

【公報種別】特許法第17条の2の規定による補正の掲載
 【部門区分】第6部門第2区分
 【発行日】令和6年8月29日(2024.8.29)

【国際公開番号】WO2022/076963
 【公表番号】特表2023-545981(P2023-545981A)
 【公表日】令和5年11月1日(2023.11.1)
 【年通号数】公開公報(特許)2023-206
 【出願番号】特願2023-520368(P2023-520368)
 【国際特許分類】

10

G 1 0 L 2 5 / 7 8 (2 0 1 3 . 0 1)
H 0 4 M 1 / 7 2 (2 0 2 1 . 0 1)
G 1 0 L 1 5 / 2 0 (2 0 0 6 . 0 1)
G 0 6 F 1 / 3 2 0 6 (2 0 1 9 . 0 1)
G 0 6 F 1 / 3 2 8 7 (2 0 1 9 . 0 1)
G 0 6 F 1 / 3 2 9 6 (2 0 1 9 . 0 1)
G 0 6 F 1 / 3 2 3 7 (2 0 1 9 . 0 1)

【F I】

G 1 0 L 2 5 / 7 8
 H 0 4 M 1 / 7 2
 G 1 0 L 1 5 / 2 0 3 5 3
 G 0 6 F 1 / 3 2 0 6
 G 0 6 F 1 / 3 2 8 7
 G 0 6 F 1 / 3 2 9 6
 G 0 6 F 1 / 3 2 3 7

20

【手続補正書】

【提出日】令和6年8月21日(2024.8.21)

【手続補正1】

【補正対象書類名】特許請求の範囲

30

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項1】

命令を記憶するように構成されたメモリと、

1つまたは複数のプロセッサとを備え、前記1つまたは複数のプロセッサは、

第1のマイクロフォンの第1の出力に対応する第1のオーディオデータと、第2のマイクロフォンの第2の出力に対応する第2のオーディオデータとを含むオーディオデータを受信することと、

40

前記オーディオデータを動的分類器に提供することと、前記動的分類器は、前記オーディオデータに対応する分類出力を生成するように構成される、
分類出力のクラスの各々を自己音声アクティビティまたは他の音アクティビティに関連付ける検証出力を生成するために検証器に前記オーディオデータを提供することと、

前記分類出力および前記検証出力に少なくとも部分的に基づいて、前記オーディオデータが自己音声アクティビティに対応するかどうかを決定することと

を行うために前記命令を実行するように構成される、デバイス。

【請求項2】

前記第1のマイクロフォンと前記第2のマイクロフォンとをさらに備え、前記第1のマイクロフォンは、前記1つまたは複数のプロセッサに結合され、ユーザの発話をキャプチ

50

ヤするように構成され、前記第2のマイクロフォンは、前記1つまたは複数のプロセッサに結合され、周囲音をキャプチャするように構成される、請求項1に記載のデバイス。

【請求項3】

前記分類出力は、前記第1のオーディオデータと前記第2のオーディオデータとの間の利得差、前記第1のオーディオデータと前記第2のオーディオデータとの間の位相差、またはそれらの組合せに基づく、請求項1に記載のデバイス。

【請求項4】

前記1つまたは複数のプロセッサは、前記第1のオーディオデータと前記第2のオーディオデータとに基づいて特徴データを生成するようにさらに構成され、ここにおいて、前記オーディオデータは、前記特徴データとして前記動的分類器に提供され、前記分類出力は、前記特徴データに少なくとも部分的に基づき、前記特徴データは、

前記第1のオーディオデータと前記第2のオーディオデータとの間の少なくとも1つの両耳間位相差と、

前記第1のオーディオデータと前記第2のオーディオデータとの間の少なくとも1つの両耳間強度差と

を含み得る、請求項1に記載のデバイス。

【請求項5】

前記1つまたは複数のプロセッサは、前記特徴データを生成する前に、前記第1のオーディオデータと前記第2のオーディオデータとを変換領域に変換するようにさらに構成され、前記特徴データは、複数の周波数についての両耳間位相差と、複数の周波数についての両耳間強度差とを含む、請求項4に記載のデバイス。

【請求項6】

前記動的分類器は、前記オーディオデータ内で表される音が、前記第2のマイクロフォンよりも前記第1のマイクロフォンに近いソースから発生しているかどうかに基づいて、特徴データのセットを適応的にクラスタ化するように構成される、請求項4に記載のデバイス。

【請求項7】

前記1つまたは複数のプロセッサは、

前記オーディオデータに基づく前記動的分類器のクラスタ化動作、または前記動的分類器の分類決定基準

を更新するようにさらに構成される、請求項1に記載のデバイス。

【請求項8】

前記動的分類器は、自己組織化マップを含む、請求項1に記載のデバイス。

【請求項9】

前記動的分類器は、オーディオデータのセットのシーケンスを受信し、前記シーケンス内のオーディオデータの以前のセットに少なくとも部分的に基づいて前記シーケンスのセットを適応的にクラスタ化するようにさらに構成される、請求項1に記載のデバイス。

【請求項10】

前記1つまたは複数のプロセッサは、前記オーディオデータが前記自己音声アクティビティに対応するという決定に回答して、音声コマンド処理動作を開始するようにさらに構成される、請求項1に記載のデバイス。

【請求項11】

前記1つまたは複数のプロセッサは、前記音声コマンド処理動作を開始するために起動信号または割込みのうちの少なくとも1つを生成するように構成され、前記1つまたは複数のプロセッサは、好ましくは、

前記動的分類器を含む常時接続電力ドメインと、

音声コマンド処理ユニットを含む第2の電力ドメインと、ここにおいて、前記起動信号は、前記音声コマンド処理ユニットをアクティブ化するために前記第2の電力ドメインを低電力モードから遷移させるように構成される、

をさらに含む、請求項10に記載のデバイス。

10

20

30

40

50

【請求項 1 2】

前記 1 つまたは複数のプロセッサは、前記第 1 のマイクロフォンと前記第 2 のマイクロフォンとを含むヘッドセットデバイスに統合され、前記ヘッドセットデバイスは、ユーザによって装着されたとき、前記第 1 のマイクロフォンにおいて、前記第 2 のマイクロフォンと比較してより大きい強度およびより少ない遅延で前記ユーザの発話をキャプチャするために、前記第 1 のマイクロフォンを前記第 2 のマイクロフォンよりも前記ユーザの口の近くに配置するように構成される、請求項 1 に記載のデバイス。

【請求項 1 3】

前記 1 つまたは複数のプロセッサは、モバイルフォン、タブレットコンピュータデバイス、ウェアラブル電子デバイス、カメラデバイス、仮想現実ヘッドセット、拡張現実ヘッドセットのうち少なくとも 1 つ、またはビークルのオペレータの発話をキャプチャするように配置される前記第 1 のマイクロフォンおよび前記第 2 のマイクロフォンを含む前記ビークルに統合される、請求項 1 に記載のデバイス。

10

【請求項 1 4】

音声アクティビティ検出の方法であって、

1 つまたは複数のプロセッサにおいて、第 1 のマイクロフォンの第 1 の出力に対応する第 1 のオーディオデータと第 2 のマイクロフォンの第 2 の出力に対応する第 2 のオーディオデータとを含むオーディオデータを受信することと、

前記 1 つまたは複数のプロセッサにおいて、前記オーディオデータに対応する分類出力を生成するために、前記オーディオデータを動的分類器に提供することと、

20

前記 1 つまたは複数のプロセッサにおいて、分類出力のクラスの各々を自己音声アクティビティまたは他の音アクティビティに関連付ける検証出力を生成するために検証器に前記オーディオデータを提供することと、

前記 1 つまたは複数のプロセッサにおいて、前記分類出力および前記検証出力に少なくとも部分的に基づいて、前記オーディオデータがユーザ音声アクティビティに対応するかどうかを決定することと
を備える、方法。

【請求項 1 5】

命令を備える非一時的コンピュータ可読媒体であって、前記命令は、1 つまたは複数のプロセッサによって実行されたとき、前記 1 つまたは複数のプロセッサに、請求項 1 4 の方法を実行させる、非一時的コンピュータ可読媒体。

30

【手続補正 2】

【補正対象書類名】明細書

【補正対象項目名】0 2 0 3

【補正方法】変更

【補正の内容】

【0 2 0 3】

[0208] 開示される態様の上記の説明は、開示される態様を当業者が作成または使用することを可能にするために与えられた。これらの態様への様々な変更は当業者には容易に明らかになり、本明細書で定義された原理は本開示の範囲から逸脱することなく他の態様に適用され得る。したがって、本開示は、本明細書に示された態様に限定されるものではなく、以下の特許請求の範囲によって定義される原理および新規の特徴に一致する可能な最も広い範囲を与えられるべきである。

40

以下に本願の出願当初の特許請求の範囲に記載された発明を付記する。

[C 1]

命令を記憶するように構成されたメモリと、

1 つまたは複数のプロセッサとを備え、前記 1 つまたは複数のプロセッサは、

第 1 のマイクロフォンの第 1 の出力に対応する第 1 のオーディオデータと、第 2 のマイクロフォンの第 2 の出力に対応する第 2 のオーディオデータとを含むオーディオデータを受信することと、

50

前記オーディオデータを動的分類器に提供することと、前記動的分類器は、前記オーディオデータに対応する分類出力を生成するように構成される、

前記分類出力に少なくとも部分的に基づいて、前記オーディオデータがユーザ音声アクティビティに対応するかどうかを決定することと

を行うために前記命令を実行するように構成される、デバイス。

[C 2]

前記第 1 のマイクロフォンと前記第 2 のマイクロフォンとをさらに備え、前記第 1 のマイクロフォンは、前記 1 つまたは複数のプロセッサに結合され、ユーザの発話をキャプチャするように構成され、前記第 2 のマイクロフォンは、前記 1 つまたは複数のプロセッサに結合され、周囲音をキャプチャするように構成される、C 1 に記載のデバイス。

10

[C 3]

前記分類出力は、前記第 1 のオーディオデータと前記第 2 のオーディオデータとの間の利得差、前記第 1 のオーディオデータと前記第 2 のオーディオデータとの間の位相差、またはそれらの組合せに基づく、C 1 に記載のデバイス。

[C 4]

前記 1 つまたは複数のプロセッサは、前記第 1 のオーディオデータと前記第 2 のオーディオデータとに基づいて特徴データを生成するようにさらに構成され、ここにおいて、前記オーディオデータは、前記特徴データとして前記動的分類器に提供され、前記分類出力は、前記特徴データに基づく、C 1 に記載のデバイス。

[C 5]

前記特徴データは、

前記第 1 のオーディオデータと前記第 2 のオーディオデータとの間の少なくとも 1 つの両耳間位相差と、

前記第 1 のオーディオデータと前記第 2 のオーディオデータとの間の少なくとも 1 つの両耳間強度差と

を含む、C 4 に記載のデバイス。

20

[C 6]

前記 1 つまたは複数のプロセッサは、前記特徴データの少なくとも 1 つの値の符号または大きさのうちの少なくとも 1 つにさらに基づいて、前記オーディオデータが前記ユーザ音声アクティビティに対応するかどうかを決定するように構成される、C 4 に記載のデバイス。

30

[C 7]

前記 1 つまたは複数のプロセッサは、前記特徴データを生成する前に、前記第 1 のオーディオデータと前記第 2 のオーディオデータとを変換領域に変換するようにさらに構成され、前記特徴データは、複数の周波数についての両耳間位相差と、複数の周波数についての両耳間強度差とを含む、C 4 に記載のデバイス。

[C 8]

前記動的分類器は、前記オーディオデータ内で表される音が、前記第 2 のマイクロフォンよりも前記第 1 のマイクロフォンに近いソースから発生しているかどうかに基づいて、特徴データのセットを適応的にクラスタ化するように構成される、C 4 に記載のデバイス。

40

[C 9]

前記 1 つまたは複数のプロセッサは、前記オーディオデータに基づいて前記動的分類器のクラスタ化動作を更新するようにさらに構成される、C 1 に記載のデバイス。

[C 1 0]

前記 1 つまたは複数のプロセッサは、前記動的分類器の分類決定基準を更新するようにさらに構成される、C 1 に記載のデバイス。

[C 1 1]

前記動的分類器は、自己組織化マップを含む、C 1 に記載のデバイス。

[C 1 2]

50

前記動的分類器は、オーディオデータのセットのシーケンスを受信し、前記シーケンス内のオーディオデータの以前のセットに少なくとも部分的に基づいて前記シーケンスのセットを適応的にクラスタ化するようにさらに構成される、C 1 に記載のデバイス。

[C 1 3]

前記 1 つまたは複数のプロセッサは、前記オーディオデータが前記ユーザ音声アクティビティに対応するという決定に应答して、音声コマンド処理動作を開始するようにさらに構成される、C 1 に記載のデバイス。

[C 1 4]

前記 1 つまたは複数のプロセッサは、前記音声コマンド処理動作を開始するために起動信号または割込みのうちの少なくとも 1 つを生成するように構成される、C 1 3 に記載のデバイス。

[C 1 5]

前記 1 つまたは複数のプロセッサは、前記動的分類器を含む常時接続電力ドメインと、音声コマンド処理ユニットを含む第 2 の電力ドメインと、ここにおいて、前記起動信号は、前記音声コマンド処理ユニットをアクティブ化するために前記第 2 の電力ドメインを低電力モードから遷移させるように構成される、をさらに含む、C 1 4 に記載のデバイス。

[C 1 6]

前記 1 つまたは複数のプロセッサに結合されたモデムをさらに備え、前記モデムは、前記動的分類器に基づいた、前記オーディオデータが前記ユーザ音声アクティビティに対応するという決定に应答して、前記オーディオデータを第 2 のデバイスに送信するように構成される、C 1 に記載のデバイス。

[C 1 7]

前記 1 つまたは複数のプロセッサは、前記第 1 のマイクロフォンと前記第 2 のマイクロフォンとを含むヘッドセットデバイスに統合され、前記ヘッドセットデバイスは、ユーザによって装着されたとき、前記第 1 のマイクロフォンにおいて、前記第 2 のマイクロフォンと比較してより大きい強度およびより少ない遅延で前記ユーザの発話をキャプチャするために、前記第 1 のマイクロフォンを前記第 2 のマイクロフォンよりも前記ユーザの口の近くに配置するように構成される、C 1 に記載のデバイス。

[C 1 8]

前記 1 つまたは複数のプロセッサは、モバイルフォン、タブレットコンピュータデバイス、ウェアラブル電子デバイス、カメラデバイス、仮想現実ヘッドセット、または拡張現実ヘッドセットのうちの少なくとも 1 つに統合される、C 1 に記載のデバイス。

[C 1 9]

前記 1 つまたは複数のプロセッサは、ピークルに統合され、前記ピークルは、前記第 1 のマイクロフォンと前記第 2 のマイクロフォンとをさらに含み、前記第 1 のマイクロフォンは、前記ピークルのオペレータの発話をキャプチャするように配置される、C 1 に記載のデバイス。

[C 2 0]

音声アクティビティ検出の方法であって、1 つまたは複数のプロセッサにおいて、第 1 のマイクロフォンの第 1 の出力に対応する第 1 のオーディオデータと第 2 のマイクロフォンの第 2 の出力に対応する第 2 のオーディオデータとを含むオーディオデータを受信することと、

前記 1 つまたは複数のプロセッサにおいて、前記オーディオデータに対応する分類出力を生成するために、前記オーディオデータを動的分類器に提供することと、

前記 1 つまたは複数のプロセッサにおいて、前記分類出力に少なくとも部分的に基づいて、前記オーディオデータがユーザ音声アクティビティに対応するかどうかを決定することと

を備える、方法。

10

20

30

40

50

[C 2 1]

前記分類出力は、前記第 1 のオーディオデータと前記第 2 のオーディオデータとの間の利得差、前記第 1 のオーディオデータと前記第 2 のオーディオデータとの間の位相差、またはそれらの組合せに基づき、C 2 0 に記載の方法。

[C 2 2]

前記動的分類器は、自己組織化マップを含む、C 2 0 に記載の方法。

[C 2 3]

前記オーディオデータが前記ユーザ音声アクティビティに対応するかどうかを決定することは、前記オーディオデータに対応する特徴データの少なくとも 1 つの値の符号または大きさのうちの少なくとも 1 つにさらに基づき、C 2 0 に記載の方法。

[C 2 4]

前記オーディオデータが前記ユーザ音声アクティビティに対応するという決定にตอบสนองして、音声コマンド処理動作を開始することをさらに備える、C 2 0 に記載の方法。

[C 2 5]

前記音声コマンド処理動作を開始するために起動信号または割込みのうちの少なくとも 1 つを生成することをさらに備える、C 2 4 に記載の方法。

[C 2 6]

前記動的分類器に基づいた、前記オーディオデータが前記ユーザ音声アクティビティに対応するという決定にตอบสนองして、前記オーディオデータを第 2 のデバイスに送信することをさらに備える、C 2 0 に記載の方法。

[C 2 7]

命令を備える非一時的コンピュータ可読媒体であって、前記命令は、1 つまたは複数のプロセッサによって実行されたとき、前記 1 つまたは複数のプロセッサに、

第 1 のマイクロフォンの第 1 の出力に対応する第 1 のオーディオデータと、第 2 のマイクロフォンの第 2 の出力に対応する第 2 のオーディオデータとを含むオーディオデータを受信することと、

前記オーディオデータに対応する分類出力を生成するために、前記オーディオデータを動的分類器に提供することと、

前記分類出力に少なくとも部分的に基づいて、前記オーディオデータがユーザ音声アクティビティに対応するかどうかを決定することと

を行わせる、非一時的コンピュータ可読媒体。

[C 2 8]

前記分類出力は、前記第 1 のオーディオデータと前記第 2 のオーディオデータとの間の利得差、前記第 1 のオーディオデータと前記第 2 のオーディオデータとの間の位相差、またはそれらの組合せに基づき、C 2 7 に記載の非一時的コンピュータ可読媒体。

[C 2 9]

第 1 のマイクロフォンの第 1 の出力に対応する第 1 のオーディオデータと、第 2 のマイクロフォンの第 2 の出力に対応する第 2 のオーディオデータとを含むオーディオデータを受信するための手段と、

動的分類器において、前記オーディオデータに対応する分類出力を生成するための手段と、

前記分類出力に少なくとも部分的に基づいて、前記オーディオデータがユーザ音声アクティビティに対応するかどうかを決定するための手段と

を備える装置。

[C 3 0]

前記分類出力は、前記第 1 のオーディオデータと前記第 2 のオーディオデータとの間の利得差、前記第 1 のオーディオデータと前記第 2 のオーディオデータとの間の位相差、またはそれらの組合せに基づき、C 2 9 に記載の装置。

10

20

30

40

50