【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３区分
【発行日】令和5年6月20日(2023.6.20)

【国際公開番号】WO2021/011331
【公表番号】特表2022-539794(P2022-539794A)
【公表日】令和4年9月13日(2022.9.13)
【年通号数】公開公報(特許)2022-169
【出願番号】特願2022-500128(P2022-500128)
【国際特許分類】
Ｇ０６Ｆ　3/01(2006.01)
Ｇ０６Ｆ　3/16(2006.01)
【ＦＩ】
Ｇ０６Ｆ　3/01　５７０
Ｇ０６Ｆ　3/16　６５０
Ｇ０６Ｆ　3/16　６７０
Ｇ０６Ｆ　3/01　５１０

【手続補正書】
【提出日】令和5年6月12日(2023.6.12)
【手続補正１】
【補正対象書類名】特許請求の範囲
【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】
マルチモーダルユーザ入力のためのデバイスであって、
第１の入力デバイスから受信された第１のデータを処理することと、前記第１のデータは、第１の入力モードに基づくユーザからの第１の入力を示し、前記第１の入力は、コマンドに対応し、
前記第１のデータを処理することに基づいて出力デバイスにフィードバックメッセージを送ることと、ここにおいて、前記フィードバックメッセージは、前記第１の入力モードとは異なる第２の入力モードに基づいて、前記第１の入力に関連するコマンドを識別する第２の入力を提供するように前記ユーザに命令する、
第２の入力デバイスから第２のデータを受信することと、前記第２のデータは、前記第２の入力を示し、
前記第１の入力のマッピングを、前記第２の入力に関連付けられた行為に対して更新することと、
を行うように構成された１つまたは複数のプロセッサを備える、デバイス。
【請求項２】
前記第１の入力モードは、スピーチモード、ジェスチャーモード、またはビデオモードのうちの１つであり、前記第２の入力モードは、前記スピーチモード、前記ジェスチャーモード、または前記ビデオモードのうちの異なる１つである、請求項１に記載のデバイス。
【請求項３】
前記フィードバックメッセージは、前記第１の入力をディスアンビギュエートするために前記第２の入力を提供するように前記ユーザに命令し、
オプションで、前記１つまたは複数のプロセッサは、前記第１の入力の認識処理に関連する確信度レベルが確信度しきい値を満たすことに失敗したことに応答して、前記フィー

ドバックメッセージを送るようにさらに構成された、請求項１に記載のデバイス。
【請求項４】
前記更新されたマッピングは、前記第１の入力と前記第２の入力との組合せを前記コマンドに関連付ける、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、マルチモーダル認識エンジンを含み、前記マルチモーダル認識エンジンは、
組み合わされた埋め込みベクトルを生成するために、前記第１の入力モードに関連する第１の埋め込みネットワークと、前記第２の入力モードに関連する第２の埋め込みネットワークとの出力を組み合わせるように構成された融合埋め込みネットワークと、
前記組み合わされた埋め込みベクトルを特定のコマンドにマッピングするように構成された分類器と、
を含む、請求項１に記載のデバイス。
【請求項６】
前記ユーザに対応する第１の埋め込みネットワークデータおよび第１の重みデータと、
第２のユーザに対応する第２の埋め込みネットワークデータおよび第２の重みデータと、前記第１の埋め込みネットワークデータは、前記ユーザと前記第２のユーザとの間の入力コマンドの差に基づいて前記第２の埋め込みネットワークデータとは異なり、前記第１の重みデータは、前記ユーザと前記第２のユーザとの間の入力モード信頼性の差に基づいて前記第２の重みデータとは異なり、
を記憶するように構成されたメモリをさらに備える、請求項５に記載のデバイス。
【請求項７】
前記第１の入力モードは、ビデオモードに対応し、前記１つまたは複数のプロセッサは、照明しきい値を下回る値を有する周辺光メトリックに応答して前記フィードバックメッセージを送るように構成された、請求項１に記載のデバイス。
【請求項８】
前記第１の入力モードは、スピーチモードに対応し、前記１つまたは複数のプロセッサは、雑音しきい値を超える値を有する雑音メトリックに応答して前記フィードバックメッセージを送るように構成された、請求項１に記載のデバイス。
【請求項９】
グラフィカルユーザインターフェースを表すように構成されたディスプレイをさらに備える、請求項１に記載のデバイス。
【請求項１０】
１つまたは複数のキーワードまたは音声コマンドを含むオーディオ入力をキャプチャするように構成された１つまたは複数のマイクロフォンをさらに備える、請求項１に記載のデバイス。
【請求項１１】
１つまたは複数のジェスチャーまたは視覚的コマンドを含むビデオ入力をキャプチャするように構成された１つまたは複数のカメラをさらに備える、請求項１に記載のデバイス。
【請求項１２】
ジェスチャー入力を示すデータを受信するように構成された１つまたは複数のアンテナをさらに備える、請求項１に記載のデバイス。
【請求項１３】
前記フィードバックメッセージをレンダリングするかまたは前記ユーザにダイレクトするように構成された１つまたは複数のラウドスピーカーをさらに備える、請求項１に記載のデバイス。
【請求項１４】
マルチモーダルユーザ入力のための方法であって、
デバイスの１つまたは複数のプロセッサにおいて、第１の入力デバイスから受信された

第１のデータを処理することと、前記第１のデータは、第１の入力モードに基づくユーザからの第１の入力を示し、前記第１の入力は、コマンドに対応し、
前記１つまたは複数のプロセッサから、前記第１のデータを処理することに基づいて出力デバイスにフィードバックメッセージを送ることと、ここにおいて、前記フィードバックメッセージは、前記第１の入力モードとは異なる第２の入力モードに基づいて、前記第１の入力に関連するコマンドを識別する第２の入力を提供するように前記ユーザに命令する、
前記１つまたは複数のプロセッサにおいて、第２の入力デバイスから第２のデータを受信することと、前記第２のデータは、前記第２の入力を示し、
前記第１の入力のマッピングを、前記第２の入力に関連付けられた行為に対して更新することと、
を備える、方法。
【請求項１５】
デバイスの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の入力デバイスから受信された第１のデータを処理することと、前記第１のデータは、第１の入力モードに基づくユーザからの第１の入力を示し、前記第１の入力が、コマンドに対応し、
前記第１のデータを処理することに基づいて出力デバイスにフィードバックメッセージを送ることと、ここにおいて、前記フィードバックメッセージは、前記第１の入力モードとは異なる第２の入力モードに基づいて、前記第１の入力に関連するコマンドを識別する第２の入力を提供するように前記ユーザに命令する、
第２の入力デバイスから第２のデータを受信することと、前記第２のデータは、前記第２の入力を示し、
前記第１の入力のマッピングを、前記第２の入力に関連付けられた前記コマンドに対して更新することと、
を行わせる命令を備える非一時的コンピュータ可読媒体。
【手続補正２】
【補正対象書類名】明細書
【補正対象項目名】０１１９
【補正方法】変更
【補正の内容】
【０１１９】
[0132]開示される実装形態の前の説明は、開示される実装形態を当業者が製作または使用することを可能にするために提供される。これらの実装形態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実装形態に適用され得る。したがって、本開示は、本明細書に示された実装形態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ような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。
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。
［Ｃ１］
マルチモーダルユーザ入力のためのデバイスであって、
第１の入力デバイスから受信された第１のデータを処理することと、前記第１のデータは、第１の入力モードに基づくユーザからの第１の入力を示し、前記第１の入力は、コマンドに対応し、
前記第１のデータを処理することに基づいて出力デバイスにフィードバックメッセージを送ることと、ここにおいて、前記フィードバックメッセージは、前記第１の入力モードとは異なる第２の入力モードに基づいて、前記第１の入力に関連するコマンドを識別する第２の入力を提供するように前記ユーザに命令する、
第２の入力デバイスから第２のデータを受信することと、前記第２のデータは、前記

第２の入力を示し、
　前記第２の入力によって識別される前記コマンドに前記第１の入力を関連付けるようにマッピングを更新することと、
　を行うように構成された１つまたは複数のプロセッサを備える、デバイス。
［Ｃ２］
　前記第１の入力モードは、スピーチモード、ジェスチャーモード、またはビデオモードのうちの１つであり、前記第２の入力モードは、前記スピーチモード、前記ジェスチャーモード、または前記ビデオモードのうちの異なる１つである、Ｃ１に記載のデバイス。
［Ｃ３］
　前記フィードバックメッセージは、前記第１の入力をディスアンビギュエートするために前記第２の入力を提供するように前記ユーザに命令する、Ｃ１に記載のデバイス。
［Ｃ４］
　前記１つまたは複数のプロセッサは、前記第１の入力の認識処理に関連する確信度レベルが確信度しきい値を満たすことに失敗したことに応答して、前記フィードバックメッセージを送るようにさらに構成された、Ｃ３に記載のデバイス。
［Ｃ５］
　前記更新されたマッピングは、前記第１の入力と前記第２の入力との組合せを前記コマンドに関連付ける、Ｃ１に記載のデバイス。
［Ｃ６］
　前記１つまたは複数のプロセッサは、マルチモーダル認識エンジンを含み、前記マルチモーダル認識エンジンは、
　組み合わされた埋め込みベクトルを生成するために、前記第１の入力モードに関連する第１の埋め込みネットワークと、前記第２の入力モードに関連する第２の埋め込みネットワークとの出力を組み合わせるように構成された融合埋め込みネットワークと、
　前記組み合わされた埋め込みベクトルを特定のコマンドにマッピングするように構成された分類器と、
　を含む、Ｃ１に記載のデバイス。
［Ｃ７］
　前記ユーザに対応する第１の埋め込みネットワークデータおよび第１の重みデータと、
　第２のユーザに対応する第２の埋め込みネットワークデータおよび第２の重みデータと、前記第１の埋め込みネットワークデータは、前記ユーザと前記第２のユーザとの間の入力コマンドの差に基づいて前記第２の埋め込みネットワークデータとは異なり、前記第１の重みデータは、前記ユーザと前記第２のユーザとの間の入力モード信頼性の差に基づいて前記第２の重みデータとは異なり、
　を記憶するように構成されたメモリをさらに備える、Ｃ６に記載のデバイス。
［Ｃ８］
　前記第１の入力モードは、ビデオモードに対応し、前記１つまたは複数のプロセッサは、照明しきい値を下回る値を有する周辺光メトリックに応答して前記フィードバックメッセージを送るように構成された、Ｃ１に記載のデバイス。
［Ｃ９］
　前記第１の入力モードは、スピーチモードに対応し、前記１つまたは複数のプロセッサは、雑音しきい値を超える値を有する雑音メトリックに応答して前記フィードバックメッセージを送るように構成された、Ｃ１に記載のデバイス。
［Ｃ１０］
　グラフィカルユーザインターフェースを表すように構成されたディスプレイをさらに備える、Ｃ１に記載のデバイス。
［Ｃ１１］
　１つまたは複数のキーワードまたは音声コマンドを含むオーディオ入力をキャプチャするように構成された１つまたは複数のマイクロフォンをさらに備える、Ｃ１に記載のデバイス。

［Ｃ１２］
１つまたは複数のジェスチャーまたは視覚的コマンドを含むビデオ入力をキャプチャするように構成された１つまたは複数のカメラをさらに備える、Ｃ１に記載のデバイス。
［Ｃ１３］
ジェスチャー入力を示すデータを受信するように構成された１つまたは複数のアンテナをさらに備える、Ｃ１に記載のデバイス。
［Ｃ１４］
前記フィードバックメッセージをレンダリングするかまたは前記ユーザにダイレクトするように構成された１つまたは複数のラウドスピーカーをさらに備える、Ｃ１に記載のデバイス。
［Ｃ１５］
前記ユーザは、ロボットまたは他の電子デバイスを含む、Ｃ１に記載のデバイス。
［Ｃ１６］
前記第１の入力デバイスと前記出力デバイスとは、仮想現実ヘッドセットまたは拡張現実ヘッドセットに組み込まれる、Ｃ１に記載のデバイス。
［Ｃ１７］
前記第１の入力デバイスと前記出力デバイスとは、車両に組み込まれる、Ｃ１に記載のデバイス。
［Ｃ１８］
マルチモーダルユーザ入力のための方法であって、
デバイスの１つまたは複数のプロセッサにおいて、第１の入力デバイスから受信された第１のデータを処理することと、前記第１のデータは、第１の入力モードに基づくユーザからの第１の入力を示し、前記第１の入力は、コマンドに対応し、
前記１つまたは複数のプロセッサから、前記第１のデータを処理することに基づいて出力デバイスにフィードバックメッセージを送ることと、ここにおいて、前記フィードバックメッセージは、前記第１の入力モードとは異なる第２の入力モードに基づいて、前記第１の入力に関連するコマンドを識別する第２の入力を提供するように前記ユーザに命令する、
前記１つまたは複数のプロセッサにおいて、第２の入力デバイスから第２のデータを受信することと、前記第２のデータは、前記第２の入力を示し、
前記１つまたは複数のプロセッサにおいて、前記第２の入力によって識別される前記コマンドに前記第１の入力を関連付けるようにマッピングを更新することと、
を備える、方法。
［Ｃ１９］
前記第１の入力モードは、スピーチモード、ジェスチャーモード、またはビデオモードのうちの１つであり、前記第２の入力モードは、前記スピーチモード、前記ジェスチャーモード、または前記ビデオモードのうちの異なる１つである、Ｃ１８に記載の方法。
［Ｃ２０］
前記フィードバックメッセージは、前記第１の入力をディスアンビギュエートするために前記第２の入力を提供するように前記ユーザに命令する、Ｃ１８に記載の方法。
［Ｃ２１］
前記フィードバックメッセージは、前記第１の入力の認識処理に関連する確信度レベルが確信度しきい値を満たすことに失敗したことに応答して送られる、Ｃ２０に記載の方法。
［Ｃ２２］
前記更新されたマッピングは、前記第１の入力と前記第２の入力との組合せを前記コマンドに関連付ける、Ｃ１８に記載の方法。
［Ｃ２３］
前記マッピングを更新することは、
前記ユーザに関連する埋め込みネットワークデータを更新すること、または

前記ユーザに関連する重みデータを更新すること、
のうちの少なくとも１つを含む、Ｃ１８に記載の方法。
［Ｃ２４］
前記第１の入力モードは、ビデオモードに対応し、前記フィードバックメッセージは、照明しきい値を下回る値を有する周辺光メトリックに応答して送られる、Ｃ１８に記載の方法。
［Ｃ２５］
前記第１の入力モードは、スピーチモードに対応し、前記フィードバックメッセージは、雑音しきい値を超える値を有する雑音メトリックに応答して送られる、Ｃ１８に記載の方法。
［Ｃ２６］
マルチモーダルユーザ入力のための装置であって、
第１の入力デバイスから受信された第１のデータを処理するための手段と、前記第１のデータは、第１の入力モードに基づくユーザからの第１の入力を示し、前記第１の入力は、コマンドに対応し、
前記第１のデータを処理することに基づいて出力デバイスにフィードバックメッセージを送るための手段と、ここにおいて、前記フィードバックメッセージは、前記第１の入力モードとは異なる第２の入力モードに基づいて、前記第１の入力に関連するコマンドを識別する第２の入力を提供するように前記ユーザに命令する、
第２の入力デバイスから第２のデータを受信するための手段と、前記第２のデータは、前記第２の入力を示し、
前記第２の入力によって識別される前記コマンドに前記第１の入力を関連付けるようにマッピングを更新するための手段と、
を備える、装置。
［Ｃ２７］
前記更新されたマッピングは、前記第１の入力と前記第２の入力との組合せを前記コマンドに関連付ける、Ｃ２６に記載の装置。
［Ｃ２８］
デバイスの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の入力デバイスから受信された第１のデータを処理することと、前記第１のデータは、第１の入力モードに基づくユーザからの第１の入力を示し、前記第１の入力が、コマンドに対応し、
前記第１のデータを処理することに基づいて出力デバイスにフィードバックメッセージを送ることと、ここにおいて、前記フィードバックメッセージは、前記第１の入力モードとは異なる第２の入力モードに基づいて、前記第１の入力に関連するコマンドを識別する第２の入力を提供するように前記ユーザに命令する、
第２の入力デバイスから第２のデータを受信することと、前記第２のデータは、前記第２の入力を示し、
前記第２の入力によって識別される前記コマンドに前記第１の入力を関連付けるようにマッピングを更新することと、
を行わせる命令を備える非一時的コンピュータ可読媒体。
［Ｃ２９］
前記第１の入力モードは、ビデオモードに対応し、前記フィードバックメッセージは、照明しきい値を下回る値を有する周辺光メトリックに応答して送られる、Ｃ２８に記載の非一時的コンピュータ可読媒体。
［Ｃ３０］
前記第１の入力モードは、スピーチモードに対応し、前記フィードバックメッセージは、雑音しきい値を超える値を有する雑音メトリックに応答して送られる、Ｃ２８に記載の非一時的コンピュータ可読媒体。