

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6328627号  
(P6328627)

(45) 発行日 平成30年5月23日 (2018. 5. 23)

(24) 登録日 平成30年4月27日 (2018. 4. 27)

(51) Int. Cl.		F I	
<b>G 1 O L</b>	<b>21/034</b>	<b>(2013. 01)</b>	G 1 O L 21/034
<b>G 1 O L</b>	<b>21/0364</b>	<b>(2013. 01)</b>	G 1 O L 21/0364
<b>G 1 O L</b>	<b>25/84</b>	<b>(2013. 01)</b>	G 1 O L 25/84
<b>H O 3 G</b>	<b>5/16</b>	<b>(2006. 01)</b>	H O 3 G 5/16 1 6 5

請求項の数 5 (全 34 頁)

(21) 出願番号	特願2015-521830 (P2015-521830)	(73) 特許権者	515010084
(86) (22) 出願日	平成25年7月12日 (2013. 7. 12)		ディーティーエス インコーポレイテッド
(65) 公表番号	特表2015-523606 (P2015-523606A)		D T S, I N C.
(43) 公表日	平成27年8月13日 (2015. 8. 13)		アメリカ合衆国 9 1 3 0 2 カリフォル
(86) 国際出願番号	PCT/US2013/050207		ニア, カラバサス, ラス ヴァージネ
(87) 国際公開番号	W02014/011959		ス ロード 5 2 2 0
(87) 国際公開日	平成26年1月16日 (2014. 1. 16)		5 2 2 0 L a s V i r g e n e s R
審査請求日	平成28年7月7日 (2016. 7. 7)		o a d, C a l a b a s a s, C a l
(31) 優先権主張番号	61/670, 991	(74) 代理人	100094569
(32) 優先日	平成24年7月12日 (2012. 7. 12)		弁理士 田中 伸一郎
(33) 優先権主張国	米国 (US)	(74) 代理人	100088694
(31) 優先権主張番号	61/671, 005		弁理士 弟子丸 健
(32) 優先日	平成24年7月12日 (2012. 7. 12)		
(33) 優先権主張国	米国 (US)		

最終頁に続く

(54) 【発明の名称】 雑音検出及びラウドネス低下検出によるラウドネスコントロール

(57) 【特許請求の範囲】

【請求項 1】

音声信号を処理するように構成されたラウドネスコントロールシステムであって、  
前記音声信号の短期的なラウドネス推定を生成するように構成されたラウドネス測定モジュールと、

前記音声信号がコンテンツ又は雑音である可能性が高いかどうかを示すコンテンツ対雑音分類を生成するように構成された雑音検出モジュールと、

前記コンテンツ対雑音分類に基づいて少なくとも1つの平滑化係数を調整し、前記短期的なラウドネス推定に基づいて、前記少なくとも1つの平滑化係数を使用して前記音声信号の長期的なラウドネス推定を生成するように構成された時間平滑化モジュールと、

前記長期的なラウドネス推定に基づいて、時間的に変化する利得を前記音声信号に適用するように構成された利得補正モジュールと、を備えるラウドネスコントロールシステム。

【請求項 2】

前記雑音検出モジュールは、周波数領域雑音検出を使用して前記コンテンツ対雑音分類を生成するように構成される、請求項 1 に記載のラウドネスコントロールシステム。

【請求項 3】

前記雑音検出モジュールは、時間領域雑音検出を使用して前記コンテンツ対雑音分類を生成するように構成される、請求項 1 に記載のラウドネスコントロールシステム。

【請求項 4】

前記少なくとも1つの平滑化係数は、前記利得補正モジュールが利得レベルを高くすることができる速度を制御するリリース平滑化係数を含む、請求項1に記載のラウドネスコントロールシステム。

【請求項5】

ラウドネス低下検出値を生成するように構成されたラウドネス低下検出モジュールをさらに備え、

前記時間平滑化モジュールは、前記ラウドネス低下検出値に基づいて前記少なくとも1つの平滑化係数を調整するようにさらに構成される、請求項1に記載のラウドネスコントロールシステム。

【発明の詳細な説明】

10

【技術分野】

【0001】

本出願は、ラウドネスコントロールシステム (loudness control systems) に関する。

【背景技術】

【0002】

関連出願の相互参照

本出願は、参照することにより本明細書に援用されている、2012年7月12日出願した米国特許仮出願第61/670,991号明細書及び2012年7月12日出願した米国特許仮出願第61/671,005号明細書の利益を主張するものである。

20

【0003】

ラウドネスコントロールシステムは、ラウドネスレベルが変化する入力音声信号から均一なラウドネスレベルを有する出力音声信号を生成するように設計され得る。これらのシステムは、音声ブロードキャストチェーン (audio broadcast chains) などのアプリケーション及びラウドネスレベルが変化する複数のコンテンツソース (content sources) が利用可能な音声再生デバイスにおいて使用され得る。ラウドネスコントロールシステムの例示的な目標は、均一な平均ラウドネスレベルを有する出力信号を自動的に供給し、リスナーが再生デバイス側のボリュームコントロールを常時調整する必要をなくすことである。

【0004】

30

ラウドネスコントロールシステムに関係しているのは、自動利得制御 (AGC) 及びダイナミックレンジ制御 (DRC) システムである。AGCシステムは、現代のラウドネスコントロールシステムの先駆けとなるもので、通信及び放送アプリケーションにおいて長い歴史を有し、初期の頃の多くの設計はアナログ回路として実装されていた。AGCシステムは、入力信号に時間的に変化する利得信号を乗算することによって機能するものであり、この利得信号は、出力信号の客観的尺度が所定の目標レベルに正規化されるように制御される。例えば二乗平均平方根 (RMS)、ピーク、振幅、又はエネルギー尺度などの客観的尺度が使用され得る。既存のAGC設計の短所の1つは、出力信号の知覚ラウドネスが予測不可能なままとなり得ることである。これは、知覚ラウドネスがRMS、ピーク、振幅、又はエネルギー準位などの客観的尺度とおおまかにしか相関しない主観的尺度であるという心理音響的な現象に起因する。そのためAGCは、出力信号のRMS値を適切に制御できるが、必ずしも知覚ラウドネスが均一であるという結果をもたらさない。

40

【0005】

DRCシステムも、ラウドネスコントロールシステムに関係するが、目標はわずかに異なる。DRCシステムでは、信号の長期的な平均レベルが既に期待レベルに正規化されている前提で、短期的なダイナミックスの修正だけを試みる。DRCシステムは、音の大きい事象が減衰され、音の静かな事象が増幅されるようにダイナミックスを圧縮することができる。これは、短期的な信号ダイナミックスを保ちながら信号の平均ラウドネスレベルを正規化するラウドネスコントロールシステムの目標と異なる。

【0006】

50

現代のラウドネスコントロールシステムでは、心理音響学及びラウドネス知覚の分野からの知見を組み込むことによって、AGC及びDRC設計への改善を試みる。ラウドネスコントロールシステムは、入力信号の知覚ラウドネスを推定することと、出力信号の平均ラウドネスレベルが所定の目標ラウドネスレベルに正規化されるように、時間的に変化する利得を制限することによって動作し得る。

【発明の概要】

【発明が解決しようとする課題】

【0007】

既存のラウドネスコントロールシステムの問題点は、所望のコンテンツと不要な雑音とが区別されず、所定の閾値を超える全ての低レベル音声コンテンツが増幅されるという点である。既存のラウドネスコントロールシステムで共通の問題となる信号は、例えば中程度の背景雑音を含む会話である。会話中に長い休止がある場合、ラウドネスコントロールシステムは、背景雑音の増幅を開始することがある。その結果、S/N比（信号対雑音比）が下がることとなり、リスナーによっては不愉快に感じることがある。ラウドネスコントロールシステムが雑音レベルの相対的増幅を回避し、これによって入力信号のS/N比を維持することが望ましい。

10

【0008】

ラウドネスコントロールシステムにおける別のチャレンジングなシナリオは、コンテンツ内の短期的な信号ダイナミクスを制限することによる悪影響を生じさせずに、均一な平均ラウドネスレベルを維持することである。ラウドネス変化に速く反応するシステムは、所望の目標レベルを一貫して達成することができるが、その代わりに短期的な信号ダイナミクスが低下する。他方でラウドネス変化にゆっくり反応するシステムは、ラウドネスレベルを効果的に制御できないか、又は入力信号のラウドネスレベルの大きな変化が生じている間にランピングなどの顕著なアーチファクト (artifacts) を示すことがある。大きくて長期的なラウドネス変化は、プログラム遷移又はコンテンツソースの変化などのコンテンツ間遷移のときに最も頻発する。ラウドネスコントロールシステム内においては、大きなコンテンツ間のラウドネス遷移を速く制御しつつコンテンツ内の短期的な信号ダイナミクスが維持されるように、コンテンツ間変動とコンテンツ内変動に個別に対処することが望ましい。

20

【課題を解決するための手段】

30

【0009】

ラウドネスコントロールシステム及び方法は、音声コンテンツを所定のラウドネスレベルに正規化することができる。音声コンテンツに中程度の背景雑音が含まれる場合、背景雑音も目標ラウドネスレベルに正規化され得る。雑音信号は、コンテンツ対雑音分類を使用して検出され、ラウドネスコントロールシステム又は方法は、雑音の検出結果に基づいて、入力信号のS/N比を維持するために調整され得る。雑音信号は、周波数領域又は時間領域内の信号分析によって検出され得る。さらにラウドネスコントロールシステムは、コンテンツが高い長期的なラウドネスレベルから低い長期的なラウドネスレベルに遷移するとき、望ましくない音声アーチファクトを発生し得る。そのようなラウドネス低下 (loudness drops) が検出される場合、ラウドネスコントロールシステムは、ラウドネスレベル間の遷移時に発生する望ましくないアーチファクトが最小になるように調整され得る。

40

【0010】

一実施形態によれば、ラウドネスコントロールシステムは、音声信号を処理するように構成され得る。ラウドネスコントロールシステムは、音声信号の短期的なラウドネス推定を生成するように構成されたラウドネス測定モジュールを備えることができる。ラウドネスコントロールシステムは、音声信号のコンテンツ対雑音分類を生成するように構成された雑音検出モジュールをさらに備えることができる。ラウドネスコントロールシステムは、コンテンツ対雑音分類結果に基づいて少なくとも1つの平滑化係数を調整し、短期的なラウドネス推定に基づいて、少なくとも1つの平滑化係数を使用して音声信号の長期的な

50

ラウドネス推定を生成するように構成された時間的平滑化モジュールをさらに備えることができる。ラウドネスコントロールシステムは、長期的なラウドネス推定に基づいて、時間的に変化する利得を音声信号に適用するように構成された利得補正モジュールをさらに備えることができる。雑音検出モジュールは、周波数領域雑音検出又は時間領域雑音検出を使用してコンテンツ対雑音分類結果を生じるように構成され得る。少なくとも1つの平滑化係数は、利得補正モジュールが利得レベルを高くすることができる速度を制御するリリース平滑化係数を含み得る。コンテンツ対雑音分類は、範囲[0, 1]上で正規化され得る。ラウドネスコントロールシステムは、ラウドネス低下検出値を生成するように構成されたラウドネス低下検出モジュールをさらに備えることができ、時間的平滑化モジュールは、ラウドネス低下検出値に基づいて少なくとも1つの平滑化係数を調整するように構成され得る。

10

## 【0011】

別の実施形態によれば、ラウドネスコントロールシステムは、音声信号を処理するように構成され得る。ラウドネスコントロールシステムは、音声信号の短期的なラウドネス推定を生成するように構成されたラウドネス測定モジュールを備えることができる。ラウドネスコントロールシステムは、ラウドネス低下検出値を生成するように構成されたラウドネス低下検出モジュールをさらに備えることができる。ラウドネスコントロールシステムは、ラウドネス低下検出値に基づいて少なくとも1つの平滑化係数を調整し、短期的なラウドネス推定に基づいて、少なくとも1つの平滑化係数を使用して音声信号の長期的なラウドネス推定を生成するように構成された時間的平滑化モジュールをさらに備えることができる。ラウドネスコントロールシステムは、長期的なラウドネス推定に基づいて、時間的に変化する利得を音声信号に適用するように構成された利得補正モジュールをさらに備えることができる。少なくとも1つの平滑化係数は、利得補正モジュールが利得レベルを高くすることができる速度を制御するリリース平滑化係数を含み得る。ラウドネス低下検出値は、範囲[0, 1]上で正規化され得る。ラウドネスコントロールシステムは、音声信号のコンテンツ対雑音分類を生成するように構成された雑音検出モジュールをさらに備えることができ、時間的平滑化モジュールは、コンテンツ対雑音分類に基づいて少なくとも1つの平滑化係数を調整するように構成され得る。

20

## 【0012】

別の実施形態によれば、システムは、周波数領域雑音検出を実行するように構成され得る。システムは、複数のチャネルを含む入力信号を受け取り、複数のチャネルを総和することによってモノ信号を生成するように構成された総和コンポーネントを備えることができる。システムは、短時間フーリエ変換(STFT)コンポーネントをさらに備えることができ、このコンポーネントは、STFTをモノ信号に適用することによって周波数領域信号を生成するように構成され得る。システムは、周波数領域信号に基づいてパワースペクトルを生成し、そのパワースペクトルをデシベル(dB)領域に変換するように構成されたデシベルコンバータをさらに備えることができる。システムは、パワースペクトルのそれぞれの周波数帯のエネルギーの時間平均を推定することによって、時間的に平滑化されたパワースペクトルを生成するように構成された時間的平滑化コンポーネントをさらに備えることができる。システムは、パワースペクトルと時間的に平滑化されたパワースペクトルとの平均差を計算することによって、パワースペクトルのスペクトル束値を計算するように構成されたスペクトル束測定コンポーネントをさらに備えることができる。システムは、サブバンドの相対エネルギーをその近傍と比較して測定することによってパワースペクトルのそれぞれのサブバンドの音調特性を推定し、それによってピークネス値を生成するように構成されたピークネス測定コンポーネントをさらに備えることができる。システムは、パワースペクトルのスペクトル束値、ピークネス値及びパワースペクトルに基づいて、雑音パワースペクトルを推定してS/N比を生成するように構成されたS/N比推定器コンポーネントをさらに備えることができる。システムは、S/N比に基づいて、平滑化されたS/N比を生成するように構成された時間的平滑化コンポーネントをさらに備えることができる。システムは、S/N比に基づいて入力信号に対するコンテンツ対雑

30

40

50

音分類値を生成するように構成されたヒステリシスコンポーネントをさらに備えることができる。S/N比推定器コンポーネントは、所望のコンテンツの成分であると推測される信号のオリジナルスペクトルから任意の時間的ダイナミクス又は音調成分を取り除くことによって、信号の雑音パワースペクトルを推定するように構成され得る。コンテンツ対雑音分類は、範囲[0, 1]上で正規化され得る。信号対雑音比推定器コンポーネントは、広帯域雑音レベル及び信号レベルを計算するように構成され得る。このシステムは、ラウドネスコントロールシステムに備えられるものとしてよく、ラウドネスコントロールシステムは、コンテンツ対雑音分類値に基づいて利得補正速度を調整するように構成された時間的平滑化コンポーネントを備えることができる。

#### 【0013】

別の実施形態によれば、システムは、時間領域雑音検出を実行するように構成され得る。システムは、複数のチャネルを含む入力信号を受け取り、複数のチャネルを総和することによってモノ信号を生成するように構成された総和コンポーネントを備えることができる。システムは、モノ信号を短期的なエンベロープ推定に変換するように構成された二乗平均平方根(RMS)コンポーネントをさらに備えることができる。システムは、短期的なエンベロープ推定に対してデシベル(dB)変換を実行するように構成されたデシベルコンバータをさらに備えることができる。システムは、短期的なエンベロープ推定の平均を取って長期的な平均エンベロープ推定を生成するように構成された平滑化フィルターをさらに備えることができる。システムは、短期的なエンベロープ推定から長期的な平均エンベロープ推定を減算してエンベロープ値を生成するように構成された減算コンポーネントをさらに備えることができる。システムは、エンベロープ値を半波整流するように構成された半波整流器コンポーネントをさらに備えることができる。システムは、エンベロープ値に基づいてオンセットエネルギーの平均及びオフセットエネルギーの平均を推定するように構成された少なくとも2つの平滑化フィルターをさらに備えることができる。システムは、オンセットエネルギーの平均とオフセットエネルギーの平均との間の正規化平方誤差を計算するように構成された正規化誤差計算器をさらに備えることができる。システムは、この正規化平方誤差を時間的に平滑化するように構成された時間的平滑化コンポーネントをさらに備えることができる。システムは、平滑化された正規化平方誤差にヒステリシスを適用してコンテンツ対雑音分類を生成するように構成されたヒステリシスコンポーネントをさらに備えることができる。平滑化フィルターは、短期的なエンベロープ推定の指数移動平均(EMA)を取るように構成され得る。時間的平滑化コンポーネントは、信号に依存する平滑化係数を使用する。平滑化係数は、アタックとリリースが異なる特性を有する。コンテンツ対雑音分類は、範囲[0, 1]上で正規化される。請求対象のシステムは、ラウドネスコントロールシステムに備えられるものとしてよく、ラウドネスコントロールシステムは、コンテンツ対雑音分類値に基づいて利得補正速度を調整するように構成された時間的平滑化コンポーネントを含むことができる。

#### 【0014】

別の実施形態によれば、システムは、ラウドネス低下検出を実行するように構成され得る。システムは、入力信号を受け取り、入力信号に基づいて短期的なラウドネス推定を計算するように構成された短期的ラウドネス測定モジュールを備えることができる。システムは、低速平滑化ラウドネス推定及び高速平滑化ラウドネス推定を計算するように構成された少なくとも2つの時間的平滑化フィルターをさらに備えることができる。システムは、低速平滑化ラウドネス推定から高速平滑化ラウドネス推定を減算して差分値を生成するように構成された減算モジュールをさらに備えることができる。システムは、差分値を半波整流して整流された差分値を生成するように構成された半波整流器モジュールをさらに備えることができる。システムは、整流された差分値を正規化して低下検出値を生成するように構成された正規化モジュールをさらに備えることができる。短期的ラウドネス測定モジュールは、ITU-R BS.1770ラウドネス尺度を使用して短期的なラウドネス推定を計算するように構成され得る。少なくとも2つの時間的平滑化フィルターは、低速平滑化係数及び高速平滑化係数をそれぞれ使用するように構成されるものとしてよく、

10

20

30

40

50

低速及び高速平滑化係数は、入力信号のダイナミクスに基づいて動的に修正され得る。低速平滑化係数及び高速平滑化係数は、信号ダイナミクスの高い尺度で入力信号に対して相互に低速化することができる。低速平滑化係数及び高速平滑化係数は、信号ダイナミクスの低い尺度で入力信号に対して相互に高速化することができる。正規化モジュールは、平行移動、拡大縮小及び飽和を使用して低下検出値を計算することができる。正規化モジュールは、[ 0 , 1 ] の範囲内で低下検出値を生成するように構成されるものとしてよく、1の低下検出値は、ラウドネス低下が検出されたことを示し、0の低下検出値は、ラウドネス低下が検出されなかったことを示す。このシステムは、ラウドネスコントロールシステムに備えられるものとしてよく、ラウドネスコントロールシステムは、低下検出値に基づいて利得補正速度を調整するように構成された時間的平滑化コンポーネントを含むことができる。

10

【図面の簡単な説明】

【 0 0 1 5 】

【図 1】入力音波が音声処理システムを通過して出力音波を発生するブロック図である。

【図 2】ラウドネスコントロールシステムのブロック図である。

【図 3】一実施形態による、周波数領域雑音検出システムのブロック図である。

【図 4 A】音楽の短いセグメントのパワースペクトルを示す図である。

【図 4 B】雑音の短いセグメントのパワースペクトルを示す図である。

【図 4 C】信号の音調及び過渡的構造が取り除かれている、音楽のセグメントに対する雑音パワースペクトルの推定を示す図である。

20

【図 4 D】信号の音調及び過渡的構造が取り除かれている、雑音のセグメントに対する雑音パワースペクトルの推定を示す図である。

【図 4 E】音楽のセグメントとその後に続く雑音のセグメントとからなる信号に対する周波数領域雑音検出システムからのコンテンツ対雑音分類出力を示す図である。

【図 5】一実施形態による、時間領域雑音検出システムのブロック図である。

【図 6 A】コンテンツ対雑音遷移上の信号エンベロープ及び平滑化された信号エンベロープを示す図である。

【図 6 B】図 6 A の信号に対応する時間領域雑音検出システムからの例示的な分類出力を示す図である。

【図 7】一実施形態による、雑音検出を行うラウドネスコントロールシステムのブロック図である。

30

【図 8】一実施形態による、ラウドネス低下検出システムのブロック図である。

【図 9】ラウドネス低下検出システム内の短期的なラウドネス推定、2つの平滑化されたフィルター出力、及び結果として得られるラウドネス低下検出信号に対する例示的な信号を d B 単位で示す図である。

【図 10 A】ラウドネス低下検出システムにおける異なる平滑化係数選択に対する、短期的なラウドネス推定、平滑化されたフィルター出力、及びラウドネス低下検出信号の例をそれぞれ示す図である。

【図 10 B】ラウドネス低下検出システムにおける異なる平滑化係数選択に対する、短期的なラウドネス推定、平滑化されたフィルター出力、及びラウドネス低下検出信号の例をそれぞれ示す図である。

40

【図 10 C】ラウドネス低下検出システムにおける異なる平滑化係数選択に対する、短期的なラウドネス推定、平滑化されたフィルター出力、及びラウドネス低下検出信号の例をそれぞれ示す図である。

【図 10 D】ラウドネス低下検出システムにおける異なる平滑化係数選択に対する、短期的なラウドネス推定、平滑化されたフィルター出力、及びラウドネス低下検出信号の例をそれぞれ示す図である。

【図 11】一実施形態による、動的平滑化係数を有するラウドネス低下検出システムのブロック図である。

【図 12 A】ラウドネス低下検出システムにおける、動的平滑化係数を有する、短期的な

50

ラウドネス推定、平滑化されたフィルター出力、及びラウドネス低下検出信号の例をそれぞれ示す図である。

【図12B】ラウドネス低下検出システムにおける、動的平滑化係数を有する、短期的なラウドネス推定、平滑化されたフィルター出力、及びラウドネス低下検出信号の例をそれぞれ示す図である。

【図13】一実施形態による、ラウドネス低下検出を行うラウドネスコントロールシステムのブロック図である。

【図14】一実施形態による、雑音検出及びラウドネス低下検出を行うラウドネスコントロールシステムのブロック図である。

【発明を実施するための形態】

【0016】

音波は、空気などの圧縮性媒質を伝搬する物体の振動によって引き起こされる一種の圧力波である。音波は、媒質（例えば空気）中の物質を周期的に変位し、物質を振動させる。音波の周波数は、1周期内の完全なサイクルの数を記述し、ヘルツ（Hz）で表される。12 Hz から 20,000 Hz の周波数範囲内の音波は、人間の耳に聞こえる。

【0017】

図1は、入力音波105が音声処理システムを通過して出力音波135を発生する流れ図100を示している。音声信号は、可聴音波を電圧として表したものである。例えばマイクロフォンなどのデバイス110は、機械的エネルギーである音圧波を受け取り、電気エネルギー又は音声信号115に変換する。同様に、スピーカー又はヘッドフォンなどのデバイス130は、電氣的音声信号125を可聴音波135に変換する。音声信号処理ブロック120は、音声信号の可聴効果を変えるために音声信号115を意図的に操作する。音声信号処理は、アナログ又はデジタル領域において実行され得る。

【0018】

アナログ音声信号は、例えば電圧、電流、又は電荷変化の形態で電気回路を流れるデータの連続的な流れによって表される。アナログ信号処理（ASP）は、様々な電氣的手段を介して電圧、電流又は電荷を変化させることによって物理的に連続的信号を変える。デジタル音声信号は、アナログ音声信号のサンプリングを通じて生成され、この信号は、記号列、典型的には2進数の列として表されるので、信号処理のためにマイクロプロセッサ及びコンピュータなどのデジタル回路を使用することができる。この場合、処理は、信号のデジタル表現に対して実行される。ラウドネスコントロールは、音声信号処理の一例である。

【0019】

本明細書で説明されている実施形態は、ラウドネスコントロールシステム及び音声信号に適用される方法に関して説明されるが、これらの概念及び機能強化は、他の音声信号処理システム、例えばAGC及びDRCのシステム及び方法にも同様に適用可能であるのは当然である。ラウドネスコントロールシステムは、ラウドネスレベルが変化する入力音声信号を操作し、リスナーにとってより心地よい均一なラウドネスレベルを有する出力音声信号を発生する働きをし得る。

【0020】

本明細書で説明されている実施形態全体を通して、いくつかの表記法が使用される。信号 $x[n]$ は、サンプルインデックス $n$ 及びサンプルレート $F_{s_n}$ を有する時系列である。信号 $x[n]$ は、複数の音声チャネル $C$ からなり、 $c$ がチャネルインデックス $0 \leq c \leq C-1$ である特定のチャネルを指定するために $x_c[n]$ と表記される。信号 $x[m]$ は、 $x[m]$ のサンプルレートが $F_{s_m} = F_{s_n} / M$ となるように、 $M$ という因数によってダウンサンプリングされた時系列とする。

【0021】

ラウドネスコントロールシステム200の高水準ブロック図が図2に示されている。ラウドネスコントロールシステム200は、少なくともラウドネス測定モジュール205、時間的平滑化モジュール210及び利得補正モジュール215の3つのコアモジュールを

10

20

30

40

50

含み得る。ラウドネスコントロールシステム200は、入力音声信号 $x[n]$ を修正して、改善されたラウドネス特性を有する出力音声信号 $y[n]$ を生成することができる。例えばラウドネスコントロールシステム200は、図1の音声処理システム100内の音声処理ブロック120の一部とすることができる。

【0022】

図2を参照すると、ラウドネス測定モジュール205は、入力信号 $x[n]$ の短いセグメントを分析して短期的なラウドネス推定 $L_{short}[m]$ を生成できることがわかる。時間的平滑化モジュール210は、短期的なラウドネス推定を時間に関して平滑化することによって、長期的な平均ラウドネスレベル $L_{ave}[m]$ の推定を提供し得る。利得補正モジュール215は、時間的に変化する補間された利得を入力信号 $x[n]$ に適用することができ、この利得は、出力信号 $y[n]$ の長期的な平均ラウドネスレベルが所定の目標ラウドネスレベルに等しくなるように制御され得る。

10

【0023】

ラウドネス測定モジュール205は、音声信号の知覚ラウドネスを推定するために如何なるプロセスをも使用することができる。そのようなプロセスの例は以下のとおりである。

【0024】

・国際電気標準会議(IEC)によって定義されているようなA、B、又はC周波数重み付けと結び付けられ得るラウドネス等価尺度( $L_{eq}$ )。

【0025】

・国際標準化機構(ISO)によって定義されている規格の基盤となったツウィッカー及びファストルラウドネスモデル。

20

【0026】

・国際電気通信連合(ITU)によって定義されているような改訂された低周波B-重み付け(RLB)周波数重み付け及び前置フィルターに結び付けられている $L_{eq}$ 尺度。

【0027】

例えばITU勧告(ITU-R)BS.1770ラウドネス測定システムは、ラウドネスコントロールシステム200のラウドネス測定モジュール205で使用され得る。ITU-R BS.1770の方法は、高度テレビジョンシステムズ委員会及びヨーロッパ放送連合を含む放送業界によって広く採用されている国際規格である。ITU-R BS.1770の実装は、必要とされる計算能力及び記憶容量は一般的に低く、リスナーによるラウドネス知覚と適切に相関することが実証されている。

30

【0028】

ラウドネス測定モジュール205は、入力信号 $x[n]$ の短いセグメント、例えば5~10ミリ秒のセグメントの知覚ラウドネスを推定することができる。その結果得られる短期的なラウドネス推定 $L_{short}[m]$ は、例えばラウドネスコントロールの設計及び実装に応じて、振幅、エネルギー又はデシベル(dB)領域で表され得る。

【0029】

ラウドネスコントロールシステム200の目的は、短期的な信号ダイナミクスを過剰に圧縮することなく、均一な平均ラウドネスレベルを有する出力信号 $y[n]$ を生成することである。したがって時間的平滑化モジュール210は、信号の長期的な平均ラウドネスレベルの推定を得るために、短期的なラウドネス推定を時間に関して平均化又は平滑化することができる。短期的なラウドネス推定に対して時間的な平滑化を実行するための方法は、例えば以下の数式に従って単極指数移動平均(EMA)フィルターを適用することである。

40

【0030】



【数 1】

$$L_{ave}[m] = L_{ave}[m-1] \cdot (1 - \alpha) + L_{short}[m] \cdot \alpha$$

ただし  $L_{short}[m]$  は、短期的なラウドネス推定であり、 $L_{ave}[m]$  は、長期的な平均ラウドネス推定であり、 $\alpha$  は、時間的な平滑化の挙動を制御する平滑化係数である。

【0031】

時間的平滑化モジュール 210 は、異なる平滑化係数  $\alpha$  値を使用して、「アタック」及び「リリース」の挙動が別々に設計され得る。アタックフェーズは、前の平均ラウドネス推定  $L_{ave}[m]$  より大きい新たに取得された短期的なラウドネス推定  $L_{short}[m]$  を参照し得る。リリースフェーズは、前の平均ラウドネス推定  $L_{ave}[m]$  より小さい新たに取得された短期的なラウドネス推定  $L_{short}[m]$  を参照し得る。したがって以下のようになる。

【0032】

【数 2】

$$\alpha = \begin{cases} \alpha_{attack}, & L_{short}[m] > L_{ave}[m-1] \\ \alpha_{release}, & L_{short}[m] \leq L_{ave}[m-1] \end{cases}$$

アタック平滑化係数  $\alpha_{attack}$  及びリリース平滑化係数  $\alpha_{release}$  は、平均ラウドネスレベルの長期的推定に近似するように設定され、人間の聴覚系の非対称的なラウドネス形成に近似するように、アタック平滑化係数  $\alpha_{attack}$  がリリース平滑化係数  $\alpha_{release}$  より速い速度に設定され得る。

【0033】

アタックリリース平滑化係数のチューニングは、アプリケーション特有のものであり、出力ラウドネスレベルの一貫性に影響し得る。比較的ゆっくりとしたアタックリリース平滑化係数では、平均ラウドネス推定が信号ラウドネスレベルを追跡する速度は遅すぎ、その結果、出力ラウドネスレベルはかなり変動する可能性がある。比較的速いアタックリリース平滑化係数では、平均ラウドネス推定が短期的な信号ダイナミクスを追跡するのが厳密すぎて、その結果、出力信号  $y[n]$  は、一貫したラウドネスレベルを有するが、信号ダイナミクスが過度に圧縮され得る。

【0034】

ラウドネスコントロールシステム 200 は、静的雑音閾値  $T_{noise,static}$  を含むものとしてよく、この閾値より低い入力信号は不要な雑音と推測され、この閾値より高い入力信号は所望のコンテンツであると推測される。ラウドネスコントロールシステムは、雑音と推測するレベルに反応することを回避するように設計され、これにより雑音の不快な増幅は低減され得る。したがって雑音閾値  $T_{noise,static}$  より低い値を測定する短期的なラウドネス推定は、長期的な平均ラウドネス推定に含まれず、事実上、平均ラウドネス推定を前の値に「凍結する」ことができる。

【0035】

短期的なラウドネス推定  $L_{short}[m]$  が静的雑音閾値  $T_{noise,static}$  より低いときに平均ラウドネス推定を凍結する一方法では、時間的平滑化フィルターに条件を追加することができる。例えば以下のように  $\alpha$  をゼロに設定することによって、平均ラウドネス推定を事実上前の値に維持することができる。

【0036】

【数3】

$$\alpha = \begin{cases} \alpha_{attack}, & L_{short}[m] > L_{ave}[m-1] \\ \alpha_{release}, & T_{noise,static} < L_{short}[m] \leq L_{ave}[m-1] \\ 0, & L_{short}[m] \leq T_{noise,static} \end{cases}$$

これは、雑音と推測される低レベル信号への反応を回避するために使用することができる多くの方法のうちの1つに過ぎない。

10

【0037】

利得補正モジュール215は、所定の目標ラウドネスレベル $Tar_{dB}$ と平均ラウドネス推定 $L_{ave,dB}[m]$ との差を取ることによって、時間的に変化する利得値 $G_{dB}[m]$ を計算することができる。下付文字dBは、ラウドネス値がデシベル領域で表されることを示している。

【0038】

【数4】

$$G_{dB}[m] = Tar_{dB} - L_{ave,dB}[m]$$

20

サンプルレート $F_{s_m}$ でダウンサンプリングされた利得値 $G_{dB}[m]$ は、線形領域に変換され、サンプルレート $F_{s_n}$ で平滑化利得信号 $G[n]$ を生成するように補間される。補間法としては、限定はしないが、例えばEMA平滑化法、線形補間法又は三次補間法が挙げられる。出力信号 $y[n]$ は、以下のように利得値 $G[n]$ に入力信号 $x[n]$ を乗算することによって生成される。

【0039】

【数5】

$$y[n] = G[n] \cdot x[n]$$

30

ラウドネスコントロールシステムは、不要な雑音を相対的に増幅することによって、例えば中程度のレベルの背景雑音を含む会話等、特定の状況下でS/N比が低減し得る。図2を参照しつつ説明されているように、ラウドネスコントロールシステム200は、雑音と推測される信号の増幅を制限する単純な方法として静的雑音閾値 $T_{noise,static}$ を含み得る。入力信号ラウドネスが静的雑音閾値 $T_{noise,static}$ より低く測定された場合、推定された平均ラウドネスレベル $L_{ave}[m]$ 及び利得信号 $G[n]$ を凍結する。この凍結機構は、信号 $x[n]$ 内の実際の雑音レベルが静的雑音閾値 $T_{noise,static}$ より低い限り、S/N比を維持する許容可能なジョブを実行することができる。しかし雑音レベルが雑音閾値 $T_{noise,static}$ より高い場合、不要な雑音が増幅され得る。現実世界の雑音は、極めて大きく、また予測できない場合があるため、静的閾値との単純な比較よりも精密な解決策を必要とする。

40

【0040】

改善は、雑音及び雑音レベルを検出する高度な方法を通じて、ラウドネスコントロールシステムに対してなされ得る。音声のセグメントが所望のコンテンツなのか不要な雑音なのかの情報は、ラウドネスコントロールシステムにとって有用な情報となり得る。雑音検出の自動的な方法は、以下で説明されているように、音声のセグメントがコンテンツであるか雑音であるかを分類するために使用され得る。

50

## 【 0 0 4 1 】

不要な雑音の種類として、限定はしないが、例えば背景雑音、周辺雑音、環境雑音及びヒスノイズが挙げられる。不要な雑音の特性は、雑音を自動的に検出するために定義され得る。不要な雑音は、以下の特性を有するものとして定義され得る。

## 【 0 0 4 2 】

・定常：雑音の信号電力及びスペクトル形状は、時間に関して十分に静止していると推測される。

## 【 0 0 4 3 】

・低レベル：雑音は、所望のコンテンツに比べて十分にレベルが低いと推測される。

## 【 0 0 4 4 】

・広域 / 平坦スペクトル：雑音のスペクトルは、十分に広域で平坦な周波数であると推測される。著しいスペクトルの山や谷（例えば音調信号）を有する信号は、所望のコンテンツであると考えられる。

## 【 0 0 4 5 】

雑音検出システム又は方法は、上記推測の1つ又は複数を使用することができる。

## 【 0 0 4 6 】

雑音検出は、自明な作業ではなく、最適性能を得るために精密な分析を必要とする場合がある。一実施形態において、周波数領域雑音検出システムは、スペクトルの定常性及び広域性の推測を有効に使うことで正確な分類結果をもたらす。しかしラウドネスコントロールシステムは、計算能力及び電力が制約されている多くのアプリケーションにおいて必要とされる。これらのアプリケーションに関して、別の実施形態によれば、より効率的な時間領域雑音検出システムは、定常性の推測を有効に使う。

## 【 0 0 4 7 】

本明細書で説明されている雑音検出のための解決策は、「ソフト」なコンテンツ対雑音分類を生成することができる。この分類は、例えば範囲 [ 0 , 1 ] 上で定義されるものとしてよく、0は雑音を示し、1はコンテンツを示し、これらの間の値は、あまり信頼できない分類である。ソフトな判定は、雑音検出を積分するシステムに柔軟性をもたらす。

## 【 0 0 4 8 】

さらに本明細書で説明されている雑音検出システムは、レベル独立であるものとしてよい。言い換えると、入力信号に適用されるスカラーオフセットは、分類を変えないということである。コンテンツ及び雑音の予想されるレベルはアプリケーション毎に大きく異なるため、信号レベルに関して強い推測をすると、アプリケーションによっては性能の低下を引き起こし得ることから、これは重要な特性である。雑音検出システムは、レベル独立であるとしても、性能を安全に改善するために幾つかの注意レベルに依存するバイアスが含まれ得る。例えば非常に大きな信号（例えばフルスケール（d B F S）に関して - 1 2 から 0 デシベル）は、1 0 0 % の信頼性でコンテンツとして解釈され得る。同様に、妥当な静的雑音閾値より低い信号（例えば - 6 0 d B F S）は、1 0 0 % の信頼性で雑音とみなし得る。

## 【 0 0 4 9 】

一実施形態によれば周波数領域雑音検出は、雑音スペクトルを推定し、S / N比を計算することによって、信号をコンテンツ又は雑音として分類することができる。高いS / N比は、信号が主に所望のコンテンツからなることを示し、低いS / N比は、信号が主に雑音からなることを示す。雑音スペクトルは、所望のコンテンツの成分であると推測される信号のオリジナルスペクトルから時間的ダイナミクス又は音調成分を取り除く試みによって推定され得る。時間的ダイナミクスを推定するためにスペクトル束が使用され、音調成分を推定するためにピーキネス尺度が使用され得る。

## 【 0 0 5 0 】

一実施形態による、周波数領域雑音検出システム 3 0 0 のブロック図が図 3 に示されている。周波数領域雑音検出システム 3 0 0 は、音声信号  $x_c[n]$  を受け取り、m をインデックスとする分類推定クラス [ m ] を出力することができる。分類クラス [ m ] は、信

10

20

30

40

50

号がコンテンツ又は雑音である可能性が高いかどうかを示す。この分類は、例えば範囲 [ 0 , 1 ] 上で定義されるものとし、0 は雑音を示し、1 はコンテンツを示し、これらの間の値はあまり信頼できない分類とし得る。しかし他の分類範囲、例えば [ - 1 , 1 ] 又は [ 0 , 1 0 0 ] を使用することができる。

【 0 0 5 1 】

周波数領域雑音検出システム 3 0 0 は、チャンネル総和コンポーネント 3 0 5、短時間フーリエ変換 ( S T F T ) コンポーネント 3 1 0、デシベルコンバータ 3 1 5、時間的平滑化コンポーネント 3 2 0、スペクトル束測定コンポーネント 3 2 5、ピーキネス測定コンポーネント 3 3 0、S / N 比推定器コンポーネント 3 3 5、時間的平滑化コンポーネント 3 4 0、正規化コンポーネント 3 4 5 及びヒステリシスコンポーネント 3 5 0 のいずれかを含むことができる。周波数領域雑音検出システム 3 0 0 は、以下でさらに詳しく説明される。

10

【 0 0 5 2 】

チャンネル総和コンポーネント 3 0 5 は、C チャンネル信号  $x_c[n]$  の全てのチャンネル ( 低周波効果 ( L F E ) チャンネルが含まれるなら、場合によっては除く ) を総和して、以下のモノ信号を生成する。

【 0 0 5 3 】

【 数 6 】

$$x_{mono}[n] = \sum_{c=0}^{C-1} x_c[n]$$

20

ただし n はサンプル時間インデックス、c はチャンネルインデックス、C はチャンネルカウントであり、場合によっては L F E チャンネルを除く。チャンネル総和コンポーネント 3 0 5 は、計算効率を改善し、リソース要求を削減することができる。

【 0 0 5 4 】

モノ信号  $x_{mono}[n]$  は、S T F T コンポーネント 3 1 0 を適用する前に、オーバーラップする窓関数を掛け合わせたフレームに分割され得る。

30

【 0 0 5 5 】

【 数 7 】

$$X_{lin}[m, k] = \sum_{f=0}^{F-1} x_{mono}[f + mM] w[f] e^{-j(\frac{2\pi}{F})kf}$$

ただし M はサンプルホップサイズ、F はサンプル窓サイズ、m はダウンサンプリングされた時間インデックス、k は  $0 \leq k < K - 1$  の周波数インデックス、 $K = ( 0.5 \cdot F + 1 )$  は固有周波数インデックスの数、w は分析窓で例えば長さ F の H a n n 窓である。S T F T コンポーネント 3 1 0 の代わりに、他の種類のフィルターバンクコンポーネントを使用することができる。

40

【 0 0 5 6 】

デシベルコンバータ 3 1 5 は、S T F T コンポーネント 3 1 0 の出力  $X_{lin}[m, k]$  からパワースペクトルを計算して、それぞれのインデックス m 及び k についてパワースペクトルを d B 領域に変換することができる。

【 0 0 5 7 】

【数 8】

$$X[m, k] = 10 \cdot \log_{10} (|X_{lin}[m, k]|^2)$$

あるいは必要な計算量及び記憶容量を減らすために、S T F Tコンポーネント 3 1 0 の均一な間隔で並ぶパワースペクトルを結合して、等価方形帯域幅 ( E R B )、臨界帯域幅又は他の何らかの知覚帯域幅に近似するサブバンドにすることができる。

【 0 0 5 8 】

時間的平滑化コンポーネント 3 2 0 は、例えば時間に関する d B スペクトル  $X [ m , k ]$  の指数移動平均を使用して、それぞれの周波数帯のエネルギーの時間平均  $X' [ m , k ]$  を推定することができる。

【 0 0 5 9 】

【数 9】

$$X'[m, k] = X'[m-1, k] \cdot (1 - \alpha_s) + X[m, k] \cdot \alpha_s$$

ただし  $\alpha_s$  は、この数式に固有であり、望ましい平滑化特性を生成するために選択され得る平滑化係数である。

【 0 0 6 0 】

スペクトル束測定コンポーネント 3 2 5 は、時間に関するスペクトル変化の尺度であるスペクトル束  $s f [ m ]$  を測定する。雑音信号は、0に近い束を測定する定常スペクトルを有する傾向があり、コンテンツ信号は、高スペクトル束の短い持続時間を示すオンセット、オフセット及び過渡成分を含むより動的なスペクトルを有する傾向がある。スペクトル束値は、短期的なスペクトル  $X [ m , k ]$  と時間的に平滑化されたスペクトル  $X' [ m , k ]$  との間の平均差として計算され得る。時間的に平滑化されたスペクトルは、スペクトル束を計算するとき、現在のフレームスペクトルの積分を防ぐために、1 フレーム分遅延させることができる。

【 0 0 6 1 】

【数 1 0】

$$sf[m] = \frac{1}{K} \sum_{k=0}^{K-1} (X[m, k] - X'[m-1, k])$$

スペクトル束  $s f [ m ]$  は、d B 領域において計算されるので、その測定はレベル独立としてよく、線形領域で実行される束の計算とは異なり、さらなる正規化を要しない。

【 0 0 6 2 】

ピーキネス  $P [ m , k ]$  は、その近傍と比較した周波数帯の相対エネルギーを測定することによって、周波数帯の音調特性を推定する。ピーキネスは、例えば典型的なコンテンツは 2 0 H z から 6 k H z の範囲内で音調成分を含み得るため、制限された周波数帯の範囲にわたって推定され得る。ピーキネス測定コンポーネント 3 3 0 は、それぞれの周波数帯  $k$  の周囲の平均エネルギー  $P_{SE} [ m , k ]$  を最初に推定することによって、ピーキネスを計算することができる。

【 0 0 6 3 】

10

20

30

40

【数 1 1】

$$P_{SE}[m, k] = \frac{1}{2W} \left( -X'[m, k] + \sum_{r=\max(k-W, 0)}^{\min(k+W, K-1)} X'[m, r] \right)$$

ただし、 $2W$ は平均する隣接周波数帯の数である。

【0064】

隣接周波数帯  $P_{SE}[m, k]$  の平均エネルギーを中心周波数帯エネルギー  $X'[m, k]$  から減算することができる。 10

【0065】

【数 1 2】

$$P_{\text{delta}}[m, k] = X'[m, k] - P_{SE}[m, k]$$

$P_{\text{delta}}[m, k]$  の大きな正値は、中心周波数帯  $k$  内の音調成分の存在を示し、 $P_{\text{delta}}[m, k]$  の負値は、隣接周波数帯内の音調成分の存在を示し得る。音調成分を対象とする雑音検出アプリケーションでは、ピーキネスを計算するときの周波数帯漏れを補償するために、負値はゼロに設定され、正値は隣接周波数帯内に分散され得る。 20

【0066】

【数 1 3】

$$P[m, k] = \sum_{r=\max(k-W, 0)}^{\min(k+W, K-1)} \max(P_{\text{delta}}[m, r], 0)$$

30

S/N比推定器コンポーネント335は、ピーキネス  $P[m, k]$  及びスペクトル束  $sf[m]$  尺度を入力パワースペクトル  $X[m, k]$  から減算することによって、雑音パワースペクトル  $N[m, k]$  を推定することができる。

【0067】

【数 1 4】

$$N[m, k] = X[m, k] - P[m, k] - |sf[m]|$$

40

雑音スペクトルは、雑音レベル  $n_{\text{wide}}[m]$  の広帯域推定を計算するために、周波数に関して平均され得る。

【0068】

【数 1 5】

$$n_{\text{wide}}[m] = \frac{1}{K} \sum_{k=0}^{K-1} N[m, k]$$

50

さらに入力信号パワースペクトルは、信号レベル  $x_{wide}[m]$  の広帯域推定を計算するために、周波数に関して平均され得る。

【 0 0 6 9 】

【 数 1 6 】

$$x_{wide}[m] = \frac{1}{K} \sum_{k=0}^{K-1} X[m, k]$$

10

S / N 比推定器コンポーネント 3 3 5 は、推定された広帯域雑音レベル  $n_{wide}[m]$  を推定された広帯域信号レベル  $x_{wide}[m]$  から減算することによって、 $snr[m]$  を計算することができる。

【 0 0 7 0 】

【 数 1 7 】

$$snr[m] = x_{wide}[m] - n_{wide}[m]$$

20

その結果得られる S / N 比、 $snr[m]$  は、非常に分散が高い可能性がある。そのため時間的平滑化コンポーネント 3 4 0 は、指数移動平均フィルターを  $snr[m]$  に適用して分散を低減し、より大きな S / N 比傾向を捕らえて平滑化された S / N 比、 $smr'[m]$  を生成し得る。

【 0 0 7 1 】

【 数 1 8 】

$$smr'[m] = smr'[m-1] \cdot (1-\alpha) + snr[m] \cdot \alpha, \text{ where } \alpha = \begin{cases} \alpha_{attack,snr}, & snr[m] > smr'[m-1] \\ \alpha_{release,snr}, & snr[m] \leq smr'[m-1] \end{cases}$$

30

平滑化係数  $\alpha_{attack,snr}$  及び  $\alpha_{release,snr}$  は、時間的平滑化コンポーネント 3 4 0 で実行される平滑化 S / N 比計算に固有であり、望ましい平滑化特性を生み出すように選択され得る。

【 0 0 7 2 】

平滑化された S / N 比の値  $smr'[m]$  は、正規化コンポーネント 3 4 5 によって中間分類値  $c[m]$  に変換され得る。例えばその値は、0 が雑音を示し、1 がコンテンツを示し、間の値はあまり信頼できない分類となるように、dB - 線形領域変換、並びに拡大縮小及び平行移動を介して範囲 [ 0 , 1 ] に正規化され得る。

【 0 0 7 3 】

【 数 1 9 】

$$c[m] = 1 - 10^{\frac{-smr'[m]}{20}}$$

40

ヒステリシスコンポーネント 3 5 0 は、ヒステリシスのモデルを適用することによって最終分類結果を計算することができる。ヒステリシスモデルは、過去の分類に基づいて最終分類にバイアスをかける。より高いコンテンツ閾値  $T_{content}$  及びより低い雑音閾値  $T_{noise}$  の 2 つの閾値が使用され、これらの閾値は、スカラーバイアス計算 ( 数式 2 1 ) に固

50

有のものであってよい。中間分類値  $c[m]$  がコンテンツ閾値  $T_{\text{content}}$  を超える場合、最終分類クラス  $[m]$  は、より低い雑音閾値  $T_{\text{noise}}$  よりも中間分類値  $c[m]$  が低くなるまでコンテンツ分類の方へバイアスされ、これによって再びコンテンツ閾値を横切るまで最終分類クラス  $[m]$  が雑音分類の方へバイアスされ得る。

【0074】

【数20】

$$\text{class}[m] = \text{saturate}(c[m] \cdot \beta[m]),$$

10

ただし、

【0075】

【数21】

$$\beta[m] = \begin{cases} \beta_{\text{content}}, & c[m] \geq T_{\text{content}} \\ \beta_{\text{noise}}, & c[m] \leq T_{\text{noise}} \\ \beta[m-1], & T_{\text{noise}} < c[m] < T_{\text{content}} \end{cases}$$

20

及び

【0076】

【数22】

$$\text{saturate}(x) = \begin{cases} x, & 0 \leq x \leq 1 \\ 1, & x > 1 \\ 0, & x < 0 \end{cases}$$

30

数式20～22について、クラス  $[m]$  は最終分類結果であり、 $\beta_{\text{content}}$  は、例えば1より大きい値として選択され得る正のバイアススカラーであり、 $\beta_{\text{noise}}$  は、例えば1より小さい値として選択され得る正のバイアススカラーである。

【0077】

3.5秒分のコンテンツ（調性音楽）とその後に続く3.5秒分の広帯域雑音からなる信号は、図3の周波数領域雑音検出システム300の動作を図解するために使用されている。図4A～4Dは、周波数帯が等価方形帯域幅（ERB）に変換された信号全体を通して、複数の時間インデックスにおけるパワースペクトルを示している。図4Aは、音楽のパワースペクトル  $X[m, k]$  を示しており、時間インデックス  $m_0$  は、スペクトル内の顕著なピークとして見られる音楽の音調の態様を示すために、信号の最初の3.5秒分で任意に選択されたものである。図4Bは、雑音のパワースペクトル  $X[m_1, k]$  を示しており、時間インデックス  $m_1$  は、顕著なピークを欠いた比較的平坦なスペクトルとして見られる広帯域雑音中の音調構造の欠如を示すために、信号の後半部で任意に選択されたものである。図4Cは、図4Aと同じ時間インデックスにおいて、音楽の推定された雑音パワースペクトル  $N[m_0, k]$  を示している。図4Aにおいてピークとして見られた音調構造は、図4Cの推定された雑音パワースペクトルでは除去されており、これは図3で説明されているピーキネス測定コンポーネント330によって行われる。また図3のスペクトル束測定コンポーネント325で行われ得るように、過渡構造も図4Cで除去されている。しかしながら単一の時間インデックスしか示されていないので、過渡構造の除去は、図4A及び4Cに明確には現れていない。図4Dは、図4Bと同じ時間インデックスに

40

50



おける雑音の推定された雑音パワースペクトル  $N[m_1, k]$  を示している。予想通り、図 4 D の推定された雑音パワースペクトルは、図 4 B の雑音の真のパワースペクトルとほぼ同一である。図 4 E は、図 3 で説明されているように、3.5 秒分の調性音楽とその後続く 3.5 秒分の広帯域雑音とからなる完全な信号に対し、周波数領域雑音検出システム 300 が出力するコンテンツ対雑音分類を示している。この例示的なシナリオでは、0 の分類は雑音を示し、1 の分類はコンテンツを示す。図 4 E に図示されているように、図 3 の周波数領域雑音検出システム 300 は、信号内のコンテンツから雑音への遷移を 1 秒以内に検出することが可能である。

【0078】

別の実施形態によれば、雑音検出は、時間領域内で実行され得る。時間領域雑音検出システム又は方法は、計算能力に対する要件が低いことが望ましいシナリオにおいて使用することができる。時間領域雑音検出システムは、典型的な雑音信号が時間に関して適度な定常性を示す信号電力を有し、典型的なコンテンツ信号が時間的に変化するダイナミックスを示す信号電力を有するという推測を有効に使うことができる。

【0079】

一実施形態による、時間領域雑音検出システム 500 のブロック図が図 5 に示されている。時間領域雑音検出システム 500 は、音声信号  $x_c[n]$  を受け取り、 $m$  をインデックスとする分類推定クラス  $[m]$  を出力することができる。そのような分類クラス  $[m]$  は、信号がコンテンツ又は雑音である可能性が高いかどうかを示す。この分類は、例えば範囲  $[0, 1]$  上で定義されるものとしてよく、0 は雑音を示し、1 はコンテンツを示し、間の値は、あまり信頼できない分類である。しかし他の分類値も使用できる。

【0080】

時間領域雑音検出システム 500 は、チャンネル総和コンポーネント 505、二乗平均平方根 (RMS) コンポーネント 510、デシベルコンバータ 515、時間的平滑化フィルタ 520、減算コンポーネント 525、半波整流コンポーネント 530、時間的平滑化コンポーネント 535 及び 540、正規化誤差計算器 545、時間的平滑化コンポーネント 550 並びにヒステリシスコンポーネント 555 のいずれかを備えることができる。時間領域雑音検出システム 500 は、以下でさらに詳しく説明される。

【0081】

チャンネル総和コンポーネント 505 は、 $C$  チャンネル信号  $x_c[n]$  の全てのチャンネル (低周波効果 (LFE) チャンネルが含まれるなら、場合によっては除く) を総和して以下のモノ信号を発生する。

【0082】

【数 23】

$$x_{mono}[n] = \sum_{c=0}^{C-1} x_c[n]$$

ただし  $n$  はサンプル時間インデックス、 $c$  はチャンネルインデックス、 $C$  はチャンネルカウントであり、場合によっては LFE チャンネルを除く。チャンネル総和コンポーネント 505 は、計算効率を改善し、要求されるリソースを削減することができる。

【0083】

二乗平均平方根 (RMS) コンポーネント 510 は、 $F$  個のサンプルの窓で二乗平均平方根 (RMS) を計算することによって、入力信号を線形領域の短期的なエンベロープ推定  $env_{lin}[m]$  に変換することができる。

【0084】

10

20

30

40

【数 2 4】

$$env_{lin}[m] = \frac{1}{F} \sum_{f=0}^{F-1} (x_{mono}^2[f + mM])$$

線形領域の短期的なエンベロープ推定  $env_{lin}[m]$  は、デシベルコンバータコンポーネント 5 1 5 を介して、dB 領域の短期的なエンベロープ推定  $env[m]$  に変換され得る。

【0 0 8 5】

【数 2 5】

$$env[m] = 10 \cdot \log_{10}(env_{lin}[m])$$

入力信号の短期的なエンベロープを推定するために、他のエンベロープ推定器又は技術も使用できることに留意されたい。信号エンベロープは、コンテンツと雑音とを区別するのに役立つ。典型的な雑音信号の短期的なエンベロープは、長期的なエンベロープの平均を中心とする対称性を示す傾向があり、典型的なコンテンツ信号の短期的なエンベロープは、かなり不規則又は非対称になる傾向がある。

【0 0 8 6】

時間的平滑化コンポーネント 5 2 0、例えば単極指数移動平均 (EMA) 平滑化フィルターは、長期的な平均エンベロープ推定  $env'[m]$  を生成するために、短期的なエンベロープ推定  $env[m]$  に適用され得る。

【0 0 8 7】

【数 2 6】

$$env'[m] = env'[m-1] \cdot (1 - \alpha_{env}) + env[m] \cdot \alpha_{env}$$

ただし平滑化係数  $\alpha_{env}$  は、長期的な平均エンベロープ推定  $env'[m]$  の計算に固有であり、望ましい平滑化特性を生成するように選択され得る。

【0 0 8 8】

減算コンポーネント 5 2 5 は、短期的なエンベロープ値から長期的な平均エンベロープ推定を減算することによって、エンベロープデルタ値を計算することができる。

【0 0 8 9】

【数 2 7】

$$env_{delta}[m] = env[m] - env'[m]$$

半波整流コンポーネント 5 3 0 は、正の半波整流をエンベロープデルタ値に適用することができ、負の値がゼロに設定され、信号中の短期的なオンセットエネルギーの推定を与えることができる。

【0 0 9 0】

10

20

30

40

【数 2 8】

$$onset[m] = \max(env_{delta}[m], 0)$$

時間的平滑化コンポーネント 5 3 5 は、オンセットエネルギーの長期的平均を推定するために、オンセットエネルギーに適用され得る。

【0 0 9 1】

【数 2 9】

$$onset'[m] = onset'[m-1] \cdot (1 - \alpha_{onset}) + onset[m] \cdot \alpha_{onset}$$

10

ただし平滑化係数  $\alpha_{onset}$  は、数式 2 9 の計算に固有であり、望ましい平滑化特性を生成するように選択され得る。

【0 0 9 2】

半波整流コンポーネント 5 3 0 は、負の半波整流をエンベロープデルタ値に適用することもでき、正の値がゼロに設定され、絶対値を取ることで、信号中の短期的なオフセットエネルギーの推定を生成することができる。

【0 0 9 3】

【数 3 0】

20

$$offset[m] = |\min(env_{delta}[m], 0)|$$

時間的平滑化コンポーネント 5 4 0 は、オフセットエネルギーの長期的平均を推定するために、オフセットエネルギーに適用され得る。

【0 0 9 4】

【数 3 1】

$$offset'[m] = offset'[m-1] \cdot (1 - \alpha_{offset}) + offset \cdot \alpha_{offset}$$

30

ただし平滑化係数  $\alpha_{offset}$  は、数式 3 1 の計算に固有であり、望ましい平滑化特性を生成するように選択され得る。

【0 0 9 5】

オンセット平均エネルギー  $onset'[m]$  及びオフセット平均エネルギー  $offset'[m]$  は、典型的な雑音信号についてはレベルが類似しており、典型的なコンテンツ信号については顕著な差があり得る。正規化誤差計算器 5 4 5 は、オンセット平均エネルギーとオフセット平均エネルギーとの間の平方誤差  $err[m]$  を計算し、それを例えば平均エネルギーの最大値で除算することによって、0 から 1 の間に正規化することができる。

40

【0 0 9 6】

【数 3 2】

$$err[m] = \left( \frac{onset'[m] - offset'[m]}{\max(onset'[m], offset'[m])} \right)^2$$

50

例えばコンテンツ信号の時間構造が不規則であると、 $err[m]$ は1に向かう傾向となり、定常雑音の時間構造が欠如していると、 $err[m]$ は0に向かう傾向となり得る。

【0097】

時間的平滑化コンポーネント550は、分散を小さくするために平方誤差 $err[m]$ を時間的に平滑することによって、コンテンツ対雑音分類の生成を補助することができる。

【0098】

【数33】

$$err'[m] = err'[m-1] \cdot (1 - \alpha_{err}) + err[m] \cdot \alpha_{err}$$

10

平滑化係数 $\alpha_{err}$ は、アタック平滑化係数 $\alpha_{attack, err}$ 及びリリース平滑化係数 $\alpha_{release, err}$ によって決定され、アタックとリリースが異なる特性を生成するために信号依存であってよい。

【0099】

【数34】

$$\alpha_{err} = \begin{cases} \alpha_{attack, err} & err[m] > class[m-1] \\ \alpha_{release, err} & err[m] \leq class[m-1] \end{cases}$$

20

時間領域雑音検出システム500内で使用されるアタック平滑化係数 $\alpha_{attack, err}$ 及びリリース平滑化係数 $\alpha_{release, err}$ は、数式34に固有のものであり、例えば図2のラウドネスコントロールシステムの時間的平滑化モジュール210によって使用されるものよりも速い場合がある。これによって雑音検出システムは、ラウドネスコントロールシステムがレベルを補正するよりも速く、コンテンツ又は雑音として信号を分類することができる。

30

【0100】

図5を参照すると、ヒステリシスコンポーネント555は、図3の周波数領域雑音検出システム300のヒステリシスコンポーネント350と似た方法で、ヒステリシスのモデルを $err'[m]$ に適用することによって、最終的なコンテンツ対雑音分類クラス $[m]$ を計算することができる。

【0101】

図6Aは、最初の半分が音楽のセグメントで後の半分が雑音のセグメントであるコンテンツ対雑音遷移からなる信号について、エンベロープ $env$ 及び平滑化されたエンベロープ $env'$ をdB単位で示している。図6Aに図示されているように、0からおおよそ3.5秒までのエンベロープ信号の前半部は、長期的な平均エンベロープ $env'$ に対する短期的なエンベロープ $env$ の不規則性を示し、3.5から7秒までの後半部は、長期的な平均エンベロープ $env'$ に対する短期的なエンベロープ $env$ の対称性を示している。図6Bは、図6Aの信号に対応する図5の時間領域雑音検出システム500からの例示的なコンテンツ対雑音分類出力を図示しており、0は雑音、1はコンテンツを示す。

40

【0102】

例えば図3の周波数領域雑音検出システム300によって生成され、或いは図5の時間領域雑音検出システム500によって生成されるような雑音検出分類結果クラス $[m]$ は、図2のラウドネスコントロールシステム200などのラウドネスコントロールシステムに組み込まれ得る。

【0103】

50

例えば図7は、一実施形態により、雑音検出モジュール720をラウドネスコントロールシステム700に組み込んだ高水準ブロック図である。ラウドネスコントロールシステム700は、ラウドネス測定モジュール705、雑音検出モジュール720、時間的平滑化モジュール710及び利得補正モジュール715を含むことができる。ラウドネス測定モジュール705及び利得補正モジュール715は、図2で説明されているラウドネス測定モジュール205及び利得補正モジュール215と同様の動作をし得る。雑音検出モジュールは、雑音検出技術を使用して、図3及び5の周波数領域及び時間領域アプローチをそれぞれ含むコンテンツ対雑音分類結果クラス[m]の生成を行うことができる。次いで時間的平滑化モジュール710は、以下で説明されているように、雑音検出モジュール720からの最終分類出力クラス[m]を考慮することができる。

10

## 【0104】

ラウドネスコントロールシステム700の時間的平滑化モジュール710は、図2のラウドネスコントロールシステム200の時間的平滑化モジュール210と同様に、別々の「アタック」及び「リリース」平滑化係数を備えることができる。リリース平滑化係数  $\alpha_{release}$  は、ラウドネスコントロールで利得レベルを高めるときの許容速度を制御し得る。高速なリリース平滑化係数  $\alpha_{release}$  値は、ラウドネスコントロールで利得レベルを速く高めることを可能にし、低速なリリース平滑化係数  $\alpha_{release}$  値は、利得レベルを高める速度を制約し得る。極端な場合、リリース平滑化係数は、ラウドネスコントロールを凍結し、利得レベルの上昇が事実上生じないようにするために、0に設定され得る。

20

## 【0105】

信号がコンテンツ又は雑音からなるかどうかに関する情報が欠如している場合、図2のラウドネスコントロールシステム200は、所望のコンテンツ及び不要な雑音に対する利得レベルを同じ速度で高めるように強制され得る。しかし図7のラウドネスコントロールシステム700は、信号がコンテンツ又は雑音からなるかどうかに関する情報を有している場合、仮に不要な雑音あったとしても、不要な雑音の利得レベルは著しく遅い速度で高めつつ所望のコンテンツの利得レベルは速い速度で高くする改善された決定を下すことができる。

## 【0106】

一実施形態において、雑音依存利得レベルは、雑音検出モジュール720から受け取ったコンテンツ対雑音分類クラス[m]に基づいて、時間的平滑化モジュール710でリリース平滑化係数値  $\alpha_{release}$  を動的に修正することによって実装され得る。

30

## 【0107】

雑音検出モジュール720が高い信頼度で所望のコンテンツとして信号を検出したとき、 $\alpha_{release}[m]$  値は、利得レベル増加の既定速度に対応する所定の値  $\alpha_{release,def}$  に設定され得る。高い信頼度で不要な雑音として信号が検出されたとき、 $\alpha_{release}[m]$  値は、利得レベルの増加を事実上生じさせないように0に設定され得る。それに加えて雑音検出の「ソフト」分類が使用される場合、信頼度の低い雑音検出は、雑音検出の信頼度に比例して利得レベルの増加を遅くしてよい。例えば範囲[0, 1]上でソフト分類を使用すると、クラス[m] = 0.5の雑音分類結果は、信号がコンテンツであることの信頼度が50%及び信号が雑音であることの信頼度が50%あることを示し得る。この場合、 $\alpha_{release}[m]$  値は、既定値と0との間の補間値に設定され、それによって利得レベルを上昇させることができる速度を中間量で制約することができる。

40

## 【0108】

## 【数35】

$$\alpha_{release}[m] = \alpha_{release,def} \cdot class[m]$$

不要な雑音に対して利得レベルの増加を許可しないことは、入力信号  $x[n]$  のS/N比を維持する効果をもたらし得る。例えばコンテンツ対雑音遷移の間において、雑音レベ

50

ルがコンテンツレベルより低い場合、ラウドネスコントロールシステム700は、等しい利得レベルをコンテンツ及び雑音の両方のセグメントに適用することができ、それは利得レベルが雑音信号について増加するのが妨げられるからである。したがって入力信号中に存在するコンテンツと雑音の相対的なレベルは、出力信号中でも維持される。

#### 【0109】

S/N比の維持は、コンテンツ対雑音分類で達成され得る唯一の機能向上ではない。雑音抑圧などの他の機能向上も同様に、雑音信号が検出されたときに相対的減衰を適用することによって、ラウドネスコントロールの範囲内で実現することができる。

#### 【0110】

別の実施形態によれば、ラウドネス低下検出システム又は方法は、図2のラウドネスコントロールシステム200などのラウドネスコントロールシステムにおいて、利得補正速度を動的に修正するために使用され得る。

#### 【0111】

ラウドネスコントロールシステム200の設計目標は、オリジナルの信号ダイナミクスを維持しながら長期的なラウドネスレベルを正規化することであるとしてよい。しかしコンテンツ内ダイナミクスを制限することによる悪影響を生じさせることなく、コンテンツ間遷移に起因する大きなラウドネス低下を制御することは簡単ではない。大きな長期的なラウドネス低下の後、速く回復するには、十分に高速な時定数を使用して、時間的平滑化モジュール210のリリース平滑化係数  $\tau_{\text{release}}$  が計算され得る。しかし短期的な信号ダイナミクスを維持するには、十分に低速な時定数を使用して、リリース平滑化係数  $\tau_{\text{release}}$  が計算され得る。これらの相反する目標に対処するために、ラウドネス低下検出モジュールは、動的かつ信号に依存する方法でリリース平滑化係数  $\tau_{\text{release}}$  を修正すべく、ラウドネスコントロールシステム200に含まれ得る。

#### 【0112】

一実施形態によれば、ラウドネス低下検出システムは、短期的な信号ダイナミクスに起因するラウドネス変動が生じている間の検出を回避しつつ、大きな長期的なラウドネス低下を確実に検出することができる。図8は、一実施形態によるラウドネス低下検出システム800のブロック図を示している。図8のラウドネス低下検出システム800は、音声信号  $x[n]$  を受け取り、 $m$  をインデックスとして、時間的に変化するラウドネス低下検出推定  $\text{drop}[m]$  を出力することができる。このような  $\text{drop}[m]$  は、顕著なラウドネスレベルの低下が生じているかどうかを示す。ラウドネス低下検出推定  $\text{drop}[m]$  は、例えば範囲  $[0, 1]$  上で定義され、0は、ラウドネス低下がないことを示し、1は、大きなラウドネス低下がちょうど生じたばかりであることを示し、これらの間の値は、より小さな又はより穏やかなラウドネス低下を示す指標である。しかし他の低下検出値も使用できる。

#### 【0113】

ラウドネス低下検出システム800は、短期的ラウドネス測定モジュール805、時間的平滑化コンポーネント810及び815、減算モジュール820、半波整流モジュール825並びに正規化モジュール830のいずれかを含み得る。

#### 【0114】

短期的ラウドネス測定モジュール805は、図2のラウドネスコントロールシステム200のラウドネス測定モジュール205と同様に、短期的なラウドネス推定を計算することができる。短期的ラウドネス測定モジュール805は、例えば既に説明したITU-R BS.1770ラウドネス尺度又はRMSを含むラウドネス測定技術を使用することができる。現在のダウンサンプリングされたインデックス  $m$  で計算された短期的なラウドネス推定は、 $L_{\text{short, dB}}[m]$  で表すことができる。

#### 【0115】

時間的平滑化コンポーネント810及び815は、時間的な平滑化を短期的なラウドネス推定  $L_{\text{short, dB}}[m]$  に適用することができる。時間的平滑化コンポーネント810及び815は、例えば異なる平滑化係数を持つ2つの指数移動平均(EMA)フィルタート

10

20

30

40

50

してよい。時間的平滑化コンポーネント 8 1 0 及び 8 1 5 の各々は、比較的低速な平滑化係数  $\alpha_{slow}$  及び比較的高速な平滑化係数  $\alpha_{fast}$  を使用して、平滑化ラウドネス推定  $\mu_{slow}[m]$  及び  $\mu_{fast}[m]$  をそれぞれ計算することができる。

【 0 1 1 6 】

【数 3 6】

$$\mu_{slow}[m] = \mu_{slow}[m-1] \cdot (1 - \alpha_{slow}) + L_{short,dB}[m] \cdot \alpha_{slow}$$

【 0 1 1 7 】

【数 3 7】

$$\mu_{fast}[m] = \mu_{fast}[m-1] \cdot (1 - \alpha_{fast}) + L_{short,dB}[m] \cdot \alpha_{fast}$$

平滑化されたラウドネス推定  $\mu_{slow}[m]$  及び  $\mu_{fast}[m]$  は、異なる速度でラウドネスダイナミクスを追跡することができる。 $\mu_{slow}[m]$  の目標は、例えば話し言葉の間の休止のような短期的なダイナミックを追跡することなく、ラウドネス推定の長期的平均に追従することとしてよい。 $\mu_{fast}[m]$  の目標は、より速くラウドネス推定の平均を追跡し、 $\mu_{fast}[m]$  のレベルが  $\mu_{slow}[m]$  に比べて十分に低いときにおけるラウドネス低下を推測できることとしてよい。減算モジュール 8 2 0 は、入力信号のラウドネスの変化を捕捉するために、平滑化ラウドネス推定  $\mu_{slow}[m]$  と  $\mu_{fast}[m]$  との差分信号  $diff[m]$  を計算し得る。

【 0 1 1 8 】

【数 3 8】

$$diff[m] = \mu_{slow}[m] - \mu_{fast}[m]$$

例えば正の  $diff[m]$  値は、ラウドネス低下を示し、負の値は、ラウドネス増加を示し得る。半波整流モジュール 8 2 5 は、正の半波整流を差分信号  $diff[m]$  に適用して、信号中のラウドネス増加の影響を受けていない間のラウドネス低下を示す信号  $diff_{rect}[m]$  を生成し得る。

【 0 1 1 9 】

【数 3 9】

$$diff_{rect}[m] = \begin{cases} diff[m], & diff[m] > 0 \\ 0.0, & diff[m] \leq 0 \end{cases}$$

正規化モジュール 8 3 0 は、整流された差分  $diff_{rect}[m]$  を正規化して、dB 範囲から所望の検出範囲に変換し、低下検出値  $drop[m]$  を生成し得る。例えば検出範囲  $[0, 1]$  について、以下のように、単純な平行移動、拡大縮小及び飽和を正規化に使用し得る。

【 0 1 2 0 】

10

20

30

40

【数 4 0】

$$drop[m] = \text{saturate} \left( \frac{diff_{rect}[m] - D_{min}}{D_{max} - D_{min}} \right), \quad D_{max} > D_{min} \geq 0$$

ただし、

【0 1 2 1】

【数 4 1】

$$\text{saturate}(x) = \begin{cases} x, & 0 \leq x \leq 1 \\ 1, & x > 1 \\ 0, & x < 0 \end{cases}$$

10

また  $D_{min}$  及び  $D_{max}$  は、例えば 0 及び 1 の検出値にそれぞれマップするラウドネス低下閾値を表す。この例では、1 のラウドネス低下検出値  $drop[m]$  は、 $D_{max}$  より大きいラウドネス低下が生じていることを示し、これは例えば音の大きなテレビコマーシャルから静かな番組に遷移するようなコンテンツ間遷移の際に生じ得る。0 の値は、低下がないことを示し、これは例えば単一のコンテンツ全体を通して共通である。0 と 1 との間の値は、中間レベルのラウドネス低下を示す。

20

【0 1 2 2】

図 9 は、2 秒の大きなラウドネス低下からなる音声信号に適用され、図 8 のラウドネス低下検出システム 800 に対する短期的なラウドネス推定  $L_{short, dB}[m]$  (実線)、2 つの平滑化されたフィルター出力  $\mu_{slow}[m]$  (鎖線) 及び  $\mu_{fast}[m]$  (破線)、並びにラウドネス低下検出信号  $drop[m]$  (下側のプロット) を示している。ここで注目すべき点は、短期的なラウドネス推定  $L_{short, dB}[m]$  (実線) は、約 -10 dB から -30 dB まで 2 秒でほぼ瞬間的に低下し、時間的に平滑化されたフィルター出力  $\mu_{fast}[m]$  (破線) は、時間的に平滑化されたフィルター出力  $\mu_{slow}[m]$  (鎖線) よりも速く -30 dB に到達することである。下側プロットのラウドネス低下検出信号  $drop[m]$  は、2 秒から始まるラウドネス低下、及び大きなラウドネス低下が生じていることを示す約 2.5 秒のピーキングを示している。平滑化係数  $_{slow}$  及び  $_{fast}$  は、比較的高速になるように相互に選択されており、これはラウドネス低下検出が可能な速度を直接的に制御する。

30

【0 1 2 3】

図 9 の例は、ラウドネス低下検出システム、例えば図 8 のシステム 800 の能力を図示したものであり、 $_{slow}$  及び  $_{fast}$  に比較的高速な値を用いることでラウドネスの大きな低下を速く識別できることを示している。しかし、これらの相互に高速な平滑化係数が同じでは、その検出性能は、会話などの非常に動的な信号については最適とは言えず、自然な信号変動がラウドネス低下として誤って検出されるといった誤検出が頻発し得る。

40

【0 1 2 4】

図 10 A ~ 10 D は、図 9 と同様に、例えば図 8 のラウドネス低下検出システム 800 などのラウドネス低下検出システムにおける短期的なラウドネス推定  $L_{short, dB}[m]$  (実線)、2 つの平滑化されたフィルター出力  $\mu_{slow}[m]$  (鎖線) 及び  $\mu_{fast}[m]$  (破線)、並びにラウドネス低下検出信号  $drop[m]$  (下側のプロット) を  $_{slow}$  及び  $_{fast}$  の異なる平滑化係数選択のそれぞれについて示している。図 9 の 2 秒のラウドネス低下からなる音声信号は、図 10 A 及び 10 C で再び使用されており、図 10 A は、相互に高速な平滑化係数  $_{slow}$  及び  $_{fast}$  を使用した結果を示し、図 10 C は、相互に低速な平

50



滑化係数  $\mu_{slow}$  及び  $\mu_{fast}$  を使用した結果を示している。図 10A 及び 10C に示されている音声信号からすれば、できる限り速くラウドネス低下を検出することがラウドネス低下検出システムにとって望ましいであろう。図 10B 及び 10D では、動的な会話のセグメントが使用されており、図 10B は、相互に高速な平滑化係数  $\mu_{slow}$  及び  $\mu_{fast}$  を使用した結果を示し、図 10D は、相互に低速な平滑化係数  $\mu_{slow}$  及び  $\mu_{fast}$  を使用した結果を示している。約 -40 dB のより静寂な周辺環境雑音で区切られた約 -10 dB の一連の大きな話し言葉でコンテンツが構成されているときは、動的音声信号における短期的なラウドネスレベル  $L_{short, dB} [m]$  の大きな変動に留意すべきである。動的音声信号は、長期的なラウドネス低下を含まないので、理想的なラウドネス低下検出システムであるならば、ラウドネス低下を検出しないであろう。

10

## 【0125】

図 10A の低下検出信号  $drop [m]$  は、図 8 のラウドネス低下検出システム 800 において、大きな長期的なラウドネス低下を含む信号に対して相互に高速な平滑化係数を用いた場合、速く正確に約 2.5 秒でラウドネス低下を検出することを可能にすることを示している。しかし図 10B の低下検出信号  $drop [m]$  は、ラウドネス低下検出システムにおいて、非常に動的な信号に対して相互に高速な平滑化係数を用いた場合、 $\mu_{fast} [m]$  が過度に速く反応して会話中の単語間の休止を追跡することとなり、多数の部分的な検出を不正確に報告する要因となることを示している。

## 【0126】

前述したように相互に高速な平滑化係数は、ラウドネス低下の誤検出が生じる可能性がより高いので、非常に動的な信号については最適ではあり得ない。図 10C 及び 10D は、相互により低速な平滑化係数を使用した結果を示している。図 10C のラウドネス低下検出信号  $drop [m]$  は、図 8 のラウドネス低下検出システム 800 において、大きな長期的なラウドネス低下を含む信号に対して相互に低速な平滑化係数を使用した場合、相互に高速な平滑化係数を使用した場合の 2.5 秒とは対照的に、約 4 秒までラウドネス低下を完全には検出できない可能性があることを示している。図 10D のラウドネス低下検出信号  $drop [m]$  は、ラウドネス低下検出システムにおいて、非常に動的な信号に対して相互に低速な平滑化係数を使用した場合、長期的なラウドネス低下がないことを正確に報告することが可能になることを示している。

20

## 【0127】

図 10C 及び 10D の例において、注目すべき点は、相互に低速な平滑化係数が使用される場合、平滑化係数  $\mu_{fast}$  は、アタック速度が比較的高速なままでリリース速度のみが遅くなるように独自に修正されており、アタック速度及びリリース速度が両方とも平滑化係数  $\mu_{slow}$  に対して等しく遅くなっていることである。 $\mu_{fast}$  に対して独立した高速なアタック速度及び低速なリリース速度を許容した場合、平滑化された結果である  $\mu_{fast} [m]$  がラウドネス推定のピークの方へバイアスされ、通常は  $\mu_{fast} [m]$  が  $\mu_{slow} [m]$  より高いままとなる。この修正は、非常に動的なコンテンツに対するラウドネス低下の誤検出率を改善し得る。

30

## 【0128】

上記の分析は、ラウドネス低下検出システムにおける平滑化係数速度のチューニングにトレードオフが存在することを示唆している。ラウドネス低下検出システムの改善は、非常に動的なコンテンツ（例えば、図 10D）のときには遅くなるように平滑化係数速度を動的に修正して、ラウドネス低下の誤検出を制限し、あまり動的でないコンテンツのときには速くなるように平滑化係数速度を動的に修正して、ラウドネス低下をより速く検出する（例えば、図 10A）ことによって達成され得る。平滑化係数を動的に修正するラウドネス低下検出システムの一例を以下説明する。

40

## 【0129】

動的平滑化係数は、ラウドネス低下検出性能を改善するために、図 8 のシステム 800 に組み込まれ得る。図 11 は、一実施形態による、動的平滑化係数を有するラウドネス低下検出システム 1100 のブロック図である。特に図 11 は、標準偏差モジュール 113

50

5をラウドネス低下検出システム1100に一体化することを示している。標準偏差モジュール1135は、信号依存の方法で時間的平滑化コンポーネント1110及び1115が平滑化係数 $\alpha_{slow}$ 及び $\alpha_{fast}$ を動的に修正し得るように、信号ダイナミックスの推定を提供することができる。ラウドネス低下検出システム1100は、ラウドネス測定モジュール1105、減算モジュール1120、半波整流モジュール1125及び正規化モジュール1130も含むことができる。

【0130】

ラウドネス低下検出システム1100は、音声信号 $x[n]$ を受け取り、 $m$ をインデックスとして、時間的に変化するラウドネス低下検出推定 $drop[m]$ を出力することができる。これにより $drop[m]$ は、顕著なラウドネスレベル低下が生じているかどうかを示す。ラウドネス低下検出推定は、例えば範囲 $[0, 1]$ 上で定義され、0はラウドネス低下が存在しないことを示し、1は大きなラウドネス低下がちょうど生じたばかりであることを示し、これらの間の値は、より小さな、又はより穏やかなラウドネス低下を示す指標である。しかし他の低下検出値も使用できる。ラウドネス測定モジュール1105、時間的平滑化コンポーネント1110及び1115、減算モジュール1120、半波整流モジュール1125並びに正規化モジュール1130は、図8で説明されているラウドネス測定モジュール805、時間的平滑化コンポーネント810及び815、減算モジュール820、半波整流モジュール825並びに正規化モジュール830と同様の動作をすることができる。

10

【0131】

前述したように平滑化されたラウドネス推定 $\mu_{slow}[m]$ 及び $\mu_{fast}[m]$ の相対的挙動は、検出されるラウドネス低下の頻度及び程度に影響を及ぼし得る。したがって平滑化係数 $\alpha_{slow}$ 及び $\alpha_{fast}$ に適切な値を使用することで、異なる種類の入力信号に対して適切な性能を達成することができる。

20

【0132】

信号ダイナミックスは、短期的なラウドネス推定の修正された標準偏差尺度を計算することによって、標準偏差モジュール1135を介して推定され得る。ラウドネス平均は、短期的なラウドネス推定 $L_{short,dB}[m]$ を時間的に平滑することによって推定され得る。平滑化係数 $\alpha_L$ は、数式42に固有であり、 $\mu_L[m]$ が所望の平均窓長に近似するように選択され得る。

30

【0133】

【数42】

$$\mu_L[m] = \mu_L[m-1] \cdot (1 - \alpha_L) + L_{short,dB}[m] \cdot \alpha_L$$

短期的なラウドネス推定とその推定された平均との差を取ることができる。

【0134】

【数43】

$$d[m] = L_{short,dB}[m] - \mu_L[m]$$

40

この差は、正に半波整流され、平方され得る。

【0135】

【数 4 4】

$$d_{rect}[m] = \begin{cases} d^2[m], & d[m] > 0 \\ 0 & d[m] \leq 0 \end{cases}$$

半波整流は、一般的な標準偏差尺度の一部でない場合があるが、ラウドネス低下とラウドネス増加とを区別するのに役立つ。差分信号  $d[m]$  は、ラウドネス低下中は負であり、したがって正の半波整流を適用することによって、その結果得られる平方された差分値は、ラウドネス増加にのみ基づき得る。この計算において効果的にラウドネス低下を取り除くことによって、レベルが低い短期的なダイナミクス及び場合によっては大きな長期的なラウドネス低下（例えば図 10 A 及び 10 C に示されているラウドネス低下）の信号では、低い平方された差分値  $d_{rect}[m]$  が得られ、レベルが高い短期的なダイナミクス（例えば、図 10 B 及び 10 D に示されているラウドネス低下）の信号では、高い平方された差分値  $d_{rect}[m]$  が得られる。

10

【0 1 3 6】

整流され、平方された差分  $d_{rect}[m]$  は、数式 4 5 に固有の平滑化係数  $\alpha_{std}$  で時間的に平滑化され、その平方根を取ることによって、短期的なラウドネス推定の標準偏差  $\sigma[m]$  の推定を得ることができる。

20

【0 1 3 7】

【数 4 5】

$$\sigma[m] = \sqrt{\sigma^2[m-1] \cdot (1 - \alpha_{std}) + d_{rect}[m] \cdot \alpha_{std}}$$

推定された標準偏差  $\sigma[m]$  は、次いで、前述した  $d_{drop}[m]$  と同様に、平行移動、拡大縮小及び飽和などの方法を使用して、例えば範囲  $[0, 1]$  に正規化され得る。

【0 1 3 8】

一例において、その結果得られる正規化標準偏差  $\sigma_{norm}[m]$  は、時間的平滑化コンポネント 1 1 1 0 及び 1 1 1 5 のそれぞれにおいて、平滑化係数  $\alpha_{slow}[m]$  及び  $\alpha_{fast}[m]$  を動的に変調するために使用され得る。例えば平滑化係数  $\alpha_{slow}[m]$  及び  $\alpha_{fast}[m]$  は、2つの所定の平滑化係数速度の間、最低速度と最高速度との間で線形補間され得る。前述したように平滑化係数  $\alpha_{slow}[m]$  は、等しいアタックリリース速度を有するのが望ましいので、最低速度と最高速度との間で単純に線形補間され得る。

30

【0 1 3 9】

【数 4 6】

$$\alpha_{slow}[m] = \alpha_{slow,max} \cdot (1 - \sigma_{norm}[m]) + \alpha_{slow,min} \cdot \sigma_{norm}[m]$$

40

ただし  $\alpha_{slow,max} > \alpha_{slow,min}$ 、換言すれば  $\alpha_{slow,max}$  は、 $\alpha_{slow,min}$  より高速である。標準偏差尺度が高い場合、例えば  $\sigma_{norm}[m] = 1$  である場合、 $\alpha_{slow}[m]$  は、低速の値  $\alpha_{slow,min}$  に設定され得る。標準偏差尺度が低い場合、例えば、 $\sigma_{norm}[m] = 0$  である場合、 $\alpha_{slow}[m]$  は、高速の値  $\alpha_{slow,max}$  に設定され得る。標準偏差尺度がこれらの間にある場合、例えば  $0 < \sigma_{norm}[m] < 1$  である場合、 $\alpha_{slow}[m]$  は、最低速度と最高速度との間で線形補間され得る。

【0 1 4 0】

前述したように性能は、正規化標準偏差に基づいてリリース係数が最低速度と最高速度との間で線形補間される間、アタック係数が高速に維持されるように、平滑化係数  $\alpha_{fast}$

50

[ m ] のアタック速度とリリース速度が独立して計算されるときに改善され得る。

【 0 1 4 1 】

【 数 4 7 】

$$\alpha_{fast}[m] = \begin{cases} \alpha_{fast,max}, & L_{short,dB}[m] > \mu_{fast}[m-1] \\ \alpha_{fast,max} \cdot (1 - \sigma_{norm}[m]) + \alpha_{fast,min} \cdot \sigma_{norm}[m], & otherwise \end{cases}$$

ただし  $\alpha_{fast,max}$  及び  $\alpha_{fast,min}$  は、所定の平滑化係数であり、 $\alpha_{fast,max} > \alpha_{fast,min}$ 、換言すれば  $\alpha_{fast,max}$  は、 $\alpha_{fast,min}$  より高速である。

10

【 0 1 4 2 】

図 1 2 A 及び 1 2 B は、これらの動的平滑化係数の修正を適用した結果の例を示している。図 1 2 A 及び 1 2 B は、図 9 及び図 1 0 A ~ 1 0 D と同様に、例えば図 1 1 のラウドネス低下検出システム 1 1 0 0 等のラウドネス低下検出システムにおける短期的なラウドネス推定  $L_{short,dB}[m]$  (実線)、2つの平滑化されたフィルター出力  $\mu_{slow}[m]$  (鎖線) 及び  $\mu_{fast}[m]$  (破線)、並びにラウドネス低下検出信号  $drop[m]$  (下側のプロット) を示している。図 1 2 A のラウドネス低下検出信号  $drop[m]$  は、真のラウドネス低下の発生から 0.5 秒以内の正確な検出を示している。図 1 2 B のラウドネス低下検出信号  $drop[m]$  は、短期的な信号ダイナミクスで誤検出がないことを示している。図 1 2 A 及び 1 2 B は、図 1 0 A ~ 1 0 D に示されている静的平滑化係数に対し、信号依存の動的平滑化係数を使用することによって行える改善を図示している。

20

【 0 1 4 3 】

図 8 のラウドネス低下検出システム 8 0 0 及び図 1 1 のラウドネス低下検出システム 1 1 0 0 は、図 2 に示されているラウドネスコントロールシステム 2 0 0 のように、ラウドネスコントロールシステムに一体化することができる。図 1 3 は、図 8 で説明されているラウドネス低下検出システム 8 0 0 又は図 1 1 で説明されているラウドネス低下検出システム 1 1 0 0 などのラウドネス低下検出モジュール 1 3 2 5 を備えるラウドネスコントロールシステム 1 3 0 0 の高水準ブロック図を示している。

【 0 1 4 4 】

ラウドネスコントロールシステム 1 3 0 0 は、ラウドネス測定モジュール 1 3 0 5、ラウドネス低下検出モジュール 1 3 2 5、時間的平滑化モジュール 1 3 1 0 及び利得補正モジュール 1 3 1 5 を含む得る。ラウドネス測定モジュール 1 3 0 5 及び利得補正モジュール 1 3 1 5 は、図 2 で説明されているラウドネス測定モジュール 2 0 5 及び利得補正モジュール 2 1 5 に関して説明されている動作と同様の動作をし得る。

30

【 0 1 4 5 】

図 2 のラウドネスコントロールシステム 2 0 0 に関して説明されているように、時間的平滑化モジュール 1 3 1 0 は、別々の「アタック」及び「リリース」平滑化係数を備えることができる。リリース平滑化係数  $\alpha_{release}$  は、ラウドネスコントロールで利得レベルを高めることができる速度を制御し得る。高速な  $\alpha_{release}$  値は、ラウドネスコントロールで利得レベルを速く高めることを可能にし、低速な  $\alpha_{release}$  値は、利得レベルを高めることができる速度を制限し得る。

40

【 0 1 4 6 】

単純なラウドネスコントロールシステムは、コンテンツ内及びコンテンツ間のダイナミクスをバランスさせるように選択された信号独立の所定値に平滑化係数  $\alpha_{release}$  を設定し得るが、これは最適な性能を妥協することになる。ラウドネス低下検出を一体化することによってラウドネスコントロールシステムは、コンテンツ間ダイナミクス及びコンテンツ内ダイナミクスの両方が適切に処理されるように、平滑化係数  $\alpha_{release}[m]$  を動的に修正することができる。ラウドネス低下検出がない間、例えば  $drop[m] = 0$  の場合、 $\alpha_{release}[m]$  は、コンテンツ内ダイナミクスを維持する所定の既定値  $\alpha_{release,def}$  に設定され得る。ラウドネス低下が検出された場合、例えば  $drop[m] =$

50

1 のとき、 $\alpha_{release}[m]$  は、所定の値  $\alpha_{release,max}$  まで加速され、それによって例えばコンテンツ間遷移の際に、利得レベルを速く高めることができる。部分的低下検出の間、例えば  $0 < drop[m] < 1$  の場合には、 $\alpha_{release}[m]$  値が極値と極値との間で線形補間され得る。

【0147】

【数48】

$$\alpha_{release}[m] = \alpha_{release,def} \cdot (1 - drop[m]) + \alpha_{release,max} \cdot drop[m]$$

10

大きなラウドネス低下に対してラウドネス低下検出値が高い場合には、より小さなラウドネス低下に比べて利得回復が速くなる可能性がある。これは、ランプの持続時間を短縮することによって顕著な「ランピング」アーチファクトを軽減する働きをし得る。

【0148】

ラウドネス低下からの回復は、固定された長さの時間内に広範囲のラウドネス低下から回復することによっても達成し得る。例えばラウドネス低下からの回復は、ラウドネス低下の及ぶ範囲にかかわらず、3秒以内に行われるのが望ましい。ラウドネス低下の推定を使用することで、ラウドネス低下の及ぶ範囲と無関係に、この長さの時間内の回復を確実にする適切な平滑化係数  $\alpha_{release}[m]$  を計算し得る。

【0149】

20

別の実施形態によれば、図3のシステム300又は図5のシステム500などの雑音検出システム、及び図8のシステム800又は図11のシステム1100などのラウドネス低下検出システムは、いずれも図2のシステム200などのラウドネスコントロールシステムに一体化され得る。図14は、一実施形態による、雑音検出及びラウドネス低下検出を行うラウドネスコントロールシステム1400のブロック図である。

【0150】

ラウドネスコントロールシステム1400は、ラウドネス測定モジュール1405、雑音検出モジュール1420、ラウドネス低下検出モジュール1425、時間的平滑化モジュール1410及び利得補正モジュール1415を含み得る。ラウドネス測定モジュール1405及び利得補正モジュール1415は、図2で説明されているラウドネス測定モジュール205及び利得補正モジュール215に関して説明されている動作と同様の動作をし得る。雑音検出モジュール1420は、図3で説明されている周波数雑音検出システム300又は図5で説明されている周波数雑音検出システム500に関して説明されている動作と同様の動作をし得る。ラウドネス低下検出モジュール1425は、図8で説明されているラウドネス低下検出システム800又は図11で説明されているラウドネス低下検出システム1100に関して説明されている動作と同様の動作をし得る。

30

【0151】

時間的平滑化モジュール1410は、図7で説明されている時間的平滑化モジュール710及び図13で説明されている時間的平滑化モジュール1310に関して説明されている動作と同様の動作をし得る。時間的平滑化モジュール1410は、図7の説明で示されているような平滑化係数を遅くすることができるコンテンツ対雑音分類値を受け取り、図13の説明で示されているような平滑化係数の速度を高めることができるラウドネス低下検出値も受け取ることができる。コンテンツ対雑音分類に基づいて平滑化係数を遅くするか、又はラウドネス低下検出に基づいて平滑化係数の速度を高めるか、或いはこれら2つの組合せを介して新しい速度を計算するかの決定は、多数のトレードオフを伴う決定であり、アプリケーション特有であり得る。一実施形態において、時間的平滑化モジュール1410のリリース平滑化係数  $\alpha_{release}[m]$  は、以下のように、数式35及び48の結果の平均を介するコンテンツ対雑音分類値とラウドネス低下検出値との一次結合によって、動的に修正され得る。

40

【0152】

50

【数 4 9】

$$\alpha_{release}[m] = \frac{\alpha_{release,def} \cdot (class[m] + 1 - drop[m]) + \alpha_{release,max} \cdot drop[m]}{2}$$

特定の組合せで特徴及び構成要素が上記説明されているが、当業者であれば、それぞれの特徴又は構成要素は単独で、あるいは他の特徴及び要素と組み合わせて使用できることを理解するであろう。本明細書で説明されている特徴及び構成要素は、いずれも個別のモジュールとして実装され得るか、又は特徴のセット若しくは一部が組み合わされて共通のプログラム可能なモジュールに実装され得る。

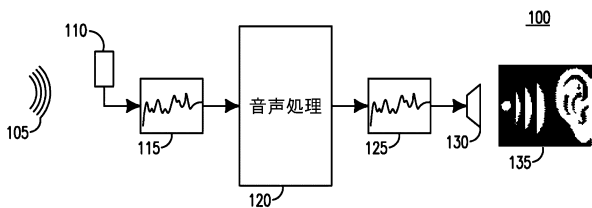
10

【 0 1 5 3】

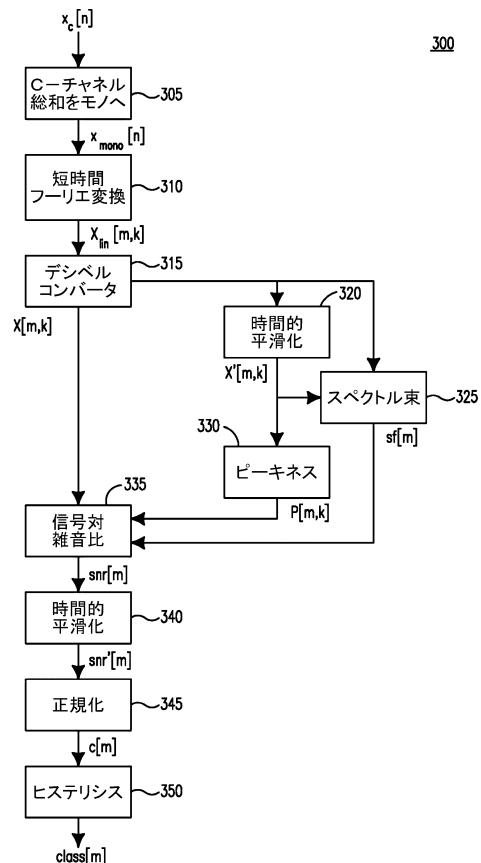
さらに本明細書で説明されているシステム及び方法は、コンピュータ又はプロセッサにより実行できるようにコンピュータ可読媒体内に組み込まれたコンピュータプログラム、ソフトウェア又はファームウェアでハードウェアに実装され得る。コンピュータ可読媒体の例としては、電子信号（有線又はワイヤレス接続で送信される）及びコンピュータ可読記憶媒体が挙げられる。コンピュータ可読記憶媒体の例としては、特にこれらに限定はされないが、リードオンリーメモリ（ROM）、ランダムアクセスメモリ（RAM）、レジスタ、キャッシュメモリ、半導体メモリデバイス、内蔵ハードディスク及びリムーバブルディスクなどの磁気媒体、光磁気媒体、並びにCD-ROMディスク及びデジタル多用途ディスク（DVD）などの光学媒体が挙げられる。

20

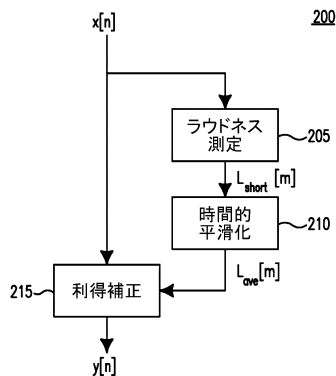
【 図 1】



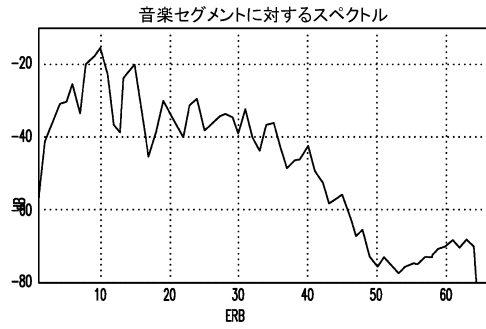
【 図 3】



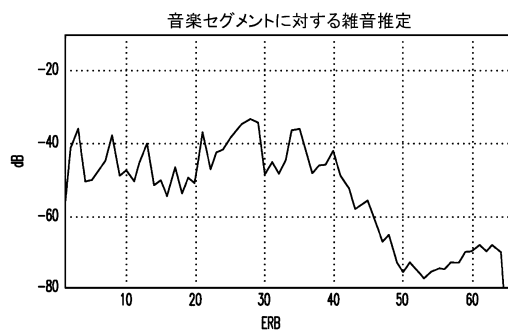
【 図 2】



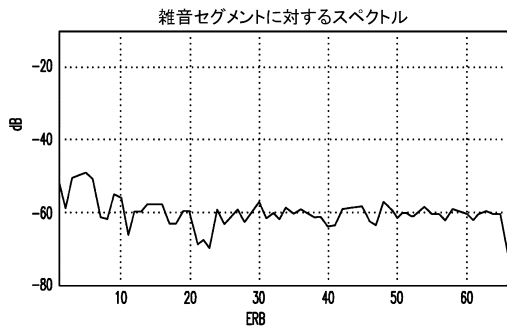
【図4A】



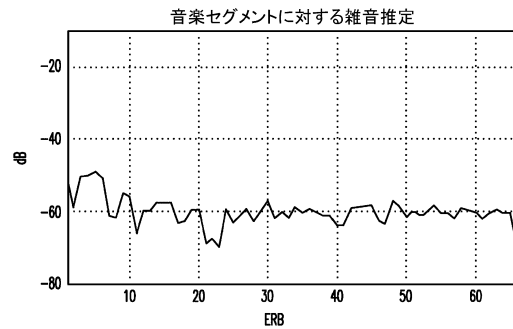
【図4C】



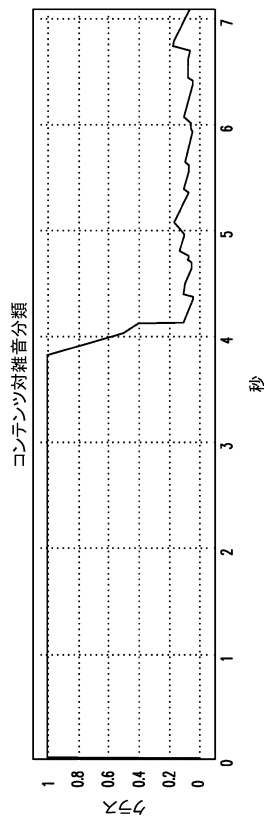
【図4B】



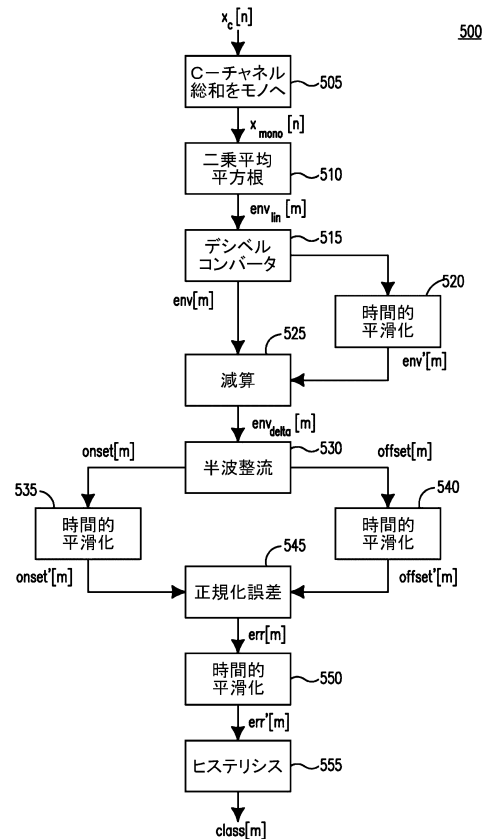
【図4D】



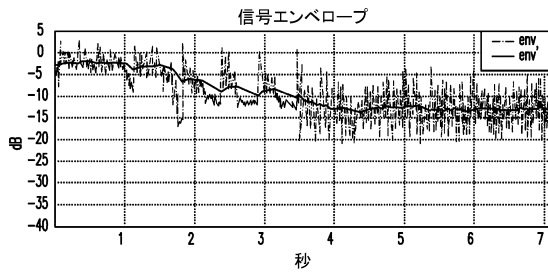
【図4E】



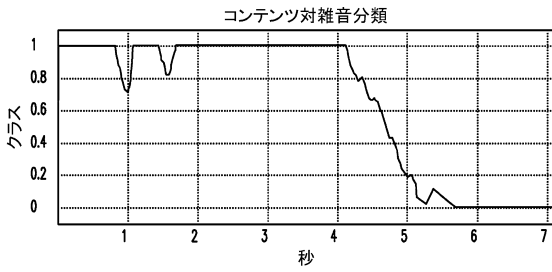
【図5】



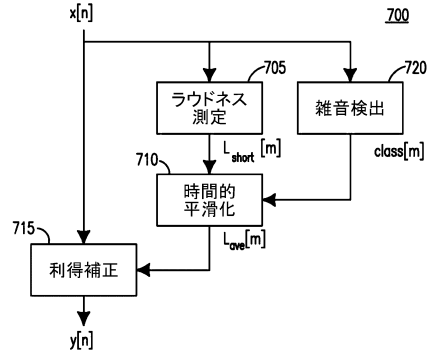
【図 6 A】



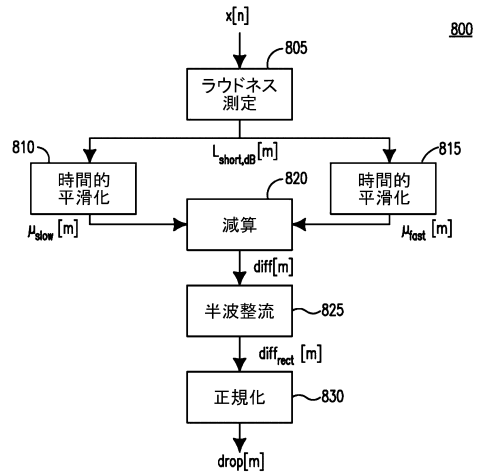
【図 6 B】



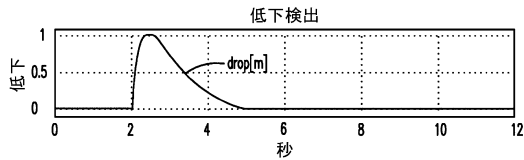
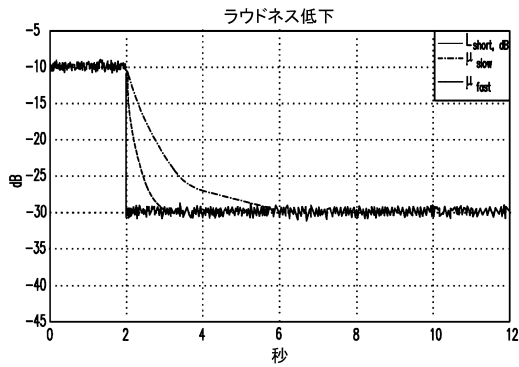
【図 7】



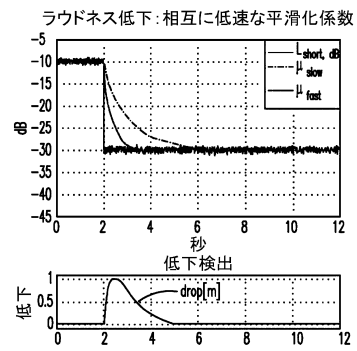
【図 8】



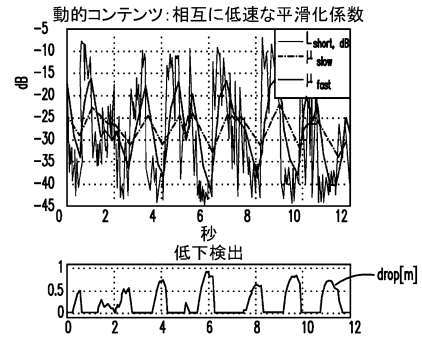
【図 9】



【図 10 A】

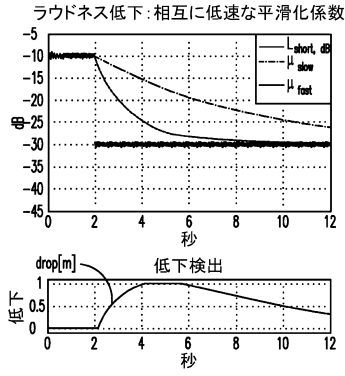


【図 10 B】

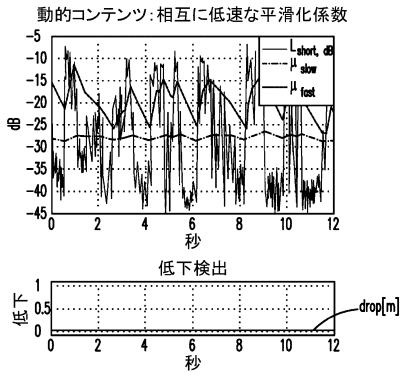




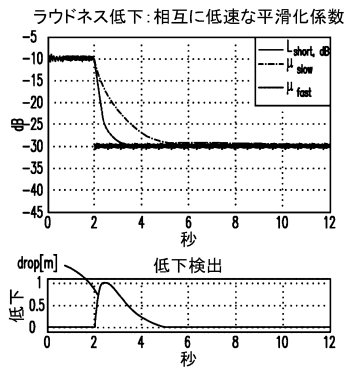
【図10C】



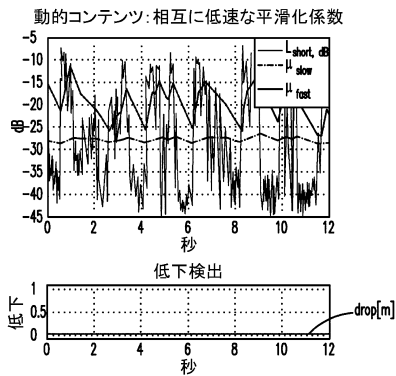
【図10D】



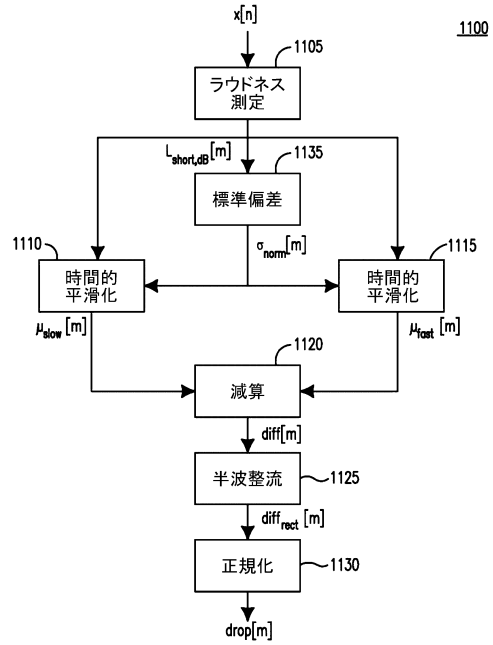
【図12A】



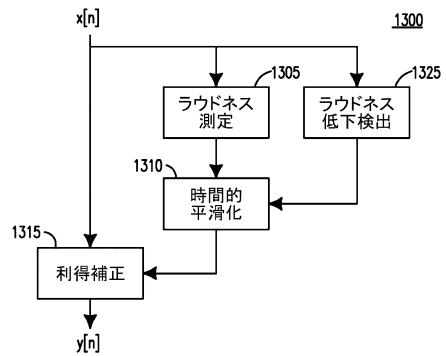
【図12B】



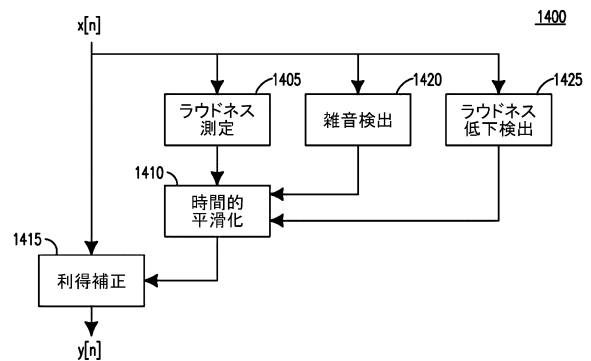
【図11】



【図13】



【図14】



## フロントページの続き

- (74)代理人 100103610  
弁理士 吉田 和彦
- (74)代理人 100067013  
弁理士 大塚 文昭
- (74)代理人 100086771  
弁理士 西島 孝喜
- (74)代理人 100109070  
弁理士 須田 洋之
- (74)代理人 100109335  
弁理士 上杉 浩
- (74)代理人 100120525  
弁理士 近藤 直樹
- (74)代理人 100158551  
弁理士 山崎 貴明
- (72)発明者 スミス, ブランドン  
アメリカ合衆国 98033 ワシントン, カークランド, レイクビュー ドライブ 600  
7 #9
- (72)発明者 ワーナー, アーロン  
アメリカ合衆国 98122 ワシントン, シアトル, 25番 アベニュー 1712
- (72)発明者 トンプソン, ジェフ  
アメリカ合衆国 98011 ワシントン, ボセル, 86番 プレイス ノースイースト 2  
0265

審査官 大野 弘

(56)参考文献 特開2009-296298(JP,A)

(58)調査した分野(Int.Cl., DB名)

G10L 21/034  
G10L 21/0364  
G10L 25/84  
H03G 5/16