(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2016-224449
(P2016-224449A)
(43)公開日　平成28年12月28日(2016.12.28)

(51)Int.Cl.		F I			テーマコード（参考）
G10L 21/034	(2013.01)	G10L	21/034		5J030
G10L 25/51	(2013.01)	G10L	25/51		
G10L 21/0364	(2013.01)	G10L	21/0364		
H03G 5/16	(2006.01)	H03G	5/16	165	

審査請求 未請求　請求項の数 18　ＯＬ　（全 85 頁）

(21)出願番号　特願2016-145567 (P2016-145567)
(22)出願日　平成28年7月25日 (2016.7.25)
(62)分割の表示　特願2016-505487 (P2016-505487)の分割
原出願日　平成26年3月17日 (2014.3.17)
(31)優先権主張番号　201310100422.1
(32)優先日　平成25年3月26日 (2013.3.26)
(33)優先権主張国　中国 (CN)
(31)優先権主張番号　61/811,072
(32)優先日　平成25年4月11日 (2013.4.11)
(33)優先権主張国　米国 (US)

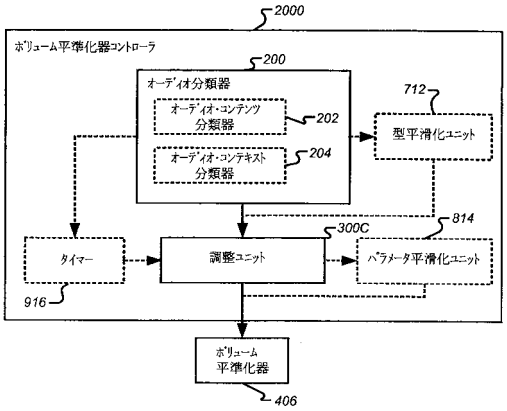
（特許庁注：以下のものは登録商標）
１．ウィンドウズ

(71)出願人　507236292
ドルビー　ラボラトリーズ　ライセンシング　コーポレイション
アメリカ合衆国　９４１０３　カリフォルニア州　サンフランシスコ　マーケット　ストリート　１２７５
(74)代理人　100107766
弁理士　伊東　忠重
(74)代理人　100070150
弁理士　伊東　忠彦
(74)代理人　100091214
弁理士　大貫　進介

最終頁に続く

(54)【発明の名称】ボリューム平準化器コントローラおよび制御方法

(57)【要約】
【課題】ボリューム平準化器コントローラおよび制御方法が提供される。
【解決手段】ある実施形態では、ボリューム平準化器コントローラは、リアルタイムでオーディオ信号のコンテンツ型を識別するためのオーディオ・コンテンツ分類器と；識別されたコンテンツ型に基づいて連続的な仕方でボリューム平準化器を調整する調整ユニットとを有する。調整ユニットは、ボリューム平準化器の動的な利得を、前記オーディオ信号の情報性のコンテンツ型と正に相関させ、ボリューム平準化器の動的な利得を、前記オーディオ信号の干渉性のコンテンツ型と負に相関させるよう構成されていてもよい。
【選択図】図２０



【特許請求の範囲】
【請求項１】
リアルタイムでオーディオ信号のコンテンツ型を識別するためのオーディオ・コンテンツ分類器と；
識別されたコンテンツ型に基づいて連続的な仕方でボリューム平準化器を調整する調整ユニットとを有する、
ボリューム平準化器コントローラ。
【請求項２】
請求項１記載のボリューム平準化器コントローラを有するオーディオ処理装置。
【請求項３】
オーディオ信号の短期的セグメントのコンテンツ型を識別する段階と；
少なくとも部分的には識別されたコンテンツ型に基づく前記短期的セグメントのコンテキスト型を識別する段階とを含む、
オーディオ分類方法。
【請求項４】
前記コンテンツ型を分類する動作が、前記短期的セグメントをコンテンツ型VoIP発話またはコンテンツ型非VoIP発話に分類することを含み、
前記コンテキスト型を識別する動作が、VoIP発話および非VoIP発話の信頼値に基づいて、前記短期的セグメントをコンテキスト型VoIPまたはコンテキスト型非VoIPに分類するよう構成されている、
請求項３記載のオーディオ分類方法。
【請求項５】
前記コンテンツ型を分類する動作がさらに、
短期的セグメントをコンテンツ型VoIPノイズおよびコンテンツ型非VoIPノイズに分類することを含み、
前記コンテキスト型を識別する動作が、VoIP発話、非VoIP発話、VoIPノイズおよび非VoIPノイズの信頼値に基づいて、前記短期的セグメントをコンテキスト型VoIPまたはコンテキスト型非VoIPに分類するよう構成されている、
請求項４記載のオーディオ分類方法。
【請求項６】
前記コンテキスト型を識別する動作が：
VoIP発話の信頼値が第一の閾値より大きい場合、前記短期的セグメントをコンテキスト型VoIPとして分類し；
VoIP発話の信頼値が、前記第一の閾値より大きくない第二の閾値より大きくない場合、前記短期的セグメントをコンテキスト型非VoIPとして分類し；
それ以外の場合には、前記短期的セグメントを最後の短期的セグメントについてのコンテキスト型として分類するよう構成されている、
請求項４記載のオーディオ分類方法。
【請求項７】
前記コンテキスト型を識別する動作が：
VoIP発話の信頼値が第一の閾値より大きい場合またはVoIPノイズの信頼値が第三の閾値より大きい場合、前記短期的セグメントをコンテキスト型VoIPとして分類し；
VoIP発話の信頼値が、前記第一の閾値より大きくない第二の閾値より大きくない場合またはVoIPノイズの信頼値が前記第三の閾値より大きくない第四の閾値より大きくない場合、前記短期的セグメントをコンテキスト型非VoIPとして分類し；
それ以外の場合には前記短期的セグメントを最後の短期的セグメントについてのコンテキスト型として分類するよう構成されている、
請求項２記載のオーディオ分類方法。
【請求項８】
前記コンテンツ型の過去の信頼値に基づいて現在の時点での前記コンテンツ型の信頼値

を平滑化することをさらに含む、請求項１記載のオーディオ分類方法。
【請求項９】
前記の平滑化する動作は、現在の短期的セグメントの信頼値と最後の短期的セグメントの平滑化された信頼値との重み付けされた和を計算することによって、現在の短期的セグメントの平滑化された信頼値を決定するよう構成されている、請求項８記載のオーディオ分類方法。
【請求項１０】
前記短期的セグメントからコンテンツ型発話を識別する段階をさらに含み、平滑化前の現在の短期的セグメントについてのVoIP発話の信頼値が、所定の信頼値として、あるいはコンテンツ型発話についての信頼値が第五の閾値より低い最後の短期的セグメントの平滑化された信頼値として、設定される、請求項９記載のオーディオ分類方法。
【請求項１１】
前記コンテキスト型を識別する動作が、特徴として、前記短期的セグメントのコンテンツ型の信頼値および前記短期的セグメントから抽出された他の特徴を使って、機械学習モデルに基づいて前記短期的セグメントを分類するよう構成されている、請求項２記載のオーディオ分類方法。
【請求項１２】
前記コンテキスト型を識別する動作が同じコンテキスト型を連続的に出力する継続時間を測定する段階をさらに含み、当該オーディオ分類方法は、新しいコンテキスト型の継続時間の長さが第六の閾値に達するまで、現在のコンテキスト型を使い続けるよう構成される、請求項６記載のオーディオ分類方法。
【請求項１３】
あるコンテキスト型から別のコンテキスト型への異なる遷移対について、異なる第六の閾値が設定される、請求項１２記載のオーディオ分類方法。
【請求項１４】
前記第六の閾値が、前記新しいコンテキスト型の信頼値と負に相関している、請求項１２記載のオーディオ分類方法。
【請求項１５】
前記第一および／または第二の閾値が、最後の短期的セグメントのコンテキスト型によって異なる、請求項６記載のオーディオ分類方法。
【請求項１６】
オーディオ信号の短期的セグメントのコンテンツ型を識別するオーディオ・コンテンツ分類器と；
少なくとも部分的には前記オーディオ・コンテンツ分類器によって識別されたコンテンツ型に基づいて前記短期的セグメントのコンテキスト型を識別するオーディオ・コンテキスト分類器とを有しており、
請求項３記載の方法を実行するよう構成されている、
オーディオ分類器。
【請求項１７】
オーディオ信号の短期的セグメントのコンテンツ型を識別するオーディオ・コンテンツ分類器と；
請求項３に従って、少なくとも部分的には前記オーディオ・コンテンツ分類器によって識別されたコンテンツ型に基づいて前記短期的セグメントのコンテキスト型を識別するオーディオ・コンテキスト分類器とを有する、
オーディオ分類器を有するオーディオ処理装置。
【請求項１８】
一つまたは複数のプロセッサによって実行されたときに請求項３記載のオーディオ分類方法を実行する命令が記憶されている非一時的なコンピュータ可読媒体。
【発明の詳細な説明】
【技術分野】

【０００１】
関連出願への相互参照
本願は2013年3月26日に出願された中国特許出願第201310100422.1号および2013年4月11日に出願された米国仮特許出願第61/811,072号の優先権を主張するものである。これら各出願はここに参照によってその全体において組み込まれる。
【０００２】
技術分野
本願は概括的にはオーディオ信号処理に関する。詳細には、本願の実施形態はオーディオ分類および処理、特にダイアログ向上器、サラウンド仮想化器、ボリューム平準化器および等化器の制御のための装置および方法に関する。
【背景技術】
【０００３】
いくつかのオーディオ改善装置は、オーディオの全体的な品質を改善し、相応してユーザーの経験を向上させるために、時間領域またはスペクトル領域のいずれかにおいてオーディオ信号を修正する傾向がある。さまざまなオーディオ改善装置がさまざまな目的のために開発されている。オーディオ改善装置のいくつかの典型的な例は次のものを含む。
【０００４】
ダイアログ向上器〔エンハンサー〕：ダイアログは、映画およびラジオまたはテレビ・プログラムにおいてストーリーを理解するための最も重要な構成要素である。特に聴力が衰えつつある高齢者のために、ダイアログの明瞭性および了解性を高めるためにダイアログを向上させる諸方法が開発された。
【０００５】
サラウンド仮想化器：サラウンド仮想化器は、PCの内部スピーカーを通じてまたはヘッドフォンを通じてサラウンド（マルチチャネル）サウンド信号がレンダリングされることができるようにする。すなわち、（スピーカーおよびヘッドフォンのような）ステレオ装置を用いて、仮想的にサラウンドの効果を生成し、消費者のために映画館の体験を提供するのである。
【０００６】
ボリューム平準化器：ボリューム平準化器は、再生時にオーディオ・コンテンツのボリュームを調整し、目標ラウドネス値に基づいて時間軸を通じてボリュームがほぼ一貫しているようにするすることをねらいとする。
【０００７】
等化器〔イコライザー〕：等化器は、「トーン」または「音色」として知られるスペクトル・バランスの一貫性を提供し、ユーザーが、ある種の音を強調したり望ましくない音を除去したりするために、個々の周波数帯域での周波数応答（利得）の全体的なプロファイル（曲線または形状）を構成設定できるようにする。伝統的な等化器では、異なる等化器プリセットが、種々の音楽ジャンルのような種々の音のために提供されてもよい。ひとたびプリセットが選択されたらまたは等化プロファイルが設定されたら、手動で等化プロファイルが修正されるまで、同じ等化利得が信号に対して適用される。対照的に、動的等化器は、オーディオのスペクトル・バランスを連続的にモニタリングし、それを所望されるトーンと比較し、オーディオの元のトーンを所望されるトーンに変換するための等化フィルタを動的に調整することによってスペクトル・バランス一貫性を達成する。
【０００８】
一般に、オーディオ改善装置はその独自の応用シナリオ／コンテキストをもつ。すなわち、オーディオ改善装置は、あらゆる可能なオーディオ信号についてではなく、ある種のコンテンツの集合についてのみ好適であることがある。異なるコンテンツは異なる仕方で処理される必要があることがあるからである。たとえば、ダイアログ向上方法は、通例、映画コンテンツに適用される。それがダイアログのない音楽に適用されたとしたら、いくつかの周波数サブバンドを誤ってブーストし、重度の音色変化および知覚的な非一貫性を導入することがありうる。同様に、音楽信号に対してノイズ抑制方法が適用されたとした

ら、強いアーチファクトが可聴となるであろう。
【０００９】
しかしながら、通例はオーディオ改善装置の集合を含むオーディオ処理システムについて、その入力は、必然的に、あらゆる可能な型のオーディオ信号でありうる。たとえば、PCに統合されたオーディオ処理システムは映画、音楽、VoIPおよびゲームを含む多様な源からオーディオ・コンテンツを受領することになる。よって、対応するコンテンツに対してよりよいアルゴリズムまたは各アルゴリズムのよりよいパラメータを適用するために、処理されるコンテンツを識別または区別することが重要になる。
【００１０】
オーディオ・コンテンツを区別して、相応してよりよいパラメータまたはよりよいオーディオ改善アルゴリズムを適用するために、伝統的なシステムは、通例、プリセットの集合を事前設計し、ユーザーは再生されるコンテンツについてのプリセットを選ぶことを求められる。プリセットは通例、映画または音楽再生のために特に設計された「映画」プリセットおよび「音楽」プリセットのように、適用されるオーディオ改善アルゴリズムおよび／またはその最良のパラメータの集合をエンコードする。
【先行技術文献】
【特許文献】
【００１１】
【特許文献１】国際公開第2008/106036号（H. Muesch、“Speech Enhancement in Entertainment Audio”）
【特許文献２】米国特許出願公開第2009/0097676A1号（A. J. Seefeldt et al.、“Calculating and Adjusting the Perceived Loudness and/or the Perceived Spectral Balance of an Audio Signal”）
【特許文献３】国際公開第2007/127023号（B.G. Grockett et al.、“Audio Gain Control Using Specific-Loudness-Based Auditory Event Detection”）
【特許文献４】国際公開第2009/011827号、（A. Seefeldt et al.、“Audio Processing Using Auditory Scene Analysis and Spectral Skewness”）
【非特許文献】
【００１２】
【非特許文献１】L. Lu, H.-J. Zhang, and S. Li、"Content-based Audio Classification and Segmentation by Using Support Vector Machines"、ACM Multimedia Systems Journal 8 (6), pp. 482-492, March, 2003
【非特許文献２】L. Lu, D. Liu, and H.-J. Zhang、“Automatic mood detection and tracking of music audio signals”、IEEE Transactions on Audio, Speech, and Language Processing, 14(1):5－18, 2006
【非特許文献３】M.F McKinney and J. Breebaart、“Features for audio and music classification”、Proc. ISMIR, 2003
【非特許文献４】G. H. Wakefield、“Mathematical representation of joint time Chroma distributions”、SPIE, 1999
【非特許文献５】Ludovic Malfait, Jens Berger, and Martin Kastner、“P.563-The ITU-T Standard for Single-Ended Speech Quality Assessment”、IEEE Transaction on Audio, Speech, and Language Processing, VOL. 14, NO. 6, November 2006
【発明の概要】
【発明が解決しようとする課題】
【００１３】
しかしながら、手動選択はユーザーにとって不便である。ユーザーは通例、あらかじめ定義されたプリセットの間で頻繁に切り換えることはせず、単にすべてのコンテンツについて一つのプリセットを使い続ける。さらに、いくつかの自動ソリューションにおいてでさえ、プリセットにおけるパラメータまたはアルゴリズム・セットアップは通例離散的で

あり（たとえば特定のコンテンツに関して特定のアルゴリズムについてオンまたはオフにするなど）、コンテンツ・ベースの連続的な仕方でパラメータを調整することはできない。

【課題を解決するための手段】

【００１４】

本願の第一の側面は、再生時にオーディオ・コンテンツに基づいて連続的な仕方でオーディオ改善装置を自動的に構成設定することである。この「自動」モードでは、ユーザーは、わざわざ異なるプリセットを選ぶことなく、簡単にコンテンツを享受することができる。他方、連続的に調整することは、遷移点における耳に聞こえるアーチファクトを避けるために、より重要である。

【００１５】

第一の側面のある実施形態によれば、オーディオ処理装置は、リアルタイムでオーディオ信号を少なくとも一つのオーディオ型に分類するオーディオ分類器と；聴衆の経験を改善するためのオーディオ改善装置と；前記少なくとも一つのオーディオ型の信頼値に基づいて連続的な仕方で前記オーディオ改善装置の少なくとも一つのパラメータを調整するための調整ユニットとを含む。

【００１６】

オーディオ改善装置は、ダイアログ向上器、サラウンド仮想化器、ボリューム平準化器および等化器のうちの任意のものであってもよい。

【００１７】

対応して、オーディオ処理方法は、リアルタイムでオーディオ信号を少なくとも一つのオーディオ型に分類し；前記少なくとも一つのオーディオ型の信頼値に基づいて連続的な仕方でオーディオ改善のための少なくとも一つのパラメータを調整することを含む。

【００１８】

第一の側面のもう一つの実施形態によれば、ボリューム平準化器コントローラが、リアルタイムでオーディオ信号のコンテンツ型を識別するためのオーディオ・コンテンツ分類器と；識別されたコンテンツ型に基づいて連続的な仕方でボリューム平準化器を調整する調整ユニットとを含む。調整ユニットは、ボリューム平準化器の動的な利得を、オーディオ信号の情報性のコンテンツ型と正に相関させ、ボリューム平準化器の動的な利得を、オーディオ信号の干渉性のコンテンツ型と負に相関させるよう構成されていてもよい。

【００１９】

上記のようなボリューム平準化器コントローラを有するオーディオ処理装置も開示される。

【００２０】

対応して、ボリューム平準化器制御方法は、リアルタイムでオーディオ信号のコンテンツ型を識別し；識別されたコンテンツ型に基づいて連続的な仕方でボリューム平準化器を調整することを含む。該調整は、ボリューム平準化器の動的な利得を、オーディオ信号の情報性のコンテンツ型と正に相関させ、ボリューム平準化器の動的な利得を、オーディオ信号の干渉性のコンテンツ型と負に相関させることによる。

【００２１】

第一の側面のさらにもう一つの実施形態によれば、等化器コントローラが、リアルタイムでオーディオ信号のオーディオ型を識別するためのオーディオ分類器と；識別されたオーディオ型の信頼値に基づいて連続的な仕方で等化器を調整する調整ユニットとを含む。

【００２２】

上記のような等化器コントローラを有するオーディオ処理装置も開示される。

【００２３】

対応して、等化器制御方法は、リアルタイムでオーディオ信号のオーディオ型を識別し；識別されたオーディオ型の信頼値に基づいて連続的な仕方で等化器を調整することを含む。

【００２４】

本願は、コンピュータ・プログラム命令が記録されたコンピュータ可読媒体であって、前記命令は、プロセッサによって実行されると、前記プロセッサが上述したオーディオ処理方法または前記ボリューム平準化器制御方法または前記等化器制御方法を実行できるようにするものをも提供する。

【００２５】

第一の側面の実施形態によれば、ダイアログ向上器、サラウンド仮想化器、ボリューム平準化器および等化器のうちの一つであってもよいオーディオ改善装置は、オーディオ信号の型および／または該型の信頼値に従って連続的に調整されてもよい。

【００２６】

本願の第二の側面は、複数のオーディオ型を識別するコンテンツ識別コンポーネントを開発することである。検出結果は、連続的な仕方でよりよいパラメータを見出すことにおいてさまざまなオーディオ改善装置の挙動を操縦／案内するために使われてもよい。

【００２７】

第二の側面のある実施形態によれば、オーディオ分類器は：それぞれオーディオ・フレームのシーケンスを含む短期的オーディオ・セグメントから短期的特徴を抽出する短期的特徴抽出器と；長期的オーディオ・セグメント内の短期的セグメントのシーケンスをそれぞれの短期的特徴を使って諸短期的オーディオ型に分類する短期的分類器と；前記長期的オーディオ・セグメント内の短期的セグメントのシーケンスに関して短期的分類器の結果の統計量を長期的特徴として計算する統計抽出器と；前記長期的特徴を使って、前記長期的オーディオ・セグメントを長期的オーディオ型に分類する長期的分類器とを含む。

【００２８】

上記のようなオーディオ分類器を有するオーディオ処理装置も開示される。

【００２９】

対応して、オーディオ分類方法は：それぞれオーディオ・フレームのシーケンスを含む短期的オーディオ・セグメントから短期的特徴を抽出し；長期的オーディオ・セグメント内の短期的セグメントのシーケンスをそれぞれの短期的特徴を使って諸短期的オーディオ型に分類し；前記長期的オーディオ・セグメント内の短期的セグメントのシーケンスに関して分類処理の結果の統計量を長期的特徴として計算し；前記長期的特徴を使って、前記長期的オーディオ・セグメントを長期的オーディオ型に分類することを含む。

【００３０】

第二の側面のもう一つの実施形態によれば、オーディオ分類器は、オーディオ信号の短期的セグメントのコンテンツ型を識別するオーディオ・コンテンツ分類器と；少なくとも部分的には前記オーディオ・コンテンツ分類器によって識別されたコンテンツ型に基づいて前記短期的セグメントのコンテキスト型を識別するオーディオ・コンテキスト分類器とを含む。

【００３１】

上記のようなオーディオ分類器を有するオーディオ処理装置も開示される。

【００３２】

対応して、オーディオ分類方法は、オーディオ信号の短期的セグメントのコンテンツ型を識別し；少なくとも部分的には識別されたコンテンツ型に基づいて前記短期的セグメントのコンテキスト型を識別することを含む。

【００３３】

本願は、コンピュータ・プログラム命令が記録されたコンピュータ可読媒体であって、前記命令は、プロセッサによって実行されると、前記プロセッサが上述したオーディオ分類方法を実行できるようにするものをも提供する。

【００３４】

第二の側面の実施形態によれば、オーディオ信号は、短期的な型またはコンテンツ型とは異なる、種々の長期的な型またはコンテキスト型に分類されてもよい。オーディオ信号の型および／または該型の信頼値は、ダイアログ向上器、サラウンド仮想化器、ボリューム平準化器または等化器のようなオーディオ改善装置を調整するためにさらに使われても

よい。
【図面の簡単な説明】
【００３５】
本願は、限定ではなく例として、付属の図面の図において例解される。図面において同様の参照符号は同様の要素を指す。
【図１】本願のある実施形態に基づくオーディオ処理装置を示す図である。
【図２】図１に示した実施形態の変形を示す図である。
【図３】図１に示した実施形態の変形を示す図である。
【図４】複数のオーディオ型を識別し、信頼値を計算する分類器の可能な構成を示す図である。
【図５】複数のオーディオ型を識別し、信頼値を計算する分類器の可能な構成を示す図である。
【図６】複数のオーディオ型を識別し、信頼値を計算する分類器の可能な構成を示す図である。
【図７】本願のオーディオ処理装置のさらなる実施形態を示す図である。
【図８】本願のオーディオ処理装置のさらなる実施形態を示す図である。
【図９】本願のオーディオ処理装置のさらなる実施形態を示す図である。
【図１０】種々のオーディオ型の間の遷移の遅延を示す図である。
【図１１】本願の実施形態に基づくオーディオ処理方法を示すフローチャートである。
【図１２】本願の実施形態に基づくオーディオ処理方法を示すフローチャートである。
【図１３】本願の実施形態に基づくオーディオ処理方法を示すフローチャートである。
【図１４】本願の実施形態に基づくオーディオ処理方法を示すフローチャートである。
【図１５】本願のある実施形態に基づくダイアログ向上器コントローラを示す図である。
【図１６】ダイアログ向上器の制御において本願に基づくオーディオ処理方法の使用を示すフローチャートである。
【図１７】ダイアログ向上器の制御において本願に基づくオーディオ処理方法の使用を示すフローチャートである。
【図１８】本願のある実施形態に基づくサラウンド仮想化器コントローラを示す図である。
【図１９】サラウンド仮想化器の制御において本願に基づくオーディオ処理方法の使用を示すフローチャートである。
【図２０】本願のある実施形態に基づくボリューム平準化器コントローラを示す図である。
【図２１】本願に基づくボリューム平準化器コントローラの効果を示す図である。
【図２２】本願のある実施形態に基づく等化器コントローラを示す図である。
【図２３】所望されるスペクトル・バランス・プリセットのいくつかの例を示す図である。
【図２４】本願のある実施形態に基づくオーディオ分類器を示す図である。
【図２５】本願のオーディオ分類器によって使用されるいくつかの特徴を示す図である。
【図２６】本願のオーディオ分類器によって使用されるいくつかの特徴を示す図である。
【図２７】本願に基づくオーディオ分類器のさらなる実施形態を示す図である。
【図２８】本願に基づくオーディオ分類器のさらなる実施形態を示す図である。
【図２９】本願に基づくオーディオ分類器のさらなる実施形態を示す図である。
【図３０】本願の実施形態に基づくオーディオ分類方法を示すフローチャートである。
【図３１】本願の実施形態に基づくオーディオ分類方法を示すフローチャートである。
【図３２】本願の実施形態に基づくオーディオ分類方法を示すフローチャートである。
【図３３】本願の実施形態に基づくオーディオ分類方法を示すフローチャートである。
【図３４】本願のもう一つの実施形態に基づくオーディオ分類器を示す図である。
【図３５】本願にさらにもう一つの実施形態に基づくオーディオ分類器を示す図である。
【図３６】本願のオーディオ分類器において使われるヒューリスティック規則を示す図で

ある。
【図３７】本願に基づくオーディオ分類器のさらなる実施形態を示す図である。
【図３８】本願に基づくオーディオ分類器のさらなる実施形態を示す図である。
【図３９】本願の実施形態に基づくオーディオ分類方法を示すフローチャートである。
【図４０】本願の実施形態に基づくオーディオ分類方法を示すフローチャートである。
【図４１】本願の実施形態を実装する例示的なシステムを示すブロック図である。
【発明を実施するための形態】
【００３６】
以下では、本願の実施形態が図面を参照しつつ記述される。明確のために、当業者に知られているが本願を理解するために必要ではないコンポーネントおよびプロセスについての表現および記述は、図面および説明において省略されることを注意しておく。
【００３７】
当業者は理解するであろうが、本願の諸側面は、システム、装置（たとえば携帯電話、ポータブル・メディア・プレーヤー、パーソナル・コンピュータ、バーサー、テレビジョン・セットトップボックスまたはデジタル・ビデオ・レコーダまたは他の任意のメディア・プレーヤー）、方法またはコンピュータ・プログラム・プロダクトとして具現されうる。よって、本願の諸側面は、ハードウェア実施形態、ソフトウェア実施形態（ファームウェア、常駐ソフトウェア、マイクロコードなどを含む）またはソフトウェアおよびハードウェア側面の両方を組み合わせる実施形態の形と取りうる。これらはみな本稿では「回路」、「モジュール」または「システム」と称されることがある。さらに、本願の諸側面は、コンピュータ可読プログラム・コードが具現された一つまたは複数のコンピュータ可読媒体において具現されたコンピュータ・プログラム・プロダクトの形を取ることがある。
【００３８】
一つまたは複数のコンピュータ可読媒体のいかなる組み合わせが利用されてもよい。コンピュータ可読媒体は、コンピュータ可読信号媒体またはコンピュータ可読記憶媒体でありうる。コンピュータ可読記憶媒体は、たとえば、電子的、磁気的、光学式、電磁式、赤外線または半導体のシステム、装置またはデバイスまたは以上のものの任意の好適な組み合わせでありうるがそれに限られない。コンピュータ可読記憶媒体のさらなる個別的な例（網羅的なリストではない）は、次のものを含む：一つまたは複数のワイヤをもつ電気的接続、ポータブル・コンピュータ・ディスケット、ハードディスク、ランダム・アクセス・メモリ（RAM）、読み出し専用メモリ（ROM）、消去可能なプログラム可能型読み出し専用メモリ（EPROMまたはフラッシュメモリ）、光ファイバー、ポータブル・コンパクトディスク読み出し専用メモリ（CD-ROM）、光記憶デバイス、磁気記憶デバイスまたは以上のものの任意の好適な組み合わせ。本稿のコンテキストでは、コンピュータ可読記憶媒体は、命令実行システム、装置またはデバイスによって使うためまたはそれらとの関連で使うためのプログラムを含むまたは記憶することができるいかなる有体の媒体であってもよい。
【００３９】
コンピュータ可読信号媒体は、たとえばベースバンドにおいてまたは搬送波の一部としてコンピュータ可読プログラム・コードが具現されている伝搬するデータ信号を含みうる。そのような伝搬する信号は、電磁的または光学的な信号またはそれらの任意の好適な組み合わせを含むがそれに限られない多様な形の任意のものを取りうる。
【００４０】
コンピュータ可読信号媒体は、命令実行システム、装置またはデバイスによって使うためまたはそれらとの関連で使うためのプログラムを通信する、伝搬させるまたは搬送することができる、コンピュータ可読記憶媒体ではないいかなるコンピュータ可読媒体であってもよい。
【００４１】
コンピュータ可読媒体上に具現されるプログラム・コードは、無線、有線、光ファイバーケーブル、RFなどまたは以上のものの任意の好適な組み合わせを含むがそれに限られな

いいかなる適切な媒体を使って伝送されてもよい。

【００４２】

本願の諸側面の動作を実行するためのコンピュータ・プログラム・コードは、ジャバ、スモールトーク、C++などといったオブジェクト指向プログラミング言語および「C」プログラミング言語といった従来型の手続き型プログラミング言語または同様のプログラミング言語を含む、一つまたは複数のプログラミング言語の任意の組み合わせで書かれてもよい。プログラム・コードは、完全にユーザーのコンピュータ上でスタンドアローンのソフトウェア・パッケージとして、部分的にユーザーのコンピュータ上で部分的にはリモート・コンピュータ上で、あるいは完全にリモート・コンピュータまたはサーバー上で実行されうる。この最後のシナリオでは、リモート・コンピュータはユーザーのコンピュータに、ローカル・エリア・ネットワーク（LAN）または広域ネットワーク（WAN）を含む任意の型のネットワークを通じて接続されてもよく、あるいは（たとえばインターネット・サービス・プロバイダーを使ってインターネットを通じて）外部コンピュータに接続がされてもよい。

【００４３】

本発明の諸側面は、本発明の実施形態に基づく方法、装置（システム）およびコンピュータ・プログラム・プロダクトのフローチャート図および／またはブロック図を参照して記述される。フローチャート図および／またはブロック図の各ブロックならびにフローチャート図および／またはブロック図のブロックの組み合わせは、コンピュータ・プログラム命令によって実装されることができることは理解されるであろう。これらのコンピュータ・プログラム命令は、汎用コンピュータ、特殊目的コンピュータまたは他のプログラム可能なデータ処理装置のプロセッサに与えられて、該コンピュータまたは他のプログラム可能なデータ処理装置のプロセッサによって実行される該命令が前記フローチャートおよび／またはブロック図の単数または複数のブロックにおいて特定されている機能／工程を実装する手段を作り出すよう、機械を生成してもよい。

【００４４】

これらのコンピュータ・プログラム命令は、コンピュータ、他のプログラム可能なデータ処理装置または他のデバイスが特定の仕方で機能するよう指令することができるコンピュータ可読媒体に記憶され、それにより、該コンピュータ可読媒体に記憶される命令は、前記フローチャートおよび／またはブロック図の単数または複数のブロックにおいて特定されている機能／工程を実装する命令を含む製造物を作り出してもよい。

【００４５】

コンピュータ・プログラム命令はコンピュータ、他のプログラム可能なデータ処理装置または他のデバイスにロードされて、該コンピュータ、他のプログラム可能な装置または他のデバイス上で一連の動作処理を実行させて、前記コンピュータまたは他のプログラム可能な装置上で実行される前記命令が前記フローチャートおよび／またはブロック図の単数または複数のブロックにおいて特定されている機能／工程を実装するためのプロセスを提供するようなコンピュータ実装されたプロセスを作り出してもよい。

【００４６】

下記では、本願の実施形態が詳細に記述される。明確のため、記述は次の構成に編成される：

第一部：オーディオ処理装置および方法

【００４７】

〈第一部：オーディオ処理装置および方法〉

図１は、再生時にオーディオ・コンテンツに基づく改善されたパラメータでの少なくとも一つのオーディオ改善装置の自動的な構成設定をサポートするコンテンツ適応的なオーディオ処理装置１００の概括的なフレームワークを示している。これは三つの主要なコンポーネントを有する：オーディオ分類器２００、調整ユニット３００、オーディオ改善装置４００である。

【００４８】

オーディオ分類器２００は、リアルタイムでオーディオ信号を少なくとも一つのオーディオ型に分類するものである。これは再生時にコンテンツのオーディオ型を自動的に識別する。オーディオ・コンテンツを識別するためには、信号処理、機械学習およびパターン認識を通じてなど、いかなるオーディオ分類技術が適用されることもできる。あらかじめ定義された目標オーディオ型の集合に関するオーディオ・コンテンツの確率を表わす信頼値がほぼ同時に推定される。

【００４９】
オーディオ改善装置４００は、オーディオ信号に対して処理を実行することによって聴衆の経験を改善するものであり、のちに詳細に論じる。
【００５０】
調整ユニット３００は、前記少なくとも一つのオーディオ型の信頼値に基づいて連続的な仕方で前記オーディオ改善装置の少なくとも一つのパラメータを調整するものである。これは、オーディオ改善装置４００の挙動を操縦するよう設計される。これは、オーディオ分類器２００から得られた結果に基づいて対応するオーディオ改善装置の最も好適なパラメータを推定する。
【００５１】
さまざまなオーディオ改善装置がこの装置において適用できる。図２は、ダイアログ向上器（DE: Dialog Enhancer）４０２、サラウンド仮想化器（SV: Surround Virtualizer）４０４、ボリューム平準化器（VL: Volume Leveler）４０６および等化器（EQ: Equalizer）４０８を含む四つのオーディオ改善装置を含む例示的なシステムを示している。各オーディオ改善装置は、オーディオ分類器２００において得られる結果（オーディオ型および／または信頼値）に基づいて連続的な仕方で自動的に調整されることができる。
【００５２】
むろん、オーディオ処理装置は、必ずしもすべての種類のオーディオ改善装置を含まなくてもよく、そのうち一つまたは複数を含むだけでもよい。他方、オーディオ改善装置は本開示において与えられている装置に限定されず、さらなる種類のオーディオ改善装置を含んでいてもよく、それらも本願の範囲内である。さらに、ダイアログ向上器（DE）４０２、サラウンド仮想化器（SV）４０４、ボリューム平準化器（VL）４０６および等化器（EQ）４０８を含む本開示において論じられるオーディオ改善装置の名称は限定をなすものではなく、そのそれぞれは同じまたは同様の機能を実現する他の任意の装置をカバーすると解釈される。
【００５３】
〈１．１節　オーディオ型〉
さまざまな種類のオーディオ改善装置を適正に制御するために、本願はさらに、オーディオ型の新たな構成を提供する。ただし、従来技術におけるオーディオ型も本願で適用可能である。
【００５４】
具体的には、オーディオ信号中の基本成分を表わす低レベルのオーディオ要素と、現実のユーザー娯楽アプリケーションにおけるたいていの一般的なオーディオ・コンテンツを表わす高レベルのオーディオ・ジャンルとを含め、異なる意味的レベルからのオーディオ型がモデル化される。前者は「コンテンツ型」と称されてもよい。基本的オーディオ・コンテンツ型は、発話、音楽（歌を含む）、背景音（または効果音）およびノイズを含んでいてもよい。
【００５５】
発話および音楽の意味は明らかである。本願におけるノイズは、意味的なノイズではなく、物理的なノイズを意味する。本願における物理的なノイズは、たとえばエアコンからのノイズや、信号伝送経路に起因するピンク・ノイズのような技術的理由により生じるノイズを含みうる。対照的に、本願における「背景音」は、聴取者の注意のコア・ターゲットの周辺で生起する聴覚イベントであってもよい効果音である。たとえば、電話の通話におけるオーディオ信号では、話者の声のほかに、通話に関係ない何らかの他の人物の声、キーボードの音、足音などのような、意図されない何らかの他の音があることがある。これらの望まれない音は、ノイズではなく、「背景音」と称される。つまり、「背景音」は、ターゲット（または聴取者の注意のコア・ターゲット）ではない、あるいはさらに望まれないものであるが、それでも何らかの内容的な意味をもつ音と定義してもよい。一方、「ノイズ」は、ターゲットオンおよび背景音を除く望まれない音と定義されてもよい。

【００５６】
時に、背景音は本当に「望まれない」のではなく、意図的に生成され、何らかの有用な情報を担う。たとえば、映画、テレビ番組またはラジオ放送番組における背景音がそうである。よって、時に、「効果音」と称されることがある。本開示では以下では、簡潔のため「背景音」のみが使用され、さらに「背景」と短縮されることもある。
【００５７】
さらに、音楽はさらに、優勢な源のない音楽と優勢な源のある音楽に分類されてもよい。音楽片において他の源よりずっと強い源（声または楽器）がある場合には、「優勢な源のある音楽」と称される。そうでない場合には、「優勢な源のない音楽」と称される。たとえば、歌声およびさまざまな楽器を伴う多声音楽では、和声的にバランスが取れているまたはいくつかの最も顕著な源のエネルギーが互いに匹敵する場合には、優勢な源のない音楽と考えられる。対照的に、ある源（たとえば声）がずっと大きく、一方他の源がずっと静かである場合には、優勢な源を含んでいると考えられる。もう一つの例として、飛び抜けた、あるいは目立つ楽器トーンは「優勢な源をもつ音楽」である。
【００５８】
音楽はさらに、種々の標準に基づく種々の型に分類されうる。音楽は、これに限られないがロック、ジャズ、ラップおよびフォークのような音楽のジャンルに基づいて分類されることができる。音楽は、声楽および器楽のように、楽器に基づいて分類されることもできる。器楽は、ピアの音楽およびギター音楽など、種々の楽器を用いて演奏されるさまざまな音楽を含みうる。他の例示的な標準は、リズム、テンポ、音楽の音色および／または属性の類似性に基づいて音楽がグループ化されることのできる他の任意の音楽的な属性を含む。たとえば、音色に基づいて、声楽はテノール、バリトン、バス、ソプラノ、メゾソプラノおよびアルトに分類されうる。
【００５９】
オーディオ信号のコンテンツ型は、複数のフレームから構成されるような短期的オーディオ・セグメントに関して分類されてもよい。一般に、オーディオ・フレームは20msのような複数ミリ秒の長さであり、オーディオ分類器によって分類されるべき短期的オーディオ・セグメントの長さは、数百ミリ秒から数秒、たとえば1秒の長さをもちうる。
【００６０】
コンテンツ適応的な仕方でオーディオ改善装置を制御するために、オーディオ信号はリアルタイムで分類されてもよい。上記のコンテンツ型については、現在の短期的オーディオ・セグメントのコンテンツ型は現在のオーディオ信号のコンテンツ型を表わす。短期的オーディオ・セグメントの長さはそれほど長くないので、オーディオ信号は、順次の、重なり合わない短期的オーディオ・セグメントとして分割されてもよい。しかしながら、短期的オーディオ・セグメントは、オーディオ信号の時間軸に沿って連続的／半連続的にサンプリングされてもよい。すなわち、短期的オーディオ・セグメントは、オーディオ信号の時間軸に沿って一つまたは複数のフレームのステップ・サイズで動く所定の長さ（短期的オーディオ・セグメントの意図される長さ）をもつ窓を用いてサンプリングされてもよい。
【００６１】
高レベルのオーディオ・ジャンルは、オーディオ信号の長期的な型を示すので「コンテキスト型」と称されることもあり、その時のサウンド・イベントの環境またはコンテキストと見なされてもよく、それは上記のようなコンテンツ型に分類されてもよい。本願によれば、コンテキスト型は、映画的メディア、音楽（歌を含む）、ゲームおよびVoIP（インターネット・プロトコル上での音声）のようなたいていの一般的なオーディオ・アプリケーションを含みうる。
【００６２】
音楽、ゲームおよびVoIPの意味は自明である。映画的メディアは映画、テレビ番組、ラジオ放送番組または上記のものと同様の他の任意のオーディオ・メディアを含んでいてもよい。映画的メディアの主要な特徴は、可能な発話、音楽およびさまざまな種類の背景音

（効果音）の混合である。

【００６３】

コンテンツ型およびコンテキスト型はいずれも音楽（歌を含む）を含むことを注意してもよいであろう。以下、本願では、それらを区別するためにそれぞれ「短期的音楽」および「長期的音楽」という言い方を使う。

【００６４】

本願のいくつかの実施形態については、他のいくつかのコンテキスト型構成も提案される。

【００６５】

たとえば、オーディオ信号は高品質オーディオ（映画的メディアおよび音楽CDなど）または低品質オーディオ（VoIP、低ビットレート・オンライン・ストリーミング・オーディオおよびユーザー生成コンテンツなど）と分類されてもよい。これらはまとめて「オーディオ品質型」と称されてもよい。

【００６６】

もう一つの例として、オーディオ信号はVoIPまたは非VoIPとして分類されてもよい。これらは上述した４コンテキスト型構成（VoIP、映画的メディア、（長期的）音楽およびゲーム）の変換と見なされてもよい。VoIPまたは非VoIPのコンテキストとの関連で、オーディオ信号は、VoIP発話、非VoIP発話、VoIPノイズおよび非VoIPノイズのようなVoIPに関係したオーディオ・コンテンツ型として分類されてもよい。VoIPオーディオ・コンテンツ型の構成は、VoIPおよび非VoIPコンテキストを区別するために特に有用である。通例、VoIPコンテキストは、ボリューム平準化器（オーディオ改善装置の一つの種類）の最も困難な応用シナリオだからである。

【００６７】

一般に、オーディオ信号のコンテキスト型は、短期的オーディオ・セグメントより長い長期的オーディオ・セグメントに関して分類されてもよい。長期的オーディオ・セグメントは、短期的オーディオ・セグメントにおけるフレーム数より多くの数の複数のフレームから構成される。長期的オーディオ・セグメントは、複数の短期的オーディオ・セグメントから構成されてもよい。一般に、長期的オーディオ・セグメントは、数秒から数十秒、たとえば10秒などの秒のオーダーの長さをもちうる。

【００６８】

同様に、適応的な仕方でオーディオ改善装置を制御するために、オーディオ信号は、リアルタイムでコンテキスト型に分類されてもよい。同様に、現在の長期的オーディオ・セグメントのコンテキスト型は、現在のオーディオ信号のコンテキスト型を表わす。長期的オーディオ・セグメントの長さは比較的長いので、オーディオ信号は、そのコンテキスト型の急激な変化を、よってオーディオ改善装置（単数または複数）の作動パラメータの急激な変化を避けるために、オーディオ信号の時間軸に沿って連続的／半連続的にサンプリングされてもよい。すなわち、長期的オーディオ・セグメントは、オーディオ信号の時間軸に沿って一つまたは複数のフレームまたは一つまたは複数の短期的セグメントのステップ・サイズで動く所定の長さ（長期的オーディオ・セグメントの意図される長さ）をもつ窓を用いてサンプリングされてもよい。

【００６９】

上記では、コンテンツ型およびコンテキスト型の両方が記述された。本願の実施形態では、調整ユニット３００は、さまざまなコンテンツ型の少なくとも一つおよび／またはさまざまなコンテキスト型の少なくとも一つに基づいて、オーディオ改善装置（単数または複数）の少なくとも一つのパラメータを調整してもよい。したがって、図３に示されるように、図１に示した実施形態のある変形では、オーディオ分類器２００は、オーディオ・コンテンツ分類器２０２またはオーディオ・コンテキスト分類器２０４またはその両方を有していてもよい。

【００７０】

上記では、（たとえばコンテキスト型について）種々の標準に基づく種々のオーディオ

型および（たとえばコンテンツ型について）種々の階層レベルでの種々のオーディオ型が言及された。しかしながら、これらの標準および階層レベルは単にここでの記述の便宜のためであって、全く限定するものではない。つまり、本願では、上述したオーディオ型の任意の二つ以上は、同時にオーディオ分類器２００によって識別され、同時に調整ユニット３００によって考慮されることができる。これについては後述する。つまり、種々の階層レベルにおけるすべてのオーディオ型は並列、あるいは同じレベルであってもよい。

【００７１】

〈１．２節　オーディオ型の信頼値および分類器の構成〉

オーディオ分類器２００は、硬判定結果を出力してもよく、あるいは調整ユニット３００はオーディオ分類器２００の結果を硬判定結果と見なしてもよい。硬判定についてでも、複数のオーディオ型がオーディオ・セグメントに割り当てられることができる。たとえば、オーディオ・セグメントは、発話および短期的音楽の混合信号でありうるので、「発話」および「短期的音楽」の両方によってラベル付けされることができる。得られたラベルは、オーディオ改善装置（単数または複数）４００を操縦するために直接使われることができる。簡単な例は、発話が存在するときにダイアログ向上器４０２を有効にし、発話が存在しないときにオフにするというものである。しかしながら、この硬判定方法は、注意深い平滑化方式（後述）なしの場合には、あるオーディオ型から別のオーディオ型への遷移点においていくらかの不自然さを導入することがある。

【００７２】

より柔軟性をもち、連続的な仕方でオーディオ改善装置のパラメータを調整するために、各ターゲット・オーディオ型の信頼値が推定されることができる（軟判定）。信頼値は、識別されるべきオーディオ・コンテンツとターゲット・オーディオ型の間の一致レベルを0から1の値で表わす。

【００７３】

先述したように、多くの分類技法は直接、信頼値を出力してもよい。信頼値は、分類器の一部と見なされてもよいさまざまな方法から計算されることもできる。たとえば、オーディオ・モデルがガウシアン混合モデル（GMM: Gaussian Mixture Models）のようないくつかの確率的モデル化技術によってトレーニングされる場合、信頼値を表わすために、次のように事後確率が使われることができる。

【００７４】

【数１】

$$p(c_i \mid x) = \frac{p(x \mid c_i)}{\sum_{i=1}^{N} p(x \mid c_i)} \qquad (1)$$

ここで、xはオーディオ・セグメントの一片であり、c_iはターゲット・オーディオ型であり、Nはターゲット・オーディオ型の数であり、p(chi)はオーディオ・セグメントxがオーディオ型c_iである確からしさであり、$p(c_o ax)$は対応する事後確率である。

【００７５】

他方、オーディオ・モデルがサポートベクターマシン（SVM: Support Vector Machine）およびアダブースト（adaBoost）のようないくつかの弁別的方法からトレーニングされる場合には、モデル比較からはスコア（実数値）だけが得られる。これらの場合、得られたスコア（理論的には－∞から∞）を期待される信頼度（0から1）にマッピングするために、通例、次のシグモイド関数が使われる。

【００７６】

【数２】

$$conf = \frac{1}{1+e^{Ay+B}} \qquad (2)$$

ここで、yはSVMまたはアダブーストからの出力スコアであり、AおよびBは何らかのよく知られた技術を使ってトレーニング・データ・セットから推定される必要のある二つのパラメータである。

【００７７】

本願のいくつかの実施形態については、調整ユニット３００は、三つ以上のコンテンツ型および／または三つ以上のコンテキスト型を使ってもよい。その場合、オーディオ・コンテンツ分類器２０２は三つ以上のコンテンツ型を識別する必要がある、および／またはオーディオ・コンテキスト分類器２０４は三つ以上のコンテキスト型を識別する必要がある。そのような状況では、オーディオ・コンテンツ分類器２０２またはオーディオ・コンテキスト分類器２０４は、ある構成で編成された分類器の群であってもよい。

【００７８】

たとえば、調整ユニット３００が四種類のコンテキスト型、映画的メディア、長期的音楽、ゲームおよびVoIPの全部を必要とする場合には、オーディオ・コンテキスト分類器２０４は以下の種々の構成をもちうる。

【００７９】

第一に、オーディオ・コンテキスト分類器２０４は、図４に示されるように編成された６個の一対一二項分類器（各分類器は一つのターゲット・オーディオ型を別のあるターゲット・オーディオ型から弁別する）と、図５に示されるように編成された３個の一対他の二項分類器（各分類器はターゲット・オーディオ型を他のオーディオ型から弁別する）と、図６に示されるように編成された４個の一対他の分類器を有していてもよい。判定有向非環状グラフ（DDAG: Decision Directed Acyclic Graph）構成のような他の構成もある。図４～図６および対応する以下の記述において、簡潔のため、「映画的メディア」の代わりに「映画」が使われていることを注意しておく。

【００８０】

各二項分類器は、その出力について信頼スコアH(x)を与える（xはオーディオ・セグメントを表わす）。各二項分類器の出力が得られたのち、それらを、識別された諸コンテキスト型の最終的な諸信頼値にマッピングする必要がある。

【００８１】

一般に、オーディオ信号はM個のコンテキスト型に分類されるとする（Mは正の整数）。通常の一対一構成はM(M－1)/2個の分類器を構築する。ここで、各分類器は二つのクラスからのデータでトレーニングされる。次いで、それぞれの一対一分類器はその好ましいクラスについて一票を投じ、最終結果は、M(M－1)/2個の分類器の分類のうちの最多票をもつクラスである。通常の一対一構成と比べ、図４における階層的な構成もM(M－1)/2個の分類器を構築することを必要とする。しかしながら、セグメントxは各階層レベルにおいて対応するクラスにある／ないと判定され、全体的なレベル・カウントはM－1なので、試験反復工程はM－1に短縮されることができる。さまざまなコンテキスト型についての最終的な信頼値は、たとえば二項分類信頼値$H_k(x)$から計算されてもよい（k＝1,2,…,6は種々のコンテキスト型を表わす）。

【００８２】

【数３】

$$C_{MOVIE} = (1 - H_1(x)) \cdot (1 - H_3(x)) \cdot (1 - H_6(x))$$

$$C_{VOIP} = H_1(x) \cdot H_2(x) \cdot H_4(x)$$

$$C_{MUSIC} = H_1(x) \cdot (1 - H_2(x)) \cdot (1 - H_5(x)) + H_3(x) \cdot (1 - H_1(x)) \cdot (1 - H_5(x)) + H_6(x) \cdot (1 - H_1(x)) \cdot (1 - H_3(x))$$

$$C_{GAME} = H_1(x) \cdot H_2(x) \cdot (1 - H_4(x)) + H_1(x) \cdot H_5(x) \cdot (1 - H_2(x)) + H_3(x) \cdot H_5(x) \cdot (1 - H_1(x))$$

図５に示した構成では、二項分類結果$H_k(x)$から最終的な信頼値へのマッピング関数は次の例のように定義できる。

【００８３】

【数４】

$$C_{MOVIE} = H_1(x)$$

$$C_{MUSIC} = H_2(x) \cdot (1 - H_1(x))$$

$$C_{VOIP} = H_3(x) \cdot (1 - H_2(x)) \cdot (1 - H_1(x))$$

$$C_{GAME} = (1 - H_3(x)) \cdot (1 - H_2(x)) \cdot (1 - H_1(x))$$

図６に示した構成では、最終的な信頼値は、対応する二項分類結果$H_k(x)$に等しくてもよく、あるいはすべてのクラスについての信頼値の和が1であることが要求されるならば、最終的な信頼値は単に推定された$H_k(x)$に基づいて規格化されることができる。

【００８４】

【数５】

$$C_{MOVIE} = H_1(x)/(H_1(x) + H_2(x) + H_3(x) + H_4(x))$$

$$C_{MUSIC} = H_2(x)/(H_1(x) + H_2(x) + H_3(x) + H_4(x))$$

$$C_{VOIP} = H_3(x)/(H_1(x) + H_2(x) + H_3(x) + H_4(x))$$

$$C_{GAME} = H_4(x)/(H_1(x) + H_2(x) + H_3(x) + H_4(x))$$

最大の信頼値をもつ一つまたは複数が最終的な識別されたクラスとして決定されることができる。

【００８５】

図４～図６に示される構成では、種々の二項分類器のシーケンスは必ずしも図示したとおりではなく、他のシーケンスであってもよいことを注意しておくべきである。そうしたシーケンスは、さまざまな応用の種々の要求に応じて、手動の割り当てまたは自動学習によって選択されてもよい。

【００８６】

上記の記述は、オーディオ・コンテキスト分類器２０４に向けられている。オーディオ・コンテンツ分類器２０２については、状況は同様である。

【００８７】

あるいはまた、オーディオ・コンテンツ分類器２０２またはオーディオ・コンテキスト分類器２０４は、同時にすべてのコンテンツ型／コンテキスト型を識別する一つの単独の

分類器として実装され、同時に対応する諸信頼値を与えてもよい。これを行なうために多くの既存の技法がある。

【００８８】

信頼値を使って、オーディオ分類器２００の出力はベクトルとして表現できる。ベクトルの各次元が各ターゲット・オーディオ型の信頼値を表わす。たとえば、ターゲット・オーディオ型が順次（発話、短期的音楽、ノイズ、背景）であれば、例示的な出力結果は(0.9,0.5,0.0,0.0)であることができる。これは、そのオーディオ・コンテンツが発話であることは90%確実であり、そのオーディオが音楽であることは50%確実であることを示す。出力ベクトル中のすべての次元の和が1である必要はないことを注意しておく（たとえば、図６からの結果は必ずしも規格化されない）。つまり、オーディオ信号は発話および短期的音楽の混合信号であってもよい。

【００８９】

のちに、第六部および第七部において、オーディオ・コンテキスト分類およびオーディオ・コンテンツ分類の新規な実装を詳細に論じる。

【００９０】

〈１．３節　オーディオ型の信頼値の平滑化〉

任意的に、各オーディオ・セグメントがあらかじめ定義されたオーディオ型に分類された後、追加的なステップは、ある型から別の型への急激なジャンプを避け、オーディオ改善装置におけるパラメータのよりなめらかな推定をするために、時間軸に沿って分類結果を平滑化するというものである。たとえば、長い抜粋が、VoIPとして分類された一つだけのセグメントを除いて映画的メディアと分類されると、急激なVoIP判定は、平滑化によって映画的メディアに修正されることができる。

【００９１】

したがって、図７に示した実施形態のある変形では、各オーディオ型について、現在の時間でのオーディオ信号の信頼値を平滑化するために、型平滑化ユニット７１２がさらに設けられる。

【００９２】

一般的な平滑化方法は、現在の実際の信頼値と最後の時点の平滑化された信頼値の重み付けされた和を計算するなど、重み付けされた平均に基づく。

【００９３】

$$\mathrm{smoothConf}(t) = \beta \cdot \mathrm{smoothConf}(t-1) + (1-\beta) \cdot \mathrm{conf}(t) \qquad (3)$$

ここで、tは現在の時間（現在のオーディオ・セグメント）、t－1は最後の時間（最後のオーディオ・セグメント）、βは重み、confおよびsmoothConfはそれぞれ平滑化の前および後の信頼値（confidence　value）である。

【００９４】

信頼値の観点からは、分類器の硬決定からの結果は、値が0か1のいずれかである信頼値を用いて表わされることもできる。すなわち、ターゲット・オーディオ型が選ばれ、オーディオ・セグメントに割り当てられる場合、対応する信頼値は1であり、そうでなければ信頼値は0である。したがって、たとえオーディオ分類器２００が信頼値を与えず、オーディオ型に関する硬決定を与えるだけであっても、調整ユニット３００の連続的な調整は、型平滑化ユニット７１２の平滑化動作を通じて、相変わらず可能である。

【００９５】

平滑化アルゴリズムは、異なる場合について異なる平滑化重みを使うことによって「非対称」であることができる。たとえば、重み付けされた和を計算するための重みは、オーディオ信号のオーディオ型の信頼値に基づいて適応的に変えられてもよい。現在のセグメントの信頼値がより大きければ、その重みもより大きくなる。

【００９６】

別の観点からは、重み付けされた和を計算するための重みは、特に、オーディオ改善装置（単数または複数）が、一つの単独のコンテンツ型の有無に基づくのではなく、オーディオ分類器２００にいおって識別される複数のコンテンツ型に基づいて調整されるときは

、あるオーディオ型から別のオーディオ型への異なる遷移対に基づいて適応的に変えられてもよい。たとえば、あるコンテキストにおいてより頻繁に現われるオーディオ型からそのコンテキストにおいてそれほど頻繁に現われない別のオーディオ型への遷移については、後者の信頼値は、あまり速く増大しないよう、平滑化されてもよい。たまたまの突発であることもありうるからである。

【００９７】

もう一つの要因は、変化レートを含む変化（増大または減少）トレンドである。あるオーディオ型が存在するようになるときに（すなわちその信頼値が増すときに）我々が遅延についてより気にするとすると、次のようにして平滑化アルゴリズムを設計できる。

【００９８】

【数６】

$$smoothConf(t)=\begin{cases} conf(t) & conf(t)\geq smoothConf(t-1) \\ \beta\cdot smoothConf(t-1)+(1-\beta)\cdot conf(t) & \text{それ以外の場合} \end{cases} \quad (4)$$

上記の公式は、平滑化された信頼値が、信頼値が増加するときには迅速に現在状態に応答し、信頼値が減少するときにはゆっくりと平滑化されていくことを許容する。平滑化関数の諸変形が同様にして簡単に設計できる。たとえば、公式(4)は、conf(t)≧smoothConf(t－1)であるときにconf(t)の重みがより大きくなるよう修正されてもよい。実際、公式(4)では、β＝0であり、conf(t)の重みは最大、すなわち1であると見なすことができる。

【００９９】

異なる観点からは、あるオーディオ型の変化トレンドを考えることは、オーディオ型の異なる遷移対を考えることの単なる個別的な例である。たとえば、型Aの信頼値を増すことは、非AからAへの遷移と見なされてもよく、型Aの信頼値を減らすことは、Aから非Aへの遷移と見なされてもよい。

【０１００】

〈１．４節　パラメータ調整〉

調整ユニット３００は、オーディオ分類器２００からの得られた結果に基づいてオーディオ改善装置（単数または複数）４００についての適正なパラメータを推定または調整するよう設計される。コンテンツ型またはコンテキスト型のいずれかをまたは合同判定のために両方を使って、異なるオーディオ改善装置について異なる調整アルゴリズムが設計されてもよい。たとえば、映画的メディアおよび長期的音楽のようなコンテキスト型情報では、上述したようなプリセットが自動的に選択され、対応するコンテンツに適用されることができる。利用可能なコンテンツ型情報を用いて、各オーディオ改善装置のパラメータは、後述する部において示されるように、より細かい仕方で調整されることができる。コンテンツ型情報およびコンテキスト情報はさらに、長期的および短期的情報をバランスさせるために、調整ユニット３００において合同で使われることができる。特定のオーディオ改善装置についての特定の調整アルゴリズムは、別個の調整ユニットと見なされてもよい。あるいは、異なる調整アルゴリズムがまとめて連合した調整ユニットと見なされてもよい。

【０１０１】

すなわち、調整ユニット３００は、少なくとも一つのコンテンツ型の信頼値および／または少なくとも一つのコンテキスト型の信頼値に基づいて、オーディオ改善装置の前記少なくとも一つのパラメータを調整するよう構成されていてもよい。特定のオーディオ改善装置について、オーディオ型のいくつかは情報性であり、オーディオ型のいくつかは干渉性である。よって、特定のオーディオ改善装置のパラメータは、情報性のオーディオ型（単数または複数）または干渉性のオーディオ型（単数または複数）の信頼値（単数または複数）と正または負に相関しうる。ここで「正に相関する」とは、オーディオ型の信頼値の増大または減少とともに、パラメータが線形または非線型に増大または減少することを

意味する。「負に相関する」とは、オーディオ型の信頼値の減少または増大とともに、パラメータが線形または非線型にそれぞれ増大または減少することを意味する。

【０１０２】

ここで、信頼値の減少および増大は、正または負の相関によって、調整されるべきパラメータに直接「伝達」される。数学では、そのような相関または「伝達」は、正比例または反比例、プラスまたはマイナス（加算または減算）演算、乗算または除算または非線型関数として具現されうる。これらすべての形の相関は「伝達関数」と称されてもよい。信頼値の増大または減少を決定するために、現在の信頼値またはその数学的変換を最後の信頼値もしくは複数の履歴信頼値またはその数学的変換と比較することもできる。本願のコンテキストにおいて、用語「比較」は、減算を通じた比較または除算を通じた比較を意味する。差が0より大きいか否かまたは比が1より大きいか否かを判定することによって増大または減少を判定できる。

【０１０３】

個別的な実装では、適正なアルゴリズム（伝達関数など）を通じてパラメータを信頼値またはその比もしくは差と直接関係させることができ、「外部観察者」が特定の信頼値および／または特定のパラメータが増大したまたは減少したことを明示的に知ることは必要でない。いくつかの個別的な例が、個別的なオーディオ改善装置についての後述する第二～五部において与えられる。

【０１０４】

前節で述べたように、同じオーディオ・セグメントに関し、分類器２００は、それぞれの信頼値をもつ複数のオーディオ型を識別することがある。オーディオ・セグメントは音楽と発話と背景音など、同時に複数の成分を含むことがあるので、それらの信頼値は必ずしも総計1にならないことがある。そのような状況では、オーディオ改善装置のパラメータは、異なるオーディオ型の間でバランスを取る必要がある。たとえば、調整ユニット３００は、前記少なくとも一つのオーディオ型の重要性に基づいて前記少なくとも一つのオーディオ型の信頼値を重み付けすることを通じて、複数のオーディオ型の少なくともいくつかを考慮するよう構成されていてもよい。特定のオーディオ型がより重要であるほど、パラメータはそれによってより大きく影響される。

【０１０５】

重みは、オーディオ型の情報性および干渉性の効果を反映することもできる。たとえば、干渉性のオーディオ型については、マイナスの重みが与えられてもよい。いくつかの個別的な例が、個別的なオーディオ改善装置についての後述する第二～五部において与えられる。

【０１０６】

本願のコンテキストにおいて、「重み」は、多項式における係数よりも広い意味をもつことを注意しておく。多項式における係数のほか、「重み」は指数または冪の形を取ることもできる。多項式における係数であるとき、重み付け係数は規格化されてもされなくてもよい。一言で言うと、重みは単に重み付けされた対象が調整されるべきパラメータに対してどのくらいの影響をもつかを表わすものである。

【０１０７】

いくつかの他の実施形態では、同じオーディオ・セグメントに含まれる複数のオーディオ型について、その信頼値が、規格化されることを通じて重みに変換されてもよい。次いで、最終的なパラメータが、各オーディオ型についてあらかじめ定義され、信頼値に基づいて重みによって重み付けされたパラメータ・プリセット値の和を計算することを通じて決定されてもよい。すなわち、調整ユニット３００は、信頼値に基づいて複数のオーディオ型の効果を重み付けすることを通じて複数のオーディオ型を考慮するよう構成されていてもよい。

【０１０８】

重み付けの個別的な例として、調整ユニットは、信頼値に基づいて少なくとも一つの優勢なオーディオ型を考慮するよう構成される。あまりに低い（閾値より低い）信頼値をも

つオーディオ型については、考慮されなくてもよい。これは、信頼値が閾値より小さい他のオーディオ型の重みが0に設定されることと等価である。いくつかの個別的な例が、個別的なオーディオ改善装置についての後述する第二～五部において与えられる。

【０１０９】

コンテンツ型およびコンテキスト型は一緒に考慮されることができる。ある実施形態では、それらは同じレベルにあると見なされることができ、それらの信頼値はそれぞれの重みをもちうる。もう一つの実施形態では、まさに名称が示すように、「コンテキスト型」は、「コンテキスト型」が位置しているコンテキストまたは環境であり、よって調整ユニット２００は、異なるコンテキスト型のオーディオ信号におけるコンテンツ型が、オーディオ信号のコンテキスト型に依存して異なる重みを割り当てられるよう構成されてもよい。一般に、いかなるオーディオ型も、別のオーディオ型のコンテキストを構成することができ、調整ユニット２００はあるオーディオ型の重みを、別のオーディオ型の信頼値を用いて修正するよう構成されてもよい。いくつかの個別的な例が、個別的なオーディオ改善装置についての後述する第二～五部において与えられる。

【０１１０】

本願のコンテキストでは、「パラメータ」は、その文字通りの意味より広い意味をもつ。単一の値をもつパラメータのほかに、パラメータは、種々のパラメータの集合、種々のパラメータからなるベクトルまたはプロファイルを含め、上述したようなプリセットを意味することもある。特に、後述する第二～五部においては、次のパラメータが論じられるが、本願はそれに限定されるものではない：ダイアログ向上のレベル、ダイアログ向上されるべき周波数帯域を決定するための閾値、背景レベル、サラウンド・ブースト量、サラウンド仮想化器についての開始周波数、ボリューム平準化器の動的利得または動的利得の範囲〔ダイナミック・ゲインのレンジ〕、オーディオ信号が新しい知覚可能なオーディオ・イベントである度合いを示すパラメータ、等化レベル、等化プロファイルおよびスペクトル・バランス・プリセット。

【０１１１】

〈１．５節　パラメータ平滑化〉

１．３節では、急激な変化を避け、よってオーディオ改善装置のパラメータの急激な変化を避けるためにオーディオ型の信頼値を平滑化することを論じた。他の措置も可能である。一つは、オーディオ型に基づいて調整されるパラメータを平滑化することであり、本節で論じる。他方は、オーディオ分類器および／または調整ユニットを、オーディオ分類器の結果の変化を遅らせるよう構成することであり、これについては１．６節で論じる。

【０１１２】

ある実施形態では、パラメータは、遷移点における可聴アーチファクトを導入しうる素速い変化を避けるために、次のように、さらに平滑化されることができる。

【０１１３】

【数７】

$$\tilde{L}(t) = \tau\tilde{L}(t-1) + (1-\tau)L(t) \qquad (3')$$

ここで、チルダ付きのL(t)は平滑化されたパラメータ、L(t)は平滑化されていないパラメータ、τは時定数を表わす係数、tは現在の時間、t－1は最後の時間である。

【０１１４】

すなわち、図８に示されるように、オーディオ処理装置は、パラメータ平滑化ユニット８１４を有していてもよい。これは、調整ユニット３００によって調整される（ダイアログ向上器４０２、サラウンド仮想化器４０４、ボリューム平準化器４０６および等化器４０８のうちの少なくとも一つのような）オーディオ改善装置のパラメータについて、現在の時間において調整ユニットによって決定されたパラメータ値および最後の時間の平滑化されたパラメータ値の重み付けされた和を計算することによって、現在の時間における調

整ユニット３００によって決定されるパラメータ値を平滑化する。

【０１１５】

時定数τは、応用の個別的な要求および／またはオーディオ改善装置４００の実装に基づく固定値であることができる。時定数τは、オーディオ型に基づいて、特に、音楽から発話、発話から音楽など、あるオーディオ型から別のオーディオ型への種々の遷移型に基づいて適応的に変更されてもよい。

【０１１６】

等化器を例に取る（さらなる詳細は第五部で言及されうる）。等化は、音楽コンテンツに適用するのには良好だが、発話コンテンツに適用するのはよくない。よって、等化のレベルを平滑化するためには、オーディオ信号が音楽から発話に遷移するときは時定数は比較的小さくてもよく、それにより発話コンテンツに対してより小さな等化レベルがより迅速に適用されることができる。他方、発話から音楽への遷移についての時定数は、遷移点における可聴アーチファクトを避けるために、比較的大きいことができる。

【０１１７】

遷移型（たとえば、発話から音楽または音楽から発話）を推定するために、コンテンツ分類結果は直接使われることができる。すなわち、オーディオ・コンテンツを音楽または発話に分類するれば、遷移型を得ることは単純になる。より連続的な仕方で遷移を推定できるよう、オーディオ型の硬決定を直接比較する代わりに、推定された平滑化されない等化レベルにも頼ることができる。一般的な発想は、平滑化されない等化レベルが増大する場合には、それは発話から音楽への（またはより音楽的への）遷移を示し、そうでない場合にはそれは音楽から発話への（またはより発話的への）遷移により近い。異なる遷移型を区別することにより、時定数は対応して設定されることができる。一例は次のようなものである。

【０１１８】

【数８】

$$\tau(t)=\begin{cases}\tau_1 & L(t)\geq L(t-1)\\ \tau_2 & L(t)<L(t-1)\end{cases} \qquad (4')$$

ここで、τ(t)は、コンテンツに依存する時間変化する時定数であり、τ1およびτ2は二つのプリセット時定数値であり、通例τ1＞τ2を満たす。直観的には、上記の関数は、等化レベルが増加するときには比較的遅い遷移を示し、等化レベルが減少するときには比較的速い遷移を示す。だが本願はこれに限定されるものではない。さらに、パラメータは等化レベルに限定されず、他のパラメータであってもよい。すなわち、パラメータ平滑化ユニット８１４は、重み付けされた和を計算するための重みが、調整ユニット３００によって決定されるパラメータ値の増加トレンドまたは減少トレンドに基づいて適応的に変えられるよう構成されてもよい。

【０１１９】

〈１．６節　オーディオ型の遷移〉

図９および図１０を参照して、オーディオ型の急激な変化を避け、よってオーディオ改善装置のパラメータの急激な変化を避けるためのもう一つの方式が記述される。

【０１２０】

図９に示されるように、オーディオ処理装置１００はさらに、オーディオ分類器２００が連続的に同じ新しいオーディオ型を出力する持続時間を測定するためのタイマー９１６を有していてもよい。調整ユニット３００は、新しいオーディオ型の持続時間の長さが閾値に達するまで、現在のオーディオ型を使い続けるよう構成されてもよい。

【０１２１】

換言すれば、図１０に示されるように、観察（または維持）フェーズが導入される。調

整ユニット３００が実際に新しいオーディオ型を使う前に、（持続時間の長さの閾値に対応する）観察フェーズにおいて、オーディオ型が本当に変化したのかどうかを確証するために、ある連続量の時間にわたってオーディオ型の変化がさらにモニタリングされる。

【０１２２】

図１０に示されるように、矢印（１）は、現在状態が型Aであり、オーディオ分類器２００の結果が変わらない状況を示す。

【０１２３】

現在状態が型Aであり、オーディオ分類器２００の結果が型Bになる場合、タイマー９１６は計時を開始する、あるいは図１０に示されるように、プロセスは観察フェーズにはいり（矢印（２））、残存（hangover）カウントcntの初期値が設定される。これは観察継続時間の長さ（前記閾値に等しい）を示す。

【０１２４】

次いで、オーディオ分類器２００が連続的に型Bを出力し、cntは連続的に減少して（矢印（３））、しまいにはcntは0に等しくなる（すなわち、新しい型Bの持続時間の長さが閾値に達する）場合、調整ユニット３００は新しいオーディオ型Bを使用しうる（矢印（４））。あるいは、換言すれば、この時点になってはじめて、オーディオ型は本当に型Bに変わったと見なされうる。

【０１２５】

そうでなく、cntが0になる前に（持続時間の長さが閾値に達する前に）オーディオ分類器２００の出力がもとの型Aに戻る場合には、観察フェーズは打ち切られ、調整ユニット３００は相変わらずもとの型Aを使う（矢印（５））。

【０１２６】

型Bから型Aへの変化は、上記のプロセスと同様であってもよい。

【０１２７】

上記のプロセスでは、閾値（または残存カウント）は用途の要件に基づいて設定されうる。これはあらかじめ定義された固定値であってもよい。これは適応的に設定されてもよい。ある変形では、閾値は、あるオーディオ型から別のオーディオ型への異なる遷移対については異なる。たとえば、型Aから型Bに変わるとき、閾値は第一の値であってもよく、型Bから型Aに変わるとき、閾値は第二の値であってもよい。

【０１２８】

もう一つの変形では、残存カウント（閾値）は、新しいオーディオ型の信頼値と負に相関していてもよい。一般的な発想は、信頼値が二つの型の間の混乱を示す場合には（たとえば、信頼値が約0.5しかないときは）、観察継続時間は長い必要がある。そうでない場合には、継続時間は比較的短くてもよい。このガイドラインに従い、例示的な残存カウント（hangover　count）は、次の公式によって設定されることができる。

【０１２９】

$$HangCnt = C \cdot |0.5 - Conf| + D$$

ここで、HangCntは残存継続時間または閾値であり、CおよびDは用途の要求に基づいて設定されることのできる二つのパラメータであり、通例、Cは負、Dは正の値である。

【０１３０】

なお、タイマー９１６（よって上記の遷移プロセス）はオーディオ処理装置の一部だがオーディオ分類器２００の外部として記述した。他のいくつかの実施形態では、まさに７．３節で述べるように、オーディオ分類器２００の一部と見なされてもよい。

【０１３１】

〈１．７節　実施形態の組み合わせおよび応用シナリオ〉

上記で論じたすべての実施形態およびその変形は、そのいかなる組み合わせにおいて実装されてもよく、異なる部／実施形態において言及されるが同じまたは同様の機能をもついかなる構成要素も同じまたは別個の構成要素として実装されてもよい。

【０１３２】

具体的には、以上において実施形態およびその変形を記述するとき、前の実施形態また

は変形においてすでに記述されたものと同様の参照符号をもつ構成要素は省略され、異なる構成よそが記述されるだけである。実際、これらの異なる構成要素は、他の実施形態または変形の構成要素と組み合わされたり、あるいは単独で別個の解決策を構成したりすることができる。たとえば、図１ないし図１０を参照して述べた解決策の任意の二つ以上が互いと組み合わされてもよい。最も完備な解決策として、オーディオ処理装置はオーディオ・コンテンツ分類器２０２およびオーディオ・コンテキスト分類器２０４の両方ならびに平滑化ユニット７１２、パラメータ平滑化ユニット８１４およびタイマー９１６を有していてもよい。

【０１３３】

先述したように、オーディオ改善装置４００は、ダイアログ向上器４０２、サラウンド仮想化器４０４、ボリューム平準化器４０６および等化器４０８を含んでいてもよい。オーディオ処理装置１００は、それらの任意の一つまたは複数を含んでいてもよく、調整ユニット３００がそれに適応されてもよい。複数のオーディオ改善装置４００に関わるとき、調整ユニット３００は、それぞれのオーディオ改善装置４００に固有の複数のサブユニット３００Ａ～３００Ｄ（図１５、図１８、図２０および図２２）を含むものと見なされてもよし、あるいは相変わらず一つの連合した調整ユニットと見なされてもよい。あるオーディオ改善装置に固有であるとき、調整ユニット３００はオーディオ分類器２００および他の可能なコンポーネントと一緒になって、その特定のオーディオ改善装置のコントローラと見なされてもよい。これについては、後述する第二部～第五部において詳細に論じる。

【０１３４】

さらに、オーディオ改善装置４００は、上述した例に限定されず、他のいかなるオーディオ改善装置を含んでいてもよい。

【０１３５】

さらに、すでに論じた任意の解決策またはそれらの任意の組み合わせは、本開示の他の部分において記述または含意される任意の実施形態とさらに組み合わされてもよい。特に、第六部および第七部において論じられるオーディオ分類器の実施形態は、オーディオ処理装置において使用されてもよい。

【０１３６】

〈１．８節　オーディオ処理方法〉

上記の実施形態におけるオーディオ処理装置を記述する過程で、いくつかのプロセスまたは方法も開示されていることは明らかである。以下では、これらの方法の概要が与えられるが、上記ですでに論じた詳細の一部は繰り返さない。ただし、これらの方法はオーディオ処理装置を記述する過程において開示されているものの、これらの方法は必ずしも記載されるコンポーネントを採用するものではなく、必ずしもそうしたコンポーネントによって実行されるのではない。たとえば、オーディオ処理装置の実施形態は、部分的または完全にハードウェアおよび／またはファームウェアを用いて実現されてもよく、一方、以下で論じるオーディオ処理方法は、オーディオ処理装置のハードウェアおよび／またはファームウェアを採用してもよいが、完全にコンピュータ実行可能プログラムによって実現されてもよい。

【０１３７】

図１１～図１４を参照して以下でこれらの方法について述べる。本方法がリアルタイムで実装されるときは、オーディオ信号のストリーミング属性に対応して、さまざまな動作が繰り返され、異なる動作は必ずしも同じオーディオ・セグメントに関してではないことに注意されたい。

【０１３８】

図１１に示される実施形態では、オーディオ処理方法が提供される。まず、処理されるべきオーディオ信号がリアルタイムで少なくとも一つのオーディオ型に分類される（動作１１０２）。前記少なくとも一つのオーディオ型の信頼値に基づいて、オーディオ改善のための少なくとも一つのパラメータが連続的に調整されることができる（動作１１０４）

。オーディオ改善は、ダイアログ向上（動作１１０６）、サラウンド仮想化（動作１１０８）、ボリューム平準化（動作１１１０）および／または等化（動作１１１２）であってもよい。対応して、前記少なくとも一つのパラメータは、ダイアログ向上処理、サラウンド仮想化処理、ボリューム平準化処理および等化処理のうちの少なくとも一つについての少なくとも一つのパラメータを含んでいてもよい。

【０１３９】

ここで、「リアルタイムで」および「連続的に」はオーディオ型が、よってパラメータも、オーディオ信号の特定の内容とともにリアルタイムで変化することを意味する。「連続的に」は、調整が、急激または離散的な調整ではなく、信頼値に基づく連続的な調整であることをも意味する。

【０１４０】

オーディオ型はコンテンツ型および／またはコンテキスト型を含んでいてもよい。対応して、調整の動作１１０４は、少なくとも一つのコンテンツ型の信頼値および少なくとも一つのコンテキスト型の信頼値に基づいて前記少なくとも一つのパラメータを調整するよう構成されていてもよい。コンテンツ型はさらに、短期的音楽、発話、背景音およびノイズのコンテンツ型のうちの少なくとも一つを含んでいてもよい。コンテキスト型はさらに、長期的音楽、映画的メディア、ゲームおよびVoIPのコンテキスト型のうちの少なくとも一つを含んでいてもよい。

【０１４１】

他のいくつかのコンテキスト型スキームも提案される。たとえば、VoIPおよび非VoIPを含むVoIP関係コンテキスト型および高品質オーディオまたは低品質オーディオを含むオーディオ品質型などである。

【０１４２】

短期的音楽は、種々の標準に従ってサブ型にさらに分類されてもよい。優勢な源の存在に依存して、優勢な源のない音楽および優勢な源のある音楽を含んでいてもよい。さらに、短期的音楽は、少なくとも一つのジャンル・ベースのクラスターまたは少なくとも一つの楽器ベースのクラスターまたは音楽のリズム、テンポ、音色および／または他の任意の音楽的属性に基づいて分類された少なくとも一つの音楽クラスターを含んでいてもよい。

【０１４３】

コンテンツ型およびコンテキスト型の両方が識別されたとき、コンテンツ型の重要性は、そのコンテンツ型が位置しているところのコンテキスト型によって決定されてもよい。すなわち、異なるコンテキスト型のオーディオ信号におけるコンテンツ型は、オーディオ信号のコンテキスト型に依存して異なる重みを割り当てられる。より一般には、あるオーディオ型が別のオーディオ型に影響してもよく、別のオーディオ型の前提であってもよい。したがって、調整１１０４の動作は、あるオーディオ型の重みを別のオーディオ型の信頼値を用いて修正するよう構成されていてもよい。

【０１４４】

オーディオ信号が同時に（すなわち、同じオーディオ・セグメントに関して）複数のオーディオ型に分類されるとき、調整１１０４の動作は、そのオーディオ・セグメントを改善するためのパラメータ（単数または複数）を調整するための識別されたオーディオ型の一部または全部を考慮してもよい。たとえば、調整１１０４の動作は、前記少なくとも一つのオーディオ型の重要性に基づいて前記少なくとも一つのオーディオ型の信頼値に重み付けするよう構成されていてもよい。あるいは、調整１１０４の動作は、前記オーディオ型の少なくともいくつかを、その信頼値に基づいてそれらに重み付けすることを通じて考慮するよう構成されていてもよい。ある特別な場合には、調整１１０４の動作は、信頼値に基づいて少なくとも一つの優勢なオーディオ型を考慮するよう構成されていてもよい。

【０１４５】

結果の急激な変化を避けるために、平滑化方式が導入されてもよい。

【０１４６】

調整されたパラメータ値は平滑化されてもよい（図１２の動作１２１４）。たとえば、

現在の時間における調整１１０４の動作によって決定されたパラメータ値が、現在の時間における調整の動作によって決定されたパラメータ値と、最後の時間の平滑化されたパラメータ値の重み付けされた和で置き換えられてもよい。このように、逐次反復される平滑化動作を通じて、パラメータ値は時間軸上で平滑化される。

【０１４７】

重み付けされた和を計算するための重みは、オーディオ信号のオーディオ型に基づいてまたはあるオーディオ型から別のオーディオ型への異なる遷移対に基づいて適応的に変更されてもよい。あるいはまた、重み付けされた和を計算するための重みは、調整の動作によって決定されたパラメータ値の増加または減少トレンドに基づいて適応的に変更される。

【０１４８】

もう一つの平滑化方式が図１３に示されている。すなわち、本方法はさらに、各オーディオ型について、現在における実際の信頼値と最後の時間の平滑化された信頼値の重み付けされた和を計算することによって、現在の時間におけるオーディオ信号の信頼値を平滑化することを含んでいてもよい（動作１３０３）。パラメータ平滑化動作１２１４と同様に、重み付けされた和を計算するための重みは、オーディオ信号のオーディオ型の信頼値に基づいて、またはあるオーディオ型から別のオーディオ型への異なる遷移対に基づいて適応的に変更されてもよい。

【０１４９】

もう一つの平滑化方式は、オーディオ分類動作１１０２の出力が変わったとしてもあるオーディオ型から別のオーディオ型への遷移を遅らせるバッファ機構である。すなわち、調整１１０４の動作は、すぐに新しいオーディオ型を使うのではなく、オーディオ分類動作１１０２の出力の安定化を待つ。

【０１５０】

具体的には、本方法は、分類動作が同じ新しいオーディオ型を連続的に出力する持続時間を測定することを含んでいてもよい（図１４の動作１４０３）。ここで、調整１１０４の動作は、新しいオーディオ型の持続時間の長さがある閾値に達する（動作１４０３５における「Y」および動作１１０４２）まで、現在のオーディオ型を使い続ける（動作１４０３５における「N」および動作１１０４１）よう構成される。具体的には、オーディオ分類動作１１０２から出力されるオーディオ型が、オーディオ・パラメータ調整動作１１０４において使われている現在のオーディオ型に関して変化するとき（動作１４０３において「Y」）、計時が始まる（動作１４０３２）。オーディオ分類動作１１０２が新しいオーディオ型を出力することを続ければ、すなわち、動作１４０３１における判断が「Y」であり続ければ、計時は続く（動作１４０３２）。最終的には、新しいオーディオ型の持続時間が閾値に達すると（動作１４０３５における「Y」）、調整動作１１０４は新しいオーディオ型を使い（動作１１０４２）、計時は、オーディオ型の次の切り換えに備えてリセットされる（動作１４０３４）。閾値に達するまでは（動作１４０３５における「N」）、調整動作１１０４は現在のオーディオ型を使い続ける（動作１１０４１）。

【０１５１】

ここで、計時はタイマーの機構（カウントアップまたはカウントダウン）を用いて実装されてもよい。計時が始まった後、ただし閾値に達する前に、オーディオ分類動作１１０２の出力が調整動作１１０４において使われている現在のオーディオ型に戻る場合には、調整動作１１０４によって使用される現在のオーディオ型に関して変化がないと見なされるべきである（動作１４０３１における「N」）。だが（オーディオ信号における分類されるべき現在のオーディオ・セグメントに対応する）現在の分類結果は、（オーディオ信号における分類されるべき前のオーディオ・セグメントに対応する）オーディオ分類動作１１０２の前の出力に関して変化し（動作１４０３３における「Y」）、よって計時は、次の変化（動作１４０３１における「Y」）が計時を開始するまで、リセットされる（動作１４０３４）。むろん、オーディオ分類動作１１０２の分類結果がオーディオ・パラメータ調整動作１１０４によって使用される現在のオーディオ型に関して変化せず（動作１

４０３１における「N」）、前の分類に関しても変化しない（動作１４０３３における「N」）場合には、そのことは、オーディオ分類が安定した状態にあることを示し、現在のオーディオ型が使用され続ける。

【０１５２】

ここで使われる閾値は、あるオーディオ型から別のオーディオ型への異なる遷移対については異なっていてもよい。というのも、状態があまり安定でないときは、一般に、オーディオ改善装置が他の状態よりもそのデフォルト状態にあることが好ましいことがありうるからである。他方、新しいオーディオ型の信頼値が比較的高い場合には、新しいオーディオ型に遷移するほうが安全である。したがって、閾値は、新しいオーディオ型の信頼値と負に相関していてもよい。信頼値が高いほど、閾値は低く、つまりオーディオ型は新しいオーディオ型により速く遷移しうる。

【０１５３】

オーディオ処理装置の実施形態と同様に、オーディオ処理方法の実施形態およびその変形の任意の組み合わせが現実的である。他方、オーディオ処理方法の実施形態およびその変形のあらゆる側面は別個の解決策であってもよい。特に、オーディオ処理方法のすべてにおいて、第六部および第七部において論じるようなオーディオ分類方法が使われてもよい。

【０１５４】

〈第二部：ダイアログ向上器コントローラおよび制御方法〉

オーディオ改善装置の一例はダイアログ向上器（DE）である。これは、特に聴力が低下しつつある高齢者のために、再生時にオーディオを断続的にモニタリングし、ダイアログの存在を検出し、ダイアログの明瞭性および了解性を高める（ダイアログを聞いて理解しやすくする）ためにダイアログを向上させることをねらいとする。ダイアログが存在するかどうかを検出するほか、ダイアログが存在し、よって（動的スペクトル再均衡化（dynamic spectral rebalancing）を用いて）相応して向上される場合、了解性に最も重要な周波数も検出される。例示的なダイアログ向上方法が特許文献１に呈示されている。その全体はここに参照によって組み込まれる。

【０１５５】

ダイアログ向上器における一般的な手動の構成設定は、通例、映画的メディア・コンテンツについては有効にされるが、音楽コンテンツについては無効にされるというものである。ダイアログ向上は、音楽信号に対しては誤ってトリガーしすぎることがあるからである。

【０１５６】

利用可能なオーディオ型情報を用いて、ダイアログ向上のレベルおよび他のパラメータが、識別されたオーディオ型の信頼値に基づいて調整されることができる。先に論じたオーディオ処理装置および方法の個別的な例として、ダイアログ向上器は、第一部で論じたすべての実施形態およびそれらの実施形態の任意の組み合わせを使用してもよい。特に、ダイアログ向上器を制御する場合、図１～図１０に示されるようなオーディオ処理装置１００におけるオーディオ分類器２００および調整ユニット３００は、図１５に示されるようなダイアログ向上器コントローラ１５００を構成してもよい。この実施形態では、調整ユニットはダイアログ向上器に固有なので、３００Ａと称されてもよい。先述した部において論じたように、オーディオ分類器２００は、オーディオ・コンテンツ分類器２０２およびオーディオ・コンテキスト分類器２０４のうちの少なくとも一つを含んでいてもよく、ダイアログ向上器コントローラ１５００はさらに、型平滑化ユニット７１２、パラメータ平滑化ユニット８１４およびタイマー９１６のうちの少なくとも一つを含んでいてもよい。

【０１５７】

したがって、この部においては、先の部ですでに記述した内容を繰り返すことはせず、単にこの部のいくつかの固有の例を与える。

【０１５８】

ダイアログ向上器については、調整可能なパラメータは、ダイアログ向上のレベル、背景レベルおよび向上されるべき周波数帯域を決定するための閾値を含むがそれに限定されない。特許文献１参照。その全体はここに参照によって組み込まれる。

【０１５９】

〈２．１節　ダイアログ向上のレベル〉

ダイアログ向上のレベルに関わるとき、調整ユニット３００Ａは、ダイアログ向上器のダイアログ向上のレベルを、発話の信頼値と正に相関させるよう構成されていてもよい。追加的または代替的に、レベルは、他のコンテンツ型の信頼値に負に相関させられてもよい。こうして、ダイアログ向上のレベルは発話信頼度に（線形または非線形に）比例するように設定されることができる。よって、ダイアログ向上は、音楽および背景音（効果音）のような非発話信号についてはそれほど効果的ではない。

【０１６０】

コンテキスト型については、調整ユニット３００Ａは、ダイアログ向上器のダイアログ向上のレベルを、映画的メディアおよび／またはVoIPの信頼値と正に相関させ、ダイアログ向上器のダイアログ向上のレベルを、長期的音楽および／またはゲームの信頼値と負に相関させるよう構成されていてもよい。たとえば、ダイアログ向上のレベルは映画的メディアの信頼値に（線形または非線形に）比例するように設定されることができる。映画的メディア信頼値が0のとき（たとえば音楽コンテンツにおいて）は、ダイアログ向上のレベルも0であり、これはダイアログ向上器を無効にすることと等価である。

【０１６１】

先の部で述べたように、コンテンツ型およびコンテキスト型は合同して考慮されてもよい。

【０１６２】

〈２．２節　向上させるべき周波数帯域の決定のための閾値〉

ダイアログ向上器の作動の間、各周波数帯域について、向上される必要があるかどうかを判定するための閾値（通例、エネルギーまたはラウドネス閾値）がある。すなわち、それぞれのエネルギー／ラウドネス閾値より上の周波数帯域が向上される。それらの閾値を調整するために、調整ユニット３００Ａは、閾値を、短期的音楽および／またはノイズおよび／または背景音の信頼値と正に相関させるおよび／または閾値を発話の信頼値と負に相関させるよう構成されていてもよい。たとえば、発話信頼度が高ければ、より信頼できる発話検出を想定して閾値を下げることができ、より多くの周波数帯域が向上されることを許容する。他方、音楽信頼値が高ければ、閾値を高くすることができ、より少数の周波数帯域が向上されるようにする（よってアーチファクトをより少なくする）。

【０１６３】

〈２．３節　背景レベルへの調整〉

ダイアログ向上器におけるもう一つのコンポーネントは、図１５に示されるような最小追跡ユニット４０２２である。これは、（SNR推定および２．２節で述べた周波数帯域閾値推定のために）オーディオ信号における背景レベルを推定するために使われる。これは、オーディオ・コンテンツ型の信頼値に基づいて調整されることもできる。たとえば、発話信頼値が高い場合、最小追跡ユニットは、背景レベルを現在の最小に設定するのにより自信を持つことができる。音楽信頼度が高い場合には、背景レベルはその現在の最小よりはやや高く設定されることができ、あるいは別の仕方では現在の最小と現在フレームのエネルギーとの、現在の最小に大きな重みをかけた重み付き平均に設定されることができる。ノイズおよび背景信頼度が高い場合には、背景レベルは現在の最小値よりずっと高く設定されることができ、あるいは別の仕方では、現在の最小と現在フレームのエネルギーとの、現在の最小に小さな重みをかけた重み付き平均に設定されることができる。

【０１６４】

こうして、調整ユニット３００Ａは、最小追跡ユニットによって推定された背景レベルに調整を割り当てるよう構成されてもよい。ここで、調整ユニットはさらに、調整を、短期的音楽および／またはノイズおよび／または背景音の信頼値と正に相関させるおよび／

または調整を発話の信頼値と負に相関させるよう構成されている。ある変形では、調整ユニット３００Ａは、調整を、短期的音楽より、ノイズおよび／または背景音の信頼値と、より正に相関させるよう構成されていてもよい。

【０１６５】

〈２．４節　実施形態の組み合わせおよび応用シナリオ〉

第一部と同様に、上記で論じたすべての実施形態およびその変形は、そのいかなる組み合わせにおいて実装されてもよく、異なる部／実施形態において言及されるが同じまたは同様の機能をもついかなる構成要素も同じまたは別個の構成要素として実装されてもよい。

【０１６６】

たとえば、２．１節ないし２．３節において述べた解決策の任意の二つ以上が互いと組み合わされてもよい。そして、これらの組み合わせは、第一部および後述する他の部において記載または含意されている任意の実施形態とさらに組み合わされてもよい。特に、それぞれの種類のオーディオ改善装置または方法に対して多くの公式が実際に適用可能であるが、それらは必ずしも本開示の各部において記載または議論されていない。そのような場合、ある部において論じられる特定の公式を、特定の用途の特定の要件に応じて関連するパラメータ、係数、冪（指数）および重みが適正に調整されるだけで他の部に適用するために、本開示の各部の間で相互参照がなされることがある。

【０１６７】

〈２．５節　ダイアログ向上器制御方法〉

第一部と同様に、上記の実施形態におけるダイアログ向上器コントローラを記述する過程で、いくつかのプロセスまたは方法も開示されていることは明らかである。以下では、これらの方法の概要が与えられるが、上記ですでに論じた詳細の一部は繰り返さない。

【０１６８】

まず、第一部で論じたオーディオ処理方法の実施形態がダイアログ向上器のために使われてもよい。ダイアログ向上器のパラメータ（単数または複数）が、オーディオ処理方法によって調整されるべきターゲットの一つである。この観点から、オーディオ処理方法はダイアログ向上器制御方法でもある。

【０１６９】

この節では、ダイアログ向上器の制御に固有の側面のみが論じられる。制御方法の一般的な側面については、第一部が参照されうる。

【０１７０】

ある実施形態によれば、オーディオ処理方法はさらに、ダイアログ向上処理を含んでいてもよく、調整１１０４の動作は、ダイアログ向上のレベルを映画的メディアおよび／またはVoIPの信頼値と正に相関させるおよびまたはダイアログ向上のレベルを長期的音楽および／またはゲームの信頼値と負に相関させることを含む。すなわち、ダイアログ向上は主として、映画的メディアまたはVoIPのコンテキストにおけるオーディオ信号に向けられる。

【０１７１】

より具体的には、調整１１０４の動作は、ダイアログ向上器のダイアログ向上のレベルを発話の信頼値と正に相関させることを含んでいてもよい。

【０１７２】

本願は、ダイアログ向上処理において向上されるべき周波数帯域を調整してもよい。図１６に示されるように、それぞれの周波数帯域が向上されるべきかどうかを決定するための閾値（通例エネルギーまたはラウドネス）が、本願に従って、識別されたオーディオ型の信頼値（単数または複数）に基づいて調整されてもよい（動作１６０２）。次いで、ダイアログ向上器内で、調整された閾値に基づいて、それぞれの閾値より上の周波数帯域が選択され（動作１６０４）、向上される（動作１６０６）。

【０１７３】

具体的には、調整１１０４の動作は、それらの閾値を、短期的音楽および／またはノイ

ズおよび／または背景音の信頼値と正に相関させるおよび／またはそれらの閾値を発話の信頼値と負に相関させることを含んでいてもよい。

【０１７４】

オーディオ処理方法（特にダイアログ向上処理）は一般にさらに、オーディオ信号における背景レベルを推定することを含む。これは一般に、ダイアログ向上器４０２において実現される最小追跡ユニット４０２２によって実装され、SNR推定または周波数帯域閾値推定において使われる。本願は、背景レベルを調整するために使われてもよい。そのような状況では、背景レベルが推定された後（動作１７０２）、背景レベルはまず、オーディオ型（単数または複数）の信頼値（単数または複数）に基づいて調整され（動作１７０４）、次いでSNR推定および／または周波数帯域閾値推定において使われる（動作１７０６）。特に、調整１１０４の動作は、推定された背景レベルに調整を割り当てるよう構成されていてもよい。ここで、調整１１０４の動作は、調整を、短期的音楽および／またはノイズおよび／または背景音の信頼値と正に相関させるおよび／または調整を発話の信頼値と負に相関させるよう構成されていてもよい。

【０１７５】

より具体的には、調整１１０４の動作は、調整を、短期的音楽よりも、ノイズおよび／または背景の信頼値と、より正に相関させるよう構成されていてもよい。

【０１７６】

オーディオ処理装置の実施形態と同様に、オーディオ処理方法の実施形態およびその変形の任意の組み合わせが現実的である。他方、オーディオ処理方法の実施形態およびその変形のあらゆる側面は別個の解決策であってもよい。さらに、この節に記載される解決策の任意の二つ以上が互いと組み合わされてもよく、これらの組み合わせがさらに、第一部または後述する他の部において記載または含意される任意の実施形態と組み合わされてもよい。

【０１７７】

〈第三部：サラウンド仮想化器コントローラおよび制御方法〉

サラウンド仮想化器は、サラウンドサウンド信号（マルチチャネル5.1および7.1など）がPCの内部スピーカーを通じてまたはヘッドフォンを通じてレンダリングされることを可能にする。すなわち、内蔵ラップトップ・スピーカーまたはヘッドフォンのようなステレオ装置を用いて、仮想的にサラウンド効果を生成し、消費者のために映画館の体験を提供するのである。サラウンド仮想化器では、マルチチャネル・オーディオ信号に関連付けられたさまざまなスピーカー位置からくる音の耳への到来をシミュレートするために、通例、頭部伝達関数（HRTF: Head Related Transfer Function）が利用される。

【０１７８】

現在のサラウンド仮想化器はヘッドフォン上でよく機能するが、組み込みスピーカーと異なるコンテンツに対して異なる仕方で機能する。一般医、映画的メディア・コンテンツはスピーカーのためにサラウンド仮想化器を有効にし、一方、音楽はそうしない。あまりに薄っぺらに聞こえることがありうるからである。

【０１７９】

サラウンド仮想化器における同じパラメータが映画的メディアおよび音楽コンテンツの両方について同時に良好な音像を作り出すことはできないので、パラメータはコンテンツに基づいてより精密に調整される必要がある。利用可能なオーディオ型情報、特に音楽信頼値および発話信頼値ならびに他の何らかのコンテンツ型情報およびコンテキスト情報を用いて、機能は本願とともに行なうことができる。

【０１８０】

第二部と同様に、第一部で論じたオーディオ処理装置および方法の個別的な例として、サラウンド仮想化器４０４は、第一部で論じたすべての実施形態およびそこで開示されたそれらの実施形態の任意の組み合わせを使用してもよい。特に、サラウンド仮想化器を制御する場合、図１～図１０に示されるようなオーディオ処理装置１００におけるオーディオ分類器２００および調整ユニット３００は、図１８に示されるようなサラウンド仮想化

器コントローラ１８００を構成してもよい。この実施形態では、調整ユニットはサラウンド仮想化器に固有なので、３００Ｂと称されてもよい。第二部と同様に、オーディオ分類器２００は、オーディオ・コンテンツ分類器２０２およびオーディオ・コンテキスト分類器２０４のうちの少なくとも一つを含んでいてもよく、サラウンド仮想化器コントローラ１８００はさらに、型平滑化ユニット７１２、パラメータ平滑化ユニット８１４およびタイマー９１６のうちの少なくとも一つを含んでいてもよい。

【０１８１】

したがって、この部においては、第一部ですでに記述した内容を繰り返すことはせず、単にこの部のいくつかの固有の例を与える。

【０１８２】

サラウンド仮想化器については、調整可能なパラメータは、サラウンド・ブースト量およびサラウンド仮想化器４０４の開始周波数を含むがそれに限定されない。

【０１８３】

〈３．１節　サラウンド・ブースト量〉

サラウンド・ブースト量に関わるとき、調整ユニット３００Ｂは、サラウンド仮想化器４０４のサラウンド・ブースト量を、ノイズおよび／または背景および／または発話の信頼値と正に相関させるおよび／またはサラウンド・ブースト量を短期的音楽の信頼値と負に相関させるよう構成されていてもよい。

【０１８４】

特に、音楽（コンテンツ型）が受け入れ可能に聞こえるようにサラウンド仮想化器４０４を修正するために、調整ユニット３００Ｂの例示的な実装は、短期的音楽信頼値に基づいてサラウンド・ブーストの量を調整することができる。たとえば、

$$SB \propto (1 - Conf_{music}) \qquad (5)$$

ここで、SBはサラウンド・ブースト量、$Conf_{music}$は短期的音楽の信頼値である。

【０１８５】

それは、音楽についてサラウンド・ブーストを減少させ、音楽が生気ないように聞こえることを防ぐ。

【０１８６】

同様に、発話信頼値も利用でき、たとえば、

$$SB \propto (1 - Conf_{music}) * Conf_{speech}{}^{\alpha} \qquad (6)$$

ここで、$Conf_{speech}$は発話の信頼値であり、αは指数の形の重み付け係数であり、1～2の範囲にあってもよい。この公式は、サラウンド・ブースト量は純粋な発話（高い発話信頼度および低い音楽信頼度）についてのみ高くなることを示す。

【０１８７】

あるいは、発話の信頼値のみを考慮することもできる

$$SB \propto Conf_{speech} \qquad (7)$$

さまざまな変形が同様にして設計できる。特に、ノイズまたは背景音について、公式(5)ないし(7)と同様の公式が構築されてもよい。さらに、それら四つのコンテンツ型の効果は任意の組み合わせにおいて一緒に考慮されてもよい。そのような状況において、ノイズおよび背景は周囲音であり、大きなブースト量をもってもより安全である。発話は、話者が通例はスクリーンの前方に座ると想定して、中程度のブースト量をもつことができる。したがって、調整ユニット３００Ｂは、サラウンド・ブースト量を、コンテンツが型発話よりも、ノイズおよび／または背景の信頼値と、より正に相関させるよう構成されていてもよい。

【０１８８】

各コンテンツ型について期待されるブースト量（これは重みと等化である）をあらかじめ定義していたとして、もう一つの代替を適用することもできる。

【０１８９】

【数９】

$$\hat{a} = \frac{a_{speech} \cdot Conf_{speech} + a_{music} \cdot Conf_{music} + a_{noise} \cdot Conf_{noise} + a_{bkg} \cdot Conf_{bkg}}{Conf_{speech} + Conf_{music} + Conf_{noise} + Conf_{bkg}} \quad (8)$$

ここで、＾付きのaは推定されたブースト量、コンテンツ型の添え字をもつαはそのコンテンツ型の期待される／あらかじめ定義されたブースト量（重み）、コンテンツ型の添え字をもつConfはそのコンテンツ型の信頼値である（ここで、bkgは「background　sound」（背景音）を表わす）。状況に依存して、a_{music}は（必須ではないが）0に設定されてもよい。これはサラウンド仮想化器４０４が純粋な音楽（コンテンツ型）については無効にされることを示す。

【０１９０】

別の観点からは、公式(8)におけるコンテンツ型の添え字をもつαはそのコンテンツ型の期待される／あらかじめ定義されたブースト量であり、対応するコンテンツ型の信頼値を、すべての識別されたコンテンツ型の信頼値の和で割った商は、対応するコンテンツ型のあらかじめ定義された／期待されるブースト量の規格化された重みと見なされてもよい。すなわち、調整ユニット３００Ｂは、それらの信頼値に基づいて複数のコンテンツ型のあらかじめ定義された諸ブースト量に重み付けすることを通じて、複数のコンテンツ型のうち少なくともいくつかを考慮するよう構成されていてもよい。

【０１９１】

コンテキスト型については、調整ユニット３００Ｂは、サラウンド仮想化器４０４のサラウンド・ブースト量を、映画的メディアおよび／またはゲームの信頼値と正に相関させ、サラウンド・ブースト量を、長期的音楽および／またはVoIPの信頼値と負に相関させるよう構成されていてもよい。次いで、(5)ないし(8)と同様の公式が構築されうる。

【０１９２】

特殊な例として、サラウンド仮想化器４０４は、純粋な映画的メディアおよび／またはゲームについては有効にされ、音楽および／またはVoIPについては無効にされることができる。一方、サラウンド仮想化器４０４のブースト量は映画的メディアおよびゲームについて異なるように設定されることができる。映画的メディアはより高いブースト量を使い、ゲームはより低い。したがって、調整ユニット３００Ｂは、サラウンド・ブースト量を、ゲームよりも、映画的メディアの信頼値と、より正に相関させるよう構成されてもよい。

【０１９３】

コンテンツ型と同様に、オーディオ信号のブースト量は、コンテキスト型の信頼値の重み付き平均に設定されることもできる。

【０１９４】

【数１０】

$$\hat{a} = \frac{a_{MOVIE} \cdot Conf_{MOVIE} + a_{MUSIC} \cdot Conf_{MUSIC} + a_{GAME} \cdot Conf_{GAME} + a_{VOIP} \cdot Conf_{VOIP}}{Conf_{MOVIE} + Conf_{MUSIC} + Conf_{GAME} + Conf_{VOIP}} \quad (9)$$

ここで、＾付きのaは推定されたブースト量、コンテキスト型の添え字をもつαはそのコンテキスト型の期待される／あらかじめ定義されたブースト量（重み）、コンテキストの添え字をもつConfはそのコンテキスト型の信頼値である。状況に依存して、a_{MUSIC}およびa_{VOIP}は（必須ではないが）0に設定されてもよい。これはサラウンド仮想化器４０４が純粋な音楽（コンテキスト型）およびまたは純粋なVoIPについては無効にされることを示す。

【０１９５】

やはりコンテンツ型と同様に、公式(9)におけるコンテキスト型の添え字をもつαはそのコンテキスト型の期待される／あらかじめ定義されたブースト量であり、対応するコンテキスト型の信頼値を、すべての識別されたコンテキスト型の信頼値の和で割った商は、対応するコンテキスト型のあらかじめ定義された／期待されるブースト量の規格化された重みと見なされてもよい。すなわち、調整ユニット３００Ｂは、それらの信頼値に基づいて複数のコンテキスト型のあらかじめ定義された諸ブースト量に重み付けすることを通じて、複数のコンテキスト型のうち少なくともいくつかを考慮するよう構成されていてもよい。

【０１９６】

〈３．２節　開始周波数〉

他のパラメータも開始周波数のようなサラウンド仮想化器において修正されることができる。一般に、オーディオ信号中の高周波数成分は空間的にレンダリングされるのに、より好適である。たとえば、音楽では、ベースがより多くのサラウンド効果をもつようにレンダリングされると、おかしく聞こえる。よって、特定のオーディオ信号について、サラウンド仮想化器は、それより上の成分が空間的にレンダリングされ、それより下の成分が保持される周波数閾値を決定する必要がある。周波数閾値は開始周波数である。

【０１９７】

本願のある実施形態によれば、サラウンド仮想化器についての開始周波数は音楽コンテンツに対しては高められることができ、それにより音楽信号についてはより多くのベースが保持されることができる。すると、調整ユニット３００Ｂは、サラウンド仮想化器の開始周波数を短期的音楽の信頼値と正に相関させるよう構成されうる。

【０１９８】

〈３．３節　実施形態の組み合わせおよび応用シナリオ〉

第一部と同様に、上記で論じたすべての実施形態およびその変形は、そのいかなる組み合わせにおいて実装されてもよく、異なる部／実施形態において言及されるが同じまたは同様の機能をもついかなる構成要素も同じまたは別個の構成要素として実装されてもよい。

【０１９９】

たとえば、３．１節および３．２節において述べた解決策の任意の二つ以上が互いと組み合わされてもよい。そして、これらの組み合わせの任意のものが、第一部、第二部および後述する他の部において記載または含意されている任意の実施形態とさらに組み合わされてもよい。

【０２００】

〈３．４節　サラウンド仮想化器制御方法〉

第一部と同様に、上記の実施形態におけるサラウンド仮想化器コントローラを記述する過程で、いくつかのプロセスまたは方法も開示されていることは明らかである。以下では、これらの方法の概要が与えられるが、上記ですでに論じた詳細の一部は繰り返さない。

【０２０１】

まず、第一部で論じたオーディオ処理方法の実施形態がサラウンド仮想化器について使用されてもよい。サラウンド仮想化器のパラメータ（単数または複数）が、オーディオ処理方法によって調整されるべきターゲットの一つである。この観点から、オーディオ処理方法はサラウンド仮想化器制御方法でもある。

【０２０２】

この節では、サラウンド仮想化器の制御に固有の側面のみが論じられる。制御方法の一般的な側面については、第一部が参照されうる。

【０２０３】

ある実施形態によれば、オーディオ処理方法はさらに、サラウンド仮想化処理を含んでいてもよく、調整する動作１１０４はサラウンド仮想化処理のサラウンド・ブースト量をノイズおよび／または背景および／または発話の信頼値と正に相関させるおよび／またはサラウンド・ブースト量を短期的音楽の信頼値と負に相関させるよう構成されていてもよ

い。
【０２０４】
特に、調整する動作１１０４はサラウンド・ブースト量をノイズおよび／または背景および／または発話の信頼値と、コンテンツ型発話よりも、より正に相関させるよう構成されていてもよい。
【０２０５】
代替的または追加的に、サラウンド・ブースト量は、コンテキスト型（単数または複数）の信頼値（単数または複数）に基づいて調整されてもよい。特に、調整する動作１１０４は、サラウンド仮想化処理のサラウンド・ブースト量を映画的メディアおよび／またはゲームの信頼値と正に相関させるおよび／またはサラウンド・ブースト量を長期的音楽および／またはVoIPの信頼値と負に相関させるよう構成されていてもよい。
【０２０６】
特に、調整する動作１１０４はサラウンド・ブースト量を映画的メディアと、ゲームよりも、より正に相関させるよう構成されていてもよい。
【０２０７】
調整されるべきもう一つのパラメータは、サラウンド仮想化処理のための開始周波数である。図１９に示されるように、開始周波数はオーディオ型（単数または複数）の信頼値（単数または複数）に基づいてまず調整され（動作１９０２）、次いで、サラウンド仮想化器は、開始周波数より上のオーディオ・コンポーネントを処理する（動作１９０４）。特に、調整する動作１１０４は、サラウンド仮想化処理の開始周波数を短期的音楽の信頼値と正に相関させるよう構成されていてもよい。
【０２０８】
オーディオ処理装置の実施形態と同様に、オーディオ処理方法の実施形態およびその変形の任意の組み合わせが現実的である。他方、オーディオ処理方法の実施形態およびその変形のあらゆる側面は別個の解決策であってもよい。さらに、この節に記載される解決策の任意の二つ以上が互いと組み合わされてもよく、これらの組み合わせがさらに、本開示の他の部において記載または含意される任意の実施形態と組み合わされてもよい。
【０２０９】
〈第四部：ボリューム平準化器コントローラおよび制御方法〉
異なるオーディオ源または同じオーディオ源の異なるピースのボリュームは時に大きく変化する。ユーザーがボリュームを頻繁に調整しなければならないので、これはわずらわしい。ボリューム平準化器（VL: Volume Leveler）は再生時のオーディオ・コンテンツのボリュームを調整し、ターゲット・ラウドネス値に基づいて時間軸上でほとんど一貫しているようにすることをねらいとする。例示的なボリューム平準化器は特許文献２、特許文献３、特許文献４に記載されている。これら三つの文書はここに全体において参照によって組み込まれる。
【０２１０】
ボリューム平準化器は、何らかの仕方でオーディオ信号のラウドネスを連続的に測定し、次いで利得（gain）の量だけ信号を修正する。gainはオーディオ信号のラウドネスを修正するためのスケーリング因子であり、通例、測定されたラウドネス、所望されるターゲット・ラウドネスおよび他のいくつかの因子の関数である。適正な利得を推定するためにいくつかの因子が、ターゲット・ラウドネスに接近しかつダイナミックレンジを維持するための基礎になる基準とともに、考慮される必要がある。それは通例、自動利得制御（AGC）、聴覚イベント検出、ダイナミックレンジ制御（DRC）のようないくつかのサブ要素を含む。
【０２１１】
制御信号は、オーディオ信号の「利得」を制御するためにボリューム平準化器において一般に適用される。たとえば、制御信号は、純粋な信号解析によって導出される、オーディオ信号の大きさの変化のインジケーターであることができる。制御信号はまた、聴覚シーン解析または特定ラウドネス・ベースの（specific-loudness-based）聴覚イベント・

検出といった音響心理学的解析を通じた、新しいオーディオ・イベントが現われるかどうかを表わすオーディオ・イベント・インジケーターであることもできる。そのような制御信号は、たとえばオーディオ信号における利得の急激な変化に起因する可能な可聴アーチファクトを低減するために利得が聴覚イベント内でほぼ一定であることを保証することによって、および利得変化の多くをイベント境界の近傍に制約することによって、利得制御のためにボリューム平準化器において適用される。

【０２１２】

しかしながら、制御信号を導出する従来の方法は、情報性の聴覚イベントを非情報性（干渉性）の聴覚イベントから区別することができない。ここで、情報性の聴覚イベントは、ダイアログおよび音楽のような、意味のある情報を含むオーディオ・イベントを表わし、ユーザーがより注意を払うことがありうる。一方、非情報性の信号は、VoIPにおけるノイズのように、ユーザーにとって意味のある情報を含まない。結果として、非情報性の信号も、大きな利得を適用され、ターゲット・ラウドネス近くにブーストされうる。それはいくつかの応用では快くない。たとえば、VoIP通話では、会話の休止に現われるノイズ信号が、ボリューム平準化器により処理された後に、しばしば大きなボリュームにブーストされる。これはユーザーによって望まれない。

【０２１３】

この問題に少なくとも部分的に対処するために、本願は、第一部で論じた実施形態に基づいてボリューム平準化器を制御することを提案する。

【０２１４】

第二部および第三部と同様に、第一部で論じたオーディオ処理装置および方法の個別的な例として、ボリューム平準化器４０６は、第一部で論じたすべての実施形態およびそこで開示されたそれらの実施形態の任意の組み合わせを使用してもよい。特に、ボリューム平準化器４０６を制御する場合、図１～図１０に示されるようなオーディオ処理装置１００におけるオーディオ分類器２００および調整ユニット３００は、図２０に示されるようなボリューム平準化器４０６コントローラ２０００を構成してもよい。この実施形態では、調整ユニットはボリューム平準化器４０６に固有なので、３００Ｃと称されてもよい。

【０２１５】

すなわち、第一部の開示に基づき、ボリューム平準化器コントローラ２０００は、オーディオ信号のオーディオ型（コンテンツ型および／またはコンテキスト型など）を連続的に識別するオーディオ分類器２００と、識別されたオーディオ型の信頼値に基づいて連続的な仕方でボリューム平準化器を調整する調整ユニット３００Ｃとを有していてもよい。同様に、オーディオ分類器２００は、オーディオ・コンテンツ分類器２０２およびオーディオ・コンテキスト分類器２０４のうちの少なくとも一つを含んでいてもよく、ボリューム平準化器コントローラ２０００はさらに、型平滑化ユニット７１２、パラメータ平滑化ユニット８１４およびタイマー９１６のうちの少なくとも一つを含んでいてもよい。

【０２１６】

したがって、この部においては、第一部ですでに記述した内容を繰り返すことはせず、単にこの部のいくつかの固有の例を与える。

【０２１７】

分類結果に基づいて、ボリューム平準化器４０６の種々のパラメータが適応的に調整されることができる。たとえば非情報性信号についての利得を低減することにより、動的利得または動的利得の範囲に直接関係したパラメータを調整することができる。信号が新しい知覚可能なオーディオ・イベントである度合いを示すパラメータを調整し、そして動的利得を間接的に制御することもできる（利得は、オーディオ・イベント内でゆっくり変化するが、二つのオーディオ・イベントの境界では急激に変化することがある）。本願では、パラメータ調整またはボリューム平準化器制御機構のいくつかの実施形態が呈示される。

【０２１８】

〈４．１節　情報性および干渉性のコンテンツ型〉

上述したように、ボリューム平準化器の制御との関連で、オーディオ・コンテンツ型は情報性のコンテンツ型および干渉性のコンテンツ型として分類されうる。
調整ユニット３００Ｃは、ボリューム平準化器の動的利得をオーディオ信号の情報性コンテンツ型と正に相関させ、ボリューム平準化器の動的利得をオーディオ信号の干渉性コンテンツ型と負に相関させるよう構成されてもよい。

【０２１９】

例として、ノイズが干渉性（非情報性）であり、それは大きなボリュームにブーストされるとわずらわしいとする。動的利得を直接制御するパラメータまたは新しいオーディオ・イベントを示すパラメータは、

$$\mathrm{GainControl} \propto 1 - \mathrm{Conf}_{noise} \qquad (10)$$

のように、ノイズ信頼値（Conf_{noise}）の減少関数に比例するよう設定されることができる。

【０２２０】

ここで、簡単のため、ボリューム平準化器における利得制御に関係するすべてのパラメータ（またはその効果）を表わすために記号GainControlを用いる。ボリューム平準化器の異なる実装は異なる基礎的な意味をもつパラメータの異なる名前を使いうるからである。単一の用語GainControlを使うことは、一般性を失うことなく、短い表現をもつことができる。本質的には、これらのパラメータを調整することは、もとの利得に線形または非線形の重みを適用することと等価である。一例として、GainControlは、GainControlが小さければ利得が小さくなるよう、利得をスケーリングするために直接使われることができる。もう一つの個別的な例として、利得は、特許文献３に記載されるイベント制御信号をGainControlを用いてスケーリングすることによって間接的に制御される。同出願はここにその全体において参照によって組み込まれる。この場合、GainControlが小さいときは、ボリューム平準化器の利得の制御は、利得が時間とともに著しく変化することを防ぐよう修正される。GainControlが大きいときは、制御は、平準化器の利得がより自由に変化することを許容されるように修正される。

【０２２１】

公式(10)において記述される利得制御（もとの利得またはイベント制御信号を直接スケーリングすること）を用いて、オーディオ信号の動的利得はそのノイズ信頼値に（線形または非線形に）相関される。信号が高い信頼値でノイズであれば、採取的な利得は、因子（$1-\mathrm{Conf}_{noise}$）のため、小さくなる。このように、ノイズ信号を快くない大きなボリュームにブーストすることを避ける。

【０２２２】

公式(10)の変形例として、（VoIPなどの）用途において背景音にも関心がない場合には、背景音も同様に扱うことができ、小さな利得によって適用されうる。制御関数は、ノイズ信頼値（Conf_{noise}）および背景信頼値（Conf_{bkg}）の両方を考慮に入れることができる。たとえば、

$$\mathrm{GainControl} \propto (1 - \mathrm{Conf}_{noise}) \cdot (1 - \mathrm{Conf}_{bkg}) \qquad (11)$$

上記の公式においては、ノイズおよび背景音の両方が望まれないので、GainControlはノイズの信頼値および背景の信頼値によって等しく影響される。これは、ノイズおよび背景音が同じ重みをもつと見なしうる。状況に依存して、両者は異なる重みをもってもよい。たとえば、ノイズおよび背景音の信頼値（またはそれらの1との差）に異なる係数または異なる指数（αおよびγ）を与えてもよい。すなわち、公式(11)は

$$\mathrm{GainControl} \propto (1 - \mathrm{Conf}_{noise})^{\alpha} \cdot (1 - \mathrm{Conf}_{bkg})^{\gamma} \qquad (12)$$

または

$$\mathrm{GainControl} \propto (1 - \mathrm{Conf}_{noise}{}^{\alpha}) \cdot (1 - \mathrm{Conf}_{bkg}{}^{\gamma}) \qquad (13)$$

と書き直されてもよい。

【０２２３】

あるいはまた、調整ユニット３００Ｃは、信頼値に基づいて少なくとも一つの優勢なコンテンツ型を考慮するよう構成されていてもよい。たとえば、

$$GainControl \propto 1 - \max(Conf_{noise}, Conf_{bkg}) \quad (14)$$

公式(11)（およびその諸変形）および公式(14)の両方は、ノイズ信号および背景音信号についての小さな利得を示し、ボリューム平準化器のもとの挙動は、（発話および音楽信号におけるように）ノイズ信頼値および背景信頼値の両方が小さくGainControlが1に近いときにのみ保持される。

【０２２４】

上記の例は、優勢な干渉コンテンツ型を考慮する。状況に依存して、調整ユニット３００Ｃは、信頼値に基づいて優勢な情報性コンテンツ型を考慮するよう構成されていてもよい。より一般には、調整ユニット３００Ｃは、識別されたオーディオ型が情報性および／または干渉性オーディオ型である／を含むか否かに関わりなく、信頼値に基づいて少なくとも一つの優勢なコンテンツ型を考慮するよう構成されていてもよい。

【０２２５】

公式(10)のもう一つの例示的な変形として、発話信号が最も情報性のコンテンツであり、ボリューム平準化器のデフォルトの挙動に対して必要な修正がより少ないとすると、制御関数はノイズ信頼値（$Conf_{noise}$）および発話信頼値（$Conf_{speech}$）の両方を

$$GainControl \propto 1 - Conf_{noise} \cdot (1 - Conf_{speech}) \quad (15)$$

として考慮することができる。この館数を用いると、小さなGainControlが得られるのは、高いノイズ信頼度および低い発話信頼度をもつ信号（たとえば純粋なノイズ）についてのみであり、発話信頼度が高い場合にはGainControlは1に近くなる（よってボリューム平準化器のもとの挙動を保持する）。より一般には、あるコンテンツ型（$Conf_{noise}$など）の重みが少なくとも一つの他のコンテンツ型（$Conf_{speech}$など）をもって修正されうると見なされることができる。上記の公式(15)において、発話の信頼度はノイズの信頼度の重み係数を変化させると見なされることができる（公式(12)および(13)における重みに比べると別の種類の重み）。換言すれば、公式(10)では$Conf_{noise}$の係数が1と見なされることができ、一方、公式(15)では、いくつかの他のオーディオ型（発話などだがそれに限られない）がノイズの信頼値の重要性に影響する。よって、$Conf_{noise}$の重みが発話の信頼値によって修正されると言うことができる。本開示のコンテキストにおいて、用語「重み」はこれを含むように解釈される。すなわち、値の重要性を示すが、必ずしも規格化されていない。１．４節が参照されてもよい。

【０２２６】

別の観点からは、公式(12)および(13)と同様に、指数の形の重みが上記の関数における信頼値に適用されて、異なるオーディオ信号の優先度（または重要性）を示すことができる。たとえば、公式(15)は次のように変更できる。

【０２２７】

$$GainControl \propto 1 - Conf_{noise}^{\alpha} \cdot (1 - Conf_{speech})^{\gamma} \quad (16)$$

ここで、αおよびγは二つの重みである。これらは、平準器パラメータを修正するためにより大きく反応することが期待される場合にはより小さく設定されることができる。

【０２２８】

公式(10)～(16)は、自由に組み合わされて、異なる応用において好適でありうるさまざまな制御関数を形成することができる。音楽信頼値のような他のオーディオ・コンテンツ型の信頼値も同様の仕方で制御関数に簡単に組み込まれることができる。

【０２２９】

GainControlが信号が新しい知覚可能なオーディオ・イベントである度合いを示すパラメータを調整し、そして動的利得を間接的に制御するために使われる場合には（利得はオーディオ・イベント内ではゆっくり変化するが、二つのオーディオ・イベントの境界では急激に変化しうる）、コンテンツ型の信頼値と最終的な動的利得との間のもう一つの伝達関数があると見なされてもよい。

【０２３０】

〈４．２節　種々のコンテキストにおけるコンテンツ型〉

公式(10)～(16)における上記の制御関数は、ノイズ、背景音、短期的音楽および発話の

ようなオーディオ・コンテンツ型の信頼値を考慮に入れるが、映画的メディアおよびVoIPなど、音がどこからくるかのオーディオ・コンテキストは考慮しない。同じオーディオ・コンテンツ型が、たとえば背景音について、異なるオーディオ・コンテキストでは異なる仕方で処理される必要があることがありうる。背景音は、自動車エンジン、爆発および拍手など、さまざまな音を含む。VoIPでは意味がないかもしれないが、映画的メディアでは重要であることがある。これは、関心のあるオーディオ・コンテキストが識別され、異なるオーディオ・コンテキストについて異なる制御関数が設計される必要があることを示している。

【０２３１】

したがって、調整ユニット３００Ｃはオーディオ信号のコンテンツ型を、オーディオ信号のコンテキスト型に基づいて情報性または干渉性と見なすよう構成されていてもよい。たとえば、ノイズ信頼値および背景信頼値を考慮し、VoIPおよび非VoIPコンテキストを区別することによって、オーディオ・コンテキスト依存制御関数は次のようなものであることができる。

【０２３２】

if オーディオ・コンテキストがVoIP

$$\text{GainControl} \propto 1 - \max(\text{Conf}_{noise}, \text{Conf}_{bkg})$$

else

$$\text{GainControl} \propto 1 - \text{Conf}_{noise} \qquad (17)$$

すなわち、VoIPコンテキストでは、ノイズおよび背景音は干渉性コンテンツ型と見なされ、一方、非VoIPコンテキストでは、背景音は情報性コンテンツ型と見なされる。

【０２３３】

もう一つの例として、発話、ノイズおよび背景の信頼値を考え、VoIPおよび非VoIPコンテキストを区別するオーディオ・コンテキスト依存制御関数は次のようなものであることができる。

【０２３４】

if オーディオ・コンテキストがVoIP

$$\text{GainControl} \propto 1 - \max(\text{Conf}_{noise}, \text{Conf}_{bkg})$$

else

$$\text{GainControl} \propto 1 - \text{Conf}_{noise} \cdot (1 - \text{Conf}_{speech}) \qquad (18)$$

ここで、発話は情報性コンテンツ型として強調される。

【０２３５】

音楽も非VoIPコンテキストにおいて重要な情報性の情報であるとすると、公式(18)の後半を

$$\text{GainControl} \propto 1 - \text{Conf}_{noise} \cdot (1 - \max(\text{Conf}_{speech}, \text{Conf}_{music}) \qquad (19)$$

と拡張できる。

【０２３６】

実のところ、(10)～(16)における制御関数のそれぞれまたはその変形は、異なる／対応するオーディオ・コンテキストにおいて適用されることができる。よって、オーディオ・コンテキスト依存制御関数を形成する多数の組み合わせを生成することができる。

【０２３７】

公式(17)および(18)において区別され、利用されるVoIPおよび非VoIPコンテキストのほか、映画的メディア、長期的音楽およびゲームまたは低品質オーディオおよび高品質オーディオのような他のオーディオ・コンテキストが同様の仕方で利用されることができる。

【０２３８】

〈４．３節　コンテキスト型〉

コンテキスト型は、ノイズのようなわずらわしい音がブーストされすぎるのを避けるようボリューム平準化器を制御するために直接使われることもできる。たとえば、VoIP信頼値が、ボリューム平準化器を、その信頼値が高いときに感度を低くするよう操縦するために使われることができる。

【０２３９】
具体的には、VoIP信頼値$Conf_{VOIP}$を用いて、ボリューム平準化器のレベルは$(1-Conf_{VOIP})$に比例するよう設定されることができる。すなわち、ボリューム平準化器はVoIPコンテンツでは（VoIP信頼値が高いときは）ほとんど非作動にされる。これは、VoIPコンテキストについてボリューム平準化器を無効にする伝統的な手動のセットアップ（プリセット）と整合する。
【０２４０】
あるいはまた、オーディオ信号の種々のコンテキストについて異なる動的利得範囲を設定することができる。一般に、VL（ボリューム平準化器）量は、オーディオ信号に適用される利得の量をさらに調整し、利得に対するもう一つの（非線形な）重みと見ることができる。ある実施形態では、セットアップは次のようなものであることができる。
【０２４１】
【表１】

	映画的メディア	長期的音楽	VOIP	ゲーム
VL量	高	中	オフ(または最低)	低

さらに、期待されるVL量が各コンテキスト型についてあらかじめ定義されているとする。たとえば、VL量は映画的メディアについては1、VoIPについては0、音楽については0.6、ゲームについては0.3と設定されるが、本願はそれに限定されない。この例によれば、映画的メディアの動的利得の範囲が100%であれば、VoIPの動的利得の範囲は60%である、などとなる。オーディオ分類器２００の分類が硬判定に基づく場合には、動的利得の範囲は上記の例のように直接設定されてもよい。オーディオ分類器２００の分類が軟判定に基づく場合には、該範囲はコンテキスト型の信頼値に基づいて調整されてもよい。
【０２４２】
同様に、オーディオ分類器２００は、オーディオ信号から複数のコンテキスト型を識別することがあり、調整ユニット３００Ｃは、前記複数のコンテンツ型の重要性に基づいて前記複数のコンテンツ型の信頼値に重み付けすることによって動的利得の範囲を調整するよう構成されていてもよい。
【０２４３】
一般に、コンテキスト型についても、適切なVL量を適応的に設定するために、(10)～(16)と同様の関数が、その中のコンテンツ型をコンテキスト型で置き換えて、ここで使用されることができる。実際、表１は異なるコンテキスト型の重要性を反映する。
【０２４４】
別の観点からは、信頼値は、１．４節で論じた規格化された重みを導出するために使われてもよい。表１の各コンテキスト型について特定の量があらかじめ定義されているとすると、公式(9)と同様の公式が適用されることもできる。なお、同様の解決策が、複数のコンテンツ型および他の任意のオーディオ型に適用されてもよい。
【０２４５】
〈４．４節　実施形態の組み合わせおよび応用シナリオ〉
第一部と同様に、上記で論じたすべての実施形態およびその変形は、そのいかなる組み合わせにおいて実装されてもよく、異なる部／実施形態において言及されるが同じまたは同様の機能をもついかなる構成要素も同じまたは別個の構成要素として実装されてもよい。たとえば、４．１節ないし４．３節において述べた解決策の任意の二つ以上が互いと組み合わされてもよい。そして、これらの組み合わせの任意のものが、第一部～第三部および後述する他の部において記載または含意されている任意の実施形態とさらに組み合わされてもよい。
【０２４６】
図２１は、もとの短期的セグメント（図２１（Ａ））、パラメータ修正なしの通常のボ

リューム平準化器によって処理された短期的セグメント（図２１（Ｂ））および本願において呈示されるボリューム平準化器によって処理された短期的セグメント（図２１（Ｃ））を比べることによって、本願で提案されるボリューム平準化器コントローラの効果を示している。見て取れるように、図２１（Ｂ）に示される通常のボリューム平準化器では、ノイズ（オーディオ信号の後半）のボリュームもブーストとされ、わずらわしい。対照的に、図２１（Ｃ）に示される新しいボリューム平準化器では、オーディオ信号の実効部分のボリュームが、ノイズのボリュームを一見してブーストすることなく、ブーストとされ、聞き手に対して良好な経験を与える。

【０２４７】

〈４．５節　ボリューム平準化器制御方法〉

第一部と同様に、上記の実施形態におけるボリューム平準化器コントローラを記述する過程で、いくつかのプロセスまたは方法も開示されていることは明らかである。以下では、これらの方法の概要が与えられるが、上記ですでに論じた詳細の一部は繰り返さない。

【０２４８】

まず、第一部で論じたオーディオ処理方法の実施形態がボリューム平準化器について使用されてもよい。ボリューム平準化器のパラメータ（単数または複数）が、オーディオ処理方法によって調整されるべきターゲットの一つである。この観点から、オーディオ処理方法はボリューム平準化器制御方法でもある。

【０２４９】

この節では、ボリューム平準化器の制御に固有の側面のみが論じられる。制御方法の一般的な側面については、第一部が参照されうる。

【０２５０】

本願によれば、ボリューム平準化器制御方法であって、リアルタイムでオーディオ信号のコンテンツ型を識別し、ボリューム平準化器の動的利得をオーディオ信号の情報性コンテンツ型と正に相関させ、ボリューム平準化器の動的利得をオーディオ信号の干渉性コンテンツ型と負に相関させることによって、識別されたコンテンツ型に基づいて連続的な仕方でボリューム平準化器を調整する方法が提供される。

【０２５１】

コンテンツ型は、発話、短期的音楽、ノイズおよび背景音を含んでいてもよい。一般に、ノイズは干渉性コンテンツ型と見なされる。

【０２５２】

ボリューム平準化器の動的利得を調整するとき、コンテンツ型の信頼値に基づいて直接調整されてもよいし、あるいはコンテンツ型の信頼値の伝達関数を介して調整されてもよい。

【０２５３】

すでに述べたように、オーディオ信号は同時に複数のオーディオ型に分類されうる。複数のコンテンツ型に関わるとき、調整動作１１０４は、複数のコンテンツ型の重要性に基づいて複数のコンテンツ型の信頼値に重み付けすることを通じて、または信頼値に基づいて複数のコンテンツ型の効果に重み付けすることを通じて、複数のオーディオ・コンテンツ型の少なくともいくつかを考慮するよう構成されてもよい。特に、調整動作１１０４は、信頼値に基づいて少なくとも一つの優勢なコンテンツ型を考慮するよう構成されていてもよい。オーディオ信号が干渉性のコンテンツ型（単数または複数）および情報性のコンテンツ型（単数または複数）の両方をｇふくむと器、調整動作は、信頼値に基づいて少なくとも一つの優勢な干渉性コンテンツ型を考慮するおよび／または信頼値に基づいて少なくとも一つの優勢な情報性コンテンツ型を考慮するよう構成されてもよい。

【０２５４】

異なるオーディオ型が互いに影響してもよい。したがって、調整動作１１０４は、あるコンテンツ型の重みを、少なくとも一つの他のコンテンツ型の信頼値を用いて修正するよう構成されていてもよい。

【０２５５】

第一部で述べたように、オーディオ信号のオーディオ型の信頼値が平滑化されてもよい。平滑化動作の詳細については、第一部を参照されたい。

【０２５６】

本方法はさらに、オーディオ信号のコンテキスト型を識別することを含んでいてもよい。ここで、調整動作１１０４は、コンテキスト型の信頼値に基づいて動的利得の範囲を調整するよう構成されていてもよい。

【０２５７】

コンテンツ型の役割は、それが位置しているコンテキスト型によって制限される。したがって、オーディオ信号について同時に（すなわち、同じオーディオ・セグメントについて）コンテンツ型情報およびコンテキスト型情報の両方が得られるとき、オーディオ信号のコンテンツ型は、オーディオ信号のコンテキスト型に基づいて情報性または干渉性と判定されてもよい。さらに、異なるコンテキスト型のオーディオ信号におけるコンテンツ型は、オーディオ信号のコンテキスト型に依存して異なる重みを割り当てられてもよい。別の観点からは、コンテンツ型の情報性の性質または干渉性の性質を反映するよう、異なる重み（より大きいまたはより小さい、正の値または負の値）を使うことができる。

【０２５８】

オーディオ信号のコンテキスト型は、VoIP、映画的メディア、長期的音楽およびゲームを含んでいてもよい。コンテキスト型VoIPのオーディオ信号においては、背景音が干渉性のコンテンツ型と見なされてもよい。一方、コンテキスト型非VoIPのオーディオ信号においては、背景および／または発話および／または音楽；が情報性のコンテンツ型と見なされる。他のコンテキスト型は高品質オーディオまたは低品質オーディオを含んでいてもよい。

【０２５９】

複数のコンテンツ型と同様に、オーディオ信号が同時に（すなわち、同じオーディオ・セグメントについて）複数のコンテキスト型情報に分類されるとき、調整動作１１０４は、複数のコンテキスト型の重要性に基づいて複数のコンテキスト型の信頼値に重み付けすることを通じて、または信頼値に基づいて複数のコンテキスト型の効果に重み付けすることを通じて、複数のオーディオ・コンテンツ型の少なくともいくつかを考慮するよう構成されてもよい。特に、調整動作は、信頼値に基づいて少なくとも一つの優勢なコンテンツ型を考慮するよう構成されていてもよい。

【０２６０】

最後に、本節で論じた方法の実施形態は、第六部および第七部で論じるオーディオ分類方法を使ってもよい。詳細な記述はここでは割愛する。

【０２６１】

オーディオ処理装置の実施形態と同様に、オーディオ処理方法の実施形態およびその変形の任意の組み合わせが現実的である。他方、オーディオ処理方法の実施形態およびその変形のあらゆる側面は別個の解決策であってもよい。さらに、この節に記載される任意の二つ以上の解決策が互いと組み合わされてもよく、これらの組み合わせがさらに、本開示の他の部において記載または含意される任意の実施形態と組み合わされてもよい。

【０２６２】

〈第五部：等化器コントローラおよび制御方法〉

等化は、通例、音楽信号に適用されて、「トーン」または「音色」として知られるそのスペクトル・バランスを調整または修正する。伝統的な等化器は、ある種の楽器を強調したりまたは望まれない音を除去したりするために、ユーザーが個々の周波数帯域における周波数応答（利得）の全体的なプロファイル（曲線または形状）を構成設定できるようにする。ウィンドウズ・メディア・プレーヤーのような一般的な音楽プレーヤーは、種々のジャンルの音楽の最良の聴取経験を得るために、各周波数帯域における利得を調整するためのグラフィック・イコライザーを提供し、ロック、ラップ、ジャズおよびフォークのような種々の音楽ジャンルについての等化器プリセットの集合をも提供する。ひとたびプリセットが選択され、プロファイルが設定されたら、プロファイルが手動で修正されるまで

、同じ等化利得が信号に対して適用される。

【０２６３】

対照的に、動的等化器は、所望される音色またはトーンに関してスペクトル・バランスの全体的な一貫性を保持するために、各周波数帯域における等化利得を自動的に調整するすべを提供する。この一貫性は、オーディオのスペクトル・バランスを連続的にモニタリングし、それを所望されるプリセット・スペクトル・バランスと比較し、オーディオの元のスペクトル・バランスを所望されるスペクトル・バランスに変換するための適用される等化利得を動的に調整することによって達成される。所望されるスペクトル・バランスは、手動で選択されるまたは処理前に事前設定される。

【０２６４】

両方の種類の等化器は、次の不都合な点を共有する。最良の等化プロファイル、所望されるスペクトル・バランスまたは関係したパラメータは手動で選択される必要があり、再生時にオーディオ・コンテンツに基づいて自動的に修正されることができない。オーディオ・コンテンツ型を決定することは、種々のオーディオ信号について全体的な良好な品質を提供するために非常に重要である。たとえば、異なるジャンルのものなど、異なる音楽片は異なる等化プロファイルを必要とする。

【０２６５】

（音楽だけでなく）任意の種類のオーディオ信号が入力されることが可能である等化器システムでは、等化器パラメータはコンテンツ型に基づいて調整される必要がある。たとえば、等化器は通例、音楽信号に対しては有効にされるが、発話信号に対しては、発話の音色を大きく変えすぎて信号が不自然に聞こえるようにしてしまうことがあるので、無効にされる。

【０２６６】

この問題に少なくとも部分的に対処するために、本願は、第一部で論じた実施形態に基づいて等化器を制御することを提案する。

【０２６７】

第二部～第四部と同様に、第一部で論じたオーディオ処理装置および方法の個別的な例として、等化器４０８は、第一部で論じたすべての実施形態およびそこで開示されたそれらの実施形態の任意の組み合わせを使用してもよい。特に、等化器４０８を制御する場合、図１～図１０に示されるようなオーディオ処理装置１００におけるオーディオ分類器２００および調整ユニット３００は、図２２に示されるような等化器４０８コントローラ２０００を構成してもよい。この実施形態では、調整ユニットは等化器４０８に固有なので、３００Ｄと称されてもよい。

【０２６８】

すなわち、第一部の開示に基づき、等化器コントローラ２２００は、オーディオ信号のオーディオ型を連続的に識別するオーディオ分類器２００と、識別されたオーディオ型の信頼値に基づいて連続的な仕方で等化器を調整する調整ユニット３００Ｄとを有していてもよい。同様に、オーディオ分類器２００は、オーディオ・コンテンツ分類器２０２およびオーディオ・コンテキスト分類器２０４のうちの少なくとも一つを含んでいてもよく、ボリューム等化器コントローラ２２００はさらに、型平滑化ユニット７１２、パラメータ平滑化ユニット８１４およびタイマー９１６のうちの少なくとも一つを含んでいてもよい。

【０２６９】

したがって、この部においては、第一部ですでに記述した内容を繰り返すことはせず、単にこの部のいくつかの固有の例を与える。

【０２７０】

〈５．１節　コンテンツ型に基づく制御〉

一般に、音楽、発話、背景音およびノイズのような一般的なオーディオ・コンテンツ型について、等化器は異なるコンテンツ型に対して異なるように設定されるべきである。伝統的なセットアップと同様に、等化器は、自動的に音楽信号に対して有効にされるが、発

話に対しては無効にされることができる。あるいはより連続的な仕方で、音楽信号に対しては高い等化レベルを、発話信号に対しては低い等化レベルを設定することができる。このようにして、等化器の等化レベルは異なるオーディオ・コンテンツについて自動的に設定されることができる。

【０２７１】

特に音楽について、優勢な源をもつ音楽片に対しては等化器はあまりうまく機能しないことが観察される。不適切な等化が適用されると、優勢な源の音色が著しく変化して不自然に聞こえることがあるからである。これを考えると、優勢の源がある音楽片に対しては弾く伊藤かレベルを設定するほうがよい。一方、優勢な源のない音楽片に対しては等化レベルは高く保つことができる。この情報を用いて、等化器は種々の音楽コンテンツについて自動的に等化レベルを設定することができる。

【０２７２】

音楽は、ジャンル、楽器およびリズム、テンポおよび音色を含む一般的な音楽特性といった種々の属性に基づいてグループ化することもできる。異なる音楽ジャンルについて異なる等化器プリセットが使用されるのと同じように、これらの音楽グループ／クラスターもそれ自身の最適等化プロファイルまたは等化器曲線（伝統的な等化器の場合）または最適な所望されるスペクトル・バランス（動的等化器の場合）をもっていてもよい。

【０２７３】

上述したように、等化器は一般には音楽コンテンツに対しては有効にされるが、発話については無効にされる。等化器は、音色変化のため、ダイアログをそれほどよく聞こえさせないことがありうるからである。それを自動的に達成する一つの方法は、等化器レベルをコンテンツに、特にオーディオ・コンテンツ分類モジュールから得られる音楽信頼値および／または発話信頼値に関係付けることである。ここで、等化レベルは、適用される等化器利得の重みとして説明できる。レベルが高いほど、適用される等化は強い。たとえば、等化レベルが1であれば、フル等化プロファイルが適用される。等化レベルが0であれば、すべての利得が対応して0dBとなり、よって非等化が適用される。等化レベルは、等化器アルゴリズムの種々の実装において種々のパラメータによって表わされることがある。このパラメータの例は、特許文献２において実装されるような等化器重みである。同文献はここにその全体において参照によって組み込まれる。

【０２７４】

等化レベルを調整するために、さまざまな制御方式が設計されることができる。たとえば、オーディオ・コンテンツ型情報では、発話信頼値または音楽信頼値が等化レベルを設定するために

$$L_{eq} \propto Conf_{music} \qquad (20)$$

または

$$L_{eq} \propto 1 - Conf_{speech} \qquad (21)$$

として使用されることができる。ここで、L_{eq}は等化レベルっであり、$Conf_{music}$および$Conf_{speech}$は音楽および発話の信頼値を表わす。

【０２７５】

すなわち、調整ユニットは３００Ｄは、等化レベルを短期的音楽の信頼値と正に相関させるまたは等化レベルを発話の信頼値と負に相関させるよう構成されていてもよい。

【０２７６】

発話信頼値および音楽信頼値はさらに、等化レベルを設定するために統合して利用されることができる。一般的な発想は、等化レベルが高いのは、音楽信頼値が高くかつ発話信頼値が低いときにのみであり、他の場合には等化レベルは低いということである。たとえば、

$$L_{eq} = Conf_{music}(1 - Conf_{speech}{}^{\alpha}) \qquad (22)$$

ここで、発話信頼値は、頻繁に起こりうる、音楽信号における0でない発話信頼値を扱うために、α乗される。上記の公式を用いれば、等化は、発話成分のない純粋な音楽信号に対してはフルに適用される（1に等しいレベルで）。第一部で述べたように、αはコンテ

ンツ型の重要性に基づく重み付け係数と見なされてもよく、典型的には1ないし2に設定されることができる。

【０２７７】

発話の信頼値により大きな重みをおくならば、調整ユニット３００Ｄは、コンテンツ型発話についての信頼値がある閾値より大きいときに等化器４０８を無効にするよう構成されていてもよい。

【０２７８】

上記の記述では、音楽および発話のコンテンツ型が例に取られている。代替的または追加的に、背景音および／またはノイズの信頼値も考慮されてもよい。特に、調整ユニット３００Ｄは、等化レベルを背景の信頼値と正に相関させるおよび／または等化レベルをノイズの信頼値と負に相関させるよう構成されていてもよい。

【０２７９】

もう一つの例として、信頼値は１．４節で論じた規格化された重みを導出するために使われてもよい。期待される等化レベルが各コンテンツ型についてあらかじめ定義されているとすると（たとえば、音楽については1、発話については0、ノイズおよび背景については0.5）、公式(8)と同様の公式が厳密に適用されることができる。

【０２８０】

等化レベルはさらに、遷移点において可聴アーチファクトを導入しうる急激な変化を避けるために、平滑化されてもよい。これは、１．５節で述べたパラメータ平滑化ユニット８１４を用いてできる。

【０２８１】

〈５．２節　音楽における優勢な源の確からしさ〉

優勢な源をもつ音楽が高い等化レベルを適用されることを避けるために、等化レベルはさらに、音楽片が優勢な源を含むかどうかを示す信頼値$\mathrm{Conf}_{\mathrm{dom}}$に相関させられてもよい。たとえば、

$$L_{eq}=1-\mathrm{Conf}_{\mathrm{dom}} \qquad (23)。$$

【０２８２】

このようにして、等化レベルは優勢な源をもつ音楽片に対しては低く、優勢な源のない音楽片については高い。

【０２８３】

ここで、優勢な源をもつ音楽の信頼値が記述されているが、優勢な源をもたない音楽の信頼値を使うこともできる。すなわち、調整ユニット３００Ｄは、等化レベルを優勢な源をもたない短期的音楽の信頼値と正に相関させるおよび／または等化レベルを優勢な源をもつ短期的音楽の信頼値と負に相関させるよう構成されてもよい。

【０２８４】

１．１節で述べたように、音楽および発話と、優勢な源をもつまたはもたない音楽とは、異なる階層レベルでのコンテンツ型であるが、並行して考えられることができる。上記のような優勢な源の信頼値および発話および音楽の信頼値を合同して考えることにより、等化レベルは公式(20)～(21)のうちの少なくとも一つを(23)と組み合わせることによって設定されることができる。一例は、これら三つの公式すべてを組み合わせて

$$L_{eq}=\mathrm{Conf}_{\mathrm{music}}(1-\mathrm{Conf}_{\mathrm{speech}})(1-\mathrm{Conf}_{\mathrm{dom}}) \qquad (24)$$

とすることである。

【０２８５】

一般性のために、公式(22)のようにして、コンテンツ型の重要性に基づく異なる重みがさらに異なる信頼値に適用されることができる。

【０２８６】

もう一つの例として、$\mathrm{Conf}_{\mathrm{dom}}$はオーディオ信号が音楽であるときにのみ計算されるとして、階段関数が次のように設計されることができる。

【０２８７】

【数１１】

$$L_{eq}=\begin{cases}(1-Conf_{dom}) & Conf_{music} > \text{閾値} \\ Conf_{music}(1-conf_{speech}^{\alpha}) & \text{それ以外}\end{cases} \quad (25)$$

この関数は、分類システムがオーディオが音楽であることをかなり確証する（音楽信頼値が閾値より大きい）場合には優勢なスコアの信頼値に基づいて等化レベルを設定し、そうでない場合には、音楽および発話信頼値に基づいて設定される。すなわち、調整ユニット３００Ｄは、短期的音楽についての信頼値が閾値より大きいときに、優勢な源がない／ある短期的音楽を考慮するよう構成されていてもよい。もちろん、公式(25)における前半または後半が公式(20)ないし(24)のように修正されてもよい。

【０２８８】

１．５節で論じたのと同じ平滑化方式が適用されることもでき、時定数αがさらに、優勢な源をもつ音楽から優勢な源のない音楽への遷移または優勢な源のない音楽から優勢な源をもつ音楽への遷移といった、遷移型に基づいて設定されることができる。この目的のために、公式(4')と同様の公式が適用されることもできる。

【０２８９】

〈５．３節　等化器プリセット〉

オーディオ・コンテンツ型の信頼値に基づいて等化レベルを適応的に調整することのほか、種々のオーディオ・コンテンツについて、そのジャンル、楽器または他の特性に依存して、適切な等化プロファイルまたは所望されるスペクトル・バランス・プリセットが自動的に選ばれることもできる。同じジャンルをもつ、同じ楽器を含むまたは同じ音楽特性をもつ音楽は同じ等化プロファイルまたは所望されるスペクトル・バランス・プリセットを共有することができる。

【０２９０】

一般性のために、同じジャンル、同じ楽器または同様の音楽アトリビュートをもつ音楽グループを表わすために用語「音楽クラスター」を使う。これは、１．１節で述べたオーディオ・コンテンツ型のもう一つの階層レベルと見なすことができる。適切な等化プロファイル、等化レベルおよび／または所望されるスペクトル・バランス・プリセットがそれぞれの音楽クラスターに関連付けられてもよい。等化プロファイルは、音楽信号に対して適用される利得曲線であり、異なる音楽ジャンル（クラシック、ロック、ジャズおよびフォークなど）について使用される等化器プリセットの任意のものであることができ、所望されるスペクトル・バランス・プリセットは各クラスターについての所望される音色を表わす。図２３は、ドルビー・ホーム・シアター技術において実装される所望されるスペクトル・バランス・プリセットのいくつかの例を示している。それぞれは、可聴周波数範囲にわたる所望されるスペクトル形状を記述する。この形状は連続的に、はいってくるオーディオのスペクトル形状と比較され、等化利得は、はいってくるオーディオのスペクトル形状の、プリセットのスペクトル形状への変換との比較から、計算される。

【０２９１】

新しい音楽片については、、最も近いクラスターが決定されることができる（硬判定）または各音楽クラスターに関する信頼値が計算されることができる（軟判定）。この情報に基づいて、適正な等化プロファイルまたは所望されるスペクトル・バランス・プリセットが所与の音楽片について決定されることができる。最も簡単な方法は、

$$P_{eq}=P_{c^*} \quad (26)$$

として、最良のマッチしたクラスターの対応するプロファイルを割り当てることである。ここで、P_{eq}は推定される等化プロファイルまたは所望されるスペクトル・バランス・プリセットであり、c^*は最良のマッチした音楽クラスター（優勢なオーディオ型）のインデックスであり、これは最も高い信頼値をもつクラスターを拾うことによって得られる。

【０２９２】
さらに、0より大きい信頼値をもつ音楽クラスターが二つ以上あることがある。つまり、音楽片は、それらのクラスターと多少なりとも同様なアトリビュートをもつ。たとえば、音楽片は複数の楽器をもつことがあり、あるいは複数のジャンルのアトリビュートをもつことがある。このことは、最も近いクラスターのみを使うのではなく、すべてのクラスターを考慮することによって適正な等化プロファイルを推定するもう一つの方法を着想させる。たとえば、重み付けされた和が使用されることができる。
【０２９３】
【数１２】

$$P_{eq} = \sum_{c=1}^{N} w_c P_c \qquad (27)$$

ここで、Nはあらかじめ定義されたクラスターの数であり、w_cは各あらかじめ定義された音楽クラスター（インデックスcをもつ）に関する設計されたプロファイルP_cの重みである。この重みは、その信頼値に基づいて1に規格化されるべきである。このようにして、推定されるプロファイルは、音楽クラスターのプロファイルの混合となる。たとえば、ジャズおよびロックのアトリビュートの両方をもつ音楽片について、推定されたプロファイルは中間の何かであろう。
【０２９４】
いくつかの用途では、公式(27)に示されるすべてのクラスターに関与させることを望まないこともある。該クラスターの部分集合――現在の音楽片に最も関係している諸クラスター――のみが考慮される必要があり、公式(27)は次のように微修正できる。
【０２９５】
【数１３】

$$P_{eq} = \sum_{c'=1}^{N'} w_{c'} P_{c'} \qquad (28)$$

ここで、N'は考慮されるべきクラスターの数であり、c'はそれらのクラスターを信頼値に基づいて降順にソートしたあとのクラスター・インデックスである。部分集合を使うことにより、最も関係した諸クラスターにより焦点を当てることができ、それほど重要でないものを除外することができる。換言すれば、調整ユニット３００Ｄは、信頼値に基づいて、少なくとも一つの優勢なオーディオ型を考慮するよう構成されてもよい。
【０２９６】
上記の記述では、音楽クラスターが例として取られている。実のところ、これらの解決策は、１．１節で論じた任意の階層レベルのオーディオ型に適用可能である。よって、一般に、調整ユニット３００Ｄは、等化レベルおよび／または等化プロファイルおよび／またはスペクトル・バランス・プリセットを各オーディオ型に割り当てるよう構成されてもよい。
【０２９７】
〈５．４節　コンテキスト型に基づく制御〉
これまでの節では、さまざまなコンテンツ型に焦点を当てている。本節で論じるさらなる実施形態では、代替的または追加的にコンテキスト型が考慮されてもよい。
【０２９８】
一般に、等化器は、明らかな音色変化のため映画的メディアにおけるダイアログをあまり良好に聞こえなくしてしまうことがあるので、音楽については有効にされるが、映画的メディアについては無効にされる。このことは、等化レベルは、長期的音楽の信頼値およ

び／または映画的メディアの信頼値に関係付けられうることを示す：

$$L_{eq} \propto Conf_{MUSIC} \quad (29)$$

または

$$L_{eq} \propto 1 - Conf_{MOVIE} \quad (30)$$

ここで、L_{eq}は等化レベル、$Conf_{MUSIC}$および$Conf_{MOVIE}$は長期的音楽および映画的メディアの信頼値を表わす。

【０２９９】

すなわち、調整ユニット３００Ｄは、等化レベルを長期的音楽の信頼値と正に相関させるまたは等化レベルを映画的メディアの信頼値と負に相関させるよう構成されていてもよい。

【０３００】

すなわち、映画的メディア信号については、映画的メディア信頼値は高く（または音楽信頼度は低い）、よって等化レベルは低い。他方、音楽信号については、映画的メディア信頼値は低く（または音楽信頼度は高い）、よって等化レベルは高い。

【０３０１】

公式(29)および(30)に示される解決策は、公式(22)ないし(25)と同様に修正されてもよく、および／または公式(22)ないし(25)に示される解決策の任意のものと組み合わされてもよい。

【０３０２】

追加的または代替的に、調整ニット３００Ｄは、等化レベルをゲームの信頼値と負に相関させるよう構成されていてもよい。

【０３０３】

もう一つの実施形態では、信頼値は、１．４節で論じた規格化された重みを導出するために使われてもよい。期待される等化レベル／プロファイルが各コンテキスト型についてあらかじめ定義されているとすると（等化プロファイルは下記の表２に示されている）、公式(9)と同様の公式適用されることもできる。

【０３０４】

【表２】

	映画的メディア	長期的音楽	VOIP	ゲーム
等化プロファイル	プロファイル1	プロファイル2	プロファイル3	プロファイル4

ここで、いくつかのプロファイルでは、映画的メディアおよびゲームのようなある種のコンテキスト型について等化器を無効にする方法として、すべての利得が0に設定されることができる。

【０３０５】

〈５．５節　実施形態の組み合わせおよび応用シナリオ〉

第一部と同様に、上記で論じたすべての実施形態およびその変形は、そのいかなる組み合わせにおいて実装されてもよく、異なる部／実施形態において言及されるが同じまたは同様の機能をもついかなる構成要素も同じまたは別個の構成要素として実装されてもよい。

【０３０６】

たとえば、５．１節ないし５．４節において述べた解決策の任意の二つ以上が互いと組み合わされてもよい。そして、これらの組み合わせの任意のものが、第一部～第四部および後述する他の部において記載または含意されている任意の実施形態とさらに組み合わされてもよい。

【０３０７】

〈５．６節　等化器制御方法〉

第一部と同様に、上記の実施形態における等化器コントローラを記述する過程で、いくつかのプロセスまたは方法も開示されていることは明らかである。以下では、これらの方法の概要が与えられるが、上記ですでに論じた詳細の一部は繰り返さない。

【０３０８】

まず、第一部で論じたオーディオ処理方法の実施形態が等化器について使用されてもよい。等化器のパラメータ（単数または複数）が、オーディオ処理方法によって調整されるべきターゲットの一つである。この観点から、オーディオ処理方法は等化器制御方法でもある。

【０３０９】

この節では、等化器の制御に固有の側面のみが論じられる。制御方法の一般的な側面については、第一部が参照されうる。

【０３１０】

諸実施形態によれば、等化器制御方法が、リアルタイムでオーディオ信号のオーディオ型を識別し、識別されたオーディオ型の信頼値に基づいて連続的な仕方で等化器を調整することを含みうる。

【０３１１】

本願の他の部分と同様に、対応する信頼値をもつ複数のオーディオ型が関与するとき、調整の動作１１０４は、複数のオーディオ型の重要性に基づいて複数のオーディオ型の信頼値に重み付けすることを通じて、または信頼値に基づいて複数のオーディオ型の効果に重み付けすることを通じて、複数のオーディオ型の少なくともいくつかを考慮するよう構成されてもよい。特に、調整動作１１０４は、信頼値に基づいて少なくとも一つの優勢なコンテンツ型を考慮するよう構成されていてもよい。

【０３１２】

第一部で述べたように、調整されたパラメータ値が平滑化されてもよい。１．５節および１．８節が参照され、ここでは詳細は記述は割愛する。

【０３１３】

オーディオ型はコンテンツ型またはコンテキスト型でありうる。コンテンツ型に関わるとき、調整動作１１０４は、等化レベルを短期的音楽の信頼値と正に相関させるおよび／または等化レベルを発話の信頼値と負に相関させるよう構成されていてもよい。追加的または代替的に、調整動作は、等化レベルを背景の信頼値と正に相関させるおよび／または等化レベルをノイズの信頼値と負に相関させるよう構成されていてもよい。

【０３１４】

コンテキスト型に関わるとき、調整動作１１０４は、等化レベルを長期的音楽の信頼値と正に相関させるおよび／または等化レベルを映画的メディアおよび／またはゲームの信頼値と負に相関させるよう構成されていてもよい。

【０３１５】

短期的音楽のコンテンツ型については、調整動作１１０４は、等化レベルを、優勢な源のない短期的音楽の信頼値と正に相関させるおよび／または等化レベルを、優勢な源をもつ短期的音楽の信頼値と負に相関させるよう構成されていてもよい。これは、短期的音楽についての信頼値がある閾値より大きいときにのみ行なわれることができる。

【０３１６】

等化レベルを調整することのほか、等化器の他の側面が、オーディオ信号のオーディオ型（単数または複数）の信頼値（単数または複数）に基づいて調整されてもよい。たとえば、調整動作１１０４は、等化レベルおよび／または等化プロファイルおよび／またはスペクトル・バランス・プリセットをそれぞれのオーディオ型に割り当てるよう構成されていてもよい。

【０３１７】

オーディオ型の具体例については第一部が参照されうる。

【０３１８】

オーディオ処理装置の実施形態と同様に、オーディオ処理方法の実施形態およびその変

形の任意の組み合わせが現実的である。他方、オーディオ処理方法の実施形態およびその変形のあらゆる側面は別個の解決策であってもよい。さらに、この節に記載される任意の二つ以上の解決策が互いと組み合わされてもよく、これらの組み合わせがさらに、本開示の他の部において記載または含意される任意の実施形態と組み合わされてもよい。

【０３１９】

〈第六部：オーディオ分類器および分類方法〉

１．１節および１．２節で述べたように、さまざまな階層レベルのコンテンツ型およびコンテキスト型を含む本願で論じられるオーディオ型は、機械学習ベースの方法を含め何らかの既存の分類方式を用いて分類または識別されることができる。この部および次の部では、本願は、これまでの部で言及されたコンテキスト型を分類するための分類器および方法のいくつかの新規な側面を提案する。

【０３２０】

〈６．１節　コンテンツ型分類に基づくコンテキスト分類器〉

これまでの部で述べたように、オーディオ分類器２００は、オーディオ信号のコンテンツ型を識別するおよび／またはオーディオ信号のコンテキスト型を識別するために使われる。したがって、オーディオ分類器２００は、オーディオ・コンテンツ分類器２０２および／またはオーディオ・コンテキスト分類器２０４を有していてもよい。オーディオ・コンテンツ分類器２０２および／またはオーディオ・コンテキスト分類器２０４を実装するための既存の技法を採用するとき、両分類器は互いから独立でありうるが、いくつかの特徴を共有していてもよく、よって該特徴を抽出するためのいくつかの方式を共有していてもよい。

【０３２１】

この部および次の第七部では、本願で提案される新規な側面に従って、オーディオ・コンテキスト分類器２０４は、オーディオ・コンテンツ分類器２０２の結果を利用してもよい。すなわち、オーディオ分類器２００は、オーディオ信号のコンテンツ型を識別するオーディオ・コンテンツ分類器２０２と；オーディオ・コンテンツ分類器２０２の結果に基づいてオーディオ信号のコンテキスト型を識別するオーディオ・コンテキスト分類器２０４とを有する。よって、オーディオ・コンテンツ分類器２０２の分類結果は、オーディオ・コンテキスト分類器２０４およびこれまでの部で論じた調整ユニット３００（または調整ユニット３００Ａないし３００Ｄ）の両方によって使用されてもよい。しかしながら、図面には示されていないものの、オーディオ分類器２００は、調整ユニット３００およびオーディオ・コンテキスト分類器２０４によってそれぞれ使用される二つのオーディオ・コンテンツ分類器２０２を含んでいてもよい。

【０３２２】

さらに、１．２節において論じられるように、特に複数のオーディオ型を分類するとき、オーディオ・コンテンツ分類器２０２またはオーディオ・コンテキスト分類器２０４は、互いと協働する分類器のグループをなしてもよい。ただし、一つの単一の分類器として実装されることも可能である。

【０３２３】

１．１節で論じたように、コンテンツ型は、一般に数フレームないし数十フレームのオーダーの長さ（たとえば1秒）をもつ短期的オーディオ・セグメントに関するオーディオ型の種類であり、コンテキスト型は、一般に数秒ないし数十秒のオーダーの長さ（たとえば10秒）をもつ長期的オーディオ・セグメントに関するオーディオ型の種類である。よって、「コンテンツ型」および「コンテキスト型」に対応して、必要なときはそれぞれ「短期」「長期」を使う。しかしながら、次の第七部で論じるように、コンテキスト型は比較的長い時間スケールでのオーディオ信号の属性を示すものであるものの、短期的オーディオ・セグメントから抽出される特徴に基づいて識別されることもできる。

【０３２４】

ここで、図２４を参照して、オーディオ・コンテンツ分類器２０２およびオーディオ・コンテキスト分類器２０４の構造に目を向ける。

【０３２５】
図２４に示されるように、オーディオ・コンテンツ分類器２０２は、それぞれオーディオ・フレームのシーケンスを含む短期的オーディオ・セグメントから短期的特徴を抽出する短期的特徴抽出器２０２２と；長期的オーディオ・セグメント中の短期的セグメントのシーケンスをそれぞれの短期的特徴を使って短期的オーディオ型に分類する短期的分類器２０２４とを有していてもよい。短期的特徴抽出器２０２２および短期的分類器２０２４の両方は、既存の技法を用いて実装されてもよいが、後述する６．３節において、短期的特徴抽出器２０２２についていくつかの修正も提案される。
【０３２６】
短期的分類器２０２４は、短期的セグメントのシーケンスの各セグメントを次の短期的オーディオ型（コンテンツ型）のうちの少なくとも一つに分類するよう構成されていてもよい：発話、短期的音楽、背景音およびノイズ。これについては１．１節で説明してある。各コンテンツ型はさらに、１．１節で論じたような、ただしそれに限定されないより低い階層レベルでのコンテンツ型にさらに分類されてもよい。
【０３２７】
当技術分野において知られているように、分類されたオーディオ型の信頼値は、短期的分類器２０２４によって得られてもよい。本願では、何らかの分類器の動作に言及するとき、明示的に記録されるか否かによらず、必要であれば同時に信頼値が得られることは理解される。オーディオ型分類の例は、非特許文献１に見出されうる。同文献はここにその全体において参照によって組み込まれる。
【０３２８】
他方、図２４に示されるように、オーディオ・コンテキスト分類器２０４は、長期的オーディオ・セグメント内の短期的セグメントのシーケンスに関して短期的分類器の結果の統計量を、長期的特徴として計算するための統計量抽出器２０２４と；長期的特徴を使って、長期的オーディオ・セグメントを長期的オーディオ型に分類する長期的分類器２０４４とを有していてもよい。同様に、統計量抽出器２０４２および長期的分類器２０４４の両方は、既存の技法を用いて実装されてもよいが、次の６．２節において統計量抽出器２０４２についていくつかの修正が提案される。
【０３２９】
長期的分類器２０４４は長期的オーディオ・セグメントを次の長期的オーディオ型（コンテキスト型）の少なくとも一つに分類するよう構成されていてもよい：映画的メディア、長期的音楽、ゲームおよびVoIP。これについては１．１節で説明してある。代替的または追加的に、長期的分類器２０４４は長期的オーディオ・セグメントを、１．１節で説明したVoIPまたは非VoIPに分類するよう構成されていてもよい。代替的または追加的に、長期的分類器２０４４は長期的オーディオ・セグメントを、１．１節で説明した高品質オーディオまたは低品質オーディオに分類するよう構成されていてもよい。実際上、用途／システムの必要性に基づいて、さまざまなターゲット・オーディオ型が選ばれることができ、トレーニングされることができる。
【０３３０】
短期的セグメントおよび長期的セグメント（ならびに６．３節で論じるフレーム）の意味および選択については、１．１節が参照されうる。
【０３３１】
〈６．２節　長期的特徴の抽出〉
図２４に示されるように、ある実施形態では、統計量抽出器２０４２のみが、短期的分類器２０２４の結果から長期的特徴を抽出するために使用される。長期的特徴として、次のうちの少なくとも一つが統計量抽出器２０４２によって計算されてもよい：分類されるべき長期的セグメント内の短期的セグメントの短期的オーディオ型の信頼値の平均および分散、短期的セグメントの重要度によって重み付けされた前記平均および分散、各短期的オーディオ型の出現頻度および分類されるべき長期的セグメント内の種々の短期的オーディオ型の間の遷移の頻度。

【０３３２】
図２５において、各短期的セグメント（長さ1s）における発話および短期的音楽信頼値の平均を示す。比較のために、セグメントは三つの異なるオーディオ・コンテキストから抽出されている：映画的メディア（図２５（Ａ））、長期的音楽（図２５（Ｂ））およびVoIP（図２５（Ｃ））。映画的メディア・コンテキストについては、発話型についてまたは音楽型について高い信頼値が得られ、これら二つのオーディオ型の間で頻繁に交替することが観察できる。対照的に、長期的音楽のセグメントは安定した高い短期的音楽の信頼値および比較的安定した低い発話信頼値を与える。一方、VoIPのセグメントは安定して低い短期的音楽の信頼値を与えるが、VoIP会話の間の休止のため揺動する発話信頼値を与える。

【０３３３】
各オーディオ型についての信頼値の分散も種々のオーディオ・コンテキストを分類するための重要な特徴である。図２６は、遺影が的メディア、長期的音楽およびVoIPオーディオ・コンテキストにおける発話、短期的音楽、背景およびノイズの信頼値の分散のヒストグラムを与えている（横軸はデータセット中の信頼値の分散であり、縦軸はデータセット中の分散値sの各ビンにおける生起数であり、これは分散値の各ビンの正規確率を示すよう規格化されることができる）。映画的メディアについては、発話、短期的音楽および背景の信頼値の分散のすべては比較的高く、幅広く分布している。これは、これらのオーディオ型の信頼値が強く変化していることを示す。長期的な音楽については、発話、短期的音楽、背景およびノイズの信頼値の分散のすべては比較的低く、狭く分布している。これは、これらのオーディオ型の信頼値が安定を保っていることを示す。発話信頼値は一定して低く保たれ、音楽信頼値は一定した高く保たれる。VoIPについては、短期的音楽の信頼値の分散は低く、狭く分布している一方、発話の信頼値の分散は比較的幅広く分布している。これは、VoIP会話の間の頻繁な休止のためである。

【０３３４】
重み付けされた平均および分散を計算する際に使われる重みについて、これらは各短期的セグメントの重要度に基づいて決定される。短期的セグメントの重要度は、そのエネルギーまたはラウドネスによって測定されてもよい。エネルギーやラウドネスは多くの既存の技法を用いて推定されることができる。

【０３３５】
分類されるべき長期的セグメントにおける各短期的オーディオ型の出現頻度は、該長期的セグメント内の短期的セグメントが分類された各オーディオ型のカウントを、長期的セグメントの長さで規格化したものである。

【０３３６】
分類されるべき長期的セグメント内の種々の短期的オーディオ型の間の遷移の頻度は、分類されるべき長期的セグメント内の隣り合う短期的セグメント間のオーディオ型変化のカウントを、長期的セグメントの長さで規格化したものである。

【０３３７】
図２５を参照して信頼値の平均および分散を論じるとき、各短期的オーディオ型の出現頻度およびそれら種々の短期型オーディオ型の間の遷移頻度も実際には触れられる。これらの特徴は、オーディオ・コンテキスト分類にも大きく関わってくる。たとえば、長期的音楽はほとんど短期的音楽オーディオ型を含み、よって短期的音楽の高い出現頻度をもつ。一方、VoIPはほとんと発話および休止を含み、よって発話またはノイズの高い出現頻度をもつ。もう一つの例として、映画的メディアは、長期的音楽またはVoIPよりも頻繁に異なる短期的オーディオ型の間で遷移し、よって一般に短期的音楽、発話および背景の間でより高い遷移頻度をもつ。VoIPは通例、他よりも発話とノイズの間でより頻繁に遷移し、よって発話とノイズの間のより高い遷移頻度をもつ。

【０３３８】
一般に、長期的セグメントは同じアプリケーション／システムでは同じ長さであると想定する。そうであれば、各短期的オーディオ型の出現カウントおよび長期的セグメント内

の異なる短期的オーディオ型の間の遷移カウントは規格化なしで直接使用されてもよい。長期的セグメントの長さが可変であれば、上述した出現頻度および遷移頻度が使われるべきである。本願の請求項は、両方の状況をカバーするものと解釈されるべきである。

【０３３９】

追加的または代替的に、オーディオ分類器２００（またはオーディオ・コンテキスト分類器２０４）はさらに、長期的オーディオ・セグメント内の短期的セグメントのシーケンスの短期的特徴に基づいて長期的オーディオ・セグメントからさらなる長期的特徴を抽出するための長期的特徴抽出器２０４６（図２７）を含んでいてもよい。換言すれば、長期的特徴抽出器２０４６は、短期的分類器２０２４の分類結果を使わず、短期的特徴抽出器２０２２によって抽出された短期的特徴を直接使って、長期的分類器２０４４によって使用されるべきいくつかの長期的特徴を導出する。長期的特徴抽出器２０４６および統計量抽出器２０４２は独立してまたは合同して使われてもよい。換言すれば、オーディオ分類器２００は、長期的特徴抽出器２０４６または統計量抽出器２０４２の一方または両方を含んでいてもよい。

【０３４０】

任意の特徴が、長期的特徴抽出器２０４６によって抽出されることができる。本願では、長期的特徴として、短期的特徴抽出器２０２２からの短期的特徴の次の統計量の少なくとも一つを計算することが提案される：平均、分散、重み付けされた平均、重み付けされた分散、高平均、低平均および高平均と低平均の間の比（コントラスト）。

【０３４１】

分類されるべき長期的セグメント内の短期的セグメントから抽出された短期的特徴の平均および分散。

【０３４２】

分類されるべき長期的セグメント内の短期的セグメントから抽出された短期的特徴の重み付けされた平均および分散。短期的特徴は、各短期的セグメントの、たった今述べたそのエネルギーまたはラウドネスを用いて測定される重要度に基づいて重み付けされる。

【０３４３】

高平均（high　average）：分類されるべき長期的セグメント内の短期的セグメントから抽出された、選択された短期的特徴の平均。短期的特徴は、次の条件のうちの少なくとも一つを満たすときに選択される：ある閾値より大きい；または他のすべての短期的特徴より低くない、短期的特徴のあらかじめ決定された割合以内、たとえば短期的特徴の上位10%以内。

【０３４４】

低平均（low　average）：分類されるべき長期的セグメント内の短期的セグメントから抽出された、選択された短期的特徴の平均。短期的特徴は、次の条件のうちの少なくとも一つを満たすときに選択される：ある閾値より小さい；または他のすべての短期的特徴より高くない、短期的特徴のあらかじめ決定された割合以内、たとえば短期的特徴の下位10%以内。

【０３４５】

コントラスト：高平均と低平均の間の比。長期的セグメント内の諸短期的特徴のダイナミックを表わす。

【０３４６】

短期的特徴抽出器２０２２は、既存の技法を用いて実装されてもよく、それによりいかなる特徴が抽出されることもできる。にもかかわらず、短期的特徴抽出器２０２２のためにいくつかの修正が次の６．３節で提案される。

【０３４７】

〈６．３節　短期的特徴の抽出〉

図２４および図２７に示されるように、短期的特徴抽出器２０２２は、短期的特徴として、次の特徴のうちの少なくとも一つを、各短期的オーディオ・セグメントから直接抽出するよう構成されていてもよい：リズム特性、中断／ミュート特性および短期的オーディ

オ品質特徴。
【０３４８】
リズム特性は、リズム強さ、リズム規則性、リズム明確性（ここにその全体において参照によって組み込まれる非特許文献２参照）および2Dサブバンド変調（ここにその全体において参照によって組み込まれる非特許文献３参照）を含みうる。
【０３４９】
中断／ミュート特性は、発話中断、シャープな減衰、ミュート長さ、不自然な無音、不自然な無音の平均、不自然な無音の全エネルギーを含みうる。
【０３５０】
短期的オーディオ品質特徴は、短期的セグメントに関するオーディオ品質特徴であり、これは下記で論じる、オーディオ・フレームから抽出されるオーディオ品質特徴と同様である。
【０３５１】
代替的または追加的に、図２８に示されるように、オーディオ分類器２００は、短期的セグメントに含まれるオーディオ・フレームのシーケンスの各フレームからフレーム・レベル特徴を抽出するフレーム・レベル特徴抽出器２０１２を有していてもよい。短期的特徴抽出器２０２２は、オーディオ・フレームのシーケンスから抽出されるフレーム・レベル特徴に基づいて短期的特徴を計算するよう構成されていてもよい。
【０３５２】
前処理として、入力オーディオ信号は、モノ・オーディオ信号にダウンミックスされてもよい。この前処理は、オーディオ信号がすでにモノ信号であれば不要である。次いで、あらかじめ定義された長さ（典型的には10ないし25ミリ秒）でフレームに分割される。対応して、各フレームからフレーム・レベル特徴が抽出される。
【０３５３】
フレーム・レベル特徴抽出器２０１２は、次の特徴のうちの少なくとも一つを抽出するよう構成されていてもよい：さまざまな短期的オーディオ型の属性を特徴付ける特徴、カットオフ周波数、静的な信号雑音比（SNR）特性、セグメントの信号雑音比（SNR）特性、基本的発話記述子および声道特性。
【０３５４】
さまざまな短期的オーディオ型（特に、発話、短期的音楽、背景音およびノイズ）の属性を特徴付ける特徴は、次の特徴のうちの少なくとも一つを含んでいてもよい：フレーム・エネルギー、サブバンド・スペクトル分布、スペクトル・フラックス（spectral flux）、メル‐周波数ケプストラム係数（MFCC: Mle-frequency Cepstral Coefficient）、ベース（bass）、残差情報（residual information）、クロマ（Chroma）特徴および零交差レート（zero-crossing rate）。
【０３５５】
MFCCの詳細については、ここに参照によってその全体において組み込まれる非特許文献１参照。クロマ特徴の詳細については、ここに参照によってその全体において組み込まれる非特許文献４参照。
【０３５６】
カットオフ周波数は、それより上ではコンテンツのエネルギーが0に近い、オーディオ信号の最高周波数を表わす。これは、帯域が限定されているコンテンツを検出するよう設計され、これは本願ではオーディオ・コンテキスト分類のために有用である。カットオフ周波数は通例、符号化によって引き起こされる。たいていの符号化器は、低ビットレートまたは中ビットレートでは高周波数を破棄するからである。たとえば、MP3コーデックは128kbpsにおいて16kHzのカットオフ周波数をもつ。もう一つの例として、多くの一般的なVoIPコーデックは8kHzまたは16kHzのカットオフ周波数をもつ。
【０３５７】
カットオフ周波数のほかに、オーディオ・エンコード・プロセスの間の信号劣化が、VoIP対非VoIPコンテキスト、高品質対低品質オーディオ・コンテキストのようなさまざまな

オーディオ・コンテキストを区別するためのさらなる特性として考慮される。よりリッチな特性を捕捉するために、客観的発話品質評価（ここに参照によってその全体において組み込まれる非特許文献５参照）のためのもののようなオーディオ品質を表わす特徴が、複数レベルにおいてさらに抽出されてもよい。オーディオ品質特徴の例は次のものを含む：
ａ）推定された背景ノイズ・レベル、スペクトル明瞭性などを含む静的なSNR特性
ｂ）スペクトル・レベル偏差、スペクトル・レベル範囲、相対ノイズ・フロアなどを含むセグメントSNR特性
ｃ）ピッチ平均、発話セクション・レベル変動、発話レベルなどを含む基本的発話記述子
ｄ）ロボット化（robotization）、ピッチ・クロス・パワー（pitch cross power）などを含む声道特性。

【０３５８】
フレーム・レベル特徴から短期的特徴を導出するために、短期的特徴抽出器２０２２は、フレーム・レベル特徴の統計量を、短期的特徴として計算するよう構成されていてもよい。

【０３５９】
フレーム・レベル特徴の統計量の例は、平均および標準偏差を含む。これは、短期的音楽、発話、背景およびノイズのようなさまざまなオーディオ型を区別するためのリズム属性を捕捉する。たとえば、発話は通例、有声音と無声音の間で音節レートで交替し、一方音楽はそのような交替がなく、発話のフレーム・レベル特徴の変動は通例、音楽より大きい。

【０３６０】
統計量のもう一つの例は、フレーム・レベル特徴の重み付けされた平均である。たとえば、カットオフ周波数について、短期的セグメント内の全オーディオ・フレームから導出されたカットオフ周波数の間の、各フレームのエネルギーまたはラウドネスを重みとして重み付けされた平均は、短期的セグメントについてのカットオフ周波数となる。

【０３６１】
代替的または追加的に、図２９に示されるように、オーディオ分類器２００は、オーディオ・フレームからフレーム・レベル特徴を抽出するフレーム・レベル特徴抽出器２０１２と、それぞれのフレーム・レベル特徴を使ってオーディオ・フレームのシーケンスの各フレームをフレーム・レベル・オーディオ型に分類するフレーム・レベル分類器２０１４とを有していてもよい。ここで、短期的特徴抽出器２０２２は、前記シーケンスのオーディオ・フレームに関するフレーム・レベル分類器２０１４の結果に基づいて短期的特徴を計算するよう構成されていてもよい。

【０３６２】
換言すれば、オーディオ・コンテンツ分類器２０２およびオーディオ・コンテキスト分類器２０４に加えて、オーディオ分類器２００はさらに、フレーム分類器２０１を有していてもよい。そのような構成では、オーディオ・コンテンツ分類器２０２は、フレーム分類器２０１のフレーム・レベルの分類結果に基づいて短期的セグメントを分類し、オーディオ・コンテキスト分類器２０４は、オーディオ・コンテンツ分類器２０２の短期的分類結果に基づいて長期的セグメントを分類する。

【０３６３】
フレーム・レベル分類器２０１４は、オーディオ・フレームのシーケンスの各フレームを何らかのクラスに分類するよう構成されていてもよい。そのクラスは、「フレーム・レベルのオーディオ型」と称されてもよい。
ある実施形態では、フレーム・レベルのオーディオ型は、上記で論じたコンテンツ型の構成と同様の構成を有していてもよく、コンテンツ型と同様の意味を有していてもよい。唯一の相違は、フレーム・レベルのオーディオ型とコンテンツ型はオーディオ信号の異なるレベルで、すなわちフレーム・レベルおよび短期的セグメント・レベルで分類されるということである。たとえば、フレーム・レベル分類器２０１４は、オーディオ・フレームのシーケンスの各フレームを次のフレーム・レベルのオーディオ型のうちの少なくとも一つ

に分類するよう構成されていてもよい：発話、音楽、背景音およびノイズ。他方、フレーム・レベルのオーディオ型は、部分的または完全にコンテンツ型の構成とは異なる、フレーム・レベルの分類により好適であり短期的分類のための短期的特徴として使われるのにより好適な構成を有していてもよい。たとえば、フレーム・レベル分類器２０１４は、オーディオ・フレームのシーケンスの各フレームを、次のフレーム・レベルのオーディオ型のうちの少なくとも一つに分類するよう構成されていてもよい：有声、無声および休止。

【０３６４】

フレーム・レベル分類の結果からいかにして短期的特徴を導出するかについて、６．２節の記述を参照することによって、同様の方式が採用されてもよい。

【０３６５】

代替として、フレーム・レベル分類器２０１４の結果に基づく短期的特徴およびフレーム・レベル特徴抽出器２０１２によって得られたフレーム・レベル特徴に直接基づく短期的特徴が短期的分類器２０２４によって使用されてもよい。したがって、短期的特徴抽出器２０２２は、オーディオ・フレームのシーケンスから抽出されたフレーム・レベル特徴およびオーディオ・フレームのシーケンスに関するフレーム・レベル分類器の結果の両方に基づく短期的特徴を計算するよう構成されていてもよい。

【０３６６】

換言すれば、フレーム・レベル特徴抽出器２０１２は、６．２節で論じたのと同様の統計量および図２８との関連で述べた、下記の特徴のうちの少なくとも一つを含む短期的特徴の両方を計算するよう構成されていてもよい：さまざまな短期的オーディオ型の属性を特徴付ける特徴、カットオフ周波数、静的な信号雑音比特性、セグメントの信号雑音比特性、基本的発話記述子および声道特性。

【０３６７】

リアルタイムで作業するために、すべての実施形態において、短期的特徴抽出器２０２２は、所定のステップ長さで長期的オーディオ・セグメントの時間次元内をスライドする移動窓を用いて形成される短期的オーディオ・セグメントに対して作用するよう構成されてもよい。短期的オーディオ・セグメントについての移動窓および長期的オーディオ・セグメントについての移動窓について、詳細は１．１節が参照されうる。

【０３６８】

〈６．４節　実施形態の組み合わせおよび応用シナリオ〉

第一部と同様に、上記で論じたすべての実施形態およびその変形は、そのいかなる組み合わせにおいて実装されてもよく、異なる部／実施形態において言及されるが同じまたは同様の機能をもついかなる構成要素も同じまたは別個の構成要素として実装されてもよい。

【０３６９】

たとえば、６．１節ないし６．３節において述べた解決策の任意の二つ以上が互いと組み合わされてもよい。そして、これらの組み合わせの任意のものが、第一部～第五部および後述する他の部において記載または含意されている任意の実施形態とさらに組み合わされてもよい。特に、第一部で論じた型平滑化ユニット７１２がオーディオ分類器２００のコンポーネントとしてこの部において、フレーム分類器２０１４またはオーディオ・コンテンツ分類器２０２またはオーディオ・コンテキスト分類器２０４の結果を平滑化するために使用されてもよい。さらに、オーディオ分類器２００の出力の急激な変化を避けるために、オーディオ分類器２００のコンポーネントとしてタイマー９１６も役割を果たしてもよい。

【０３７０】

〈６．５節　オーディオ分類方法〉

第一部と同様に、上記の実施形態におけるオーディオ分類器を記述する過程で、いくつかのプロセスまたは方法も開示されていることは明らかである。以下では、これらの方法の概要が与えられるが、上記ですでに論じた詳細の一部は繰り返さない。

【０３７１】

図３０に示されるある実施形態では、オーディオ分類方法が提供される。（互いに重なり合うまたは重なり合わない）短期的オーディオ・セグメントのシーケンスからなる長期的オーディオ・セグメントの長期的オーディオ型（つまりコンテンツ型）を識別するために、短期的オーディオ・セグメントがまず短期的オーディオ型に、つまりコンテンツ型に分類され（動作３００４）、長期的オーディオ・セグメント内の短期的セグメントのシーケンスに関する分類動作の結果の統計量を計算する（動作３００６）ことによって長期的特徴が得られる。次いで、朝敵特徴を使って長期的分類（動作３００８）が実行されてもよい。短期的オーディオ・セグメントは、オーディオ・フレームのシーケンスを含んでいてもよい。むろん、短期的セグメントの短期的オーディオ型を識別するために、それらのセグメントから短期的特徴が抽出される必要がある。

【０３７２】

短期的オーディオ型（コンテンツ型）は、発話、短期的音楽、背景音およびノイズを含んでいてもよいが、それに限られない。

【０３７３】

長期的特徴は、それらの短期的セグメントの信頼値の平均および分散、それらの短期的セグメントの重要度によって重み付けされた前記平均および分散、各短期的オーディオ型の出現頻度および異なる短期的オーディオ型の間の遷移の頻度。

【０３７４】

ある平均では、図３１に示されるように、さらなる長期的特徴が、長期的オーディオ・セグメント内の短期的セグメントのシーケンスの短期的特徴に直接基づいて得られてもよい（動作３１０７）。そのような長期的特徴は、短期的特徴の次の統計量を含んでいてもよいが、それに限られない：平均、分散、重み付けされた平均、重み付けされた分散、高平均、低平均および高平均と低平均の間の比。

【０３７５】

短期的特徴を抽出するためには種々の方法がある。一つは、分類されるべき短期的オーディオ・セグメントから短期的特徴を直接抽出することである。そのような特徴は、リズム特性、中断／ミュート特性および短期的オーディオ品質特徴を含むがそれに限られない。

【０３７６】

第二の方法は、各短期的セグメントに含まれるオーディオ・フレームからフレーム・レベルの特徴を抽出して（図３２の動作３２０１）、次いでフレーム・レベルの特徴に基づいて短期的特徴を計算する、たとえば短期的特徴としてフレーム・レベルの特徴の統計量を計算することである。フレーム・レベル特徴は：さまざまな短期的オーディオ型の属性を特徴付ける特徴、カットオフ周波数、静的な信号雑音比特性、セグメントの信号雑音比特性、基本的発話記述子および声道特性を含んでいてもよいがそれに限られない。さまざまな短期的オーディオ型の属性を特徴付ける特徴はさらに、フレーム・エネルギー、サブバンド・スペクトル分布、スペクトル・フラックス（spectral flux）、メル‐周波数ケプストラム係数（MFCC: Mle-frequency Cepstral Coefficient）、ベース（bass）、残差情報（residual information）、クロマ（Chroma）特徴および零交差レート（zero-crossing rate）を含んでいてもよい。

【０３７７】

第三の方法は、長期的特徴の抽出と同様の仕方で短期的特徴を抽出することである：分類されるべき短期的セグメント内のオーディオ・フレームからフレーム・レベル特徴を抽出した（動作３２０１）後、それぞれのフレーム・レベル特徴を使って各オーディオ・フレームをフレーム・レベルのオーディオ型に分類する（図３３における動作３２０１１）；短期的特徴は、フレーム・レベル・オーディオ型（任意的に信頼値を含む）に基づいて短期的特徴を計算することによって抽出されうる（動作３００２）。フレーム・レベル・オーディオ型は、短期的オーディオ型（コンテンツ型）と同様の属性および構成を有していてもよく、発話、音楽、背景音およびノイズをも含んでいてもよい。

【０３７８】

上記第二の方法および第三の方法は、図３３において破線矢印で示されるように一緒に組み合わされてもよい。

【０３７９】

第一部で論じたように、短期的オーディオ・セグメントおよび長期的オーディオ・セグメントの両方が移動窓を用いてサンプリングされてもよい。すなわち、短期的特徴を抽出する動作（動作３００２）は、所定のステップ長さで長期的オーディオ・セグメントの時間次元においてスライドする移動窓を用いて形成される短期的オーディオ・セグメントに対して実行されてもよく、長期的特徴を抽出する動作（動作３１０７）および短期的オーディオ型の統計量を計算する動作（動作３００６）は、所定のステップ長さでオーディオ信号の時間次元においてスライドする移動窓を用いて形成される長期的オーディオ・セグメントに対して実行されてもよい。

【０３８０】

オーディオ処理装置の実施形態と同様に、オーディオ処理方法の実施形態およびその変形の任意の組み合わせが現実的である。他方、オーディオ処理方法の実施形態およびその変形のあらゆる側面は別個の解決策であってもよい。さらに、この節に記載される任意の二つ以上の解決策が互いと組み合わされてもよく、これらの組み合わせがさらに、本開示の他の部において記載または含意される任意の実施形態と組み合わされてもよい。特に、６．４節においてすでに論じたように、オーディオ型の平滑化方式および遷移方式がここで論じたオーディオ分類方法の一部であってもよい。

【０３８１】

〈第七部：VoIP分類器および分類方法〉

第六部では、少なくとも部分的にはコンテンツ型分類の結果に基づいてオーディオ信号をオーディオ・コンテキスト型に分類するための新規なオーディオ分類器が提案されている。第六部で論じた実施形態において、長期的特徴は、数秒ないし数十秒の長さの長期的セグメントから抽出される。よって、オーディオ・コンテキスト分類は長いレイテンシーを引き起こしうる。オーディオ・コンテキストがリアルタイムでまたはほぼリアルタイムでたとえば短期的セグメント・レベルにおいて分類されうることが望まれる。

【０３８２】

〈７．１節　短期的セグメントに基づくコンテキスト分類〉

したがって、図３４に示されるように、オーディオ信号の短期的セグメントのコンテンツ型を同定するためのオーディオ・コンテンツ分類器２０２Ａと、少なくとも部分的には前記オーディオ・コンテンツ分類器によって識別されたコンテンツ型に基づいて短期的セグメントのコンテキスト型を識別するオーディオ・コンテキスト分類器２０４Ａとを有するオーディオ分類器２００Ａが提供される。

【０３８３】

ここで、オーディオ・コンテンツ分類器２０２Ａは、第六部ですでに述べた技法を採用してもよいが、下記の７．２節で論じる種々の技法を採用してもよい。また、オーディオ・コンテキスト分類器２０４Ａは、第六部ですでに述べた技法を採用してもよいが、違いとして、コンテキスト分類器２０４Ａは、オーディオ・コンテンツ分類器２０２Ａからの結果の統計量を使うのではなく、オーディオ・コンテンツ分類器２０２Ａの結果を直接使ってもよい。オーディオ・コンテキスト分類器２０４およびオーディオ・コンテンツ分類器２０２Ａはいずれも同じ短期的セグメントを分類しているからである。さらに、第六部と同様に、オーディオ・コンテンツ分類器２０２Ａからの結果に加えて、オーディオ・コンテキスト分類器２０４は、短期的セグメントから直接抽出された他の特徴を使ってもよい。すなわち、オーディオ・コンテキスト分類器２０４Ａは、特徴として、短期的セグメントのコンテンツ型の信頼値および短期的セグメントから抽出された他の特徴を使って、機械学習モデルに基づいて短期的セグメントを分類するよう構成されてもよい。短期的セグメントから抽出される特徴については、第六部が参照されうる。

【０３８４】

オーディオ・コンテンツ分類器２００Ａは、同時に、短期的セグメントを、VoIP発話／

ノイズおよび／または非VoIP発話／ノイズよりも多くのオーディオ型としてラベル付けしてもよく（VoIP発話／ノイズおよび／または非VoIP発話／ノイズについては下記の７．２節で論じる）、複数のオーディオ型のそれぞれは、１．２節において論じたその独自の信頼値をもちうる。これは、よりリッチな情報が補足できるので、よりよい分類精度を達成できる。たとえば、発話および短期的音楽の信頼値の合同情報は、どの程度そのオーディオ・コンテンツが発話および背景音楽の混合でありそうかを明らかにし、それにより純粋なVoIPコンテンツから弁別されることができる。

【０３８５】

〈７．２節　VoIP発話およびVoIPノイズを使った分類〉

本願のこの側面は、短い判断レイテンシーのために現在の短期的セグメントを分類することを要求されるVoIP／非VoIP分類システムにおいて特に有用である。

【０３８６】

この目的のために、図３４に示されるように、オーディオ分類器２００Ａは、VoIP／非VoIP分類のために特別に設計される。VoIP／非VoIPを分類するために、オーディオ・コンテキスト分類器２０４による最終的な堅牢なVoIP／非VoIP分類のための中間結果を生成するようVoIP発話分類器２０２６および／またはVoIPノイズ分類器が開発される。

【０３８７】

VoIP短期的セグメントはVoIP発話およびVoIPノイズを交互に含むであろう。発話の短期的セグメントをVoIP発話または非VoIP発話に分類するためには高い精度が達成できるが、ノイズの短期的セグメントをVoIPノイズまたは非VoIPノイズに分類するためにはそうではないことが観察される。このように、発話とノイズの間の差を考慮せず、よってこれら二つのコンテンツ型（発話およびノイズ）の特徴が一緒に混ざったままで短期的セグメントをVoIP（VoIP発話およびVoIPノイズを含むが、VoIP発話およびVoIPノイズは個々に同定されていない）および非VoIPに直接分類することによって、弁別性がぼかされていると結論できる。

【０３８８】

分類器にとって、VoIPノイズ／非VoIPノイズ分類よりもVoIP発話／非VoIP発話分類についてより高い精度を達成することは理にかなっている。発話がノイズより多くの情報を含み、カットオフ周波数のような特徴は発話を分類するためにより効果的だからである。アダブースト・トレーニング／プロセスから得られる重みランク付けに従って、VoIP／非VoIP分類についての上位の重みを付けられた諸短期的特徴は：対数エネルギーの標準偏差、カットオフ周波数、リズム強さの標準偏差およびスペクトル・フラックスの標準偏差である。対数エネルギーの標準偏差、リズム強さの標準偏差およびスペクトル・フラックスの標準偏差は一般に、非VoIP発話についてよりVoIP発話について高くなる。一つのありそうな理由は、映画的メディアまたはゲームのような非VoIPコンテキストにおける多くの短期的発話セグメントは通例、背景音楽または効果音のような他の音と混ざっており、該背景音楽または効果音については上記の特徴の値はより低いということである。一方、カットオフ特徴は一般に、非VoIP発話についてよりVoIP発話について低くなる。このことは、多くの一般的なVoIPコーデックによって導入される低いカットオフ周波数を示す。

【０３８９】

したがって、ある実施形態では、オーディオ・コンテンツ分類器２０２Ａは、短期的セグメントをコンテンツ型VoIP発話またはコンテンツ型非VoIP発話に分類するVoIP発話分類器２０２６を有していてもよく、オーディオ・コンテキスト分類器２０４は、VoIP発話および非VoIP発話の信頼値に基づいて短期的セグメントをコンテキスト型VoIPまたはコンテキスト型非VoIPに分類するよう構成されていてもよい。

【０３９０】

もう一つの実施形態では、オーディオ・コンテンツ分類器２０２Ａはさらに、短期的セグメントをコンテンツ型VoIPノイズまたはコンテンツ型非VoIPノイズに分類するVoIPノイズ分類器２０２８を有していてもよく、オーディオ・コンテキスト分類器２０４は、VoIP発話、非VoIP発話、VoIPノイズおよび非VoIPノイズの信頼値に基づいて短期的セグメント

をコンテキスト型VoIPまたはコンテキスト型非VoIPに分類するよう構成されていてもよい。

【０３９１】

VoIP発話、非VoIP発話、VoIPノイズおよび非VoIPノイズのコンテンツ型は、第六部、１．２節および７．１節で論じた既存の技法を用いて同定されてもよい。

【０３９２】

あるいはまた、オーディオ・コンテンツ分類器２０２Ａはさらに図３５に示される階層構造を有していてもよい。すなわち、発話／ノイズ分類器２０２５からの結果を活用してまず短期的セグメントを発話またはノイズ／背景に分類するのである。

【０３９３】

単にVoIP発話分類器２０２６を使う実施形態に基づいて、短期的セグメントが発話／ノイズ分類器２０２５（このような状況ではこれは単に発話分類器である）によって発話と判定される場合、VoIP発話分類器２０２６は、それがVoIP発話または非VoIP発話のいずれであるかを分類することに進み、二値分類結果を計算する。それ以外の場合には、VoIP発話の信頼値が低い、あるいはVoIP発話についての決定が不確かであると見なされてもよい。

【０３９４】

単にVoIPノイズ分類器２０２８を使う実施形態に基づいて、短期的セグメントが発話／ノイズ分類器２０２５（このような状況ではこれは単にノイズ（背景）分類器である）によってノイズと判定される場合、VoIPノイズ分類器２０２８は、それをVoIPノイズまたは非VoIPノイズに分類し、二値分類結果を計算することに進む。それ以外の場合には、VoIPノイズの信頼値が低い、あるいはVoIPノイズについての決定が不確かであると見なされてもよい。

【０３９５】

ここで、一般に発話は情報性のコンテンツ型であり、ノイズ／背景は干渉性のコンテンツ型であるので、たとえ短期的セグメントがノイズでなくても、前段の実施形態において、短期的セグメントがコンテキスト型VoIPではないと確定的に判定することはできない。短期的セグメントが発話でなければ、単にVoIP発話分類器２０２６を使う実施形態では、それはおそらくコンテキスト型VoIPではないであろう。したがって、一般に、担任VoIP発話分類器２０２６を使う実施形態は独立して実現されうる。一方、単位VoIPノイズ分類器２０２８を使う他方の実施形態は、たとえばVoIP発話分類器２０２６を使う実施形態と協働する補足的実施形態として使用されうる。

【０３９６】

すなわち、VoIP発話分類器２０２６およびVoIPノイズ分類器２０２８の両方が使用されてもよい。短期的セグメントが発話／ノイズ分類器２０２５によって発話と判定される場合、VoIP発話分類器２０２６はそれがVoIP発話または非VoIP発話のいずれであるかを分類することに進み、二値分類結果を計算する。短期的セグメントが発話／ノイズ分類器２０２５によってノイズと判定される場合、VoIPノイズ分類器２０２８はそれをVoIPノイズまたは非VoIPノイズに分類し、二値分類結果を計算することに進む。それ以外の場合には、短期的セグメントが非VoIPと分類されうると見なされてもよい。

【０３９７】

発話／ノイズ分類器２０２５、VoIP発話分類器２０２６およびVoIPノイズ分類器２０２８の実装は、いかなる既存の技法を採用してもよく、第一部ないし第六部で論じたオーディオ・コンテンツ分類器２０２であってもよい。

【０３９８】

上記に従って実装されたオーディオ・コンテンツ分類器２０２Ａが最終的に短期的セグメントを発話、ノイズおよび背景のいずれにも、あるいはVoIP発話、非VoIP発話、VoIPノイズおよび非VoIPノイズのいずれにも分類しない場合、つまり、すべての関連する信頼値が低い場合、オーディオ・コンテンツ分類器２０２Ａ（およびオーディオ・コンテキスト分類器２０４）はその短期的セグメントを非VoIPと分類してもよい。

【０３９９】
短期的セグメントを、VoIP発話分類器２０２６およびVoIPノイズ分類器２０２８の結果に基づいてVoIPまたは非VoIPのコンテキスト型に分類するために、オーディオ・コンテキスト分類器２０４は、７．１節で論じた機械学習ベースの技法を採用してもよく、修正として、７．１節ですでに論じたような、短期的セグメントから直接抽出された短期的特徴および／またはVoIP関係のコンテンツ型以外のコンテンツ型に向けられた他のオーディオ・コンテンツ分類器（単数または複数）の結果を含むより多くの特徴が使用されてもよい。
【０４００】
上記の機械学習ベースの技法のほかに、VoIP／非VoIP分類への代替的なアプローチは、ドメイン知識を活用し、分類結果をVoIP発話およびVoIPノイズとの関連で利用するヒューリスティック規則であることができる。
【０４０１】
時刻tの現在の短期的セグメントがVoIP発話または非VoIP発話として決定される場合、分類結果はVoIP／非VoIP分類結果として直接取られる。VoIP／非VoIP発話分類は先に論じたように堅牢だからである。すなわち、短期的セグメントがVoIP発話であると判定される場合は、それはコンテキスト型VoIPである。短期的セグメントが非VoIP発話であると判定される場合は、それはコンテキスト型非VoIPである。
【０４０２】
VoIP発話分類器２０２６が先述した発話／ノイズ分類器２０２５によって判定される発話に関してVoIP発話／非VoIP発話に関する二値判定をするとき、VoIP発話および非VoIP発話の信頼値は相補的でありうる。すなわち、それらの和は1である（0が100%否を表わし、1が100%肯定を表わす場合）。そしてVoIP発話と非VoIP発話を区別するための信頼値の閾値は実際に同じ点を指しうる。VoIP発話分類器２０２６が二値分類器でない場合には、VoIP発話および非VoIP発話の信頼値は相補的ではなく、VoIP発話と非VoIP発話を区別するための信頼値の閾値は必ずしも同じ点を指さないことがある。
【０４０３】
しかしながら、VoIP発話または非VoIP発話信頼度が閾値に近く、該閾値のまわりで揺動する場合、VoIP／非VoIP分類結果は、頻繁に切り替わりすぎることがありうる。そのような揺動を避けるために、バッファ方式が提供されてもよい。VoIP発話および非VoIP発話についての両方の閾値がより大きく設定されてもよく、よって現在のコンテンツ型から他方のコンテンツ型に切り替わるのはそれほど容易ではなくなる。記述の簡単のため、非VoIP発話についての信頼値をVoIP発話の信頼値に変換してもよい。すなわち、信頼値が高ければ、短期的セグメントはVoIP発話により近いと見なされ、信頼値が低ければ、短期的セグメントは非VoIP発話により近いと見なされる。上記のような非二値分類器については、非VoIP発話についての高い信頼値は必ずしもVoIP発話の低い信頼値を意味しないが、そのような単純化は、本解決策の本質をよく反映でき、二値分類器の言辞を用いて記述される関連する請求項は、非二値分類器についての等価な解決策をカバーすると解釈される。
【０４０４】
バッファ方式は図３６に示されている。二つの閾値Th1およびTh2（Th1≧Th2）の間にバッファ領域がある。VoIP発話の信頼値v(t)がこの領域において低下するとき、コンテキスト分類は変化しない。これは図３６における左側および右側の矢印によって示されている。信頼値v(t)が大きいほうの閾値Th1より大きいときにのみ、短期的セグメントはVoIPと分類され（図３６の下部の矢印によって示されるように）、信頼値が小さいほうの閾値Th2より大きくないときにのみ、短期的セグメントは非VoIPと分類されることになる（図３６の上部の矢印によって示されるように）。
【０４０５】
VoIPノイズ分類器２０２８が代わりに使われる場合、状況は同様である。解決策をより堅牢にするために、VoIP発話分類器２０２６およびVoIPノイズ分類器２０２８が合同して使用されてもよい。次いで、オーディオ・コンテキスト分類器２０４Ａは：VoIP発話の信

頼値が第一の閾値より大きい場合またはVoIPノイズの信頼値が第三の閾値より大きい場合、短期的セグメントをコンテキスト型VoIPとして分類し；VoIP発話の信頼値が、前記第一の閾値より大きくない第二の閾値より大きくない場合またはVoIPノイズの信頼値が前記第三の閾値より大きくない第四の閾値より大きくない場合、短期的セグメントをコンテキスト型非VoIPとして分類し；それ以外の場合には短期的セグメントを最後の短期的セグメントについてのコンテキスト型として分類するよう構成されていてもよい。

【０４０６】

ここで、第一の閾値は、第二の閾値に等しくてもよく、第三の閾値は第四の閾値に等しくてもよい。これは特に二値VoIP発話分類器および二値VoIPノイズ分類器についてそうであるが、それに限られない。しかしながら、一般に、VoIPノイズ分類結果はそれほど堅牢ではないので、第三および第四の閾値は互いに等しくないほうがよいであろう。両者は0.5から遠いべきである（0は非VoIPノイズである高い信頼度を示し、1はVoIPノイズである高い信頼度を示す）。

【０４０７】

〈７．３節　平滑化ゆらぎ〉

急速な揺動〔ゆらぎ〕を避けるために、もう一つの解決策は、オーディオ・コンテンツ分類器によって決定される信頼値を平滑化することである。したがって、図３７に示されるように、型平滑化ユニット２０３Ａがオーディオ分類器２００Ａに含まれてもよい。先に論じた四つのVoIP関係のコンテンツ型のそれぞれの信頼値について、１．３節で論じた平滑化方式が採用されてもよい。

【０４０８】

あるいはまた、７．２節と同様に、VoIP発話および非VoIP発話は、相補的な信頼値を有する対と見なされてもよい。VoIPノイズおよび非VoIPノイズも相補的な信頼値をもつ対と見なされてもよい。そのような状況では、各対のうち一つのみが平滑化される必要があり、１．３節で論じた平滑化方式が採用されてもよい。

【０４０９】

VoIP発話の信頼値を例に取ると、公式(3)が次のように書き直されてもよい。

【０４１０】

$$v(t)=\beta\cdot v(t-1)+(1-\beta)\cdot voipSpeechConf(t) \qquad (3")$$

ここで、v(t)は時刻tにおける平滑化されたVoIP発話信頼値であり、v(t－1)は最後の時点における平滑化されたVoIP発話信頼値であり、voipSpeechConfは平滑化前の現在時刻tにおけるVoIP発話信頼度であり、αは重み付け係数である。

【０４１１】

ある変形では、上記のように発話／ノイズ分類器２０２５がある場合、短期的セグメントについての発話の信頼値が低ければ、その短期的セグメントはVoIP発話として堅牢に分類されることはできず、VoIP発話分類器２０２６を実際に機能させることなく、voipSeechConf(t)＝v(t－1)と直接設定することができる。

【０４１２】

あるいはまた、上記の状況において、不確かなケースを示してvoipSpeechConf(t)＝0.5（または0.4～0.5のような0.5より高くない他の値）と設定することができる（ここで、信頼度＝1がVoIPである高い信頼性を示し、信頼度＝0がVoIPではない高い信頼性を示す）。

【０４１３】

したがって、この変形によれば、図３７に示されるように、オーディオ・コンテンツ分類器２００Ａはさらに、短期的セグメントの発話のコンテンツ型を識別するための発話／ノイズ分類器２０２５を有していてもよく、型平滑化ユニット２０３Ａは、平滑化前の現在の短期的セグメントについてのVoIP発話の信頼値を、（0.5または0.4～0.5などの他の値のような）所定の信頼値として、あるいは発話／ノイズ分類器によって分類されるコンテンツ型発話についての信頼値が第五の閾値より低い最後の短期的セグメントの平滑化された信頼値として、設定するよう構成されていてもよい。そのような状況では、VoIP発話

分類器２０２６は作用してもしなくてもよい。あるいはまた、信頼値の設定は、VoIP発話分類器２０２６によって行なわれてもよい。これは、この作用が型平滑化ユニット２０３Ａによってなされる解決策と等価であり、請求項は両方の状況をカバーするものと解釈されるべきである。さらに、ここで、「発話／ノイズ分類器によって分類されるコンテンツ型発話についての信頼値が第五の閾値より低い」という言辞を使っているが、保護範囲はそれに限定されず、短期的セグメントが発話以外のコンテンツ型に分類される状況に等価である。

【０４１４】

VoIPノイズの信頼値について、状況は同様であり、詳細な説明はここでは割愛する。

【０４１５】

急激なゆらぎを避けるために、さらにもう一つの解決策は、オーディオ・コンテキスト分類器２０４Ａによって決定される信頼値を平滑化するというものであり、１．３節で論じた平滑化方式が採用されてもよい。

【０４１６】

急激なゆらぎを避けるために、さらにもう一つの解決策は、VoIPと非VoIPの間のコンテキスト型の遷移を遅らせるというものであり、１．６節で述べたのと同じ方式が使われてもよい。１．６節で述べたように、タイマー９１６はオーディオ分類器の外部であってもよいし、あるいはオーディオ分類器の一部としてその内部であってもよい。したがって、図３８に示されるように、オーディオ分類器２００Ａはさらにタイマー９１６を有していてもよい。そして、オーディオ分類器は、新しいコンテキスト型の持続時間の長さが第六の閾値に達するまで現在のコンテキスト型を出力し続けるよう構成される（コンテキスト型はオーディオ型のインスタンスである）。１．６節を参照することにより、詳細な説明はここでは割愛しうる。

【０４１７】

追加的または代替的に、VoIPと非VoIPの間の遷移を遅らせるもう一つの方式として、VoIP／非VoIP分類について先述した第一および／または第二の閾値が、最後の短期的セグメントのコンテキスト型に依存して異なっていてもよい。すなわち、新しい短期的セグメントのコンテキスト型が最後の短期的セグメントのコンテキスト型と異なるときは、第一および／または第二の閾値はより大きくなり、一方、新しい短期的セグメントのコンテキスト型が最後の短期的セグメントのコンテキスト型と同じときはより小さくなる。このようにして、コンテキスト型は現在のコンテキスト型に維持される傾向があり、よってコンテキスト型の急激なゆらぎがある程度抑制されうる。

【０４１８】

〈７．４節　実施形態の組み合わせおよび応用シナリオ〉

第一部と同様に、上記で論じたすべての実施形態およびその変形は、そのいかなる組み合わせにおいて実装されてもよく、異なる部／実施形態において言及されるが同じまたは同様の機能をもついかなる構成要素も同じまたは別個の構成要素として実装されてもよい。

【０４１９】

たとえば、７．１節ないし７．３節において述べた解決策の任意の二つ以上が互いと組み合わされてもよい。そして、これらの組み合わせの任意のものが、第一部～第六部において記載または含意されている任意の実施形態とさらに組み合わされてもよい。特に、この部で論じた諸実施形態およびそれらの任意の組み合わせは、第四部で論じたオーディオ処理装置／方法またはボリューム平準化器／制御方法の実施形態と組み合わされてもよい。

【０４２０】

〈７．５節　VoIP分類方法〉

第一部と同様に、上記の実施形態におけるオーディオ分類器を記述する過程で、いくつかのプロセスまたは方法も開示されていることは明らかである。以下では、これらの方法の概要が与えられるが、上記ですでに論じた詳細の一部は繰り返さない。

【０４２１】
図３９に示されるある実施形態では、オーディオ分類方法が、オーディオ信号の短期的セグメントのコンテンツ型を識別し（動作４００４）、次いで、少なくとも識別されたコンテンツ型に基づいて短期的セグメントのコンテキスト型を識別する（動作４００８）ことを含む。
【０４２２】
オーディオ信号のコンテキスト型を動的にかつ高速に識別するために、この部におけるオーディオ分類方法は、コンテキスト型VoIPおよび非VoIPを識別する際に特に有用である。そのような状況では、短期的セグメントはまずコンテンツ型VoIP発話またはコンテンツ型非VoIP発話に分類されてもよく、コンテキスト型を識別する上記動作は、VoIP発話および非VoIP発話の信頼値に基づいて、短期的セグメントをコンテキスト型VoIPまたはコンテキスト型非VoIPに分類するよう構成される。
【０４２３】
あるいはまた、短期的セグメントはまずコンテンツ型VoIPノイズまたはコンテンツ型非VoIPノイズに分類されてもよく、コンテキスト型を識別する上記動作は、VoIPノイズおよび非VoIPノイズの信頼値に基づいて、短期的セグメントをコンテキスト型VoIPまたはコンテキスト型非VoIPに分類するよう構成されてもよい。
【０４２４】
発話およびノイズは合同して考えられてもよい。そのような状況では、コンテキスト型を識別する上記動作は、VoIP発話、非VoIP発話、VoIPノイズおよび非VoIPノイズの信頼値に基づいて、短期的セグメントをコンテキスト型VoIPまたはコンテキスト型非VoIPに分類するよう構成されてもよい。
【０４２５】
短期的セグメントのコンテキスト型を識別するために、短期的セグメントのコンテンツ型の信頼値および短期的セグメントから抽出された他の特徴の両方を特徴として取り入れて、機械学習モデルが使われてもよい。
【０４２６】
コンテキスト型を識別する上記動作は、ヒューリスティック規則に基づいて実現されてもよい。VoIP発話および非VoIP発話のみが関わるときは、ヒューリスティック規則は次のようなものである：VoIP発話の信頼値が第一の閾値より大きければ短期的セグメントをコンテキスト型VoIPとして分類し；VoIP発話の信頼値が第二の閾値より大きくなければ短期的セグメントをコンテキスト型非VoIPとして分類し、ここで、第二の閾値は第一の閾値より大きくなく；それ以外の場合には、短期的セグメントを最後の短期的セグメントについてのコンテキスト型として分類する。
【０４２７】
VoIPノイズおよび非VoIPノイズのみが関わる状況についてのヒューリスティック規則も同様である。
【０４２８】
発話およびノイズの両方が関わるときは、ヒューリスティック規則は次のようなものである：VoIP発話の信頼値が第一の閾値より大きい場合またはVoIPノイズの信頼値が第三の閾値より大きい場合には短期的セグメントをコンテキスト型VoIPとして分類し；VoIP発話の信頼値が、第一の閾値より大きくない第二の閾値より大きくない場合またはVoIPノイズの信頼値が、第三の閾値より大きくない第四の閾値より大きくない場合には、短期的セグメントをコンテキスト型非VoIPとして分類し；それ以外の場合には、短期的セグメントを最後の短期的セグメントについてのコンテキスト型として分類する。
【０４２９】
１．３節および１．８節で論じた平滑化方式がここで採用されてもよく、詳細な説明は割愛する。１．３節で述べた平滑化方式への修正として、平滑化動作４１０６の前に、本方法はさらに、短期的セグメントからのコンテンツ型発話を識別する段階（図４０の動作４００４０）を含んでいてもよい。ここで、コンテンツ型発話についての信頼値は第五の

閾値より低い（動作４００４１における「No」）場合には、平滑化前の現在の短期的セグメントについてのVoIP発話の信頼値は、所定の信頼値または最後の短期的セグメントの平滑化された信頼値として設定される（図４０の動作４００４４）。

【０４３０】

そうではなく、コンテンツ型発話を識別する動作が短期的セグメントを発話として堅牢に判断する場合には（動作４００４１における「Yes」）、短期的セグメントは、平滑化動作４１０６の前に、さらにVoIP発話または非VoIP発話に分類される（動作４００４２）。

【０４３１】

実のところ、平滑化方式を使わなくても、本方法は、コンテンツ型発話および／またはノイズを最初に識別してもよく、端的的セグメントが発話またはノイズとして分類されるとき、短期的セグメントをVoIP発話および非VoIP発話の一方またはVoIPノイズまたは非VoIPノイズの一方に分類するようさらなる分類が実装される。次いで、コンテキスト型を識別する動作が行なわれる。

【０４３２】

１．６節および１．８節で述べたように、そこで論じた遷移方式は、ここで記述されるオーディオ分類方法の一部として取り入れられてもよく、詳細は割愛する。手短かに言うと、本方法はさらに、コンテキスト型を識別する動作が同じコンテキスト型を連続的に出力する継続時間を測定することを含んでいてもよい。オーディオ分類方法は、新しいコンテキスト型の継続時間の長さが第六の閾値に達するまで、現在のコンテキスト型を出力し続けるよう構成される。

【０４３３】

同様に、あるコンテキスト型から別のコンテキスト型への異なる遷移対について、異なる第六の閾値が設定されてもよい。さらに、第六の閾値は、新しいコンテキスト型の信頼値と負に相関されてもよい。

【０４３４】

VoIP／非VoIP分類に特に向けられるオーディオ分類方法における遷移方式への修正として、現在の短期的セグメントについての第一ないし第四の閾値の任意の一つまたは複数が、最後の短期的セグメントのコンテキスト型に依存して異なるように設定されてもよい。

【０４３５】

オーディオ処理装置の実施形態と同様に、オーディオ処理方法の実施形態およびその変形の任意の組み合わせが現実的である。他方、オーディオ処理方法の実施形態およびその変形のあらゆる側面は別個の解決策であってもよい。さらに、この節に記載される任意の二つ以上の解決策が互いと組み合わされてもよく、これらの組み合わせがさらに、本開示の他の部において記載または含意される任意の実施形態と組み合わされてもよい。特に、ここに記載したオーディオ分類方法は、先述したオーディオ処理方法、特にボリューム平準化器制御方法において使用されてもよい。

【０４３６】

本願の「発明を実施するための形態」の冒頭で論じたように、本願の実施形態はハードウェアまたはソフトウェアまたは両方において具現されうる。図４１は、本願の諸側面を実装する例示的なシステムを示すブロック図である。

【０４３７】

図４１において、中央処理ユニット（CPU）４２０１は、読み出し専用メモリ（ROM）４２０２に記憶されたプログラムまたは記憶セクション４２０８からランダム・アクセス・メモリ（RAM）４２０３にロードされたプログラムに従ってさまざまなプロセスを実行する。RAM ４２０３では、CPU ４２０１が該さまざまなプロセスなどを実行するときに必要とされるデータも必要に応じて記憶される。

【０４３８】

CPU ４２０１、ROM ４２０２およびRAM ４２０３はバス４２０４を介して互いに接続される。入出力インターフェース４２０５もバス４２０４に接続される。

【０４３９】
次のコンポーネントが入出力インターフェース４２０５に接続される：キーボード、マウスなどを含む入力部４２０６；陰極線管（CRT）、液晶ディスプレイ（LCD）などのディスプレイおよびラウドスピーカーなどを含む出力部４２０７；ハードディスクなどを含む記憶部４２０８；およびLANカード、モデムなどのようなネットワーク・インターフェース・カードを含む通信部４２０９。通信部４２０９は、インターネットのようなネットワークを介して通信プロセスを実行する。
【０４４０】
ドライブ４２１０も必要に応じて入出力インターフェース４２０５に接続される。磁気ディスク、光ディスク、光磁気ディスク、半導体メモリなどのようなリムーバブル媒体４２１１がドライブ４２１０に必要に応じてマウントされる。それにより、そこから読み込まれるコンピュータ・プログラムが必要に応じて記憶部４２０８にインストールされる。
【０４４１】
上記のコンポーネントがソフトウェアによって実装される場合、該ソフトウェアをなすプログラムはインターネットのようなネットワークまたはリムーバブル媒体４２１１のような記憶媒体からインストールされる。
【０４４２】
本願で使われる用語は単に具体的な実施形態を記述するためのものであり、本願を限定することは意図されていないことを注意されたい。本稿での用法では、文脈がそうでないことを明確に示すのでない限り、単数形は複数も含むことが意図されている。さらに、本明細書において使われるときの「含む」および／または「有する」という用語は、述べられている特徴、整数、動作、段階、要素および／またはコンポーネントの存在を示すが、一つまたは複数の他の特徴、整数、動作、段階、要素、コンポーネントおよび／またはそれらの群の存在または追加を排除するものではない。
【０４４３】
請求項におけるあらゆる手段または動作に機能を加えた要素の対応する構造、材料、工程および等価物は、はっきりと請求項に記載されている他の請求項記載の要素との組み合わせにおいて機能を実行するための任意の構造、材料または工程を含むことが意図されている。本願の記述は、例解および説明のために提示されたが、網羅的であることや開示される形の応用に限定されることは意図されていない。本願の範囲および精神から外れることなく、多くの修正および変形が当業者には明白となるであろう。実施形態は、本願の原理および実際的な応用を最もよく説明するためおよび当業者が、考えられている具体的な用途に適したさまざまな修正をもつさまざまな実施形態について本願を理解できるようにするために選ばれ、記述された。
【０４４４】
いくつかの態様を記載しておく。
〔態様１〕
リアルタイムでオーディオ信号のコンテンツ型を識別するためのオーディオ・コンテンツ分類器と；
識別されたコンテンツ型に基づいて連続的な仕方でボリューム平準化器を調整する調整ユニットとを有するボリューム平準化器コントローラであって、
前記調整ユニットは、前記ボリューム平準化器の動的な利得を、前記オーディオ信号の情報性のコンテンツ型と正に相関させ、前記ボリューム平準化器の動的な利得を、前記オーディオ信号の干渉性のコンテンツ型と負に相関させるよう構成されている、
ボリューム平準化器コントローラ。
〔態様２〕
前記オーディオ信号の前記コンテンツ型が、発話、短期的音楽、ノイズおよび背景音のうちの一つを含む、態様１記載のボリューム平準化器コントローラ。
〔態様３〕
ノイズが干渉性のコンテンツ型と見なされる、態様１記載のボリューム平準化器コント

ローラ。
〔態様４〕
前記調整ユニットが、前記コンテンツ型の信頼値に基づいて前記ボリューム平準化器の動的な利得を調整するよう構成されている、態様１記載のボリューム平準化器コントローラ。
〔態様５〕
前記調整ユニットが、前記コンテンツ型の信頼値の伝達関数を介して前記動的な利得を調整するよう構成されている、態様４記載のボリューム平準化器コントローラ。
〔態様６〕
前記オーディオ・コンテンツ分類器が前記オーディオ信号を、対応する信頼値をもつ複数のコンテンツ型に分類するよう構成されており、前記調整ユニットが、前記複数のコンテンツ型の重要性に基づいて前記複数のコンテンツ型の前記信頼値を重み付けすることを通じて前記複数のオーディオ型の少なくともいくつかを考慮するよう構成されている、態様１記載のボリューム平準化器コントローラ。
〔態様７〕
前記オーディオ・コンテンツ分類器が前記オーディオ信号を、対応する信頼値をもつ複数のコンテンツ型に分類するよう構成されており、前記調整ユニットが、あるコンテンツ型の重みを少なくとも一つの他のコンテンツ型の信頼値を用いて修正するよう構成されている、態様１記載のボリューム平準化器コントローラ。
〔態様８〕
前記オーディオ・コンテンツ分類器が前記オーディオ信号を、対応する信頼値をもつ複数のコンテンツ型に分類するよう構成されており、前記調整ユニットが、前記信頼値に基づいて前記複数のコンテンツ型の効果を重み付けすることを通じて前記複数のオーディオ型の少なくともいくつかを考慮するよう構成されている、態様１記載のボリューム平準化器コントローラ。
〔態様９〕
前記調整ユニットが、前記信頼値に基づいて少なくとも一つの優勢なコンテンツ型を考慮するよう構成されている、態様８記載のボリューム平準化器コントローラ。
〔態様１０〕
前記オーディオ・コンテンツ分類器が前記オーディオ信号を、対応する信頼値をもつ複数の干渉性のコンテンツ型および／または複数の情報性のコンテンツ型に分類するよう構成されており、前記調整ユニットが、前記信頼値に基づいて少なくとも一つの優勢な干渉性のコンテンツ型および／または少なくとも一つの優勢な情報性のコンテンツ型を考慮するよう構成されている、態様１記載のボリューム平準化器コントローラ。
〔態様１１〕
各コンテンツ型について、前記オーディオ信号の過去の信頼値に基づいて現在の時点での前記オーディオ信号の信頼値を平滑化するための型平滑化ユニットをさらに有する、態様１ないし１０のうちいずれか一項記載のボリューム平準化器コントローラ。
〔態様１２〕
前記型平滑化ユニットは、現在の実際の信頼値と最後の時点での平滑化された信頼値との重み付けされた和を計算することによって、現時点での前記オーディオ信号の平滑化された信頼値を決定するよう構成されている、態様１１記載のボリューム平準化器コントローラ。
〔態様１３〕
前記オーディオ信号のコンテキスト型を識別するオーディオ・コンテキスト分類器をさらに有しており、前記調整ユニットは、前記コンテキスト型の信頼値に基づいて前記動的な利得の範囲を調整するよう構成されている、態様１ないし１０のうちいずれか一項記載のボリューム平準化器コントローラ。
〔態様１４〕
前記オーディオ信号のコンテキスト型を識別するオーディオ・コンテキスト分類器をさ

らに有しており、前記調整ユニットは、前記オーディオ信号の前記コンテキスト型に基づいて前記オーディオ信号の前記コンテンツ型を情報性または干渉性と見なすよう構成されている、態様１ないし１０のうちいずれか一項記載のボリューム平準化器コントローラ。

〔態様１５〕

前記オーディオ信号の前記コンテキスト型が、VoIP、映画的メディア、長期的音楽およびゲームのうちの一つを含む、態様１４記載のボリューム平準化器コントローラ。

〔態様１６〕

コンテキスト型VoIPのオーディオ信号においては、背景音が干渉性コンテンツ型と見なされ、一方、コンテキスト型非VoIPのオーディオ信号においては、背景音および／または発話および／または音楽が情報性コンテンツ型と見なされる、態様１４記載のボリューム平準化器コントローラ。

〔態様１７〕

前記オーディオ信号の前記コンテキスト型が高品質オーディオまたは低品質オーディオを含む、態様１４記載のボリューム平準化器コントローラ。

〔態様１８〕

異なるコンテキスト型のオーディオ信号におけるコンテンツ型が、前記オーディオ信号のコンテキスト型に依存して異なる重みを割り当てられる、態様１４記載のボリューム平準化器コントローラ。

〔態様１９〕

前記オーディオ・コンテキスト分類器が前記オーディオ信号を、対応する信頼値をもつ複数のコンテキスト型に分類するよう構成されており、前記調整ユニットが、前記複数のコンテキスト型の重要性に基づいて前記複数のコンテキスト型の前記信頼値を重み付けすることを通じて前記複数のコンテキスト型の少なくともいくつかを考慮するよう構成されている、態様１４記載のボリューム平準化器コントローラ。

〔態様２０〕

前記オーディオ・コンテキスト分類器が前記オーディオ信号を、対応する信頼値をもつ複数のコンテキスト型に分類するよう構成されており、前記調整ユニットが、前記信頼値に基づいて前記複数のコンテキスト型の効果を重み付けすることを通じて前記複数のコンテキスト型の少なくともいくつかを考慮するよう構成されている、態様１４記載のボリューム平準化器コントローラ。

〔態様２１〕

前記オーディオ・コンテンツ分類器が前記オーディオ信号の短期的セグメントに基づいて前記コンテンツ型を識別するよう構成されており、

前記オーディオ・コンテキスト分類器が少なくとも部分的には前記オーディオ・コンテンツ分類器によって識別されたコンテンツ型に基づく前記オーディオ信号の短期的セグメントに基づいて前記コンテキスト型を識別するよう構成されている、態様１４記載のボリューム平準化器コントローラ。

〔態様２２〕

前記オーディオ・コンテンツ分類器が、短期的セグメントをコンテンツ型VoIP発話またはコンテンツ型非VoIP発話に分類するVoIP発話分類器を有しており、

前記オーディオ・コンテキスト分類器は、VoIP発話および非VoIP発話の信頼値に基づいて、前記短期的セグメントをコンテキスト型VoIPまたはコンテキスト型非VoIPに分類するよう構成されている、態様２１記載のボリューム平準化器コントローラ。

〔態様２３〕

前記オーディオ・コンテンツ分類器がさらに、

短期的セグメントをVoIPノイズのコンテンツ型および非VoIPノイズのコンテンツ型に分類するVoIPノイズ分類器を有しており、

前記オーディオ・コンテキスト分類器は、VoIP発話、非VoIP発話、VoIPノイズおよび非VoIPノイズの信頼値に基づいて、前記短期的セグメントをコンテキスト型VoIPまたはコンテキスト型非VoIPに分類するよう構成されている、

態様２２記載のボリューム平準化器コントローラ。
〔態様２４〕
前記オーディオ・コンテキスト分類器が：
VoIP発話の信頼値が第一の閾値より大きい場合、前記短期的セグメントをコンテキスト型VoIPとして分類し；VoIP発話の信頼値が、前記第一の閾値より大きくない第二の閾値より大きくない場合、前記短期的セグメントをコンテキスト型非VoIPとして分類し；それ以外の場合には、前記短期的セグメントを最後の短期的セグメントについてのコンテキスト型として分類するよう構成されている、態様２２記載のボリューム平準化器コントローラ。
〔態様２５〕
前記オーディオ・コンテキスト分類器が：
VoIP発話の信頼値が第一の閾値より大きい場合またはVoIPノイズの信頼値が第三の閾値より大きい場合、前記短期的セグメントをコンテキスト型VoIPとして分類し；VoIP発話の信頼値が、前記第一の閾値より大きくない第二の閾値より大きくない場合またはVoIPノイズの信頼値が前記第三の閾値より大きくない第四の閾値より大きくない場合、前記短期的セグメントをコンテキスト型非VoIPとして分類し；それ以外の場合には前記短期的セグメントを最後の短期的セグメントについてのコンテキスト型として分類するよう構成されている、態様２３記載のボリューム平準化器コントローラ。
〔態様２６〕
前記コンテンツ型の過去の信頼値に基づいて現在の時点での前記コンテンツ型の信頼値を平滑化するための型平滑化ユニットをさらに有する、態様２１ないし２５のうちいずれか一項記載のボリューム平準化器コントローラ。
〔態様２７〕
前記型平滑化ユニットは、現在の短期的セグメントの信頼値と最後の短期的セグメントの平滑化された信頼値との重み付けされた和を計算することによって、現在の短期的セグメントの平滑化された信頼値を決定するよう構成されている、態様２６記載のボリューム平準化器コントローラ。
〔態様２８〕
前記オーディオ・コンテンツ分類器が前記短期的セグメントの発話のコンテンツ型を識別する発話／ノイズ分類器をさらに有しており、前記型平滑化ユニットは、平滑化前の現在の短期的セグメントについてのVoIP発話の信頼値を、所定の信頼値として、あるいは前記発話／ノイズ分類器によって分類されるコンテンツ型発話についての信頼値が第五の閾値より低い最後の短期的セグメントの平滑化された信頼値として、設定するよう構成されている、態様２６記載のボリューム平準化器コントローラ。
〔態様２９〕
前記オーディオ・コンテキスト分類器が、特徴として、前記短期的セグメントのコンテンツ型の信頼値および前記短期的セグメントから抽出された他の特徴を使って、機械学習モデルに基づいて前記短期的セグメントを分類するよう構成されている、態様２２または２３記載のボリューム平準化器コントローラ。
〔態様３０〕
前記オーディオ・コンテキスト分類器が同じコンテキスト型を連続的に出力する継続時間を測定するタイマーをさらに有しており、前記調整ユニットは、新しいコンテキスト型の継続時間の長さが第六の閾値に達するまで、現在のコンテキスト型を使い続けるよう構成される、態様１４ないし２９のうちいずれか一項記載のボリューム平準化器コントローラ。
〔態様３１〕
あるコンテキスト型から別のコンテキスト型への異なる遷移対について、異なる第六の閾値が設定される、態様３０記載のボリューム平準化器コントローラ。
〔態様３２〕
前記第六の閾値が、前記新しいコンテキスト型の信頼値と負に相関している、態様３０

記載のボリューム平準化器コントローラ。
〔態様３３〕
前記第一および／または第二の閾値が、最後の短期的セグメントのコンテキスト型によって異なる、態様２４または２５記載のボリューム平準化器コントローラ。
〔態様３４〕
態様１ないし３３のうちいずれか一項記載のボリューム平準化器コントローラを有するオーディオ処理装置。
〔態様３５〕
オーディオ信号の短期的セグメントのコンテンツ型を識別するオーディオ・コンテンツ分類器と；
少なくとも部分的には前記オーディオ・コンテンツ分類器によって識別されたコンテンツ型に基づいて前記短期的セグメントのコンテキスト型を識別するオーディオ・コンテキスト分類器とを有する、
オーディオ分類器。
〔態様３６〕
前記オーディオ・コンテンツ分類器が、前記短期的セグメントをコンテンツ型VoIP発話またはコンテンツ型非VoIP発話に分類するVoIP発話分類器を有しており、
前記オーディオ・コンテキスト分類器は、VoIP発話および非VoIP発話の信頼値に基づいて、前記短期的セグメントをコンテキスト型VoIPまたはコンテキスト型非VoIPに分類するよう構成されている、
態様３５記載のオーディオ分類器。
〔態様３７〕
前記オーディオ・コンテンツ分類器がさらに、
前記短期的セグメントをコンテンツ型VoIPノイズおよびコンテンツ型非VoIPノイズに分類するVoIPノイズ分類器を有しており、
前記オーディオ・コンテキスト分類器は、VoIP発話、非VoIP発話、VoIPノイズおよび非VoIPノイズの信頼値に基づいて、前記短期的セグメントをコンテキスト型VoIPまたはコンテキスト型非VoIPに分類するよう構成されている、
態様３６記載のオーディオ分類器。
〔態様３８〕
前記オーディオ・コンテキスト分類器が：
VoIP発話の信頼値が第一の閾値より大きい場合、前記短期的セグメントをコンテキスト型VoIPとして分類し；VoIP発話の信頼値が、前記第一の閾値より大きくない第二の閾値より大きくない場合、前記短期的セグメントをコンテキスト型非VoIPとして分類し；それ以外の場合には、前記短期的セグメントを最後の短期的セグメントについてのコンテキスト型として分類するよう構成されている、態様３７記載のオーディオ分類器。
〔態様３９〕
前記オーディオ・コンテキスト分類器が：
VoIP発話の信頼値が第一の閾値より大きい場合またはVoIPノイズの信頼値が第三の閾値より大きい場合、前記短期的セグメントをコンテキスト型VoIPとして分類し；VoIP発話の信頼値が、前記第一の閾値より大きくない第二の閾値より大きくない場合またはVoIPノイズの信頼値が前記第三の閾値より大きくない第四の閾値より大きくない場合、前記短期的セグメントをコンテキスト型非VoIPとして分類し；それ以外の場合には前記短期的セグメントを最後の短期的セグメントについてのコンテキスト型として分類するよう構成されている、
態様３７記載のオーディオ分類器。
〔態様４０〕
前記コンテンツ型の過去の信頼値に基づいて現在の時点での前記コンテンツ型の信頼値を平滑化するための型平滑化ユニットをさらに有する、態様３５ないし３９のうちいずれか一項記載のオーディオ分類器。

〔態様４１〕
前記型平滑化ユニットは、現在の短期的セグメントの信頼値と最後の短期的セグメントの平滑化された信頼値との重み付けされた和を計算することによって、現在の短期的セグメントの平滑化された信頼値を決定するよう構成されている、態様４０記載のオーディオ分類器。
〔態様４２〕
前記オーディオ・コンテンツ分類器が前記短期的セグメントからコンテンツ型発話を識別する発話／ノイズ分類器をさらに有しており、前記型平滑化ユニットは、平滑化前の現在の短期的セグメントについてのVoIP発話の信頼値を、所定の信頼値として、あるいは前記発話／ノイズ分類器によって分類されるコンテンツ型発話についての信頼値が第五の閾値より低い最後の短期的セグメントの平滑化された信頼値として、設定するよう構成されている、態様４１記載のオーディオ分類器。
〔態様４３〕
前記オーディオ・コンテキスト分類器が、特徴として、前記短期的セグメントのコンテンツ型の信頼値および前記短期的セグメントから抽出された他の特徴を使って、機械学習モデルに基づいて前記短期的セグメントを分類するよう構成されている、態様３６または３７記載のオーディオ分類器。
〔態様４４〕
前記オーディオ・コンテキスト分類器が同じコンテキスト型を連続的に出力する継続時間を測定するタイマーをさらに有しており、当該オーディオ分類器は、新しいコンテキスト型の継続時間の長さが第六の閾値に達するまで、現在のコンテキスト型を出力し続けるよう構成される、態様３８または３９記載のオーディオ分類器。
〔態様４５〕
あるコンテキスト型から別のコンテキスト型への異なる遷移対について、異なる第六の閾値が設定される、態様４４記載のオーディオ分類器。
〔態様４６〕
前記第六の閾値が、前記新しいコンテキスト型の信頼値と負に相関している、態様４４記載のオーディオ分類器。
〔態様４７〕
前記第一および／または第二の閾値が、最後の短期的セグメントのコンテキスト型によって異なる、態様３８または３９記載のオーディオ分類器。
〔態様４８〕
態様３５ないし４７のうちいずれか一項記載のオーディオ分類器を有するオーディオ処理装置。
〔態様４９〕
リアルタイムでオーディオ信号のコンテンツ型を識別する段階と；
識別されたコンテンツ型に基づいて連続的な仕方でボリューム平準化器を調整することを、前記ボリューム平準化器の動的な利得を、前記オーディオ信号の情報性のコンテンツ型と正に相関させ、前記ボリューム平準化器の動的な利得を、前記オーディオ信号の干渉性のコンテンツ型と負に相関させることによって行なう段階とを含む、
ボリューム平準化器制御方法。
〔態様５０〕
前記オーディオ信号の前記コンテンツ型が、発話、短期的音楽、ノイズおよび背景音のうちの一つを含む、態様４９記載のボリューム平準化器制御方法。
〔態様５１〕
ノイズが干渉性のコンテンツ型と見なされる、態様４９記載のボリューム平準化器制御方法。
〔態様５２〕
前記調整する動作が、前記コンテンツ型の信頼値に基づいて前記ボリューム平準化器の動的な利得を調整するよう構成されている、態様４９記載のボリューム平準化器制御方法

。
〔態様５３〕
前記調整する動作が、前記コンテンツ型の信頼値の伝達関数を介して前記動的な利得を調整するよう構成されている、態様５２記載のボリューム平準化器制御方法。
〔態様５４〕
前記オーディオ信号が、対応する信頼値をもつ複数のコンテンツ型に分類され、前記調整する動作が、前記複数のコンテンツ型の重要性に基づいて前記複数のコンテンツ型の前記信頼値を重み付けすることを通じて前記複数のオーディオ型の少なくともいくつかを考慮するよう構成されている、態様４９記載のボリューム平準化器制御方法。
〔態様５５〕
前記オーディオ信号が、対応する信頼値をもつ複数のコンテンツ型に分類され、前記調整する動作が、あるコンテンツ型の重みを少なくとも一つの他のコンテンツ型の信頼値を用いて修正するよう構成されている、態様４９記載のボリューム平準化器制御方法。
〔態様５６〕
前記オーディオ信号が、対応する信頼値をもつ複数のコンテンツ型に分類され、前記調整する動作が、前記信頼値に基づいて前記複数のコンテンツ型の効果を重み付けすることを通じて前記複数のコンテンツ型の少なくともいくつかを考慮するよう構成されている、態様４９記載のボリューム平準化器制御方法。
〔態様５７〕
前記調整する動作が、前記信頼値に基づいて少なくとも一つの優勢なコンテンツ型を考慮するよう構成されている、態様５６記載のボリューム平準化器制御方法。
〔態様５８〕
前記オーディオ信号が、対応する信頼値をもつ複数の干渉性のコンテンツ型および／または複数の情報性のコンテンツ型に分類され、前記調整する動作が、前記信頼値に基づいて少なくとも一つの優勢な干渉性のコンテンツ型および／または少なくとも一つの優勢な情報性のコンテンツ型を考慮するよう構成されている、態様５６記載のボリューム平準化器制御方法。
〔態様５９〕
各コンテンツ型について、前記オーディオ信号の過去の信頼値に基づいて現在の時点での前記オーディオ信号の信頼値を平滑化する段階をさらに含む、態様４９ないし５８のうちいずれか一項記載のボリューム平準化器制御方法。
〔態様６０〕
前記の型平滑化の動作は、現在の実際の信頼値と最後の時点での平滑化された信頼値との重み付けされた和を計算することによって、現時点での前記オーディオ信号の平滑化された信頼値を決定するよう構成されている、態様５９記載のボリューム平準化器制御方法。
〔態様６１〕
前記オーディオ信号のコンテキスト型を識別する段階をさらに含み、前記調整する動作は、前記コンテキスト型の信頼値に基づいて前記動的な利得の範囲を調整するよう構成されている、態様４９ないし５８のうちいずれか一項記載のボリューム平準化器制御方法。
〔態様６２〕
前記オーディオ信号のコンテキスト型を識別する段階をさらに含み、前記調整する動作は、前記オーディオ信号の前記コンテキスト型に基づいて前記オーディオ信号の前記コンテンツ型を情報性または干渉性と見なすよう構成されている、態様４９ないし５８のうちいずれか一項記載のボリューム平準化器制御方法。
〔態様６３〕
前記オーディオ信号の前記コンテキスト型が、VoIP、映画的メディア、長期的音楽およびゲームのうちの一つを含む、態様６２記載のボリューム平準化器制御方法。
〔態様６４〕
コンテキスト型VoIPのオーディオ信号においては、背景音が干渉性コンテンツ型と見な

され、一方、コンテキスト型非VoIPのオーディオ信号においては、背景音および／または発話および／または音楽が情報性コンテンツ型と見なされる、態様６２記載のボリューム平準化器制御方法。

〔態様６５〕

前記オーディオ信号の前記コンテキスト型が高品質オーディオまたは低品質オーディオを含む、態様６２記載のボリューム平準化器制御方法。

〔態様６６〕

異なるコンテキスト型のオーディオ信号におけるコンテンツ型が、前記オーディオ信号のコンテキスト型に依存して異なる重みを割り当てられる、態様６２記載のボリューム平準化器制御方法。

〔態様６７〕

前記オーディオ信号が、対応する信頼値をもつ複数のコンテキスト型に分類され、前記調整する動作が、前記複数のコンテキスト型の重要性に基づいて前記複数のコンテキスト型の前記信頼値を重み付けすることを通じて前記複数のコンテキスト型の少なくともいくつかを考慮するよう構成されている、態様６２記載のボリューム平準化器制御方法。

〔態様６８〕

前記オーディオ信号が、対応する信頼値をもつ複数のコンテキスト型に分類され、前記調整する動作が、前記信頼値に基づいて前記複数のコンテキスト型の効果を重み付けすることを通じて前記複数のコンテキスト型の少なくともいくつかを考慮するよう構成されている、態様６２記載のボリューム平準化器制御方法。

〔態様６９〕

前記コンテンツ型を識別する動作が、前記オーディオ信号の短期的セグメントに基づいて前記コンテンツ型を識別するよう構成されており、

前記コンテキスト型を識別する動作が、少なくとも部分的には識別されたコンテンツ型に基づく前記オーディオ信号の短期的セグメントに基づいて前記コンテキスト型を識別するよう構成されている、

態様６２記載のボリューム平準化器制御方法。

〔態様７０〕

コンテンツ型を識別する動作が、短期的セグメントをコンテンツ型VoIP発話またはコンテンツ型非VoIP発話に分類することを含み、

コンテキスト型を識別する動作が、VoIP発話および非VoIP発話の信頼値に基づいて、前記短期的セグメントをコンテキスト型VoIPまたはコンテキスト型非VoIPに分類するよう構成されている、態様６９記載のボリューム平準化器制御方法。

〔態様７１〕

コンテンツ型を識別する動作がさらに、

短期的セグメントをコンテンツ型VoIPノイズおよびコンテンツ型非VoIPノイズに分類することを含み、

コンテキスト型を識別する動作は、VoIP発話、非VoIP発話、VoIPノイズおよび非VoIPノイズの信頼値に基づいて、前記短期的セグメントをコンテキスト型VoIPまたはコンテキスト型非VoIPに分類するよう構成されている、

態様７０記載のボリューム平準化器制御方法。

〔態様７２〕

コンテキスト型を識別する動作が：

VoIP発話の信頼値が第一の閾値より大きい場合、前記短期的セグメントをコンテキスト型VoIPとして分類し；

VoIP発話の信頼値が、前記第一の閾値より大きくない第二の閾値より大きくない場合、前記短期的セグメントをコンテキスト型非VoIPとして分類し；

それ以外の場合には、前記短期的セグメントを最後の短期的セグメントについてのコンテキスト型として分類するよう構成されている、

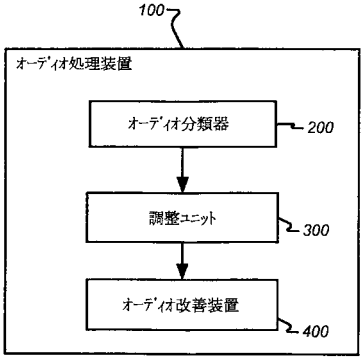
態様７０記載のボリューム平準化器制御方法。

〔態様７３〕
コンテキストを識別する動作が：
VoIP発話の信頼値が第一の閾値より大きい場合またはVoIPノイズの信頼値が第三の閾値より大きい場合、前記短期的セグメントをコンテキスト型VoIPとして分類し；
VoIP発話の信頼値が、前記第一の閾値より大きくない第二の閾値より大きくない場合またはVoIPノイズの信頼値が前記第三の閾値より大きくない第四の閾値より大きくない場合、前記短期的セグメントをコンテキスト型非VoIPとして分類し；
それ以外の場合には前記短期的セグメントを最後の短期的セグメントについてのコンテキスト型として分類するよう構成されている、態様７１記載のボリューム平準化器制御方法。
〔態様７４〕
前記コンテンツ型の過去の信頼値に基づいて現在の時点での前記コンテンツ型の信頼値を平滑化する段階をさらに含む、態様６９ないし７３のうちいずれか一項記載のボリューム平準化器制御方法。
〔態様７５〕
前記の型平滑化の動作は、現在の短期的セグメントの信頼値と最後の短期的セグメントの平滑化された信頼値との重み付けされた和を計算することによって、現在の短期的セグメントの平滑化された信頼値を決定するよう構成されている、態様７４記載のボリューム平準化器制御方法。
〔態様７６〕
前記短期的セグメントの発話のコンテンツ型を識別する段階をさらに含み、平滑化前の現在の短期的セグメントについてのVoIP発話の信頼値が、所定の信頼値として、あるいはコンテンツ型発話についての信頼値が第五の閾値より低い最後の短期的セグメントの平滑化された信頼値として、設定される、態様７５記載のボリューム平準化器制御方法。
〔態様７７〕
特徴として、前記短期的セグメントのコンテンツ型の信頼値および前記短期的セグメントから抽出された他の特徴を使って、機械学習モデルに基づいて、前記短期的セグメントが分類される、態様７０または７１記載のボリューム平準化器制御方法。
〔態様７８〕
コンテキスト型を識別する動作が同じコンテキスト型を連続的に出力する継続時間を測定することをさらに含み、前記調整する動作は、新しいコンテキスト型の継続時間の長さが第六の閾値に達するまで、現在のコンテキスト型を使い続けるよう構成される、態様６２ないし７７のうちいずれか一項記載のボリューム平準化器制御方法。
〔態様７９〕
あるコンテキスト型から別のコンテキスト型への異なる遷移対について、異なる第六の閾値が設定される、態様７８記載のボリューム平準化器制御方法。
〔態様８０〕
前記第六の閾値が、前記新しいコンテキスト型の信頼値と負に相関している、態様７８記載のボリューム平準化器制御方法。
〔態様８１〕
前記第一および／または第二の閾値が、最後の短期的セグメントのコンテキスト型によって異なる、態様７２または７３記載のボリューム平準化器制御方法。
〔態様８２〕
オーディオ信号の短期的セグメントのコンテンツ型を識別する段階と；
少なくとも部分的には識別されたコンテンツ型に基づいて前記短期的セグメントのコンテキスト型を識別する段階とを含む、
オーディオ分類方法。
〔態様８３〕
コンテンツ型を分類する動作が、前記短期的セグメントをコンテンツ型VoIP発話またはコンテンツ型非VoIP発話に分類することを含み、

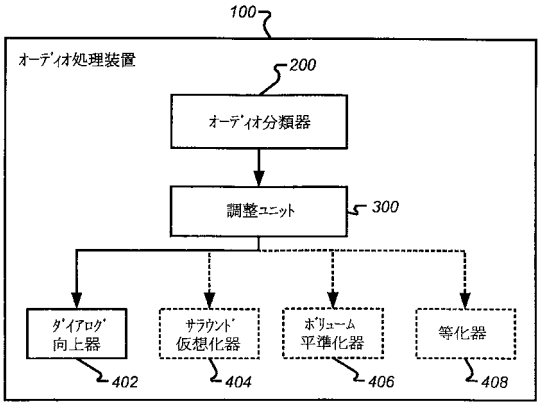
コンテキスト型を識別する動作が、VoIP発話および非VoIP発話の信頼値に基づいて、前記短期的セグメントをコンテキスト型VoIPまたはコンテキスト型非VoIPに分類するよう構成されている、
態様８２記載のオーディオ分類方法。
〔態様８４〕
コンテンツ型を分類する動作がさらに、
前記短期的セグメントをコンテンツ型VoIPノイズまたはコンテンツ型非VoIPノイズに分類することを含み、
コンテキスト型を識別する動作が、VoIP発話、非VoIP発話、VoIPノイズおよび非VoIPノイズの信頼値に基づいて、前記短期的セグメントをコンテキスト型VoIPまたはコンテキスト型非VoIPに分類するよう構成されている、
態様８３記載のオーディオ分類方法。
〔態様８５〕
コンテキスト型を識別する動作が：
VoIP発話の信頼値が第一の閾値より大きい場合、前記短期的セグメントをコンテキスト型VoIPとして分類し；
VoIP発話の信頼値が、前記第一の閾値より大きくない第二の閾値より大きくない場合、前記短期的セグメントをコンテキスト型非VoIPとして分類し；
それ以外の場合には、前記短期的セグメントを最後の短期的セグメントについてのコンテキスト型として分類するよう構成されている、
態様８３記載のオーディオ分類方法。
〔態様８６〕
コンテキスト型を識別する動作が：
VoIP発話の信頼値が第一の閾値より大きい場合またはVoIPノイズの信頼値が第三の閾値より大きい場合、前記短期的セグメントをコンテキスト型VoIPとして分類し；
VoIP発話の信頼値が、前記第一の閾値より大きくない第二の閾値より大きくない場合またはVoIPノイズの信頼値が前記第三の閾値より大きくない第四の閾値より大きくない場合、前記短期的セグメントをコンテキスト型非VoIPとして分類し；
それ以外の場合には前記短期的セグメントを最後の短期的セグメントについてのコンテキスト型として分類するよう構成されている、
態様８４記載のオーディオ分類方法。
〔態様８７〕
前記コンテンツ型の過去の信頼値に基づいて現在の時点での前記コンテンツ型の信頼値を平滑化する段階をさらに含む、態様８２ないし８６のうちいずれか一項記載のオーディオ分類方法。
〔態様８８〕
前記の型平滑化の動作は、現在の短期的セグメントの信頼値と最後の短期的セグメントの平滑化された信頼値との重み付けされた和を計算することによって、現在の短期的セグメントの平滑化された信頼値を決定するよう構成されている、態様８７記載のオーディオ分類方法。
〔態様８９〕
前記短期的セグメントからコンテンツ型発話を識別する段階をさらに含み、平滑化前の現在の短期的セグメントについてのVoIP発話の信頼値が、所定の信頼値として、あるいはコンテンツ型発話についての信頼値が第五の閾値より低い最後の短期的セグメントの平滑化された信頼値として、設定される、態様８８記載のオーディオ分類方法。
〔態様９０〕
コンテキスト型を識別する動作が、特徴として、前記短期的セグメントのコンテンツ型の信頼値および前記短期的セグメントから抽出された他の特徴を使って、機械学習モデルに基づいて前記短期的セグメントを分類するよう構成されている、態様８３または８４記載のオーディオ分類方法。

〔態様９１〕
コンテキスト型を識別する動作が同じコンテキスト型を連続的に出力する継続時間を測定する段階をさらに含み、当該オーディオ分類方法は、新しいコンテキスト型の継続時間の長さが第六の閾値に達するまで、現在のコンテキスト型を出力し続けるよう構成される、態様８５または８６記載のオーディオ分類方法。
〔態様９２〕
あるコンテキスト型から別のコンテキスト型への異なる遷移対について、異なる第六の閾値が設定される、態様９１記載のオーディオ分類方法。
〔態様９３〕
前記第六の閾値が、前記新しいコンテキスト型の信頼値と負に相関している、態様９１記載のオーディオ分類方法。
〔態様９４〕
前記第一および／または第二の閾値が、最後の短期的セグメントのコンテキスト型によって異なる、態様８５または８６記載のオーディオ分類方法。
〔態様９５〕
プロセッサによって実行されると該プロセッサがボリューム平準化器制御方法を実行できるようにするコンピュータ・プログラム命令が記録されたコンピュータ可読媒体であって、前記ボリューム平準化器制御方法は、
リアルタイムでオーディオ信号のコンテンツ型を識別する段階と；
識別されたコンテンツ型に基づいて連続的な仕方でボリューム平準化器を調整することを、前記ボリューム平準化器の動的な利得を、前記オーディオ信号の情報性のコンテンツ型と正に相関させ、前記ボリューム平準化器の動的な利得を、前記オーディオ信号の干渉性のコンテンツ型と負に相関させることによって行なう段階とを含む、
コンピュータ可読媒体。
〔態様９６〕
プロセッサによって実行されると該プロセッサがオーディオ分類方法を実行できるようにするコンピュータ・プログラム命令が記録されたコンピュータ可読媒体であって、前記オーディオ分類方法は、
オーディオ信号の短期的セグメントのコンテンツ型を識別する段階と；
少なくとも部分的には識別されたコンテンツ型に基づいて前記短期的セグメントのコンテキスト型を識別する段階とを含む、
コンピュータ可読媒体。

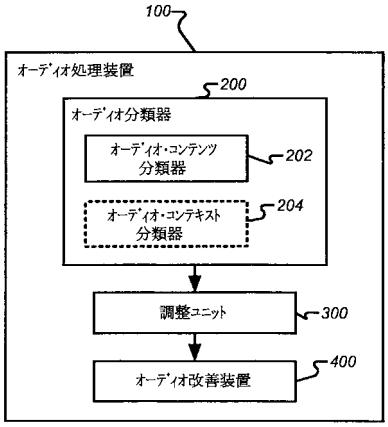
【図１】


100
オーディオ処理装置
オーディオ分類器
200
調整ユニット
300
オーディオ改善装置
400

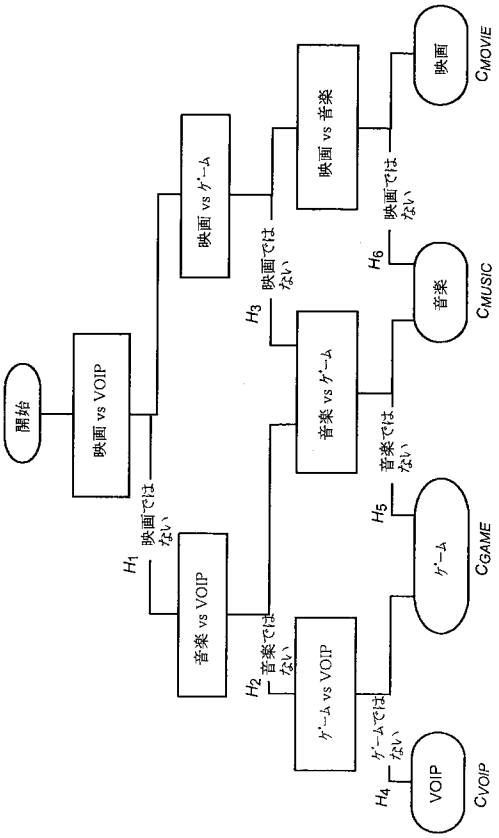
【図２】


100
オーディオ処理装置
200
オーディオ分類器
調整ユニット
300
ダイアログ向上器
402
サラウンド仮想化器
404
ボリューム平準化器
406
等化器
408

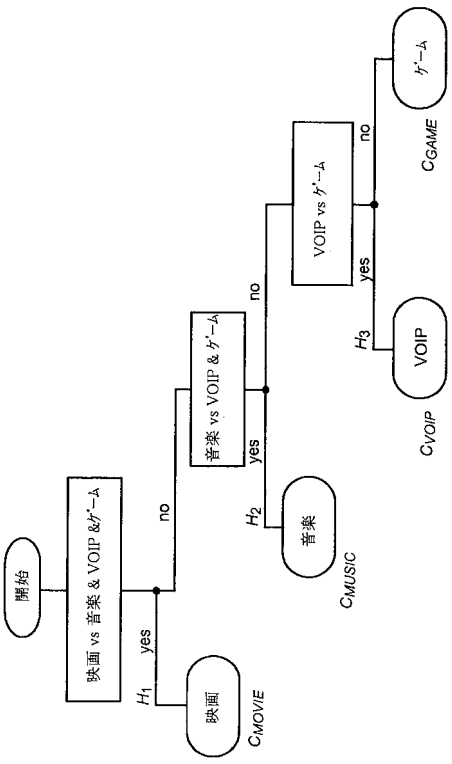
【図３】


100
オーディオ処理装置
200
オーディオ分類器
オーディオ・コンテンツ分類器
202
オーディオ・コンテキスト分類器
204
調整ユニット
300
オーディオ改善装置
400

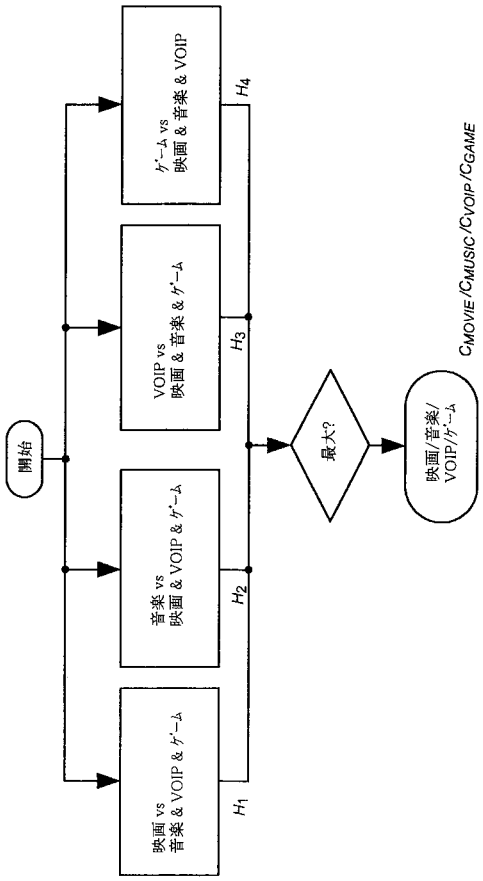
【図４】


開始
映画 vs VOIP
映画ではない
H1
音楽 vs VOIP
音楽ではない
H2
ゲーム vs VOIP
ゲームではない
H4
VOIP
CVOIP
ゲーム
CGAME
H5
音楽ではない
音楽 vs ゲーム
映画ではない
H3
映画 vs ゲーム
映画ではない
H6
映画 vs 音楽
音楽
CMUSIC
映画
CMOVIE

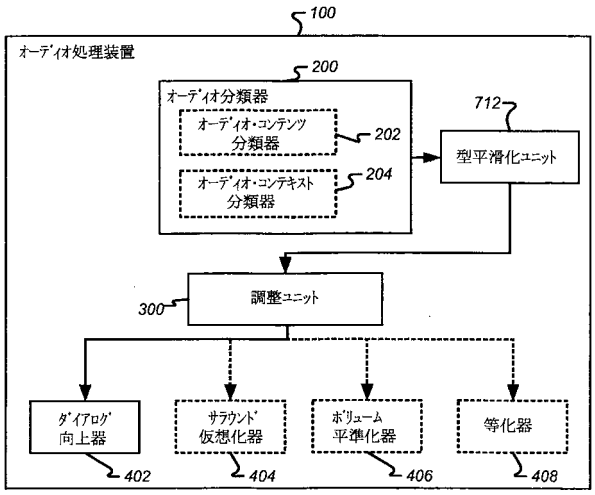
【図５】


開始
映画 vs 音楽 & VOIP &ゲーム
yes
H1
映画
CMOVIE
no
音楽 vs VOIP & ゲーム
yes
H2
音楽
CMUSIC
no
VOIP vs ゲーム
yes
H3
VOIP
CVOIP
no
ゲーム
CGAME

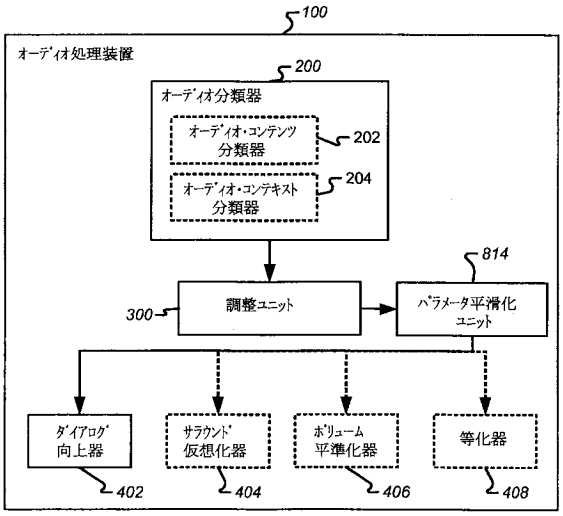
【図６】


開始
ゲーム vs 映画 & 音楽 & VOIP
VOIP vs 映画 & 音楽 & ゲーム
音楽 vs 映画 & VOIP & ゲーム
映画 vs 音楽 & VOIP & ゲーム
H4
H3
H2
H1
最大?
映画/音楽/VOIP/ゲーム
CMOVIE/CMUSIC/CVOIP/CGAME

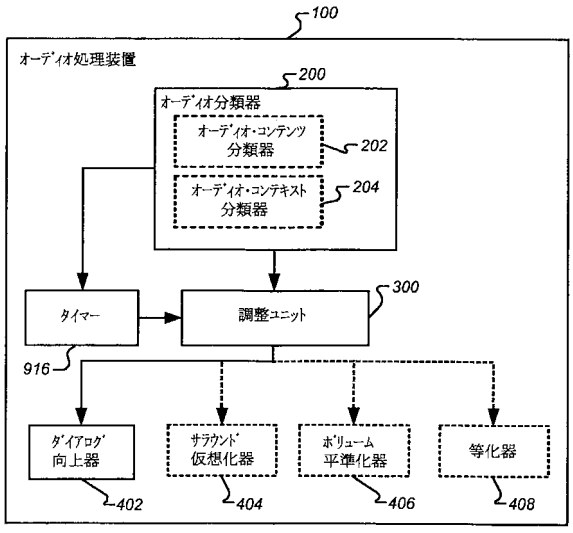
【図７】


100
オーディオ処理装置
200
オーディオ分類器
オーディオ・コンテンツ 分類器
202
オーディオ・コンテキスト 分類器
204
712
型平滑化ユニット
300
調整ユニット
ダイアログ向上器
サラウンド仮想化器
ボリューム平準化器
等化器
402
404
406
408

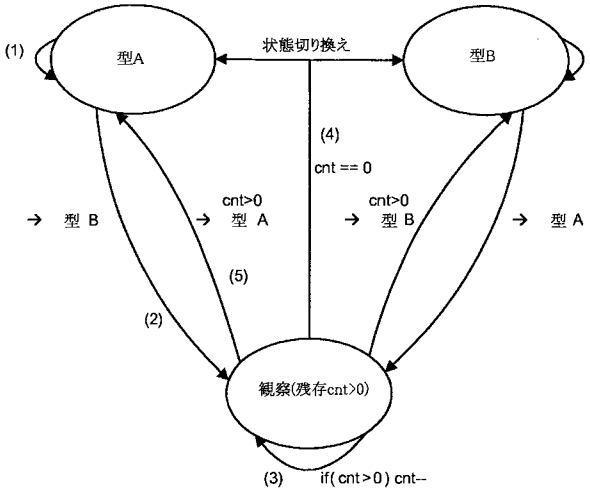
【図８】


100
オーディオ処理装置
200
オーディオ分類器
オーディオ・コンテンツ 分類器
202
オーディオ・コンテキスト 分類器
204
814
300
調整ユニット
パラメータ平滑化ユニット
ダイアログ向上器
サラウンド仮想化器
ボリューム平準化器
等化器
402
404
406
408

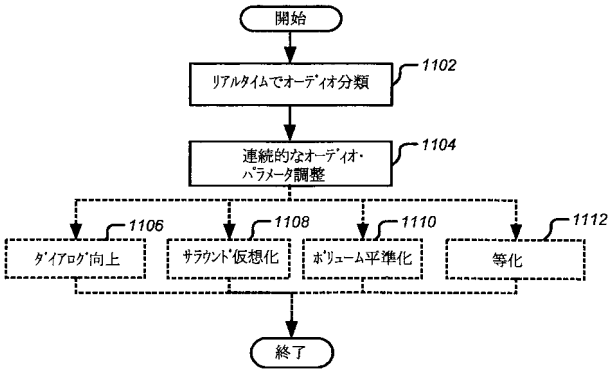
【図９】


100
オーディオ処理装置
200
オーディオ分類器
オーディオ・コンテンツ 分類器
202
オーディオ・コンテキスト 分類器
204
タイマー
916
300
調整ユニット
ダイアログ向上器
サラウンド仮想化器
ボリューム平準化器
等化器
402
404
406
408

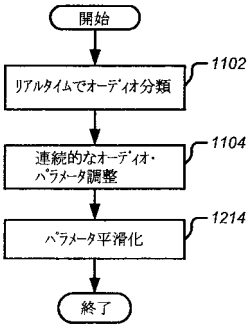
【図１０】



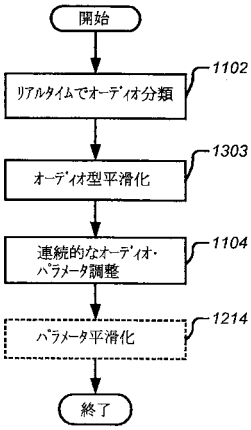
【図１１】



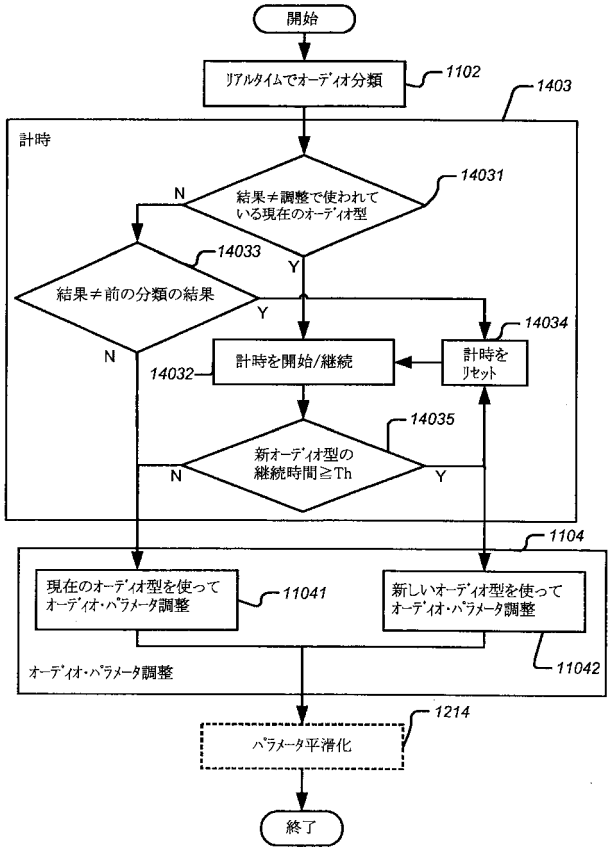
【図１２】



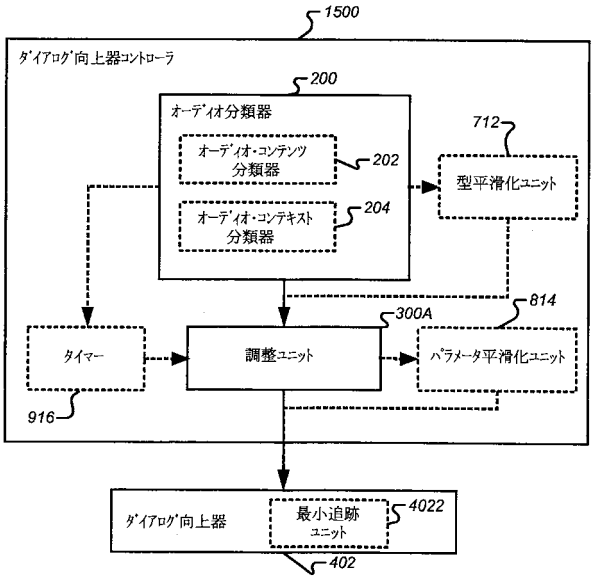
【図１３】



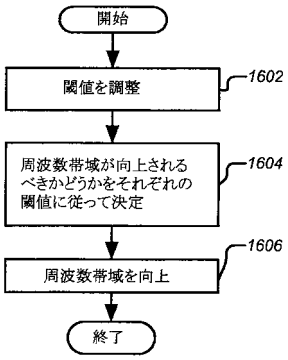
【図１４】



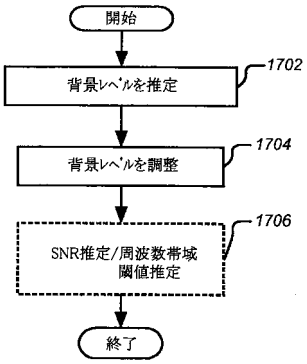
【図１５】



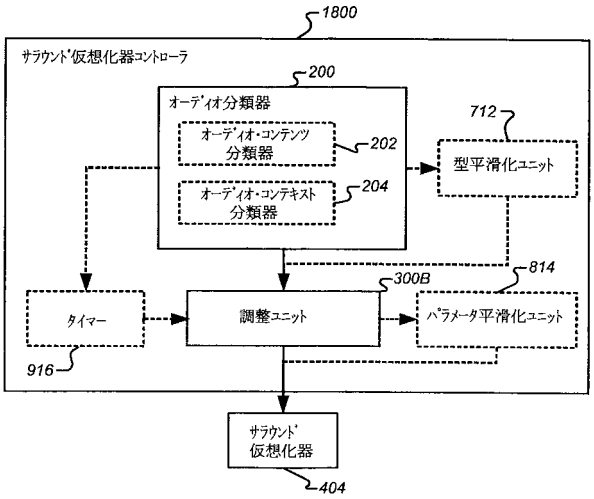
【図１６】



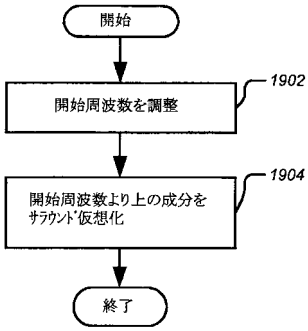
【図１７】



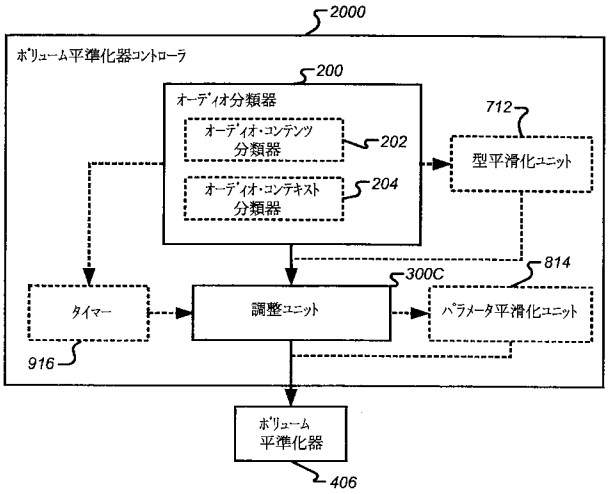
【図１８】



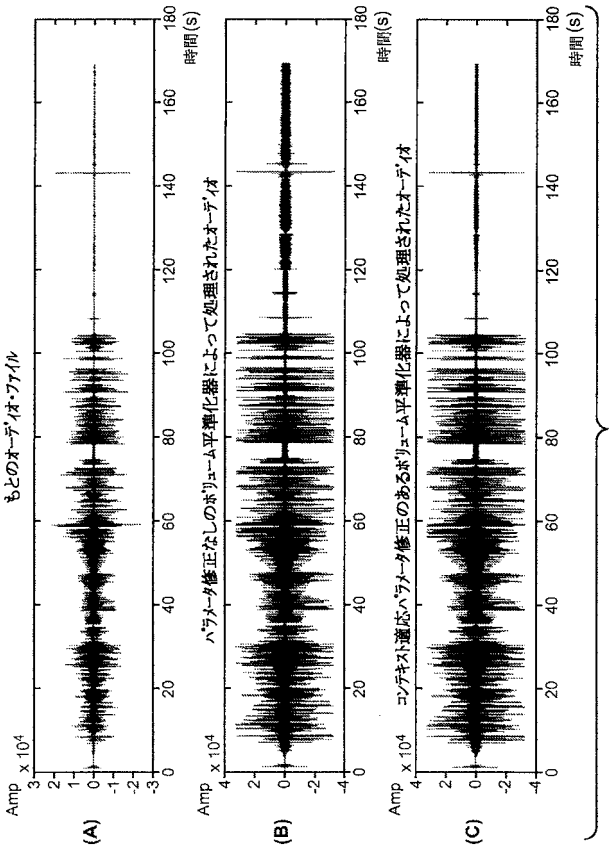
【図１９】



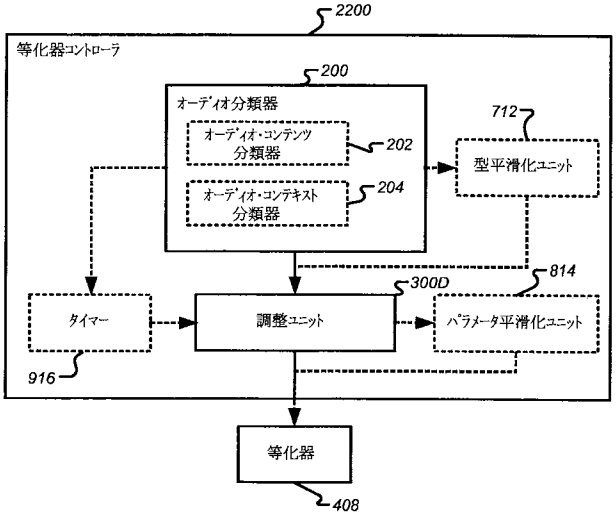
【図２０】



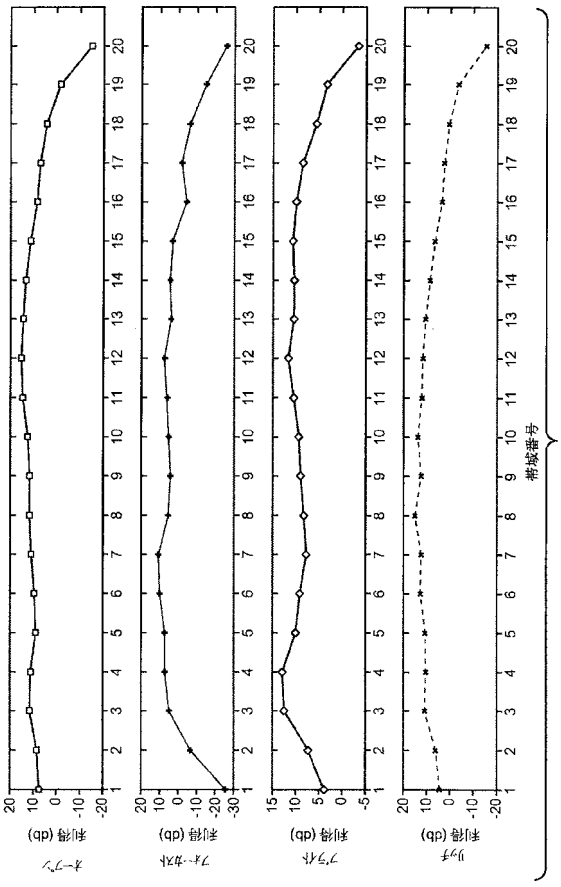
【図２１】



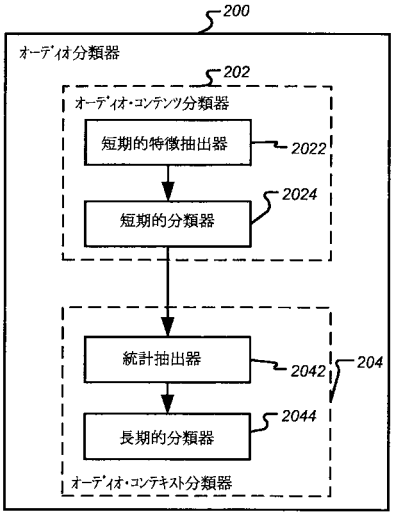
【図２２】



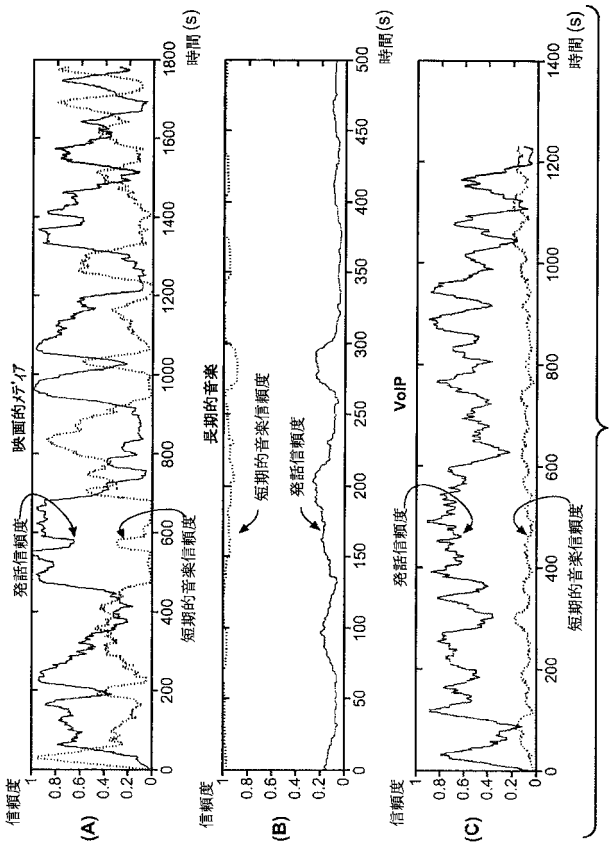
【図２３】



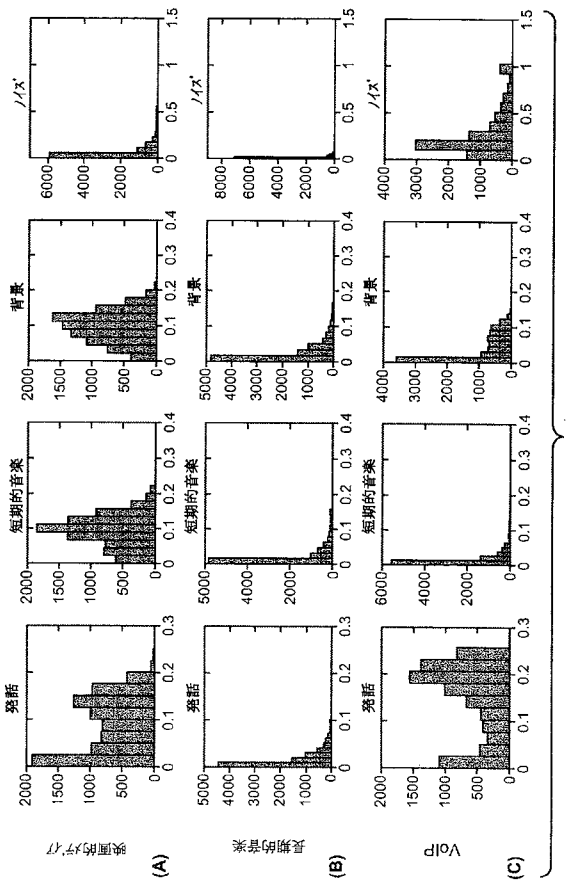
【図２４】



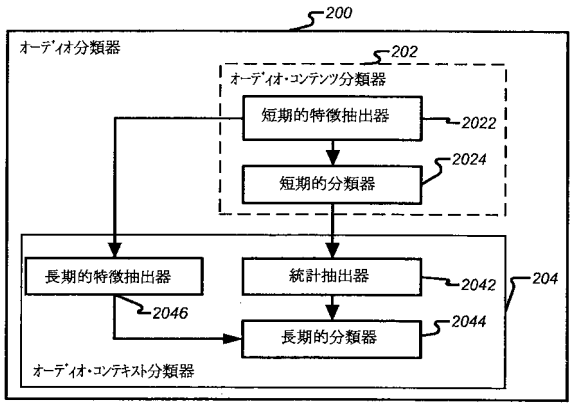
【図２５】



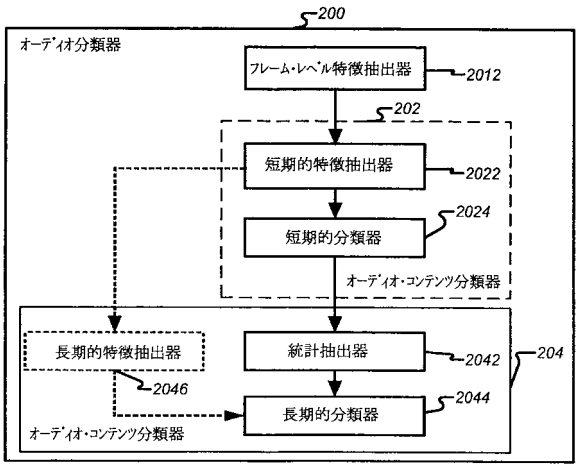
【図２６】



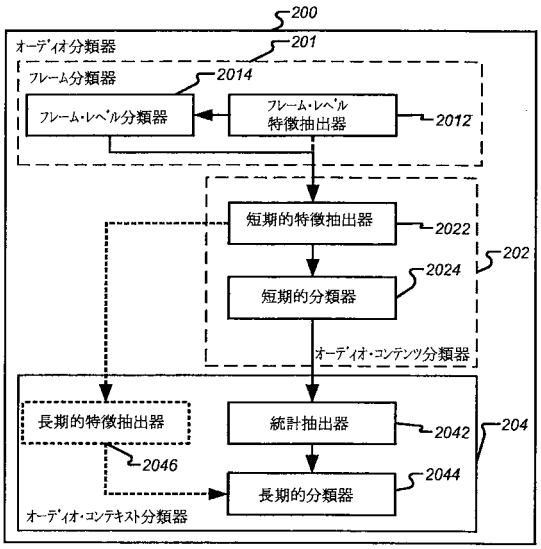
【図２７】



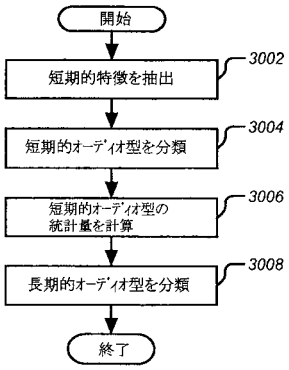
【図２８】



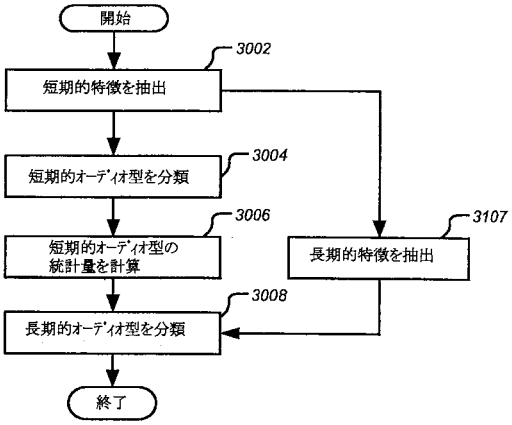
【図２９】



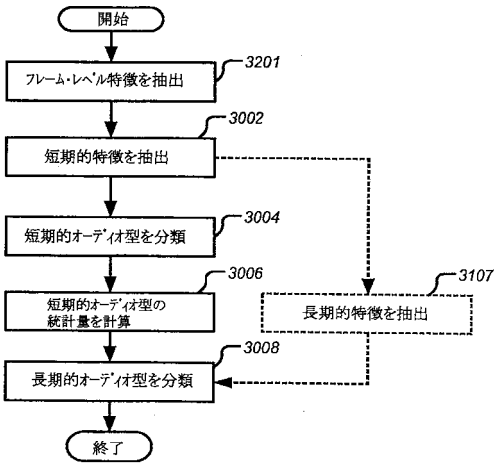
【図３０】



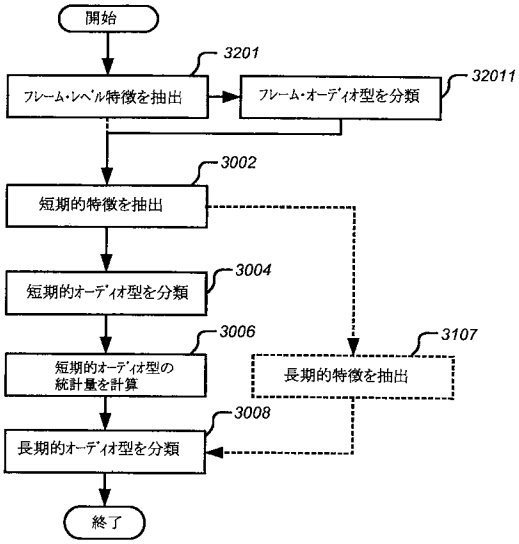
【図３１】



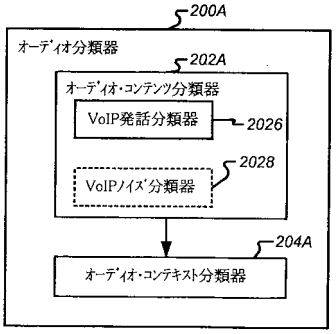
【図３２】



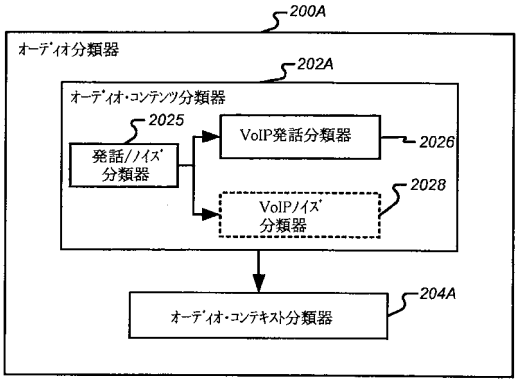
【図３３】



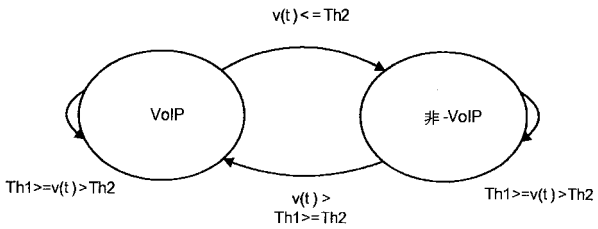
【図３４】



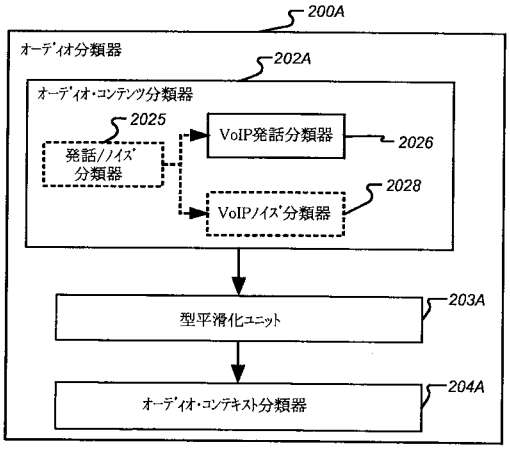
【図３５】



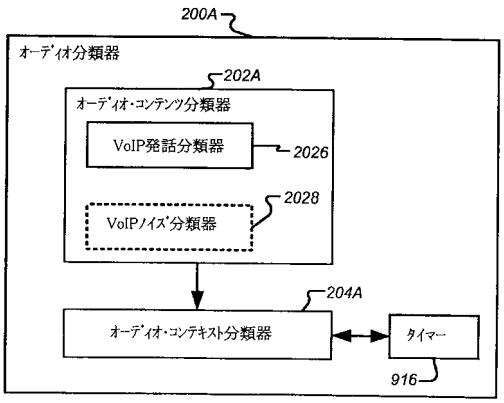
【図３６】



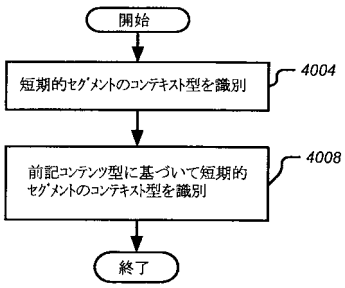
【図３７】



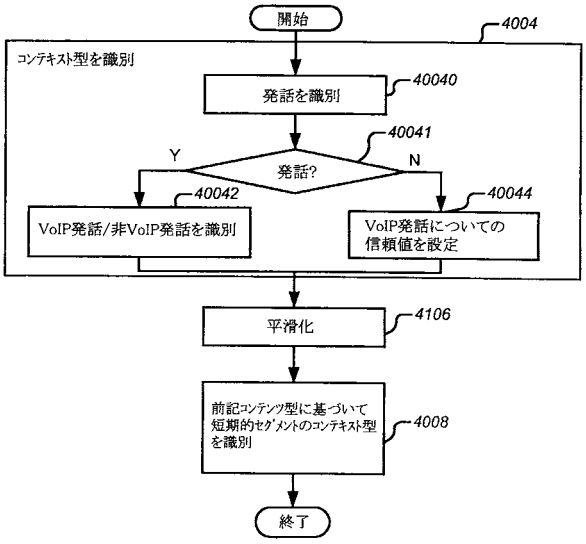
【図３８】



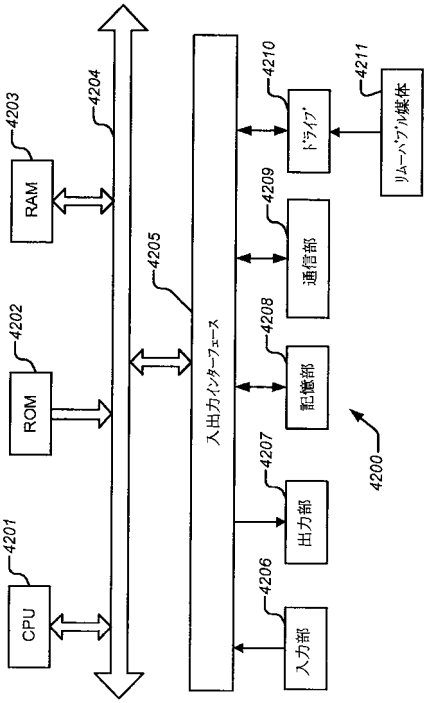
【図３９】



【図４０】


開始
4004
コンテキスト型を識別
発話を識別
40040
40041
Y
発話?
N
40042
40044
VoIP発話/非VoIP発話を識別
VoIP発話についての
信頼値を設定
平滑化
4106
前記コンテンツ型に基づいて
短期的セグメントのコンテキスト型
を識別
4008
終了

【図４１】


4203
RAM
4204
4205
4202
ROM
入出力インターフェース
4210
ドライブ
4211
リムーバブル媒体
4209
通信部
4208
記憶部
4200
4207
出力部
4201
CPU
4206
入力部

フロントページの続き

(72)発明者　ワン，ジュン
中華人民共和国，ベイジン　１０００２０，シャオヤン・ディストリクト，イースト・サード・リング・ミドル・ロード，ナンバー　１，ワールド・フィナンシャル・センター　ドルビー　ラボラトリーズ　インターナショナル　サービシーズ（ベイジン）　カンパニー　リミテッド内
(72)発明者　ルー，リエ
中華人民共和国，ベイジン　１０００２０，シャオヤン・ディストリクト，イースト・サード・リング・ミドル・ロード，ナンバー　１，ワールド・フィナンシャル・センター　ドルビー　ラボラトリーズ　インターナショナル　サービシーズ（ベイジン）　カンパニー　リミテッド内
(72)発明者　シーフェルドット，アラン
アメリカ合衆国，カリフォルニア州　９４１０３－４８１３，サンフランシスコ，ポットレロ　アヴェニュ　１００，ドルビー　ラボラトリーズ，インコーポレイテッド内

Ｆターム(参考) 5J030 AA01　AA15　AC10　AC20