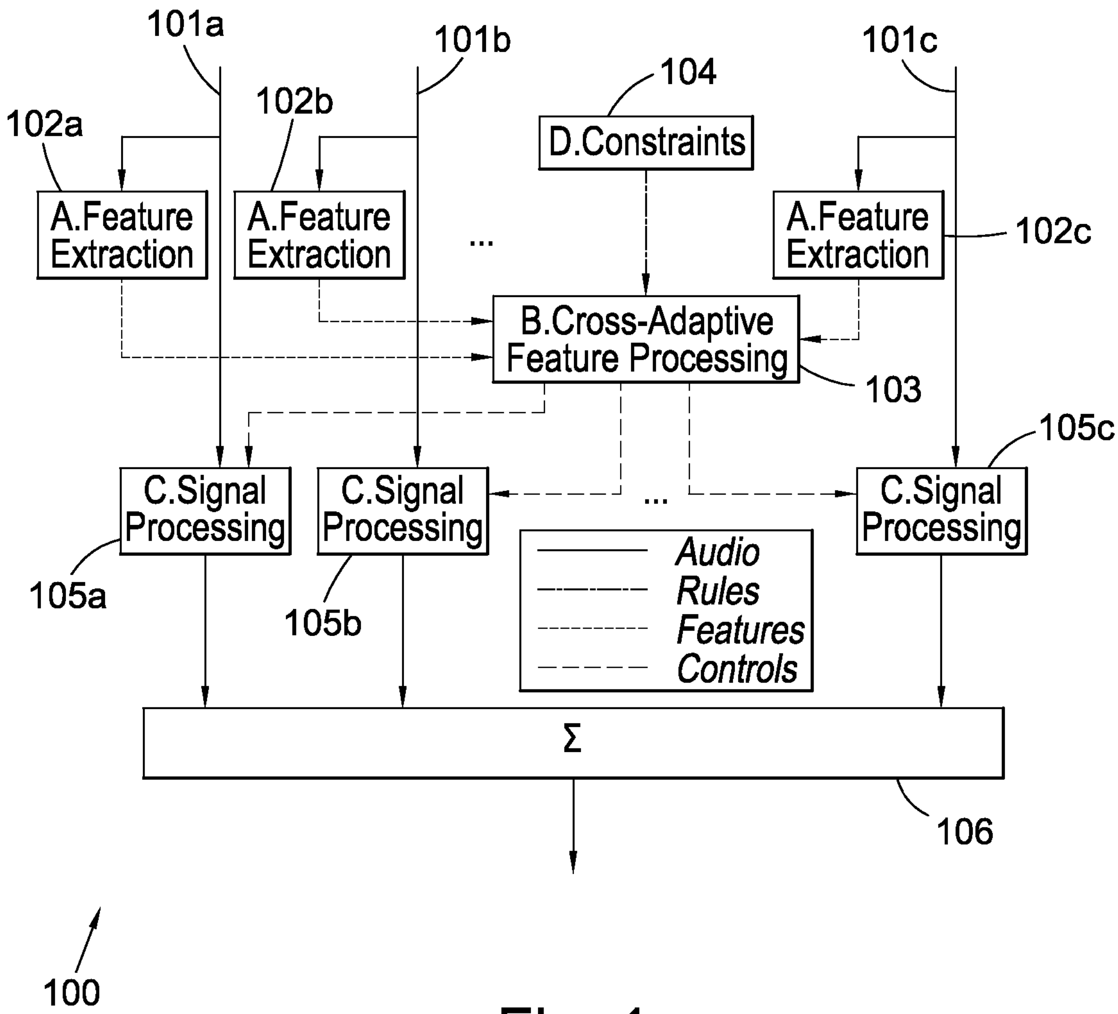
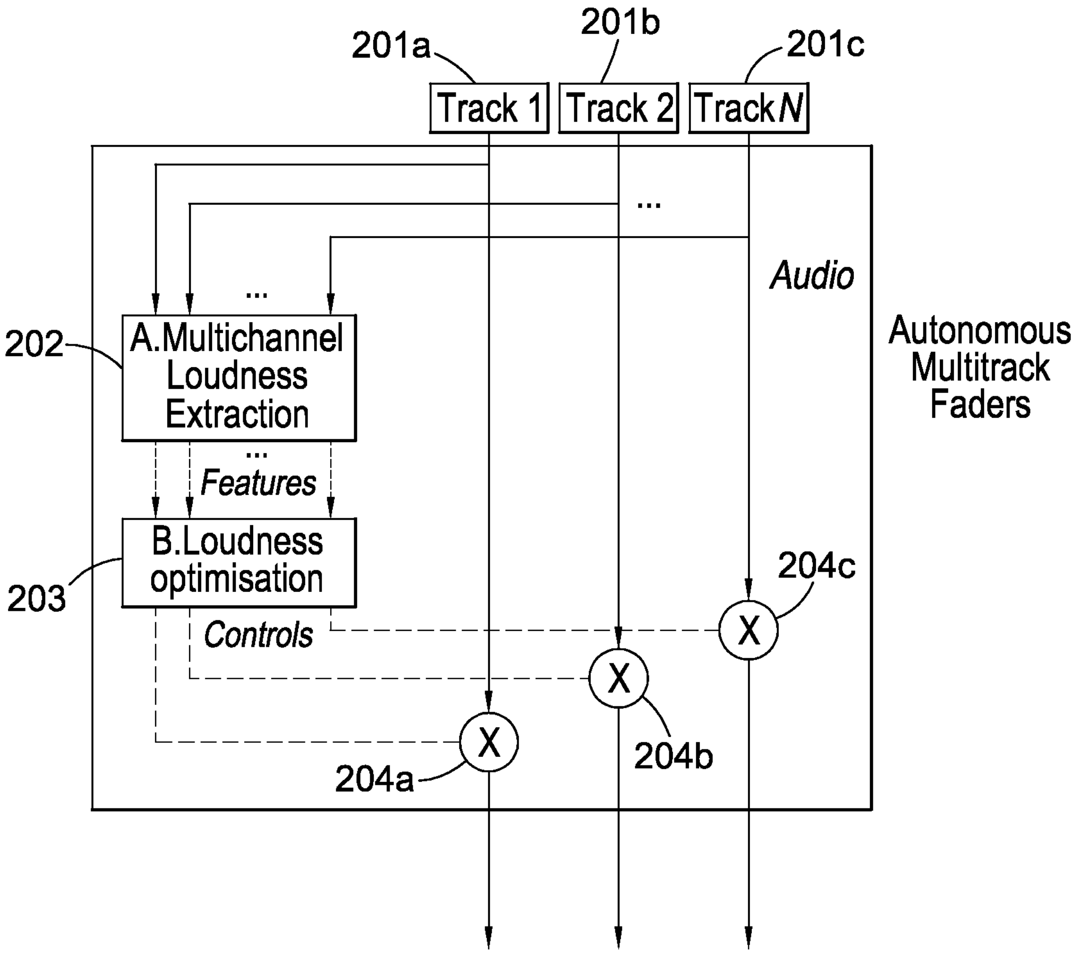

101a
101b
101c
104
102a
102b
D.Constraints
A.Feature Extraction
A.Feature Extraction
...
A.Feature Extraction
102c
B.Cross-Adaptive Feature Processing
103
105c
C.Signal Processing
C.Signal Processing
...
C.Signal Processing
105a
105b
Audio
Rules
Features
Controls
Σ
106
100

Fig. 1

19 09 13



201a
201b
201c
Track 1
Track 2
Track N
...
Audio
Autonomous
Multitrack
Faders
202
A.Multichannel
Loudness
Extraction
Features
B.Loudness
optimisation
203
Controls
204a
204b
204c
X

Fig. 2

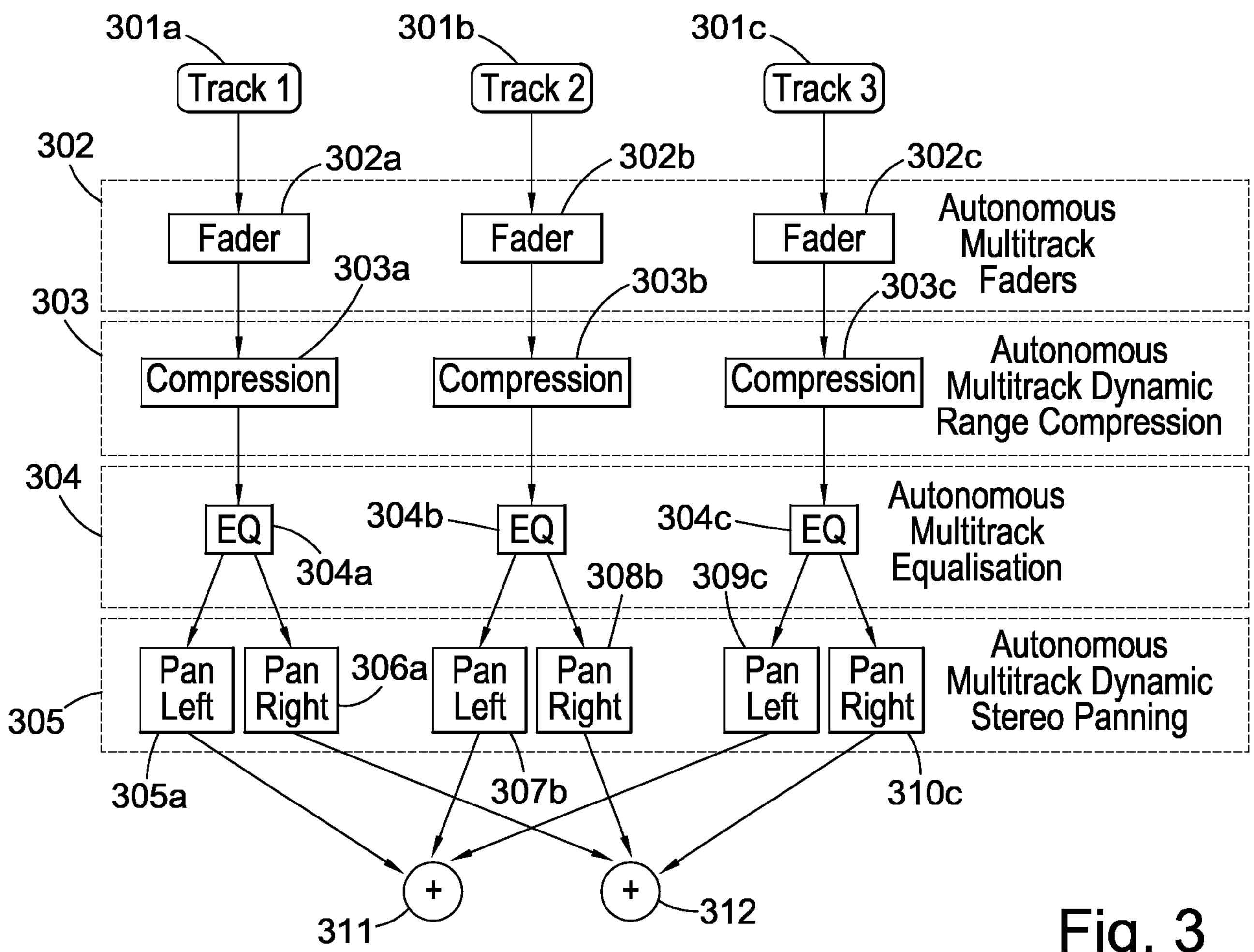

301a
301b
301c
Track 1
Track 2
Track 3
302
302a
302b
302c
Fader
Fader
Fader
Autonomous
Multitrack
Faders
303
303a
303b
303c
Compression
Compression
Compression
Autonomous
Multitrack Dynamic
Range Compression
304
304a
304b
304c
EQ
EQ
EQ
Autonomous
Multitrack
Equalisation
305
305a
306a
307b
308b
309c
310c
Pan
Left
Pan
Right
Autonomous
Multitrack Dynamic
Stereo Panning
+
311
312

Fig. 3

AUDIO PROCESSING

Field of Invention

The present invention relates to the field of audio processing. More specifically, embodiments of the invention relate to multi-track audio processing. In particular, embodiments of the invention relate to autonomous multi-track audio production.

Background to the Invention

In the field of both sound recording and live sound production it is common to process multiple tracks of audio simultaneously, each track corresponding to a separate audio signal. Many live and studio multi-track audio production tasks require dynamic user adjustment of various sound editing and manipulation parameters in order to combine multiple audio tracks into a high quality mixture. In a studio environment, such multi-track processing can be time consuming, even to a very skilled audio engineer. Furthermore, in a live environment, the real-time nature of the processing means that there is scope for error when determining how to edit and adjust parameters of the multiple tracks.

A need has therefore developed to assist engineers, particularly in the live-mixing environment, in minimising the difficulty of their work.

Some digital mixing desks have been developed that allow for particular loudness fader levels, or pan positions to be saved. As such, when a live audio engineer is dealing with multiple bands, or multiple arrangements of instruments, he can pre-set the parameters, for example during a sound check, and thereby reduce some of the complexity of his job. However, there is still a need in such systems for the engineer to adjust these parameters during a performance due to fluctuations in track levels.

To date, no systems have been developed that attempt to assist the audio engineer with managing the control of multiple parameters for multiple tracks. There is therefore a need to provide a multi-track audio production system capable of overcoming at least some of the aforementioned problems.

Summary of Invention

Embodiments of the present invention attempt to mitigate at least some of the above-mentioned problems.

In accordance with an aspect of the invention there is provided a method for mixing a plurality of audio signals within an audio mixing system. The method comprises receiving a plurality of audio signals, extracting at least one audio feature from each of the plurality of audio signals, determining, for each of the audio signals, an associated processing control function in accordance with a plurality of the extracted audio features, processing each audio signal in accordance with each associated processing control function, and outputting the plurality of processed audio signals.

The processing control functions may be determined in accordance with at least one processing rule. The at least one processing rule may be set prior to performing the mixing method. Furthermore, the at least one processing rule may comprise a plurality of rules, wherein at least one processing rule is provided for one audio signal of the audio signals, and at least one processing rule is provided for a plurality of audio signals of the audio signals.

The audio features may be extracted in parallel from the audio signals. Furthermore, the audio features may be extracted on a frame-by-frame basis. A plurality of the extracted audio features may be used to determine each processing control function derive from a frame. A plurality of the extracted audio features deriving from an audio signal of the audio signals may be filtered over a plurality of frames. The extracted features may be filtered using a moving average filter.

The audio features may include one or more of a loudness, a loudness range, a spectral masking, spectral centroid, a spatial masking, and a spectral balance. Audio characteristics such as stereo pan position and equalisation may be obtained from the audio features. Equalisation may be obtained from the spectral centroid and the spectral balance.

One of the audio features may be an audio signal loudness. Furthermore, the determining of each processing control function may further comprise determining a number of the audio signals that are currently active, summing the loudness of each of the currently active audio signals, dividing the result of the loudness summation by the number of currently active audio signals to determine an average loudness, and determining each processing control function in accordance with the average loudness.

The determining of each processing control function may further comprise determining frequency characteristics associated with each audio signal, setting the processing control function so that a stereo pan position of each audio signal is set in accordance with the determined frequency characteristics of the respective audio signal.

The stereo pan position of each audio signal may be set so that the higher the frequency characteristics of an audio signal, the further the audio signal is moved from a central pan position.

The determining of each processing control function may further comprise identifying a first audio signal and the frequency characteristics associated with the first audio signal, determining a second audio signal of the audio signals having associated frequency characteristics closest to the frequency characteristics associated with the first audio signal, and setting the processing control function so that a stereo pan position of one of the first and second audio signals is moved to an opposing stereo pan position compared to a stereo pan position of the other of the first and second audio signals.

In addition, the setting of the processing control function so that a stereo pan position of each audio signal is set in accordance with the determined frequency characteristics of the respective audio signal may further comprise combining the frequency characteristics of the audio signals, determining a stereo pan position of a centre point of the combined frequency characteristics, determining if the stereo pan position of the centre of the combined frequency characteristics is within predetermined limits, and adjusting the stereo pan position of one or more of the audio signals in order to bring the stereo pan position of the centre point of the combined frequency characteristics within the predetermined limits when the centre of the combined frequency characteristics is determined not to be within the predetermined limits.

The setting the processing control function so that a stereo pan position of each audio signal is set in accordance with the determined frequency characteristics of the respective audio signal may also further comprise combining the audio signals into a combined stereo audio signal, determining a peak magnitude of a left channel and a right channel of the combined stereo audio signal, determining if a ratio of the peak magnitude of the left and right channels of the combined stereo audio signal is within predetermined limits, and adjusting the stereo pan positions of one or more of the audio signals so that the ratio of the peak magnitude of the left and right channels of the combined stereo audio signal is within the predetermined limits when the ratio of the peak magnitude of the left and right channels of the combined stereo audio signal is determined not to be within the predetermined limits.

Furthermore, the determining of the processing control function may further comprise determining a maximum pan range of the audio signals, and setting the processing control function to adjust the stereo pan position of each audio signal to optimise the panning width.

The method may further comprise combining the plurality of processed audio signals to provide an output audio signal.

The method may be performed in real-time.

In accordance with another aspect of the invention there is provided apparatus arranged to perform one or more of the methods disclosed herein.

In accordance with yet another aspect of the invention there is provided a computer readable medium implementable on a computer and operable, in use, to perform one or more of the methods disclosed herein.

Embodiments of the invention relate a method and corresponding apparatus for automatically producing mixed content from multi-track audio.

Embodiments of the invention provide multiple inputs to and multiple outputs from the system.

Embodiments of the invention relate to extraction of features from a plurality of target audio tracks, and then analysing these features and the relationship between these features. Pre-determined constraints may dictate how the target tracks are modified based on the relationships of all target tracks and their features. Processing instructions may then be provided, in accordance with analysis of the tracks and the relationship between the tracks along with the pre-determined constraints. The plurality of audio tracks may then be processed in accordance with these instructions.

Embodiments of the invention provide a system in which features of audio signals are extracted and controls for processing the audio signals are updated on a sample-by-sample basis. This minimises system latency. The extraction of features and updating of controls for processing may be performed as a side chain operation in real-time, which does not interrupt the signal flow.

Embodiments of the invention are arranged to extract features of the audio signal(s) such as gain, loudness, loudness range, spectral masking, spatial masking, spectral balance, or spatial balance. These features and their relationship to each other may then be analysed. The system may adjust the content of the audio streams based on one or more optimisation criteria.

Embodiments of the invention provide an autonomous processing system, whereby control values for each audio track are yielded simultaneously without the need for manual control.

Embodiments of the invention provide multi-track audio production techniques which comprise a set of rules, which allow for automatic editing and combining of many parallel tracks of audio. Such techniques may be real-time operations.

Embodiments of the invention use priority schemas to set relationships between different audio sources.

Embodiments of the invention utilise perceptual and psycho-acoustic phenomena when determining how to process the audio tracks.

Embodiments of the invention relate to a multi-track mixing method comprising the steps of: inputting multiple audio signals to multiple tracks; extracting one or more features from each audio track; analysing the extracted features to determine the relationship between each audio signal; modifying the audio signal of two or more tracks according to a predetermined target model; and outputting one or more modified audio signal(s).

Embodiments of the present invention offer significant practical advantages by allowing autonomous control of a desired audio feature, such as loudness.

Embodiments of the invention can utilise any signal feature, not just signal level and can do this with many tracks simultaneously.

Embodiments of the present invention are implemented in software. Other embodiments of the invention are implemented in hardware. In particular, some embodiments of the invention are implemented within a solid-state device.

It should be appreciated that while spectral flux, crest factor, spectral centroid, and spectral spread are known features of audio signals, such features are not known for being used in audio mixing as they are in embodiments of the invention.

Brief Description of the Drawings

Exemplary embodiments of the invention shall now be described with reference to the drawings in which:

Figure 1 is a block diagram illustrating the functional components of a multi-track mixing system;

Figure 2 is a block diagram illustrating the functional components of a loudness processing portion of a multi-track mixing system; and

Figure 3 is a block diagram illustrating the functional components of an alternative multi-track mixing system.

Throughout the description and the drawings, like reference numerals refer to like parts.

Specific Description

A system arranged to automatically produce mixed audio content from multi-track audio shall now be described with reference to Figure 1.

The system 100 is arranged to perform the automated audio mixing by carrying out the following steps:

1. **Receive input signals:** digital audio signals from multiple tracks are received at an input of the system and routed to multiple parallel signal processing channels of the system;
2. **Feature extraction:** each of the digital audio signals is analysed and specific features of each of the digital audio signals are extracted;
3. **Feature Analysis:** the extracted features and the relationship between extracted features of different signals is analysed, then in accordance with one

or more processing control rules the processing required for each track is determined;

4. **Signal Processing:** The audio signals are then processed in accordance with the feature analysis; and
5. **Output processed signals:** the processed signals are then output as modified digital audio signals corresponding to each track.

The automated mixing process, including each of the above-mentioned steps, shall now be described in detail.

An input of the system is arranged to receive a plurality of stereo digital audio signals 101a, 101b, 101c, each signal corresponding to an audio track to be processed. Each stereo audio signal has a left and a right channel. The input of the system receives each track as a separate audio signal. The system is arranged to accept any number of input audio tracks; the number of tracks only being limited by the processing capability of the system and the requirements of the audio to be output.

In this embodiment of the invention the received audio data is processed in real-time. Such real-time processing is particularly useful when the received signals are real-time signals recorded live or deriving from streamed content.

Feature extraction is performed on the streaming audio in real-time as the audio is received. The features of the audio to be extracted includes features or characteristics of the audio signal such as gain loudness, loudness range, spectral masking, spatial masking, spectral balance, spatial balance, and others, as will be discussed in more detail.

The received audio signals are passed into a parallel processing operation or side-chain for the extraction and analysis of audio features. A plurality of feature extraction modules 102a, 102b, 102c provide such parallel feature extraction.

Instantaneous feature values are extracted by the feature extraction modules 102a, 102b, 102c on a sample-by-sample or frame-by-frame basis, depending on implementation. In the latter case, frame size is as low as required to ensure real-time operation with minimal latency. Accumulative averaging is applied to features to implement real-time feature estimation, the rate of which adjusts according to frame size and sample rate, which is carried out closely following the latest update of the feature value.

The extracted stream of data indicative of the certain features of an audio signal is smoothed over time using an exponential moving average filter with associated time attack and release constants, as shown by equation 1:

$$F_m(n+1) = (1-\alpha)F'_m(n+1) + \alpha F_m(n)$$

Equation 1

Wherein F'_m represents an instantaneous estimation of a feature from the m^{th} track, F represents the smoothed feature estimation, n is the current sample being processed, and α is a constant between 0 and 1 that determines the weighting of recent samples in the smoothed feature estimation. Alpha values adjust according to frame size/sample rate ratio to ensure a non-varying filter response.

The cross-adaptive multi-track feature processing module 103, shown in Figure 1, receives each of the features extracted by each of the feature extraction modules 102a, 102b, 102c. The cross-adaptive processing module determines processing control functions which dictate the processing operations to be applied to each of the tracks. The processing control functions are also determined based on pre-determined constraints or rules, along with the extracted features. The predetermined constraints may be set by a user prior to starting the mixing process and stored in a constraints module 104. The processing rules may set certain required relationships between tracks, or upper/lower limits for specific features. Constraints include, but are not limited to, the following:

- For autonomous multi-track faders, all active sources tend towards equal perceived loudness;
- For autonomous multi-track stereo positioning, all tracks are positioned such that spatial and spectral balance is maintained;
- For autonomous multi-track dynamic range compression, compressors are applied on each track such that variation in loudness range of active sources is minimised; and
- For autonomous multi-track equalisation, filters are applied on each track such that spectral bandwidth of sources does not overlap.

The cross-adaptive feature processing block 103 includes a feedback operation to ensure convergence towards the desired features in the output. That is, the controls produced by the cross-adaptive feature processing block may be analysed before they are applied. If they fail to produce the desired result within a given tolerance, then the control values are adjusted before they are applied.

The processing control functions take the form of time varying filters, such as gains, delays, and infinite impulse response filters. More specifically, a control vector is utilised which is a weighted sum of previous control vectors and a function of the extracted features. In the case of loudness faders, multi-track processing is used to derive a decibel level control for each track. The result of this processing is then converted back to the linear domain, and applied as a time varying gain to each track, as discussed below. Similarly, in the case of autonomous stereo positioning, multi-track processing is used to derive a panning position for each track, which is then applied as two gains, producing a left and a right output for stereo positioning.

Once the above-mentioned control functions have been determined they are used to process each of the tracks in the parallel signal processing modules 105a, 105b, 105c. Each track is then output by the respective processing block 105a, 105b, 105c as a separate audio signal which has been processed in accordance with the controls determined by the cross-adaptive processing module 103. Each processed signal is then combined by a summation process into a single audio output in the output module 106. The output can be of any suitable format, but in this embodiment of the invention is a stereo output.

The three main aspects of audio signals to be mixed are: the relative loudness levels of each track on a frame-by-frame basis; the relative loudness of the audio signal over a period of time; and the stereo panning of each track (for mixing of stereo audio signals). Hence, the automated feature extraction and processing for each of these aspects of an audio signal shall now be considered in detail.

Figure 2 shows how the system 100 extracts loudness and loudness range to allow for independent control of the relative loudness levels of multiple audio tracks.

Audio signals corresponding to multiple tracks 201a, 201b, 201c of the multi-track audio processing system have information relating to their loudness extracted by a multi-channel loudness extraction module 202 at each sample of frame. The multi-channel loudness extraction module 202 takes the perceptual loudness of all tracks into consideration when determining the associated loudness. A loudness optimisation module 203 then determines the control functions to be applied to one or more of the tracks, as appropriate, in accordance with the perceptual loudness determination. The tracks to have their loudness altered are then altered by the respective processing modules 204a, 204b, and 204c.

For the purpose of the calculations the standard loudness model, ITU-R BS.1770-2, has been modified with several novel improvements designed to enhance its use in the system, as discussed below.

Filter coefficients are mapped to continuous time values, and then remapped to a sample rate of the multi-track audio. This allows the loudness measure to work with any sample rate.

While the loudness is extracted at each sample or frame in accordance with the EBU R 128 standard, the standard extraction technique is customised for smooth real-time loudness estimation. An overlapping window approach, as in EBU R 128 and in the gating of ITU-R BS.1770-2, is not used for estimation of loudness. Instead, the exponential moving average filter of Equation 1 is applied, with a time constant set equal to the 3000ms window length specified in EBU R 128 for gating, i.e., detection of an active signal, and for short-term loudness estimation.

During the feature extraction, the system is arranged to distinguish between silences and wanted audio and is also able to accommodate for silent portions within the audio, with a binary state of activity for each track determined by the current loudness value immediately after its calculation. The silence threshold can be estimated based on relative levels in the audio stream. Adaptive thresholds are used to gate out any periods of silence or background noise. In particular, two independent gates are used for silence detection, a gate for determining when silence begins, set at -30LUFS (Loudness Units Full Scale) and -25LUFS for determining when silence ends. Use of two gates in this way prevents random noise resulting in fluctuating loudness estimation, and hence prevents over adjustment of the control vector in the processing stage. This silence processing is carried out as part of the feature extraction block, and included in loudness estimation to ensure that it does not interfere with signal flow through the system. Thus, computationally intensive operations can be performed without any interference with the real-time signal flow.

Information relating to masking between tracks is also obtained for determination of equalization curves, based on a novel, low computational complexity, measurement of multi-track masking of frequency content in the spectral domain derived from a measure of overlapping spectral spread of each track. Spectral spread and spectral centroid are measured on a frame by frame basis for each track. Thus a bandwidth is identified, with a mid-point frequency at the spectral centroid and covering a frequency range given by the spectral spread. The amount of overlap in bandwidth between sources is used as an indicator of masking.

The processing control functions that are developed from the extracted information for the loudness adjustment processing also take into consideration a psychoacoustic model, which has been validated by subjective evaluation and shown to improve the perceived quality of the resultant mixed audio. The psychoacoustic model includes: biquadratic Infinite Impulse Response (IIR) filters which are applied to the audio signals that correspond to the perceived loudness of each signal at different frequencies; a shelving 'pre-filter' which models the acoustic effect of the head, i.e. a head related transfer function (HRTF); and a high pass 'RLB filter' to model the response of the human ear.

In addition to ensuring a good balance of the relative loudness of each track of a multi-track audio stream at any instant in time, it is also important to monitor and adjust the relative loudness of each track, and the mix as a whole, over time. Without such monitoring and adjustment unnatural or unwanted jumps in loudness may occur.

Loudness range is therefore extracted for production of multi-track dynamic range compression. The loudness range extraction means is adapted from the EBU R 128 standard and is customised for smooth real-time estimation. Furthermore, the spectral flux and crest factor are extracted for estimation of dynamic range compression parameters.

The system 100 utilises these extracted features to provide multi-track dynamic loudness processing as follows. A count of currently active tracks is maintained, as determined by a noise gate. That is, a noise gate is used to determine which tracks have a signal below a threshold loudness indicative of an inactive track. The controls applied to inactive tracks are kept constant until the tracks becoming active again. Hence, smoothing of a track's variables begins only when it first becomes active to prevent a fade effect. The total absolute loudness of all tracks is then summed and divided by the track count to produce a dynamic target loudness, or average loudness, that allows for intended fluctuations in the overall mix signal level. An exponential moving average filter is used to provide a smoothly varying value. Controls (in the decibel domain) are then calculated as a ratio of the track loudness to the average loudness. As with other variables, the decibel level control values are smoothed using an EMA filter. Lead track(s), typically vocals, can be selected for a gain boost to ride over the rest of the mix.

The extraction and processing of the stereo positioning of the multi-track audio signals shall now be considered in detail.

The stereo positioning functionality of the multi-track audio processing system employs the same loudness and noise gate processing used for determining active tracks so that inactive tracks bypass the processing and the track's parameters therefore remain constant. The stereo pan processing then works as follows.

Spectral centroid information is extracted for determination of the stereo positioning of a signal. This spectral centroid defines a panning factor associated with an audio signal, set by a ratio against the maximum spectral centroid found over all channels, with the result that higher frequencies are progressively panned further from the centre of the panning space.The spectral centroid of each active track is calculated using a Fast Fourier Transform (FFT). Initially, as a track enters for the first time, a determination is made of which of the other tracks has the closest spectral centroid. Then each couple of tracks related by the closeness of their spectral centroid are then set to opposing left or right channels of the stereo output. Hence, this operation prevents tracks from crossing over the 0.5 centre point. If this process results in a poor distribution of tracks over the left and right channels then the system will perform this operation again until an acceptable distribution is provided. As source distribution problems typically occur on the rare occasion that multiple tracks enter the mix at exactly the same time, repeating the operation after the tracks have entered ensures a better source distribution. As a fail-safe measure the program will only allow a small number of attempts before the distribution is fixed.

The system is also arranged to determine a maximum spectral centroid range found over both channels of all tracks. A maximum spectral centroid parameter indicative of this range is then stored in a memory of or associated with the system and updated as and when this range changes. The panning factors of each audio track can then be adjusted to further spread the tracks across the left and right channels so that the full panning space is utilised even when the full frequency spectrum is not. The extent of the panning width can therefore be controlled by the user or automatically in order to maximise the use of the available spatial audio width.

Spectral and spatial balance are also analysed at the output of the stereo mix in order improve the quality of the audio output.

For spectral balance, FFTs of the left and right output channels are split into five frequency bands, and the magnitude of each band's frequency bins are cumulatively summed per channel. The arctangent of the ratio between the left and right summation produces a spectral balance angle for each band. If the spectral balance angle of a mix is below 0.45 or above 0.55, then pan locations of each track are adjusted to push the spectral balance angle back towards the 0.5 centre point. The adjustment of pan locations is performed with a ratio so that tracks with spectral centroid locations closest to the centre frequency of each band are moved the furthest, and those tracks outside a certain bandwidth (determined by a fixed Q-factor of 0.3 from the centre frequency) are not moved at all.

Spatial balance is improved by analyzing the ratio of the peak magnitude of output left and right channels, and moving all sources indiscriminately by a small factor, provisionally0.02, if the ratio of the peak magnitude of output left and right channels is outside the same allowed tolerance as described for the spectral balance above (below 0.45 or above 0.55). Typically, however, the adjustment of pan locations for spectral balance ensures the overall spatial balance is kept within the allowed tolerance.

Combined, the system continuously adapts to produce a balanced mix, with the intent to maximize panning as far as possible up to the limits determined by each track's spectral centroid. All parameters, including the final pan controls are passed through EMA filters to ensure that they vary smoothly. Lead track(s), typically vocals, can be selected to bypass the panning algorithm and be fixed in the centre of the mix.

Figure 3 illustrates an alternative embodiment of the invention which performs the processing and mixing as a series operation for autonomous, real-time, low latency multi-track audio production.

Each track 301a, 301b, 301c is received by the multi-track processing system and firstly undergoes loudness processing in the loudness processing module 302. A loudness processor 302a, 302b, 302c is associated with each individual track and performs the actual processing of the loudness characteristics of the associated track.

The tracks 301a, 301b, 301c are then processed by the compression module 303 and each compression processor 303a, 303b, 303c of the compressions module 303 associated with each track, and then by an EQ module 304 and the associated EQ processors 304a, 304b, 304c The processed audio signals corresponding to each of the tracks 301a, 301b, 301c are then processed by the stereo panning module 305 and each of the respective left and right stereo panning processors 305a, 306a, 307b, 308b, 309c, 310c. Finally, the processed signals are then combined into left and right signals to be output by the system.

The various different systems, operations, and methods described above may be implemented in hardware or by a computer program. When implemented by a computer program a computer could be provided having a memory to store the computer program, and a processor to implement the computer program. The computer program may include computer code arranged to instruct a computer to perform the functions of one or more of the various methods described above. The computer program and/or the code for performing such methods may be provided to an apparatus, such as a computer, on a computer readable medium. The computer readable medium could be, for example, an electronic, magnetic, optical, electromagnetic, infrared, or semiconductor system, or a propagation medium for data transmission, for example for downloading the code over the Internet. Non-limiting examples of a physical computer readable medium include semiconductor or solid state memory, magnetic tape, a removable computer diskette, a random access memory (RAM), a read-only memory (ROM), a rigid magnetic disc, and an optical disk, such as a CD-ROM, CD-R/W or DVD.

An apparatus such as a computer may be configured in accordance with such computer code to perform one or more processes in accordance with the various methods discussed above.

The audio input signals may be received from either a multi-track computer soundcard or from stored tracks on a computer. The output content may be produced as stored audio, or outputted directly to the soundcard where it may immediately be played out as analogue sound through loudspeakers.

It will be appreciated that input audio signal may also be a continuous digital data stream from which the different audio signals are extracted. Furthermore, in alternative embodiments of the invention the system is arranged to receive one or more analogue audio signals and one or more digital-to-analogue converters (DAC) of the system convert the analogue signal(s) to digital signal(s).

While the above-described embodiments of the invention relate to real-time processing it will be appreciated that embodiments of the invention relate to non-real-time processing. In such embodiments, the received audio data may be stored in a memory of the system, or an external memory associated with the system, and then processed at any time.

While the embodiments of the invention described so far relate to stereo audio signals, it will be appreciated that other embodiments of the invention relate to mixing of monaural audio signals.

The above-described embodiments of the invention are provided as an example only, and it will therefore be appreciated that various other arrangements could be provided without departing from the scope of the invention, which is only limited by the scope of the appended claims.

Claims:

1. A method for mixing a plurality of audio signals within an audio mixing system, the method comprising:

receiving a plurality of audio signals;

extracting at least one audio feature from each of the plurality of audio signals;

determining, for each of the audio signals, an associated processing control function in accordance with a plurality of the extracted audio features;

processing each audio signal in accordance with each associated processing control function; and

outputting the plurality of processed audio signals.

2. The method according to claim 1, wherein the processing control functions are determined in accordance with at least one processing rule.

3. The method according to claim 2, wherein the at least one processing rule is set prior to performing the mixing method.

4. The method according to claim 2 or claim 3, wherein the at least one processing rule comprises a plurality of rules, wherein at least one processing rule is provided for one audio signal of the audio signals, and at least one processing rule is provided for a plurality of audio signals of the audio signals.

5. The method according to any preceding claim, wherein the audio features are extracted in parallel from the audio signals.

6. The method according to any preceding claim, wherein the audio features are extracted on a frame-by-frame basis.

7. The method according to claim 6, wherein a plurality of the extracted audio features used to determine each processing control function derive from a frame.

8. The method according to claim 6 or claim 7, wherein a plurality of the extracted audio features deriving from an audio signal of the audio signals are filtered over a plurality of frames.

9. The method according to claim 8, wherein the extracted features are filtered using a moving average filter.

10. The method according to any preceding claim, wherein the audio features include one or more of a loudness, a loudness range, a spectral masking, spectral centroid, a spatial masking, and a spectral balance.

11. The method according to any preceding claim, wherein one of the audio features is an audio signal loudness, and the determining of each processing control function further comprises:

determining a number of the audio signals that are currently active;

summing the loudness of each of the currently active audio signals;

dividing the result of the loudness summation by the number of currently active audio signals to determine an average loudness; and determining each processing control function in accordance with the average loudness.

12. The method according to any preceding claim, wherein the determining of each processing control function further comprises:

determining frequency characteristics associated with each audio signal; and

setting the processing control function so that a stereo pan position of each audio signal is set in accordance with the determined frequency characteristics of the respective audio signal.

13. The method according to claim 12, wherein the stereo pan position of each audio signal is set so that the higher the frequency characteristics of an audio signal, the further the audio signal is moved from a central pan position.

14. The method according to claim 12 or claim 13, wherein the determining of each processing control function further comprises:

identifying a first audio signal and the frequency characteristics associated with the first audio signal;

determining a second audio signal of the audio signals having associated frequency characteristics closest to the frequency characteristics associated with the first audio signal; and

setting the processing control function so that a stereo pan position of one of the first and second audio signals is moved to an opposing stereo pan position compared to a stereo pan position of the other of the first and second audio signals.

15. The method according to any one of claims 12 to 14, wherein the setting of the processing control function so that a stereo pan position of each audio signal is set in accordance with the determined frequency characteristics of the respective audio signal further comprises:

combining the frequency characteristics of the audio signals;

determining a stereo pan position of a centre point of the combined frequency characteristics;

determining if the stereo pan position of the centre of the combined frequency characteristics is within predetermined limits; and

adjusting the stereo pan position of one or more of the audio signals in order to bring the stereo pan position of the centre point of the combined frequency characteristics within the predetermined limits when the centre of the combined frequency characteristics is determined not to be within the predetermined limits.

16. The method according to any one of claims 12 to 15, wherein the setting the processing control function so that a stereo pan position of each audio signal is set in accordance with the determined frequency characteristics of the respective audio signal further comprises:

combining the audio signals into a combined stereo audio signal;

determining a peak magnitude of a left channel and a right channel of the combined stereo audio signal;

determining if a ratio of the peak magnitude of the left and right channels of the combined stereo audio signal is within predetermined limits; and

adjusting the stereo pan positions of one or more of the audio signals so that the ratio of the peak magnitude of the left and right channels of the combined stereo audio signal is within the predetermined limits when the ratio of the peak magnitude of the left and right channels of the combined stereo audio signal is determined not to be within the predetermined limits.

17. The method according to any one or claims 12 to 16, wherein the determining of the processing control function further comprises:

determining a maximum pan range of the audio signals; and

setting the processing control function to adjust the stereo pan position of each audio signal to optimise the panning width.

18. The method according to any preceding claim, further comprising combining the plurality of processed audio signals to provide an output audio signal.

19. The method according to any preceding claim, wherein the method is performed in real-time.

20. Apparatus arranged to perform the method of any one of claims 1 to 19.

21. A computer readable medium implementable on a computer and operable, in use, to perform the method of any one of claims 1 to 19.

INTELLECTUAL PROPERTY OFFICE

Application No:	GB1208012.3	**Examiner:**	Rhiannon Jenkins
Claims searched:	1-21	**Date of search:**	8 November 2013

Patents Act 1977: Search Report under Section 17

Documents considered to be relevant:

Category	Relevant to claims	Identity of document and passage or figure of particular relevance
X	1-9 & 18-21	US 2012/130516 A1 (KELLETT P ET AL) - See the figures and paragraphs [0029] to [0030]
X	1-5, 10, 11 & 19-21	US 7825322 B1 (ADOBE SYSTEMS INC) - See the figures and columns 4 to 7
X	1-5, 10 & 18-21	WO 2011/034520 A1 (HEWLETT PACKARD DEVELOPMENT CO) - See the figures and paragraphs [0015] to [0030]
X,E	1-5 & 18-21	GB 2500790 A (JAGUAR LAND ROVER LTD) - See the figures and pages 3 to 8
X	1-5 & 18-21	US 2009/049979 A1 (APPLE INC) - See the figures and paragraphs [0017] to [0023]

Categories:

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.

Field of Search:

Search of GB, EP, WO & US patent documents classified in the following areas of the UKC^{X} :

Worldwide search of patent documents classified in the following areas of the IPC

G10H; H04S

The following online and other databases have been used in the preparation of this search report

EPODOC, WPI

INTELLECTUAL
PROPERTY OFFICE

International Classification:

Subclass	Subgroup	Valid From
G10H	0001/02	01/01/2006
H04R	0005/04	01/01/2006
H04S	0007/00	01/01/2006