(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7536083号
(P7536083)

(45)発行日　令和6年8月19日(2024.8.19)

(24)登録日　令和6年8月8日(2024.8.8)

(51)国際特許分類

Ｇ１０Ｌ	*21/0364(2013.01)*	Ｇ１０Ｌ	21/0364	
Ｇ１０Ｌ	*25/30 (2013.01)*	Ｇ１０Ｌ	25/30	
Ｈ０４Ｓ	*7/00 (2006.01)*	Ｈ０４Ｓ	7/00	３２０
Ｈ０４Ｒ	*3/00 (2006.01)*	Ｈ０４Ｓ	7/00	３３０
		Ｈ０４Ｓ	7/00	３４０

ＦＩ

請求項の数　30　（全45頁）　最終頁に続く

(21)出願番号　特願2022-507528(P2022-507528)
(86)(22)出願日　令和2年7月31日(2020.7.31)
(65)公表番号　特表2022-544138(P2022-544138A)
(43)公表日　令和4年10月17日(2022.10.17)
(86)国際出願番号　PCT/EP2020/071700
(87)国際公開番号　WO2021/023667
(87)国際公開日　令和3年2月11日(2021.2.11)
審査請求日　令和4年3月29日(2022.3.29)
(31)優先権主張番号　19190381.4
(32)優先日　令和1年8月6日(2019.8.6)
(33)優先権主張国・地域又は機関
欧州特許庁(EP)
前置審査

(73)特許権者　591037214
フラウンホッファー－ゲゼルシャフト　ツァ　フェルダールング　デァ　アンゲヴァンテン　フォアシュンク　エー．ファオ
ドイツ連邦共和国　８０６８６　ミュンヘン　ハンザシュトラッセ　２７ツェー
(73)特許権者　508097906
イルメナウテヒニッシェ大学
ドイツ連邦共和国　９８６９３　イルメナウ　エーレンベルクシュトラーセ　２９
(74)代理人　100079577
弁理士　岡田　全啓
(72)発明者　スポラー　トーマス
ドイツ連邦共和国　９８６９３　イルメナウ　エーレンベルクシュトラーセ　３１
最終頁に続く

(54)【発明の名称】　選択的受聴を補助するためのシステムおよび方法

(57)【特許請求の範囲】

【請求項１】

選択的受聴を補助するためのシステムであって、

聴覚環境の少なくとも２つの受信済みマイクロフォン信号を処理する機械学習モデルを用いて、１つ以上のオーディオソースのオーディオソース信号部分を検出するための検出器（１１０）と、

前記１つ以上のオーディオソースの各々に位置情報を割り当てるための位置決定器（１２０）と、

所与の音源タイプの任意のインスタンスが聴覚シーンに存在するか否かを決定する音源検出を行うことによって、前記１つ以上のオーディオソースのうちのオーディオソースに既定のクラスのセットからクラスラベルを割り当てるためのオーディオタイプ分類器（１３０）と、

前記オーディオソースの前記クラスラベルに応じて、前記１つ以上のオーディオソースのうちの前記オーディオソースの前記オーディオソース信号部分を変化させて、前記少なくとも１つのオーディオソースの修正済みオーディオ信号部分を得るための信号部分修正器（１４０）と、

前記１つ以上のオーディオソースの各オーディオソースに対して、このオーディオソースの前記位置情報およびユーザの頭の向きに応じて複数の両耳室内インパルス応答（ｂｉｎａｕｒａｌ　ｒｏｏｍ　ｉｍｐｕｌｓｅ　ｒｅｓｐｏｎｓｅｓ）を生成し、前記複数の両耳室内インパルス応答に応じて、かつ前記少なくとも１つのオーディオソースの前記修正

済みオーディオ信号部分に応じて少なくとも２つのラウドスピーカー信号を生成するための信号生成器（１５０）と、
を備える、システム。
【請求項２】
前記位置決定器（１２０）は、前記１つ以上のオーディオソースのそれぞれについて、キャプチャ画像または録画ビデオに応じた前記位置情報を決定するように構成されている、請求項１に記載のシステム。
【請求項３】
前記検出器（１１０）は、前記少なくとも２つの受信済みマイクロフォン信号に応じて、前記聴覚環境の１つ以上の音響特性を決定するように構成されている、請求項１または請求項２に記載のシステム。
【請求項４】
前記信号生成器（１５０）は、前記聴覚環境の前記１つ以上の音響特性に応じて、前記複数の両耳室内インパルス応答を決定するように構成されている、請求項３に記載のシステム。
【請求項５】
前記信号部分修正器（１４０）は、オーディオソース信号部分が修正される前記オーディオソースを、以前に学習したユーザシナリオに依存して選択し、前記以前に学習したユーザシナリオに依存してこれを修正するように構成されている、請求項１ないし４のいずれかに記載のシステム。
【請求項６】
前記システムは、２つ以上の以前に学習したユーザシナリオからなるグループから前記以前に学習されたユーザシナリオを選択するためのユーザインタフェース（１６０）を含む、請求項５に記載のシステム。
【請求項７】
前記検出器（１１０）および／または前記位置決定器（１２０）および／または前記オーディオタイプ分類器（１３０）および／または前記信号部分修正器（１４０）および／または前記信号生成器（１５０）が、ハフ変換（Ｈｏｕｇｈ　ｔｒａｎｓｆｏｒｍａｔｉｏｎ）を用いてまたは複数のＶＬＳＩチップを用いてまたは複数のメモリスタを用いて並列信号処理を行うよう構成されている、請求項１ないし６のいずれかに記載のシステム。
【請求項８】
システムが、聴覚能力が制限されている及び／又は聴覚が損傷しているユーザのための補聴器として機能する聴覚装置（１７０）を含み、前記聴覚装置（１７０）が、前記少なくとも２つのラウドスピーカー信号を出力するための少なくとも２つのラウドスピーカー（１７１、１７２）を含む、請求項１ないし７のいずれかに記載のシステム。
【請求項９】
前記システムが、前記少なくとも２つのラウドスピーカー信号（１８１，１８２）を出力するための少なくとも２つのラウドスピーカー（１８１，１８２）と、前記少なくとも２つのラウドスピーカー（１８１，１８２）を収容するハウジング構造（１８３）とを含み、前記少なくとも１つのハウジング構造（１８３）がユーザの頭部または前記ユーザの他の任意の身体部分に固定されるのに適している、請求項１ないし７のいずれかに記載のシステム。
【請求項１０】
前記システムが、前記少なくとも２つのラウドスピーカー信号を出力するための少なくとも２つのラウドスピーカー（１８１、１８２）を含むヘッドフォン（１８０）を含む、請求項１ないし７のいずれかに記載のシステム。
【請求項１１】
前記検出器（１１０）と前記位置決定器（１２０）と前記オーディオタイプ分類器（１３０）と前記信号部分修正器（１４０）と前記信号生成器（１５０）が前記ヘッドフォン（１８０）に統合されている、請求項１０に記載のシステム。

【請求項１２】
前記システムが、前記検出器（１１０）と前記位置決定器（１２０）と前記オーディオタイプ分類器（１３０）と前記信号部分修正器（１４０）と前記信号生成器（１５０）とを含むリモート装置（１９０）を含み、
前記リモート装置は、前記ヘッドフォンから空間的に分離されている、請求項１０に記載のシステム。
【請求項１３】
前記リモート装置（１９０）は、スマートフォンである、請求項１２に記載のシステム。
【請求項１４】
選択的受聴を補助するための方法であって、
聴覚環境の少なくとも２つの受信済みマイクロフォン信号を処理する機械学習モデルを用いて、１つ以上のオーディオソースのオーディオソース信号部分を検出するステップと、
前記１つ以上のオーディオソースの各々に位置情報を割り当てるステップと、
聴覚シーンに所与の音源タイプの任意のインスタンスが存在するかを決定する音源検出を実行することによって、前記１つ以上のオーディオソースのオーディオソースに既定のクラスのセットからクラスラベルを割り当てるステップと、
前記少なくとも１つのオーディオソースの修正済みオーディオ信号部分を得るために、前記オーディオソースの前記クラスラベルに応じて、前記１つ以上のオーディオソースのうちの前記オーディオソースの前記オーディオソース信号部分を変化させるステップと、
前記１つ以上のオーディオソースの各オーディオソースに対して、このオーディオソースの前記位置情報およびユーザの頭の向きに応じて複数の両耳室内インパルス応答（ｂｉｎａｕｒａｌ　ｒｏｏｍ　ｉｍｐｕｌｓｅ　ｒｅｓｐｏｎｓｅｓ）を生成し、前記複数の両耳室内インパルス応答に応じて、かつ前記少なくとも１つのオーディオソースの前記修正済みオーディオ信号部分に応じて少なくとも２つのラウドスピーカー信号を生成するステップと、
を備える、方法。
【請求項１５】
請求項１４に記載の方法を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。
【請求項１６】
１つ以上の室内音響パラメータを決定するための装置であって、
前記装置は、１つ以上のマイクロフォン信号を含むマイクロフォンデータを取得するように構成されており、
前記装置は、ユーザの位置および／または向きに関する追跡データを取得するように構成されており、
前記装置は、前記マイクロフォンデータに応じて、かつ前記追跡データに応じて、前記１つ以上の室内音響パラメータを決定するように構成されている、装置を含む、請求項１ないし１３のいずれかに記載のシステム。
【請求項１７】
前記装置は、前記マイクロフォンデータに応じて、かつ前記追跡データに応じて、前記１つ以上の室内音響パラメータを決定するために機械学習を用いるように構成されている、請求項１６に記載のシステム。
【請求項１８】
前記装置は、前記装置がニューラルネットワークを採用するように構成されている点で、機械学習を用いるように構成されている、請求項１７に記載のシステム。
【請求項１９】
前記装置は、機械学習のためにクラウドベースの処理（ｃｌｏｕｄ－ｂａｓｅｄ　ｐｒｏｃｅｓｓｉｎｇ）を採用するように構成されている、請求項１７または１８に記載のシステム。
【請求項２０】

前記１つ以上の室内音響パラメータは残響時間を含む、請求項１６ないし１９のいずれかに記載のシステム。
【請求項２１】
前記１つ以上の室内音響パラメータは直間比（ｄｉｒｅｃｔ－ｔｏ－ｒｅｖｅｒｂｅｒａｎｔ　ｒａｔｉｏ）を含む、請求項１６ないし２０のいずれかに記載のシステム。
【請求項２２】
前記追跡データは、前記ユーザの前記位置を分類するためのｘ座標、ｙ座標、およびｚ座標を含む、請求項１６ないし２１のいずれかに記載のシステム。
【請求項２３】
前記追跡データは、前記ユーザの前記向きを分類するためのピッチ座標、ヨー座標、およびロール座標を含む、請求項１６ないし２２のいずれかに記載のシステム。
【請求項２４】
前記装置が、前記１つ以上のマイクロフォン信号を時間領域から周波数領域に変換するように構成されており、
前記装置が、前記周波数領域における前記１つ以上のマイクロフォン信号の１つ以上の特徴を抽出するように構成されており、
前記装置は、前記１つ以上の特徴に応じて、前記１つ以上の室内音響パラメータを決定するように構成されている、請求項１６ないし２３のいずれかに記載のシステム。
【請求項２５】
前記装置は、前記１つ以上の特徴を抽出するためにクラウドベースの処理（ｃｌｏｕｄ－ｂａｓｅｄ　ｐｒｏｃｅｓｓｉｎｇ）を用いるように構成されている、請求項２４に記載のシステム。
【請求項２６】
前記装置が、前記１つ以上のマイクロフォン信号を記録するための複数のマイクロフォンのマイクロフォン設備を含む、請求項１６ないし２５のいずれかに記載のシステム。
【請求項２７】
前記マイクロフォン設備は、ユーザの身体に装着されるように構成される、請求項２６に記載のシステム。
【請求項２８】
前記信号部分修正器（１４０）が、前記１つ以上の室内音響パラメータのうちの少なくとも１つの室内音響パラメータに応じて、前記１つ以上のオーディオソースのうちの前記オーディオソースの前記オーディオソース信号部分の前記変化を実行するように構成されている、および／または、
前記信号生成器（１５０）は、前記１つ以上のオーディオソースの各オーディオソースについての前記複数の両耳室内インパルス応答のうちの少なくとも１つの生成を、前記１つ以上の室内音響パラメータの前記少なくとも１つに依存して実行するように構成されている、請求項１６ないし２７のいずれかに記載のシステム。
【請求項２９】
１つ以上のマイクロフォン信号を含むマイクロフォンデータを取得するステップと、
ユーザの位置および／または向きに関する追跡データを取得するステップと、
前記マイクロフォンデータに応じて、かつ前記追跡データに応じて、１つ以上の室内音響パラメータを決定するステップと、
を含む、請求項１４に記載の方法。
【請求項３０】
請求項２９に記載の方法を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。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、空間の記録、分析、再生、知覚の側面、特に両耳分析および合成に関するも

のである。
【背景技術】
【０００２】
選択的受聴（ＳＨ）とは、リスナーが聴覚の中で特定の音源または複数の音源に注意を向ける能力のことである。逆に言えば、興味のない音源へのリスナーのフォーカスが減少することを意味する。
【０００３】
このように、人間のリスナーは大音量の環境下でもコミュニケーションをとることができる。これは通常、異なる側面を利用する。両耳で聞く場合、音に関して、方向に依存した時間差とレベル差、そして方向に依存した異なるスペクトルの色付けが存在する。このため、聴覚は音源の方向を判断し、その方向に集中することができる。
【０００４】
また、自然界の音源、特に音声の場合、異なる周波数の信号部分が一時的に結合している。これにより、片耳で聞く場合でも、聴覚は異なる音源を分離することができる。両耳聴音では、この２つを併用する。さらに、局所的に発生する大音量の騒音は、積極的に無視することができる。
【０００５】
文献上、選択的受聴の概念は、アシストリスニング（ａｓｓｉｓｔｅｄ　ｌｉｓｔｅｎｉｎｇ）［１］、仮想および増幅された聴覚環境［２］といった他の用語と関連している。アシストリスニングは、仮想、増幅、ＳＨのアプリケーションを含む、より広い用語である。
【０００６】
先行技術によれば、従来の聴覚装置はモノラル方式で動作する。すなわち、右耳と左耳とに対する信号処理は、周波数応答と動的圧縮に関して完全に独立している。その結果、両耳の信号間の時間差、レベル差、周波数差が失われてしまう。
【０００７】
最近のいわゆる両耳装用型聴覚装置（補聴器）は、２つの聴覚装置の補正係数をカップリングしている。多くの場合、複数のマイクロフォンを備えているが、通常は「最も音声に近い」信号を持つマイクロフォンのみが選択され、ビームフォーミングは計算されない。複雑な聴覚状況では、望ましい音信号と望ましくない音信号とが同じように増幅されるため、望ましい音成分に焦点を当てることはサポートされていない。
【０００８】
電話機などのハンズフリー機器の分野では、現在すでに複数のマイクロフォンが使用されており、それぞれのマイクロフォン信号からいわゆるビームが計算され、ビームの方向から来る音は増幅され、他の方向からの音は低減される。今日の方法は、背景の一定の音（例えば、車のエンジン音や風切り音）を学習し、さらにビームを通してよく局在化された大きな騒音を学習し、これらを使用信号から差し引く（例：一般化サイドローブキャンセラー（ｓｉｄｅ　ｌｏｂｅ　ｃａｎｃｅｌｌｅｒ））。電話システムでは、音声の静的特性を検出する検出器を使用して、音声のように構成されていないものをすべて抑圧することもある。ハンズフリー機器では、最終的にモノラル信号のみが伝送されるため、状況を把握するのに役立つ空間情報が伝送経路で失われ、特に複数の話者が相互に通話する場合に「そこに人がいる」かのような錯覚が生じる。非音声信号を抑制することで、会話相手の音響環境に関する重要な情報が失われ、コミュニケーションに支障をきたす可能性がある。
【０００９】
人間は本来、「選択的受聴」が可能であり、周囲の個々の音源に意識的に焦点を合わせることができる。人工知能（ＡＩ）による自動的な選択的受聴のシステムは、まずその基礎となる概念を学習する必要がある。音響シーンの自動分解（シーンデコンポジション（シーン分解）：ｓｃｅｎｅ　ｄｅｃｏｍｐｏｓｉｔｉｏｎ）には、まずすべてのアクティブな音源を検出・分類し、さらにそれらを別々のオーディオ（音声、音）オブジェクトと

して処理、増幅、あるいは弱めることができるように分離することが必要である。

【００１０】

聴覚シーン解析の研究分野では、録音されたオーディオ信号に基づいて、足音、拍手、叫び声などの時間的に配置された音響イベントと、コンサート、レストラン、スーパーマーケットなどのよりグローバルな音響シーンを検出及び分類しようとするものである。この場合、現在の手法では、もっぱら人工知能（ＡＩ）やディープラーニングの分野の手法が用いられている。これは、大量の学習量に基づいて、オーディオ信号の特徴的なパターンを検出するために学習するディープニューラルネットワークのデータ駆動型（ｄａｔａ－ｄｒｉｖｅｎ　ｌｅａｒｎｉｎｇ）の学習を含む［７０］。とりわけ、画像処理（コンピュータビジョン）や音声処理（自然言語処理）の研究分野の進歩に触発されて、スペクトログラム表現の２次元パターン検出のための畳み込みニューラルネットワーク（ｃｏｎｖｏｌｕｔｉｏｎａｌ　ｎｅｕｒａｌ　ｎｅｔｗｏｒｋｓ）と音の時間的モデリングのための再帰層（ｒｅｃｕｒｒｅｎｔ　ｌａｙｅｒｓ）（リカレントニューラルネットワーク：ｒｅｃｕｒｒｅｎｔ　ｎｅｕｒａｌ　ｎｅｔｗｏｒｋｓ）の混合が、原則として使用される。

【００１１】

オーディオ解析の場合、具体的に対処すべき課題が山積している。ディープラーニングモデル（深層学習モデル）は、その複雑さゆえに、非常にデータを必要とする。画像処理や音声処理の研究分野とは対照的に、オーディオ処理で利用できるデータセットは比較的に少ない。最大のデータセットはＧｏｏｇｌｅのＡｕｄｉｏＳｅｔデータセット［８３］で、約２００万の音例と６３２種類の音イベントクラスがあるが、研究で使われるほとんどのデータセットはかなり小さい。この少量の学習データは、たとえば、転送学習を使用して対処できる。この場合、大きなデータセットで事前学習されたモデルは、その後、ユースケース用に決定された新しいクラスを使用して小さなデータセットに微調整される（微調整（ｆｉｎｅ－ｔｕｎｉｎｇ））［７７］。さらに、半教師付き学習の手法を利用することで、一般に大量に存在する注釈のないオーディオデータも学習対象とする。

【００１２】

さらに、画像処理との大きな違いは、同時に聞こえる音響イベントの場合、（画像の場合のように）サウンドオブジェクトのマスキングがなく、位相に依存した複雑なオーバーラップがあることである。ディープラーニングにおける現在のアルゴリズムは、いわゆる「注意」のメカニズムを使用しており、例えば、モデルが特定の時間セグメントまたは周波数範囲において分類に焦点を当てることを可能にする［２３］。音イベントの検出は、その継続時間に関する高い分散によってさらに複雑である。アルゴリズムは、ピストルの発砲のような非常に短いイベントから、列車の通過のような長いイベントまで、確実に検出できる必要がある。

【００１３】

学習データの収録時の音響条件に強く依存するため、空間的な残響やマイクロフォンの位置が異なるような新しい音響環境では、モデルが予期せぬ挙動を示すことがよくある。この問題を軽減するために、さまざまな解決方法が開発されている。たとえば、データ補強法は、異なる音響条件のシミュレーション［６８］や異なる音源の人工的なオーバーラップによって、より高いロバスト性とモデルの不変性を達成しようとするものである。さらに、複雑なニューラルネットのパラメータを別の方法で調節することで、学習データに対する過剰学習や特殊化を回避し、同時に未知のデータに対するより良い一般化を達成することができる。近年では、以前に学習したモデルを新しい応用条件に適応させるための「領域適応（ｄｏｍａｉｎ　ａｄａｐｔｉｏｎ）」［６７］のためのさまざまなアルゴリズムが提案されている。このプロジェクトで計画しているヘッドフォン内での使用シナリオでは、音源検出アルゴリズムのリアルタイム能力（ｒｅａｌ－ｔｉｍｅ　ｃａｐａｂｉｌｉｔｙ）が重要な意味を持つ。ここでは、ニューラルネットワークの複雑さと、基盤となるコンピューティングプラットフォームの計算処理の最大可能回数との間のトレードオフが必然的に行われなければならない。たとえ音イベントの持続時間が長い場合でも、対応する音源分離を開始するためには、できるだけ早く検出する必要がある。

【００１４】
フラウンホーファーＩＤＭＴでは、近年、自動音源探査（ａｕｔｏｍａｔｅｄ ｓｏｕｎｄ ｓｏｕｒｃｅ ｄｅｔｅｃｔｉｏｎ）の分野で多くの研究を行っている。研究プロジェクト「ＳｔａｄｔＬａｒｍ」では、騒音レベルを測定し、都市内のさまざまな場所で録音されたオーディオ信号に基づいて、１４種類の音響シーンとイベントのクラスを分類できる分散型のセンサーネットワーク（ｄｉｓｔｒｉｂｕｔｅｄ ｓｅｎｓｏｒ ｎｅｔｗｏｒｋ）が開発された［６９］。この場合、センサーでの処理は、組み込みプラットフォームであるｒａｓｐｂｅｒｒｙ Ｐｉ ３（ラズベリーパイスリー）上でリアルタイムに実行される。また、先行研究として、オートエンコーダーネットワークをベースにしたスペクトログラムのデータ圧縮のための新しいアプローチを検討した［７１］。近年、音楽信号処理（音楽情報検索）の分野でディープラーニングの手法を用いることで、楽曲の書き起こし［７６］、［７７］、コード検出［７８］、楽器検出［７９］などのアプリケーションで大きな進展があった。産業用オーディオ処理の分野では、新しいデータセットが確立され、電気モータの音響状態の監視などにディープラーニングの手法が用いられている［７５］。

【００１５】
本実施形態で扱うシナリオは、当初は数や種類が不明で、常に変化する可能性がある複数の音源を想定している。音源分離では、複数のスピーカのような類似した特性をもつ複数の音源が特に大きな課題となる［８０］。

【００１６】
また、高い空間分解能を得るためには、複数のマイクロフォンをアレイ状に配置する必要がある［７２］。このような録音シナリオでは、従来のモノラル（１チャネル）やステレオ（２チャネル）のオーディオ記録とは異なり、リスナーのまわりの音源を正確に位置特定（ｌｏｃａｌｉｚａｔｉｏｎ）することができる。

【００１７】
音源分離アルゴリズムでは、音源間のひずみやクロストークなどのアーチファクトが残るが［５］、これは一般に聴取者に邪魔なものとして知覚される可能性がある。このようなアーチファクトは、トラックを再ミキシングすることで、部分的にマスキングされ、低減される［１０］。

【００１８】
「ブラインド」音源分離を強化するために、検出された音源の数や種類、または推定された空間位置などの追加情報がしばしば使用される（インフォームド・ソース分離（ｉｎｆｏｒｍｅｄ ｓｏｕｒｃｅ ｓｅｐａｒａｔｉｏｎ）［７４］）。複数の講演者が活動している会議の場合、現在の分析システムは、講演者の数を同時に推定し、それぞれの時間的活動を決定し、その後、音源分離の手段によって講演者を分離することができる［６６］。

【００１９】
フラウンホーファーＩＤＭＴでは、近年、音源分離アルゴリズムの知覚に基づく評価に関する多くの研究が行われている［７３］。

【００２０】
音楽信号処理の分野では、独奏楽器の基本周波数推定を追加情報として利用し、独奏楽器と伴奏楽器を分離するリアルタイム可能なアルゴリズムが開発されている［８１］。また、ディープラーニングを用いた複雑な楽曲から歌唱を分離する手法も提案されている［８２］。また、産業用音声解析のコンテキストで適用するために、特別な音源分離アルゴリズムが開発されている［７］。

【００２１】
ヘッドフォンは周囲の音響感覚に大きな影響を与える。ヘッドフォンの構造によって、耳に入射する音は異なる程度に減衰される。インイヤー型ヘッドフォンは、耳のチャンネルを完全に塞ぐ［８５］。外耳を音響的に囲む密閉型ヘッドフォンは、リスナーを外部環境から強く遮断する。開放型および半開放型ヘッドフォンは、音を完全または部分的に通

過させることができる［８４］。日常生活の多くの用途において、ヘッドフォンは、その構造タイプで可能なものよりも強く、望ましくない周囲の音を分離することが望まれる。

【００２２】

外部からの干渉影響は、アクティブノイズコントロール（ＡＮＣ：ａｃｔｉｖｅ　ｎｏｉｓｅ　ｃｏｎｔｒｏｌ）により、さらに減衰させることができる。これは、ヘッドフォンのマイクロフォンによって入射音信号を記録することによって、かつ、これらの音部分とヘッドフォンを透過した音部分が干渉によって互いに打ち消し合うようにラウドスピーカで再生することによって実現される。これにより、全体として周囲から強く隔離された音響を実現することができる。しかし、日常的な場面では危険も伴うため、オンデマンドでインテリジェントにこの機能をオンにできることが望まれている。

【００２３】

最初の製品は、パッシブアイソレーション（ｐａｓｓｉｖｅ　ｉｓｏｌａｔｉｏｎ）を減らすために、マイクフォン信号をヘッドフォンに通すことを可能にする。このように、プロトタイプ［８６］以外にも、”トランスペアレントリスニング（ｔｒａｎｓｐａｒｅｎｔ　ｌｉｓｔｅｎｉｎｇ）”の機能を宣伝している製品がすでに存在する。例えば、ゼンハイザーはヘッドセット（Ｓｅｎｎｈｅｉｓｅｒ）”ＡＭＢＥＯ”［８８］で、ブラギは製品”Ｔｈｅ　Ｄａｓｈ　Ｐｒｏ”でその機能を提供している。しかし、この可能性はまだ始まりに過ぎない。将来的には、この機能を大幅に拡張し、周囲の音を完全にオン・オフするだけでなく、個々の信号部分（例えば、音声やアラーム信号のみ）を要求に応じて排他的に聞こえるようにすることができる。フランスのオロサウンド社（Ｏｒｏｓｏｕｎｄ）は、ヘッドセット「ティルデイヤホン（Ｔｉｌｄｅ　Ｅａｒｐｈｏｎｅｓ）」［８９］を装着した人が、スライダーでＡＮＣの強さを調節できるようにしている。また、ＡＮＣを作動させた状態で、会話相手の音声を導くこともできる。ただし、これは、会話相手が６０°の円錐の中に対面している場合にのみ機能する。方向性に依存しない適応は不可能である。

【００２４】

特許出願公開ＵＳ２０１５　１９５６４１　Ａ１（参照：［９１］）には、ユーザの聴覚環境を生成するために実装された方法が開示されている。この場合、方法は、ユーザの周囲聴覚環境を表す信号を受信することと、周囲聴覚環境における複数の音タイプのうちの少なくとも１つの音タイプを識別するように、マイクロプロセッサを使用して信号を処理することとを含む。さらに、本方法は、複数の音タイプのそれぞれに対するユーザの好みを受信することと、周囲聴覚環境における音タイプごとに信号を修正することと、ユーザの聴覚環境を生成するように修正した信号を少なくとも１つのラウドスピーカに出力することと、を含む。

【発明の概要】

【発明が解決しようとする課題】

【００２５】

請求項１によるシステム、請求項１６による方法、請求項１７によるコンピュータプログラム、請求項１８による装置、請求項３２による方法、及び請求項３３によるコンピュータプログラムが提供される。

【課題を解決するための手段】

【００２６】

選択的受聴を支援するためのシステムが提供される。このシステムは、聴覚環境の少なくとも２つの受信済みマイクロフォン信号を用いて、１つ以上のオーディオソースのオーディオソース信号部分を検出するための検出器を含む。さらに、このシステムは、１つ以上のオーディオソースの各々に位置情報を割り当てるための位置決定器を含む。さらに、このシステムは、１つ以上のオーディオソースの各々のオーディオソース信号部分にオーディオ信号タイプを割り当てるためのオーディオタイプ分類器を含む。さらに、本システムは、少なくとも１つのオーディオソースの修正済みオーディオ信号部分を得るように、少なくとも１つのオーディオソースのオーディオソース信号部分のオーディオ信号タイプに応じて、１つ以上のオーディオソースのうちの少なくとも１つのオーディオソースのオ

ーディオソース信号部分を変化させるための信号部分修正器を含んでいる。さらに、このシステムは、１つ以上のオーディオソースの各オーディオソースに対して、このオーディオソースの位置情報およびユーザの頭の向きに応じて複数の両耳室内インパルス応答（ｂｉｎａｕｒａｌ　ｒｏｏｍ　ｉｍｐｕｌｓｅ　ｒｅｓｐｏｎｓｅｓ）を生成し、複数の両耳室内インパルス応答に応じて、かつ少なくとも１つのオーディオソースの修正済みオーディオ信号部分に応じて少なくとも２つのラウドスピーカー信号を生成するための信号生成器と、を含む。

【００２７】

さらに、選択的受聴を補助する方法を提供する。本方法は以下を含む。

－聴覚環境の少なくとも２つの受信済みマイクロフォン信号を用いて、１つ以上のオーディオソースのオーディオソース信号部分を検出するステップ。

－１つ以上のオーディオソースの各々に位置情報を割り当てるステップ。

－１つ以上のオーディオソースの各々のオーディオソース信号部分にオーディオ信号タイプを割り当てるステップ。

－少なくとも１つのオーディオソースの修正済みオーディオ信号部分を得るように、少なくとも１つのオーディオソースのオーディオソース信号部分のオーディオ信号タイプ　に応じて、１つ以上のオーディオソースのうちの少なくとも１つのオーディオソース　のオーディオソース信号部分を変化させるステップ。

－１つ以上のオーディオソースの各オーディオソースに対して、このオーディオソース位置情報およびユーザの頭の向きに応じて複数の両耳室内インパルス応答（ｂｉｎａｕ　ｒａｌ　ｒｏｏｍ　ｉｍｐｕｌｓｅ　ｒｅｓｐｏｎｓｅｓ）を生成し、複数の両耳室　内インパルス応答に応じて、かつ少なくとも１つのオーディオソースの修正済みオー　ディオ信号部分に応じて少なくとも２つのラウドスピーカー信号を生成するステップ　。

【００２８】

さらに、上記方法を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・プログラムが提供される。

【００２９】

さらに、１つ以上の室内音響パラメータを決定するための装置が提供される。本装置は、１つ以上のマイクロフォン信号を含むマイクロフォンデータを取得するように構成される。さらに、本装置は、ユーザの位置及び／又は方向に関する追跡データを取得するように構成される。さらに、本装置は、マイクロフォンデータに応じて、および追跡データに応じて、１つ以上の室内音響パラメータを決定するように構成されている。

【００３０】

さらに、１つ以上の室内音響パラメータを決定するための方法が提供される。本方法は以下を含む。

－１つ以上のマイクロフォン信号を含むマイクロフォンデータを取得するステップ。

－　ユーザの位置および／または方向に関する追跡データを取得するステップ。

－マイクロフォンデータおよび追跡データに応じて、１つ以上の室内音響パラメータを決定するステップ。

【００３１】

さらに、上記方法を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も提供される。

【００３２】

とりわけ、実施形態は、正常な聴覚を有する人々および損傷した聴覚を有する人々のために、音質および生活の質の向上（例えば、所望の音がより大きくなる、所望の音がより小さくなる、より良い音声理解度）が達成されるように、技術システムにおいて聴覚補助のための異なる技術を取り入れ、組み合わせることに基づくものである。

【００３３】

続いて、本発明の好ましい実施形態について、図面を参照しながら説明する。

【図面の簡単な説明】

【００３４】
【図１】図１は、実施形態による選択的受聴を支援するシステムを示す図である。
【図２】図２は、実施形態によるシステムを示し、さらに、ユーザインタフェースを含む。
【図３】図３は、対応する２つのラウドスピーカーを有する聴覚装置を含む、実施形態によるシステムを示す。
【図４】図４は、ハウジング構造および２つのラウドスピーカーを含む、実施形態によるシステムを示す図である。
【図５】図５は、２つのラウドスピーカーを有するヘッドフォンを含む、実施形態によるシステムを示す図である。
【図６】図６は、検出器と位置決定器と音声タイプ分類器と信号部分修正器と信号生成器を含むリモート装置１９０を含む、実施形態によるシステムを示す図である。
【図７】図７は、５つのサブシステムを含む、一実施形態によるシステムを示す図である。
【図８】図８は、実施形態による対応するシナリオを示す図である。
【図９】図９は、４つの外部音源を有する実施形態によるシナリオを示す図である。
【図１０】図１０は、実施形態によるＳＨアプリケーションの処理ワークフローを示す図である。
【発明を実施するための形態】
【００３５】
現代の生活では、メガネが多くの人の知覚を助けている。聴覚については、補聴器があるが、正常な聴覚を持つ人でも、多くの場面で知能システムによる補助を受けることができる。
【００３６】
周囲がうるさくて、特定の音だけが気になり、選択的に聞きたいと思うことはよくある。人間の脳はすでにその能力に長けているが、今後さらに知的な支援を受けることで、この選択的な聞き取りを大幅に向上させることができる。このような「インテリジェント・ヒアラブル（ｉｎｔｅｌｌｉｇｅｎｔ　ｈｅａｒａｂｌｅｓ）」（聴覚装置、補聴器）を実現するためには、技術システムが（音響）環境を分析し、個々の音源を識別して、個別に処理できるようにする必要がある。このテーマについては、従来から研究されているが、音響環境全体をリアルタイム（耳に対して透過）に、かつ高音質（通常の音響環境と区別がつかないほど）で解析及び処理することは、従来技術では実現されていない。
【００３７】
以下では、機械リスニングのための改良されたコンセプトを紹介する。
【００３８】
図１は、実施形態による選択的聴力を支援するシステムを示す図である。
【００３９】
本システムは、聴覚環境（またはリスニング環境）の少なくとも２つの受信済みマイクロフォン信号を用いて、１つ以上のオーディオソースのオーディオソース信号部分を検出するための検出器１１０を含む。
【００４０】
さらに、システムは、１つ以上のオーディオソースの各々に位置情報を割り当てるための位置決定器１２０を含む。
【００４１】
さらに、このシステムは、１つ以上のオーディオソースの各々のオーディオソース信号部分にオーディオ信号タイプを割り当てるためのオーディオタイプ分類器１３０を含む。
【００４２】
さらに、本システムは、少なくとも１つのオーディオソースの修正済みオーディオ信号部分を得るように、少なくとも１つのオーディオソースのオーディオソース信号部分のオーディオ信号タイプに応じて、１つ以上のオーディオソースの少なくとも１つのオーディオソースのオーディオソース信号部分を変化させるための信号部分修正器１４０を含んでいる。

【００４３】
さらに、このシステムは、１つ以上のオーディオソースの各オーディオソースについて、このオーディオソースの位置情報およびユーザの頭の向きに応じて複数の両耳室内インパルス応答（ｂｉｎａｕｒａｌ　ｒｏｏｍ　ｉｍｐｕｌｓｅ　ｒｅｓｐｏｎｓｅｓ）を生成するための信号生成器１５０と、複数の両耳室内インパルス応答に応じて、かつ少なくとも１つのオーディオソースの修正済みオーディオ信号部分に応じて少なくとも２つのラウドスピーカー信号を生成するための信号生成器１５０とを含む。
【００４４】
実施形態によれば、例えば、検出器１１０は、ディープラーニングモデルを使用して、１つ以上のオーディオソースのオーディオソース信号部分を検出するように構成されてもよい。
【００４５】
実施形態において、例えば、位置決定器１２０は、１つ以上のオーディオソースのそれぞれについて、キャプチャされた画像または記録されたビデオに依存する位置情報を決定するように構成されてもよい。
【００４６】
実施形態によれば、例えば、位置決定器１２０は、１つ以上のオーディオソースのそれぞれについて、ビデオ内の人物の唇の動きを検出することによって、および唇の動きに応じて１つ以上のオーディオソースのうちの１つのオーディオソース信号部分に同じものを割り当てることによって、ビデオに応じた位置情報を決定するよう構成されてもよい。
【００４７】
実施形態において、例えば、検出器１１０は、少なくとも２つの受信済みマイクロフォン信号に応じて、聴覚環境の１つ以上の音響特性を決定するように構成されてもよい。
【００４８】
実施形態によれば、例えば、信号生成器１５０は、聴覚環境の１つ以上の音響特性に応じて、複数の両耳室内インパルス応答を決定するように構成されてもよい。
【００４９】
実施形態において、例えば、信号部分修正器１４０は、オーディオソース信号部分が以前に学習したユーザシナリオに応じて修正される少なくとも１つのオーディオソースを選択し、以前に学習したユーザシナリオに応じてそれを修正するように構成されてもよい。
【００５０】
実施形態によれば、例えば、システムは、２つ以上の以前に学習されたユーザシナリオのグループから以前に学習されたユーザシナリオを選択するためのユーザインタフェース１６０を含んでもよい。図２は、そのようなユーザインタフェース１６０を追加的に含む、実施形態によるそのようなシステムを示す。
【００５１】
実施形態において、例えば、検出器１１０および／または位置決定器１２０および／またはオーディオタイプ分類器１３０および／または信号修正器１４０および／または信号生成器１５０は、ハフ変換（Ｈｏｕｇｈ　ｔｒａｎｓｆｏｒｍａｔｉｏｎ）を使用して、または複数のＶＬＳＩチップを採用して、または複数のメモリスタを採用して、並列信号処理を行うように構成されてもよい。
【００５２】
実施形態によれば、例えば、システムは、聴覚能力が制限されている及び／又は聴覚が損傷しているユーザのための補聴器として機能する聴覚装置１７０を含み得、聴覚装置は、少なくとも２つのラウドスピーカー信号を出力するための少なくとも２つのラウドスピーカー１７１、１７２を含む。図３は、対応する２つのラウドスピーカー１７１、１７２を有するこのような聴覚装置１７０を含む、一実施形態によるこのようなシステムを示す。
【００５３】
実施形態では、例えば、システムは、少なくとも２つのラウドスピーカー信号を出力するための少なくとも２つのラウドスピーカー１８１、１８２と、少なくとも２つのラウド

スピーカーを収容するハウジング構造１８３とを含み、少なくとも１つのハウジング構造１８３は、ユーザの頭部１８５又はユーザの他の身体部分に固定されるのに適している。図４は、そのようなハウジング構造１８３と２つのラウドスピーカー１８１、１８２とを含む対応するシステムを示している。

【００５４】

実施形態によれば、例えば、システムは、少なくとも２つのラウドスピーカー信号を出力するための少なくとも２つのラウドスピーカー１８１、１８２を含むヘッドフォン１８０を含んでもよい。図５は、実施形態による、２つのラウドスピーカー１８１、１８２を有する対応するヘッドフォン１８０を示す。

【００５５】

実施形態において、例えば、検出器１１０と位置決定器１２０とオーディオタイプ分類器１３０と信号部分修正器１４０と信号生成器１５０は、ヘッドフォン１８０に統合されてもよい。

【００５６】

図６に示される実施形態によれば、例えば、システムは、検出器１１０と位置決定器１２０とオーディオタイプ分類器１３０と信号部分修正器１４０と信号生成器１５０とを含むリモート装置１９０を含むことができる。この場合、リモート装置１９０は、例えば、ヘッドフォン１８０から空間的に分離されてもよい。

【００５７】

実施形態において、例えば、リモート装置１９０は、スマートフォンであってもよい。

【００５８】

実施形態は、必ずしもマイクロプロセッサを使用しないが、特に人工ニューラルネットワークのためにも、ハフ変換などの並列信号処理ステップ、ＶＬＳＩチップ、またはエネルギー効率の良い実現のためのメモリスタなどを使用する。

【００５９】

実施形態では、聴覚環境は、空間的に捕捉され、再現され、これは、一方では、入力信号の表現のために複数の信号を使用し、他方では、空間的な再現も使用する。

【００６０】

実施形態において、信号分離は、ディープラーニング（ＤＬ）モデル（例えば、ＣＮＮ、ＲＣＮＮ、ＬＳＴＭ、シャムネットワーク（Ｓｉａｍｅｓｅ　ｎｅｔｗｏｒｋ））によって実施され、同時に、少なくとも２つのマイクロフォンチャネルからの情報を処理し、ここで、各可聴体に少なくとも１つのマイクロフォンが存在する。本発明によれば、（個々の音源に応じた）複数の出力信号が、相互分析を通じてそれぞれの空間位置とともに決定される。記録手段（マイクロフォン）が頭部に接続されている場合、オブジェクトの位置は頭部の動きによって変化する。これにより、例えば音源の方を向くことで、重要な音やそうでない音に自然に焦点を合わせることができる。

【００６１】

いくつかの実施形態では、信号解析のためのアルゴリズムは、例えばディープラーニングアーキテクチャに基づくものである。あるいは、これは、局在化、検出、および音の分離のために、分析器によるバリエーション、または分離されたネットワークによるバリエーションを使用する。一般化相互相関（相関バーサス時間オフセット）の代替的な使用は、頭部による周波数依存のシャドウイング／アイソレーションに対応し、位置特定、検出、および音源分離を向上させる。

【００６２】

実施形態によれば、異なるソースカテゴリ（例えば、音声、車両、男性／女性／子供の声、警告音など）が、トレーニング段階において検出器によって学習される。ここで、音源分離ネットワークは、高い信号品質に関してもトレーニングされ、また、高い精度の位置特定に関しても、ターゲット刺激による位置特定ネットワークと同様にトレーニングされる。

【００６３】

例えば、上記のトレーニングステップでは、マルチチャンネルのオーディオデータを使用する。通常、最初のトレーニングラウンドは、ラボでシミュレーションまたは録音されたオーディオデータを使用して実施される。これに続いて、異なる自然環境（リビングルーム、教室、駅、（工業）生産環境など）でのトレーニングが実行され、すなわち伝達学習とドメイン適応が行われる。

【００６４】

代替的または追加的に、位置検出器は、音源の視覚的位置も決定するように、１つ以上のカメラに結合され得る。音声の場合、唇の動きと音源分離器から来る音声信号が相関し、より正確な位置特定を実現することができる。

【００６５】

学習後、ネットワーク・アーキテクチャ及び関連するパラメータを持つＤＬモデルが作成される。

【００６６】

いくつかの実施形態では、可聴化（オーラリゼーション）は両耳合成（バイノーラル合成）によって実施される。両耳合成は、望ましくない成分を完全に削除することはできないが、知覚できるが邪魔にならない程度に減らすことが可能であるというさらなる利点を提供する。これにより、完全に音を消した場合には聞き逃してしまうような、予期しない別の音源（警告信号、叫び声など）を知覚することができるというさらなる利点がある。

【００６７】

いくつかの実施形態によれば、聴覚環境の分析は、オブジェクトを分離するためだけでなく、音響特性（例えば残響時間、初期時間ギャップ）を分析するためにも使用される。これらの特性は、両耳合成において、あらかじめ保存された（おそらく個別化された）両耳室内インパルス応答（ＢＲＩＲ）を実際の部屋（または空間）に適合させるために用いられる。部屋の発散を減らすことで、リスナーは最適化された信号を理解する際のリスニングの労力が大幅に軽減される。室内発散を最小化することは、聴覚事象の外在化に影響を与え、その結果、モニター室での空間オーディオ再生の妥当性に影響を与える。音声の理解や最適化された信号の一般的な理解については、従来技術では解決策がない。

【００６８】

実施形態では、どの音源を選択するかを決定するために、ユーザインタフェースが使用される。本発明によれば、これは、「真正面からの音声を増幅する」（１人との会話）、「＋６０度の範囲での音声を増幅する」（グループでの会話）、「音楽を抑制して音楽を増幅する」（コンサートの来場者の声を聞きたくない）、「すべてを消音する」（一人になりたい）、「すべての叫びと警告音を抑える」等の異なるユーザシナリオを事前に学習することによって行われる。

【００６９】

いくつかの実施形態は、使用されるハードウェアに依存しない、すなわち、開放型及び密閉型ヘッドフォンを使用することができる。信号処理は、ヘッドフォンに内蔵されてもよいし、外部装置に内蔵されてもよいし、スマートフォンに内蔵されてもよい。任意で、音響的に記録され処理された信号の再生に加えて、スマートフォンから直接信号を再生してもよい（例えば、音楽、電話）。

【００７０】

他の実施形態では、「ＡＩ支援による選択的受聴」のためのエコシステムが提供される。実施形態は、「パーソナライズされた聴覚現実」（ｐｅｒｓｏｎａｌｉｚｅｄ　ａｕｄｉｔｏｒｙ　ｒｅａｌｉｔｙ：ＰＡＲｔｙ）に言及する。このようなパーソナライズされた環境では、リスナーは、定義された音響オブジェクトを増幅、低減、または修正することが可能である。個人の要求に合わせた音体験を作り出すために、一連の分析および合成プロセスが実行される。ターゲット変換フェーズの研究作業は、このための不可欠な要素を形成している。

【００７１】

いくつかの実施形態は、実際の音環境の分析と個々の音響オブジェクトの検出、利用可

能なオブジェクトの分離、追跡、および編集可能性、ならびに修正済み音響シーンの再構成および再現を実現する。

【００７２】

実施形態では、音イベントの検出、音イベントの分離、及びいくつかの音イベントの抑制が実現される。

【００７３】

実施形態では、ＡＩ手法（特に、ディープラーニングベースの手法）が使用される。

【００７４】

本発明の実施形態は、空間オーディオの記録、信号処理、および再生のための技術開発に寄与する。

【００７５】

例えば、実施形態は、対話するユーザを有するマルチメディアシステムにおける空間性及び三次元性を生成する。

【００７６】

この場合、実施形態は、空間的な聴覚／リスニングの知覚的および認知的なプロセスに関する研究知識に基づいている。

【００７７】

いくつかの実施形態では、以下の概念のうちの２つ以上を使用する。

【００７８】

シーン分解：実環境の空間音響検出（ｓｐａｔｉａｌ－ａｃｏｕｓｔｉｃａｌ ｄｅｔｅｃｔｉｏｎ）と、パラメータ推定（ｐａｒａｍｅｔｅｒ ｅｓｔｉｍａｔｉｏｎ）および／または位置依存音場解析（ｐｏｓｉｔｉｏｎ－ｄｅｐｅｎｄｅｎｔ ｓｏｕｎｄ ｆｉｅｌｄ ａｎａｌｙｓｉｓ）が含まれる。

【００７９】

シーン表現：これは、オブジェクトおよび／または環境の表現と識別、および／または効率的な表現と保存を含む。

【００８０】

シーンの組み合わせと再生：これは、オブジェクトと環境の適合と変化、および／またはレンダリングおよび可聴化を含む。

【００８１】

品質評価：これは、技術的および／または聴覚的な品質測定を含む。

【００８２】

マイクロフォンの位置決め：マイクロフォンアレイや適切な音声信号処理の適用を含む。

【００８３】

信号のコンディショニング：特徴抽出およびＭＬ（機械学習）用データセット生成も含まれる。

【００８４】

室内音響、環境音響の推定：これは、音源分離やＭＬに必要な、室内音響パラメータの計測および推定および／または室内音響特性の提供を含む。

【００８５】

可聴化：これは、環境への聴覚的適応を伴う空間的なオーディオ再生、および／または、検証および評価、および／または、機能証明および品質推定を含む。

【００８６】

図８は、一実施形態による対応するシナリオを示す図である。

【００８７】

実施形態では、音源の検出、分類、分離、位置特定、および強調のための概念を組み合わせており、各分野における最近の進歩が強調され、それらの間の結合が示されている。

【００８８】

以下では、実生活におけるＳＨに必要な柔軟性と堅牢性を提供するように、音源の結合／検出／分類／位置特定および分離／強調を行うことができる首尾一貫した概念を提供す

る。
【００８９】
さらに、実施形態は、実生活における聴覚シーンのダイナミクスを扱う際に、リアルタイム性能に適した低レイテンシーを有する概念を提供する。
【００９０】
いくつかの実施形態は、ディープラーニング、機械リスニング、及びスマートヘッドフォン（スマートヒアラブル）の概念を使用し、リスナーが聴覚シーンを選択的に変更することを可能にする。
【００９１】
実施形態では、ヘッドフォンやイヤホンなどの聴覚装置によって、聴覚シーンにおける音源を選択的に強化、減衰、抑制、または変更する可能性をリスナーに提供する。
【００９２】
図９は、４つの外部音源を有する実施形態によるシナリオを示す。
【００９３】
図９では、ユーザが聴覚シーンの中心である。この場合、４つの外部音源（Ｓ１～Ｓ４）がユーザの周囲で活動している。ユーザインタフェースにより、リスナーは聴覚シーンに影響を与えることができる。音源Ｓ１－Ｓ４は、対応するスライダを用いて、減衰、改善、抑制することができる。図１に見られるように、リスナーは、聴覚シーンに保持すべきまたは抑制すべき音源または音イベントを定義することができる。図１では、都市の背景音は抑制し、警報や電話の着信音は保持する。ユーザは常に、聴覚装置を通して音楽やラジオなどの追加オーディオストリームを再生することができる。
【００９４】
ユーザは通常、システムの中心となり、コントロールユニットによって聴覚シーンを制御する。ユーザは、図９に示すようなユーザインタフェース、または音声コントロール、ジェスチャー、視線方向など、あらゆるタイプのインタラクションを用いて聴覚シーンを変更することが可能である。ユーザがシステムにフィードバックを与えると、次のステップは検出／分類／位置特定の段階からなる。場合によっては、検出のみが必要なこともある。例えば、ユーザが聴覚シーンで発生するあらゆる音声を維持したい場合などである。例えば、火災報知器の音は聴き取りたいが、電話の着信音やオフィス内の騒音は聴き取りたくない場合など、分類が必要な場合もある。また、音源の位置のみがシステムにとって重要な場合もある。例えば、図９の４つの音源がそうである。ユーザは、音源の種類や特性に関係なく、ある方向から来る音源を除去するか、減衰させるかを決定することができる。
【００９５】
図１０は、実施形態に係るＳＨアプリケーションの処理ワークフローを示す図である。
【００９６】
まず、図１０の分離／強調段階において、聴覚シーンが修正される。これは、特定の音源（例えば、特定の音源群）を抑制するか、減衰させるか、または増強するかのいずれかによって行われる。図１０に示すように、ＳＨにおける追加の処理代替は、ノイズ制御であり、聴覚シーンにおける背景ノイズを除去するか、または最小化する目標を有する。騒音制御のための最も一般的で広く普及している技術は、おそらくアクティブ・ノイズ・コントロール（ａｃｔｉｖｅ　ｎｏｉｓｅ　ｃｏｎｔｒｏｌ：ＡＮＣ）であろう［１１］。
【００９７】
選択的受聴は、仮想音源をシーンに追加することなく、実際の音源のみが聴覚シーンにおいて変更されるアプリケーションに選択的受聴を限定することで、仮想および拡張聴覚環境と区別される。
【００９８】
機械リスニングの観点から、選択的受聴のアプリケーションには、音源の自動検出、位置特定、分類、分離、強調の技術が必要である。選択的受聴にまつわる用語をさらに明確にするために、以下の用語を定義し、それらの違いと関係を強調する。

【００９９】
実施形態では、例えば、音源位置特定が使用され、これは聴覚シーンにおける音源の位置を検出する能力を指す。音声処理のコンテキストでは、音源位置は通常、与えられた音源の到来方向（ｄｉｒｅｃｔｉｏｎ　ｏｆ　ａｒｒｉｖａｌ：ＤＯＡ）を指し、これは、２次元座標（方位角）またはそれが仰角を含む場合には３次元座標として与えられることができる。また、音源からマイクロフォンまでの距離を位置情報として推定するシステムもある［３］。音楽処理のコンテキストでは、位置はしばしば、最終的な混合物における音源のパンニング（ｐａｎｎｉｎｇ）を意味し、通常、度単位の角度として与えられる［４］。

【０１００】
実施形態によれば、例えば、音源検出が使用され、これは所与の音源タイプの任意のインスタンスが聴覚シーンに存在するかどうかを決定する能力を指す。検出タスクの一例は、任意のスピーカーがシーンに存在するかどうかを決定することである。この場合、シーン内のスピーカーの数やスピーカーの識別子を決定することは、音源検出の範囲外である。検出は、クラスが”音源存在”と”音源不在”に対応する２値分類タスクとして理解することができる。

【０１０１】
実施形態では、例えば、音源分類が使用され、これは、所定の音源または所定の音イベントに所定のクラスのセットからクラスラベルを割り当てる。分類タスクの例としては、与えられた音源が音声、音楽、または環境ノイズのいずれに対応するかを決定することが挙げられる。音源の分類と検出は、密接に関連した概念である。分類システムでは、「クラスなし」を可能なラベルの１つとして考慮することによって、検出段階を含む場合がある。このような場合、システムは暗黙のうちに音源の有無を検出することを学習し、どの音源もアクティブであるという十分な証拠がない場合にクラスラベルを割り当てることを余儀なくされることはない。

【０１０２】
実施形態によれば、例えば、音源分離が使用され、これは音声混合物または聴覚シーンから所定の音源を抽出することを指す。音源分離の例としては、歌手の他に他の楽器が同時に演奏されているオーディオ混合物から、歌声を抽出することが挙げられる［５］。音源分離は、選択的受聴のシナリオにおいて、リスナーにとって関心のない音源を抑制することができるため、重要な意味を持つ。音源分離システムのなかには、混合物から音源を抽出する前に、暗黙のうちに検出タスクをおこなうものがある。しかし、これは必ずしもルールではないため、これらのタスクの区別を強調する。また、分離は音源強調［６］や分類［７］などの他のタイプの分析の前処理段階として機能することが多い。

【０１０３】
実施形態では、例えば、さらに一歩進んで、オーディオ信号中の音源の特定のインスタンスを識別することを目的とする音源識別が使用される。スピーカ識別は、おそらく、今日、音源識別の最も一般的な用途である。このタスクの目的は、特定の話者がシーンに存在するかどうかを識別することである。図１の例では、ユーザは聴覚的なシーンで保持すべき音源の１つとして「話者Ｘ」を選択したことになる。このためには、音声の検出や分類を超える技術が必要であり、このような正確な識別を可能にする話者認識モデル（ｓｐｅａｋｅｒ－ｓｐｅｃｉｆｉｃ　ｍｏｄｅｌｓ）が求められている。

【０１０４】
実施形態によれば、例えば音源強調が使用され、これは聴覚シーンにおける所定の音源の顕著性を増加させるプロセスを指す［８］。音声信号の場合、その目標は、しばしば、その知覚的品質及び明瞭度を高めることである。音声強調の一般的なシナリオは、ノイズに汚染された音声のデノイズ（ｄｅ－ｎｏｉｓｉｎｇ）である［９］。音楽処理のコンテキストでは、音源強調はリミックスの概念に関連しており、ミックス内の１つの楽器（音源）をより際立たせるために実行されることがよくある。リミックスアプリケーションでは、音源分離フロントエンドを使って、個々の音源にアクセスし、ミックスの特性を変化

させることが多い［１０］。音源強調の前に音源分離のステージがある場合もあるが、そうでない場合もある。

【０１０５】

音源の検出、分類、および特定の分野において、例えば、実施形態のいくつかは、音響シーンおよびイベントの検出および分類のような、以下の概念のいずれかを使用する［１８］。このコンテキストにおいて、家庭環境におけるオーディオイベント検出（ａｕｄｉｏ ｅｖｅｎｔ ｄｅｔｅｃｔｉｏｎ：ＡＥＤ）のための方法が提案されており、そこでは、１０秒間の録音内で所定のサウンドイベントの時間境界を検出することが目標である［１９］、［２０］。この特定のケースでは、猫、犬、話し声、警報、流水など１０種類の音イベントが検討された。また、ポリフォニック（多音）な音イベント（複数のイベントが同時に発生する音）を検出する方法も文献において提案されている［２１］［２２］。［２１］では、双方向性長短期記憶（ｂｉ－ｄｉｒｅｃｔｉｏｎａｌ ｌｏｎｇ ｓｈｏｒｔ－ｔｅｒｍ ｍｅｍｏｒｙ：ＢＬＳＴＭ）リカレントニューラルネットワーク（ＲＮＮ）に基づくバイナリ活動検出器（ｂｉｎａｒｙ ａｃｔｉｖｉｔｙ ｄｅｔｅｃｔｏｒｓ）を用いて、実生活のコンテキストから合計６１の音イベントを検出する多声音イベント検出手法が提案されている。

【０１０６】

いくつかの実施形態は、例えば、弱くラベル付け（分類）されたデータに対処するために、分類のために信号の特定の領域に焦点を当てる時間的注意メカニズム（ｔｅｍｐｏｒａｌ ａｔｔｅｎｔｉｏｎ ｍｅｃｈａｎｉｓｍｓ）を組み込む［２３］。分類におけるノイズの多いラベルの問題は、クラスラベルが非常に多様であり、高品質の注釈が非常にコストがかかる選択的受聴アプリケーションに特に関連している［２４］。音イベントの分類タスクにおけるノイズの多いラベルは［２５］で扱われており、カテゴリのクロスエントロピーに基づくノイズに強い損失関数、および、ノイズの多いデータと手動でラベル付けされたデータの両方を評価する方法が提示されている。同様に、［２６］は畳み込みニューラルネットワーク（ＣＮＮ）に基づくオーディオイベント分類のためのシステムを提示しており、トレーニング例の複数のセグメントに対するＣＮＮの予測コンセンサスに基づくノイズラベルの検証ステップを組み込んでいる。

【０１０７】

例えば、いくつかの実施形態では、音イベントの検出と位置特定が同時に実現される。したがって、いくつかの実施形態は、［２７］のようなマルチラベル分類タスクとして検出を行い、位置は、各音イベントの到来方向（ＤＯＡ）の３－Ｄ座標として与えられる。

【０１０８】

いくつかの実施形態は、ＳＨのために音声区間検出（ｖｏｉｃｅ ａｃｔｉｖｉｔｙ ｄｅｔｅｃｔｉｏｎ）及び話者認識／識別の概念を使用する。音声区間検出は、デノイズオートエンコーダ［２８］、リカレントニューラルネットワーク［２９］、または生の波形を使用するエンドツーエンドシステム［３０］を使用して、ノイズ環境下で対処されてきた。話者認識アプリケーションについては、多くのシステムが文献で提案されており［３１］、その大部分は、例えばデータの増強や認識を容易にする改良された埋め込みによって、異なる条件に対する堅牢性を高めることに焦点を当てている［３２］～［３４］。したがって、いくつかの実施形態は、これらの概念を使用している。

【０１０９】

さらなる実施形態では、音イベント検出のために楽器の分類のための概念を使用する。モノフォニックとポリフォニックの両方における楽器の分類は、文献［３５］、［３６］で扱われている。［３５］では、３秒間のオーディオセグメントで優勢な楽器が１１の楽器クラスの間で分類され、いくつかの集計方法が提案されている。同様に、［３７］では、１秒という細かい時間分解能で楽器を検出することができる楽器区間検出（ｉｎｓｔｒｕｍｅｎｔ ａｃｔｉｖｉｔｙ ｄｅｔｅｃｔｉｏｎ）のための方法を提案している。歌声解析の分野でも多くの研究がなされている。特に、歌声がアクティブであるオーディオ録音のセグメントを検出するタスクのために、［３８］などの方法が提案されている。いく

つかの実施形態は、これらの概念を使用する。

【０１１０】

実施形態の中には、音源位置特定に関して以下で説明する概念のいずれかを用いるものがある。音源位置特定は、現実のアプリケーションでは聴覚シーンにおける音源の数が通常知られていないため、音源計数の問題と密接に関連している。シーン内の音源数が既知であることを前提に動作するシステムもある。例えば、アクティブ強度ベクトルのヒストグラムを使用して音源の位置を特定する［３９］のモデルがそうである。教師ありの観点から、［４０］は、入力表現として位相マップを使用して聴覚シーン内の複数のスピーカーのＤＯＡを推定するＣＮＮベースのアルゴリズムを提案している。一方、この文献では、シーン内の音源の数とその位置情報を共同で推定するものがいくつかある。これは［４１］で、ノイズや残響のある環境における複数話者の位置特定を行うシステムが提案されている。このシステムでは、複素数値の混合ガウスモデル（Ｇａｕｓｓｉａｎ Ｍｉｘｔｕｒｅ Ｍｏｄｅｌ：ＧＭＭ）を用いて音源の数とその位置を推定する。そこに記載された概念は、いくつかの実施形態によって使用される。

【０１１１】

音源位置特定アルゴリズムは、聴覚シーンの周囲の大きな空間をスキャンすることが多いため、計算量が多くなることがある［４２］。位置特定アルゴリズムにおける計算要求を低減するために、いくつかの実施形態は、クラスタリングアルゴリズム（ｃｌｕｓｔｅｒｉｎｇ ａｌｇｏｒｉｔｈｍｓ）［４３］を使用することによって、又は、被制御応答出力位相変化（ｓｔｅｅｒｅｄ ｒｅｓｐｏｎｓｅ ｐｏｗｅｒ ｐｈａｓｅ ｔｒａｎｓｆｏｒｍ：ＳＲＰ－ＰＨＡＴ）に基づくものなどの確立された方法について多重解像度探索（ｍｕｌｔｉ－ｒｅｓｏｌｕｔｉｏｎ ｓｅａｒｃｈｅｓ）［４２］を行うことによって、探索空間を低減する概念を使用する。また、スパース性（ｓｐａｒｓｉｔｙ）制約を課し、与えられた時間－周波数領域で１つの音源だけが優勢であることを仮定する方法もある［４４］。より最近では、生波形から直接方位を検出するエンド・ツー・エンドのシステムが［４５］で提案されている。これらの概念を利用した実施形態もある。

【０１１２】

いくつかの実施形態は、特に音声分離と音楽分離の分野から、音源分離（ＳＳＳ）のために後に説明された概念を使用する。

【０１１３】

特に、いくつかの実施形態は、話者非依存型分離の概念を使用する。分離は、シーン内の話者に関するいかなる事前情報なしにそこで実行される［４６］。また、いくつかの実施形態は、分離を実行するために、話者の空間的位置を評価する［４７］。

【０１１４】

選択的受聴のアプリケーションにおける計算性能の重要性を考えると、低遅延を達成することを具体的な目的として行われた研究は特に関連性が高い。利用可能な学習データがほとんどない状態で低遅延音声分離（１０ｍｓ未満）を行うための研究がいくつか提案されている［４８］。周波数領域でのフレーミング分析による遅延を避けるために、時間領域で適用するフィルタを注意深く設計することによって分離問題にアプローチするシステムもある［４９］。また、エンコーダ・デコーダのフレームワークを用いて時間領域の信号を直接モデル化することにより、低遅延の分離を実現するシステムもある［５０］。一方、周波数領域分離のアプローチでフレーミング遅延を低減しようとするシステムもある［５１］。これらの概念は、いくつかの実施形態によって採用されている。

【０１１５】

いくつかの実施形態では、リード楽器－伴奏分離のための概念［５２］など、オーディオ混合物から音楽ソースを抽出する音楽サウンド分離（music sound separation：ＭＳＳ）のための概念を用いる。これらのアルゴリズムは、クラスラベルに関係なく、混合物中の最も顕著な音源を取り出し、それを残りの伴奏から分離することを試みる。また、歌声の分離のための概念を用いる実施形態もある［５３］。ほとんどの場合、歌声の特徴を捉えるために、特定の音源モデル［５４］またはデータ駆動型モデル［５５］のいずれか

が使用される。［５５］で提案されたようなシステムは、分離を達成するために分類または検出段階を明示的に組み込んでいないにもかかわらず、これらのアプローチのデータ駆動型の性質により、これらのシステムは、分離前に一定の精度で歌声を検出することを暗黙的に学習することが可能である。音楽領域における別のクラスのアルゴリズムでは、分離前に音源の分類や検出を試みず、音源の位置のみを用いて分離を行うことを試みている［４］。

【０１１６】

実施形態の中には、アクティブノイズキャンセル（ＡＮＣ：Ａｃｔｉｖｅ Ｎｏｉｓｅ Ｃｏｎｔｒｏｌ）の概念を用いたものがある。ＡＮＣシステムは、ほとんどの場合、アンチノイズ信号を導入してそれを打ち消すことにより、ヘッドフォンユーザのバックグラウンドノイズを除去することを目的としている［１１］。ＡＮＣはＳＨの特殊なケースと考えることができ、同様に厳しい性能要件に直面している［１４］。自動車のキャビン［５６］や産業用シナリオ［５７］など、特定の環境におけるアクティブノイズ制御に焦点を当てた研究も行われている。［５６］では、ロードノイズやエンジンノイズなど異なるタイプのノイズのキャンセレーションを分析し、異なるタイプのノイズに対応できる統一的なノイズコントロールシステムの必要性を訴えている。また、特定の空間領域で騒音をキャンセルするＡＮＣシステムの開発に焦点を当てた研究もある。［５８］では、ノイズフィールドを表現するためのベース関数として球面調和（ｓｐｈｅｒｉｃａｌ ｈａｒｍｏｎｉｃｓ）を使用して、空間領域上のＡＮＣに取り組んでいる。いくつかの実施形態は、本明細書に記載される概念を使用する。

【０１１７】

実施形態の中には、音源強調のための概念を用いるものがある。

【０１１８】

音声強調のコンテキストでは、最も一般的なアプリケーションの１つは、ノイズによって破損された音声の強調である。多くの研究が、単一チャネル音声強調の位相処理に焦点を当てている［８］。ディープニューラルネットワークの観点からは、［５９］ではデノイズオートエンコーダで、［６０］ではディープニューラルネットワーク（ＤＮＮ）を用いてクリーンな音声とノイズのある音声との間の非線形回帰問題として、［６１］ではＧｅｎｅｒａｔｉｖｅ Ａｄｖｅｒｓａｒｉａｌ Ｎｅｔｗｏｒｋｓ（ＧＡＮｓ）を用いてエンドツーエンドシステムとして、音声ノイズ除去の問題を扱ってきた。多くの場合、音声協調は自動音声認識（ａｕｔｏｍａｔｉｃ ｓｐｅｅｃｈ ｒｅｃｏｇｎｉｔｉｏｎ：ＡＳＲ）システムのフロントエンドとして適用され、［６２］ではＬＳＴＭ ＲＮＮを用いて音声協調がアプローチされている。音声強調はまた、最初に音声を抽出し、次に分離された音声信号に強調技術を適用することである音源分離アプローチと関連してしばしば行われる［６］。本明細書に記載された概念は、実施形態のいくつかによって使用される。

【０１１９】

音楽に関連する音源強調は、多くの場合、音楽のリミックスを作成するためのアプリケーションを指す。音声の協調では、音声がノイズ源によってのみ破損していることが前提となっていることが多いのに対して、音楽アプリケーションでは、強調する音源と他の音源（楽器）が同時に演奏されていることが前提となっている場合がほとんどである。このため、音楽リミックスアプリケーションは、常に音源分離の段階を経て提供される。たとえば、［１０］では、初期のジャズ録音を、リード－伴奏と和声－打楽器の分離技術を適用してリミックスし、混合音におけるより良いサウンドバランスを達成した。同様に、［６３］では、歌声とバックトラックの相対的なラウドネスを変更するために、異なる歌声分離アルゴリズムを使用することを研究し、最終混合物にわずかだが聴き取れる歪みを導入することによって、６ｄＢ増加することが可能であることを示している。［６４］では、音源分離技術を応用して新しいミックスを実現し、人工内耳装用者の音楽知覚を向上させる方法を研究している。そこで述べられたコンセプトは、いくつかの実施形態で用いられている。

【０１２０】

選択的受聴アプリケーションの最大の課題の一つは、処理時間に関する厳しい要件に関連するものである。ユーザにとって自然で知覚的な品質を維持するためには、完全な処理ワークフローを最小限の遅延で実行する必要がある。システムの最大許容遅延時間は、アプリケーションと聴覚シーンの複雑さによって大きく異なる。たとえば、ＭｃＰｈｅｒｓｏｎらは、インタラクティブ音楽インタフェースの許容遅延の基準として１０ｍｓを提案している［１２］。ネットワーク経由の音楽演奏の場合、［１３］の著者らは、２０－２５ｍｓから５０－６０ｍｓの範囲で遅延が知覚できるようになることを報告している。しかし、アクティブノイズコントロール／キャンセル（ＡＮＣ）技術では、より良いパフォーマンスを得るために超低遅延処理が要求される。このようなシステムでは、許容可能な遅延時間は周波数と減衰に依存するが、２００Ｈｚ以下の周波数を約５ｄＢ減衰させると、１ｍｓまで低くすることができる［１４］。ＳＨアプリケーションで最後に考慮すべきは、変更された聴覚シーンの知覚品質である。さまざまなアプリケーションで音質を確実に評価するための方法論については、かなりの量の研究が行われている［１５］、［１６］、［１７］。しかしながら、ＳＨの課題は、処理の複雑さと知覚的品質との間の明確なトレードオフを管理することである。いくつかの実施形態は、そこに記載された概念を使用する。

【０１２１】

いくつかの実施形態は、［４１］に記載されているように計数／計算および位置特定についての概念、［２７］に記載されているように位置特定および検出についての概念、［６５］に記載されているように分離および分類についての概念、［６６］に記載されているように分離および計数についての概念を使用する。

【０１２２】

いくつかの実施形態は、［２５］、［２６］、［３２］、［３４］に記載されているように、現在の機械リスニング法の堅牢性を強化するための概念を使用し、ここで、新しい新たな方向性として、ドメイン適応［６７］及び複数の装置で記録されたデータセットでのトレーニング［６８］が含まれる。

【０１２３】

実施形態のいくつかは、［４８］に記載されているような機械リスニング法の計算効率を高めるための概念、または生の波形を扱うことができる［３０］、［４５］、［５０］、［６１］に記載されている概念を使用している。

【０１２４】

いくつかの実施形態は、シーン内の音源を選択的に修正できるようにするために、検出／分類／位置決めおよび分離／強調を組み合わせた統一最適化スキームを実現し、独立した検出、分離、位置決め、分類、および強調方法は信頼でき、ＳＨに必要な堅牢性および柔軟性を提供する。

【０１２５】

いくつかの実施形態は、アルゴリズムの複雑さと性能の間に良好なトレードオフが存在する、リアルタイム処理に適している。

【０１２６】

いくつかの実施形態は、ＡＮＣと機械リスニングを組み合わせている。例えば、聴覚シーンが最初に分類され、その後、ＡＮＣが選択的に適用される。

【０１２７】

さらなる実施形態は、以下に提供される。

【０１２８】

仮想オーディオオブジェクトで実際の聴覚環境を増強するためには、オーディオオブジェクトの各位置から部屋内のリスナーの各位置への伝達関数が十分に知られている必要がある。

【０１２９】

伝達関数は、音源の特性、オブジェクトとユーザとの間の直接的な音、室内で発生するすべての反射音をマッピングする。リスナーが実際にいる部屋の音響特性を正しい空間オ

ーディオ再生を保証するためには、伝達関数はさらに、リスナーの部屋の音響特性を十分な精度でマッピングする必要がある。

【０１３０】

室内の異なる位置にある個々のオーディオオブジェクトの表現に適したオーディオシステムでは、多数のオーディオオブジェクトが存在する場合、個々のオーディオオブジェクトを適切に検出し分離することが課題となっている。また、部屋の録音位置や聴取位置において、オブジェクトの音声信号が重なることがある。また、部屋のオブジェクトやリスニング位置が変わると、部屋の音響やオーディオ信号の重なりが変化する。

【０１３１】

相対的な移動に伴い、室内音響のパラメータを十分に高速に推定する必要がある。ここでは、高精度よりも低遅延での推定が重要である。また、音源と受信機の位置が変わらない場合（静的な場合）には、高い精度が要求される。提案システムでは、オーディオ信号のストリームから室内音響パラメータ、室内形状、リスナー位置を推定（または抽出）する。オーディオ信号は、音源と受信機が任意の方向に移動でき、かつ音源及び／又は受信機の向きを任意に変えられる実環境で記録される。

【０１３２】

オーディオ信号ストリームは、１つまたは複数のマイクロフォンを含む任意のマイクロフォンセットアップの結果であってもよい。そのストリームは、前処理および／またはさらなる分析のために、信号処理ステージに供給される。その後、出力は特徴抽出ステージに送られる。この段階では、室内音響パラメータ、例えばＴ６０（残響時間）、ＤＲＲ（直接残響比（Ｄｉｒｅｃｔ－ｔｏ－Ｒｅｖｅｒｂｅｒａｎｔ　Ｒａｔｉｏ））、その他を推定する。

【０１３３】

２つ目のデータストリームは、６ＤｏＦセンサー（「６つの自由度」：部屋の中の位置と視線方向のそれぞれを３次元で捉える）により、マイクロフォンセットアップの向きと位置を捉えて生成される。この位置データストリームは、６ＤｏＦ信号処理ステージに送られ、前処理やさらなる分析が行われる。

【０１３４】

６ＤｏＦ信号処理、オーディオ特徴抽出ステージ、および前処理されたマイクロフォンストリームの出力は、機械学習ブロックに送られ、聴覚空間またはリスニングルーム（サイズ、形状、反射面）および部屋でのマイクロフォンフィールドの位置が推定される。さらに、よりロバストな推定を可能にするために、ユーザ行動モデルを適用する。このモデルでは、人間の動作の制限（例えば連続動作、速度など）や、異なる種類の動作の確率分布を考慮する。

【０１３５】

実施形態の中には、任意のマイクロフォン設備（配置）を用い、かつ、ユーザの位置・姿勢情報を付加することで、室内音響パラメータのブラインド推定を実現し、さらに機械学習手法によるデータ解析を行うものもある。

【０１３６】

例えば、実施形態によるシステムは、音響的に拡張された現実（ａｃｏｕｓｔｉｃａｌｌｙ　ａｕｇｍｅｎｔｅｄ　ｒｅａｌｉｔｙ：ＡＡＲ）に使用されてもよい。この場合、推定されたパラメータから仮想の部屋のインパルス応答を合成する必要がある。

【０１３７】

いくつかの実施形態は、記録された信号から残響を除去することを含む。このような実施形態の例としては、健聴者のための補聴器や聴覚障害者のための補聴器がある。この場合、推定されたパラメータの助けを借りて、マイクロフォンセットアップの入力信号から残響が除去され得る。

【０１３８】

さらなる応用として、現在の聴覚空間とは別の部屋で生成されたオーディオシーンの空間合成がある。この目的のために、オーディオシーンの一部である室内音響パラメータは

、聴覚空間の室内音響パラメータに関して適応される。
【０１３９】
両耳合成の場合、この目的のために、利用可能なＢＲＩＲは、聴覚空間の異なる音響学パラメータに適合される。
【０１４０】
一実施形態では、１以上の室内音響パラメータを決定するための装置が提供される。
【０１４１】
本装置は、１つ以上のマイクロフォン信号を含むマイクロフォンデータを取得するように構成される。
【０１４２】
さらに、本装置は、ユーザの位置及び／又は向きに関する追跡データを取得するように構成される。
【０１４３】
さらに、本装置は、マイクロフォンデータに依存して、かつ追跡データに依存して、１つ以上の室内音響パラメータを決定するように構成される。
【０１４４】
実施形態によれば、例えば、装置は、マイクロフォンデータに依存して、および追跡データに依存して、１つ以上の室内音響パラメータを決定するために機械学習を採用するように構成されてもよい。
【０１４５】
実施形態において、例えば、装置は、ニューラルネットワークを採用するように構成されてもよいという点で、機械学習を採用するように構成されてもよい。
【０１４６】
実施形態によれば、例えば、装置は、機械学習のためにクラウドベースの処理を採用するように構成されてもよい。
【０１４７】
実施形態において、例えば、１つ以上の室内音響パラメータは、残響時間を含んでもよい。
【０１４８】
実施形態によれば、例えば、１つ以上の室内音響パラメータは、直接－残響比（ｄｉｒｅｃｔ－ｔｏ－ｒｅｖｅｒｂｅｒａｎｔ　ｒａｔｉｏ）を含んでもよい。
【０１４９】
実施形態において、例えば、追跡データは、ユーザの位置をラベル付けするためのｘ座標、ｙ座標、及びｚ座標を含んでもよい。
【０１５０】
実施形態によれば、例えば、追跡データは、ユーザの向きをラベル付けするために、ピッチ座標、ヨー座標、及びロール座標を含んでもよい。
【０１５１】
実施形態において、例えば、装置は、１つ以上のマイクロフォン信号を時間領域から周波数領域に変換するように構成されてもよく、例えば、装置は、周波数領域における１つ以上のマイクロフォン信号の１つ以上の特徴を抽出するように構成されてもよく、例えば、装置は、１つ以上の特徴に依存して１つ以上の室内音響パラメータを決定するように構成されてもよい。
【０１５２】
実施形態によれば、例えば、装置は、１つ以上の特徴を抽出するためにクラウドベースの処理を採用するように構成されてもよい。
【０１５３】
実施形態において、例えば、装置は、複数のマイクロフォン信号を記録するための複数のマイクロフォンのマイクロフォン設備を含んでもよい。
【０１５４】

実施形態によれば、例えば、マイクロフォン設備は、ユーザの身体に装着されるように構成されてもよい。

【０１５５】

実施形態において、例えば、図1の上述のシステムは、1つまたは複数の室内音響パラメータを決定するための上述の装置をさらに含んでもよい。

【０１５６】

実施形態によれば、例えば、信号部分修正器１４０は、１つ以上の室内音響パラメータの少なくとも１つに応じて、１つ以上のオーディオソースの少なくとも１つのオーディオソースのオーディオソース信号部分の変動を実行するように構成されてもよく；および／または、信号生成器１５０は、１つ以上の室内音響パラメータの少なくとも１つに応じて、１つ以上のオーディオソースの各オーディオソースに対する複数の両耳室内インパルス応答の少なくとも１つを発生することを実行するように構成されてもよい。

【０１５７】

図７は、５つのサブシステム（サブシステム１～５）を含む、一実施形態によるシステムを示す。

【０１５８】

サブシステム１は、１つ、２つ、または複数の個別のマイクロフォンのマイクロフォンセットアップを含み、１つ以上のマイクロフォンが利用可能な場合は、マイクロフォンフィールドに結合されることがある。マイクロフォン／複数のマイクロフォンの互いに対する位置決め及び相対的配置は任意であってよい。マイクロフォンの設備は、ユーザが装着する装置の一部であってもよいし、関心のある部屋に配置された別個の装置であってもよい。

【０１５９】

さらに、サブシステム１は、ユーザの位置及び／又は向きに関する追跡データを取得するためのトラッキング装置を含む。例えば、ユーザの位置及び／又は向きに関する追跡データは、ユーザの移動位置や室内におけるユーザの頭部姿勢を決定するために使用される場合がある。最大６ＤｏＦ（６自由度、例えばｘ座標、ｙ座標、ｚ座標、ピッチ角、ヨー角、ロール角）が測定される場合がある。

【０１６０】

この場合、例えば、トラッキング装置が追跡データを測定するように構成されていてもよい。トラッキング装置は、ユーザの頭部に配置してもよいし、複数のサブ装置に分割して必要なＤｏＦを測定してもよいし、ユーザに載せても載せなくてもよい。

【０１６１】

このように、サブシステム１は、マイクロフォン信号入力インタフェース１０１と位置情報入力インタフェース１０２とを含む入力インタフェースを表している。

【０１６２】

サブシステム２は、録音されたマイクロフォン信号の信号処理を行う。これは、周波数変換及び／又は時間領域ベースの処理を含む。さらに、これは、フィールド処理を実現するために、異なるマイクロフォン信号を組み合わせる方法も含まれる。システム４からのフィードバックは、サブシステム２における信号処理のパラメータを適応させるために可能である。マイクロフォン信号（複数可）の信号処理ブロックは、マイクロフォン（複数可）が内蔵されている装置の一部であってもよいし、別装置の一部であってもよい。また、クラウドベースの処理の一部であってもよい。

【０１６３】

さらに、サブシステム２は、記録された追跡データに対する信号処理を含む。これは、周波数変換および／または時間領域ベースの処理を含む。さらに、ノイズ抑制、平滑化、内挿（補間法）、外挿（補外法）を採用することにより、信号の技術的品質を高める方法を含む。さらに、より高いレベルの情報を導き出すための方法も含まれる。これは、速度、加速度、経路方向、アイドル時間、移動範囲、移動経路などを含む。さらに、これは、近未来の移動経路の予測、近未来の速度の予測などを含む。トラッキング信号の信号処理

ブロックは、トラッキング装置の一部であってもよいし、別の装置の一部であってもよい。また、クラウドベースの処理の一部であってもよい。

【０１６４】

サブシステム３は、処理されたマイクロフォン（複数）の特徴を抽出することを含む。

【０１６５】

特徴抽出ブロックは、ユーザのウェアラブル装置の一部であってもよいし、別個の装置の一部であってもよい。また、クラウドベースの処理の一部であってもよい。

【０１６６】

サブシステム２及び３は、例えば、検出器１１０、オーディオタイプ分類器１３０、及び信号部分修正器１４０を、それらのモジュール１１１及び１２１と共に実現する。例えば、サブシステム３、モジュール１２１は、オーディオ分類の結果をサブシステム２、モジュール１１１に出力してもよい（フィードバック）。例えば、サブシステム２、モジュール１１２は、位置決定器１２０を実現する。さらに、実施形態では、サブシステム２、３は、例えば、サブシステム２、モジュール１１１が両耳室内インパルス応答及びラウドスピーカー信号を生成することによって、信号生成器１５０を実現することもできる。

【０１６７】

サブシステム４は、処理されたマイクロフォン信号（複数可）、マイクロフォン信号の抽出された特徴、および処理された追跡データを使用して、室内音響パラメータを推定する方法とアルゴリズムを含んでいる。このブロックの出力は、アイドルデータとしての室内音響パラメータ、およびサブシステム２におけるマイクロフォン信号処理のパラメータの制御と変動である。機械学習ブロック１３１は、ユーザの装置の一部であってもよいし、別の装置の一部であってもよい。また、クラウドベースの処理の一部であってもよい。

【０１６８】

さらに、サブシステム４は、室内音響アイドルデータパラメータの後処理を含む（例えば、ブロック１３２において）。これは、外れ値の検出、新しいパラメータへの個々のパラメータの組み合わせ、平滑化、外挿（補外法）、内挿（補間法）、および妥当性検証を含む。このブロックはまた、サブシステム２からの情報を取得する。これは、近未来の音響パラメータを推定するために、部屋におけるユーザの近未来の位置を含む。このブロックは、ユーザの装置の一部であってもよいし、別の装置の一部であってもよい。また、クラウドベースの処理の一部であってもよい。

【０１６９】

サブシステム５は、（例えば、メモリ１４１における）下流システムのための室内音響パラメータの記憶と割り当てを含む。パラメータの割り当ては、ジャストインタイムで実現されてもよく、及び／又は、時間応答が記憶されてもよい。保存は、ユーザ上またはユーザの近くに位置する装置で実行されてもよく、またはクラウドベースのシステムで実行されてもよい。

【０１７０】

以下、本発明の実施形態に係るユースケースを説明する。

【０１７１】

実施形態のユースケースは、ホームエンターテイメントであり、家庭環境におけるユーザに関するものである。例えば、ユーザは、テレビ、ラジオ、ＰＣ、タブレットなどの特定の再生装置に集中することを希望し、他の妨害源（他のユーザの装置、又は子供、工事騒音、街頭騒音）を抑制することを希望する。この場合、ユーザは好みの再生装置の近くに位置し、その装置、またはその位置を選択する。ユーザの位置に関係なく、ユーザが選択を取り消すまで、選択された装置、または音源の位置が音響的に強調される。

【０１７２】

例えば、ユーザは目的の音源の近くに移動する。ユーザは適切なインタフェースを介してターゲット音源を選択し、それに応じてヒアラブルは、ユーザ位置、ユーザの視線方向、およびターゲット音源に基づいて、妨害ノイズがある場合でもターゲット音源をよく理解できるようにオーディオ再生を適合させる。

【０１７３】
また、特に邪魔な音源の近くにユーザが移動する。ユーザは適切なインタフェースを介してこの妨害音源を選択し、ヒアラブル（聴覚装置）はそれに応じて、妨害音源を明示的にチューニングするように、ユーザ位置、ユーザ視線方向、および妨害音源に基づいてオーディオ再生を適合させる。
【０１７４】
さらなる実施形態の使用例は、ユーザが複数のスピーカーの間に位置するカクテルパーティーである。
【０１７５】
多くのスピーカーが存在する場合、例えば、ユーザは１人（または数人）のスピーカーに集中し、他の妨害源をチューニングしたい、または減衰させたい。このような場合、ヒアラブルの制御は、ユーザとの対話をほとんど必要としないはずである。生体信号や、会話の困難さ（頻繁な質問、外国語、強い方言）を示す検出可能な指標に基づく選択性の強さの制御は、オプションとなる。
【０１７６】
例えば、スピーカーはランダムに配置され、リスナーと相対的に移動する。また、定期的に会話が途切れたり、新しいスピーカーが加わったり、他のスピーカーがその場を離れたりする。可能性としては、音楽のような妨害音は比較的に大きな音である。選択されたスピーカーは音響的に強調され、発言の一時停止や位置または姿勢の変化の後に再び認識される。
【０１７７】
例えば、ヒアラブルは、ユーザの近くにあるスピーカーを認識する。適切な制御可能性（例えば、視聴方向、注意制御）を通じて、ユーザは、好ましいスピーカを選択することができる。ヒアラブルは、ユーザの視線方向及び選択されたターゲット音源に応じてオーディオ再生を適応させ、妨害ノイズがある場合でもターゲット音源をよく理解することができるようにする。
【０１７８】
あるいは、（従来）好ましくないスピーカから直接話しかけられた場合、自然なコミュニケーションを確保するために、少なくとも聞き取れるようにする必要がある。
【０１７９】
別の実施形態の別の使用例は、自動車におけるものであり、ユーザは自分の（あるいは１つの）自動車内に位置している。運転中、ユーザは、邪魔な騒音（風、モーター、乗客）の隣でそれらをよりよく理解できるように、ナビゲーション装置、ラジオ、または会話相手などの特定の再生装置に自分の音響的注意を積極的に向けたいと考えている。
【０１８０】
例えば、ユーザと対象音源は自動車内の一定の位置に配置されている。ユーザは基準システムに対して静止しているが、車両自体は動いている。そのため、適応したトラッキングソリューションが必要となる。選択された音源の位置は、ユーザが選択を取り消すか、警告信号によって装置の機能が停止するまで、音響的に強調される。
【０１８１】
例えば、ユーザが自動車に乗り込むと、周囲の状況が装置によって検出される。適切な制御可能性（例えば速度認識）を通じて、ユーザはターゲット音源を切り替えることができ、ヒアラブルはユーザの視線方向と選択されたターゲット音源に応じてオーディオ再生を適応させ、外乱音の場合でもターゲット音源をよく理解することができるようにする。
【０１８２】
また、例えば、交通関連の警告信号により、通常の流れが中断され、ユーザの選択がキャンセルされる。その後、通常のフローの再スタートが実行される。
【０１８３】
さらなる実施形態の別の使用例は、ライブ音楽であり、ライブ音楽イベントのゲストに関するものである。例えば、コンサート又はライブ音楽演奏のゲストは、ヒアラブルの助

けを借りて演奏への集中力を高めることを望み、邪魔な行動をする他のゲストをチューニングすることを望む。さらに、オーディオ信号自体を最適化することも可能で、例えば、不利なリスニング位置や部屋の音響のバランスをとることができる。

【０１８４】

例えば、ユーザは多くの妨害源の間に位置しているが、ほとんどの場合、演奏は比較的大きな音量である。ターゲット音源は固定された位置か、少なくとも定義された領域に配置されているが、ユーザは非常に移動しやすいかもしれない（例えば、ユーザはダンスをしているかもしれない）。選択された音源位置は、ユーザが選択を取り消すまで、あるいは警告信号によって装置の機能が停止されるまで、音響的に強調される。

【０１８５】

例えば、ユーザはステージエリアやミュージシャン（複数可）をターゲット音源として選択する。適切な制御可能性を通じて、ユーザはステージ／ミュージシャンの位置を定義することができ、ヒアラブルは、ユーザの視線方向及び選択されたターゲット音源に従ってオーディオ再生を適応させ、妨害ノイズの場合にもターゲット音源をよく理解することができるようにすることができる。

【０１８６】

あるいは、例えば、警告情報（例えば、避難、野外イベントの場合の雷雨の接近など）及び警告信号により、通常のフローが中断され、ユーザの選択がキャンセルされることもある。その後、通常のフローの再開がある。

【０１８７】

別の実施形態のさらなる使用例は、主要なイベント、および主要なイベントでのゲストへの関心である。したがって、主要なイベント（例えば、フットボールスタジアム、アイスホッケースタジアム、大型コンサートホールなど）において、ヒアラブルを使用して、さもなければ群衆の騒音にかき消されるであろう家族メンバーおよび友人の声を強調することが可能である。

【０１８８】

例えば、多くの参加者が集まる大きなイベントが、スタジアムや大きなコンサートホールで行われるとする。あるグループ（家族、友人、学校のクラス）がそのイベントに参加し、大勢の人が歩き回るイベント会場の外や中にいる。一人または数人の子供がグループとのアイコンタクトを失い、騒音による高いノイズレベルにもかかわらず、グループを呼び出す。その後、ユーザは音声認識をオフにし、ヒアラブルはもはや音声（複数可）を増幅しなくなる。

【０１８９】

例えば、グループの一人がヒアラブルで行方不明の子供の声を選択する。ヒアラブルはその音声の位置を特定する。そして、ヒアラブルが音声を増幅し、ユーザは増幅された音声に基づいて行方不明の子供を（より迅速に）回復させることができる。

【０１９０】

あるいは、例えば、行方不明の子供もヒアラブルを装着し、両親の音声を選択する。ヒアラブルは、両親の音声（複数可）を増幅する。増幅された音声によって、子供は両親の居場所を特定することができる。こうして、子供は親の元へ歩いて帰ることができる。あるいは、例えば、行方不明の子供もヒアラブルを装着し、両親の声を選択する。ヒアラブルは両親の声を探し出し、ヒアラブルはその声までの距離をアナウンスする。このようにして、子供は両親をより簡単に見つけることができる。オプションとして、距離のアナウンスのために、ヒアラブルからの人工音声の再生が提供されてもよい。

【０１９１】

例えば、音声（複数可）の選択的増幅のためのヒアラブルのカップリングが提供され、音声プロファイルが保存される。

【０１９２】

さらなる実施形態の使用例は、レクリエーションスポーツであり、レクリエーションアスリートに関するものである。スポーツ中に音楽を聴くことは人気があるが、しかし、そ

れはまた危険を伴う。警告信号や他の道路使用者の声が聞こえないかもしれない。音楽の再生に加えて、ヒアラブルは、警告信号または叫び声に反応し、音楽の再生を一時的に中断させることができる。さら使用場面は、少人数で行うスポーツの場合である。スポーツグループのヒアラブルを接続することで、スポーツ中に他の騒音を抑制しながら良好なコミュニケーションを確保することができる。

【０１９３】

例えば、ユーザは移動しており、可能性のある警告信号は多くの妨害源と重複している。警告信号のすべてがユーザに関係しているわけではないことは問題である（街中の遠隔サイレン、路上でのクラクション）。そこで、ヒアラブルは、ユーザが選択を解除するまで、自動的に音楽再生を停止し、通信相手の警告信号を音響的に強調する。その後、音楽は通常通り再生される。

【０１９４】

例えば、ユーザがスポーツをしながら、ヒアラブルを介して音楽を聴いているとする。ユーザに関する警告信号や叫び声が自動的に検出され、ヒアラブルは音楽の再生を中断させる。ヒアラブルは、対象となる音源／音響環境をよく理解できるように、オーディオ再生を適応させる。その後、ヒアラブルは自動的に（例えば警告信号の終了後に）、あるいはユーザの要求に応じて、音楽の再生を継続する。

【０１９５】

また、例えば、グループのアスリートがヒアラブルを接続してもよい。グループのメンバー間の音声の理解度は最適化され、他の邪魔なノイズは抑制される。

【０１９６】

別の実施形態の別の使用例は、いびきの抑制であり、いびきによって妨害される睡眠を望むすべての人々に関係する。パートナーがいびきをかく人は、夜間の休息が妨げられ、睡眠に問題がある。ヒアラブルは、いびき音を抑制し、夜の休息を確保し、家庭の平和を提供するため、安息を提供する。同時に、ヒアラブルは、他の音（赤ちゃんの泣き声、アラーム音など）を通すので、外界から音響的に完全に隔離されるわけではない。例えば、いびきの検出が可能である。

【０１９７】

例えば、ユーザはいびき音によって睡眠障害を起こしている。そのとき、ヒアラブルを使用することで、ユーザは再びよく眠れるようになり、ストレス軽減の効果が期待できる。

【０１９８】

例えば、ユーザは、睡眠時にヒアラブルを装着する。ヒアラブルをスリープモードに切り替えると、すべてのいびき音が抑制される。睡眠後、再びヒアラブルの電源を切る。

【０１９９】

あるいは、工事音、芝刈り機の音など、他の音も睡眠中に抑制することができる。

【０２００】

さらなる実施形態のユースケースは、日常生活におけるユーザのための診断装置である。ヒアラブルは、嗜好（例えば、どの音源を選択したか、どの減衰／増幅を選択したか）を記録し、使用期間を介して傾向のあるプロファイルを作成する。このデータにより、聴覚能力に関する変化について結論を導き出すことができるかもしれない。この目的は、聴覚の喪失をできるだけ早期に発見することである。

【０２０１】

例えば、ユーザが日常生活で持ち歩く、あるいは前述のユースケースで数カ月または数年にわたり持ち歩く。ヒアラブルは、選択された設定に基づき分析を行い、警告や推奨事項をユーザに出力する。

【０２０２】

例えば、ユーザが長期間（数ヶ月から数年）にわたってヒアラブルを装着する。装置は、聴覚の好みに基づいて分析を作成し、装置は、聴覚の喪失を発症した場合に、推奨および警告を出力する。

【０２０３】

別の実施形態のさらなる使用例は、治療装置であり、日常生活における聴覚障害を持つユーザに関するものである。聴覚装置への移行装置としての役割において、潜在的な患者はできるだけ早期に支援され、したがって、認知症は予防的に治療される。他の可能性としては、集中力トレーナーとしての使用（例えば、ＡＤＨＳのための）、耳鳴りの治療、およびストレス軽減が挙げられる。

【０２０４】

例えば、リスナーは聴覚の問題や注意欠陥があり、ヒアラブルをヒアリング装置として一時的／暫定的に使用する。聴覚の問題に応じて、例えば、すべての信号の増幅（難聴）、好みの音源に対する高い選択性（注意欠陥）、治療音の再生（耳鳴りの治療）などにより、ヒアラブルにより軽減することができる。

【０２０５】

ユーザは独自に、または医師の助言により、治療の形態を選択し、好みの調整を行い、ヒアラブルは選択された治療を実行する。

【０２０６】

また、ヒアラブルは、ＵＣ－ＰＲＯ１から聴覚の問題を検出し、検出された問題に基づいて、ヒアラブルが自動的に再生を調整し、ユーザに通知する。

【０２０７】

さらなる実施形態の使用例は、公共部門における仕事であり、公共部門の従業員に関するものである。仕事中に高いレベルの騒音にさらされる公共部門の従業員（病院、小児科医、空港カウンター、教育者、レストラン産業、サービスカウンターなど）は、より良いコミュニケーションのため１人または数人のみの発話を強調し、および例えばストレスの軽減を通じて、仕事中のより良い安全のためにヒアラブルを着用する。

【０２０８】

例えば、従業員は職場環境で高いレベルの騒音にさらされ、バックグラウンドノイズにもかかわらず、落ち着いた環境に切り替えることができないまま、クライアントや患者、同僚と話さなければならない。病院の従業員は、医療機器の音やビープ音（またはその他の業務に関連する騒音）を通して高いレベルの騒音にさらされており、それでも患者や同僚とコミュニケーションを取らなければならない。小児科医や教育者は、子供たちの騒音や叫び声の中で働いており、親と話ができなければならない。空港のカウンターでは、コンコース内の騒音が大きい場合、従業員は航空会社の乗客の声を理解することが困難である。来客の多いレストランでは、ウェイターが客の注文を聞き取るのが難しい。そこで、例えば、ユーザが音声選択をオフにすると、ヒアラブルは音声を増幅しなくなる。

【０２０９】

例えば、人は装着されたヒアラブルの電源を入れる。ユーザは、ヒアラブルを近くの声の音声選択に設定し、ヒアラブルは近くの声、または近くのいくつかの声を増幅し、同時にバックグラウンドノイズを抑える。これにより、ユーザは関連する音声をより理解することができる。

【０２１０】

あるいは、人はヒアラブルを連続的なノイズ抑制に設定する。ユーザは、利用可能な音声を検出する機能をオンにして、同じ音声を増幅させる。このように、ユーザはより低いレベルのノイズで作業を続けることができる。Ｘメートル離れた場所から直接話しかけられた場合、ヒアラブルはその音声を増幅する。これにより、ユーザは、低騒音で相手と会話することができる。会話終了後、ヒアラブルは、ノイズ抑制モードに元に戻るように切り替わり、作業終了後、ユーザはヒアラブルを再び電源オフにする。

【０２１１】

別の実施形態の別の使用例は、乗客の輸送であり、乗客の輸送のための自動車におけるユーザに関するものである。例えば、乗客輸送機のユーザ及び運転手は、運転中、乗客にできるだけ気を取られないようにしたいと思う。乗客が主な妨害要因であるとはいえ、乗客とのコミュニケーションは時折必要である。

【０２１２】

例えば、ユーザまたは運転手と、外乱源とは、自動車内の一定の位置に配置される。ユーザは基準システムに対して静止しているが、車両自体は移動している。そのため、適応したトラッキングソリューションが必要となる。したがって、通信が行われない限り、乗客の音や会話は、デフォルトで音響的に抑制される。

【０２１３】

例えば、ヒアラブルは、デフォルトで乗員の外乱音を抑制する。ユーザは、適切な制御可能性（音声認識、車両内のボタン）を介して、抑制を手動で解除することができる。ここで、ヒアラブルは、選択に従ってオーディオ再生を適合させる。

【０２１４】

あるいは、ヒアラブルは、乗客がドライバーに積極的に話しかけていることを検出し、ノイズ抑制を一時的に解除する。

【０２１５】

さらなる実施形態の別のユースケースは、学校および教育であり、授業中の教師および生徒に関するものである。一例では、ヒアラブルは２つの役割を持ち、そこでは装置の機能が部分的に結合されている。教師／スピーカーの装置は、邪魔なノイズを抑制し、生徒からのスピーチ／質問を増幅する。また、リスナーのヒアラブルは、教師の装置を通して制御することができる。これにより、特に重要な内容については、より大きな声で話すことなく、強調することができる。生徒は、教師の話をよりよく理解できるように、また、邪魔なクラスメートを排除できるように、ヒアラブルを設定することができる。

【０２１６】

例えば、教師と生徒が閉じた空間の決められた場所に配置されている（これがルールである）。すべての装置が互いに結合されていれば、相対的な位置は交換可能であり、その結果、音源分離が容易になる。選択された音源は、ユーザ（教師／生徒）が選択を取り消すか、警告信号によって装置の機能が中断されるまで、音響的に強調される。

【０２１７】

例えば、教師又はスピーカーがコンテンツを提示し、装置が邪魔なノイズを抑制する。教師は生徒の質問を聞きたいと思い、ヒアラブルのフォーカスを質問者に変更する（自動的または適切な制御可能性を介して）。通信後、すべての音は再び抑制される。さらに、例えば、クラスメートによって邪魔されたと感じている生徒が、それらを音響的にチューニングすることが提供され得る。例えば、さらに、教師から遠くに座っている生徒は、教師の音声を増幅してもよい。

【０２１８】

あるいは、例えば、教師および生徒の装置が結合されてもよい。生徒の装置の選択性は、教師装置を介して一時的に制御されてもよい。特に重要なコンテンツの場合、教師は、自分の声を増幅するために、生徒用装置の選択性を変更する。

【０２１９】

別の実施形態のさらなるユースケースは、軍隊であり、兵士に関するものである。一方では、フィールドにおける兵士間の言語コミュニケーションは、無線を介して行われ、他方では、叫び声や直接接触によって行われる。無線は、異なる部隊や小集団の間で通信が行われる場合に、主に使用される。あらかじめ決められた無線での作法が使われることが多い。叫び声や直接の接触は、主に分隊やグループ内でのコミュニケーションに使われる。兵士の任務中には、音響的に難しい条件（例えば、人々の叫び声、武器の騒音、悪天候など）があり、両方の通信ルートが損なわれる可能性がある。イヤホン付きの無線機は、兵士の装備の一部であることが多い。オーディオ再生という目的の他に、より高いレベルの音圧に対する保護機能も備えている。これらの装置には、環境信号をキャリアーの耳に届けるために、しばしばマイクロフォンが装備されている。また、アクティブノイズサプレッション（能動的騒音抑制）もこのシステムの一部である。機能拡張により、外乱ノイズをインテリジェントに減衰させ、指向性のある再生で音声を選択的に強調することで、騒音環境下での兵士の叫び声や直接のコンタクトを可能にする。このためには、部屋／フィールド内の兵士の相対的な位置を知っておく必要がある。さらに、音声信号と外乱ノイ

ズは、空間的および内容的に互いに分離されなければならない。このシステムは、低いささやき声から叫び声や爆発音まで、高いＳＮＲレベルに対応する必要がある。このようなシステムの利点は、騒音環境下での兵士同士の言語コミュニケーション、聴覚保護具の維持、無線作法の放棄、傍受セキュリティ（無線ソリューションではないため）などが挙げらる。

【０２２０】

例えば、ミッション中の兵士同士の叫び声や直接のコンタクトは、邪魔なノイズのために複雑になることがある。この問題は現在、近距離および長距離の無線ソリューションによって対処されている。新システムは、各スピーカーのインテリジェントかつ空間的な強調と周囲のノイズの減衰により、近距離での叫び声や直接のコンタクトを可能にする。

【０２２１】

例えば、兵士が任務についている場合である。叫び声や話し声は自動的に検知され、システムはそれらを増幅し、同時に周囲の雑音を減衰させる。システムは、ターゲット音源をよく理解できるように、空間的なオーディオ再生を適応させる。

【０２２２】

あるいは、例えば、システムは、グループの兵士を知ることができる。これらのグループのメンバーのオーディオ信号のみが通される。

【０２２３】

さらなる実施形態の使用例は、警備員や保安員に関するものである。したがって、例えば、ヒアラブルは、犯罪の予防的な検出のために、大きなイベント（祝典、抗議活動）を混乱させるのに使用されてもよい。ヒアラブルの選択性は、キーワード、例えば、助けを求める叫び声や暴力への呼びかけによって制御される。これは、オーディオ信号の内容分析（例：音声認識）を前提としている。

【０２２４】

例えば、警備員は多くの大きな音源に囲まれており、警備員とすべての音源が移動している可能性がある。助けを求めている人の声は、通常の聴力では聞こえないか、限られた範囲（ＳＮＲが悪い）しか聞こえない。手動または自動で選択された音源は、ユーザが選択を取り消すまで音響的に強調される。オプションとして、仮想サウンドオブジェクトを関心のある音源の位置／方向に配置し、その場所を簡単に見つけられるようにする（例えば、１回限りの助けを求める場合のために）。

【０２２５】

例えば、ヒアラブルは、潜在的な危険源のある音源を検出する。警備員は、どの音源、またはどのイベントを追いかけたいかを選択する（例えば、タブレットでの選択を通じて）。その後、ヒアラブルは、妨害音の場合でも音源をよく理解し、位置を特定できるように、オーディオ再生を適応させる。

【０２２６】

あるいは、例えば、対象となる音源が無音である場合には、その音源に向かう／その音源から離れた位置にローカリゼーション（ｌｏｃａｌｉｚａｔｉｏｎ）信号を配置するようにしてもよい。

【０２２７】

別の実施形態の別の使用例は、ステージ上のコミュニケーションであり、ミュージシャンに関するものである。ステージ上では、リハーサルやコンサート（バンド、オーケストラ、合唱団、ミュージカルなど）において、単一の楽器（グループ）が、他の環境ではまだ聞こえていたにもかかわらず、困難な音響条件により聞こえない場合がある。この場合、重要な（伴奏の）声が聞こえなくなるため、相互作用が損なわれる。ヒアラブルは、これらの声を強調し、再び聞こえるようにし、その結果、個々のミュージシャンの相互作用を改善し、あるいは確実にすることができるかもしれない。使用することで、個々のミュージシャンの騒音への暴露を減らし、例えばドラムを減衰させることで聴覚の喪失を防ぎ、ミュージシャンは重要なことをすべて同時に聞くことができるようになるかもしれない。

【０２２８】

例えば、ヒアラブルを装着していないミュージシャンは、ステージ上の少なくとも１人の他の声が聞こえなくなる。この場合、ヒアラブルを使用することができる。リハーサル、またはコンサートの終了後、ユーザは、ヒアラブルの電源を切った後、ヒアラブルを取り外す。

【０２２９】

一例では、ユーザは、ヒアラブルをオンにする。ユーザは、増幅される１つ以上の所望の楽器を選択する。一緒に音楽を作るとき、選択された音楽楽器は増幅され、したがって、ヒアラブルによって再び聞こえるようにされる。音楽を作った後、ユーザはヒアラブルを再びオフにする。

【０２３０】

別の例では、ユーザはヒアラブルをオンにする。ユーザは、音量を下げなければならない所望の楽器を選択する。一緒に音楽を作るとき、選択された楽器の音量は、ユーザが適度な音量でのみ聞くことができるように、ヒアラブルによって下げられる。

【０２３１】

例えば、音楽機器のプロファイルは、ヒアラブルに記憶され得る。

【０２３２】

さらなる実施形態の別の使用例は、エコシステムの意味での聴覚装置のソフトウェアモジュールとしてのソース分離であり、聴覚装置の製造業者、または聴覚装置のユーザに関するものである。製造業者は、音源分離を聴覚装置の追加ツールとして使用することができ、顧客に提供することができる。このように、聴覚装置もまた、この開発から利益を得ることができる。また、他の市場／装置（ヘッドフォン、携帯電話など）に対するライセンスモデルも考えられる。

【０２３３】

例えば、聴覚装置のユーザは、複雑な聴覚状況下で異なる音源を分離すること、例えば、特定のスピーカーに焦点を合わせることが困難であると言われている。外部の追加システム（例えば、ブルートゥース（登録商標）による携帯ラジオセットからの信号の転送、ＦＭ機器または誘導聴力機器による教室での選択的信号転送）がなくても選択的に聞くことができるように、ユーザは選択的受聴のための追加機能を備えた聴覚装置を使用する。このように、外部からの働きかけがなくても、ユーザは音源分離によって個々の音源に集中することができる。最後に、ユーザは追加機能をオフにし、聴覚装置で通常通り聞き続ける。

【０２３４】

例えば、ある聴覚装置のユーザが、選択的受聴のための追加機能が組み込まれた新しい聴覚装置を手に入れたとする。ユーザは聴覚装置に選択的受聴のための機能を設定する。次に、ユーザはプロファイルを選択する（例えば、最も大きい／近い音源を増幅する、身の回りの特定の声の音声認識を増幅する（ＵＣ－ＣＥ５－主要イベントのような））。聴覚装置は設定されたプロファイルに従ってそれぞれの音源を増幅し、同時に要求に応じてバックグラウンドノイズを抑制する。聴覚装置のユーザは、単なる「ノイズ」／音響ソースの乱雑さではなく、複雑な聴覚シーンから個々の音源を聞くことができる。

【０２３５】

あるいは、聴覚装置のユーザは、選択的受聴のための付加機能を、自身の聴覚装置用のソフトウェア等として入手する。ユーザは、自分の聴覚装置に付加機能をインストールする。そして、ユーザは、聴覚装置に選択的受聴のための機能を設定する。ユーザはプロファイル（最も大きな／最も近い音源を増幅する、身の回りの特定の声の音声認識を増幅する（ＵＣ－ＣＥ５－大きなイベントなど））を選択し、聴覚装置は設定したプロファイルに従ってそれぞれの音源を増幅し、同時に要求に応じて背景ノイズを抑圧する。この場合、聴覚装置のユーザは、単なる「ノイズ」／音響ソースの乱雑さではなく、複雑な聴覚シーンからの個々のソースを聞くことができる。

【０２３６】

例えば、ヒアラブルは、保存可能な音声プロファイルを提供してもよい。

【０２３７】
さらなる実施形態の使用例は、プロスポーツであり、競技中のアスリートに関するものである。バイアスロン、トライアスロン、サイクリング、マラソンなどのスポーツにおいて、プロのアスリートは、コーチの情報またはチームメイトとのコミュニケーションに依存する。しかし、集中力を高めるために、大きな音（バイアスロンの射撃音、大歓声、パーティーのホーンなど）から身を守りたいという場面もある。ヒアラブルは、関連する音源の完全自動選択（特定の声の検出、典型的な妨害音に対する音量制限）を可能にするよう、それぞれのスポーツ／アスリートに適合させることができる。
【０２３８】
例えば、ユーザは非常に移動しやすく、妨害音の種類はスポーツに依存する可能性がある。激しい身体的負担のため、アスリートによる装置の制御は不可能か、または限定的な範囲にとどまる。しかし、ほとんどのスポーツでは、あらかじめ決められた手順があり（バイアスロン：ランニング、射撃）、重要なコミュニケーションパートナー（トレーナー、チームメイト）をあらかじめ定義することができる。ノイズは全般的に、あるいは活動の特定の局面で抑制される。アスリート、チームメイトおよびコーチのコミュニケーションは常に重視される。
【０２３９】
例えば、アスリート（スポーツ選手）は、スポーツの種類に応じて特別に調整されたヒアラブルを使用する。このヒアラブルは、特にそれぞれのスポーツの種類で高度な注意が必要とされる状況において、外乱音を完全に自動的（事前調整）に抑制する。また、トレーナーやチームメンバーが聞こえる範囲にいる場合は、ヒアラブルは全自動でトレーナーやチームメンバーを強調表示する（事前調整済み）。
【０２４０】
さらなる実施形態の使用例は、聴覚トレーニングであり、音楽学生、プロのミュージシャン、趣味のミュージシャンに関するものである。音楽リハーサル（例えば、オーケストラで、バンドで、アンサンブルで、音楽レッスンで）のために、フィルタリングされた方法で個々の声を追跡できるように、ヒアラブルは選択的に使用される。特にリハーサルの最初のうちは、楽曲の最終録音を聴き、自分の声を追跡するこに役立つ。作曲によっては、前景の声だけを聴くため、背景の声がよく聴こえないことがある。ヒアラブルであれば、楽器を基準に声を選択的に強調するなど、より的を射た練習ができる。
【０２４１】
また、（向上心のある）音大生は、助けを借りずに複雑な楽曲から個々の声を最終的に抽出できるようになるまで、段階的に個々の声の強調を少なくしていくことで、受験に向けた選択的な聴覚能力のトレーニングにヒアラブルを利用することができる。
【０２４２】
さらに考えられる使用例としては、例えば、歌い手などが近くにない場合なら、カラオケが挙げられる。カラオケにサインするための楽器バージョンだけを聞くために、歌声（複数可）をオンデマンドで音楽の一部から抑制することができる。
【０２４３】
例えば、ミュージシャンが楽曲から歌声を学び始めるとする。ミュージシャンは、ＣＤプレーヤーや他の再生媒体で楽曲の録音を聴く。練習が終われば、再びヒアラブルの電源を切る。
【０２４４】
一例として、ユーザはヒアラブルをオンにする。ユーザは、増幅させたい楽器を選択する。音楽を聴くとき、ヒアラブルは楽器の中の声を増幅し、残りの楽器の音量を下げ、したがって、ユーザは自分の声をよりよく追跡することができる。
【０２４５】
別の例では、ユーザはヒアラブルをオンにする。ユーザは、抑制する所望の楽器を選択する。音楽作品を聴くとき、選択された音楽作品の中の声（複数可）が抑制され、残りの声だけが聴こえるようになる。ユーザは、録音された声に気を取られることなく、他の声

と一緒に自分の楽器で発声練習をすることができる。
【０２４６】
実施例では、ヒアラブルは、保存された楽器プロファイルを提供することができる。
【０２４７】
別の実施形態の別の使用例は、仕事中の安全であり、大音量の環境における労働者に関するものである。機械ホールや建設現場のような大音量の環境にいる作業員は、騒音から身を守らなければならないが、警告信号を感知し、同僚とコミュニケーションをとることもできなければならない。
【０２４８】
例えば、ユーザが非常に大きな音の環境にいて、対象となる音源（警告信号、同僚）が妨害音よりかなり小さいかもしれない。ユーザは移動することができるが、妨害音は静止していることが多い。聴覚保護具と同様に、ノイズは恒常的に低減され、ヒアラブルは警告信号を完全に自動で強調する。スピーカーソースを増幅することで、同僚とのコミュニケーションが確保される。
【０２４９】
例えば、ユーザは仕事中、聴覚保護具としてヒアラブルを使用する。警告信号（火災報知器など）は音響的に強調され、ユーザは必要に応じて作業を停止する。
【０２５０】
あるいは、例えば、ユーザは仕事中で、聴覚保護具としてヒアラブルを使用する。同僚とのコミュニケーションの必要性がある場合、適切なインタフェース（ここでは例えば視線制御）の助けを借りて、コミュニケーションの相手が選択され、音響的に強調される。
【０２５１】
さらなる実施形態の別の使用例は、ライブ翻訳者のためのソフトウェアモジュールとしてのソース分離であり、ライブ翻訳者のユーザに関するものである。ライブ翻訳者は、話された外国語をリアルタイムで翻訳し、ソース分離のための上流ソフトウェアモジュールから利益を得ることができる。特に、複数の話者が存在する場合、ソフトウェアモジュールは、ターゲット話者を抽出し、潜在的に翻訳を改善することができる。
【０２５２】
例えば、ソフトウェアモジュールはライブトランスレーター（スマートフォンの専用装置やアプリ）の一部である。例えば、ユーザは、装置のディスプレイを通してターゲットスピーカーを選択することができる。ユーザとターゲット音源は、翻訳の間、動かないか、少ししか動かないことが有利である。選択された音源の位置は音響的に強調されるため、潜在的に翻訳を改善することができる。
【０２５３】
例えば、ユーザが外国語で会話をしたい、あるいは外国語の話し手の話を聞きたいと思ったとする。ユーザは適切なインタフェース（例えばディスプレイ上のＧＵＩ）を通じてターゲットスピーカーを選択し、ソフトウェアモジュールは翻訳機でさらに使用するためにオーディオ録音を最適化する。
【０２５４】
別の実施形態のさらなる使用例は、救援部隊の作業時の安全であり、消防士、市民保護、警察、緊急サービスに関するものである。救援隊にとって、良好なコミュニケーションは、任務を成功裏に処理するために不可欠である。周囲の騒音が大きいにもかかわらず、救援隊が聴覚保護具を携帯することは、コミュニケーションを不可能にするため、救援隊にとってしばしば不可能である。例えば、消防士は、大きなモーター音にもかかわらず、命令を正確に伝え、理解できるようにしなければならないが、これは部分的に無線で行われる。このように、救援隊は聴覚保護条例を遵守できないほど大きな騒音にさらされている。一方では、ヒアラブルは、救援部隊の聴覚保護を提供し、他方では、救援部隊間の通信を依然として可能にすることになる。さらに、ヘルメットや保護具を装着していても、救助隊は音響的に環境から切り離されないので、より良い支援を提供できる可能性がある。救援隊は、より良いコミュニケーションをとることができ、また、危険を予測すること

ができる（例えば、火災の種類を聞き分けるなど）。
【０２５５】
例えば、ユーザは強い周囲の騒音にさらされているため、聴覚保護具を着用することができず、それでも他の人とコミュニケーションをとることができなければならない。その場合、ヒアラブルを使用する。任務が終わった後、または危険な状況が終わった後、ユーザはヒアラブルを再び外す。
【０２５６】
例えば、ミッションの間、ユーザはヒアラブルを装着している。ユーザヒアラブルをオンする。ヒアラブルは周囲の騒音を抑制し、近くにいる同僚や他の話者（例えば、火災の被害者）の音声を増幅する。
【０２５７】
または、ユーザは、ミッション中にヒアラブルを装着する。ユーザは、ヒアラブルをオンにすると、ヒアラブルは周囲の騒音を抑制し、無線で同僚の音声を増幅する。
【０２５８】
該当する場合、ヒアラブルは、運用仕様に従った運用のための構造的適性を満たすように特別に設計されている。可能性としては、ヒアラブルは、無線装置へのインタフェースを構成する。
【０２５９】
いくつかの側面が装置のコンテキスト内で説明されたとしても、前記側面は対応する方法の説明も表すと理解されるので、装置のブロックまたは構造的構成要素は、対応する方法ステップまたは方法ステップの特徴としても理解されることになる。それとの類推により、方法ステップのコンテキスト内で、または方法ステップとして説明されてきた側面は、対応するブロックまたは詳細または対応する装置の特徴の説明も表す。方法ステップのいくつかまたはすべては、マイクロプロセッサ、プログラマブルコンピュータ、または電子回路などのハードウェア装置を使用しながら実行されてもよい。いくつかの実施形態では、最も重要な方法ステップのいくつかまたはいくつかが、そのような装置によって実行されてもよい。
【０２６０】
特定の実装要件に応じて、本発明の実施形態は、ハードウェアで実装されてもよいし、ソフトウェアで実装されてもよい。実装は、デジタル記憶媒体、例えばフロッピー（登録商標）ディスク、ＤＶＤ、ブルーレイディスク、ＣＤ、ＲＯＭ、ＰＲＯＭ、ＥＰＲＯＭ、ＥＥＰＲＯＭまたはフラッシュメモリ、ハードディスク、またはそれぞれの方法が実行されるようにプログラム可能なコンピュータシステムと協力する、または協力し得る電子的に読み取り可能な制御信号をその上に格納している他の磁気または光学メモリを使用しながら行われることができる。このため、デジタル記憶媒体は、コンピュータ読み取り可能であってもよい。
【０２６１】
本発明によるいくつかの実施形態は、このように、本明細書に記載された方法のいずれかが実行されるようなプログラマブルコンピュータシステムと協力することができる電子的に読み取り可能な制御信号を含んでいるデータキャリアを含んでいる。
【０２６２】
一般に、本発明の実施形態は、プログラムコードを有するコンピュータプログラム製品として実施することができ、その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上で実行される場合に、いずれかの方法を実行するのに有効である。
【０２６３】
また、プログラムコードは、例えば、機械読み取り可能な担体に格納されてもよい。
【０２６４】
他の実施形態は、本明細書に記載された方法のいずれかを実行するためのコンピュータプログラムを含み、前記コンピュータプログラムは、機械可読キャリアに格納される。すなわち、このように本発明方法の実施形態は、コンピュータプログラムがコンピュータ上

で実行される場合、本明細書に記載された方法のいずれかを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である。

【０２６５】

このように本発明方法のさらなる実施形態は、本明細書に記載のいずれかの方法を実行するためのコンピュータプログラムが記録されたデータキャリア（またはデジタル記憶媒体またはコンピュータ読取可能な媒体）である。データキャリア、デジタル記憶媒体、または記録媒体は、典型的には有形であり、または不揮発性である。

【０２６６】

したがって、本発明方法のさらなる実施形態は、本明細書に記載された方法のいずれかを実行する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のシーケンスである。データストリームまたは信号のシーケンスは、例えば、データ通信リンク、例えばインターネットを介して送信されるように構成されてもよい。

【０２６７】

さらなる実施形態は、本明細書に記載の方法のいずれかを実行するように構成または適応された、例えばコンピュータまたはプログラマブルロジック装置などの処理ユニットを含む。

【０２６８】

さらなる実施形態は、本明細書に記載された方法のいずれかを実行するためのコンピュータプログラムがインストールされたコンピュータを含む。

【０２６９】

本発明に係る更なる実施形態は、本明細書に記載された方法の少なくとも１つを実行するためのコンピュータプログラムを受信機に送信するように構成された装置またはシステムを含む。送信は、例えば、電子的または光学的であってよい。受信機は、例えば、コンピュータ、モバイル装置、メモリ装置、または同様の装置であってもよい。装置又はシステムは、例えば、コンピュータ・プログラムを受信機に送信するためのファイルサーバを含んでもよい。

【０２７０】

いくつかの実施形態では、プログラマブルロジック装置（例えば、フィールドプログラマブルゲートアレイ、ＦＰＧＡ）は、本明細書に記載される方法の機能性の一部又は全部を実行するために使用されてもよい。いくつかの実施形態では、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、本明細書に記載される方法のいずれかを実行してもよい。一般に、本方法は、いくつかの実施形態において、任意のハードウェア装置によって実行される。前記ハードウェア装置は、コンピュータプロセッサ（ＣＰＵ）のような普遍的に適用可能な任意のハードウェアであってもよいし、ＡＳＩＣのような方法に特有のハードウェアであってもよい。

【０２７１】

上述した実施形態は、単に本発明の原理の例示を示すに過ぎない。当業者であれば、本明細書に記載された配置および詳細の修正および変形を理解することが理解される。このため、本発明は、説明および実施形態の議論によって本明細書に示された特定の詳細によってではなく、以下の請求項の範囲によってのみ限定されることが意図される。

【０２７２】

REFERENCES

[1] V. Valimaki, A. Franck, J. Ramo, H. Gamper, and L. Savioja, "Assisted listening using a headset: Enhancing audio perception in real, augmented, and virtual environments," IEEE Signal Processing Magazine, volume 32, no. 2, pp. 92-99, March 2015.

[2] K. Brandenburg, E. Cano, F. Klein, T. Kollmer, H. Lukashevich, A. Neidhardt , U. Sloma, and S. Werner, "Plausible augmentation of auditory scenes using dynamic binaural synthesis for personalized auditory realities," in Proc. of AES International Conference on Audio for Virtual and Augmented Re

ality, Au gust 2018.
[3] S. Argentieri, P. Dans, and P. Soures, "A survey on sound source localizati on in robotics: From binaural to array processing methods," Computer Speech Language, volume 34, no. 1, pp. 87-112, 2015.
[4] D. FitzGerald, A. Liutkus, and R. Badeau, "Projection-based demixing of s pa tial audio," IEEE/ACM Trans. on Audio, Speech, and Language Processin g, volu me 24, no. 9, pp. 1560-1572, 2016.
[5] E. Cano, D. FitzGerald, A. Liutkus, M. D. Plumbley, and F. Stoter, "Musical source separation: An introduction," IEEE Signal Processing Magazine, volum e 36, no. 1, pp. 31-40, January 2019.
[6] S. Gannot, E. Vincent, S. Markovich-Golan, and A. Ozerov, "A consolidated p erspective on multimicrophone speech enhancement and source separatio n," IEE E/ACM Transactions on Audio, Speech, and Language Processing, vo lume 25, no. 4, pp. 692-730, April 2017.
[7] E. Cano, J. Nowak, and S. Grollmisch, "Exploring sound source separation fo r acoustic condition monitoring in industrial scenarios," in Proc. of 25th E uropean Signal Processing Conference (EUSIPCO), August 2017, pp. 2264 -2268 .
[8] T. Gerkmann, M. Krawczyk-Becker, and J. Le Roux, "Phase processing for sing le-channel speech enhancement: History and recent advances," IEEE Si gnal Pro cessing Magazine, volume 32, no. 2, pp. 55-66, March 2015.
[9] E. Vincent, T. Virtanen, and S. Gannot, Audio Source Separation and Spee ch Enhancement. Wiley, 2018.
[10] D. Matz, E. Cano, and J. Abeser, "New sonorities for early jazz recordings using sound source separation and automatic mixing tools," in Proc. of the 1 6th International Society for Music Information Retrieval Conference. Malaga , Spain: ISMIR, October 2015, pp. 749-755.
[11] S. M. Kuo and D. R. Morgan, "Active noise control: a tutorial review," Pr oc eedings of the IEEE, volume 87, no. 6, pp. 943-973, June 1999.
[12] A. McPherson, R. Jack, and G. Moro, "Action-sound latency: Are our tools fa st enough?" in Proceedings of the International Conference on New Inter faces for Musical Expression, July 2016.
[13] C. Rottondi, C. Chafe, C. Allocchio, and A. Sarti, "An overview on networ ke d music performance technologies," IEEE Access, volume 4, pp. 8823-88 43, 20 16.
[14] S. Liebich, J. Fabry, P. Jax, and P. Vary, "Signal processing challenges fo r active noise cancellation headphones," in Speech Communication; 13th ITG- S ymposium, October 2018, pp. 1-5.
[15] E. Cano, J. Liebetrau, D. Fitzgerald, and K. Brandenburg, "The dimensions o f perceptual quality of sound source separation," in Proc. of IEEE Interna ti onal Conference on Acoustics, Speech and Signal Processing (ICASSP), Ap ril 2 018, pp. 601-605.
[16] P. M. Delgado and J. Herre, "Objective assessment of spatial audio qualit y using directional loudness maps," in Proc. of IEEE International Confere nce on Acoustics, Speech and Signal Processing (ICASSP), May 2019, pp. 6 21-625.
[17] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for i ntelligibility prediction of time-frequency weighted noisy speech," IEEE T ra nsactions on Audio, Speech, and Language Processing, volume 19, no. 7, pp. 2 125-2136, September 2011.

[18] M. D. Plumbley, C. Kroos, J. P. Bello, G. Richard, D. P. Ellis, and A. Mesaros, Proceedings of the Detection and Classification of Acoustic Scenes and Events 2018 Workshop (DCASE2018). Tampere University of Technology. Laboratory of Signal Processing, 2018.
[19] R. Serizel, N. Turpault, H. Eghbal-Zadeh, and A. Parag Shah, "Large- Scale Weakly Labeled Semi-Supervised Sound Event Detection in Domestic Environments," July 2018, submitted to DCASE2018 Workshop.
[20] L. JiaKai, "Mean teacher convolution system for dcase 2018 task 4," DCASE2018 Challenge, Tech. Rep., September 2018.
[21] G. Parascandolo, H. Huttunen, and T. Virtanen, "Recurrent neural networks for polyphonic sound event detection in real life recordings," in Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), March 2016, pp. 6440-6444.
[22] E. C, Cakir and T. Virtanen, "End-to-end polyphonic sound event detection using convolutional recurrent neural networks with learned time-frequency representation input," in Proc. of International Joint Conference on Neural Networks (IJCNN), July 2018, pp. 1-7.
[23] Y. Xu, Q. Kong, W. Wang, and M. D. Plumbley, "Large-Scale Weakly Supervised Audio Classification Using Gated Convolutional Neural Network," in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 2018, pp. 121-125.
[24] B. Frenay and M. Verleysen, "Classification in the presence of label noise: A survey," IEEE Transactions on Neural Networks and Learning Systems, volume 25, no. 5, pp. 845-869, May 2014.
[25] E. Fonseca, M. Plakal, D. P. W. Ellis, F. Font, X. Favory, and X. Serra, "Learning sound event classifiers from web audio with noisy labels," in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 2019.
[26] M. Dorfer and G. Widmer, "Training general-purpose audio tagging networks with noisy labels and iterative self-verification," in Proceedings of the Detection and Classification of Acoustic Scenes and Events 2018 Workshop (DCASE 2018), Surrey, UK, 2018.
[27] S. Adavanne, A. Politis, J. Nikunen, and T. Virtanen, "Sound event localization and detection of overlapping sources using convolutional recurrent neural networks," IEEE Journal of Selected Topics in Signal Processing, pp. 1- 1, 2018.
[28] Y. Jung, Y. Kim, Y. Choi, and H. Kim, "Joint learning using denoising variational autoencoders for voice activity detection," in Proc. of Interspeech, September 2018, pp. 1210-1214.
[29] F. Eyben, F. Weninger, S. Squartini, and B. Schuller, "Real-life voice activity detection with LSTM recurrent neural networks and an application to hollywood movies," in Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing, May 2013, pp. 483-487.
[30] R. Zazo-Candil, T. N. Sainath, G. Simko, and C. Parada, "Feature learning with raw-waveform CLDNNs for voice activity detection," in Proc. of INTERSPEECH, 2016.
[31] M. McLaren, Y. Lei, and L. Ferrer, "Advances in deep neural network approaches to speaker recognition," in Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), April 2015, pp. 4814-4818.

[32] D. Snyder, D. Garcia-Romero, G. Sell, D. Povey, and S. Khudanpur, "X-vec tor s: Robust DNN embeddings for speaker recognition," in Proc. of IEEE In ternat ional Conference on Acoustics, Speech and Signal Processing (ICASS P), April 2018, pp. 5329-5333.
[33] M. McLaren, D. Castan, M. K. Nandwana, L. Ferrer, and E. Yilmaz, "How t o tr ain your speaker embeddings extractor," in Odyssey, 2018.
[34] S. O. Sadjadi, J. W. Pelecanos, and S. Ganapathy, "The IBM speaker recog nit ion system: Recent advances and error analysis," in Proc. of Interspeec h, 20 16, pp. 3633-3637.
[35] Y. Han, J. Kim, and K. Lee, "Deep convolutional neural networks for pred omi nant instrument recognition in polyphonic music," IEEE/ACM Transacti ons on A udio, Speech, and Language Processing, volume 25, no. 1, pp. 208 -221, Janua ry 2017.
[36] V. Lonstanlen and C.-E. Cella, "Deep convolutional networks on the pitch sp iral for musical instrument recognition," in Proceedings of the 17th Int erna tional Society for Music Information Retrieval Conference. New York, USA: IS MIR, 2016, pp. 612-618.
[37] S. Gururani, C. Summers, and A. Lerch, "Instrument activity detection in po lyphonic music using deep neural networks," in Proceedings of the 19th Inter national Society for Music Information Retrieval Conference. Paris, F rance: ISMIR, September 2018, pp. 569-576.
[38] J. Schlutter and B. Lehner, "Zero mean convolutions for level-invariant s in ging voice detection," in Proceedings of the 19th International Society f or Music Information Retrieval Conference. Paris, France: ISMIR, Septembe r 2018 , pp. 321-326.
[39] S. Delikaris-Manias, D. Pavlidi, A. Mouchtaris, and V. Pulkki, "DOA estim at ion with histogram analysis of spatially constrained active intensity vec tor s," in Proc. of IEEE International Conference on Acoustics, Speech and Signa l Processing (ICASSP), March 2017, pp. 526-530.
[40] S. Chakrabarty and E. A. P. Habets, "Multi-speaker DOA estimation using dee p convolutional networks trained with noise signals," IEEE Journal of S elect ed Topics in Signal Processing, volume 13, no. 1, pp. 8- 21, March 2019.
[41] X. Li, L. Girin, R. Horaud, and S. Gannot, "Multiple-speaker localization b ased on direct-path features and likelihood maximization with spatial sparsi ty regularization," IEEE/ACM Transactions on Audio, Speech, and Language P ro cessing, volume 25, no. 10, pp. 1997-2012, October 2017.
[42] F. Grondin and F. Michaud, "Lightweight and optimized sound source loc aliza tion and tracking methods for open and closed microphone array conf iguration s," Robotics and Autonomous Systems, volume 113, pp. 63 - 80, 2019.
[43] D. Yook, T. Lee, and Y. Cho, "Fast sound source localization using two-le ve l search space clustering," IEEE Transactions on Cybernetics, volume 46, no. 1, pp. 20-26, January 2016.
[44] D. Pavlidi, A. Griffin, M. Puigt, and A. Mouchtaris, "Real-time multiple so und source localization and counting using a circular microphone array," IEE E Transactions on Audio, Speech, and Language Processing, volume 21, no. 1 0, pp. 2193-2206, October 2013.
[45] P. Vecchiotti, N. Ma, S. Squartini, and G. J. Brown, "End-to-end binaural s ound localisation from the raw waveform," in Proc. of IEEE International Con ference on Acoustics, Speech and Signal Processing (ICASSP), May 2019, pp. 4 51-455.

[46] Y. Luo, Z. Chen, and N. Mesgarani, "Speaker-independent speech separation with deep attractor network," IEEE/ACM Transactions on Audio, Speech, and Language Processing, volume 26, no. 4, pp. 787-796, April 2018.
[47] Z. Wang, J. Le Roux, and J. R. Hershey, "Multi-channel deep clustering: Discriminative spectral and spatial embeddings for speaker-independent speech separation," in Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), April 2018, pp. 1-5.
[48] G. Naithani, T. Barker, G. Parascandolo, L. BramslLw, N. H. Pontoppidan, and T. Virtanen, "Low latency sound source separation using convolutional recurrent neural networks," in Proc. of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), October 2017, pp. 71-75.
[49] M. Sunohara, C. Haruta, and N. Ono, "Low-latency real-time blind source separation for hearing aids based on time-domain implementation of online independent vector analysis with truncation of non-causal components," in Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), March 2017, pp. 216-220.
[50] Y. Luo and N. Mesgarani, "TaSNet: Time-domain audio separation network for real-time, single-channel speech separation," in Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), April 2018, pp. 696-700.
[51] J. Chua, G. Wang, and W. B. Kleijn, "Convolutive blind source separation with low latency," in Proc. of IEEE International Workshop on Acoustic Signal Enhancement (IWAENC), September 2016, pp. 1-5.
[52] Z. Rafii, A. Liutkus, F. Stoter, S. I. Mimilakis, D. FitzGerald, and B. Pardo, "An overview of lead and accompaniment separation in music," IEEE/ACM Transactions on Audio, Speech, and Language Processing, volume 26, no. 8, pp. 1307-1335, August 2018.
[53] F.-R. Stoter, A. Liutkus, and N. Ito, "The 2018 signal separation evaluation campaign," in Latent Variable Analysis and Signal Separation, Y. Deville, S. Gannot, R. Mason, M. D. Plumbley, and D. Ward, Eds. Cham: Springer International Publishing, 2018, pp. 293-305.
[54] J.-L. Durrieu, B. David, and G. Richard, "A musically motivated midlevel representation for pitch estimation and musical audio source separation," Selected Topics in Signal Processing, IEEE Journal of, volume 5, no. 6, pp. 1180-1191, October 2011.
[55] S. Uhlich, M. Porcu, F. Giron, M. Enenkl, T. Kemp, N. Takahashi, and Y. Mitsufuji, "Improving music source separation based on deep neural networks through data augmentation and network blending," in Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017.
[56] P. N. Samarasinghe, W. Zhang, and T. D. Abhayapala, "Recent advances in active noise control inside automobile cabins: Toward quieter cars," IEEE Signal Processing Magazine, volume 33, no. 6, pp. 61-73, November 2016.
[57] S. Papini, R. L. Pinto, E. B. Medeiros, and F. B. Coelho, "Hybrid approach to noise control of industrial exhaust systems," Applied Acoustics, volume 125, pp. 102 - 112, 2017.
[58] J. Zhang, T. D. Abhayapala, W. Zhang, P. N. Samarasinghe, and S. Jiang, "Active noise control over space: A wave domain approach," IEEE/ACM Transaction s on Audio, Speech, and Language Processing, volume 26, no. 4, pp. 774-786, April 2018.

[59] X. Lu, Y. Tsao, S. Matsuda, and C. Hori, "Speech enhancement based on deep denoising autoencoder," in Proc. of Interspeech, 2013.
[60] Y. Xu, J. Du, L. Dai, and C. Lee, "A regression approach to speech enhancement based on deep neural networks," IEEE/ACM Transactions on Audio, Speech, and Language Processing, volume 23, no. 1, pp. 7-19, January 2015.
[61] S. Pascual, A. Bonafonte, and J. Serra, "SEGAN: speech enhancement generative adversarial network," in Proc. of Interspeech, August 2017, pp. 3642-3646.
[62] F. Weninger, H. Erdogan, S. Watanabe, E. Vincent, J. Le Roux, J. R. Hershey, and B. Schuller, "Speech enhancement with LSTM recurrent neural networks and its application to noise-robust ASR," in Latent Variable Analysis and Signal Separation, E. Vincent, A. Yeredor, Z. Koldovsky, and P. Tichavsky, Eds. Cham: Springer International Publishing, 2015, pp. 91-99.
[63] H. Wierstorf, D. Ward, R. Mason, E. M. Grais, C. Hummersone, and M. D. Plumbley, "Perceptual evaluation of source separation for remixing music," in Proc. of Audio Engineering Society Convention 143, October 2017.
[64] J. Pons, J. Janer, T. Rode, and W. Nogueira, "Remixing music using source separation algorithms to improve the musical experience of cochlear implant users," The Journal of the Acoustical Society of America, volume 140, no. 6, pp. 4338-4349, 2016.
[65] Q. Kong, Y. Xu, W. Wang, and M. D. Plumbley, "A joint separation-classification model for sound event detection of weakly labelled data," in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), March 2018.
[66] T. v. Neumann, K. Kinoshita, M. Delcroix, S. Araki, T. Nakatani, and R. Haeb-Umbach, "All-neural online source separation, counting, and diarization for meeting analysis," in Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2019, pp. 91-95.
[67] S. Gharib, K. Drossos, E. Cakir, D. Serdyuk, and T. Virtanen, "Unsupervised adversarial domain adaptation for acoustic scene classification," in Proceedings of the Detection and Classification of Acoustic Scenes and Events Workshop (DCASE), November 2018, pp. 138-142.
[68] A. Mesaros, T. Heittola, and T. Virtanen, "A multi-device dataset for urban acoustic scene classification," in Proceedings of the Detection and Classification of Acoustic Scenes and Events Workshop, Surrey, UK, 2018.
[69] J. Abeser, M. Gotze, S. Kuhnlenz, R. Grafe, C. Kuhn, T. Claus, H. Lukashevich, "A Distributed Sensor Network for Monitoring Noise Level and Noise Sources in Urban Environments," in Proceedings of the 6th IEEE International Conference on Future Internet of Things and Cloud (FiCloud), Barcelona, Spain, pp. 318-324., 2018.
[70] T. Virtanen, M. D. Plumbley, D. Ellis (Eds.), "Computational Analysis of Sound Scenes and Events," Springer, 2018.
[71] J. Abeser, S. Ioannis Mimilakis, R. Grafe, H. Lukashevich, "Acoustic scene classification by combining autoencoder-based dimensionality reduction and convolutional neural net-works," in Proceedings of the 2nd DCASE Workshop on Detection and Classification of Acoustic Scenes and Events, Munich, Germany, 2017.
[72] A. Avni, J. Ahrens, M. Geierc, S. Spors, H. Wierstorf, B. Rafaely, "Spatial perception of sound fields recorded by spherical microphone arrays with varying spatial resolution," Journal of the Acoustic Society of America, 1

33(5), pp. 2711-2721, 2013.
[73] E. Cano, D. FitzGerald, K. Brandenburg, "Evaluation of quality of sound source separation algorithms: Human perception vs quantitative metrics," in Proceedings of the 24th European Signal Processing Conference (EUSIPCO), pp. 1758-1762, 2016.
[74] S. Marchand, "Audio scene transformation using informed source separation," The Journal of the Acoustical Society of America, 140(4), p. 3091, 2016.
[75] S. Grollmisch, J. Abeser, J. Liebetrau, H. Lukashevich, "Sounding industry: Challenges and datasets for industrial sound analysis (ISA)," in Proceedings of the 27th European Signal Processing Conference (EUSIPCO) (submitted), A Coruna, Spain, 2019.
[76] J. Abeser, M. Muller, "Fundamental frequency contour classification: A comparison between hand-crafted and CNN-based features, " in Proceedings of the 44th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2019.
[77] J. Abeser, S. Balke, M. Muller, "Improving bass saliency estimation using label propagation and transfer learning," in Proceedings of the 19th International Society for Music Information Retrieval Conference (ISMIR), Paris, France, pp. 306-312, 2018.
[78] C.-R. Nagar, J. Abeser, S. Grollmisch, "Towards CNN-based acoustic modeling of seventh chords for recognition chord recognition," in Proceedings of the 16th Sound & Music Computing Conference (SMC) (submitted), Malaga, Spain, 2019.
[79] J. S. Gomez, J. Abeser, E. Cano, "Jazz solo instrument classification with convolutional neural networks, source separation, and transfer learning", in Proceedings of the 19th International Society for Music Information Retrieval Conference (ISMIR), Paris, France, pp. 577- 584, 2018.
[80] J. R. Hershey, Z. Chen, J. Le Roux, S. Watanabe, "Deep clustering: Discriminative embeddings for segmentation and separation," in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 31-35, 2016.
[81] E. Cano, G. Schuller, C. Dittmar, "Pitch-informed solo and accompaniment separation towards its use in music education applications", EURASIP Journal on Advances in Signal Processing, 2014:23, pp. 1-19.
[82] S. I. Mimilakis, K. Drossos, J. F. Santos, G. Schuller, T. Virtanen, Y. Bengio, "Monaural Singing Voice Separation with Skip-Filtering Connections and Recurrent Inference of Time-Frequency Mask," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Calgary, Canada, S.721-725, 2018.
[83] J. F. Gemmeke, D. P. W. Ellis, D. Freedman, A. Jansen, W. Lawrence, R. C. Moore, M. Plakal, M. Ritter, "Audio Set: An ontology and human-labeled dataset for audio events," in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, USA, 2017.
[84] Kleiner, M. "Acoustics and Audio Technology,". 3rd ed. USA: J. Ross Publishing, 2012.
[85] M. Dickreiter, V. Dittel, W. Hoeg, M. Wohr, M. "Handbuch der Tonstudiotechnik," A. medienakademie (Eds). 7th edition, Vol. 1., Munich: K.G. Saur Verlag, 2008.

[86] F. Muller, M. Karau. "Transparant hearing," in: CHI ,02 Extended Abstracts on Human Factors in Computing Systems (CHI EA '02), Minneapolis, USA , pp. 730- 731, April 2002.
[87] L. Vieira. "Super hearing: a study on virtual prototyping for hearables and hearing aids," Master Thesis, Aalborg University, 2018. Available: https://projekter.aau.dk/projekter/files/287515943/MasterThesis__Luis.pdf.
[88] Sennheiser, "AMBEO Smart Headset," [Online]. Available: https://de-de.sennheiser.com/finalstop [Accessed: March 1, 2019].
[89] Orosound "Tilde Earphones" [Online]. Available: https://www.orosound.com/tilde-earphones/ [Accessed; March 1, 2019].
[90] Brandenburg, K., Cano Ceron, E., Klein, F., Kollmer, T., Lukashevich, H., Neidhardt, A., Nowak, J., Sloma, U., und Werner, S., "Personalized aud i tory reality," in 44. Jahrestagung fur Akustik (DAGA), Garching bei Munchen , Deutsche Gesellschaft fur Akustik (DEGA), 2018.
[91] US 2015 195641A1, Application date: January 6, 2014; published on July 9, 2015.

【図面】

【図１】

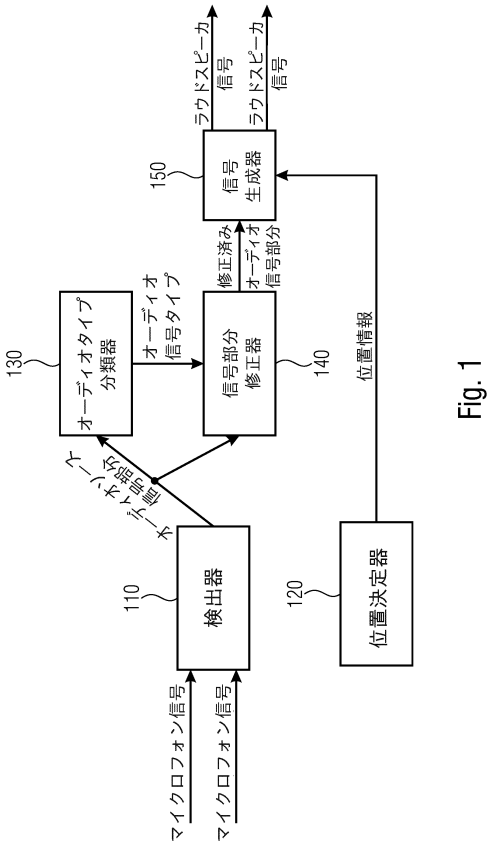

Fig. 1

【図２】

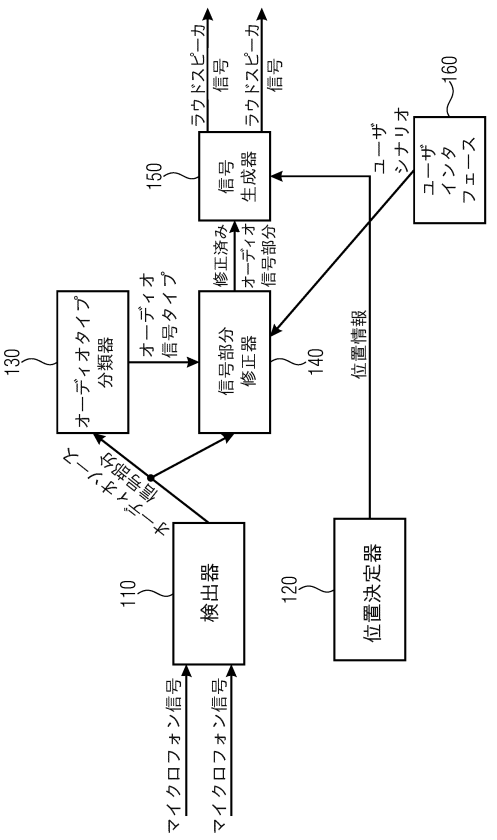

Fig. 2

【図３】

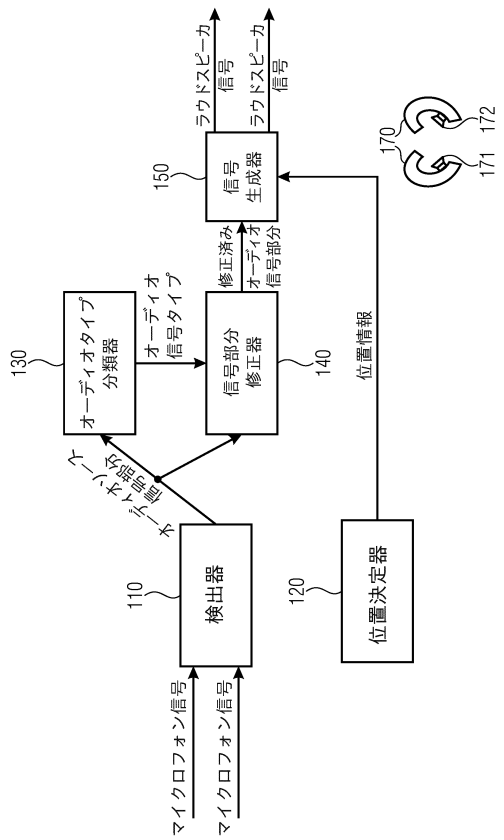

Fig. 3

【図４】

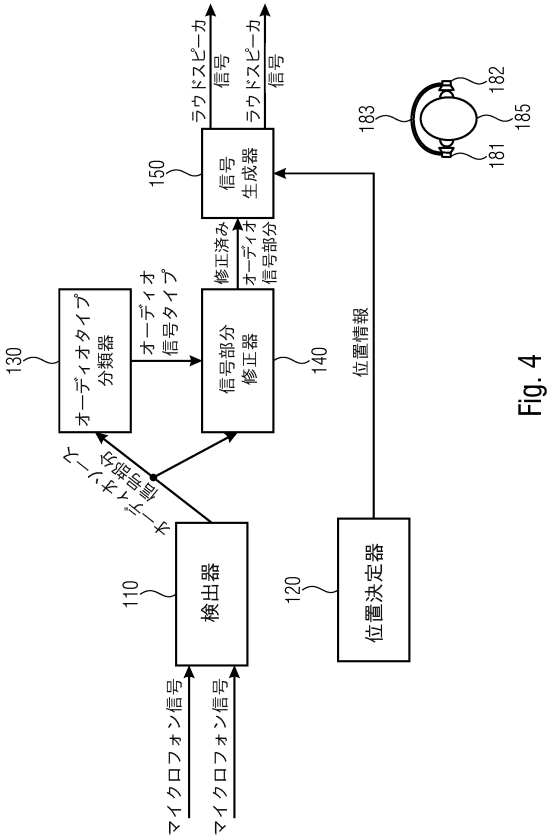

Fig. 4

【図５】

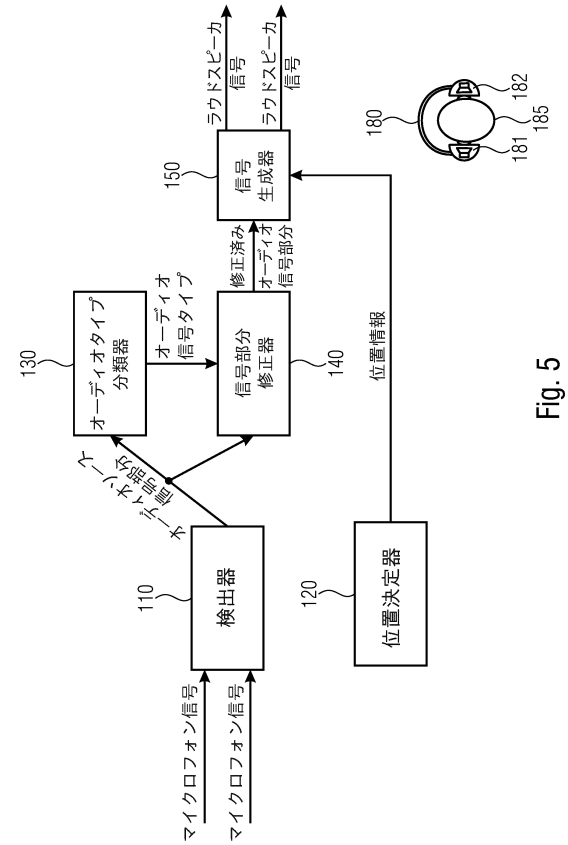

Fig. 5

【図６】

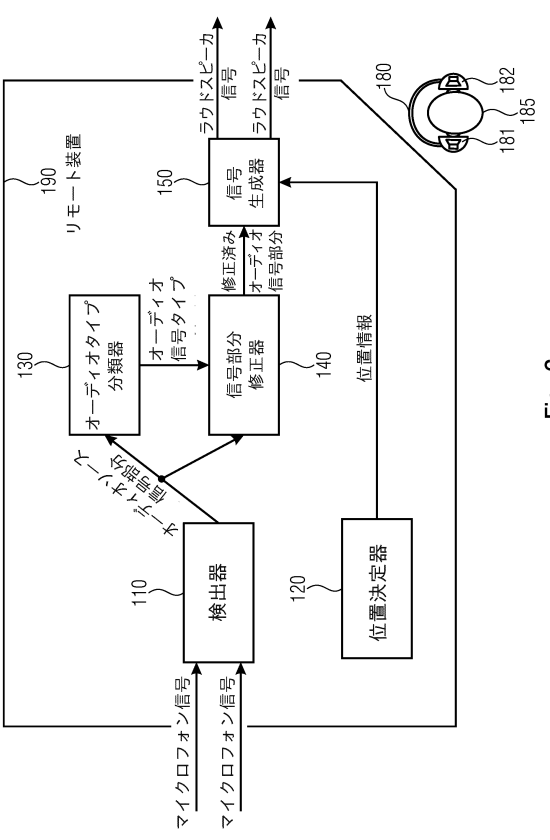

Fig. 6

【図７】

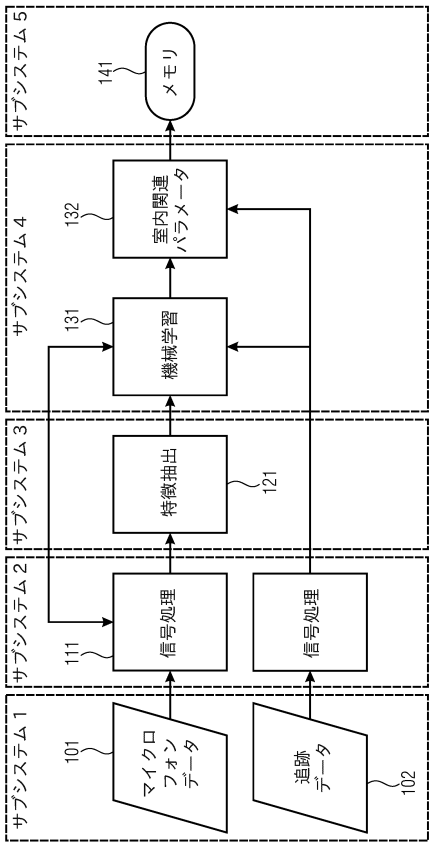

Fig. 7

【図８】

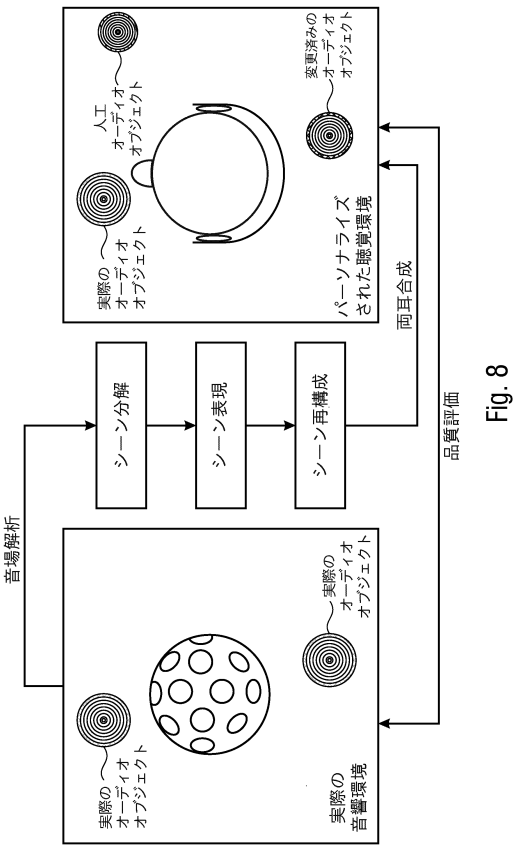

Fig. 8

【図９】

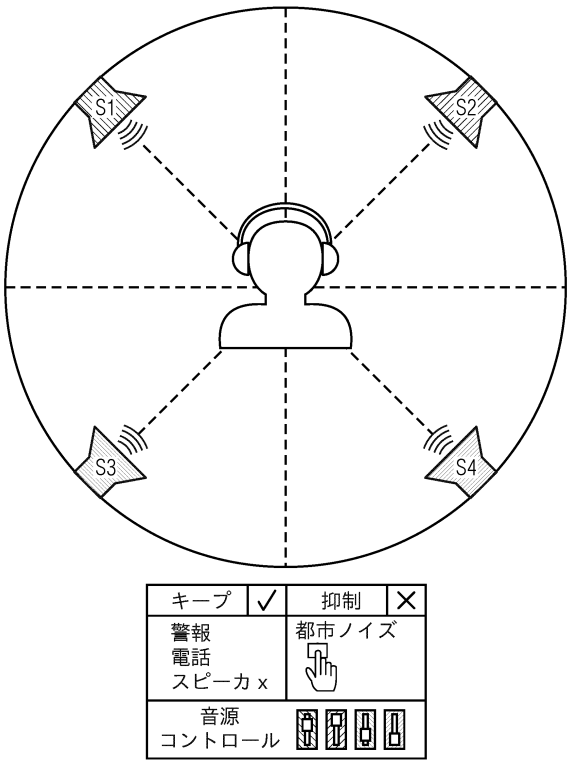

Fig. 9

【図１０】

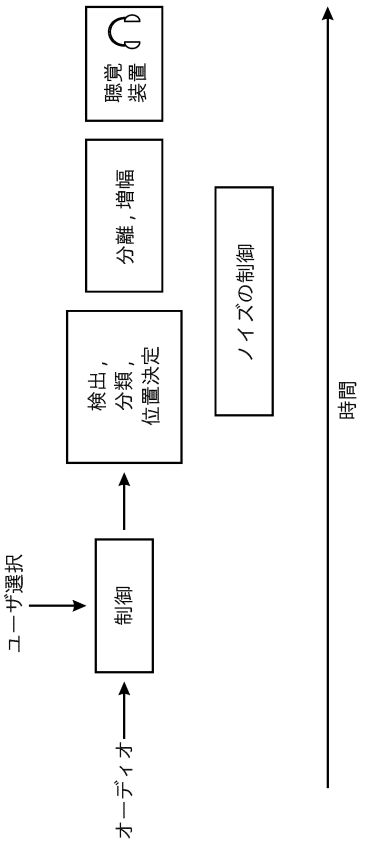

Fig. 10

フロントページの続き

(51)国際特許分類　　ＦＩ
　　Ｈ０４Ｒ　　3/00　　３２０

フラウンホッファー－インスティチュート フュア ディギターレ メーディエンテヒノロギー ＩＤＭＴ 内
(72)発明者　フィッシャー ゲオルグ
ドイツ連邦共和国 ９８６９３ イルメナウ エーレンベルクシュトラーセ ３１ フラウンホッファー－インスティチュート フュア ディギターレ メーディエンテヒノロギー ＩＤＭＴ 内
(72)発明者　ルカシェビッチ ハンナ
ドイツ連邦共和国 ９８６９３ イルメナウ エーレンベルクシュトラーセ ３１ フラウンホッファー－インスティチュート フュア ディギターレ メーディエンテヒノロギー ＩＤＭＴ 内
(72)発明者　クライン フロリアン
ドイツ連邦共和国 ９８６９３ イルメナウ エーレンベルクシュトラーセ ２９ イルメナウ大学 内
(72)発明者　ヴェルナー シュテファン
ドイツ連邦共和国 ９８６９３ イルメナウ エーレンベルクシュトラーセ ２９ イルメナウ大学 内
(72)発明者　ネイトハルト アンニカ
ドイツ連邦共和国 ９８６９３ イルメナウ エーレンベルクシュトラーセ ２９ イルメナウ大学 内
(72)発明者　スロマ ウルリケ
ドイツ連邦共和国 ９８６９３ イルメナウ エーレンベルクシュトラーセ ２９ イルメナウ大学 内
(72)発明者　シュナイダーウィンド クリスティアン
ドイツ連邦共和国 ９８６９３ イルメナウ エーレンベルクシュトラーセ ２９ イルメナウ大学 内
(72)発明者　シュティルナート クラウディア
ドイツ連邦共和国 ９８６９３ イルメナウ エーレンベルクシュトラーセ ２９ イルメナウ大学 内
(72)発明者　カノ セロン エステファニア
ドイツ連邦共和国 ９８６９３ イルメナウ エーレンベルクシュトラーセ ３１ フラウンホッファー－インスティチュート フュア ディギターレ メーディエンテヒノロギー ＩＤＭＴ 内
(72)発明者　アベッサー ヤコブ
ドイツ連邦共和国 ９８６９３ イルメナウ エーレンベルクシュトラーセ ３１ フラウンホッファー－インスティチュート フュア ディギターレ メーディエンテヒノロギー ＩＤＭＴ 内
(72)発明者　スラデチェック クリストフ
ドイツ連邦共和国 ９８６９３ イルメナウ エーレンベルクシュトラーセ ３１ フラウンホッファー－インスティチュート フュア ディギターレ メーディエンテヒノロギー ＩＤＭＴ 内
(72)発明者　ブランデンブルク カールハインツ
ドイツ連邦共和国 ９８６９３ イルメナウ ツア アクツィエン ３

審査官　大野 弘

(56)参考文献　特表２０１７－５０７５５０（ＪＰ，Ａ）
特開２０１２－０１９５０６（ＪＰ，Ａ）
特開２０１０－０７９１８２（ＪＰ，Ａ）
特開２０１０－０５０７５５（ＪＰ，Ａ）
特表２０１４－５０５４２０（ＪＰ，Ａ）
特開２０１７－０９２７３２（ＪＰ，Ａ）

(58)調査した分野　(Int.Cl.，ＤＢ名)
Ｇ１０Ｌ　２１／０３６４
Ｇ１０Ｌ　２５／３０
Ｈ０４Ｓ　７／００
Ｈ０４Ｒ　３／００