

(19)日本国特許庁(JP)

(12)公開特許公報(A)

(11)公開番号

特開2023-2650  
(P2023-2650A)

(43)公開日 令和5年1月10日(2023.1.10)

(51)国際特許分類

F I

H 0 4 M	3/428(2006.01)	H 0 4 M	3/428
G 1 0 L	15/22 (2006.01)	G 1 0 L	15/22 3 0 0 Z
G 1 0 L	13/00 (2006.01)	G 1 0 L	13/00 1 0 0 M
H 0 4 M	3/50 (2006.01)	H 0 4 M	3/50 A
H 0 4 M	3/56 (2006.01)	H 0 4 M	3/56 B

審査請求 有 請求項の数 13 O L 外国語出願 (全37頁)

(21)出願番号 特願2022-165657(P2022-165657)  
 (22)出願日 令和4年10月14日(2022.10.14)  
 (62)分割の表示 特願2021-544219(P2021-544219)の分割  
 原出願日 令和2年5月5日(2020.5.5)  
 (31)優先権主張番号 62/843,660  
 (32)優先日 令和1年5月6日(2019.5.6)  
 (33)優先権主張国・地域又は機関 米国(US)

(71)出願人 502208397  
 グーグル エルエルシー  
 Google LLC  
 アメリカ合衆国 カリフォルニア州 9 4  
 0 4 3 マウンテン ビュー アンフィシ  
 アター パークウェイ 1 6 0 0  
 1 6 0 0 Amphitheatre P  
 arkway 9 4 0 4 3 Mounta  
 in View, CA U.S.A.  
 (74)代理人 100108453  
 弁理士 村山 靖彦  
 (74)代理人 100110364  
 弁理士 実広 信哉  
 (74)代理人 100133400  
 弁理士 阿部 達彦

(特許庁注：以下のものは登録商標)  
1. BLUETOOTH

最終頁に続く

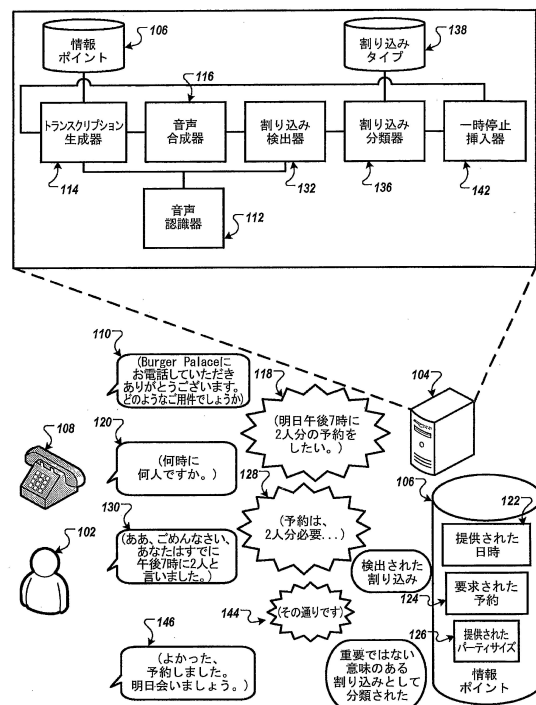
(54)【発明の名称】 自動通話システム

(57)【要約】

【課題】本発明の目的は、自動通話システムを提供することである。

【解決手段】自動通話システムのための方法、システム、および装置が開示される。いくつかの実装形態は、電話通話を開始し、ユーザとの電話での会話を行うためにボットを使用することに向けられている。ボットは、電話通話中に合成音声を提供している間、割り込まれる場合がある。割り込みは、複数の異なる割り込みタイプのうちの1つに分類することができ、ボットは、割り込みタイプに基づいて割り込みに反応することができる。いくつかの実装形態は、電話での会話中に第1のユーザが第2のユーザによって保留にされたことを判定し、第1のユーザが電話通話を切ったとの判定に回答して電話通話をアクティブ状態に維持することに向けられている。第1のユーザには、第2のユーザが通話に再参加したときに通知することができ、第1のユーザに関連付けられたボットは、第2のユーザが電話通話に再参加したことを第1のユーザに通知することができる。

【選択図】図1



## 【特許請求の範囲】

## 【請求項 1】

1つまたは複数のプロセッサによって実装される方法であって、

ボットを使用してユーザとの電話通話を開始するステップであって、前記ボットが、電話通話を開始し、電話での会話を行うように構成されている、ステップと、

前記ユーザの対応するコンピューティングデバイスにおける出力のために、前記ボットの合成音声を提供するステップと、

前記ボットの前記合成音声を提供している間、

前記ユーザから、前記ボットの前記合成音声に割り込むユーザ発話を受信するステップと、

前記合成音声に割り込む前記ユーザ発話の受信に応答して、前記受信したユーザ発話を所与のタイプの割り込みとして分類するステップであって、前記所与のタイプの割り込みが、複数の異なるタイプの割り込みのうちの1つである、ステップと、

前記所与のタイプの割り込みに基づいて、前記ユーザの前記対応するコンピューティングデバイスにおける出力のために、前記ボットの前記合成音声を提供し続けるかどうかを決定するステップと

を含む、方法。

## 【請求項 2】

前記所与のタイプの割り込みが無意味な割り込みであり、前記受信したユーザ発話を前記無意味な割り込みとして分類するステップが、

前記受信したユーザ発話が、背景雑音、肯定の単語もしくは句、またはフィラー単語もしくは句のうちの1つまたは複数を含むことを判定するために、前記受信したユーザ発話に対応するオーディオデータまたは前記受信したユーザ発話に対応するトランスクリプションを処理するステップと、

前記受信したユーザ発話が、背景雑音、肯定の単語もしくは句、またはフィラー単語もしくは句のうちの1つまたは複数を含むと判定したことに基づいて、前記受信したユーザ発話を前記無意味な割り込みとして分類するステップと

を含む、

請求項1に記載の方法。

## 【請求項 3】

前記ボットの前記合成音声を提供し続けるかどうかを決定するステップが、

前記受信したユーザ発話を前記無意味な割り込みとして分類したことに基づいて、前記ボットの前記合成音声を提供し続けると判定するステップを含む、

請求項2に記載の方法。

## 【請求項 4】

前記所与のタイプの割り込みが重要ではない意味のある割り込みであり、前記受信したユーザ発話を前記重要ではない意味のある割り込みとして分類するステップが、

前記ボットによって知られており、まだ提供されていない情報に対する要求を前記受信したユーザ発話を含むことを判定するために、前記受信したユーザ発話に対応するオーディオデータまたは前記受信したユーザ発話に対応するトランスクリプションを処理するステップと、

前記ボットによって知られており、まだ提供されていない前記情報に対する前記要求を前記受信したユーザ発話を含むと判定したことに基づいて、前記受信したユーザ発話を前記重要ではない意味のある割り込みとして分類するステップと

を含む、

請求項1に記載の方法。

## 【請求項 5】

前記ボットの前記合成音声を提供し続けるかどうかを決定するステップが、

前記ユーザ発話を前記重要ではない意味のある割り込みとして分類したことに基づいて、出力のために前記ボットの前記合成音声を提供することを中止するための、前記合成音

10

20

30

40

50

声の残りの部分における時間的ポイントを決定するステップと、

前記合成音声の前記残りの部分が前記受信した発話に応答するかどうかを判定するステップと、

前記残りの部分が前記受信したユーザ発話に応答しないとの判定に応答して、

出力のために、前記受信したユーザ発話に応答した、まだ提供されていない前記合成音声の追加部分を提供するステップと、

出力のために、前記合成音声の前記追加部分を提供した後、出力のために、前記時間的ポイントから前記ボットの前記合成音声の前記残りの部分を提供し続けるステップとを含む、

請求項4に記載の方法。

10

【請求項6】

前記残りの部分が前記受信したユーザ発話に応答するとの判定に応答して、

出力のために、前記時間的ポイントから前記ボットの前記合成音声の前記残りの部分を提供し続けるステップをさらに含む、

請求項5に記載の方法。

【請求項7】

前記所与のタイプの割り込みが重要な意味のある割り込みであり、前記受信したユーザ発話を前記重要な意味のある割り込みとして分類するステップが、

前記受信したユーザ発話が、前記ボットが前記合成音声を繰り返すことに対する要求、または前記ボットを保留にする要求を含むことを判定するために、前記受信したユーザ発話に対応するオーディオデータまたは前記受信したユーザ発話に対応するトランスクリプションを処理するステップと、

20

前記受信したユーザ発話が、前記ボットが前記合成音声を繰り返すことに対する前記要求、または前記ボットを保留にする前記要求を含むと判定したことに基づいて、前記受信したユーザ発話を前記重要ではない意味のある割り込みとして分類するステップとを含む、

請求項1に記載の方法。

【請求項8】

前記ボットの前記合成音声を提供し続けるかどうかを決定するステップが、

出力のために、前記ボットの前記合成音声の現在の単語または用語の残りの部分を提供するステップと、

30

出力のために、前記現在の単語または用語の前記残りの部分を提供した後に、出力のために、前記ボットの前記合成音声を提供することを中止するステップとを含む、

請求項7に記載の方法。

【請求項9】

前記受信したユーザ発話を前記所与のタイプの割り込みとして分類するステップが、

前記所与のタイプの割り込みを判定するために、機械学習モデルを使用して、前記受信したユーザ発話に対応するオーディオデータまたは前記受信したユーザ発話に対応するトランスクリプションを処理するステップを含む、

40

請求項1に記載の方法。

【請求項10】

複数のトレーニングインスタンスを使用して前記機械学習モデルをトレーニングするステップをさらに含み、前記トレーニングインスタンスの各々が、トレーニングインスタンス入力と、対応するトレーニングインスタンス出力とを含み、

各トレーニングインスタンス入力が、割り込み発話に対応するトレーニングオーディオデータまたは前記割り込み発話に対応するトランスクリプションを含み、

各対応するトレーニングインスタンス出力が、前記割り込み発話内に含まれる割り込みの前記タイプに対応するグラウンドトゥールズラベルを含む、

請求項9に記載の方法。

50

## 【請求項 1 1】

前記機械学習モデルを使用して、前記受信したユーザ発話に対応する前記オーディオデータまたは前記受信したユーザ発話に対応する前記トランスクリプションを処理するステップが、前記ユーザ発話の前記オーディオデータまたは前記トランスクリプションとともに受信されたときに出力されている前記合成音声を処理するステップをさらに含む、請求項9に記載の方法。

## 【請求項 1 2】

前記受信したユーザ発話を前記所与のタイプの割り込みとして分類するステップが、前記受信したユーザ発話のトークンを、前記複数の異なる割り込みタイプの各々に関連付けられた1つまたは複数の用語と一致させる1つまたは複数のルールを使用して、前記受信したユーザ発話に対応するオーディオデータまたは前記受信したユーザ発話に対応するトランスクリプションを処理するステップを含む、請求項1に記載の方法。

10

## 【請求項 1 3】

前記ボットを使用して前記ユーザとの前記電話通話を開始するステップが、前記電話通話を開始するために、前記ボットに関連付けられた所与のユーザからユーザ入力を受信することに応答する、請求項1に記載の方法。

## 【請求項 1 4】

前記電話通話を開始するための前記ユーザ入力が、前記ユーザの前記対応するコンピューティングデバイスにおける出力のために提供される前記合成音声内に含まれるべきである情報ポイントを含む、請求項13に記載の方法。

20

## 【請求項 1 5】

1つまたは複数のプロセッサによって実装される方法であって、  
 第1のユーザおよび第2のユーザが電話通話に参与していることを判定するステップであって、前記第1のユーザが対応する第1のコンピューティングデバイスに関連付けられており、前記第2のユーザが対応する第2のコンピューティングデバイスに関連付けられている、ステップと、  
 前記第2のユーザが前記第1のユーザを保留にしたことを判定するステップと、  
 前記第1のユーザが前記対応する第1のコンピューティングデバイスを切ったことを判定するステップと、  
 前記第1のユーザが前記対応する第1の電話を切ったとの判定に応答して、前記第1のユーザが保留にされている間、  
 前記電話通話をアクティブ状態に維持するステップと、  
 前記第2のユーザが前記電話通話に再参加したかどうかを判定するステップと  
 前記第2のユーザが前記電話通話に再参加したとの判定に応答して、  
 前記第2のユーザとの前記電話通話を継続するために、前記第1のユーザの前記対応する第1のコンピューティングデバイスに関連付けられたボットを使用するステップと、  
 前記対応する第1のコンピューティングデバイスにおける出力のために、前記第1のユーザが前記電話通話に再参加するための通知を提供するステップと、  
 前記第1のユーザが前記電話通話に再参加したことを判定するステップと、  
 前記第1のユーザが前記電話通話に再参加したとの判定に応答して、前記第2のユーザと前記ボットとの間の前記電話での会話を中止するステップと  
 を含む、  
 方法。

30

40

## 【請求項 1 6】

前記第2のユーザとの前記電話通話を継続するために、前記第1のユーザの前記対応する第1のコンピューティングデバイスに関連付けられた前記ボットを使用するステップが、  
 前記第1のユーザが電話通話におけるアクティブな参加者ではないことを示す前記ボットの合成音声を生成するステップと、

50

前記第2のユーザの前記対応する第2のコンピューティングデバイスにおける出力のために、前記ボットの前記合成音声を提供するステップとを含む、  
請求項15に記載の方法。

【請求項17】

前記第1のユーザが前記電話通話に再参加する前に、前記第2のユーザからユーザ発話を受信するステップと、

前記ユーザ発話のトランスクリプションを生成するステップと、

前記ユーザ発話の前記トランスクリプションを、前記第1のユーザが前記電話通話に再参加するための前記通知内に含めるステップと

10

をさらに含む、

請求項15に記載の方法。

【請求項18】

前記電話通話を前記アクティブ状態に維持するステップが、

前記対応する第1のコンピューティングデバイスの1つまたは複数のマイクロフォンをミュートするステップと、

前記対応する第1のコンピューティングデバイスと前記対応する第2のコンピューティングデバイスとの間の電話接続を維持するステップと

を含む、

請求項15に記載の方法。

20

【請求項19】

1つまたは複数のコンピュータと、

前記1つまたは複数のコンピュータによって実行されたときに、前記1つまたは複数のコンピュータに請求項1から18のいずれか一項に記載の方法を実行させるように動作可能な命令を記憶した1つまたは複数の記憶デバイスとを備えるシステム。

【請求項20】

1つまたは複数のコンピュータによって実行可能であり、実行時に、前記1つまたは複数のコンピュータに請求項1から18のいずれか一項に記載の方法を実行させる命令を備えるソフトウェアを記憶した非一時的なコンピュータ可読記録媒体。

30

【発明の詳細な説明】

【技術分野】

【0001】

関連出願の相互参照

2016年6月13日に出願した米国特許出願第62/349,396号および2017年6月13日に  
出願した米国特許出願第15/621,869号の内容は、参照によりその全体が組み込まれる。

【0002】

本明細書は、自然言語処理に関する。

【背景技術】

40

【0003】

ユーザは、人間の対話なしでは容易に取得されないタイプの情報を収集する必要がある場合がある。例えば、複数の事業所または組織からのデータを検証または収集するために、ユーザは、情報を収集するために事業所または組織の各々に電話をかける必要がある場合がある。ウェブ検索エンジンは、サービスまたは事業所のための連絡先情報を提供することによってそのようなタスクでユーザを支援することができるが、ユーザは、依然として自分自身でタスクを完了するためにサービスまたは事業所に電話をかけなければならない。

【0004】

複数の事業所または組織から収集された情報のデータベースを維持するために、人間の

50

オペレータは、データを収集するために多数の事業所への自動通話を開始することができるが、通話先(例えば、同じ料理を提供する特定の町のすべてのレストラン)を選択し、電話をかけることは、手動で実行する場合、時間がかかる可能性がある。さらに、いつ電話をかけるか、または電話をかけるかどうかを決定することは、一般に、検証、更新、または補足情報の必要性を確認するために、既存のデータの人間の分析を必要とする。

【0005】

ユーザは、予約を行う、またはサービスを採用するなどのタスクを実行することを望む場合もある。しかしながら、一般に、ユーザが所望のタスクを完了するために対話しなければならない人が存在する。例えば、ユーザは、ウェブサイトを持たない小さいレストランにおいて予約を行うために、ホステスに電話をかけて話をする必要がある場合がある。場合によっては、ユーザが自分で電話をかけたとしても、ユーザは、しばしば限られたユーザの応答のセットのみを受け付ける自動連絡網に遭遇する場合がある。

10

【発明の概要】

【課題を解決するための手段】

【0006】

いくつかの実装形態は、電話通話を開始し、ユーザとの電話での会話を行うためにボットを使用することに向けられている。ボットは、電話通話中に合成音声を提供している間、割り込まれる場合がある。割り込みは、複数の異なる割り込みタイプのうちの1つに分類することができ、ボットは、割り込みタイプに基づいて割り込みに反応することができる。いくつかの実装形態は、第1のユーザが電話での会話中に第2のユーザによって保留にされたと判定し、第1のユーザが電話を切ったとの判定にตอบสนองして、電話通話をアクティブ状態に維持することに向けられている。第1のユーザには、第2のユーザが通話に再参加したときに通知することができ、第1のユーザに関連付けられたボットは、第2のユーザが電話通話に再参加したことを第1のユーザに通知することができる。

20

【0007】

いくつかの実装形態において、1つまたは複数のプロセッサによって実装される方法が提供され、方法は、電話通話を開始し、電話での会話を行うように構成されたボットを使用してユーザとの電話通話を開始するステップと、ユーザの対応するコンピューティングデバイスにおける出力のために、ボットの合成音声を提供するステップとを含む。方法は、ボットの合成音声を提供している間、ユーザから、ボットの合成音声に割り込むユーザ発話を受信するステップと、合成音声に割り込むユーザ発話の受信にตอบสนองして、受信したユーザ発話を、複数の異なるタイプの割り込みのうちの所与のタイプの割り込みとして分類するステップと、所与のタイプの割り込みに基づいて、ユーザの対応するコンピューティングデバイスにおける出力のために、ボットの合成音声を提供し続けるかどうかを決定するステップとをさらに含む。

30

【0008】

本明細書で開示されている技術のこれらおよび他の実装形態は、以下の特徴のうちの1つまたは複数を含むことができる。

【0009】

いくつかの実装形態において、所与のタイプの割り込みは、無意味な割り込みである。受信したユーザ発話を無意味な割り込みとして分類するステップは、受信したユーザ発話が、背景雑音、肯定の単語もしくは句、またはフィラー単語もしくは句のうちの1つまたは複数を含むことを判定するために、受信したユーザ発話に対応するオーディオデータまたは受信したユーザ発話に対応するトランスクリプションを処理するステップと、受信したユーザ発話が、背景雑音、肯定の単語もしくは句、またはフィラー単語もしくは句のうちの1つまたは複数を含むと判定したことに基づいて、受信したユーザ発話を無意味な割り込みとして分類するステップとを含む。

40

【0010】

それらの実装形態のいくつかのバージョンにおいて、ボットの合成音声を提供し続けるかどうかを決定するステップは、受信したユーザ発話を無意味な割り込みとして分類した

50

ことに基づいて、ボットの合成音声を提供し続けると判定するステップを含む。

【0011】

いくつかの実装形態において、所与のタイプの割り込みは、重要ではない意味のある割り込みである。受信したユーザ発話を重要ではない意味のある割り込みとして分類するステップは、ボットによって知られており、まだ提供されていない情報に対する要求を、受信したユーザ発話を含むことを判定するために、受信したユーザ発話に対応するオーディオデータまたは受信したユーザ発話に対応するトランスクリプションを処理するステップと、ボットによって知られており、まだ提供されていない情報に対する要求を受信したユーザ発話を含むと判定したことに基づいて、受信したユーザ発話を重要ではない意味のある割り込みとして分類するステップとを含む。

10

【0012】

それらの実装形態のいくつかのバージョンにおいて、ボットの合成音声を提供し続けるかどうかを決定するステップは、ユーザ発話を重要ではない意味のある割り込みとして分類したことに基づいて、出力のためにボットの合成音声を提供することを中止するための、合成音声の残りの部分における時間的ポイントを決定するステップと、合成音声の残りの部分が受信した発話に回答するかどうかを判定するステップと、残りの部分が受信したユーザ発話に回答しないとの判定に回答して、出力のために、受信したユーザ発話に回答した、まだ提供されていない合成音声の追加部分を提供するステップと、出力のために、合成音声の追加部分を提供した後、出力のために、時間的ポイントからボットの合成音声の残りの部分を提供し続けるステップとを含む。

20

【0013】

それらの実装形態のいくつかのさらなるバージョンにおいて、方法は、残りの部分が受信したユーザ発話に回答するとの判定に回答して、出力のために、時間的ポイントからボットの合成音声の残りの部分を提供し続けるステップをさらに含む。

【0014】

いくつかの実装形態において、所与のタイプの割り込みは、重要な意味のある割り込みである。受信したユーザ発話を重要な意味のある割り込みとして分類するステップは、受信したユーザ発話が、ボットが合成音声を繰り返すことに対する要求、またはボットを保留にする要求を含むことを判定するために、受信したユーザ発話に対応するオーディオデータまたは受信したユーザ発話に対応するトランスクリプションを処理するステップと、受信したユーザ発話が、ボットが合成音声を繰り返すことに対する要求、またはボットを保留にする要求を含むと判定したことに基づいて、受信したユーザ発話を重要ではない意味のある割り込みとして分類するステップとを含む。

30

【0015】

それらの実装形態のいくつかのバージョンにおいて、ボットの合成音声を提供し続けるかどうかを決定するステップは、出力のために、ボットの合成音声の現在の単語または用語の残りの部分を提供するステップと、出力のために、現在の単語または用語の残りの部分を提供した後、出力のために、ボットの合成音声を提供することを中止するステップとを含む。

【0016】

いくつかの実装形態において、受信したユーザ発話を所与のタイプの割り込みとして分類するステップは、所与のタイプの割り込みを判定するために、機械学習モデルを使用して、受信したユーザ発話に対応するオーディオデータまたは受信したユーザ発話に対応するトランスクリプションを処理するステップを含む。

40

【0017】

それらの実装形態のいくつかのバージョンにおいて、方法は、複数のトレーニングインスタンスを使用して機械学習モデルをトレーニングするステップをさらに含む。トレーニングインスタンスの各々は、トレーニングインスタンス入力と、対応するトレーニングインスタンス出力とを含み、各トレーニングインスタンス入力は、割り込み発話に対応するトレーニングオーディオデータまたは割り込み発話に対応するトランスクリプションを含

50

み、各対応するトレーニングインスタンス出力は、割り込み発話内に含まれる割り込みのタイプに対応するグラウンドトゥーラベルを含む。それらの実装形態のいくつかのさらなるバージョンにおいて、機械学習モデルを使用して、受信したユーザ発話に対応するオーディオデータまたは受信したユーザ発話に対応するトランスクリプションを処理するステップは、ユーザ発話がオーディオデータまたはトランスクリプションとともに受信されたときに出力されている合成音声を処理するステップをさらに含む。

【0018】

いくつかの実装形態において、受信したユーザ発話を所与のタイプの割り込みとして分類するステップは、受信したユーザ発話のトークンを、複数の異なる割り込みタイプの各々に関連付けられた1つまたは複数の用語と一致させる1つまたは複数のルールを使用し、受信したユーザ発話に対応するオーディオデータまたは受信したユーザ発話に対応するトランスクリプションを処理するステップを含む。

10

【0019】

いくつかの実装形態において、ボットを使用してユーザとの電話通話を開始するステップは、電話通話を開始するために、ボットに関連付けられた所与のユーザからユーザ入力を受信することに対応する。それらの実装形態のいくつかのバージョンにおいて、電話通話を開始するためのユーザ入力は、ユーザの対応するコンピューティングデバイスにおける出力のために提供される合成音声内に含まれるべきである情報ポイントを含む。

【0020】

いくつかの実装形態において、1つまたは複数のプロセッサによって実装される方法が提供され、方法は、第1のユーザおよび第2のユーザが電話通話に参与していることを判定するステップを含む。第1のユーザは、対応する第1のコンピューティングデバイスに関連付けられており、第2のユーザは、対応する第2のコンピューティングデバイスに関連付けられている。方法は、第2のユーザが第1のユーザを保留にしたことを判定するステップと、第1のユーザが対応する第1のコンピューティングデバイスを切ったことを判定するステップと、第1のユーザが対応する第1のデバイスを切ったとの判定に対応して、第1のユーザが保留にされている間、電話通話をアクティブ状態に維持するステップと、第2のユーザが電話通話に再参加したかどうかを判定するステップとをさらに含む。方法は、第2のユーザが電話通話に再参加したと判定したことに応答して、第2のユーザとの電話通話を継続するために、第1のユーザの対応する第1のコンピューティングデバイスに関連付けられたボットを使用するステップと、対応する第1のコンピューティングデバイスにおける出力のために、第1のユーザが電話通話に再参加するための通知を提供するステップと、第1のユーザが電話通話に再参加したことを判定するステップと、第1のユーザが電話通話に再参加したとの判定に対応して、第2のユーザとボットとの間の電話での会話を中止するステップとをさらに含む。

20

30

【0021】

本明細書で開示されている技術のこれらおよび他の実装形態は、以下の特徴のうちの1つまたは複数を含むことができる。

【0022】

いくつかの実装形態において、第2のユーザとの電話通話を継続するために、第1のユーザの対応する第1のコンピューティングデバイスに関連付けられたボットを使用するステップは、第1のユーザが電話通話におけるアクティブな参加者ではないことを示すボットの合成音声を生成するステップと、第2のユーザの対応する第2のコンピューティングデバイスにおける出力のために、ボットの合成音声を提供するステップとを含む。

40

【0023】

いくつかの実装形態において、方法は、第1のユーザが電話通話に再参加する前に、第2のユーザからユーザ発話を受信するステップと、ユーザ発話のトランスクリプションを生成するステップと、ユーザ発話のトランスクリプションを、第1のユーザが電話通話に再参加するための通知内に含めるステップとをさらに含む。

【0024】

50



いくつかの実装形態において、電話通話をアクティブ状態に維持するステップは、対応する第1のコンピューティングデバイスの1つまたは複数のマイクロフォンをミュートするステップと、対応する第1のコンピューティングデバイスと対応する第2のコンピューティングデバイスとの間の電話接続を維持するステップとを含む。

【0025】

加えて、いくつかの実装形態は、1つまたは複数のコンピューティングデバイスの1つまたは複数のプロセッサ(例えば、中央処理装置(CPU)、グラフィカル処理ユニット(GPU)、および/またはテンソル処理ユニット(TPU))を含み、1つまたは複数のプロセッサは、関連するメモリ内に記憶された命令を実行するように動作可能であり、命令は、前述の方法のいずれかの実行を引き起こすように構成される。いくつかの実装形態は、前述の方法のいずれかを実行するために1つまたは複数のプロセッサによって実行可能なコンピュータ命令を記憶する1つまたは複数の非一時的コンピュータ可読記憶媒体も含む。いくつかの実装形態は、前述の方法を実行するために1つまたは複数のプロセッサによって実行可能な命令を含むコンピュータプログラム製品も含む。

10

【0026】

本明細書でより詳細に説明する前述の概念および追加の概念のすべての組合せは、本明細書で開示されている主題の一部であると考えられることが理解されるべきである。例えば、本開示の終わりに現れる特許請求された主題のすべての組合せは、本明細書で開示されている主題の一部であると考えられる。

【0027】

20

上記の説明は、本明細書で開示されているいくつかの実装形態のみの概要として提供されている。それらの実装形態、および他の実装形態について、本明細書でさらに詳細に説明する。

【図面の簡単な説明】

【0028】

【図1】ユーザおよび通話開始システムのボットが電話での会話を行っている間にユーザから受信した割り込みを処理するための例示的なシステムを示す図である。

【図2】あるユーザが保留中である電話通話を監視し、他のユーザが電話通話に再参加したときに保留中のユーザに通知する例示的なシステムを示す図である。

【図3A】ユーザおよび通話開始システムのボットが電話での会話を行っている間にユーザから受信した割り込みを処理するための例示的なシステムのフローチャートである。

30

【図3B】あるユーザが保留中である電話通話を監視し、他のユーザが通話に再参加したときに保留中のユーザに通知するための例示的なプロセスのフローチャートである。

【図4】コンピューティングデバイスおよびモバイルコンピューティングデバイスの例を示す図である。

【発明を実施するための形態】

【0029】

様々な図面中の同じ参照番号および名称は、同じ要素を示す。

【0030】

図1は、人間の代表者102および通話開始システム104のボットが電話での会話を行っている間に代表者102から受信した割り込みを処理するための例示的なシステム100を示す。簡単に、以下でより詳細に説明するように、通話開始システム104は、電話通話を行う(例えば、ユーザの代わりに補助通話(assisted call)を開始することによって、ユーザのためのタスクを実行している。代表者102は、ユーザの代わりに電話通話に回答し、通話開始システム104のボットとの電話での会話を行う。電話通話中に、代表者102は、ボットに割り込む場合がある(例えば、補助通話中)。ボットは、割り込みを識別し、割り込みを複数の異なる割り込みタイプのうちの1つに分類し、割り込みのタイプに基づいてユーザの代わりに電話での会話を継続することができる。

40

【0031】

図1に示すように、ユーザは、デジタルアシスタントが翌日の午後7時に2人分Burger

50

Palaceの予約を行うことを要求することによって、デジタルアシスタントと対話してもよい。Burger Palaceは、デジタルアシスタントが予約を要求することができるオンラインポータルを持たない小さいレストランである場合がある。代わりに、将来の顧客は、予約を行うためにBurger Palaceに電話をかけなければならない。この場合、デジタルアシスタントは、通話開始システム104がユーザの代わりにレストランとの電話通話を開始し、予約を行うことを要求する。図1に示すシステム100は、ネットワーク(例えば、LAN、WAN、WiFi、Bluetooth、および/または他のネットワーク)上のサーバによって実装されるものとして示されているが、それは、例のためのものであり、限定することを意図したものではないことが理解されるべきである。例えば、システム100は、補助通話を開始したユーザのコンピューティングデバイスにおいてローカルに実装することができ、ならびに/またはネットワーク上に分散された方法でコンピューティングデバイスおよびサーバによって実装することができる。

#### 【0032】

デジタルアシスタントは、ユーザによって提供された要求内に含まれるデータを通話開始システム104に提供することができる。情報は、予約の要求された日時(例えば、明日の午後7時)と、要求された事業所(例えば、Burger Palace)と、パーティ内の人数(例えば、2人)とを含んでもよい。レストランの予約以外の要求について、情報は、要求されたサービス提供者(例えば、航空会社、公益事業提供者、および/または任意の他のサービス提供者)の名前、サービス提供者に対する要求の説明(例えば、サービス/予約を行う/変更する/中止する)、および/またはユーザの代わりにタスクを実行する際に代表者102によって求められる可能性がある任意の他の情報を含んでもよい。通話開始システム104は、このデータの情報ポイント106として記憶してもよい。情報ポイント106は、例えば、通話開始システム104が電話通話中に代表者102に提供すべき情報、または通話開始システム104が、代表者102が電話通話中に提供することを要求すべき情報を含む。

#### 【0033】

例えば、通話開始システム104がBurger Palaceとの電話通話を開始すると仮定し、代表者102が電話108に応答すると仮定する。さらに、代表者102が「Thank you for calling Burger Palace. How can I help you?(Burger Palaceにお電話していただきありがとうございます。どのようなご用件でしょうか。)」という発話110を話すと仮定する。通話開始システム104のポットは、発話110を検出し、発話110のオーディオデータを音声認識器112に提供することができる。音声認識器112は、発話110のトランスクリプションを生成し、発話110のトランスクリプションをトランスクリプション生成器114に提供することができる。

#### 【0034】

トランスクリプション生成器114は、代表者102の発話110に回答するトランスクリプションを生成することができる。さらに、トランスクリプション生成器114は、情報ポイント106のうちの1つに含まれる情報を提供することが代表者102の発話110に対する適切な応答であるかどうかを判定するために、情報ポイント106にアクセスしてもよい。情報ポイント106のうちの1つに含まれる情報が代表者102の発話110に対する適切な応答であるかどうかを判定するために、トランスクリプション生成器114は、ルールベースの手法および/または機械学習ベースの手法を含む、様々な技法を使用してもよい。いくつかの実装形態において、トランスクリプション生成器114は、代表者102の発話110のトランスクリプション内のキーワードまたは句を識別してもよい。トランスクリプション生成器114は、代表者102の発話110のトランスクリプションをトークン化し、トークン化された用語の中からキーワードを識別してもよい。トランスクリプション生成器114は、次いで、代表者102の発話110の主題を決定するために、それらのキーワードまたは句を使用してもよい。トランスクリプション生成器114は、応答のトランスクリプションを生成するために、代表者102の発話110の主題を使用してもよい。

#### 【0035】

いくつかの実装形態において、トランスクリプション生成器114は、代表者102の発話110の主題および/または適切な応答を決定するために、機械学習を使用してトレーニングされたモデルを使用する。通話開始システム104は、以前の会話のログを含むトレーニングデータにアクセスしてもよい。以前の会話は、レストラン事業所、航空事業所、政府機関などの事業所もしくは組織のタイプに固有のもの、および/または他の事業所もしくは組織に固有の会話であってもよい。対応する会話内の発話の各々は、キーワードラベルを含む場合がある。キーワードラベルは、発話内の用語、発話内の用語の意味的表現、および/または発話もしくはそのトランスクリプションに注釈を付けるための他のタイプのラベルを含むことができる。通話開始システム104は、発話をキャプチャするオーディオデータ(および/またはそのトランスクリプション)が機械学習モデル全体の入力として適用されたときに発話のキーワードを識別するために機械学習モデルをトレーニングするためにトレーニングデータを使用することができる。いくつかの追加のおよび/または代替の実装形態において、通話開始システム104は、発話をキャプチャするオーディオデータ(および/またはそのトランスクリプション)が機械学習モデル全体の入力として適用されたときに適切な応答のためのキーワードを生成するために機械学習モデルをトレーニングするためにトレーニングデータを使用することができる。

10

20

30

40

50

**【0036】**

トランスクリプション生成器114は、情報ポイント106と組み合わせて発話110に対する適切な応答を生成するために、これらの機械学習モデルおよび/またはルールベースの手法の任意の組合せを使用してもよい。より具体的には、トランスクリプション生成器114は、適切な応答を生成するために機械学習モデルを使用し、現在の会話に固有である可能性がある任意の詳細を埋めるために情報ポイント106を使用してもよい。図1に示すように、トランスクリプション生成器114は、トランスクリプション生成器114が情報ポイント106のうちの一つまたは複数を適用する応答シェルを生成するために、代表者102の発話110のトランスクリプションを分析することができる。例えば、トランスクリプション生成器114は、「I'd like to blank (空所にしたい)」などの応答シェルを生成してもよい。トランスクリプション生成器114は、要求されたタスク(例えば、予約を行う)と、日時(例えば、明日の午後7時)と、パーティサイズ(例えば、2人)とを含む情報ポイント106で空所内を埋めてもよい。この例では、トランスクリプション生成器114は、代表者102の発話110のトランスクリプションを処理することに応答して、「I'd like to make a reservation for two people tomorrow at 7pm(明日午後7時に2人分の予約をしたい)」のトランスクリプションを生成することができる。

**【0037】**

トランスクリプション生成器114は、代表者102の発話110に応答するトランスクリプション生成器114によって生成されたトランスクリプションを音声合成器116に提供することができる。音声合成器116は、代表者102の発話110に応答する受信したトランスクリプションに対応するオーディオデータを含む合成音声を生成することができる。いくつかの実装形態において、音声合成器116は、合成音声をいくつかの異なる音声で出力するように構成されてもよい。例えば、音声合成器116は、予約を要求したユーザに似た音声、デジタルアシスタント用の予約された音声、トランスクリプション生成器114が適切な応答を生成することができない場合にボットを引き継いでもよいオペレータの音声、または予約を要求したユーザによって選択された、もしくは通話開始システム104によって選択された別の音声で、合成音声(または会話中のその一部)を出力するように構成されてもよい。

**【0038】**

通話開始システム104は、代表者102の発話110の受信に応答して、音声合成器116によって生成された「I'd like to make a reservation for two people tomorrow at 7pm(明日午後7時に2人分の予約をしたい)」の合成音声118を出力することができる。この時点で、トランスクリプション生成器114は、ボットがその情報を合成音声118の一部として代表者102に提供したので、提供された日時122、要求された予約124

、および提供されたパーティサイズ126の情報ポイント106を、満たされたとしてマークしてもよい。さらに、図1の例において、代表者102は、電話108を介して合成音声118を聞き、発話120に应答すると仮定する。発話102の検出に应答して、通話開始システム104は、発話110に関して上記で説明したのと同じまたは同様の方法において、代表者102の発話120のトランスクリプションを生成するために音声認識器112を使用し、適切な应答のトランスクリプションを生成するためにトランスクリプション生成器114を使用する。

#### 【0039】

この例では、トランスクリプション生成器114は、代表者102がすでに提供されている情報を要求していると判定する。トランスクリプション生成器114は、発話120のトランスクリプションが「what time(何時)」および「how many people(何人)」などのキーワードまたは句を含むと判定してもよい。トランスクリプション生成器114は、これらの2つの句が、合成音声118内に含まれた、提供された日時122および提供されたパーティサイズ126の情報ポイント106に対応すると判定してもよい。それにもかかわらず、トランスクリプション生成器114は、これらの特定の情報ポイント106を代表者102に再び提供するために、別のトランスクリプションを生成することができる。例えば、トランスクリプション生成器114は、発話112を受信したことに应答して「The reservation should be for two people at 7pm(予約は、午後7時に2人分必要です)」のトランスクリプションを生成してもよい。さらに、トランスクリプション生成器114は、このトランスクリプションを音声合成器116に提供することができる。

#### 【0040】

音声合成器116は、発話120に应答して生成されたトランスクリプションに対応するオーディオデータを含む合成音声128を生成することができる。さらに、通話開始システム104は、代表者102の発話120の受信に应答して音声合成器116によって生成された「The reservation should be for two people at 7pm(予約は、午後7時に2人分必要です)」の合成音声128を出力することができる。さらに、図1の例において、代表者102は、電話108を介して合成音声128を聞き、合成音声128を出力する途中で、代表者102は、割り込み、「Oh I'm sorry, you already said two people at 7pm(ああ、ごめんなさい、あなたはすでに午後7時に2人と言いました)」の発話113を話すとは定する。

#### 【0041】

割り込み検出器132は、割り込みが発生したかどうかを判定するために、代表者102からの入力オーディオデータおよび/またはその対応するトランスクリプションを継続的に分析してもよい。割り込み検出器132は、入力オーディオデータの信号強度をしきい値と比較することができる。通話開始システム104が合成音声を出力している間に、入力オーディオの信号強度が特定のしきい値を超える(例えば周辺雑音レベルを超える)場合、割り込み検出器132は、割り込みが発生したと判定してもよい。いくつかの実装形態において、割り込み検出器132は、音声認識器112によって生成された発話130のトランスクリプション、および/または発話130のエネルギーレベルを分析してもよい。この例では、音声認識器112が発話130のトランスクリプションを生成することができる場合、通話開始システム104は、代表者102または別の人が話しており、電話108がその音声を検出していると判定してもよく、割り込み検出器132は、割り込みが存在すると判定してもよい。この技法を使用することによって、割り込み検出器132は、背景雑音のみが存在する場合、誰かが背景で話している場合、または代表者102がボットに対して話していないことを示す可能性が高い、代表者102が低い音量で話している場合、割り込みを識別しない場合がある。

#### 【0042】

例えば、発話130の対応するオーディオデータの信号強度が55デシベルであると仮定し、しきい値が40デシベルであると仮定する。この例では、音声認識器112は、信号強度(例えば、55デシベル)がしきい値(例えば、40デシベル)を満たすことに基づいて、代

10

20

30

40

50

表者102の発話130のトランスクリプションを生成することができる。音声認識器112が発話130を検出したこと、および/または発話130の対応するオーディオデータの信号強度がしきい値を満たすことに応答して、トランスクリプションを生成したに基づいて、割り込み検出器132は、発話130が会話中の割り込み発話(以下、「割り込み発話130」と呼ぶ)であると判定することができる。

【0043】

割り込み分類器136は、割り込みを、複数の異なる割り込みタイプ138からの所与のタイプの割り込みとして分類することができる。割り込みタイプ138は、相互に排他的とすることができ、例えば、無意味な割り込み、重要ではない意味のある割り込み、重要な意味のある割り込み、および/または他のタイプの割り込みを含むことができる。無意味な割り込みは、通話開始システム104が会話のコースを変更する必要のない割り込みであってもよい。例えば、割り込み検出器132によって検出された所与の割り込みは、検出された割り込みが背景雑音、ささやき、背景の会話を含むか、または代表者102が「right(正しい)」、「ok(オーケー)」、「uh-huh(うん)」、または他の同様の肯定/フィルター単語および/もしくは句などの単語を使用することによってボットが言っていることを確認していると判定された場合、無意味な割り込みとして分類することができる。別の例として、割り込み検出器132によって検出された所与の割り込みは、代表者102がボットの言ったことを繰り返していると判定された場合、無意味な割り込みとして分類することができる。例えば、ボットが「I'd like to make a reservation for two people tomorrow(明日2人分の予約をしたいです)」と言うと仮定し、さらに、ボットが言い終わって「at 7pm(午後7時に)」と言う前に、代表者102が「two people, tomorrow(明日2人分)」と言うと仮定する。この例では、代表者102がボットが言ったことを繰り返すので、割り込み分類器136は、割り込みを無意味な割り込みとして分類することができる。

【0044】

重要ではない意味のある割り込みは、通話開始システム104が会話のコースを変更する必要はあるが、ボットがすぐに話すのを停止することを必要としない割り込みであってもよい。例えば、割り込み検出器132によって検出された所与の割り込みは、検出された割り込みがまだ提供されていない情報ポイント106に対する要求を含むと判定された場合、重要ではない割り込みとして分類することができる。例えば、ボットが「I'd like to make a reservation for two people tomorrow(明日2人分の予約をしたい)」と言うと仮定し、さらに、ボットが言い終わり、「at 7pm(午後7時に)」と言う前に、代表者102が「at what time?(何時に)」と言うと仮定する。この例では、代表者102がボットに知られている情報(例えば、提供された日時122)を要求しているので、割り込み分類器136は、割り込みを重要ではない割り込みとして分類し、割り込みを一時停止することなく「7PM(午後7時)」の合成音声を出力することができる。いくつかの実装形態において、重要ではない意味のある割り込みを検出した後、ボットは、合成音声のすべてを出力する前に合成音声の自然な一時停止に到達する場合があります、合成音声の残りを出力することを停止する場合があります。例えば、割り込み検出器132によって検出された所与の割り込みは、検出された割り込みが「excuse me(すみません)」、「please repeat(繰り返してください)」、「slow down(ゆっくりお願いします)」、ならびに/またはボットが先行する合成音声を明確化するか、または再び述べることを要求する他の単語および/もしくは句を代表者102が述べていることを含むと判定された場合、意味のない割り込みとして分類することができる。例えば、ボットが「I'd like to make a reservation for two people tomorrow(明日2人分の予約をしたい)」と言うと仮定し、さらに、ボットが言い終わり、「at 7pm(午後7時に)」と言う前に、代表者102が「pardon me(失礼します)」と言うと仮定する。この例では、代表者102がボットに合成音声を再出力することを要求しているので、割り込み分類器136は、割り込みを重要ではない割り込みとして分類することができ、ボットは、合成音声を再び出力する前に、合成音声の残りの部分を出力し終わることができる。

10

20

30

40

50

## 【 0 0 4 5 】

重要な意味のある割り込みは、通話開始システム104が会話のコースを変更することを必要とし、ボットがすぐに話すのをやめることを必要とする割り込みであってもよい。いくつかの実装形態において、重要な意味のある割り込みが検出された後、ボットは、ボットが現在話している単語を終了し、次いで、合成音声の残りの部分を出力することを停止してもよい。例えば、ボットが「I'd like to make a reservation for two people tomorrow(明日2人分の予約をしたい)」と言うと仮定し、さらに、ボットが言い終わり、「at 7pm(午後7時に)」と言う前に、代表者102が「please hold(お待ちください)」「stop talking(話すのをやめてください)」「wait, wait, wait(待って、待って、待って)」、ならびに/またはボットが合成音声の出力を停止すべきことを示す他の同様の単語および/もしくは句を言うことと仮定する。この例では、代表者102は、ボットが合成音声の提供を停止することを要求しているため、割り込み分類器136は、割り込みを重要な意味のある割り込みとして分類することができる。したがって、代表者102が割り込み発話を話し終えた後、ボットは、適切なアクション(例えば、会話を終了する、合成音声の出力を停止する、最近出力された合成音声を繰り返す、および/または検出された割り込みに応答する他のアクション)で割り込み発話に応答してもよい。

10

## 【 0 0 4 6 】

割り込み分類器136は、様々な分類技法を使用して、検出された割り込みを分類してもよい。いくつかの実装形態において、割り込み分類器136は、代表者102の割り込み発話を割り込みタイプ136のうちの一つとして分類するために、一つまたは複数の機械学習モデルを使用して代表者102の割り込み発話を処理することができる。通話開始システム104は、トレーニングインスタンスを使用して一つまたは複数の機械学習モデルをトレーニングすることができる。トレーニングインスタンスの各々は、トレーニングインスタンス入力と、対応するトレーニングインスタンス出力とを含むことができる。トレーニングインスタンス入力は、割り込み発話をキャプチャするオーディオデータ(および/またはそのトランスクリプション)を含むことができ、対応するトレーニングインスタンス出力は、割り込み発話の分類に対応するグラウンドトゥールズラベル(例えば、無意味な割り込み、重要ではない意味のある割り込み、重要な意味のある割り込み、および/または他のタイプの割り込み)を含むことができる。例えば、第1のトレーニングインスタンス入力は、「wait, I can't hear you(待ってください、聞こえません)」の発話をキャプチャするオーディオデータ(および/またはそのトランスクリプション)を含むことができ、第1のトレーニングインスタンス出力は、重要な意味のある割り込みに対応するグラウンドトゥールズラベルを含むことができる。さらに、第2のトレーニングインスタンス入力は、「tomorrow at(明日)」の発話をキャプチャするオーディオデータ(および/またはそのトランスクリプション)を含むことができ、第1のトレーニングインスタンス出力は、無意味な割り込みに対応するグラウンドトゥールズラベルを含むことができる。いくつかの実装形態において、トレーニングインスタンス入力は、割り込み発話が検出されたときに出力されていた合成音声の合成音声をキャプチャするオーディオデータ(および/またはそのトランスクリプション)をさらに含んでもよい。

20

30

## 【 0 0 4 7 】

いくつかの追加および/または代替の実装形態において、割り込み分類器136は、代表者102の割り込み発話を割り込みタイプ136のうちの一つとして分類するために、一つまたは複数のルールを使用して代表者102の割り込み発話を処理することができる。それらの実装形態のいくつかのバージョンにおいて、割り込み分類器136は、割り込み発話をトークン化し、トークンを様々な用語のグループと比較してもよい。トークンが第1のグループの用語を含む場合、割り込み分類器136は、割り込みを重要な意味のある割り込みとして分類してもよい。さらに、トークンが第1のグループ内に含まれる用語とは異なる第2のグループからの用語を含む場合、割り込み分類器136は、割り込みを重要ではない意味のある割り込みとして分類してもよい。さらに、トークンが第1のグループと第2のグループの両方の用語とは異なる第3のグループの用語を含む場合、割り込み分類器

40

50

136は、割り込みを無意味な割り込みとして分類してもよい。割り込みを分類するためのこの段階的な手法を使用することによって、所与の割り込みを無意味な割り込みではなく重要ではない意味のある割り込みとして分類することによって引き起こされる会話における追加の一時停止は、通話開始システム104と代表者の両方が会話の交換された主題を処理することを可能にし、それによって、迅速かつ効率的な方法で会話を終了することができるので、割り込み分類器136は、割り込みを分類し過ぎて失敗する傾向にある場合がある。様々な実装形態において、割り込み分類器136は、割り込みの用語が第1または第2の用語のグループのいずれにも含まれない場合、割り込みを無意味な割り込みとして分類してもよいので、割り込み分類器136は、第3の用語のグループを使用しなくてもよい。さらに、様々な実装形態において、割り込み分類器136は、割り込みの用語が第1、第2、または第3の用語のグループ内にない場合、割り込みがないと判定してもよい。

10

## 【0048】

それらの実装形態のいくつかのバージョンにおいて、割り込み分類器136がトークンと比較するグループの各々における用語は、固定されてもよいが、それらの実装形態の他のバージョンにおいて、トークンは、電話での会話に応じて変化してもよい。例えば、第1の用語のグループは、「stop talking(話すのをやめてください)」および「can't hear(聞こえません)」を含んでもよく、第2の用語のグループは、「excuse me(すいません)」、「I'm sorry(ごめんなさい)」、および/または合成音声内に含まれる用語と意味的に類似した用語を含んでもよい。合成音声内に含まれる用語と意味的に類似した用語は、情報ポイント106と意味的に類似した用語を含んでもよい。例えば、図1に示す情報ポイント106は、「two people(2人)」、「7pm(午後7時)」、および「tomorrow(明日)」などの、ユーザに提供された情報を含むことができる。この例では、第2のグループの意味的に類似した用語は、「two people(2人)」に対して「three people(3人)」、「7pm(午後7時)」に対して「8pm(午後8時)」、および「tomorrow(明日)」に対して「next week(来週)」を含んでもよい。このように、割り込みの用語が、代表者102がボットを誤解したことを示唆する任意の意味的に類似した用語を含む場合、割り込み分類器136は、割り込みを重要ではない意味のある割り込みとして分類してもよい。第3の用語のグループは、「right(正しい)」、「uh-huh(うん)」、および/または代表者102が情報ポイント106内に含まれる情報(例えば、「two people(2人)」、「7pm(午後7時)」、および「tomorrow(明日)」)を理解していることに適合する他のフレーズ/肯定用語などの用語を含んでもよい。

20

30

## 【0049】

さらに、図1に示す例において、代表者102は、ボットが合成音声128の「for two(2人分)」を言った後に、割り込み発話130を話し始めることがある。割り込み検出器132は、音声認識器112が割り込み発話130内の音声を認識したことに基づいて、および/または割り込み発話130に対応するオーディオデータのエネルギーレベルが上記で論じたようにしきい値を超えていることに基づいて割り込みを検出してもよい。割り込み分類器136は、割り込み発話130および/またはボットの合成音声128のトランスクリプション最初の部分を、トレーニングされた機械学習モデルおよび/またはルールベースのモデル全体の入力として適用してもよい。最初の部分は、割り込み発話130の「oh, I'm sorry,(ああ、ごめんなさい)」を含む場合がある。割り込み分類器136が機械学習モデルを利用する実装形態において、割り込み分類器136は、割り込み発話130(およびオプションで割り込み発話130の直前の合成音声128)に対応するオーディオデータを処理することに基づいて、割り込み発話130内に含まれる割り込みが重要ではない意味のある割り込みであると分類してもよい。割り込み分類器136が1つまたは複数のルールを含む実装形態において、割り込み分類器136は、割り込み発話130の最初の部分をトークン化してもよく、トークンは、「oh(ああ)」および「I'm sorry(ごめんなさい)」という用語を含む場合がある。さらに、割り込み分類器136は、音声認識器が追加の単語をトランスクリプションするとき、割り込み発話130をトークン化し続けてもよい。「oh(ああ)」および「I'm sorry(ごめんなさい)」という用語は、第1のグループ内のどの用語

40

50

とも一致しない場合があるが、「I'm sorry(ごめんなさい)」という用語は、第2のグループ内の用語と一致する場合がある。この場合、割り込み分類器136は、図1に示す割り込みを重要ではない意味のある割り込みとして分類してもよい。

【0050】

割り込みが分類されると、一時停止挿入器142は、ボットが合成音声を出力することを停止すべきかどうか、および/またはいつ停止すべきかを識別することができる。例えば、割り込みが無意味な割り込みである場合、一時停止挿入器142は、ボットの合成音声内に一時停止を挿入する必要がないと判定し、割り込みが発生しなかったかのようにボットが合成音声を出力し続けてもよいと判定してもよい。別の例として、割り込みが重要な意味のある割り込みである場合、一時停止挿入器142は、ボットが現在の単語および/または句を出力することを完了したときに、合成音声内に一時停止が挿入される必要があると判定してもよい。この例では、一時停止挿入器142は、代表者102に出力されるべき音声合成器116への用語および/または単語を提供することを停止するために、トランスクリプション生成器114に指示を提供することができる。さらに、トランスクリプション生成器114は、代表者102が割り込み発話130を話すことを完了するまで、音声合成器116に追加のトランスクリプションを提供する必要はない。

10

【0051】

一時停止挿入器142から一時停止命令を受信すると、トランスクリプション生成器114は、情報ポイント106が代表者に提供されたかどうかを示すために、情報ポイント106を更新してもよい。例えば、トランスクリプション生成器114は、情報ポイント106の所与の情報ポイントを含む合成音声の出力後に、所与の情報ポイントを満たされたものとして更新してもよい。図1の例では、トランスクリプション生成器114は、提供された日時122、要求された予約124、および提供されたパーティサイズ126に関する情報ポイントを、ボットがそれらの情報ポイント106を含む合成音声をすでに出力しているので、それらが満たされていることを示すように更新してもよい。特に、それらの情報ポイント106は、それらが満たされていることを示すように更新されるが、代表者102によって要求された場合、および/または代表者102による誤解がある場合、ボットは、それらの情報ポイント106を依然として繰り返し提供してもよいが、それらの情報ポイント106は、ボットによって2回自動的に提供されなくてもよい。

20

【0052】

上記のように、図1の例において、割り込み分類器136は、割り込みを重要ではない意味のある割り込みとして分類する。この例では、一時停止識別器142は、自然な一時停止(すなわち、割り込み発話130が検出されたときに出力されている合成音声における一時停止)のための合成音声128内の位置を決定してもよい。自然な一時停止は、文の終わり、前置詞句の終わり、前置詞句の前、接続詞の前、および/または話者が自然に一時停止する可能性がある音声の任意の他の同様の部分であってもよい。一時停止識別器142は、合成音声128内の「for two people(2人分)」という前置詞句の後(すなわち、発話120に回答する「at 7pm(午後7時に)」の部分を提供する前)の自然な一時停止位置を識別してもよい。さらに、一時停止挿入器142は、発話120に対する回答のトランスクリプションを音声合成器116に提供することを停止するようにトランスクリプション生成器114に指示してもよい。

30

40

【0053】

さらに、図1に示すように、ボットは、「for two people(2人分)」を出力した後、合成音声128を出力することを停止してもよく、代表者102は、割り込み発話130を話し続けてもよい。音声認識器112は、割り込み発話130の残りの部分をトランスクリプションすることができ、トランスクリプション生成器114は、「Oh, I'm sorry, you already said two people at 7pm(ああ、ごめんなさい、あなたはすでに午後7時に2人と言いました)」の割り込み発話130のトランスクリプションを受信することができる。さらに、トランスクリプション生成器114は、「that's correct(その通りです)」というトランスクリプションを生成することによって、割り込み発話130に対する回答

50



を生成することができ、音声合成器116は、「That's correct(その通りです)」の合成音声144を生成することができる。通話開始システム104は、次いで、合成音声144を代表者の電話108に送信することができる。

【0054】

通話開始システム104が合成音声144を代表者102の電話108に送信した後、通話開始システム104は、割り込みが完了したと判定してもよい。この時点で、トランスクリプション生成器114は、情報ポイント106のすべてが満たされているかどうかを判定することができる。そうである場合、トランスクリプション生成器114は、電話通話を完了するためにさらなるトランスクリプションを生成することができる。割り込み後に依然として不完全なままである情報ポイント106が存在する場合、トランスクリプション生成器114は、不完全な情報ポイントの情報を代表者102に提供するトランスクリプションを生成することができ、音声合成器116は、トランスクリプションに対応する合成音声を生成することができ、合成音声は、代表者102の電話108において出力することができる。

10

【0055】

図1の例において、代表者102は、「Great, I made the reservation. See you tomorrow.(よかった、予約しました。明日会いましょう。)」の発話146を話すことによって、合成音声144に応答する。音声認識器は、発話146に対して音声認識を実行することができ、音声認識器112は、発話146のトランスクリプションをトランスクリプション生成器114に提供することができ、トランスクリプション生成器114は、情報ポイント106が満たされていると判定することができる。情報ポイント106のすべてが満たされているとの判定に応答して、および/または(例えば、発話146内に含まれる認識されたテキストに基づいて)代表者が電話通話を終了しているとの判定に応答して、トランスクリプション生成器114は、電話通話を終了するためにトランスクリプションを生成することができる。トランスクリプション生成器114は、トランスクリプションを音声合成器116に提供することができる。通話開始システム104は、合成された発話(例えば、「Thank you, see you tomorrow(ありがとうございました、明日お会いしましょう)」)を電話108に送信し、電話通話を終了する。

20

【0056】

いくつかの実装形態において、トランスクリプション生成器114は、割り込み発話130が検出される前に、電話での会話を合成音声128の主題に戻してもよい。この場合、トランスクリプション生成器114は、合成音声128の任意の不完全な情報ポイント106(例えば、「7pm(午後7時)」)を完了するトランスクリプションを生成し、満たされていない情報ポイント106に基づいて、さらなる合成音声を生成してもよい。トランスクリプション生成器114は、トランスクリプションを音声合成器116に提供することができる。通話開始システム104は、合成音声を電話108に送信する。例えば、割り込み発話130が「you already said two people(あなたはすでに2人と言いました)」のみを示し、「7pm(午後7時)」を含む以前の合成音声118を認識しなかった場合、合成音声144は、提供された日時122を含む合成音声(例えば、「That's correct, and 7pm(その通り、午後7時)」)も含んでもよい。

30

40

【0057】

通話開始システム104の様々な実装形態において、ボットは、ユーザの代わりにボットと会話を行うことに同意することを代表者102に要求する合成音声を、代表者102の電話108において出力させることができる。いくつかの実装形態において、ボットは、代表者102が補助通話に応答したときに、合成音声を出力させることができる。例えば、代表者102の発話110の検出に応答して、ボットは、「Hello, this is bot calling on behalf of John Doe, do you consent to monitoring of this call(こんにちは、こちらはJohn Doeの代わりに通話しているボットですが、この通話を監視することに同意しますか)」という合成音声を発生させることができる。ボットが代表者102から同意を受け取った場合、ボットは、次いで、合成音声118を出力することができる。

50

しかしながら、ボットが代表者102から同意を受け取らなかった場合、ボットは、補助通話を終了することができ、代表者102が補助通話に同意しなかったことをユーザに通知することができる。

【0058】

図2は、第1のユーザ202と第2のユーザ204との間の電話通話を監視する例示的なシステム200を示す。例えば、電話通話の第1のユーザ202は、保留中である場合があり、システム200は、第2のユーザ204が電話通話に再参加したときに、保留中の第1のユーザ202に通知することができる。簡単に、以下でより詳細に説明するように、第1のユーザ202は、第2のユーザ204を通話してもよく、第2のユーザ204は、第1のユーザ202を保留にしてもよい。第1のユーザ202は、保留に回答して、電話206を切ってもよい。通話監視システム210は、第1のユーザ202が保留中に電話206を切ったことを検出するために、電話通話を監視することができる。第1のユーザ202は、電話206を切るが、通話監視システム210は、第1のユーザ206が電話206を切ったことに回答して電話通話をアクティブ状態に維持することができ、第2のユーザ204が電話通話に再参加したときに第1のユーザ202に通知することができる。さらに、図2に示すシステム200は、ネットワーク(例えば、LAN、WAN、Bluetooth、および/または他のネットワーク)上のサーバによって実装されるものとして示されているが、それは、例のためのものであり、限定することを意図したものではないことが理解されるべきである。例えば、システム200は、第1のユーザ202のコンピューティングデバイスにおいてローカルに実装することができ、ならびに/またはネットワーク上に分散された方法で第1のユーザ202のコンピューティングデバイスおよびサーバによって実装することができる。

10

20

【0059】

第1のユーザ202は、通話監視システム210がユーザ202の電話通話を監視することを可能にするために、電話206における通話設定を構成することができる。図2に示すように、第1のユーザ202が、対応するコンピューティングデバイス(例えば、電話206)を使用して第2のユーザ204に電話通話を行い、通話監視システム210が有効にされていると仮定する。この例では、通話監視システム210は、第1のユーザ202と第2のユーザ204との間の電話通話を監視することができる。電話通話の最初の部分の間、電話206は、電話206が電話208に接続されている第1の状態226にあってもよい。さらに、第2のユーザ204が電話208を使用して電話通話に回答し、「Thank you for calling Cable Company. How may I direct your call(Cable Companyにお電話していただきありがとうございます。お電話をどのようにおつなぎいたしましょうか)」という発話212を話すと仮定し、第1のユーザ202が第2のユーザ204に「I'd like to cancel my account(私のアカウントをキャンセルしたい)」という発話214で回答し、第2のユーザ204が第1のユーザ202に「Thank you, please hold(ありがとうございます、お待ちください)」という発話216で回答すると仮定する。

30

【0060】

第1のユーザ202と第2のユーザ204との間のこの交換中に、通話監視システム210は、電話通話を監視することができ、第1のユーザ202と第2のユーザ204の両方の発話212、214、および216に対して音声認識を実行するために、音声認識器244を使用することができる。通話監視システム210は、電話通話の主題を判定するために、発話212、214、および216の対応するトランスクリプションを分析することができる。さらに、通話監視システム210は、電話通話の主題および/または電話通話の他の詳細を情報ポイント248内に記憶することができる。この例では、通話監視システム210は、「cancel account for cable company(ケーブル会社のアカウントをキャンセルする)」という情報ポイントを記憶することができ、第1のユーザ202が第2のユーザ204からケーブルアカウントがキャンセルされたことの確認を受信する前に通話が保留にされたので、この情報ポイントを最初は満たされていないとラベル付けすることができる。

40

【0061】

通話監視システム210の通話モニタ236は、保留検出器238を含むことができる。保

50

留検出器238は、第2のユーザ204が第1のユーザ202を保留にしたことを判定することができる。保留検出器238は、例えば、第1のユーザ202が保留にされていることを明示する第2のユーザ204からの発話(例えば、発話216内の「please hold(保留にしてください)」)を検出したことに基づいて、無音のしきい値持続時間を検出した(例えば、第1のユーザ202または第2のユーザ204の発話が30秒、45秒、および/または他の持続時間の間検出されない)ことに基づいて、別のポットが第2のユーザの代わりに通話を引き継いだことを(例えば、信号エネルギー、音声識別などに基づいて)検出したことに基づいて、第1のユーザ202が保留にされたことを示す通話に関連するメタデータを検出したことに基づいて、および/または通話が保留にされたことを判定するための他の技法に基づいて、第2のユーザ204が第1のユーザ202を保留にしたことを判定することができる。この時点で、通話監視システム210は、ユーザが電話を切ろうとした場合であっても、通話をアクティブ状態に維持することができる。

#### 【0062】

図2に示すように、第1のユーザ202は、保留で待ちたくないので、電話206が第1の状態226にある間に電話206を切ろうとする。特に、第1のユーザ202が電話206を切ったとしても、通話監視システム210は、通話が完全に切断されるのを防止することができる。むしろ、通話監視システム210は、電話206を第1の状態226(例えば、接続されている)から第2の状態228(例えば、ミュート)に遷移させることができる。第2の状態228において、第1のユーザ202と第2のユーザ204との間の通話は、アクティブのままであるが、電話206のマイクロフォンは、ミュートされている。電話206は、バックグラウンドにおいてアクティブな電話通話が存在することを示し、マイクロフォンがミュートされていることも示してもよい。第1のユーザ202は、通話監視システム210が電話通話を第2の状態228に維持している間、他の活動を実行するために電話206を使用することができる場合がある。いくつかの実装形態において、第1のユーザ202は、電話が第1の状態226または第2の状態228にある間に、通話監視システム210に通話を完全に切断させる(例えば、発話された、タッチされた、および/またはタイプされた)追加の入力を提供することができる。

#### 【0063】

いくつかの実装形態において、通話監視システム210は、第1のユーザ202が電話206を第1の状態226(例えば、接続されている)から第2の状態228(例えば、ミュート)に遷移させるために電話206を切ろうとすることを必要としない場合がある。むしろ、保留検出器238は、第2のユーザ204が第1のユーザ202を保留にしたと判定することができる、それに応答して、通話モニタ236は、電話206を第1の状態226(例えば、接続されている)から第2の状態228(例えば、ミュート)に自動的に遷移させることができ、それによって、電話通話をアクティブ状態に維持する。アクティブ状態は、例えば、第1のユーザ202の電話206がネットワーク(ボイスオーバーインターネットプロトコル(VoIP)、公衆交換電話網(PSTN)、および/または他の電話通信プロトコル)を介して第2のユーザ204の電話208との通信を維持していることとすることができる。

#### 【0064】

図2に示すように、第2のユーザ204が電話通話に再参加すると仮定する。この時点で、保留検出器238は、第1のユーザ202がもはや保留にされていないと判定し、保留検出器238は、第1のユーザ202がもはや保留にされていないことの指示を通知生成器242に提供することができる。この例では、通知生成器242は、電話206に、第1のユーザ228がもはや保留にされていないことを示す第1のユーザ202のための可聴通知234を(視覚的通知に加えて、またはその代わりに)出力させる(例えば、電話206は、着信のように振動する/鳴る(ring)、電話206は、着信通知のように振動する/鳴る(ding)、電話206は、電話206のディスプレイ上に視覚的通知をレンダリングさせる、など)。いくつかの追加のおよび/または代替の実装形態において、通知生成器242は、第1のユーザ202が電話206を使用していると判定してもよい。この場合、通知生成器242は、(可聴通知に加えて、またはその代わりに)電話206のディスプレイ上に視覚的通知を提示することが

できる。

【0065】

第2のユーザ204が電話通話に再参加するとき、第2のユーザ204が「Hello, I understand you want to cancel your account(こんにちは、あなたが自分のアカウントをキャンセルしたいことを了解しました)」という発話220を話すと仮定する。いくつかの実装形態において、図2に示すように、第1のユーザ202は、通知234を受信したことに応答して通話にまだ再参加していない場合がある。この場合、通話監視システム210のボットは、第1のユーザ202が電話通話に再参加するのを待っている間に、第2のユーザ204の電話208に出力するために、「Hello, sorry. She will be right back(こんにちは、ごめんなさい。彼女はすぐに戻ってきます)」という合成音声218を生成し  
10  
てもよい。通話監視システム210は、図1に関して上記でより詳細に説明したのと同様の方法において、合成音声を生成することができる。例えば、音声認識器244は、音声認識器244を使用して、第2のユーザ204の発話220に対して音声認識を実行することができ、トランスクリプション生成器240は、発話220に応答するトランスクリプションを生成することができ、音声合成器246は、発話に応答するトランスクリプションに対応するオーディオデータを含む合成音声を生成することができ、通話監視システムは、合成音声を第2のユーザの電話208において出力させることができる。それらの実装形態のいくつかのバージョンにおいて、通話監視システム210は、第2のユーザ204の発話220のトランスクリプションを視覚的にレンダリングさせることができ、および/または第2のユーザ204の発話220をキャプチャするオーディオデータを第1のユーザの電話206にお  
20  
いて聴覚的にレンダリングさせることができる。

【0066】

いくつかの実装形態において、第2のユーザ204の発話220に応答するトランスクリプションは、通知234が第1のユーザ202に出力されたことを示すことができ、および/または第1のユーザ202が電話通話にすぐに戻ることを示すことができる。上記のように、トランスクリプション生成器240は、「Hello, sorry. She will be right back(こんにちは、ごめんなさい。彼女はすぐに戻ってきます)」というトランスクリプションを生成し、トランスクリプションを音声合成器246に提供することができ、音声合成器246は、合成音声218に対応するオーディオデータ224を含む合成音声を生成することが  
30  
できる。通話監視システム210は、オーディオデータ224を電話208に直接送信することができ、または電話208に送信するために電話206に送信することができる。

【0067】

いくつかの実装形態において、トランスクリプション生成器240は、音声合成器246に提供されるトランスクリプションを生成するために、第2のユーザ204の発話220のトランスクリプションと、電話通話の情報ポイント248とを使用してよい。この例では、トランスクリプション生成器240は、Cable Companyとのアカウントをキャンセルすることに関連する詳細を含む情報ポイント248にアクセスしてもよい。別の例として、トランスクリプション生成器240は、「Hello, sorry, she is looking for her account information and will be right back(こんにちは、ごめんなさい。彼女は自分のアカウント情報を探しており、すぐに戻ってきます)」というトランスクリプションを生成し、トランスクリプションを音声合成器246に提供することができ、音声合成器246は、合成音声に対応するオーディオデータを含む合成音声を生成することが  
40  
できる。再び、通話監視システム210は、オーディオデータを電話208に送信するか、または電話208に送信するために電話206に送信することができる。

【0068】

いくつかの実装形態において、通話監視システム210は、第1のユーザ202が通話に再参加するまで、第1のユーザの代わりに第2のユーザ204との会話を継続することができる。トランスクリプション生成器240は、電話206において検出された第2のユーザ204の追加の情報ポイントおよび/または追加の発話に関連するトランスクリプションを生成するために、図1に記載のものと同様の技術を使用してもよい。この場合、トランスクリ  
50

ブション生成器240は、第1のユーザ202が第2のユーザ204との会話を継続する準備ができていないことを示すトランスクリプションを生成してもよい。例えば、トランスクリプション生成器240は、「Hello, sorry. I'm still looking for my account information(こんにちは、ごめんなさい。まだ自分のアカウント情報を探しています)」または「Hello. Please give me a moment while my computer boots up(こんにちは。私のコンピュータが起動するまでしばらくお待ちください)」というトランスクリプションを生成してもよい。それらの実装形態のいくつかのバージョンにおいて、ポットは、第1のユーザ202のユーザプロファイルにアクセスし、ユーザプロファイル内に含まれる情報(例えば、第1のユーザ202のアドレス、第1のユーザ202の名前、ケーブル会社に関連付けられた第1のユーザ202のアカウント番号など)を提供することができる。

10

【0069】

いくつかの追加のおよび/または代替の実装形態において、ポットは、第1のユーザがしきい値持続時間内に通話に再参加しない場合、第2のユーザ204を保留にし、および/または電話通話を終了してもよい。しきい値持続時間は、固定または動的であってもよい。しきい値持続時間が動的である実装形態において、しきい値持続時間は、第1のユーザ202と電話206との対話に基づいてもよい。例えば、第1のユーザ202が30秒間通知234と対話しない場合、ポットは、第2のユーザ204を保留にしてもよい。この例では、第1のユーザ202が、第2のユーザ204が保留にされてから60秒以内に通話に再参加しない場合、ポットは、通話が終了されるべきであると判定してもよい。別の例として、第1のユーザ202が通知234と対話するが、通話にまだ再参加していない場合、ポットは、60秒後に第2のユーザ204を保留にしてもよい。この例では、第1のユーザ202が、第2のユーザ204が保留にされてからさらに60秒以内に通話に再参加しない場合、ポットは、通話が終了されるべきであると判定してもよい。

20

【0070】

図2に示すように、第1のユーザ202が通知234を受信したことに応答して電話通話に再参加し、「Hi, I'm here(こんにちは、私はここにいます)」という発話222を話すと仮定する。第1のユーザ202が通話に再参加すると、電話206は、第2の状態228(例えば、「ミュート」)から第3の状態230(例えば、「接続された」または「再接続された」)に遷移することができ、第3の状態230では、電話206は、電話通話に接続されたままであり、第1のユーザ202の電話206のマイクロフォンは、ミュートされていない。通話監視システム210および通話モニタ236は、電話通話を監視することに帰ることができる。

30

【0071】

いくつかの追加のおよび/または代替の実装形態において、図2には示されていないが、通話監視システム210は、第1のユーザ202が第2のユーザ204の発話220の検出前(または検出中)に通知234を受信したことに応答して通話に再参加したと判定することができる。通話監視システム210は、例えば、電話206が第2の状態228から第3の状態230に遷移したと判定したことに基づいて、音声識別を使用して電話206において検出された音声発話が第1のユーザ202に対応すると判定したこと(およびオプションで、音声発話が第2のユーザ204の発話220に反応していると判定したこと)に基づいて、および/または第1のユーザ202が通話に再参加したことを判定するために他の技法を使用して、第2のユーザ202の発話220の検出前(または検出中)に第1のユーザ202が通話に再参加したことを判定することができる。それらの実装形態のいくつかのバージョンにおいて、通話監視システム210は、ポットが合成音声218を生成する必要がない、および/または第2のユーザ204との会話に関与する必要がないと判定することができる。例えば、第1のユーザ202が通知234を受信してすぐに、第2のユーザ204が発話220を話している間に電話通話に再参加すると仮定する。この例では、第1のユーザ202がすでに通話に再参加しているので、通話監視システム210は、第2のユーザ204との会話に関与するためにポットを呼び出さなくてもよい。それらの実装形態のいくつかのさらなるバージョンにおいて、通話監視システム210は、第1のユーザ202が第2のユーザ204の発話220の全

40

50

体を認識することを確実にするために、依然として第2のユーザ204の発話220のトランスクリプションを第1のユーザ202の電話206上に視覚的にレンダリングさせてもよい。

【0072】

通話監視システム210の様々な実装形態において、ボットは、第1のユーザ202と第2のユーザ204との間の会話がボットによって監視されることに同意するように第2のユーザ204に要求する合成音声を発生させることができる。いくつかの実装形態において、ボットは、第2のユーザ204が通話に応答したときに、合成音声を出力させることができる。例えば、第2のユーザの発話212を検出したことに応答して、ボットは、「Hello, do you consent to the bot joining this call(こんにちは、ボットがこの呼出に参加することに同意しますか)」という合成音声をレンダリングさせることができる。ボットが第2のユーザ204から同意を受け取った場合、ボットは、会話に参加することができる。しかしながら、ボットが第2のユーザ204から同意を受け取らない場合、ボットは、通話に参加しない場合がある。ボットが第2のユーザ204から通話に参加することの同意を受け取らない場合であっても、ボットは、依然として通話を監視することができる。例えば、ボットは、音声活動を検出するようにトレーニングされた音声活動検出器(VAD)、特定の単語および/または句(例えば、「Hello(こんにちは)」、「I'm sorry for wait(お待たせして申し訳ありません)」など)を検出するようにトレーニングされたワード検出器を使用して、第2のユーザ204が通話に再参加したかどうか、および/または第2のユーザ204が通話に再参加したことを判定するために、依然として通話を監視してもよい。しかしながら、ボットは、第2のユーザ204の任意の発話を処理するために音声認識器244を使用しなくてもよい。

10

20

【0073】

図3Aは、(例えば、図1に関して上記でより詳細に説明したように)ユーザおよび通話開始システムのボットが電話での会話に関与している間にユーザ(または代表者)から受信した割り込みを処理するための例示的なプロセス300aのフローチャートである。一般に、プロセス300aは、コンピューティングデバイスの対応するユーザに関連付けられたボットを使用して、ユーザ(または代表者)との電話での会話を開始することができる。電話での会話中、ユーザ(または代表者)は、ボットが合成音声を出力している間にボットに割り込んでよい。ボットは、割り込みを処理するための適切な方法を決定するために、割り込みを識別および分類することができる。割り込みを処理した後、ボットは、割り込みの前に話し合っていた主題に会話を戻すか、または会話を進めることができる。簡略にするために、プロセス300aの動作について、1つまたは複数のプロセッサを含むシステム(例えば、図1のシステム100)によって実行されるものとして説明する。システムは、例えば、サーバ、ボットに関連付けられた対応するユーザのコンピューティングデバイス、および/またはそれらの組合せによって実装することができる。

30

【0074】

ブロック310aにおいて、システムは、電話通話を開始し、ボットを使用して電話での会話を行うように構成された通話開始システムによって、ユーザとの電話通話を開始する。システムは、所与のユーザの対応するコンピューティングデバイスにおいて、所与のユーザの代わりに通話を開始することを要求する(例えば、発話された、タイプされた、および/またはタッチされた)ユーザ入力の検出に応答して、通話を開始することができる。ボットは、対応するコンピューティングデバイスの所与のユーザの代わりにユーザとの会話に関与することができる。いくつかの実装形態において、ボットは、ユーザとの電話通話を開始する前に、所与のユーザから情報を求めることができる。例えば、所与のユーザが「Make a reservation at Burger Palace(Burger Palaceの予約を行ってください)」というユーザ入力を提供した場合、ボットは、ボットに予約を行うことを要求した所与のユーザに、日時情報とパーティサイズ情報とを求めることができる。ユーザは、会話中に関与している事業所または代理店に関連付けられた別のボットまたは人間の代表者とすることができる。上記の例を続けると、ユーザは、Burger Palaceに関連付けられた別のボットまたは人間の従業員とすることができる。ユーザとの通話を開始するこ

40

50

とについて、本明細書で(例えば、図1に関して)より詳細に説明している。

【0075】

ブロック320aにおいて、システムは、通話開始システムによる出力のために、ボットの合成音声を提供する。合成音声は、例えば、通話の開始時にユーザが会話中にボットが関与することに同意することに対する要求、通話を開始するためにユーザ入力を提供した所与のユーザの代わりに実行されるべきタスクに関連する情報、および/またはボットとユーザとの間の会話を促進するための他の合成音声に対応するオーディオデータを含むことができる。合成音声は、合成音声がユーザによって聴覚的に知覚され得るように、ユーザに関連付けられたコンピューティングデバイスのスピーカを介して出力することができる。合成音声を提供することについて、本明細書で(例えば、図1に関して)より詳細に説明している。

10

【0076】

ブロック330aにおいて、出力のためにボットの合成音声を提供している間、システムは、ユーザから、通話開始システムによって、ユーザ発話を受信する。システムは、ボットが合成音声を出力している間も、ユーザの発話を継続的に監視することができる。言い換えると、システムは、ボットがユーザに関連付けられたコンピューティングデバイスにおいて合成音声を出力している間、ユーザの発話を検出することができる。ボットが合成音声を出力している間にユーザ発話を受信することについて、本明細書で(例えば、図1に関して)より詳細に説明している。

【0077】

ブロック340aにおいて、出力のためにボットの合成音声を提供している間に、ユーザ発話とボットの合成音声の出力された部分とに基づいて、システムは、出力のためにボットの合成音声を提供することを継続するか、または出力のためにボットの合成音声を提供することを中止するかを決定する。システムは、ブロック330aにおいて受信したユーザ発話を複数の異なるタイプの割り込み(例えば、無意味な割り込み、重要ではない意味のある割り込み、重要な意味のある割り込み、および/または他のタイプの割り込み)のうちの1つに分類することができる。さらに、システムは、ユーザ発話の分類に基づいて、合成音声の提供を継続するかまたは中止するかを決定することができる。システムは、機械学習モデルおよび/または1つもしくは複数のルールを使用して、ブロック330aにおいて受信したユーザ発話(およびオプションで合成音声)を処理することに基づいて、ユーザ発話を複数の異なるタイプの割り込みのうちの1つに分類することができる。さらに、システムは、ブロック330aにおいてユーザ発話を受信する前に、ボットに会話を合成音声内に含まれる主題に戻させることができ、ユーザ発話がブロック330aにおいて受信されなかったかのように、会話を継続することができ、および/またはブロック330aにおいて受信したユーザ発話に回答する合成音声をさらに生じさせるために会話の方向を変更することができる。ユーザ発話を複数の異なる割り込みタイプのうちの1つに分類し、合成音声の提供を中止するかどうかを決定することについて、本明細書で(例えば、図1に関して)より詳細に説明している。

20

30

【0078】

図3Bは、あるユーザが保留中である電話通話を監視し、他のユーザが通話に再参加したときに保留中のユーザに通知するための例示的なプロセスのフローチャートである。一般に、プロセス300bは、2人のユーザ間の電話での会話を監視する。ユーザのうちの一方は、他のユーザを保留にする。待機する代わりに、保留中のユーザは、電話を切るようにする。プロセス300bは、電話通話をアクティブ状態に維持し、他のユーザが通話に再参加したときに保留中のユーザに通知する。プロセス300bは、保留中だったユーザが通話に戻るのを待っている間、再参加したユーザとの会話を行うためにボットを使用してもよい。簡略にするために、プロセス300bの動作について、1つまたは複数のプロセッサを含むシステム(例えば、図2のシステム200)によって実行されるものとして説明する。システムは、例えば、サーバ、ボットに関連付けられた対応するユーザのコンピューティングデバイス、および/またはそれらの組合せによって実装することができる。

40

50

## 【 0 0 7 9 】

ブロック310bにおいて、システムは、第1のユーザおよび第2のユーザが電話通話に  
関与していると判定する。システムは、第1のユーザに関連付けられた対応する第1の電  
話の状態に基づいて、第1のユーザおよび第2のユーザが電話通話に関与していると判定  
することができる。対応する第1の電話の状態は、例えば、第1のユーザおよび第2のユ  
ーザが電話通信プロトコル(例えば、VoIP、PSTN、および/または他のプロトコル)を使  
用して接続されていることを示すことができる。第1のユーザおよび第2のユーザが電話  
通話に関与していると判定することについて、本明細書で(例えば、図2に関して)より詳  
細に説明している。

## 【 0 0 8 0 】

ブロック320bにおいて、システムは、第2のユーザが第1のユーザを保留にしたと判  
定する。システムは、会話のユーザ発話を処理することに基づいて、第2のユーザが第1  
のユーザを保留にしたと判定することができる。例えば、システムは、会話のユーザ発話  
に対応するオーディオデータのストリームを処理することができる。(例えば、音声認識  
器を使用して)オーディオデータのストリームを処理することに基づいて、システムは、  
オーディオデータのストリームに対応する認識されたテキストが、第2のユーザが第1の  
ユーザを保留にしたことを示すと判定することができる。例えば、システムは、第2のユ  
ーザが、「Please hold(保留にしてください)」、「Will you please hold while I  
transfer your call(通話を転送する間、保留にしてください)」、および/または第2  
のユーザが第1のユーザを保留にしたことを示す他の句を述べたことを判定するために、  
会話のユーザ発話を処理することができる。第2のユーザが第1のユーザを保留にしたと  
判定することについて、本明細書で(例えば、図2に関して)より詳細に説明している。

## 【 0 0 8 1 】

ブロック330bにおいて、システムは、第1のユーザが対応する第1の電話を切ったと  
判定する。システムは、第1のユーザが電話通話を終了するために対応する第1の電話に  
おいてユーザ入力を提供したと判定したことに基づいて、第1のユーザが対応する第1の  
電話を切ったと判定することができる。例えば、システムは、選択されると電話通話を終  
了させるボタン(例えば、物理ボタン、および/またはグラフィカルユーザインターフェ  
ース上のソフトボタン)に向けられたユーザ入力に基づいて、第1のユーザが対応する第1  
の電話をロック状態にしたことに基づいて、および/または対応する第1の電話と他の  
対話に基づいて、第1のユーザが対応する第1の入力を切ったと判定することができる。  
第1のユーザが対応する第1の電話を切ったと判定することについて、本明細書で(例えば  
、図2に関して)より詳細に説明している。

## 【 0 0 8 2 】

ブロック340bにおいて、第1のユーザが対応する第1の電話を切ったと判定したこと  
に基づいて、第1のユーザが保留にされている間、システムは、電話通話をアクティブ状  
態に維持する。システムは、対応する第1の電話を第1の状態から第2の状態に遷移させ  
ることによって、電話通話のアクティブ状態を維持することができる。第2の状態にお  
いて、対応する第1のコンピューティングデバイスのマイクロフォンをミュートすること  
ができ、第1のユーザは、システムがバックグラウンドにおいて第2のユーザの対応する第  
2の電話との電話接続を維持している間、対応する第1の電話と対話することができる。  
対応する第1の電話をアクティブ状態に維持することについて、本明細書で(例えば、図2  
に関して)より詳細に説明している。

## 【 0 0 8 3 】

ブロック350bにおいて、システムは、第2のユーザが電話通話に再参加したと判定す  
る。システムは、第2のユーザが通話に再参加したことを判定することができる。いくつ  
かの実装形態において、システムは、第2のユーザのさらなるユーザ発話を検出したこと  
に基づいて、第2のユーザが通話に再参加したと判定することができる。システムは、第  
2のユーザが通話に再参加したことを判定するために、音声活動検出器、音声認識器、お  
よび/または他の構成要素を使用することができる。それらの実装形態のいくつかのパー

10

20

30

40

50



ジョンにおいて、システムは、音声識別を使用して、ユーザ発話が(すなわち、音楽が再生されている間に対応するオーディオデータ、対話型音声応答(IVR)システムのものに対応するオーディオデータなどとは対照的に)第2のユーザからのものであると判定することができる。第2のユーザが電話通話に再参加したと判定することについて、本明細書で(例えば、図2に関して)より詳細に説明している。特に、ブロック340bとブロック350bとの間に切れ目がある。この切れ目は、システムが、ブロック350bにおいて第2のユーザが電話通話に再参加することを積極的に監視しながら、ブロック340bにおいて電話通話のアクティブ状態を維持することができることを示す。

【0084】

ブロック360bにおいて、第2のユーザが電話通話に再参加したと判定したことに基  
10  
いて、システムは、第2のユーザと、人間との電話での会話を行うように構成されたボットとの間の電話での会話を行う。システムは、第2のユーザが電話通話に再参加したときにユーザ発話を提供したと判定したことに応答して、ボットに合成音声を出力させることができる。いくつかの実装形態において、システムは、第1のユーザがまだ電話通話に再参加していないと判定したことに応答して、ボットに合成音声を出力させるだけである。合成音声は、ボットが会話におけるアクティブな参加者であり、第1のユーザの代わりであることを示すことができる。いくつかの追加のおよび/または代替の実装形態において、ボットは、電話通話中に第2のユーザに伝達されるべき情報ポイント(例えば、名前情報、20  
アドレス情報、アカウント情報、および/または第1のユーザに関連付けられた他の情報)を含む合成音声を出力することができる。ボットと第2のユーザとの間で電話での会話を行うことについて、本明細書で(例えば、図2に関して)より詳細に説明している。

【0085】

ブロック370bにおいて、第2のユーザとボットとの間で電話での会話を行っている間、システムは、対応する第1の電話への出力のために、第1のユーザが電話通話に再参加するための通知を提供する。第1のユーザのための通知は、第2のユーザが通話に再参加したことを示すことができ、例えば、対応する第1の電話においてレンダリングされる可聴および/または視覚的通知を含むことができる。さらに、ブロック360bにおいて出力される合成音声は、第1のユーザが、第2のユーザが電話通話に再参加したことを通知されたことを示すことができる。いくつかの実装形態において、通知は、第2のユーザが電話通話に再参加した後の第2のユーザのユーザ発話および/またはボットの合成音声のト  
30  
ランスクリプション(および/またはそれに対応するオーディオデータ)をさらに含むことができる。第1のユーザが電話通話に再参加するための通知を提供することについて、本明細書で(例えば、図2に関して)より詳細に説明している。

【0086】

ブロック380bにおいて、システムは、第1のユーザが電話通話に再参加したと判定する。システムは、対応する第1の電話が第2の状態から、第1のユーザがここで会話におけるアクティブな参加者であることを示す第3の状態に遷移したことに基  
40  
づいて、第1のユーザが電話通話に再参加したと判定することができる。第1のユーザが電話通話に再参加したと判定することについて、本明細書で(例えば、図2に関して)より詳細に説明している。

【0087】

ブロック390bにおいて、第1のユーザが電話通話に再参加したと判定したことに基  
いて、システムは、第2のユーザとボットとの間で電話での会話を行うことを中止する。ボットは、第1のユーザが電話通話に再参加したとの判定に応答して、電話での会話を行うことを停止することができる。いくつかの実装形態において、第1のユーザが通話に再参加したときにボットが合成音声を出力している場合、ボットは、(例えば一時停止挿入器142に関して図1で論じたのと同様に)合成音声の出力を終了したとき、および/または合成音声における自然な一時停止に達したときに、合成音声を提供することを中止することができる。電話での会話を行うことを中止することについて、本明細書で(例えば、図  
2に関して)より詳細に説明している。

10

20

30

40

50

## 【 0 0 8 8 】

図4は、ここで説明されている技法を実装するために使用することができるコンピューティングデバイス400およびモバイルコンピューティングデバイス450の例を示す。コンピューティングデバイス400は、ラップトップ、デスクトップ、ワークステーション、携帯情報端末、サーバ、ブレードサーバ、メインフレーム、および他の適切なコンピュータなどの様々な形態のデジタルコンピュータを表すことを意図している。モバイルコンピューティングデバイス450は、携帯情報端末、携帯電話、スマートフォン、および他の同様のコンピューティングデバイスなどの様々な形態のモバイルデバイスを表すことを意図している。ここに示す構成要素、それらの接続および関係、ならびにそれらの機能は、単なる例であることを意味しており、限定であることを意味していない。

10

## 【 0 0 8 9 】

コンピューティングデバイス400は、プロセッサ402と、メモリ404と、記憶デバイス406と、メモリ404および複数の高速拡張ポート410に接続する高速インターフェース408と、低速拡張ポート414および記憶デバイス406に接続する低速インターフェース412とを含む。プロセッサ402、メモリ404、記憶デバイス406、高速インターフェース408、高速拡張ポート410、および低速インターフェース412の各々は、様々なバスを使用して相互接続され、一般的なマザーボード上に取り付けられるか、または必要に応じて他の方法で取り付けられてもよい。プロセッサ402は、高速インターフェース408に結合されたディスプレイ416などの外部入力/出力デバイス上にGUIのためのグラフィカル情報を表示するために、メモリ404内または記憶デバイス406上に記憶された命令を含む、コンピューティングデバイス400内で実行するための命令を処理することができる。他の実装形態において、必要に応じて、複数のメモリおよびメモリのタイプとともに、複数のプロセッサおよび/または複数のバスが使用されてもよい。また、(例えば、サーババンク、ブレードサーバのグループ、またはマルチプロセッサシステムとして)複数のコンピューティングデバイスが接続され、各デバイスが必要な動作の一部を提供してもよい。

20

## 【 0 0 9 0 】

メモリ404は、コンピューティングデバイス400内の情報を記憶する。いくつかの実装形態において、メモリ404は、揮発性メモリユニットである。いくつかの実装形態において、メモリ404は、不揮発性メモリユニットである。メモリ404は、磁気ディスクまたは光ディスクなどの別の形態のコンピュータ可読媒体であってもよい。

30

## 【 0 0 9 1 】

記憶デバイス406は、コンピューティングデバイス400のための大容量ストレージを提供することができる。いくつかの実装形態において、記憶デバイス406は、フロッピーディスクデバイス、ハードディスクデバイス、光ディスクデバイス、もしくはテープデバイス、フラッシュメモリもしくは他の同様のソリッドステートメモリデバイス、またはストレージエリアネットワークもしくは他の構成においてデバイスを含むデバイスのアレイなどのコンピュータ可読媒体であるか、またはそれらを含んでもよい。命令は、情報キャリア内に記憶することができる。命令は、1つまたは複数の処理デバイス(例えば、プロセッサ402)によって実行されると、上記で説明した方法などの1つまたは複数の方法を実行する。命令は、コンピュータまたは機械可読媒体(例えば、メモリ404、記憶デバイス406、またはプロセッサ402上のメモリ)などの1つまたは複数の記憶デバイスによって記憶することもできる。

40

## 【 0 0 9 2 】

高速インターフェース408は、コンピューティングデバイス400のための帯域集約的な動作を管理し、低速インターフェース412は、より低い帯域集約的な動作を管理する。そのような機能の割り当ては、単なる例である。いくつかの実装形態において、高速インターフェース408は、メモリ404、(例えば、グラフィックプロセッサまたはアクセラレータを介して)ディスプレイ416、および様々な拡張カード(図示せず)を受け入れてもよい高速拡張ポート410に結合される。実装形態において、低速インターフェース412

50

は、記憶デバイス406および低速拡張ポート414に結合される。様々な通信ポート(例えば、USB、Bluetooth、イーサネット、ワイヤレスイーサネット)を含んでもよい。低速拡張ポート414は、キーボード、ポインティングデバイス、スキャナ、または、例えば、ネットワークアダプタを介してスイッチもしくはルータなどのネットワークングデバイスなどの、1つまたは複数の入力/出力デバイスに結合されてもよい。

**【0093】**

コンピューティングデバイス400は、図に示すように、いくつかの異なる形態において実装されてもよい。例えば、標準的なサーバ420として実装されてもよく、またはそのようなサーバのグループにおいて複数回実装されてもよい。加えて、それは、ラップトップコンピュータ422のようなパーソナルコンピュータにおいて実装されてもよい。それはまた、ラックサーバシステム424の一部として実装されてもよい。代替的には、コンピューティングデバイス400からの構成要素は、モバイルコンピューティングデバイス450などのモバイルデバイス(図示せず)内の他の構成要素と組み合わせられてもよい。そのようなデバイスの各々は、コンピューティングデバイス400およびモバイルコンピューティングデバイス450のうちの1つまたは複数を含んでもよく、システム全体は、互いに通信する複数のコンピューティングデバイスから構成されてもよい。

10

**【0094】**

モバイルコンピューティングデバイス450は、数ある構成要素の中で、プロセッサ452と、メモリ464と、ディスプレイ454などの入力/出力デバイスと、通信インターフェース466と、トランシーバ468とを含む。モバイルコンピューティングデバイス450はまた、追加のストレージを提供するために、マイクロドライブまたは他のデバイスなどの記憶デバイスを設けてもよい。プロセッサ452、メモリ464、ディスプレイ454、通信インターフェース466、およびトランシーバ468の各々は、様々なバスを使用して相互接続され、構成要素のうちのいくつかは、一般的なマザーボード上に取り付けられるか、または必要に応じて他の方法で取り付けられてもよい。

20

**【0095】**

プロセッサ452は、メモリ464内に記憶された命令を含む、モバイルコンピューティングデバイス450内の命令を実行することができる。プロセッサ452は、別個の複数のアナログおよびデジタルプロセッサを含むチップのチップセットとして実装されてもよい。プロセッサ452は、例えば、ユーザインターフェースの制御、モバイルコンピューティングデバイス450によって実行されるアプリケーション、およびモバイルコンピューティングデバイス450によるワイヤレス通信などの、モバイルコンピューティングデバイス450の他の構成要素の調整を提供してもよい。

30

**【0096】**

プロセッサ452は、ディスプレイ454に結合された制御インターフェース458およびディスプレイインターフェース456を介してユーザと通信してもよい。ディスプレイ454は、例えば、TFT(薄膜トランジスタ液晶ディスプレイ)ディスプレイ、もしくはOLED(有機発光ダイオード)ディスプレイ、または他の適切なディスプレイ技術であってもよい。ディスプレイインターフェース456は、グラフィカル情報および他の情報をユーザに提示するためにディスプレイ454を駆動するための適切な回路を備えてもよい。制御インターフェース458は、ユーザからコマンドを受信し、それらをプロセッサ452に提出するために変換してもよい。加えて、外部インターフェース462は、他のデバイスとモバイルコンピューティングデバイス450との近距離通信を可能にするように、プロセッサ452との通信を提供してもよい。外部インターフェース462は、例えば、いくつかの実装形態では有線通信、または他の実装形態ではワイヤレス通信を提供してもよく、また、複数のインターフェースが使用されてもよい。

40

**【0097】**

メモリ464は、モバイルコンピューティングデバイス450内の情報を記憶する。メモリ464は、コンピュータ可読媒体、揮発性メモリユニット、または不揮発性メモリユニットのうちの1つまたは複数として実装することができる。また、拡張メモリ474が設け

50

られ、例えば、SIMM(シングルインラインメモリモジュール)カードインターフェースを含んでもよい拡張メモリインターフェース472を介して、モバイルコンピューティングデバイス450に接続されてもよい。拡張メモリ474は、モバイルコンピューティング450のための追加の記憶空間を提供してもよく、またはモバイルコンピューティングデバイス450のためのアプリケーションもしくは他の情報を記憶してもよい。具体的には、拡張メモリ474は、上記で説明したプロセスを実行または補足するための命令を含んでもよく、また、安全な情報を含んでもよい。したがって、例えば、拡張メモリ474は、モバイルコンピューティングデバイス450のためのセキュリティモジュールとして設けられてもよく、モバイルコンピューティングデバイス450の安全な使用を可能にする命令でプログラムされてもよい。加えて、安全なアプリケーションが、識別情報をハッキングできない方法でSIMMカード上に配置するなどして、追加情報とともにSIMMカードを介して提供されてもよい。

**【0098】**

メモリは、例えば、上記で論じたように、フラッシュメモリおよび/またはNVRAMメモリ(不揮発性ランダムアクセスメモリ)を含んでもよい。いくつかの実装形態において、命令は、情報キャリア内に記憶される。命令は、1つまたは複数の処理デバイス(例えば、プロセッサ452)によって実行されると、上記で説明した方法などの1つまたは複数の方法を実行する。命令はまた、1つまたは複数のコンピュータまたは機械可読媒体(例えば、メモリ464、拡張メモリ474、またはプロセッサ452上のメモリ)などの1つまたは複数の記憶デバイスによって記憶することもできる。いくつかの実装形態において、命令は、例えば、トランシーバ468または外部インターフェース462を介して、伝搬信号において受信することができる。

**【0099】**

モバイルコンピューティングデバイス450は、必要に応じてデジタル信号処理回路を含んでもよい通信インターフェース466を介してワイヤレスで通信してもよい。通信インターフェース466は、とりわけ、GSM音声通話(汎欧州デジタル移動電話方式)、SMS(ショートメッセージサービス)、EMS(拡張メッセージングサービス)、もしくはMMSメッセージング(マルチメディアメッセージングサービス)、CDMA(符号分割多元接続)、TDMA(時分割多元接続)、PDC(パーソナルデジタルセルラ)、WCDMA(登録商標)(広帯域符号分割多元接続)、CDMA2000、またはGPRS(汎用パケット無線サービス)などの、様々なモードまたはプロトコルの下での通信を提供してもよい。そのような通信は、例えば、無線周波数を使用するトランシーバ468を介して発生してもよい。加えて、Bluetooth、WiFi、または他のそのようなトランシーバ(図示せず)などを使用して、短距離通信が発生してもよい。加えて、GPS(全地球測位システム)レシーバモジュール470は、追加のナビゲーションおよび位置関連のワイヤレスデータをモバイルコンピューティングデバイス450に提供してもよく、このワイヤレスデータは、モバイルコンピューティングデバイス450上で実行されるアプリケーションによって適宜使用されてもよい。

**【0100】**

モバイルコンピューティングデバイス450はまた、オーディオコーデック460を使用して聴覚的に通信してもよく、オーディオコーデック460は、ユーザから発話情報を受信し、それを使用可能なデジタル情報に変換してもよい。オーディオコーデック460は、同様に、例えば、モバイルコンピューティングデバイス450のハンドセット内のスピーカを介するなどして、ユーザのための可聴音を生成してもよい。そのような音は、音声電話通話からの音を含んでもよく、録音された音(例えば、音声メッセージ、音楽ファイルなど)を含んでもよく、また、モバイルコンピューティングデバイス450上で動作するアプリケーションによって生成された音を含んでもよい。

**【0101】**

モバイルコンピューティングデバイス450は、図に示すように、いくつかの異なる形態において実装されてもよい。例えば、モバイルコンピューティングデバイス450は、携帯電話480として実装されてもよい。モバイルコンピューティングデバイス450はま

た、スマートフォン482、携帯情報端末、または他の同様のモバイルデバイスの一部として実装されてもよい。

【0102】

本明細書で説明されているシステムおよび技法の様々な実装形態は、デジタル電子回路、集積回路、特別に設計されたASIC(特定用途向け集積回路)、コンピュータハードウェア、ファームウェア、ソフトウェア、および/またはそれらの組合せにおいて実現することができる。これらの様々な実装形態は、記憶システムからデータおよび命令を受信し、記憶システムにデータおよび命令を送信するように結合された、専用または汎用であってもよい少なくとも1つのプログラム可能なプロセッサと、少なくとも1つの入力デバイスと、少なくとも1つの出力デバイスとを含むプログラム可能なシステム上で実行可能および/または解釈可能な1つまたは複数のコンピュータプログラムにおける実装形態を含むことができる。

10

【0103】

これらのコンピュータプログラム(プログラム、ソフトウェア、ソフトウェアアプリケーション、またはコードとしても知られる)は、プログラム可能なプロセッサのための機械命令を含み、高水準の手続き型および/もしくはオブジェクト指向プログラミング言語において、ならびに/またはアセンブリ/機械語において実装することができる。本明細書で使用されている場合、機械可読媒体およびコンピュータ可読媒体という用語は、機械可読信号として機械命令を受信する機械可読媒体を含む、機械命令および/またはデータをプログラム可能なプロセッサに提供するために使用される任意のコンピュータプログラム製品、装置、および/またはデバイス(例えば、磁気ディスク、光ディスク、メモリ、プログラマブルロジックデバイス(PLD))を指す。機械可読信号という用語は、機械命令および/またはデータをプログラム可能なプロセッサに提供するために使用される任意の信号を指す。

20

【0104】

ユーザとの対話を提供するために、本明細書で説明されているシステムおよび技法は、情報をユーザに表示するためのディスプレイデバイス(例えば、CRT(陰極線管)またはLCD(液晶ディスプレイ)モニタ)と、ユーザが入力をコンピュータに提供することができるキーボードおよびポインティングデバイス(例えば、マウスまたはトラックボール)とを有するコンピュータ上に実装することができる。同様にユーザとの対話を提供するために他の種類のデバイスを使用することができ、例えば、ユーザに提供されるフィードバックは、任意の形態の感覚フィードバック(例えば、視覚的フィードバック、聴覚的フィードバック、または触覚的フィードバック)とすることができ、ユーザからの入力、音響、音声、または触覚入力を含む任意の形式において受信することができる。

30

【0105】

本明細書で説明されているシステムおよび技法は、バックエンド構成要素(例えば、データサーバとして)を含むコンピューティングシステム、またはミドルウェア構成要素(例えば、アプリケーションサーバ)を含むコンピューティングシステム、またはフロントエンド構成要素(例えば、ユーザが本明細書で説明されているシステムおよび技法の実装形態と対話することができるグラフィカルユーザインターフェースまたはウェブブラウザを有するクライアントコンピュータ)を含むコンピューティングシステム、またはそのようなバックエンド構成要素、ミドルウェア構成要素、もしくはフロントエンド構成要素の任意の組合せを含むコンピューティングシステムにおいて実装することができる。システムの構成要素は、任意の形式または媒体のデジタルデータ通信(例えば、通信ネットワーク)によって相互接続することができる。通信ネットワークの例は、ローカルエリアネットワーク(LAN)、ワイドエリアネットワーク(WAN)、およびインターネットを含む。いくつかの実装形態において、本明細書で説明されているシステムおよび技法は、音声認識および他の処理がデバイス上で直接実行される組み込みシステム上に実装することができる。

40

【0106】

コンピューティングシステムは、クライアントとサーバとを含むことができる。クライ

50

アントおよびサーバは、一般に、互いに離れた場所にあり、通常、通信ネットワークを介して対話する。クライアントおよびサーバの関係は、それぞれのコンピュータ上で実行され、互いにクライアント～サーバの関係を有するコンピュータプログラムによって生じる。

【0107】

いくつかの実装形態について上記で詳細に説明したが、他の変更が可能である。例えば、クライアントアプリケーションについて、デリゲート(delegate)にアクセスするものとして説明したが、他の実装形態において、デリゲートは、1つまたは複数のサーバ上で実行されるアプリケーションなどの、1つまたは複数のプロセッサによって実装される他のアプリケーションによって使用されてもよい。加えて、図に示す論理フローは、所望の結果を達成するために、示されている特定の順序、または連続した順序を必要としない。加えて、他のアクションが提供されてもよく、または説明したフローからアクションが削除されてもよく、説明したシステムに他の構成要素が追加されてもよく、または説明したシステムから削除されてもよい。したがって、他の実装形態は、添付の特許請求の範囲内にある。

10

【符号の説明】

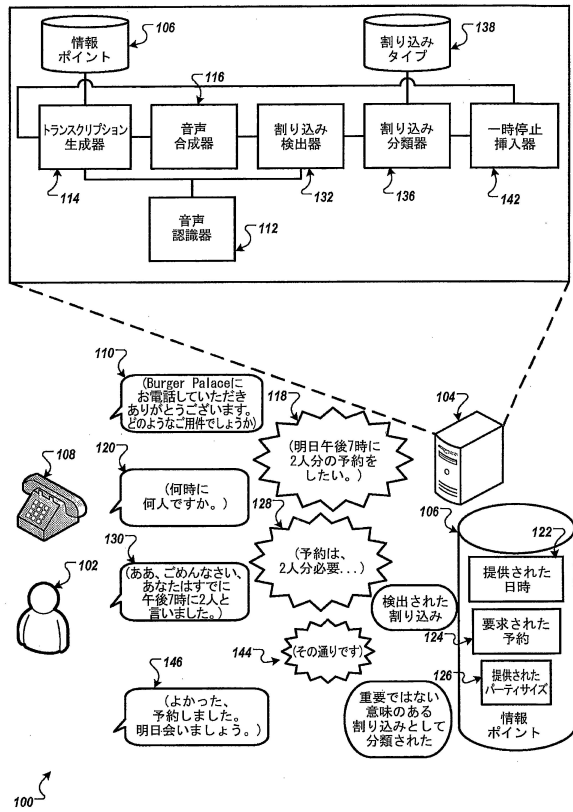
【0108】

100	システム	
102	代表者	
104	通話開始システム	20
106	情報ポイント	
108	電話	
110	発話	
112	音声認識器	
114	トランスクリプション生成器	
116	音声合成器	
118	合成音声	
120	発話	
122	提供された日時	
124	要求された予約	30
126	提供されたパーティサイズ	
128	合成音声	
130	発話、割り込み発話	
132	割り込み検出器	
136	割り込み分類器	
138	割り込みタイプ	
142	一時停止挿入器	
144	合成音声	
146	発話	
200	システム	40
202	第1のユーザ	
204	第2のユーザ	
206	電話	
208	電話	
210	通話監視システム	
212	発話	
214	発話	
216	発話	
218	合成音声	
220	発話	50

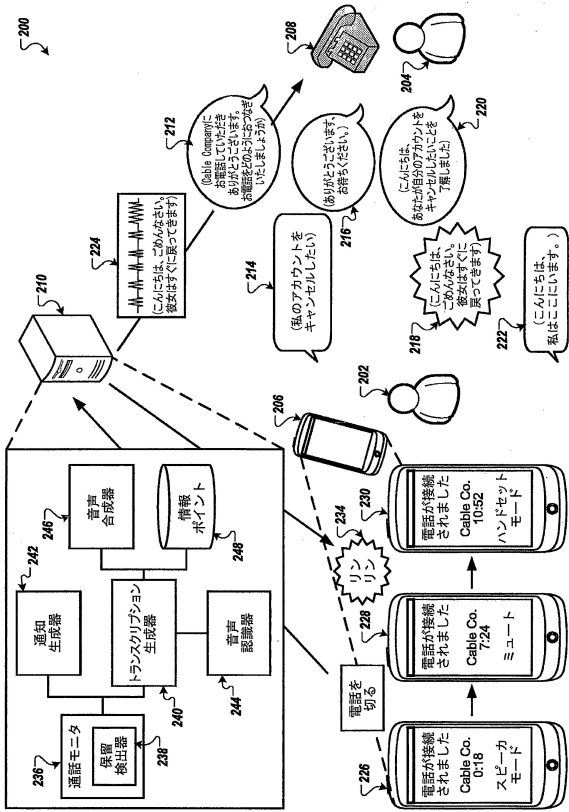
222	発話	
224	オーディオデータ	
226	第1の状態	
228	第2の状態	
230	第3の状態	
234	通知	
236	通話モニタ	
238	保留検出器	
240	トランスクリプション生成器	
242	通知生成器	10
244	音声認識器	
246	音声合成器	
248	情報ポイント	
400	コンピューティングデバイス	
402	プロセッサ	
404	メモリ	
406	記憶デバイス	
408	高速インターフェース	
410	高速拡張ポート	
412	低速インターフェース	20
414	低速拡張ポート	
416	ディスプレイ	
420	サーバ	
422	ラップトップコンピュータ	
424	ラックサーバシステム	
450	モバイルコンピューティングデバイス	
452	プロセッサ	
454	ディスプレイ	
456	ディスプレイインターフェース	
458	制御インターフェース	30
460	オーディオコーデック	
462	外部インターフェース	
464	メモリ	
466	通信インターフェース	
468	トランシーバ	
470	GPS(全地球測位システム)レシーバモジュール	
472	拡張メモリインターフェース	
474	拡張メモリ	
480	携帯電話	
482	スマートフォン	40

【 図 面 】

【 図 1 】



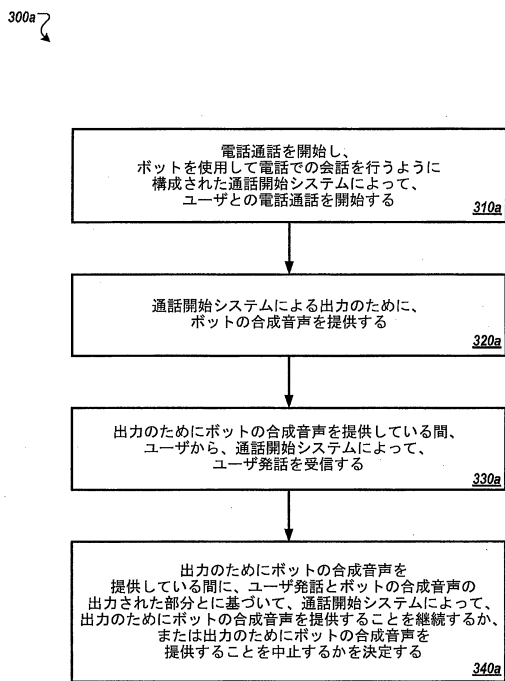
【 図 2 】



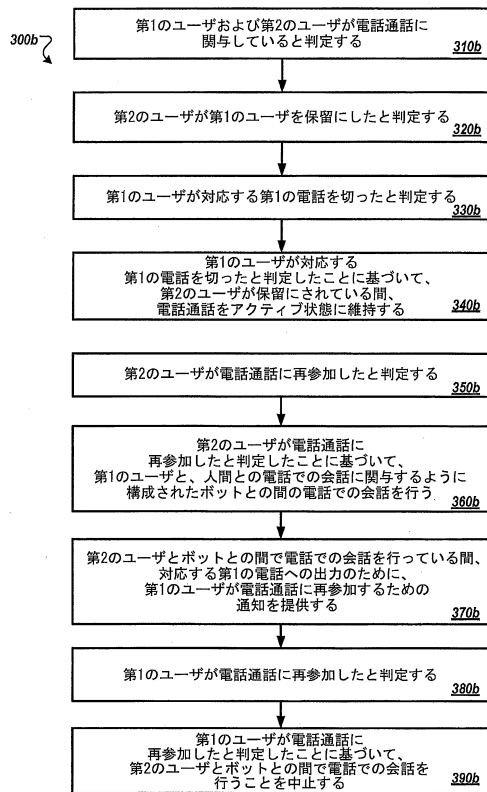
10

20

【 図 3 A 】



【 図 3 B 】



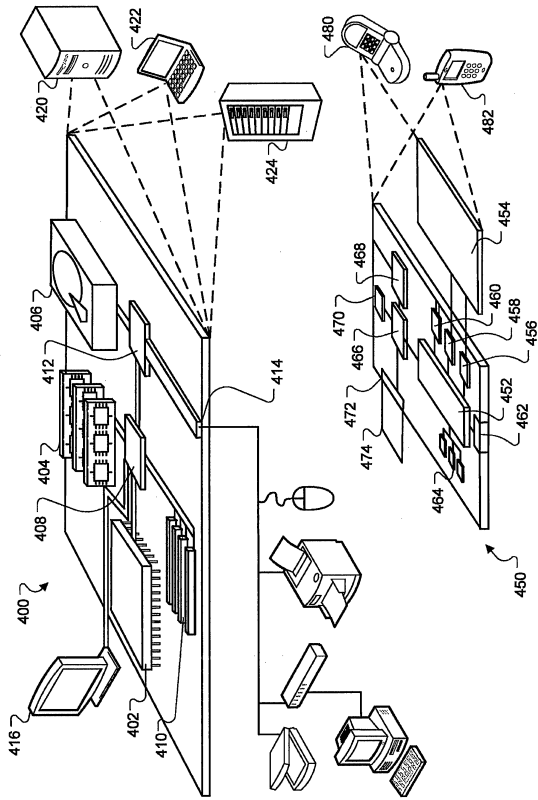
30

40

50



【 図 4 】



10

20

30

40

50

## 【 手続補正書 】

【 提出日 】 令和 4 年 10 月 17 日 ( 2022.10.17 )

## 【 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 】

1 つまたは複数のプロセッサによって実装される方法であって、

第 1 のユーザおよび第 2 のユーザが電話通話に参与していることを判定するステップであって、前記第 1 のユーザが対応する第 1 のコンピューティングデバイスに関連付けられており、前記第 2 のユーザが対応する第 2 のコンピューティングデバイスに関連付けられている、ステップと、

前記第 2 のユーザが前記第 1 のユーザを保留にしたことを判定するステップと、

前記第 1 のユーザが前記対応する第 1 のコンピューティングデバイスを切ったことを判定するステップと、

前記第 1 のユーザが前記対応する第 1 の電話を切ったとの判定に回答して、前記第 1 のユーザが保留にされている間、

前記電話通話をアクティブ状態に維持するステップと、

前記第 2 のユーザが前記電話通話に再参加したかどうかを判定するステップと

前記第 2 のユーザが前記電話通話に再参加したとの判定に回答して、  
前記第 1 のユーザの前記対応する第 1 のコンピューティングデバイスに関連付けられたボットに、前記第 2 のユーザとの前記電話通話を継続させるステップと、

前記ボットに、前記対応する第 1 のコンピューティングデバイスにおける出力のために、前記第 1 のユーザが前記電話通話に再参加するための通知を提供させるステップと、

前記第 1 のユーザが前記電話通話に再参加したことを判定するステップと、

前記第 1 のユーザが前記電話通話に再参加したとの判定に回答して、前記ボットに前記第 2 のユーザとの前記電話での会話を中止させるステップと

を含む、

方法。

## 【 請求項 2 】

前記第 1 のユーザの前記対応する第 1 のコンピューティングデバイスに関連付けられた前記ボットに、前記第 2 のユーザとの前記電話通話を継続させるステップが、

前記ボットに、前記第 1 のユーザが電話通話におけるアクティブな参加者ではないことを示す前記ボットの合成音声を生成させるステップと、

前記ボットに、前記第 2 のユーザの前記対応する第 2 のコンピューティングデバイスにおける出力のために、前記ボットの前記合成音声を提供させるステップと

を含む、

請求項 1 に記載の方法。

## 【 請求項 3 】

前記第 1 のユーザの前記対応する第 1 のコンピューティングデバイスに関連付けられた前記ボットに、前記第 2 のユーザとの前記電話通話を継続させるステップが、

前記ボットに、前記電話通話を開始する前記第 1 のユーザに関する 1 つまたは複数の詳細を伝達する前記ボットの合成音声を生成させるステップと、

前記ボットに、前記第 2 のユーザの前記対応する第 2 のコンピューティングデバイスにおける出力のために、前記ボットの前記合成音声を提供させるステップと

を含む、

請求項 1 に記載の方法。

## 【 請求項 4 】

10

20

30

40

50

前記第1のユーザが前記電話通話に再参加していないとの判定に応答して、  
前記第1のユーザの前記対応する第1のコンピューティングデバイスに関連付けられた前記ボットに、前記第2のユーザを保留にさせるステップをさらに含む、  
請求項1に記載の方法。

【請求項5】

前記第1のユーザの前記対応する第1のコンピューティングデバイスに関連付けられた前記ボットに、前記第2のユーザを保留にさせることについてのしきい値持続時間以内に、  
前記第1のユーザが前記電話通話に再参加していないとの判定に応答して、  
前記ボットに前記電話通話を終了させるステップをさらに含む、  
請求項4に記載の方法。

10

【請求項6】

前記第1のユーザが前記電話通話に再参加する前に、前記第2のユーザからユーザ発話を受信するステップと、  
前記ボットに、前記ユーザ発話のトランスクリプションを生成させるステップと、  
前記ボットに、前記第1のユーザが前記電話通話に再参加するための前記通知を前記ユーザ発話の前記トランスクリプションに含めさせるステップと  
 をさらに含む、  
 請求項1に記載の方法。

【請求項7】

前記電話通話を前記アクティブ状態に維持するステップが、  
 前記対応する第1のコンピューティングデバイスの1つまたは複数のマイクロフォンをミュートするステップと、  
 前記対応する第1のコンピューティングデバイスと前記対応する第2のコンピューティングデバイスとの間の電話接続を維持するステップと  
 を含む、  
 請求項1に記載の方法。

20

【請求項8】

前記第1のユーザが前記電話通話に再参加したとの判定に応答して、  
前記対応する第1のコンピューティングデバイスの前記1つまたは複数のマイクロフォンをミュートしないステップ  
をさらに含む、請求項7に記載の方法。

30

【請求項9】

前記第1のユーザの前記対応する第1のコンピューティングデバイスに関連付けられた前記ボットに、前記第2のユーザとの前記電話通話を継続させるステップが、  
前記第1のユーザが前記対応する第1のコンピューティングデバイスでの出力のために前記電話通話に再参加するための前記通知を前記ボットに提供させることについてのしきい値持続時間以内に、前記第1のユーザが前記電話通話に再参加したとの判定に基づいて、  
前記ボットに前記ボットの任意の合成音声を生成させることを控えるステップ  
 を含む、請求項1に記載の方法。

【請求項10】

前記ボットが、前記第1のコンピューティングデバイスから離れたリモートサーバにおいて実装される、請求項1に記載の方法。

40

【請求項11】

前記ボットが、前記第1のコンピューティングデバイスにおいてローカルに実装される、  
請求項1に記載の方法。

【請求項12】

1つまたは複数のコンピュータと、  
 前記1つまたは複数のコンピュータによって実行されたときに、前記1つまたは複数のコンピュータに請求項1から11のいずれか一項に記載の方法を実行させるように動作可能な命令を記憶した1つまたは複数の記憶デバイスと

50

を備えるシステム。

【請求項13】

1つまたは複数のコンピュータによって実行可能であり、実行時に、前記1つまたは複数のコンピュータに請求項1から11のいずれか一項に記載の方法を実行させる命令を備えるソフトウェアを記憶した非一時的なコンピュータ可読記録媒体。

【外国語明細書】

2023002650000007.pdf

10

20

30

40

50

---

フロントページの続き

- (72)発明者 アザフ・アハロニ  
アメリカ合衆国・カリフォルニア・94043・マウンテン・ビュー・アンフィシアター・パーク  
ウェイ・1600
- (72)発明者 エイアル・セガリス  
アメリカ合衆国・カリフォルニア・94043・マウンテン・ビュー・アンフィシアター・パーク  
ウェイ・1600
- (72)発明者 ヤニフ・レヴィアタン  
アメリカ合衆国・カリフォルニア・94043・マウンテン・ビュー・アンフィシアター・パーク  
ウェイ・1600