

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5259759号
(P5259759)

(45) 発行日 平成25年8月7日(2013.8.7)

(24) 登録日 平成25年5月2日(2013.5.2)

(51) Int.Cl.

F 1

H04S 3/00 (2006.01)

H04S 3/00

Z

G10L 21/034 (2013.01)

G10L 21/02

301B

G10L 25/81 (2013.01)

G10L 11/00

402G

請求項の数 14 (全 14 頁)

(21) 出願番号 特願2011-52503 (P2011-52503)
 (22) 出願日 平成23年3月10日 (2011.3.10)
 (62) 分割の表示 特願2011-505219 (P2011-505219)
 原出願日 平成21年4月17日 (2009.4.17)
 (65) 公開番号 特開2011-172235 (P2011-172235A)
 (43) 公開日 平成23年9月1日 (2011.9.1)
 審査請求日 平成23年4月7日 (2011.4.7)
 (31) 優先権主張番号 61/046,271
 (32) 優先日 平成20年4月18日 (2008.4.18)
 (33) 優先権主張国 米国(US)

(73) 特許権者 507236292
 ドルビー ラボラトリーズ ライセンシング コーポレーション
 アメリカ合衆国 カリフォルニア州 94
 103-4813 サンフランシスコ ポ
 トレロ アベニュー 100
 (74) 代理人 100068755
 弁理士 恩田 博宣
 (74) 代理人 100105957
 弁理士 恩田 誠
 (74) 代理人 100142907
 弁理士 本田 淳
 (74) 代理人 100071010
 弁理士 山崎 行造

最終頁に続く

(54) 【発明の名称】サラウンド体験に対する影響を最小限にしてマルチチャンネルオーディオにおけるスピーチの聴覚性を維持するための方法及び装置

(57) 【特許請求の範囲】

【請求項 1】

マルチチャンネルオーディオ信号におけるスピーチの可聴性を改善する方法であって、減衰係数を生成するためにマルチチャンネルオーディオ信号の第1の特性と第2の特性とを比較するステップであって、前記第1の特性は、スピーチオーディオ及びノンスピーチオーディオを含有するマルチチャンネルオーディオ信号の第1番目のチャンネルに相当し、前記第1の特性は、前記第1番目のチャンネルの信号の第1のパワースペクトルに相当し、前記第2の特性は、主としてノンスピーチオーディオを含有するマルチチャンネルオーディオ信号の第2番目のチャンネルに相当し、前記第2の特性は、前記第2番目のチャンネルの信号の第2のパワースペクトルに相当し、前記第1の特性と前記第2の特性とを比較するステップは、

予測明瞭度を生成するために前記第1のパワースペクトルと前記第2のパワースペクトルとに基づいて明瞭度の予測を行うステップと、

前記予測明瞭度が基準を満たすまで前記第2のパワースペクトルに適用するゲインを調整するステップと、

前記予測明瞭度が基準をいったん満たしたときの減衰係数として、調整が完了した前記ゲインを用いるステップと、

を含むことを特徴とする、比較するステップと、

調整済みの減衰係数を生成するためにスピーチ尤度値に従って減衰係数を調整するステップと、

10

20

前記調整済みの減衰係数を用いて前記第2番目のチャンネルを減衰させるステップと、を具備することを特徴とする方法。

【請求項2】

前記第1の特性と前記第2の特性とを生成するために前記マルチチャンネルオーディオ信号を処理するステップをさらに具備することを特徴とする請求項1に記載の方法。

【請求項3】

前記スピーチ尤度値を生成するために前記第1番目のチャンネルを処理するステップをさらに具備することを特徴とする請求項1に記載の方法。

【請求項4】

前記第2番目のチャンネルは、複数の第2番目のチャンネルのうちの1つであり、前記第2の特性は、複数の第2の特性のうちの1つであり、前記減衰係数は、複数の減衰係数のうちの1つであり、前記調整済みの減衰係数は、複数の調整済みの減衰係数のうちの1つであり。10

前記複数の減衰係数を生成するために前記第1の特性と前記複数の第2の特性とを比較するステップと、

前記複数の調整済みの減衰係数を生成するために前記スピーチ尤度値に従って前記複数の減衰係数を調整するステップと、

前記複数の調整済みの減衰係数を用いて前記複数の第2番目のチャンネルを減衰させるステップと、

をさらに具備することを特徴とする請求項1に記載の方法。20

【請求項5】

前記マルチチャンネルオーディオ信号は、主としてノンスピーチオーディオを含有する第3番目のチャンネルを具備し、

追加の減衰係数を生成するために前記第1の特性と第3の特性とを比較するステップであって、該第3の特性は前記第3番目のチャンネルに相当することを特徴とするステップと、

調整済みの追加の減衰係数を生成するために、前記スピーチ尤度値に従って前記追加の減衰係数を調整するステップと、

前記調整済みの減衰係数用いて前記第3番目のチャンネルを減衰させるステップと、30

をさらに具備することを特徴とする請求項1に記載の方法。

【請求項6】

前記第2のワースペクトルは、複数の帯域を有し、前記第1の特性と前記第2の特性とを比較するステップは、計算音量を生成するために前記第2のワースペクトルに基づいて音量の計算を行うステップを具備し、前記ゲインを調整するステップは前記予測明瞭度が明瞭度の基準を満たし前記計算音量が音量の基準を満たすまで、前記第2のワースペクトルの各帯域にそれぞれ適用される複数のゲインを調整するステップをさらに具備し、前記ゲインを用いるステップは、前記予測明瞭度が前記明瞭度の基準をいったん満たし、前記計算音量が前記音量の基準をいったん満たしたときのそれぞれの各帯域の減衰係数として、調整が行完了した前記複数のゲインを用いるステップ、

を具備することを特徴とする請求項1に記載の方法。40

【請求項7】

マルチチャンネルオーディオ信号におけるスピーチの可聴性を改善する回路を有する装置であって、

減衰係数を生成するためにマルチチャンネルオーディオ信号の第1の特性と第2の特性とを比較する比較回路であって、前記第1の特性は、スピーチオーディオ及びノンスピーチオーディオを含有するマルチチャンネルオーディオ信号の第1番目のチャンネルに相当し、前記第1の特性は、前記第1番目のチャンネルの信号の第1のワースペクトルに相当し、前記第2の特性は、主としてノンスピーチオーディオを含有するマルチチャンネルオーディオ信号の第2番目のチャンネルに相当し、前記第2の特性は、前記第2番目のチャンネルの信号の第2のワースペクトルに相当し、前記比較回路は、50

予測明瞭度を生成するために前記第1のパワースペクトルと前記第2のパワースペクトルとに基づいて明瞭度の予測を行うよう構成されている明瞭度予測回路と、

前記予測明瞭度が基準を満たすまで前記第2のパワースペクトルに適用するゲインを調整するよう構成されているゲイン調整回路と、

前記予測明瞭度が基準をいったん満たしたときの減衰係数として、調整が完了した前記ゲインを選択するよう構成されているゲイン選択回路と、

を含むことを特徴とする比較回路と、

調整済みの減衰係数を生成するためにスピーチ尤度値に従って減衰係数を調整するよう構成されている乗算器と、

前記調整済みの減衰係数を用いて前記第2番目のチャンネルを減衰させるように構成されている増幅器と、

を具備することを特徴とする装置。

【請求項8】

前記第2パワースペクトルは、複数の帯域を有し、前記比較回路は、

計算音量を生成するために前記第2のパワースペクトルに基づいて音量の計算を行うよう構成されている音量計算回路と、

前記予測明瞭度が明瞭度の基準を満たし前記計算音量が音量の基準を満たすまで、前記第2のパワースペクトルの各帯域にそれぞれ適用される複数のゲインを調整し、前記予測明瞭度が前記明瞭度の基準をいったん満たし、前記計算音量が前記音量の基準をいったん満たしたときのそれぞれの各帯域の減衰係数として、調整が完了した前記複数のゲインを使用するよう構成されている最適化回路と、

をさらに具備することを特徴とする請求項7に記載の装置。

【請求項9】

前記第1番目のチャンネルの前記第1のパワースペクトルを計算するように構成されている第1のパワースペクトル密度計算器と、

前記第2番目のチャンネルの前記第2のパワースペクトルを計算するように構成されている第2のパワースペクトル密度計算器と、

をさらに具備することを特徴とする請求項7に記載の装置。

【請求項10】

前記第1番目のチャンネルを第1の複数のスペクトル成分に分割するように構成されている第1のフィルターバンクと、

前記第1の複数のスペクトル成分から前記第1のパワースペクトルを計算するように構成されている第1のパワー推定器バンクと、

前記第2番目のチャンネルを第2の複数のスペクトル成分に分割するように構成されている第2のフィルターバンクと、

前記第2の複数のスペクトル成分から前記第2のパワースペクトルを計算するように構成されている第2のパワー推定器バンクと、

をさらに具備することを特徴とする請求項7に記載の装置。

【請求項11】

前記スピーチ尤度値を生成するために前記第1番目のチャンネルを処理するように構成されているスピーチ決定プロセッサをさらに具備することを特徴とする請求項7に記載の装置。

【請求項12】

有形の記録媒体に組み込まれた、マルチチャンネルオーディオ信号におけるスピーチの可聴性を改善するためのコンピュータプログラムであって、該コンピュータプログラムは、装置に、

減衰係数を生成するためにマルチチャンネルオーディオ信号の第1の特性と第2の特性とを比較するステップであって、前記第1の特性は、スピーチオーディオ及びノンスピーチオーディオを含有するマルチチャンネルオーディオ信号の第1番目のチャンネルに相当し、前記第1の特性は、前記第1番目のチャンネルの信号の第1のパワースペクトルに相

10

20

30

40

50

当し、前記第2の特性は、主としてノンスピーチオーディオを含有するマルチチャンネルオーディオ信号の第2番目のチャンネルに相当し、前記第2の特性は、前記第2番目のチャンネルの信号の第2のパワースペクトルに相当し、

予測明瞭度を生成するために前記第1のパワースペクトルと前記第2のパワースペクトルとに基づいて明瞭度の予測を行うステップと、

前記予測明瞭度が基準を満たすまで前記第2のパワースペクトルに適用するゲインを調整するステップと、

前記予測明瞭度が基準をいったん満たしたときの減衰係数として、調整が完了した前記ゲインを用いるステップと、

を含むことを特徴とする、比較するステップと、

10

調整済みの減衰係数を生成するためにスピーチ尤度値に従って減衰係数を調整するステップと、

前記調整済みの減衰係数を用いて前記第2番目のチャンネルを減衰させるステップと、
を具備する処理を実行させることを特徴とするコンピュータプログラム。

【請求項13】

マルチチャンネルオーディオ信号におけるスピーチの可聴性を改善する装置であって、
減衰係数を生成するためにマルチチャンネルオーディオ信号の第1の特性と第2の特性
とを比較する比較手段であって、前記第1の特性は、スピーチ及びノンスピーチオーディ
オを含有するマルチチャンネルオーディオ信号の第1番目のチャンネルに相当し、前記第
1の特性は、前記第1番目のチャンネルの信号の第1のパワースペクトルに相当し、前記
第2の特性は、主としてノンスピーチオーディオを含有するマルチチャンネルオーディオ
信号の第2番目のチャンネルに相当し、前記第2の特性は、前記第2番目のチャンネルの
信号の第2のパワースペクトルに相当し、前記比較手段は、

20

予測明瞭度を生成するために前記第1のパワースペクトルと前記第2のパワースペクト
ルとに基づいて明瞭度の予測を行う明瞭度予測手段と、

前記予測明瞭度が基準を満たすまで前記第2のパワースペクトルに適用するゲインを調
整するゲイン調整手段と、

前記予測明瞭度が基準をいったん満たしたときの減衰係数として、調整が完了した前記
ゲインを使用するゲイン使用手段と、

を含むことを特徴とする、比較手段と、

30

調整済みの減衰係数を生成するためにスピーチ尤度値に従って減衰係数を調整する調整
手段と、

前記調整済みの減衰係数を用いて前記第2番目のチャンネルを減衰させる減衰手段と、
を具備することを特徴とする装置

【請求項14】

前記第2パワースペクトルは、複数の帯域を有し、前記比較手段は、計算音量を生成す
るために前記第2のパワースペクトルに基づいて音量の計算を行うように構成されている
音量計算手段をさらに具備し、前記ゲイン調整手段は、前記予測明瞭度が明瞭度の基準を
満たし前記計算音量が音量の基準を満たすまで、前記第2のパワースペクトルの各帯域に
それぞれ適用される複数のゲインを調整する手段に相当し、前記ゲイン使用手段は、前記
予測明瞭度が前記明瞭度の基準をいったん満たし、前記計算音量が前記音量の基準をいっ
たん満たしたときのそれぞれの各帯域の減衰係数として、調整が行われた前記複数のゲイ
ンを使用する手段に相当することを特徴とする請求項13に記載の装置。

40

【発明の詳細な説明】

【技術分野】

【0001】

本出願は、2008年4月1日出願の米国暫定特許出願番号61/046,271に基づく優先権を主張し、この暫定特許出願のすべてを参照として本明細書に組み込むものとする。

【0002】

50

本発明は、一般的にはオーディオ信号処理関し、特に、エンターテインメントオーディオに囲まれた状態での、会話や談話の明瞭度の改善に関する。

【0003】

特記なき場合は、ここに記載の方法は、本出願の特許請求の範囲に対する先行技術ではなく、ここに記載されているという理由で先行技術であるとは認めない。

【背景技術】

【0004】

多数の、同時的なオーディオのチャンネル（サラウンドサウンド）をもつ現代的なエンターテインメントオーディオは、計り知れない娯楽的価値を持つ巨大でリアルなサウンド環境を聴衆に提供する。このような環境では、会話、音楽、効果音のような多くのサウンド要素が同時に出現し、リスナーの注意を得るために競い合う。聴衆によっては、特に、聴覚能力が衰えた又は認知処理が遅くなった聴衆には、会話や談話は、大きな競合サウンド要素が存在するプログラム部分では、聞こえにくいかもしれない。このような状況において、競合するサウンドのレベルが低下することは、リスナーには有益である。

10

【0005】

音楽や効果音が会話を圧倒することがあるとの認識は新しいものではなく、そのような状況を改善する方法はいくつか提案されている。しかし、以下に説明するように、提案されている方法は現在の放送実務とは相容れないか、又は全体的なエンターテインメントに必要に高い代価を求めるか、又はその両方となる。

【0006】

20

映画やテレビジョンのサラウンドオーディオを生じさせるときに、会話や談話の大部分を1つのチャンネル（中央チャンネル、スピーチチャンネルと称される）だけに配分することに固執する習慣が一般的となっている。音楽、環境音、及び効果音は、通常スピーチチャンネルと残りすべてのチャンネル（例えば、左[L]、右[R]左サラウンド[r s]、及び右サラウンド[r s]、ノンスピーチチャンネルと称されることもある）に混入される。結果として、スピーチチャンネルがオーディオプログラムに含まれるスピーチの大部分とノンスピーチオーディオのかなりの部分を伝達する一方、ノンスピーチチャンネルがノンスピーチオーディオの大部分を伝達するが、少量のスピーチも伝達することがある。このような従来の構成における会話や談話の認知を助けるひとつの簡単な方法は、ノンスピーチチャンのレベルを、スピーチチャンネルに比べて、例えば6dBほど恒久的に減少させることである。この方法は簡単で効果的であり今日よく用いられている（SRS[サウンドリトリバルシステム]による会話の明瞭化又はサラウンドデコーダーにおける修正したダウンミックス方程式）。しかし、この方法には少なくとも1つの欠点がある。すなわち、ノンスピーチチャンネルに一定の減衰を加えることで、スピーチの聴取を妨げないレベルの静かな環境音を聞こえないレベルにまで下げてしまうことがある。妨げとならない環境音を減衰させることにより、スピーチを理解するという付随的効果もなく、プログラムの審美的バランスを変えてしまう。

30

【0007】

代替的な解決策が、Vaudrey及びSaundersによる一連の特許（米国特許番号7,266,501、米国特許番号6,772,127、米国特許番号6,912,501、及び米国特許番号6,650,755）に記載されている。周知のとおり、これらの方法は、コンテンツの生成と分配を修正するものである。この構成によれば、消費者は2つの別々のオーディオ信号を受け取る。これらの信号の第1番目は「プライマリーコンテンツ」オーディオを具備する。この信号は多くの場合スピーチが優勢であるが、コンテンツの製作者が望むなら、他の形式の信号も含むことができる。第2番目の信号は、「セカンダリーコンテンツ」オーディオを具備し、残りのすべてのサウンド要素により構成されている。ユーザーは、各信号のレベルを手動で調整することにより、又は、ユーザーが選択したパワー比率を自動的に維持することにより、これら2つの信号の相対的レベルの制御ができる。この構成は、妨害とはならない環境音を必要に減衰させることを制限することができるが、従来の生成方法と分配方法には適用できないことが広く用いられる

40

50

ことの妨げとなっている。

【0008】

スピーチとノンスピーチオーディオとの相対的レベルを管理するための他の方法の例が、Bennettにより、米国特許出願番号20070027682中に提案されている。

【0009】

背景技術のすべての例において、他の欠点の中でも、会話を強化することがコンテンツの創作者が意図するリスニング体験に影響を与えることを最小限にするための手段を提供しないという制限を有する点で共通している。従って、ノンスピーチオーディオ成分の可聴性を維持したまま、スピーチを理解できるようにするために、従来の混合がなされたマルチチャンネルエンターテインメントプログラム中のノンスピーチオーディオチャンネルのレベルを制限する方法を提供することが本発明の目的である。10

【0010】

従って、スピーチの可聴性を維持する方法を改善することが必要となる。本発明は、マルチチャンネルオーディオ信号におけるスピーチの可聴性を改善する装置及び方法を提供することによりこれらの問題を解決するものである。

【発明の概要】

【0011】

本発明の実施の形態によりスピーチの可聴性が改善される。1つの実施の形態では、本発明はマルチチャンネルオーディオ信号におけるスピーチの可聴性を改善する方法を含む。該方法は、減衰係数を生成するためにマルチチャンネルオーディオ信号の第1の特性と第2の特性とを比較するステップを含む。第1の特性は、スピーチ及びノンスピーチオーディオを含有するマルチチャンネルオーディオ信号の第1番目のチャンネルに相当し、第2の特性は、主としてノンスピーチオーディオを含有するマルチチャンネルオーディオ信号の第2番目のチャンネルに相当する。該方法は、調整済みの減衰係数を生成するためにスピーチ尤度値に従って減衰係数を調整するステップをさらに含む。該方法は、この調整済みの減衰係数を用いて前記第2番目のチャンネルを減数させるステップをさらに含む。20

【0012】

本発明の第1の特徴は、一般的なエンターテインメントプログラムのスピーチチャンネルが、プログラムの継続期間の実質的な部分にノンスピーチ信号を伝達することの観察に基づく。その結果、本発明の第1の特徴によれば、ノンスピーチオーディオによるスピーチオーディオのマスキングを、(a)スピーチチャンネルにおける信号パワーに対するノンスピーチチャンネルにおける信号パワーの比率が所定の閾値を越えないように制限するために必要な、ノンスピーチチャンネルにおける信号の減衰量を決定するステップ、(b)スピーチ中のスピーチチャンネルにおける信号の尤度に単調に関係する係数により減衰量を縮小拡大するステップ、及び(c)縮小拡大した減衰量を適用するステップにより、制御することができる。30

【0013】

本発明の第2の特徴は、スピーチ信号のパワーとマスキング信号のパワーとの比率は、スピーチの明瞭度を予測するためには貧弱な判断材料であるとの観察に基づく。その結果、本発明の第2の特徴によれば、所定のレベルの明瞭度を維持するために必要な、ノンスピーチチャンネルにおける信号の減衰量を、心理音響に基づく明瞭度の予測モデルによりノンスピーチ信号が存在するところでのスピーチ信号の明瞭度を予測することにより、計算する。40

【0014】

本発明の第3の特徴は、周波数全体にわたって減衰量を変化させることができるのなら、(a)明瞭度の所定のレベルが種々の減衰のパターンにより達成することができること、及び(b)異なる減衰のパターンにより異なる音量のレベル又は異なるノンスピーチオーディオの主要部をもたらすことができることの観察に基づく。その結果、本発明の第3の特徴によれば、予測したスピーチの明瞭度の所定のレベルを達成する制限の下で、音量50

を最大にするか又はノンスピーチオーディオの主要部の他の測度を最大にする減衰のパターンを見つけることにより、ノンスピーチオーディオによるスピーチオーディオのマスキングを制御する

本発明の実施の形態は方法又はプロセスとして実行することができる。この方法は、ハードウェア又はソフトウェア又はそれらの組み合わせとしての電子回路により実施することができる。このプロセスを実施するために用いられる回路は、（特定の作業のみ行う）専用回路であっても（1以上の特定の作業を行うようにプログラムされた）汎用回路であってもよい。

【0015】

以下の詳細な説明と添付図により、本発明の本質及び利点をよく理解できる。

10

【図面の簡単な説明】

【0016】

【図1】本発明の一実施の形態による、信号プロセッサを示す。

【図2】本発明の他の実施の形態による、信号プロセッサを示す。

【図3】本発明の他の実施の形態による、信号プロセッサを示す。

【図4A】図1～3の実施の形態のさらなる変形を示すブロック図である。

【図4B】図1～3の実施の形態のさらなる変形を示すブロック図である。

【発明を実施するための形態】

【0017】

ここに記載したのは、スピーチの可聴性を維持するための技術である。以下の記載において、説明目的で、本発明を完全に理解してもらうために多くの実施例と具体的な詳細を述べる。しかしながら、当業者にとって、特許請求の範囲として定義される本発明は、以下に記載の実施例単独又は実施例の組み合わせにおける特徴のいくつか又はすべてを含み、さらに、ここに記載した特徴及び概念を修正したもの及び均等なものを含むことは自明である。

20

【0018】

種々の方法とプロセスを以下に記載する。これらは、主として理解しやすくするような順序で記載する。特定のステップについては、様々な実施の形態において、違う順序又は必要に応じた平行して行うことができることは理解できよう。特定のステップが他のステップの前又は他のステップの後でなければならないときは、文脈から明らかでない場合、具体的にその旨を指摘する。

30

【0019】

本発明の第1の形態についてその原理を図1に示す。図1を参照すると、スピーチチャンネル（101）及び2つのノンスピーチチャンネル（102及び103）からなるマルチチャンネル信号を受け取っている。これらのチャンネルの各々の信号のパワーは信号推定器（104, 105, 及び106）のバンクで測定し対数スケールで表現する。これらのパワー推定器は、漏れ積分回路のような平滑化機構を含むことができ、測定したパワーレベルが1つの文又は節全体にわたる平均化されたパワーレベルを反映することができる。スピーチチャンネル中の信号のパワーレベルを、（加算器107及び108により）ノンスピーチチャンネルの各々から減算し、2つの信号形式同士のパワーレベルの差異の測度を得る。比較回路109で、スピーチチャンネル中の信号のパワーレベルより少なくともd B少ないパワーレベルがのこるようにノンスピーチチャンネルを減衰するようなd Bの数値を、各ノンスピーチチャンネルに対して定める。（記号 α は、変数であり、スクリプト α を意味する。）1つの実施の形態によれば、この実施例では、閾値（回路110に保存されている）をパワーレベルの差異に加え（この中間的な結果はマージンと称される）、（リミッター111及び112により）その結果がゼロ以下になるように制限する。この結果は、パワーレベルをd Bだけスピーチチャンネルのパワーレベルより低くしておくために、ノンスピーチチャンネルに適用しなければならないd B単位で表したゲイン（又は否定減衰（negated attenuation））となる。に適した値は15 dBである。この値は他の実施形態において必要に応じて調整することが

40

50

できる。

【0020】

対数スケール(dB)で表現した測度と線形スケールで表現した測度との間に一意的な関係があるので、図1と等価な回路は、パワー、ゲイン、及び閾値がすべて線形スケールで表現することができるように、構成することができる。この実施の形態では、すべてのレベルの差異を線形測度の比率で置き換えることができる。代替的実施の形態では、パワー測度を、信号の絶対値のような信号強度に対応する測度で置き換えることができる。

本発明の第1の形態において特筆すべき特徴は、実際にスピーチがなされたスピーチチャンネルにおける信号の尤度に単調に関連する値により導き出されたゲインを拡大縮小することである。さらに図1を参照して、制御信号(113)を受け取りゲインが(乗算器114及び115により)乗算される。拡大縮小されたゲインは、(增幅器116及び117により)対応するノンスピーチチャンネルに適用され、修正した信号L及びR(118及び119)を作る。制御信号(113)は、通常、スピーチがなされたスピーチチャンネルにおける信号の尤度の、自動的に導き出された測度となる。スピーチ信号となった信号の尤度を自動的に決定する様々な方法を用いることができる。1つの実施の形態によれば、スピーチ尤度130は、Cチャンネル101における情報からスピーチ尤度値p(113)を生成する。このような機構の1つの例は、Robinson及びVintonによる「Automated Speech/Other Discrimination for Loudness Monitoring」(2005年5月のAudio Engineering Society, Preprint number 6437 of Convention 118)に記載されている。あるいは、この制御信号(113)は、例えば、手動で作成し、コンテンツ制作者によりオーディオ信号と共にエンジニアに送信することができる。

【0021】

本発明の属する技術の分野における通常の知識を有する者(当業者)はこの構成をどんな数の入力チャンネルにも拡張することができることを理解するであろう。

【0022】

本発明の第2の形態の原理は、図2に示されている。図2を参照すると、1つのスピーチチャンネル(101)及び2つのノンスピーチチャンネル(102 及び 103)からなるマルチチャンネル信号を受け取っている。これらのチャンネルの各々の信号のパワーは信号推定器(201 , 202 , 及び 203)のバンクで測定される。図1の対応する部分とは異なり、これらのパワー推定器は周波数全体にわたって信号パワーの分布を測定し、単数ではなくパワースペクトルを結果として生じさせる。このパワースペクトルの周波数分解能は、明瞭度予測モデル(205 及び 206 、未説明)の周波数分解能と理想的にマッチする。

【0023】

パワースペクトルは比較回路204に送られる。このブロックの目的は、各ノンスピーチチャンネルに適用する減衰量を定め、ノンスピーチチャンネルの信号がスピーチチャンネルの信号の明瞭度を所定の基準以下に下がらないようにすることである。この機能はスピーチ信号(201)とノンスピーチ信号(202 及び 203)のパワースペクトルから、スピーチの明瞭度を予測する明瞭度予測回路(205 及び 206)を採用することにより達成することができる。この明瞭度予測回路205及び206は、選択とトレードオフの設計に従い適切な明瞭度予測モデルを組み込むことができる。例として、ANSI S3.5-1997(スピーチ明瞭度指数の計算方法)に定められたスピーチ明瞭度指数及び、Muessch及びBuusによるスピーチ認識感度(「スピーチ明瞭度予測のための統計的決定理論の使用。Iモデル構造」Journal of the Acoustic al Society of America , 2001 , Vol 109 , P2896-2909)がある。明瞭度予測モデルの出力は、スピーチチャンネル中のスピーチ以外の信号が平滑化されるとき、意味を持たないことは明らかである。それにもかかわらず、明瞭度予測モデルの出力に続くものは、予測スピーチ明瞭度と称される。このような誤りを

10

20

30

40

50

理解することで、スピーチがなされた信号(113、未説明)の尤度に関するパラメータで、比較回路204からのゲイン出力を縮小拡大することによる、次の処理を行うことの説明とすることができます。

【0024】

明瞭度予測モデルは一般に、ノンスピーチ信号のレベルを下げた結果として、上がるか又は変化しないスピーチ明瞭度の予測をおこなう。図2の処理フローを続けると、比較回路207及び208で、予測明瞭度を基準値と比較する。ノンスピーチ信号のレベルが低く予測明瞭度が基準を超えている場合、0dBに初期化されているゲインパラメータが回路209又は210から取り出され比較回路204の出力として回路211及び212に供給される。基準を満たさない場合は、ゲインパラメータを、ある定められた量だけ減少させ、明瞭度の予測を繰り返す。適切なステップサイズは1dBである。ここに記載したような繰り返しを予測明瞭度が基準値に一致するまで又は基準値を越えるまで続けられる。スピーチチャンネル中の信号は、ノンスピーチチャンネル中に信号がない場合でも明瞭度の基準に到達することができないようにすることも可能である。このような状態になる例として、スピーチ信号が非常に低レベルである場合、又は厳しく帯域幅が制限されている場合がある。このようなことが起こると、ノンスピーチチャンネルに適用するゲインをどれだけ減少させても予測スピーチ明瞭度に影響を与えることはなく、基準値を満たさないという状態になってしまう。このような状態では、(205, 206), (207, 208), 及び(209, 210)により形成されるループが永久に続くことなり、このようなループを中断させるためには付加的なロジック(不図示)を適用する必要がある。そのようなロジックの簡単な例は、繰り返しの数を数え、繰り返しが所定の数を越えたらループを抜け出すようにすることである。10

【0025】

図2の処理フローを続けて、制御信号(113)を受け取り、(乗算器113及び115により)ゲインを乗算する。制御信号(113)は、一般に、スピーチが行われたスピーチチャンネルにおける信号の自動的に導き出された尤度の測度となる。スピーチ信号となる信号の尤度を自動的に測定する方法は、自明であり、図1に関して既に説明したとおりである(スピーチ尤度プロセッサ130参照)。縮小拡大したゲインは(増幅器116及び117により)対応するノンスピーチチャンネルに適用され、修正した信号R'及びL'(118及び119)を生成する。20

【0026】

本発明の第3の形態の原理は、図3に示されている。ここで図3を参照すると、1つのスピーチチャンネル(101)と2つのノンスピーチチャンネル((102及び103)からなるマルチチャンネル信号を受け取っている。この3つの信号の各々は(フィルターバンク301, 302, 及び303により)スペクトル成分に分割される。スペクトル分析は時間領域Nチャンネルフィルターバンクにより実行することができる。1つの実施の形態によれば、フィルターバンクは周波数領域を1/3オクターブの帯域に区分するか、又は、人の内耳で生じるようなフィルタリングに似せる。ここで、信号はN個のサブ信号からなることを太線で示している。図3の処理はサイドブランチ処理として知られている。信号経路の後に、ノンスピーチチャンネルを形成するN個のサブ信号はそれぞれN個のゲイン値のセットの1要素により(増幅器116及び117により)縮小拡大される。これらのゲインの導出については後述する。次に、縮小拡大されたサブ信号は、再結合して単一のオーディオ信号となる。これは(回路313及び314による)単純な加算により行われる。代替的に、分析フィルターバンクに適合する合成フィルターバンクを用いることができる。この処理の結果として、修正した信号R'及びL'(118及び119)が得られる。40

【0027】

ここで、図3の処理におけるサイドブランチ経路について説明すると、各フィルターバンクの出力は、N個のパワー推定器の対応するバンク(304, 305, 及び306)で使用可能となる。結果として生じたスペクトルは、N次元のゲインベクトルを出力として

10

20

30

40

50

有する、最適化回路（307及び308）の入力となる。この最適化では、明瞭度予測回路（309及び310）及び音量計算回路（311及び312）の両方を採用して、スピーチ信号の予測される明瞭度の所定のレベルを維持しながらノンスピーチチャンネルの音量を最大化するゲインベクトルを見つける。明瞭度を予測する適切なモデルは図2に関連して既に説明したとおりである。音量計算回路311及び312は、選択とトレードオフの設計に従い適切な音量予測モデルを組み込むことができる。適切なモデルの例として、米国規格ANSI S3.4-2007「Procedure for the Computation of Loudness of Steady Sounds」及びドイツ規格DIN 45631「Berechnung des Lautstaerkepegels und der Lautheit aus dem Geräuschspektrum」がある。10

【0028】

利用可能な計算資源及び課せられた制約条件に応じて、最適化回路（307及び308）の形及び複雑さは大きく変化する。1つの実施の形態によれば、N個の自由パラメータの、繰り返し、多次元制約最適化を用いることができる。各パラメータは、ノンスピーチチャンネルの周波数帯域の1つに適用するゲインを表す。N次元検索スペースにおける最急勾配法のような標準的な技法を、最大値を見つけるために適用することができる。他の実施の形態においては、計算的に厳しくないアプローチではゲイン対周波数関数を、異なるスペクトル勾配のセット又はシェルフィルターのような、可能性のあるゲイン対周波数関数の小さなセットのメンバーに限定する。この追加の限定により、最適化の問題は、小さな数の1次元の最小化に縮減することができる。さらに他の実施の形態では、可能なゲインの非常に小さなセットに対して徹底検索を実行する。この後者のアプローチは、一定の計算負荷と検索速度が必要な場合に、実時間のアプリケーションで特に好ましいであろう。20

【0029】

当業者は、本発明の追加的な実施の形態による最適化に組み込むことのできる付加的な構成を容易に認識することができる。1つの実施例では、修正したノンスピーチチャンネルの音量が修正前の音量より大きくならないよう制限する。他の実施例では、再構成フィルターバンク（313, 314）に時間的なエイリアシングが生じる潜在的 possibility を制限するため又は好ましくない音色の変化が生じる可能性を減少させるために隣接する周波数帯域同士でのゲインの差の制限が組み込まれる。好ましい制約条件は、フィルターバンクの技術的な実施の形態と、明瞭度の完全と音色の変化との間のトレードオフをどのように選択するかとに左右される。説明を簡単にするために、これらの制約は図3から省略した。30

【0030】

図3の処理フローを続けると、制御信号p（113）を受け取り、（乗算器114及び115で）ゲイン関数を乗算する。制御信号（113）は、通常、自動的に導き出したスピーチが行われたスピーチチャンネルにおける信号の尤度の測度となる。スピーチが行われた信号の尤度の自動的な計算の方法については図1と関連して既に説明した（スピーチ尤度プロセッサ130参照）。縮小拡大させたゲインは、先に説明したように、（増幅器116及び117）により、対応するノンスピーチチャンネルに適用する。40

【0031】

図4A及び4Bは、図1～3に示した形態の変形を示したものである。なお、当業者であれば、図1～3に記載した発明の要素を組み合わせるいくつかの方法を考え付くであろう。

【0032】

図4Aは、図1の構成を、L, C, 及びRの1以上の周波数サブ帯域にも適用することができることを示している。具体的には、信号L, C, 及びRは、それぞれフィルターバンク（441, 442, 及び443）を通過し、サブ帯域の3つのセット、すなわち{L₁, L₂, ..., L_n}, {C₁, C₂, ..., C_n}, 及び{R₁, R₂, ...} 50

, R_n)、を生じさせる。サブ帯域のマッチングは、図1に示すn個のインスタンスの回路125に送られ、処理されたサブ信号は(加算回路451及び452により)再結合される。個々の閾値nは、各サブ帯域で選択することができる。nが対応する周波数領域にあるスピーチのキューブの平均値に比例するところ、すなわち、極端な周波数スペクトルにおける帯域が、対応する優勢なスピーチ周波数の帯域より低い閾値に割り当てられるところに良い選択が設定される。本発明のこの実施の形態により、コンピュータの複雑さとパフォーマンスとの間の良好なトレードオフがもたらされる。

【0033】

図4Bは、もう1つの変形を示す。例えば、計算負荷を軽減するために、5チャンネル(C, L, R, I_s, 及び r_s)を有する一般的なサラウンドサウンド信号を、図3に示した回路325によりL及びR信号を処理し、図1に示した回路125により、L及びR信号より一般にパワーの少ない、I_s及びr_sを処理することにより、改善することができる。

10

【0034】

上記説明において、用語「スピーチ」(又はスピーチオーディオ又はスピーチチャンネル又はスピーチ信号)及び用語「ノンスピーチ」(又はノンスピーチオーディオ又はノンスピーチチャンネル又はノンスピーチ信号)が使われている。当業者であればこれらの用語は、相互に違っていることを示すために主として用い、チャンネルのコンテンツを完全に説明するためにほとんど用いないことを理解するであろう。例えば、映画のレストランのシーンで、スピーチチャンネルには主として1つのテーブルでの会話が含まれ、ノンスピーチチャンネルには、他のテーブルでの会話が含まれるであろう(つまり、両方とも、専門家でない人が使う用語としての「スピーチ」が含まれる)。他のテーブルでの会話であっても、本発明の実施の形態では、減衰させる。

20

【0035】

(実施形態)

[実施形態]

本発明は、ハードウェア又はソフトウェア又は両方を組み合わせたもの(例えば、プログラマブルロジックアレー)で実施することができる。特に記載がない限り、本発明の一部として含まれているアルゴリズムは本質的に、特定のコンピュータや他の装置と関連付けられるものではない。特に、種々の汎用機をこの記載に従って書かれたプログラムと共に用いてもよい、あるいは、要求の方法を実行するために、より特化した装置(例えば、集積回路)を構成することが便利かもしれない。このように、本発明は、それぞれ少なくとも1つのプロセッサ、少なくとも1つの記憶システム(揮発性及び非揮発性メモリー及び/又は記憶素子を含む)、少なくとも1つの入力装置又は入力ポート、及び少なくとも1つの出力装置又は出力ポートを具備する、1つ以上のプログラマブルコンピュータシステム上で実行される1つ以上のコンピュータプログラムにより実現することができる。ここに記載した機能を遂行し、出力情報を出力させるために入力データにプログラムコードを適用する。この出力情報は、公知の方法で、1以上の出力装置に適用される。

30

【0036】

このようなプログラムの各々は、コンピュータシステムとの通信のために、必要とされるどんなコンピュータ言語(機械語、アセンブリ、又は、高級な、手続言語、論理型言語、又は、オブジェクト指向言語を含む)ででも実現することができる。いずれにせよ、言語はコンパイル言語であってもインタープリタ言語であってもよい。

40

【0037】

このようなコンピュータプログラムの各々は、ここに記載の手順を実行するために、コンピュータにより記憶媒体又は記憶装置を読み込んだとき、コンピュータを設定し動作させるための、汎用プログラマブルコンピュータ又は専用プログラマブルコンピュータにより、読み込み可能な記憶媒体又は記憶装置(例えば、半導体メモリー又は半導体媒体、又は磁気媒体又は光学媒体)に保存又はダウンロードすることが好ましい。本発明のシステムはまた、コンピュータプログラムにより構成されるコンピュータにより読み込み可能な

50

記憶媒体として実行することを考えることもできる。ここで、この記憶媒体は、コンピュータシステムを、ここに記載した機能を実行するために、具体的にあらかじめ定めた方法で動作させる。

【0038】

どのように本発明の形態を実施するかについての実施例と共に、本発明の多くの実施の形態について記載した。上記実施例及び実施の形態は、唯一の実施形態であるとみなしてはならず、以下の特許請求の範囲で定義した本発明の柔軟性と優位性を示すため記載したものである。上記記載と以下の特許請求の範囲の記述に基づき、他の構成、実施形態、実施例、及びその均等物は、当業者にとって自明であり、当業者であれば特許請求の範囲で定義した本発明の精神と技術範囲を逸脱することなく採用することができよう。

10

【符号の説明】

【0039】

101	スピーチチャンネル
102, 103	ノンスピーチチャンネル
201, 202, 203	パワースペクトル密度
205, 206	明瞭度予測
304, 305, 306	N帯域パワースペクトル
309, 310	明瞭度予測

【図1】

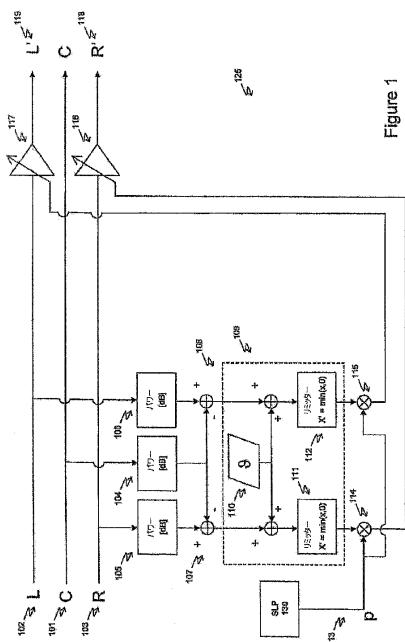


Figure 1

【図2】

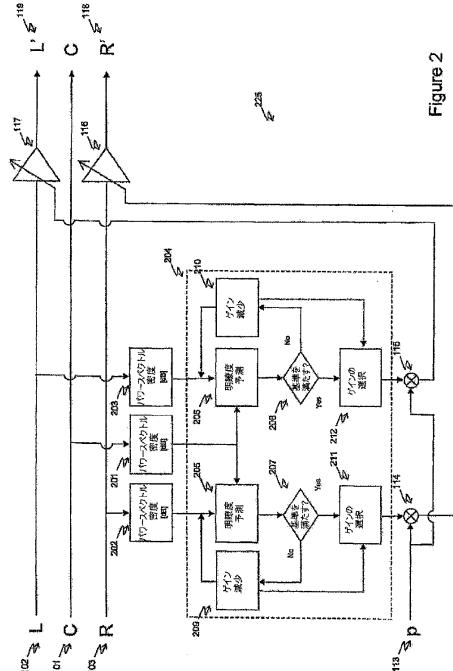


Figure 2

【図3】

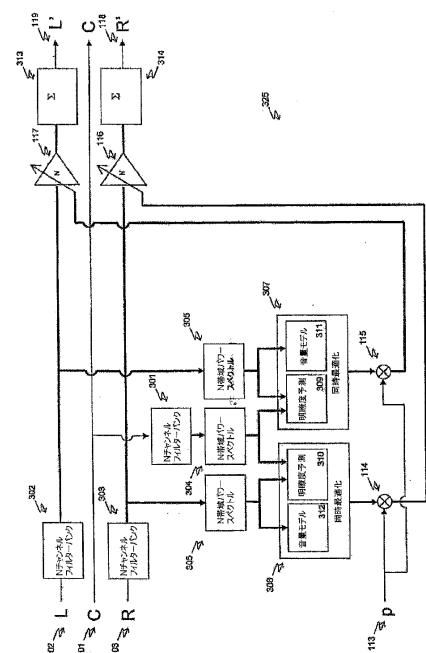


Figure 3

【図4A】

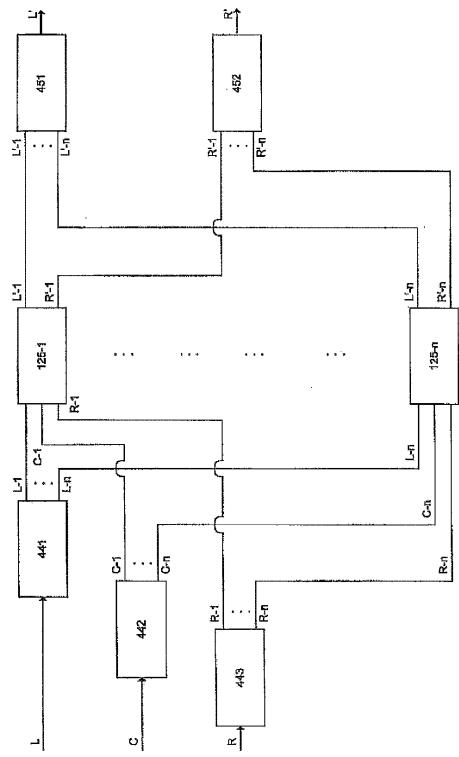


Figure 4A

【図4B】

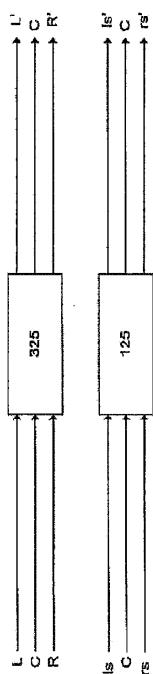


Figure 4B

フロントページの続き

(74)代理人 100121762
弁理士 杉山 直人

(74)代理人 100126767
弁理士 白銀 博

(74)代理人 100118647
弁理士 赤松 利昭

(74)代理人 100138438
弁理士 尾首 亘聰

(72)発明者 ミュシュ、ハンネス
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、ポトレロ・アベニュー 100

審査官 菊池 充

(56)参考文献 特開2007-142856 (JP, A)
特表2005-502247 (JP, A)
特表2005-531990 (JP, A)
特開平09-084198 (JP, A)
特開2007-208755 (JP, A)
特表平09-500503 (JP, A)
米国特許第6862567 (US, B1)
特開2002-335490 (JP, A)
特開2007-158873 (JP, A)
特開2001-245237 (JP, A)

(58)調査した分野(Int.Cl., DB名)

H04S 1/00 - 7/00

G10L 11/00 - 13/08

G10L 19/00 - 21/06