(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7580495号
(P7580495)

(45)発行日　令和6年11月11日(2024.11.11)

(24)登録日　令和6年10月31日(2024.10.31)

(51)国際特許分類
Ｇ１０Ｌ　25/69　(2013.01)

ＦＩ
Ｇ１０Ｌ　25/69

請求項の数　20　（全22頁）

(21)出願番号	特願2022-573351(P2022-573351)
(86)(22)出願日	令和2年5月29日(2020.5.29)
(65)公表番号	特表2023-530225(P2023-530225A)
(43)公表日	令和5年7月14日(2023.7.14)
(86)国際出願番号	PCT/EP2020/065035
(87)国際公開番号	WO2021/239255
(87)国際公開日	令和3年12月2日(2021.12.2)
審査請求日	令和5年1月30日(2023.1.30)

(73)特許権者	500242786 フラウンホファー ゲセルシャフト ツール フェールデルンク ダー アンゲヴァンテン フォルシュンク エー．ファオ． ドイツ連邦共和国 ８０６８６ ミュンヘン，ハンサシュトラーセ ２７ツェー
(74)代理人	100108453 弁理士　村山 靖彦
(74)代理人	100110364 弁理士　実広 信哉
(74)代理人	100133400 弁理士　阿部 達彦
(72)発明者	ヤン・レニース－ホッホムート ドイツ・２６１２９・オルデンブルク・マリー－キュリー－シュトラーセ・２・
最終頁に続く

(54)【発明の名称】　初期オーディオ信号を処理するための方法および装置

(57)【特許請求の範囲】

【請求項１】

対象部分(AS_TP)および副部分(AS_SP)を含む初期オーディオ信号(AS)を処理するための方法(100)であって、

a．前記初期オーディオ信号(AS)を受信するステップと、

b．第1の信号調整器を使用することによって、受信した前記初期オーディオ信号(AS)を調整して(110，110a)、第1の調整されたオーディオ信号(1st MOD AS)を取得し、

第2の信号調整器を使用することによって、受信した前記初期オーディオ信号(AS)を調整して(110，110b)、第2の調整されたオーディオ信号(2nd MOD AS)を取得するステップと、

c．前記第1の調整されたオーディオ信号を評価基準に対して評価して(120，120a)、前記評価基準の満足度を表す第1の評価値(1st PSV)を取得し、

前記第2の調整されたオーディオ信号を前記評価基準に対して評価して(120，120b)、前記評価基準の満足度を表す第2の評価値(2nd PSV)を取得するステップと、

d．それぞれの前記第1のまたは第2の評価値(1st PSV，2nd PSV)によって決まる前記第1のまたは第2の調整されたオーディオ信号(1st MOD AS，2nd MOD AS)を選択するステップ(130)と

を含み、

選択する前記ステップが、複数の独立した第1の評価値および独立した第2の評価値に基づいて、または少なくとも2つの独立した評価基準に基づいて行われ、

前記評価基準が、知覚的類似度であり、前記ステップcが、
受信した前記初期オーディオ信号(AS)を前記第1の調整されたオーディオ信号(1st MOD AS)と比較して(120，120a)、前記初期オーディオ信号(AS)と前記第1の調整されたオーディオ信号(1st MOD AS)との間の前記知覚的類似度を表す第1の評価値として第1の知覚的類似度値(1st PSV)を取得するサブステップと、
受信した前記初期オーディオ信号(AS)を前記第2の調整されたオーディオ信号(2nd MOD AS)と比較して(120，120b)、前記初期オーディオ信号(AS)と前記第2の調整されたオーディオ信号(2nd MOD AS)との間の前記知覚的類似度を表す第2の評価値として第2の知覚的類似度値(2nd PSV)を取得するサブステップと
を含む、方法(100)。

【請求項２】
前記評価基準が、
- 第1のおよび第2の知覚的類似値によって表される知覚的類似度であって、前記第1のおよび第2の知覚的類似値が、それぞれの前記第1のおよび第2の調整されたオーディオ信号と前記初期オーディオ信号(AS)との間の知覚的類似度を表す、知覚的類似度、
- 対象またはしきい値と比較するための語音明瞭度の計算された値の形式での語音明瞭度、
- ラウドネス値によって表されるラウドネス、
- サウンドパターン、
- 空間度
を含むグループの中にある、請求項1に記載の方法(100)。

【請求項３】
前記第1の調整されたオーディオ信号に対する少なくとも2つの独立した評価基準についての満足度を表すそれぞれの第1の評価値、および前記第2の調整されたオーディオ信号に対する前記少なくとも2つの独立した評価基準についての満足度を表すそれぞれの第2の評価値が決定されるように、前記少なくとも2つの独立した評価基準が別々に評価された後に、前記選択が、重み付けされた第1のおよび第2の評価値に基づいて行われる、請求項1または2に記載の方法(100)。

【請求項４】
前記第1の知覚的類似度値(1st PSV)が前記第2の知覚的類似度値(2nd PSV)よりも高いとき、前記第1の調整されたオーディオ信号(1st MOD AS)のより高い知覚的類似度を示すように、前記第1の調整されたオーディオ信号(1st MOD AS)が選択され、
前記第2の知覚的類似度値(2nd PSV)が前記第1の知覚的類似度値(1st PSV)よりも高いとき、前記第2の調整されたオーディオ信号(2nd MOD AS)のより高い知覚的類似度を示すように、前記第2の調整されたオーディオ信号(2nd MOD AS)が選択される、
請求項1から3のいずれか一項に記載の方法(100)。

【請求項５】
前記ステップdの選択によって決まる前記第1のまたは第2の調整されたオーディオ信号(1st MOD AS，2nd MOD AS)を出力するステップをさらに含む、請求項1から4のいずれか一項に記載の方法(100)。

【請求項６】
前記初期オーディオ信号(AS)を出力するステップは、それぞれの前記第1のまたは第2の知覚的類似度値(1st PSV，2nd PSV)がしきい値を下回るとき、前記第1のまたは第2の調整されたオーディオ信号(1st MOD AS，2nd MOD AS)を出力する代わりに行われ、前記しきい値を下回ると、それぞれの第1のまたは第2の調整されたオーディオ信号(1st MOD AS，2nd MOD AS)は、前記初期オーディオ信号(AS)との類似が十分でないと示される、請求項3に記載の方法(100)。

【請求項７】
前記対象部分(AS_TP)が前記初期オーディオ信号(AS)の語音部分であり、前記副部分(AS_SP)が前記初期オーディオ信号(AS)の周囲雑音部分である、請求項1から6のいずれか一

項に記載の方法(100)。
【請求項８】
前記第1のおよび/または第2の調整されたオーディオ信号(1st MOD AS，2nd MOD AS)が、前景に移動した前記対象部分(AS_TP)、および背景に移動した前記副部分(AS_SP)、ならびに/または前景に移動した前記対象部分(AS_TP)として語音部分、および背景に移動した前記副部分(AS_SP)として周囲雑音部分を含む、請求項1から7のいずれか一項に記載の方法(100)。
【請求項９】
比較する前記ステップが、知覚モデル、PEAQモデル、POLQAモデル、および/またはPEMO-Qモデルを使用することによって、前記第1のおよび/または第2の評価値(1st PSV，2nd PSV)を抽出するステップを含む、請求項1から8のいずれか一項に記載の方法(100)。
【請求項１０】
前記第1のおよび/または第2の評価値(1st PSV，2nd PSV)が、前記第1のもしくは第2の調整されたオーディオ信号(1st MOD AS，2nd MOD AS)の物理パラメータ、前記第1のもしくは第2の調整されたオーディオ信号(1st MOD AS，2nd MOD AS)の音量レベル、前記第1のもしくは第2の調整されたオーディオ信号(1st MOD AS，2nd MOD AS)についての心理音響的音響パラメータ、前記第1のもしくは第2の調整されたオーディオ信号(1st MOD AS，2nd MOD AS)のラウドネス情報、前記第1のもしくは第2の調整されたオーディオ信号(1st MOD AS，2nd MOD AS)のピッチ情報、ならびに/または前記第1のもしくは第2の調整されたオーディオ信号(1st MOD AS，2nd MOD AS)の知覚されたソース幅情報によって決まる、請求項1から9のいずれか一項に記載の方法(100)。
【請求項１１】
前記第1のおよび/もしくは第2の信号調整器が、前記初期オーディオ信号(AS)についてのSNR増加、ダイナミック圧縮、および／もしくはSNR減少を行うように構成され、ならびに/または
調整する前記ステップは、前記初期オーディオ信号(AS)が別個の対象部分(AS_TP)および別個の副部分(AS_SP)を含む場合、前記対象部分(AS_TP)を増加させるステップと、前記対象部分(AS_TP)についての周波数重み付けを増加させるステップと、前記対象部分(AS_TP)をダイナミックに圧縮するステップと、前記副部分(AS_SP)を減少させるステップと、前記副部分(AS_SP)についての周波数重み付けを減少させるステップとを含み、ならびに/または
調整する前記ステップは、前記初期オーディオ信号(AS)が組み合わさった対象部分(AS_TP)と副部分(AS_SP)とを含む場合、前記対象部分(AS_TP)および前記副部分(AS_SP)の分離を行うステップを含む、
請求項1から10のいずれか一項に記載の方法(100)。
【請求項１２】
選択する前記ステップ(130)が、以下の因子、すなわち、
- 聴覚障害のある人の難聴のグレード、
- 個人の聴覚性能、
- 個人の周波数依存聴覚性能、
- 個人の嗜好、
- 信号調整率に関する個人の嗜好
のうちの1つまたは複数を考慮に入れて行われる、請求項1から11のいずれか一項に記載の方法(100)。
【請求項１３】
調整する前記ステップ(110)、および/または比較する前記ステップ(120)が、以下の因子、すなわち、
- 聴覚障害のある人の難聴のグレード、
- 個人の聴覚性能、

- 個人の周波数依存聴覚性能、
- 個人の嗜好、
- 信号調整率に関する個人の嗜好

のうちの1つまたは複数を考慮に入れて行われる、請求項1から12のいずれか一項に記載の方法(100)。

【請求項１４】

個人の嗜好を定義する最適化対象に関する情報を受信するステップをさらに含み、

前記評価基準が、前記最適化対象によって決まり、または調整する前記ステップおよび/もしくは評価する前記ステップおよび/もしくは選択する前記ステップが、前記最適化対象によって決まり、または選択する前記ステップについて独立した評価基準を表す独立した第1のおよび第2の評価値の重み付けが、前記最適化対象によって決まる、請求項1から13のいずれか一項に記載の方法(100)。

【請求項１５】

比較する前記ステップ(120)が、

前記初期オーディオ信号(AS)の全体、および前記第1のおよび第2の調整されたオーディオ信号(1st MOD AS，2nd MOD AS)の全体について、ならびに/または

前記初期オーディオ信号(AS)の前記対象部分(AS_TP)、および前記第1のおよび第2の調整されたオーディオ信号(1st MOD AS，2nd MOD AS)のそれぞれの対象部分(AS_TP)について、ならびに/または

前記初期オーディオ信号(AS)の前記副部分(AS_SP)、および前記第1のおよび第2の調整されたオーディオ信号(1st MOD AS，2nd MOD AS)のそれぞれの副部分(AS_SP)について、

行われる、請求項1から13のいずれか一項に記載の方法(100)。

【請求項１６】

前記初期オーディオ信号(AS)が、複数の時間フレームを含み、

前記ステップaからdが、各時間フレームについて繰り返され、および/または

前記ステップaからdが、前記初期オーディオ信号(AS)のシーンの時間部分または時間フレームについて繰り返される、

請求項1から15のいずれか一項に記載の方法(100)。

【請求項１７】

複数の時間フレームを含む前記初期オーディオ信号(AS)の適応が、前記適応が必要である時間フレームについて、および知覚的連続性を維持するために他の時間フレームについて行われ、または複数の時間フレームを含む前記初期オーディオ信号(AS)の適応が、前記適応が必要である時間フレームについて、および知覚的連続性を維持するために他の時間フレームについて補間された形で行われ、ならびに/または

第1のおよび第2の後続の時間フレームの適応が、前記第1の後続の時間フレームと前記第2の後続の時間フレームとの間の遷移が知覚的連続性を維持するように形成されるように行われる、

請求項1から16のいずれか一項に記載の方法(100)。

【請求項１８】

初期ステップをさらに含み、前記初期ステップは、

語音部分を決定するために初期オーディオ部分を分析するステップ(21)と、

前記初期オーディオ信号(AS)の語音明瞭度について評価するために前記語音部分と周囲雑音部分とを比較するステップと、

前記語音明瞭度について示す値がしきい値を下回る場合、調整する前記ステップのための前記第1のおよび/または第2の信号調整器をアクティブ化するステップと

を含む、請求項1から17のいずれか一項に記載の方法(100)。

【請求項１９】

コンピュータにおいて動作すると、前記コンピュータに請求項1から18のいずれか一項に記載の方法を行わせるプログラムコードを有するコンピュータプログラム。

【請求項２０】
対象部分(AS_TP)および副部分(AS_SP)を含む初期オーディオ信号(AS)を処理するための装置であって、
前記初期オーディオ信号(AS)を受信するためのインターフェースと、
受信した前記初期オーディオ信号(AS)を調整して(110)、第1の調整されたオーディオ信号(1st MOD AS)を取得するための第1の信号調整器(11)、および受信した前記初期オーディオ信号(AS)を調整して、第2の調整されたオーディオ信号(2nd MOD AS)を取得するための第2の信号調整器(11)と、
前記第1の調整されたオーディオ信号を評価規準に対して評価して(120，120a)、前記評価規準の満足度を表す第1の評価値(1st PSV)を取得するための、および前記第2の調整されたオーディオ信号を前記評価規準に対して評価して(120，120b)、前記評価規準の満足度を表す第2の評価値(2nd PSV)を取得するための評価器と、
それぞれの前記第1のまたは第2の評価値(1st PSV，2nd PSV)によって決まる前記第1のまたは第2の調整されたオーディオ信号(1st MOD AS，2nd MOD AS)を選択するため(130)の選択器(13)と
を備え、
前記選択が、複数の独立した第1の評価値および独立した第2の評価値に基づいて、または少なくとも2つの独立した評価基準に基づいて行われ、
前記評価基準が、知覚的類似度であり、前記評価器が、
受信した前記初期オーディオ信号(AS)を前記第1の調整されたオーディオ信号(1st MOD AS)と比較して(120，120a)、前記初期オーディオ信号(AS)と前記第1の調整されたオーディオ信号(1st MOD AS)との間の前記知覚的類似度を表す第1の評価値として第1の知覚的類似度値(1st PSV)を取得し、
受信した前記初期オーディオ信号(AS)を前記第2の調整されたオーディオ信号(2nd MOD AS)と比較して(120，120b)、前記初期オーディオ信号(AS)と前記第2の調整されたオーディオ信号(2nd MOD AS)との間の前記知覚的類似度を表す第2の評価値として第2の知覚的類似度値(2nd PSV)を取得する
ように構成される、装置。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明の実施形態は、(レコーディング、または生データのような)初期オーディオ信号を処理するための方法、および対応する装置に関する。好ましい実施形態は、語音(speech)明瞭度を改善するための、および放送オーディオ素材を傾聴するための(方法およびアルゴリズムによる)手法に関する。
【背景技術】
【０００２】
オーディオ媒体およびオーディオビジュアル媒体(たとえば、映画、TV、ラジオ、ポッドキャスト、YouTube(登録商標)動画)を制作し放送する場合、たとえば、過度な背景サウンド(レコーディングにおける音楽、サウンド効果、雑音など)が追加されることにより、最終的なサウンドミキシングにおける語音明瞭度の高さが十分であることがつねに保証されるとは限らない。
【０００３】
このことは、特に、聴覚障害をもつ人たちにとって問題となるが、語音明瞭度を改善することは、正常聴覚をもつ人たちまたはネイティブスピーカでないリスナーにとっても有利になる。
【０００４】
オーディオ媒体およびオーディオビジュアル媒体を制作する際の基本問題は、背景信号(音楽、サウンド効果、雰囲気)が制作物の重要なサウンド美学の一部を作り上げていること、すなわち、これを、できるだけ排除すべきである「干渉雑音」と見なすことができない

ことである。そのため、この用途の語音明瞭度の改善または傾聴努力の低減を目的とするすべての方法では、サウンド制作の高い品質要件と創造的側面に対処するためには、元の意図されたサウンドの性質だけはできるだけ変えないようにすることをさらに考慮すべきである。しかしながら、現在のところ、良好な明瞭度とサウンドシーン/レコーディングの維持との間の最適なトレードオフを保証するための技術的な方法またはツールは一つも存在していない。

【０００５】

しかしながら、オーディオ媒体およびオーディオビジュアル媒体における語音明瞭度の改善(または傾聴努力の低減)を基本的にもたらすことができる様々な技術的手法が存在する。

【０００６】

専門のサウンドエンジニアがオルタナティブなオーディオミックスを手作業で制作することが、1つの解決策になり得、それにより、エンドユーザは、元のミックスと語音明瞭度が改善したミックスとのうちから自由に選択するようになり得る。語音明瞭度が改善したミックスは、たとえば、聴覚損失(hearing loss)シミュレーションを採用して、意図したミックスが対象の聴覚損失を伴うリスナーにも適していることを確認することによって、制作されることもあり得る[1]。しかしながら、そのような手作業での方法は、非常にコストが大きくなり、制作されたオーディオ/オーディオビジュアル媒体の大部分に適用することができなくなる。

【０００７】

自動信号強調をもたらす別の解決策として、望ましくない信号部分(たとば、干渉雑音)を低減または除去するための様々な方法が存在するが、これらは、本発明の技術的手法とは異なる。

【０００８】

ミックスド信号の干渉雑音低減方法による語音明瞭度の改善:そのような方法は、対象信号(たとえば、語音)ならびに干渉信号(たとえば、背景雑音)をともに含むミックスド信号を、干渉雑音のできるだけ大部分を除去する一方、対象信号は、理論上、そのままに維持するように処理することを目的としている(たとえば、[2]による方法)。これらの方法では、ミックスド信号における対象成分および干渉雑音成分のそれぞれの部分を推定しなくてはならないので、方法は、信号成分の物理的特性に関する仮定につねに基づいている。そのようなアルゴリズムは、たとえば、補聴器およびモバイルフォンにおいて使用されており、これは、従来技術であり、継続的に開発が進められている。

【０００９】

近年、ミックスド信号における異なるソースを分離することを目的とする機械学習(ニューロンネットワーク)に基づく方法がますます提示されてきている。大量のデータに基づいて、これらの方法は、特定の問題(たとえば、ミックスにおいていくつかの話者を分離すること[3])向けに訓練され、基本的には、オーディオビジュアル媒体における雰囲気/音楽から会話を抽出するのに使用され得、そのため、SNRが改善されたリミックスの基礎を提供することができる。[4]においては、ユーザに語音と背景との比率を自分で調節するオプションを与えるためのそのような手法が提示されている。

【００１０】

語音信号を事前処理することによる語音明瞭度の改善:いくつかの用途においては、対象信号(たとえば、語音)は、他の信号部分とは分離しており、そのため、これは、上述のミックスド信号ではなく、この方法では、どの信号成分が対象と干渉雑音とに対応しているかについて推定する必要はない。これは、たとえば、列車駅アナウンスの場合である。同時に、信号処理レベルでは、干渉雑音が影響を受ける可能性はなく、すなわち、干渉雑音(たとえば、駅アナウンスの明瞭度に干渉する通過列車の雑音)を除去または低減することは不可能である。そのような用途のシナリオの場合、対象信号を、その明瞭度が、現在存在する干渉雑音において最適になるまたは改善されるように、適応的に事前処理する方法が存在する(たとえば、[5]の方法)。そのような方法は、たとえば、対象信号のバンドパスフ

ィルタリング、周波数依存増幅、時間遅延、および/またはダイナミック圧縮を使用し、基本的には、背景雑音/雰囲気を(大幅に)修正すべきでないとき、オーディオビジュアル媒体にも適用可能になる。

【００１１】

別個のオーディオオブジェクトとして対象および背景雑音をエンコーディングすること:さらには、オーディオ信号をエンコーディングし送信する際に、対象信号に関する情報をパラメトリックにエンコーディングして、対象信号のエネルギーを受信機におけるデコーディング中に別個に調節可能にする方法が存在する。対象オブジェクト(たとえば、語音)のエネルギーを他のオーディオオブジェクト(たとえば、雰囲気)に対して相対的に高めると、結果的に語音明瞭度を改善することになり得る[11]。

【００１２】

ミックスド信号における語音信号の検出とレベル適応:その上、ミックスド信号における語音箇所を特定し、これらの箇所を、語音明瞭度の改善目的、たとえば、それらの音量を上げる目的により、調整する技術的システムが存在する。調整タイプに応じて、これは、それ以上の干渉雑音がミックスド信号に同時に存在しない場合しか、語音明瞭度を改善しない[12]。

【００１３】

語音を主として含まないチャンネルを下げること:1つのチャンネル(通常は、センター)が語音情報の大部分を含み、他のチャンネル(たとえば、左/右)が主に背景雑音を含むようなやり方でミックスされるマルチチャンネルオーディオ信号においては、1つの技術的解決策は、非語音チャンネルを固定利得だけ(たとえば、6dBだけ)減衰させ、そのやり方で信号対雑音比を改善すること(たとえば、サウンド検索システム(sound retrieval system、SRS)会話透明度(dialog clarity)またはサラウンドデコーダについてダウンミックスルールの適応)にある。

【００１４】

そのような諸方法においては、すでに非常に低く、実際には語音明瞭度に悪影響をまったく与えない背景雑音部分も減衰されることが起きる可能性がある。このことは、サウンドエンジニアが意図する雰囲気がもはや知覚できないので、サウンド美学全体の印象を低減させることになり得る。これを防止するために、米国特許第8,577,676 B2号には、非語音チャンネルが、語音明瞭度の測定基準が特定のしきい値に達するというその効果までのみ下げられるが、それ以上は下げない方法が記載されている。さらには、米国特許第8,577,676 B2号には、複数の周波数依存減衰が計算され、それぞれの減衰が、語音明瞭度の測定基準が特定のしきい値に達する効果を有する方法について開示されている。次いで、背景雑音のラウドネスを最大化するオプションは、複数のオプションから選択される。これは、このことにより、元のサウンドの性質ができるだけ最良に維持されるという仮定に基づいている。

【００１５】

それに基づいて、米国特許出願公開第2016/0071527 A1号には、非語音チャンネルが、一般仮定に反して、関連する語音情報も含み、そのため下げると明瞭度に悪影響が出る可能性があり得るとき、非語音チャンネルを下げない、またはそれほど下げない方法について記載されている。この文献にはまた、複数の周波数依存減衰が計算され、背景雑音のラウドネスを最大化するものが選択される方法が含まれている(やはり、これが元のサウンドの性質をできるだけ最良に維持するという仮定に基づいている)。

【００１６】

両方の米国特許文献には、それらの独立請求項において、本明細書に説明する本発明には必要でない非常に特異的な方法(たとえば、語音の発生の確率による低下係数をスケーリングすること)について記載されている。そのため、本発明は、米国特許第8,577,676 B2号および米国特許出願公開第2016/0071527 A1号に開示されている技術を使用することなく実現することができる。

【００１７】

米国特許第8,195,454 B2号には、音声アクティビティ検出(voice activity detection、VAD)を使用することによって、オーディオ信号において、語音が生じる部分を検出する方法について記載されている。次いで、1つまたはいくつかのパラメータがこれらの部分について修正され(たとえば、ダイナミックレンジ制御(dynamic range control)、ダイナミックイコライゼーション(dynamic equalization)、スペクトルシャープニング(spectral sharpening)、周波数移調(frequency transposition)、語音抽出、雑音低減、または他の語音増強アクション)、それにより、語音明瞭度の測定基準(たとえば、語音明瞭度指数(speech intelligibility index、SII)、[6])は、最大化されるか、または所望のしきい値よりも引き上げられるかのいずれかになる。この場合、聴覚損失またはリスナーの嗜好もしくは傾聴環境における雑音もまた、考慮され得る。

【００１８】

米国特許第8,271,276 B1号には、先行する時間セグメントに依存する増幅因子による語音セグメントのラウドネスまたはレベルの適応について記載されている。これは、本明細書に記載の本発明の中核に関係するものではなく、単に、本明細書に説明する本発明が、先行するセグメントに従って、語音として特定されるセグメントのラウドネスまたはレベルを単純に変えた場合のみ関係することになる。ソース分離、背景雑音の低下、スペクトル変動、ダイナミック圧縮など、語音セグメントの増幅を越えるオーディオ信号の適応は、含まれない。そのため、米国特許第8,271,276 B1号に開示される諸ステップもまた悪影響を与えるものではない。

【先行技術文献】

【特許文献】

【００１９】

【文献】米国特許第8,577,676号明細書

【文献】米国特許出願公開第2016/0071527号明細書

【文献】米国特許第8,195,454号明細書

【文献】米国特許第8,271,276号明細書

【発明の概要】

【発明が解決しようとする課題】

【００２０】

本発明の目的は、(語音)明瞭度とサウンドシーンの維持との間のトレードオフを改善することを可能にする概念を提供することである。

【課題を解決するための手段】

【００２１】

この目的は、独立請求項の主題によって解決される。

【００２２】

本発明の一実施形態は、対象部分(たとえば、語音部分)および副部分(たとえば、周囲雑音)を含む初期オーディオ信号を処理するための方法を提供する。この方法は、次の4つのステップ:

1.初期オーディオ信号を受信するステップ、

2.第1の信号調整器を使用することによって、受信した初期オーディオ信号を調整して、第1の調整されたオーディオ信号を取得し、第2の信号調整器を使用することによって、受信した初期オーディオ信号を調整して、第2の調整されたオーディオ信号を取得するステップ、

3.第1の調整されたオーディオ信号を評価基準に対して評価して、評価基準の満足度を表す第1の評価値を取得し、第2の調整されたオーディオ信号を評価基準に対して評価して、評価基準の満足度を表す第2の評価値を取得するステップ、

4.それぞれの第1のまたは第2の評価値によって決まる第1のまたは第2の調整されたオーディオ信号を選択するステップ

を含む。

【００２３】

諸実施形態によれば、評価基準は、知覚的類似度、語音明瞭度、ラウドネス、サウンドパターン、および空間度を含むグループのうちの1つまたは複数とすることができる。選択するステップが、諸実施形態によれば、独立した評価基準を表す複数の独立した第1のおよび第2の評価値に基づいて行われ得ることに留意されたい。評価基準および特に選択するステップは、いわゆる最適化対象によって決まってもよい。したがって、この方法は、諸実施形態によれば、個人の嗜好を定義する最適化対象に関する情報を受信するステップを含み、評価基準は、最適化対象によって決まり、または調整するステップおよび/もしくは評価するステップおよび/もしくは選択するステップは、最適化対象によって決まり、または選択するステップについて独立した評価基準を表す独立した第1のおよび第2の評価値の重み付けは、最適化対象によって決まる。

【００２４】

たとえば、最適化対象が、2つの要素、たとえば、最適な語音明瞭度と、初期オーディオ信号と調整されたオーディオ信号との間の許容可能な知覚的類似度との組合せである場合、選択のための重み付けを行ってもよい。たとえば、これらの2つの基準、すなわち、語音明瞭度および知覚的類似度は、別個に評価されてもよく、それにより、評価基準についてのそれぞれの評価値が決定され、次いで、重み付けされた評価値に基づいて選択が行われる。重み付けは、最適化対象によって決まり、最適化対象は、逆に、個人の嗜好によって設定され得る。

【００２５】

諸実施形態によれば、適応させるステップ、評価するステップ、および選択するステップは、ニューロニューラルネットワーク/人口知能を使用することによって行ってもよい。

【００２６】

好ましい実施形態によれば、語音明瞭度は、2つ以上の使用される調整器によって十分な形で改善すると仮定される。別の観点から表すと、このことは、語音明瞭度の十分に高い改善を可能にする、または語音の明瞭度が十分である信号を出力する調整器のみが考慮に入れられることを意味する。次のステップにおいては、異なって調整された信号間の選択が行われる。この選択については、知覚的類似度が評価基準として使用され、それにより、ステップ3および4(上記の方法参照)を次のように行うことが可能になる:

3.受信した初期オーディオ信号を第1の調整されたオーディオ信号と比較して、初期オーディオ信号と第1の調整されたオーディオ信号との間の知覚的類似度を表す第1の知覚的類似度値を取得し、受信した初期オーディオ信号を第2の調整されたオーディオ信号と比較して、初期オーディオ信号と第2の調整されたオーディオ信号との間の知覚的類似度を表す第2の知覚的類似度値を取得するサブステップ、ならびに

4.それぞれの第1のまたは第2の知覚的類似度値よって決まる第1のまたは第2の調整されたオーディオ信号を選択するサブステップ。

【００２７】

本発明の一実施形態によれば、第1の知覚的類似度値が第2の知覚的類似度値よりも高いとき(高い第1の知覚的類似度値は、第1の調整されたオーディオ信号のより高い知覚的類似度を示す)、第1の調整されたオーディオ信号が選択され、逆に、第2の知覚的類似度値が第1の知覚的類似度値よりも高いとき(高い第2の知覚的類似度値は、第2の調整されたオーディオ信号のより高い知覚的類似度を示す)、第2の調整されたオーディオ信号が選択される。さらなる実施形態によれば、知覚的類似度の代わりに、ラウドネス値のような別の値を使用してもよい。

【００２８】

比較するステップ3、および知覚的類似度値に基づいて選択するステップ4を有するこの適応方法は、さらなる実施形態によれば、ステップ2の後、ステップ3の前に、別の最適化基準に対して、たとえば、音声明瞭度に対して、第1のおよび第2の調整された信号を評価する追加のステップによって増強することができる。上述したように、この場合、たとえば語音明瞭度が低すぎるとき、この第1の評価基準は(十分に)満たされていないので、いくつかの調整された信号が考慮に入れられないことがあり得る。あるいは、すべての評価基

準を、重み付けなしまたは重み付けありを選択するステップ中に、考慮に入れることが可能である可能性もある。この重み付けは、ユーザによって選択され得る。諸実施形態によれば、この方法は、選択によって決まる第1のまたは第2の調整されたオーディオ信号を出力するステップをさらに含む。

【００２９】

本発明の一実施形態が、一方法を提供し、ここでは、対象部分が初期オーディオ信号の語音部分であり、副部分がオーディオ信号の周囲雑音部分である。

【００３０】

本発明の諸実施形態は、異なる語音明瞭度オプションが、複数の影響因子よって決まる、たとえば、入力オーディオストリームまたは入力オーディオシーンによって決まるその改善効果に関して異なることを定義することに基づいている。最適な語音明瞭度アルゴリズムもまた、1つのオーディオストリーム内でシーンごとに異なることがある。そのため、本発明の諸実施形態は、オーディオ信号の異なる調整、特に、初期オーディオ信号と調整されたオーディオ信号との間の知覚的類似度に関して分析して、最も高い知覚的類似度を有する調整器/調整されたオーディオ信号を選択する。このシステム/概念により、初めて、サウンド全体が、必要なだけ、ただしできるだけ知覚的に変えずに、両方の要件を満たすこと、すなわち、初期信号の語音明瞭度を改善し(または傾聴努力を低減させ)、同時にサウンド美学構成要素にできるだけ影響を与えないようにすることが可能になる。このサウンド美学の維持は、自動化された方法においてはこれまでは考慮されてこなかったユーザの受諾の重要な構成要素を表しているので、これは、境界条件としてのみ明瞭度を改善するのにこれまでは使用されてきた非自動方法と比較して努力および費用がかなり低減されること、ならびにそれらの方法に対する付加価値が大きいことを表している。

【００３１】

一実施形態によれば、初期オーディオ信号を出力するステップは、それぞれの第1のまたは第2の知覚的類似度値がしきい値を下回るとき、第1のまたは第2の調整されたオーディオ信号を出力する代わりに行われる。「下回る(below)」は、調整された信号は、初期オーディオ信号との類似が十分でないことを示す。これは、システムが、語音明瞭度または傾聴努力についてのサウンドミックスの自動試験のいずれも可能にし、同時に、サウンド全体が効果的な形で知覚的に変えられることを保証するので、有利である。

【００３２】

本発明の一実施形態が、一方法を提供し、ここでは、比較するステップは、PEAQモデル[8]、POLQAモデル[9]、および/またはPEMO-Qモデル[10]のような(知覚)モデルを使用することによって、第1のおよび/または第2の知覚的類似度値を抽出することを含む。PEAQ、POLQA、PEMO-Qが、2つのオーディオ信号の知覚的類似度を出力するように訓練された特定のモデルであることに留意されたい。諸実施形態によれば、処理の程度は、さらなるモデルによって制御される。

【００３３】

一実施形態によれば、第1のおよび/または第2の知覚的類似度値は、第1のもしくは第2の調整されたオーディオ信号の物理パラメータ、第1のもしくは第2の調整されたオーディオ信号の音量レベル、第1のもしくは第2の調整されたオーディオ信号についての心理音響パラメータ、第1のもしくは第2の調整されたオーディオ信号のラウドネス情報、第1のもしくは第2の調整されたオーディオ信号のピッチ情報、ならびに/または第1のもしくは第2の調整されたオーディオ信号の知覚されたソース幅情報によって決まることに留意されたい。

【００３４】

本発明の一実施形態が、一方法を提供し、ここでは、第1のおよび/もしくは第2の信号調整器は、(たとえば、初期オーディオ信号についての)SNR増加、(たとえば、初期オーディオ信号の)ダイナミック圧縮を行うように構成され、ならびに/または調整するステップは、初期オーディオ信号が、別個の対象部分および別個の副部分を含む場合、対象部分を増加させること、対象部分についての周波数重み付けを増加させること、対象部分をダイ

ナミックに圧縮すること、副部分を減少させること、対象部分についての周波数重み付けを減少させることを含み、あるいは、調整するステップは、初期オーディオ信号が、組み合わさった対象部分と副部分とを含む場合、対象部分および副部分の分離を行うことを含む。概して、これは、本発明の一実施形態が、一方法を提供することを意味し、ここで、第1のおよび/または第2の調整されたオーディオ信号は、前景に移動した対象部分、および背景に移動した副部分、ならびに/または前景に移動した対象部分として語音部分、および背景に移動した副部分として周囲雑音を含む。

【００３５】

一実施形態によれば、選択するステップは、聴覚障害のある人の難聴のグレード、個人の聴覚性能;個人の周波数依存聴覚性能;個人の嗜好、および/または信号調整率に関する個人の嗜好のような1つまたは複数のさらなる因子を考慮に入れて行われる。同様に、諸実施形態によれば、調整および/または比較するステップは、聴覚障害のある人の難聴のグレード、個人の聴覚性能;個人の周波数依存聴覚性能;個人の嗜好、および/または信号調整率に関する個人の嗜好のような1つまたは複数の因子を考慮に入れて行われる。したがって、選択するステップ、調整するステップ、および/または比較するステップは、やはり、個人の聴覚または個人の嗜好を考慮することができる。

【００３６】

諸実施形態によれば、処理を制御するためのモデルは、たとえば、聴覚損失または個人の嗜好に関して構成され得る。

【００３７】

一実施形態によれば、比較するステップは、初期オーディオ信号の全体、および第1のおよび第2の調整されたオーディオ信号の全体について、または第1のおよび第2の調整されたオーディオ信号のそれぞれの対象部分と比較される個々のオーディオ信号の対象部分について、または第1のおよび第2の調整されたオーディオ部分の副部分と比較される初期オーディオ信号の副部分について行われる。

【００３８】

本発明の一実施形態は、一方法を提供し、ここで、この方法は、語音部分を決定するために初期オーディオ部分を分析する初期ステップと、初期オーディオ信号の語音明瞭度について評価するために語音部分と周囲雑音部分とを比較する初期ステップと、語音明瞭度について示す値がしきい値を下回る場合、調整するステップのための第1のおよび/または第2の信号調整器をアクティブ化する初期ステップとをさらに含む。したがって、処理が、語音が発生する箇所でのみ行われることは有利である。この場合、調整されたサウンドミックスは、この語音部分について生成され、サウンドミックスは、特定の知覚的測定基準を満たす、または最大化することを目的としている。

【００３９】

本発明の一実施形態は、一方法を提供し、ここで、初期オーディオ信号は、複数の時間フレームまたはシーンを含み、基本ステップが、各時間フレームまたはシーンについて繰り返される。

【００４０】

諸実施形態によれば、第1の時間フレームが、第1の調整器を使用して適応し、第2の時間フレームについては、別の調整器が選択されることがあり得る。知覚的連続性を保証するために、時間フレーム間の遷移、または2つの時間フレームの適応部分が、挿入可能である。たとえば、第1の時間フレームの終わりおよび後続の時間フレームの始まりは、その適応性能に関して適応する。たとえば、2つの適応方法間のある種の補間が適用され得る。さらなる実施形態によれば、知覚的連続性を可能にするために、すべてのまたは複数の後続の時間フレームについて、同じ調整器が使用されることがあり得る。さらなる実施形態によれば、時間フレームの適応は、たとえば明瞭度性能の観点から、必要とされる適応がない場合であっても行われることもあり得る。しかしながら、これにより、それぞれの時間フレーム間の知覚的類似度を保証することが可能になる。

【００４１】

本発明の一実施形態は、上記の方法に従って、コンピュータにおいて動作するとき、実行するためのプログラムコードを有するコンピュータプログラムを提供する。

【００４２】

本発明の別の実施形態は、初期オーディオ信号を処理するための装置を提供する。この装置は、初期オーディオ信号を受信するためのインターフェースと、それぞれの調整されたオーディオ信号を取得するように初期オーディオ信号を処理するためのそれぞれの調整器と、それぞれの調整されたオーディオ信号の評価を行うための評価器と、それぞれの第1のまたは第2の評価値によって決まる第1のまたは第2の調整されたオーディオ信号を選択するための選択器とを備える。

【００４３】

さらなる詳細については、従属請求項の主題によって定義される。以下に、本発明の諸実施形態について、添付の図を参照しながら、詳細に論じる。

【図面の簡単な説明】

【００４４】

【図１】基本の実施形態に従って、オーディオ信号を、オーディオ信号の語音部分のような対象部分の再生品質を改善するように処理するための方法シーケンスを概略的に示す図である。

【図２】増強された実施形態を示す概略的フローチャートである。

【図３】一実施形態に従って、オーディオ信号を処理するためのデコーダの概略的ブロック図である。

【発明を実施するための形態】

【００４５】

続いて、以下に、本発明の諸実施形態について、添付の図を参照して論じ、同一のまたは類似の機能を有する物には、同一の符号が与えられている。

【００４６】

図1は、3つのステップ/ステップグループ110、120、および130を含む方法100を示す概略的フローチャートを示している。方法100は、初期オーディオ信号ASの処理を可能にする目的を有し、調整されたオーディオ信号MOD ASを出力する結果を有することができる。出力されたオーディオ信号MOD ASの可能な結果は、オーディオ信号ASの処理が必要でないことがあり得るので、仮定法が使用される。その場合、オーディオ信号と調整されたオーディオ信号とは同じである。

【００４７】

3つの基本ステップ110および120は、ここでは、ソナーステップ110a、110bなど、および120aが、並行して、または互いに順次行われるので、ステップグループと解釈される。

【００４８】

ステップ110のグループ内で、オーディオ信号ASは、異なる調整器/処理手法を使用することによって、別個に処理される。ここでは、符号110a、110bによってマーク付けされている第1のおよび第2の調整器を適用する2つの例示的なステップが示されている。両方のステップは、並行して、または互いに順次行われ、オーディオ信号ASの処理を行うことができる。オーディオ信号は、たとえば、1つのオーディオトラックを含むオーディオ信号であってよく、このオーディオトラックは、2つの信号部分を含む。たとえば、オーディオトラックは、語音信号部分(対象部分)と、周囲雑音信号部分(副部分)とを含み得る。これらの2つの部分は、符号AS_TPおよびAS_SPによってマーク付けされている。この実施形態においては、AS_TPは、この信号部分AS_TPを増幅させて語音明瞭度を高めるために、オーディオ信号ASから抽出すべき、またはオーディオ信号AS内で特定すべきであることが仮定される。この処理は、複数のオーディオトラック、たとえば、AS_SPについて1つのオーディオトラック、およびAS_TPについて1つのオーディオトラックを含むオーディオ信号ASを分離することなく、2つの部分AS_SPおよびAS_TPを含んだたった1つのオーディオトラックを有するオーディオ信号について行うことができる。

【００４９】
上述したように、たとえば、AS_TP部分を増幅させることによって、またはAS_SP部分を減少させることによって、語音明瞭度を改善することを完全に可能にするオーディオ信号ASの複数の可能な調整が存在する。さらなる例は、非語音チャンネルを下げること、ダイナミックレンジ制御、ダイナミックイコライゼーション、スペクトルシャープニング、周波数移調、語音抽出、雑音低減、または従来技術の文脈で論じた他の語音増強アクションである。これらの調整の効率性は、複数の因子によって決まり、たとえば、レコーディング自体、ASの形式(たとえば、1つだけのオーディオトラックを有する形式、または複数のオーディオトラックを有する形式)によって決まり、または複数の他の因子によって決まる。最適な語音明瞭度を可能にするために、少なくとも2つの信号調整が信号ASに適用される。第1のステップ110a内で、受信した初期オーディオ信号ASは、第1の調整器を使用することによって調整されて、第1の調整されたオーディオ信号1st MOD ASが取得される。ステップ110aとは関係なく、受信した初期オーディオ信号ASの第2の調整は、第2の調整器を使用することによって行われて、第2の調整されたオーディオ信号2nd MOD ASが取得される。たとえば、第1の調整器は、ダイナミックレンジ制御に基づいていてよく、第2の調整器は、スペクトルシャープニングに基づいていてもよい。もちろん、たとえば、ダイナミックイコライゼーション、周波数再移調、語音抽出、雑音低減、もしくは語音増強アクション、またはそのような調整器の組合せに基づいて、他の調整器が、第1のおよび/もしくは第2の調整器の代わりに、または第3の調整器(図示せず)として使用されてもよい。すべての手法が、異なる結果的に生じた調整されたオーディオ信号1st MOD ASおよび2nd MOD ASをもたらすことができ、それらは、語音明瞭度に関して、および初期オーディオ信号ASとの類似度に関して異なっていてもよい。これらの2つのパラメータ、またはこれらの2つのパラメータのうちの少なくとも一方が、次のステップ120内で評価される。

【００５０】
詳細には、ステップ120a内で、第1の調整されたオーディオ信号1st MOD ASは、類似度を見出すために、元のオーディオ信号ASと比較される。同様に、ステップ120b内で、第2の調整されたオーディオ信号2nd MOD ASは、初期オーディオ信号ASと比較される。比較については、ステップ120を行うエンティティは、オーディオ信号ASを直接受信し、第1の/第2のMOD ASを受信する。この比較の結果は、それぞれ第1の知覚的類似度値および第2の知覚的類似度値である。2つの値は、符号1st PSVおよび2nd PSVによってマーク付けされている。両方の値は、それぞれの第1の/第2の調整されたオーディオ信号1st MOD AS、2nd MOD ASと、初期オーディオ信号ASとの間の知覚的類似度を表している。語音明瞭度の改善が十分であるという仮定の下、より高い類似度を示す第1の/第2のPSVを有する第1のまたは第2の調整されたオーディオ信号が選択される。これは、選択するステップ130によって行われる。

【００５１】
選択の結果は、諸実施形態によれば、出力され/転送され得、それにより、方法100は、元の信号との類似度が最も高いそれぞれの調整されたオーディオ信号1st MOD ASまたは2nd MOD ASを出力することが可能になる。わかり得るように、調整されたオーディオ信号MOD ASは、なおも、2つの部分AS_SP'およびAS_TP'を含む。AS_SP'およびAS_TP'内で「'」によって示されているように、2つの部分AS_SP'およびAS_TP'の両方、または少なくとも一方が調整される。たとえば、AS_TP'についての増幅が高められてもよい。

【００５２】
さらなる実施形態によれば、ステップ120内で、増強された評価が行われることがあり得る。この場合、次いで、第1のまたは第2の調整器によって行われる調整(ステップ110aおよび110b参照)が十分で語音明瞭度を改善するかどうかがさらに証明される。たとえば、AS_TP'とAS_SP'との比が、AS_TPとAS_SPとの比よりも大きいかどうかが分析され得る。

【００５３】

上記の実施形態は、この方法100の目的が、MOD ASが改善された語音明瞭度を有することという仮定から始まっている。さらなる実施形態によれば、調整の目的は、異なっていてもよい。たとえば、AS_TPという部分が、調整された信号MOD AS全体内で強調すべき、別の部分、概して対象部分であってもよい。これは、AS_TP'を強調/増幅させることによって、および/またはAS_SP'を調整することによって行うことができる。

【００５４】

また、図1の上記の実施形態は、知覚的類似度の文脈で論じている。この手法が、他の評価基準について、より一般的に使用され得ることに留意されたい。図1は、評価基準が知覚的類似度であるという仮定から始まっている。しかしながら、さらなる実施形態によれば、やはり別の評価基準も、代わりに追加的に使用され得る。たとえば、語音明瞭度が、評価基準として使用され得る。そのような場合においては、第1の調整されたオーディオ信号1st MOD ASの評価が、ステップ120aの代わりに行われ、ステップ120bにおいては、第2の調整されたオーディオ信号2nd MOD ASの評価が行われる。評価するこれらの2つのステップ120aおよび120bの結果は、それぞれの第1のおよび第2の評価値である。その後、ステップ130がそれぞれの評価値に基づいて行われる。

【００５５】

さらなる評価基準は、ラウドネスまたは可聴空間広大性(auditory spaciousness)などとすることができる。

【００５６】

図2を参照しながら、機能を増強したさらなる実施形態について後述する。

【００５７】

図2は、AS_TP(語音S)およびAS_SP(周囲雑音N)という2つの部分を含むオーディオ信号ASを処理することを可能にする概略的フローチャートを示している。ここでは、信号調整器11は、信号ASを処理するのに使用され、それにより、選択エンティティ13は、調整された信号モードASを出力することができる。この実施形態においては、調整器は、異なる調整1、2、…、Mを行う。これらの調整は、複数の異なるモデルに基づいて、3つの調整された信号1st MOD AS、2nd MOD AS、およびM MOD ASを生成する。各信号1st MOD AS、2nd MOD AS、およびM MOD ASについては、S1'、N1'、S2'、N2'、およびSN'、NNM'という2つの部分が例示されている。1st MOD AS、2nd MOD AS、およびM MOD ASの出力信号は、評価器12によって、初期信号ASに対するその視点の類似度に関して評価される。したがって、1つまたは複数の評価器段12は、信号AS、ならびにそれぞれの調整された信号1st MOD AS、2nd MOD AS、およびM MOD ASを受信する。この評価器12の出力は、それぞれの類似度情報とともに、それぞれの調整信号1st MOD AS、2nd MOD AS、およびM MOD ASである。この類似度情報に基づいて、位置段13は、出力すべき調整された信号MOD ASを決定する。

【００５８】

諸実施形態によれば、信号ASは、語音が存在するか否かを判定するように分析器21によって分析されてもよい。この決定ステップには、初期オーディオ信号AS内に語音が存在しない、または調整すべき信号が存在しない場合に、21がマーク付けされる。初期/元のオーディオ信号ASは、信号として、すなわち、調整なしで(N-MOD AS参照)使用される。

【００５９】

語音が存在する場合には、第2の分析器22が、語音明瞭度を改善する必要があるかどうかを分析する。この決定点には、符号22sがマーク付けされている。調整が必要でない場合には、元の信号ASは、出力すべき信号として使用される(N-MOD AS参照)。調整が推奨される場合には、信号調整器11がイネーブルされる。

【００６０】

この構造に基づいて、オーディオおよびオーディオビジュアル媒体における語音明瞭度の改善が可能である。ここでは、処理すべきサウンドミックスは、完成したミックスであるか、または別個のオーディオトラックもしくはサウンドオブジェクト(たとえば、会話、音楽、残響、効果)から構成されているかのいずれかであってよい。第1のステップにおい

ては、信号は、語音の存在に関して分析される(符号21、21s参照)。語音アクティブ箇所は、物理パラメータまたは心理音響パラメータに関して、たとえば語音明瞭度(SIIなど)または傾聴努力の計算された値の形態で、たとえば[7]に提示されているミックスド信号のための手法に基づいて、さらに分析されることになる(符号22、22s参照)。この評価に基づいて、パラメータを対象またはしきい値と比較することによって、語音明瞭度が十分であるかどうか、またはサウンド適応が必要かどうかの決定が行われる。適応が必要でない場合、サウンドミキシングが、通常通り行われる、または元のミックスASが維持される。適応が必要である場合、所望の明瞭度が得られるように、オーディオトラックまたは異なるオーディオトラックを調整するアルゴリズムが適用されることになる。ここまでは、この方法は、米国特許第8,195,454 B2号および米国特許第8,271,276 B1号に開示されている手法に類似しているが、それぞれの請求項1に記載されている詳細に限定するものではない。

【００６１】

このことは、諸実施形態によれば、たとえば米国特許第8,577,676 B2号および米国特許出願公開第2016/0071527 A1号に記載されている非語音チャンネルのラウドネスの最大化を越えるサウンド低減方法のモデルベースの選択13が、この概念により行われることを意味する。選択については、さらなるモデル段12が適用され、このモデル段12は、元のミックスASと、物理パラメータおよび/または心理音響パラメータに基づいて異なる方式で修正されるミックス(1st MOD AS、2nd MOD AS、M MOD AS)との間の知覚的類似度をシミュレートする。ここでは、元のミックスAS、ならびに異なるタイプの修正されたミックス1st MOD AS、2nd MOD AS、M MOD ASは、さらなるモデル段12への入力として機能する。

【００６２】

サウンドシーンをできるだけ最良に維持するという目標を得るために、知覚的に最も目立たない信号調整により所望の明瞭度を得るサウンド適応のためのその方法が選択され得る(符号13参照)。

【００６３】

諸実施形態によれば、計器的な形で知覚的類似度を測定することができ、本明細書に使用され得る可能なモデルは、たとえば、PEAQ [8]、POLQA [9]、またはPemo-Q [10]である。また、または追加的に、さらなる物理的測定基準(たとえば、レベル)、もしくは心理音響的測定基準(たとえば、ラウドネス、ピッチ、知覚されたソース幅)が、知覚的類似度を評価するのに使用されてもよい。

【００６４】

典型的には、オーディオストリームは、時間領域に沿って構成されている異なるシーンを含む。そのため、諸実施形態によれば、異なるサウンド適応が、最小限の侵入的な知覚的効果を有するために、オーディオトラックASにおいて異なる時間で行われることがあり得る。たとえば、語音AS_TPおよび背景雑音AS_TPが、すでに、明らかに異なるスペクトルを有する場合、単純なSNR適応は、背景雑音の真正性を最良の可能な効果まで維持するので、最良の解決策とすることができる。さらなる話者が対象語音を重ね合わせる場合、最適化対象を満たすための他の方法(たとえば、ダイナミック圧縮)が、より良いこともあり得る。

【００６５】

さらなる実施形態によれば、このモデルベースの選択は、計算においてオーディオ素材の将来のリスナーの起こり得る聴覚障害を、たとえばオーディオグラム、個人のラウドネス関数の形態で、または個人のサウンド嗜好を入力する形態で考慮することができる。それによって、語音明瞭度は、正常の聴覚能力をもつ人たちだけでなく、特定の形態の聴覚障害(たとえば、加齢に関係する聴覚損失)をもつ人たちについても保証され、また元のバージョンと処理されたバージョンとの間の知覚的類似度が個々に異なる可能性があることも考慮する。

【００６６】

語音明瞭度および知覚的類似度の、モデルによる分析、ならびにそれぞれの信号処理は、サウンドミックス全体について行っても、もしくはミックスの一部(個々のシーン、個々の会話)についてのみ行ってもよく、またはミックス全体に沿って短い時間窓で行ってもよく、それにより、サウンド適応が行われなくてはならないかどうかの決定が各窓について行うことが可能になる。

【００６７】

以下に、そのような処理の例について、例示的に論じる。

i.サウンド適応なし: 傾聴モデルの分析により、十分に高い語音明瞭度が保証されていることが示されている場合、さらなるサウンド適応は行わない。あるいは、異なるシーン間の知覚的差異を回避するために、以下の適応が行われる。また、処理なしと、以下の選択された処理との間の「補間」が行われることもある。両方のモデルにより、異なる時間フレーム/シーンにわたって知覚的連続性が可能になる。

会話および背景雑音の別個のオーディオトラックについては、次のステップが可能である。

ii.サウンド信号を適応させるステップ: 語音信号のオーディオトラックのみが、たとえばレベルを上げることによって、周波数重み付けおよび/または単一チャンネルもしくはマルチチャンネルのダイナミック圧縮によって、語音明瞭度を改善するように処理される。

iii.干渉雑音を適応させるステップ: 語音を含まないオーディオトラックのうちの1つまたはいくつかは、たとえば、レベルを下げることによって、周波数重み付けおよび/または単一チャンネルもしくはマルチチャンネルのダイナミック圧縮によって、語音明瞭度を改善するように処理される。しかしながら、背景雑音を完全になくすことが、結果的に語音明瞭度の改善をもたらすことになることが自明な場合は、音楽、効果などの設計が創造的なサウンド設計の必須の一部でもあるので、サウンド美学という理由から実用的ではない。

iv.すべてのオーディオトラックを適応させるステップ: 語音信号のオーディオトラックと他のオーディオトラックのうちの1つまたはいくつかとはともに、語音明瞭度を改善するための上記に記載の方法によって処理される。

【００６８】

適応には、たとえばニューロンネットワークを使用した人工知能が使用され得ることに留意されたい。すでにミックスされたオーディオ信号においては(すなわち、会話および背景雑音のための非別個のオーディオトラック)においては、ステップii～ivは、たとえば、ミックスを語音と1つまたはいくつかの背景雑音とに分離するソース分離方法が事前に使用されるときにも行うことができることに留意されたい。その場合、語音明瞭度の改善は、たとえば、改善されたSNRにおいて別個の信号をリミックスすること、あるいは語音信号および/または背景雑音もしくは背景雑音の一部を、周波数重み付けまたは単一チャンネルもしくはマルチチャンネルのダイナミック圧縮によって、調整することにある可能性もあり得る。ここでは、やはり、必要に応じて語音明瞭度を改善することと、同時に元のサウンドをできるだけ最良に維持することとの両方を行うサウンド適応が選択されることになる。ソース分離のための方法は、語音アクティビティを検出するための明示的な段なしに適用されることがあり得る。

【００６９】

諸実施形態によれば、それぞれの処理の選択は、人工知能/ニューロンネットワークを使用することによって行ってもよいことに留意されたい。この人工知能/ニューロンネットワークは、たとえば、選択のための複数の因子、たとえば知覚値およびラウドネス値、または個人の傾聴嗜好に対する一致を表す値が存在する場合に、使用され得る。

【００７０】

上記では、異なる時間フレーム/シーンにわたって知覚的連続性を維持するために、必要でない場合であっても、シーンの適応を行うことが可能であることを論じてきた。別の変形形態によれば、複数のまたはすべてのシーンについての適応を選択することが可能である。さらには、異なるシーン間で、異なって適応した、もしくは適応したシーンと、適応されていないシーンとの間のある種の遷移が、知覚的連続性を維持するために統合可能で

あることに留意すべきである。

【００７１】

諸実施形態によれば、知覚的類似度(符号12参照)に基づいた評価および最適化は、対象言語、背景雑音、または語音と背景雑音とのミックスに関係し得る。たとえば、処理された語音信号、処理された背景雑音、またはそれぞれの元の信号に対する処理されたミックスの知覚的類似度について、それぞれの信号についての特定の信号調整度を上回らないようにすることができる、異なるしきい値が存在することもあり得る。さらなる境界条件は、(音楽などの)背景雑音が、先行時点または後続時点に対して、それほど知覚的に変化しないことであり得、そうでなければ、たとえば、語音が存在した瞬間に、音楽が下げられ過ぎる、もしくはその周波数成分に変化が生じる、または俳優の語音が映画の流れの中でそれほど変化しないことがある場合、知覚の連続性が攪乱されることになる。そのような境界条件は、上記に記載のモデルに基づいて検討されることもあり得る。

【００７２】

このことは、所望の明瞭度改善が、語音および/または背景雑音の知覚的類似度にそれほど干渉することなく、得られない可能性があるという効果を有することもある。ここでは、(可能性として構成可能な)決定段が、どの対象を得るべきか、またはトレードオフを見つけるべきかおよびどのように見つけるべきかを決定することが可能になる。

【００７３】

ここでは、所望の語音明瞭度および元の物に対する知覚的類似度が得られたことを検証するために、処理は、繰り返し行うことができ、すなわち、傾聴モデルの検討は、サウンド適応の後に、再度、行うことができる。

【００７４】

処理は、(傾聴モデルの計算に応じて)、オーディオ素材の継続時間全体について行っても、またはその一部(たとえば、シーン、会話)についてのみ行ってもよい。

【００７５】

諸実施形態は、すべてのオーディオ媒体およびオーディオビジュアル媒体(概して、映画、ラジオ、ポッドキャスト、オーディオレンダリング)について使用可能である。可能な商用用途は、たとえば、

i.顧客が自分のオーディオ素材をロードし、自動化された語音明瞭度改善をアクティブ化し、処理された信号をダウンロードするインターネットベースのサービス。これは、サウンド適応方法およびサウンド適応の程度の顧客固有の選択によって拡張させることができる。そのようなサービスは、すでに存在しているが、語音明瞭度に関するサウンド適応のための傾聴モデルは使用されていない(上記2.(V.)下参照)。

ii.ファイルされたまたは現在制作中のサウンドミックスの補正を可能にする、たとえばデジタルオーディオワークステーション(digital audio workstations、DAW)に統合された、サウンド制作のためのツールについてのソフトウェアソリューション。

iii.オーディオ素材における、所望の語音明瞭度に対応していない箇所を特定し、可能性として、推奨のサウンド適応調整を選択するためにユーザに提供するテストアルゴリズム。

iv.たとえば、サウンドバー、ヘッドフォン、テレビジョンデバイス、またはストリーミングされたオーディオコンテンツを受信するデバイスなど、放送チェーンのリスナーの側のエンドデバイスに統合されたソフトウェアおよび/またはハードウェア。

【００７６】

図1の文脈で論じた方法、または、図2の文脈で論じた概念は、プロセッサを使用することによって実装され得る。このプロセッサは、図3によって示されている。

【００７７】

図3は、2つの段、信号調整器11と評価器/選択器12および13との中のプロセッサ10を示している。調整器は、インターフェースからオーディオ信号を受信し、異なるモデルに基づいて調整を行って、調整されたオーディオ信号MOD ASを取得する。評価器/選択器12は、インターフェースからオーディオ信号を受信し、異なるモデルに基づいて調整を行って、調整されたオーディオ信号MOD ASを取得する。評価器/選択器12、13は、類似度

を評価し、この情報に基づいて最も高い類似度または高い類似度、およびMOD ASを出力するために十分である改善された語音明瞭度を有する信号を選択する。

【００７８】

もちろん、これらの2つの段11、12および13は、1つのプロセッサによって実装され得る。

【００７９】

いくつかの態様について、装置の文脈で説明しているが、これらの態様はまた、対応する方法の説明も表していることは明らかであり、ブロックまたはデバイスが、方法ステップまたは方法ステップの特徴に対応する。同様に、方法ステップの文脈で説明する態様もまた、対応するブロックもしくは品目、または対応する装置の特徴の説明を表している。方法ステップのうちのいくつかまたはすべてが、たとえば、マイクロプロセッサ、プログラマブルコンピュータ、または電子回路のようなハードウェア装置によって(またはハードウェア装置を使用して)実行され得る。いくつかの実施形態においては、最も重要な方法ステップのうちのいくつかまたは1つもしくは複数がそのような装置によって実行されてもよい。

【００８０】

本発明のエンコーディングされたオーディオ信号は、デジタルストレージ媒体において記憶され得、またはワイヤレス伝送媒体もしくはインターネットなどのワイヤードの伝送媒体など、伝送媒体において伝送され得る。

【００８１】

特定の実装要件に応じて、本発明の諸実施形態は、ハードウェアにおいて、またはソフトウェアにおいて実装され得る。実装形態は、デジタルストレージ媒体、たとえば、電子的に可読な制御信号を記憶したフロッピーディスク、DVD、Blu-Ray、CD、ROM、PROM、EPROM、EEPROM、またはFLASH(登録商標)メモリを使用して行うことができ、それらは、プログラマブルコンピュータシステムと協働して(または協働することができて)、それぞれの方法が実行されることになる。そのため、デジタルストレージ媒体は、コンピュータ可読とすることができる。

【００８２】

本発明によるいくつかの実施形態は、プログラマブルコンピュータシステムと協働することができる電子的に可読な制御信号を有するデータキャリアを含み、それにより、本明細書に説明した方法うちの1つが行われることになる。

【００８３】

概して、本発明の実施形態は、プログラムコードを含むコンピュータプログラム製品として実装可能であり、プログラムコードは、コンピュータプログラム製品がコンピュータにおいて動作するとき、方法のうちの1つを行うように動作する。プログラムコードは、たとえば、機械可読キャリアにおいて記憶され得る。

【００８４】

他の実施形態は、機械可読キャリアに記憶されている、本明細書に説明した方法のうちの1つを行うためのコンピュータプログラムを含む。

【００８５】

言い換えれば、そのため、本発明の一実施形態は、コンピュータにおいて動作するとき、本明細書において説明した方法のうちの1つを行うためのプログラムコードを有するコンピュータプログラムである。

【００８６】

そのため、本発明の方法のさらなる実施形態は、本明細書に説明する方法のうちの1つを行うためのコンピュータプログラムを記録したデータキャリア(またはデジタルストレージ媒体、またはコンピュータ可読媒体)である。データキャリア、デジタルストレージ媒体、または記録された媒体は、典型的には、有形および/または非一時的である。

【００８７】

そのため、本発明の方法のさらなる実施形態は、本明細書に説明する方法のうちの1つ

を行うためのコンピュータプログラムを表すデータストリームまたは信号シーケンスである。データストリームまたは信号シーケンスは、たとえば、データ通信接続を介して、たとえば、インターネットを介して、転送されるように構成されていてもよい。

【００８８】

さらなる実施形態は、本明細書に説明する方法のうちの1つを行うように構成されている、または適合されている、処理手段、たとえば、コンピュータ、またはプログラマブル論理デバイスを含む。

【００８９】

さらなる実施形態は、本明細書に説明する方法のうちの1つを行うためのコンピュータプログラムをインストールしたコンピュータを含む。

【００９０】

本発明によるさらなる実施形態は、本明細書に説明する方法のうちの1つを行うためのコンピュータプログラムを受信機に(たとえば、電子的に、または光学的に)転送するように構成されている装置またはシステムを含む。受信機は、たとえば、コンピュータ、モバイルデバイス、またはメモリデバイスなどとすることができる。装置またはシステムは、たとえば、コンピュータプログラムを受信機に転送するためのファイルサーバを含むことができる。

【００９１】

いくつかの実施形態においては、プログラマブル論理デバイス(たとえば、フィールドプログラマブルゲートアレイ)を使用して、本明細書に説明する方法の機能のうちのいくつかまたはすべてを行ってもよい。いくつかの実施形態においては、フィールドプログラマブルゲートアレイは、マイクロプロセッサと協働して、本明細書に説明する方法のうちの1つを行ってもよい。概して、方法は、好ましくは、任意のハードウェア装置によって行われる。

【００９２】

上述した実施形態は、単に本発明の原理について説明しているにすぎない。本明細書に説明した装置および詳細の修正形態および変形形態が、当業者にとっては明らかになることを理解されたい。そのため、添付の特許請求の範囲の範囲によってのみ限定がなされ、本明細書における実施形態の記述および説明によって提示される具体的な詳細によって限定がなされないことが意図される。

【００９３】

(参考文献)

[1] Simon, C. and Fassio, G., (2012), Optimierung audiovisueller Medien fuer Hoergeschaedigte, In: Fortschritte der Akustik - DAGA 2012, Darmstadt, March 2012

[2] Ephraim, Y. and Malah, D., (1984), Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator, IEEE Transactions on Acoustics Speech and Signal Processing, 32(6):1109-1121

[3] Kolbaek, M., Yu, D., Tan, Z-H., and Jensen, J., (2017), Multitalker Speech Separation With Utterance-Level Permutation Invariant Training of Deep Recurrent Neural Networks, IEEE Transactions on Audio, Speech and Language Processing, 25(10), 1901-1913, https://doi.org/10.1109/TASLP.2017.2726762

[4] Jouni, P., Torcoli, M., Uhle, C., Herre, J., Disch, S., Fuchs, H., (2019), Source Separation for Enabling Dialogue Enhancement in Object-based Broadcast with MPEG-H, JAES 67, 510-521, https://doi.org/10.17743/jaes.2019.0032

[5] Sauert, B. and Vary, P., (2012), Near end listening enhancement in the presence of bandpass noises, In: Proc. der ITG-Fachtagung Sprachkommunikation, Braunschweig, September 2012

[6] ANSI S3.5, (1997), Methods for calculation of speech intelligibility index

[7] Huber, R., Pusch, A., Moritz, N., Rennies, J., Schepker, H., Meyer, B.T., (2018), Objective Assessment of a Speech Enhancement Scheme with an Automatic Speech Recognition-Based System, ITG-Fachbericht 282: Speech Communication, 10-12, October 2018 in Oldenburg, 86-90
[8] ITU-R Recommendation BS.1387: Method for objective measurements of perceived audio quality (PEAQ)
[9] ITU-T Recommendation P.863: Perceptual objective listening quality assessment
[10] Huber, R. and Kollmeier, B., (2006), PEMO-Q - A New Method for Objective Audio Quality Assessment Using a Model of Auditory Perception, IEEE Transactions on Audio, Speech, and Language Processing, 14(6), 1902-1911
[11] NetMix player of Fraunhofer IIS, http://www.iis.fraunhofer.de/de/bf/amm/forschundentw/forschaudiomulti/dialogenhanc.html
[12] https://auphonic.com/

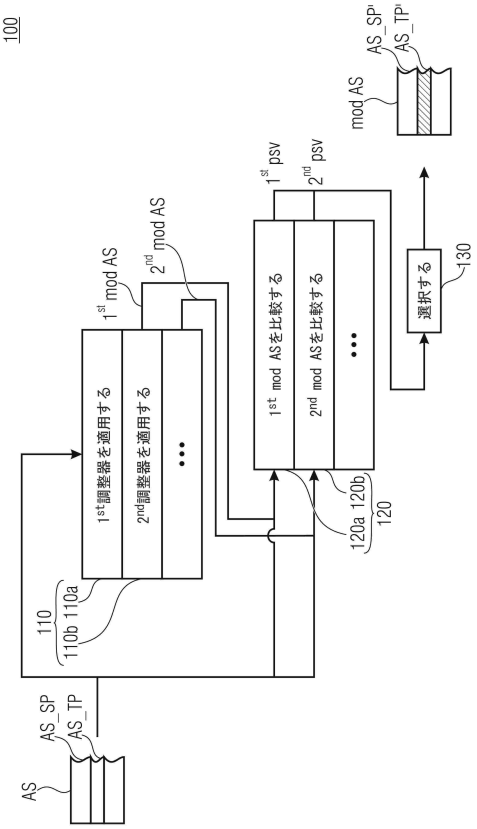
【符号の説明】

【００９４】

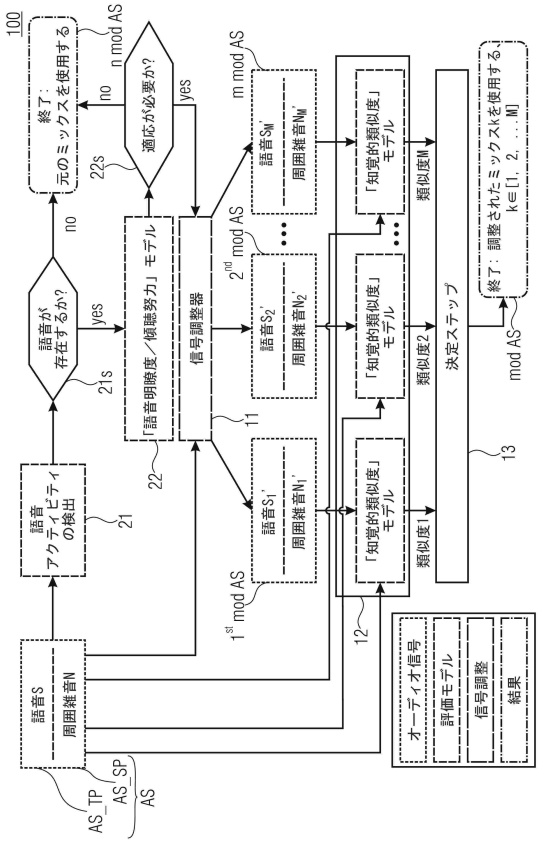
11 信号調整器
12 評価器
13 選択器
21 分析器
22 第2の分析器
100 方法
AS 初期オーディオ信号
AS_TP 語音、対象部分
AS_SP 周囲雑音、副部分
AS_SP', AS_TP' 部分
MOD AS 調整されたオーディオ信号
1st MOD AS, 2nd MOD AS, M MOD AS 調整信号
1st PSV, 2nd PSV 知覚的類似度値

【図面】

【図１】



【図２】



【図３】

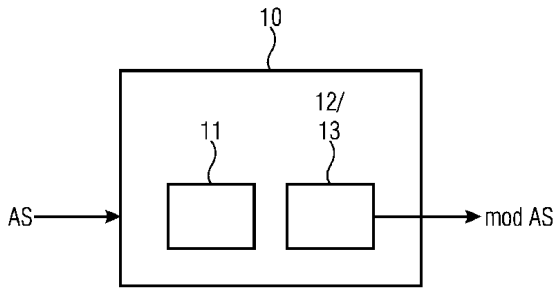


Fig. 3

フロントページの続き

フラウンホファー－インスティテュート・フュア・ディギタル・メディエンテクノロジー・イーデーエムテー・インスティテューツテイル・フアー・スプラッチ－ウント・アウディオテクノロジー・ハーエスアー内
(72)発明者　ヨハンナ・バウムガルトナー－クローネ
ドイツ・２６１２９・オルデンブルク・マリー－キュリー－シュトラーセ・２・フラウンホファー－インスティテュート・フュア・ディギタル・メディエンテクノロジー・イーデーエムテー・インスティテューツテイル・フアー・スプラッチ－ウント・アウディオテクノロジー・ハーエスアー内
審査官　堀　洋介

(56)参考文献　特開２０１０－１６０２４６（ＪＰ，Ａ）
米国特許出願公開第２０１１／０２２４９７６（ＵＳ，Ａ１）
特開２０００－０９９０９６（ＪＰ，Ａ）
特表２０１３－５００４９８（ＪＰ，Ａ）

(58)調査した分野　(Int.Cl.，ＤＢ名)
Ｇ１０Ｌ　２１／００－２１／０２７２
Ｇ１０Ｌ　２５／６０－２５／６９
Ｇ１０Ｌ　１９／００