

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第5124573号
(P5124573)

(45) 発行日 平成25年1月23日 (2013. 1. 23)

(24) 登録日 平成24年11月2日 (2012. 11. 2)

(51) Int. Cl.

F I

G 1 0 L 15/10 (2006. 01)

G 1 0 L 15/10 5 0 0 Z

G 1 0 L 15/00 (2013. 01)

G 1 0 L 15/00 2 0 0 A

G 1 0 L 25/78 (2013. 01)

G 1 0 L 11/00 4 0 2 L

請求項の数 11 (全 14 頁)

(21) 出願番号 特願2009-519430 (P2009-519430)
 (86) (22) 出願日 平成19年5月15日 (2007. 5. 15)
 (65) 公表番号 特表2009-543158 (P2009-543158A)
 (43) 公表日 平成21年12月3日 (2009. 12. 3)
 (86) 国際出願番号 PCT/US2007/011567
 (87) 国際公開番号 W02008/008117
 (87) 国際公開日 平成20年1月17日 (2008. 1. 17)
 審査請求日 平成22年5月17日 (2010. 5. 17)
 (31) 優先権主張番号 11/485, 011
 (32) 優先日 平成18年7月12日 (2006. 7. 12)
 (33) 優先権主張国 米国 (US)

(73) 特許権者 500046438
 マイクロソフト コーポレーション
 アメリカ合衆国 ワシントン州 9805
 2-6399 レッドモンド ワン マイ
 クロソフト ウェイ
 (74) 代理人 100077481
 弁理士 谷 義一
 (74) 代理人 100088915
 弁理士 阿部 和夫
 (72) 発明者 アレハンドロ アセロ
 アメリカ合衆国 98052 ワシントン
 州 レッドモンド ワン マイクロソフト
 ウェイ マイクロソフト コーポレーシ
 ョン インターナショナル パテンツ内

最終頁に続く

(54) 【発明の名称】 音声認識機能を使用した応答マシンの検出

(57) 【特許請求の範囲】

【請求項 1】

コール受信側が実際の人間であるか、応答マシンであるかを確かめるために、コンピュータで実施される方法であって、

コール受信側から可聴応答を受信し、言語モデルを有する音声認識機能により前記可聴応答を処理して、認識された音声を示すテキスト形式の出力を提供するステップと、

前記可聴応答に関連する非ワード特徴を確認するとともに、実際の人間によって使用されるフレーズと応答マシンによって使用されるフレーズとに関してトレーニングされた分類機能を用いて、前記テキスト形式の前記認識された音声を示す出力を処理し、前記コール受信側が実際の人間であるか応答マシンであるかを示す出力を提供するステップであって、前記可聴応答に関連する前記非ワード特徴は、少なくとも前記可聴応答の持続時間を含み、前記認識された音声を示す出力の前記処理は、前記非ワード特徴と前記認識された音声を示す出力の統計分析とに基づくものであり、前記可聴応答の前記持続時間が選択された範囲内にあるときに、前記認識された音声を示す出力を前記分類機能によって分析して、前記認識された音声を示す出力が前記実際の人間によって使用されるフレーズに統計的により一致するか、あるいは前記応答マシンによって使用されるフレーズに統計的により一致するかを判断することを含む、ステップと

を含むことを特徴とする方法。

【請求項 2】

前記非ワード特徴はさらに、前記可聴応答における沈黙の存在を含み、前記認識された

10

20

音声を示す出力を処理することは、前記可聴応答において最初に沈黙が存在するかどうかを確かめて、該沈黙の存在を基礎として使用し、前記コール受信側が実際の人間であるか応答マシンであるかを示す出力を提供することを含むことを特徴とする、請求項1に記載の方法。

【請求項3】

前記認識された音声を示す出力を処理することは、前記可聴応答において、選択した長さの期間の沈黙が最初に存在するかどうかを確かめて、該沈黙の存在を基礎として使用し、前記コール受信側が実際の人間であるか応答マシンであるかを示す出力を提供することを含むことを特徴とする、請求項2に記載の方法。

【請求項4】

前記非ワード特徴はさらに、前記音声認識機能が前記可聴応答内のいずれかの音声を認識できたかどうかの表示を含み、前記認識された音声を示す出力を処理することは、前記音声認識機能が前記可聴応答内のいずれかの音声を認識できたかどうかの前記表示を基礎として使用し、前記コール受信側が実際の人間であるか応答マシンであるかを示す出力を提供することを含むことを特徴とする、請求項1に記載の方法。

【請求項5】

コール受信側が実際の人間であるか応答マシンであるかを確かめるためのプログラムを記録したコンピュータ読取可能な記録媒体であって、前記プログラムはコンピュータに、
コール受信側から可聴応答を受信し、言語モデルを有する音声認識機能により前記可聴

応答を処理して、認識された音声を示すテキスト形式の出力を提供することと、
前記可聴応答に関連する非ワード特徴を確認するとともに、実際の人間によって使用されるフレーズと応答マシンによって使用されるフレーズとに関してトレーニングされた分類機能を用いて、前記テキスト形式の前記認識された音声を示す出力を処理し、前記コール受信側が実際の人間であるか応答マシンであるかを示す出力を提供することと

を実行させ、前記可聴応答に関連する前記非ワード特徴は、少なくとも前記可聴応答の持続時間を含み、前記認識された音声を示す出力の前記処理は、前記非ワード特徴と前記テキスト形式の前記認識された音声を示す出力のワードのフレーズの統計分析とに基づくものであり、前記可聴応答の前記持続時間が選択された範囲内にあるときに、前記認識された音声を示す出力を前記分類機能によって分析して、前記認識された音声を示す出力が前記実際の人間によって使用されるフレーズに統計的により一致するか、あるいは前記応答マシンによって使用されるフレーズに統計的により一致するかを判断することを含むことを特徴とするコンピュータ読取可能な記録媒体。

【請求項6】

前記非ワード特徴はさらに、前記可聴応答における沈黙の存在を含み、前記認識された音声を示す出力を処理することは、前記可聴応答において最初に沈黙が存在するかどうかを確かめて、該沈黙の存在を基礎として使用し、前記コール受信側が実際の人間であるか応答マシンであるかを示す出力を提供することを含むことを特徴とする、請求項5に記載のコンピュータ読取可能な記録媒体。

【請求項7】

前記認識された音声を示す出力を処理することは、前記可聴応答において、選択した長さの期間の沈黙が最初に存在するかどうかを確かめて、該沈黙の存在を基礎として使用し、前記コール受信側が実際の人間であるか応答マシンであるかを示す出力を提供することを含むことを特徴とする、請求項6に記載のコンピュータ読取可能な記録媒体。

【請求項8】

前記非ワード特徴はさらに、前記音声認識機能が前記可聴応答内のいずれかの音声を認識できたかどうかの表示を含み、前記認識された音声を示す出力を処理することは、前記音声認識機能が前記可聴応答の中の音声を認識できたかどうかの前記表示を基礎として使用し、前記コール受信側が実際の人間であるか応答マシンであるかを示す出力を提供することを含むことを特徴とする、請求項5に記載のコンピュータ読取可能な記録媒体。

【請求項9】

前記非ワード特徴はさらに、前記音声認識機能が前記コール受信側によるバージン・イベントを検出したことを含み、前記バージン・イベントが検出された場合、前記プログラムは、前記コンピュータに、前記コール受信側に対するメッセージ全体がプレイされるまでメッセージをリプレイすることをさらに実行させ、前記バージン・イベントを検出することは、前記コール受信側による挨拶中の沈黙期間に続く各部分を検出することを含むことを特徴とする、請求項 5 に記載のコンピュータ読取可能な記録媒体。

【請求項 10】

前記コール受信側に対するメッセージをリプレイすることは、前記コール受信側が前記応答マシンであるときに前記コール受信側に対するメッセージをリプレイすることを含むことを特徴とする、請求項 9 に記載のコンピュータ読取可能な記録媒体。

10

【請求項 11】

コール受信側へのメッセージを応答マシンに残すためにコンピュータで実施される方法であって、

コール受信側から可聴応答を受信し、言語モデルを有する音声認識機能で前記可聴応答を処理し、認識された音声を示す出力を提供するステップと、

前記可聴応答に関連する非ワード特徴を確認するとともに、実際の人間によって使用されるフレーズと応答マシンによって使用されるフレーズとに関してトレーニングされた分類機能を用いて、前記テキスト形式の前記認識された音声を示す出力を処理し、前記コール受信側が実際の人間であるか応答マシンであるかを検出するステップであって、前記可聴応答に関連する前記非ワード特徴は、少なくとも前記可聴応答の持続時間を含み、前記認識された音声を示す出力の前記処理は、前記非ワード特徴と前記テキスト形式の前記認識された音声を示す出力のワードのフレーズの統計分析とに基づくものであり、前記可聴応答の前記持続時間が選択された範囲内にあるときに、前記認識された音声を示す出力を前記分類機能によって分析して、前記認識された音声を示す出力が前記実際の人間によって使用されるフレーズに統計的により一致するか、あるいは前記応答マシンによって使用されるフレーズに統計的により一致するかを判断することを含む、ステップと、

20

前記コール受信側が前記応答マシンであるとき、

前記音声認識機能を用いて前記応答マシンによるバージン・イベントを検出するステップであって、前記バージン・イベントを検出することは、前記応答マシンによる挨拶中の沈黙期間に続く各部分を検出することを含む、ステップと、前記応答マシンによる前記バージン・イベントを検出する毎に、前記応答マシンにおいて前記コール受信側へのメッセージ全体がプレイされるまでメッセージを繰り返しリスタートするステップと

30

を含むこと特徴とする方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、音声認識機能を使用して応答マシンを検出することに関する。

【背景技術】

【0002】

40

以下の説明は、単に一般的な背景情報のために提供されるものであり、特許請求する主題の範囲を決定する際の助けとして使用されることは意図されていない。

【0003】

自動化されたアウトバウンド電話コールシステム(outbound telephone calling system)の使用は、極めて普通のことである。このようなシステムは、マーケティング目的で使われるほかに、いくつか他の適用の例を挙げると、アポイントメントを確認または変更するのに患者に連絡するために医院やクリニックによって使用されること、スケジュール変更を患者に知らせるために学校によって使用されること、寄付を得るためにチャリティによって使用されること、通知または他の情報を提供するために政府機関によって使用されることがある。

50

【 0 0 0 4 】

多くの事例において、電話のコールの受信側 (r e c i p i e n t) が実際の人間 (a c t u a l p e r s o n) であるか、応答マシン (a n s w e r i n g m a c h i n e) であるかを自動的に確かめることが必要になることもあれば、役立つこともある。実際の人間が応答したのか、または応答マシンが使用されたのかに応じて、異なるアクションがアウトバウンド電話コールシステムによってとられることがある。しかし、このタスク、すなわち、コール分析 (c a l l a n a l y s i s) は困難であり、現在は不正確である。

【 先行技術文献 】

【 非特許文献 】

10

【 0 0 0 5 】

【 非特許文献 1 】 D. Yu, Y.-C. Ju, Y. Wang および A. Acero 著 「 N-Gram Based Filler Model for Robust Grammar Authoring 」、 Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, May 2006

【 発明の概要 】

【 発明が解決しようとする課題 】

【 0 0 0 6 】

コール分析は、一般に、ハードウェアのスイッチレベルで行われる。この分析は、受信側が初期にコールをピックアップしたときから該コールがテレフォニアプリケーション (t e l e p h o n y a p p l i c a t i o n) に接続されるまでの短い間隔 (i n t e r v a l) を使用して実施される。この間隔の間に、受信側が話し始めると、システムは、受信側が実際の人間であるか、応答マシンであるかについて判断するために、例えば、可聴信号 (a u d i b l e s i g n a l) のエネルギー含量、強度、または他の信号パラメータについて受信した可聴信号を処理することになる。重要なことは、コールがこの段階でピックアップされたことがテレフォニアプリケーションには分かっていないので、どのような初期プロンプト (i n i t i a l p r o m p t) も出されないことを理解することである。従って、回線の反対側では、受信側がコールに応答して「ハロー」などの挨拶をしたとしても、システムはコール分析を行なっているので、受信側には沈黙 (s i l e n c e) が返答として聞こえるだけである。多くの事例では、受信側はこの後、単に電話を切ることになる。

20

30

【 課題を解決するための手段 】

【 0 0 0 7 】

以下の課題を解決するための手段および添付の要約書は、以下に発明を実施するための形態において詳述される概念の選択を簡略化した形式で紹介するために提供される。この課題を解決するための手段および添付の要約書は、特許請求される主題の重要な特徴または本質的特徴を特定するように意図されたものではなく、特許請求される主題の範囲を決定する際の助けとして使用されるように意図されたものでもない。さらに、特許請求される主題は、背景技術において指摘した不都合点のいずれか、またはすべてを解決するための実施形態には限定されない。

【 0 0 0 8 】

40

応答マシン検出モジュール (a n s w e r i n g m a c h i n e d e t e c t i o n m o d u l e) は、コール受信側が実際の人間であるか、応答マシンであるかを判断するのに使用される。応答マシン検出モジュールは、音声認識機能 (s p e e c h r e c o g n i z e r)、およびコール分析モジュール (c a l l a n a l y s i s m o d u l e) を含む。音声認識機能は、コールに対するコール受信側の可聴応答 (a u d i b l e r e s p o n s e) を受信する。音声認識機能は、その可聴応答を処理し、認識された音声を示す出力を提供する。コール分析モジュールは、音声認識機能の出力を処理し、コール受信側が実際の人間であるか、応答マシンであるかを示す出力を生成する。一実施形態では、コール分析モジュールは、音声認識機能からの出力の統計分析を提供してコール受信側が実際の人間であるか応答マシンであるかを判断する、分類モジュール (c

50

lassifier module)を含むことが可能である。

【0009】

また、発信側(caller)からのメッセージ全体が応答マシンによって記録されたことを確かめる手法も説明されている。具体的には、音声認識機能は、応答マシンによるバージイン(barge-in)・イベントを検出するように動作し、バージイン・イベントが検出された場合には、そのメッセージが応答マシンにリプレイされる。これにより、メッセージが一回または複数回リプレイされることがあるが、これが特に利点があるのは、応答マシンの挨拶が終了していないため応答マシンがメッセージを記録する準備状態にないことが任意のバージイン・イベントによって示されるからである。各バージイン・イベントのあとでメッセージをリプレイすることにより、最後のバージイン・イベントのあとメッセージをプレイすると、応答マシンがメッセージを記録する準備状態になったときメッセージ全体が記録されることが保証される。

10

【図面の簡単な説明】

【0010】

【図1】ダイヤルシステムを表す概略ブロック図である。

【図2】図1に示すダイヤルシステムのコンポーネントを表す概略ブロック図である。

【図3】コールを処理するための概略ブロック図である。

【図4】コールを処理するためのフローチャートを示す図である。

【図5】コールを処理するためのより詳細なフローチャートを示す図である。

【図6】例示的なコンピューティング環境を示す図である。

20

【発明を実施するための形態】

【0011】

図1および図2は、アウトバウンドコールまたはダイヤルシステム100の概略ブロック図である。システム100は、例えば、以下で図6に関連して説明されるコンピューティング環境のひとつの中で実現され、または他の適切なコンピューティング環境の中で実現される。図1に図示されるように、アウトバウンドコールシステム100は、いずれか1つまたは複数のコール受信側102(図には、コール受信側102-1乃至102-Mが示されている)にコールを出す。本明細書において使用される「コール受信側」は、実際の人間、または応答マシンのいずれかとすることができる。コール受信側102-1乃至102-Mのひとつにコールを出したあと、アウトバウンドコールシステム100は、コール受信側が実際の人間であるかどうか、または応答マシンが使用されているかどうかを確かめる。

30

【0012】

図2は、一例示の実施形態にかかるアウトバウンドコールシステム100のコンポーネントまたはモジュールをより詳細に図示する。図示されるように、アウトバウンドコールシステム100は、1つまたは複数の受信側102にコールを行なう。応答マシン検出モジュール106を使用して、コール受信側102が実際の人間であるか、応答マシンであるかを判断する。典型的なアプリケーションでは、アプリケーションモジュール104は、例えば、電話番号データベース108から取り出されたコール受信側の電話番号のダイヤリング(dialing)を開始する。次いで、応答マシン検出モジュール106を使用して、コール受信側が実際の人間であるか、応答マシンであるかを確かめる。検出モジュール106の出力に従って異なるアクションがアウトバウンドコールシステム100によってとられることになる。例えば、コール受信側が実際の人間である場合、アプリケーションモジュール104は、コールをオペレータ110に転送することがあり、自動会話で受信側と話し中のこともある。あるいは、アプリケーションモジュール104は、コール受信側102の応答マシンにメッセージを残すのに使用されることがある。図2では、コール受信側102、電話番号データベース108、およびオペレータ110を破線で示し、これら自体はアウトバウンドコールシステム100の一部でないか、またはアウトバウンドコールシステム100内に必ずしも必要でないことを図示している。さらに、アウトバウンドコールシステム100のブロック図は、理解を容易にすることを目的とし、応

40

50

答マシン検出モジュール106とそのコンポーネントとの例示的な関係だけでなく、アウトバウンドコールシステム100の他のコンポーネントとの例示的な関係も提供するための概略図である。これは単なる1つの例示実施形態であり、別の実施形態における他の構成ではアウトバウンドコールシステム100および応答マシン検出モジュール106の1つまたは複数のコンポーネントが省略され、および/または組み合わされることがあることを理解されたい。

【0013】

図示の実施形態において、応答マシン検出モジュール106は、音声認識機能120およびコール分析モジュール122を含む。図3および図4は、それぞれ方法200の概略図およびフローチャートであり、これらのモジュールのオペレーションを図示している。ステップ202において、音声認識機能120は、アウトバウンドコールシステム100によって出されたコールに対するコール受信側102の可聴応答124を受信する。音声認識機能120は、該可聴応答124を処理し、認識された音声を示す出力126を提供する。この出力126は、一実施形態では、一般的にテキストデータの形式のフレーズ(*phrase*)である。本明細書において使用されるフレーズとは、コール受信側によって提供されるコンテキストの複数のワードである。

【0014】

ステップ204において、音声認識機能120からの出力126は、コール分析モジュール122に対する入力として提供される。コール分析モジュール122は、音声認識機能120の出力126を処理して、コール受信側が人間であるか応答マシンであるかを示す出力128を生成する。コール分析モジュール122は、出力126におけるフレーズのコンテンツについて統計分析を提供する分類モジュール123を含むことができる。音声認識機能120および分類モジュール123は、周知のモジュールであり、様々な周知の技法を使用して実装することができる。しかし、音声認識機能120および分類モデル132によって使用される言語モデル130(例えば、*N-gram*、文脈自由文法(*context-free grammar*)、ハイブリッド(*hybrid*)など)は、典型的に、コールに回答するとき実際の人間と応答マシンの両方によって使用されるフレーズまたは挨拶のみに基づいてトレーニングされる。例えば、人間は、電話コールに「ハロー」、「どのようなご用件ですか?」、「こちらは*Steve*です」という挨拶で回答することもあれば、単に名前だけで回答することもある。これに対して、応答マシンは、「こちらは応答マシンです。ただいま電話に出られません。メッセージをどうぞ。」という挨拶や、または単に「メッセージをどうぞ」という挨拶のこともある。しかし、コンテンツベースの応答マシン検出では、「ただいま電話に出られません」または「メッセージをどうぞ」などの重要なフレーズを認識することが必要になることがある。同様に、受信側の位置で電話スクリーンシステム(*phone screen system*)を通る場合には、「2を押してください(*press 2*)」などのフレーズを認識することが必要になることがある。これらのようなフレーズに対する認識精度を高めるために、言語モデル130は、重要なフレーズについてトレーニングすることができ、重要なフレーズに含まれていないワードをキャプチャするために*n-gram filler*モデルを用いてこのモデルをスムーズにすることができる。*n-gram filler*モデルの例は、非特許文献1に記載されている。十分な数の例(包括的な例またはアプリケーション固有の例のどちらか)を使用すると、言語モデル130と分類モデル132の両方を、適切にトレーニングすることができる。

【0015】

コール分析モジュール122およびステップ204に関して、非ワード特徴(*non-word features*)136を、音声認識機能120の出力126に加えて、またはその代替として分析に使用することもできる。非ワード特徴136の例には、コール受信側102が「バーズイン」するかどうか(すなわち、アプリケーション104がダイアログを実行しているときにプレイされているプロンプトがインタラプトされたか)、コールに回答する際にコール受信側102によって行なわれた可聴応答の持続時間、および

10

20

30

40

50

音声認識機能 120 が可聴応答を有効なフレーズとして認識できたか否かが含まれるが、これらには限定されない。図 3 において、非ワード特徴は、非ワード特徴ジェネレータ (non-word feature generator) 128 から生成され、該非ワード特徴ジェネレータ 128 は、可聴応答 124 またはアウトバウンドコールシステム 100 からのコールに関する他の入力情報 138 を受信して、処理することができる。図 4 において、非ワード特徴 136 の受信はステップ 203 で示されている。

【0016】

このポイントにおいて、音声認識機能 120 は、完全な可聴応答を認識できることは必要ではなく、むしろ、コール受信側 102 におけるノイズ環境、記録された挨拶の中のノイズまたは電話システムからのノイズ、ならびに応答の性質に起因して、可聴応答 124 の一部分または複数の部分だけが認識可能であり、従ってコール受信側 102 が実際の人間であるか、応答マシンであるかを確かめるために使用されることに留意されたい。

【0017】

図 5 は、図 3 のステップ 204 を実行する際にコール分析モジュール 122 によって実行される、例示的な方法 300 を図示するフローチャートである。ステップ 302 は、コールが出された後に音声認識が開始されることを表す。受信側が実際の人間であるか応答マシンであるかを確かめるのに任意的に使用されることが可能な第 1 の非ワード特徴は、コールがピックアップされたあと初めに沈黙があるかどうかである。具体的には、実際の人間は、例えば、話し始める前に 1 秒または 2 秒待つことがあるのでこれがわかる。従って、ステップ 304 で、コールがピックアップされたあと相当の沈黙が存在すると判断される場合、例えば、タイマによって測定するなどして、ステップ 302 から或る選択時間を越えると、コール受信側 102 が実際の人間であると判断することができる。

【0018】

他方、ステップ 302 でコールがピックアップされたあと相当の沈黙が存在しない場合、すなわち、ステップ 302 が存在していない場合、可聴応答の持続時間を使用して、コール受信側が実際の人間であるか否かを判断することができる。ステップ 306 および 308 は、可聴応答の持続時間を含む非ワード特徴を処理することを示している。ステップ 306 で、可聴応答の持続時間が非常に短い場合（例えば、およそ 1 秒未満）、コールに応答したのが実際の人間である可能性が高い。これに対して、可聴応答の持続時間がステップ 308 で示すように非常に長い場合（例えば、およそ 4 秒またはそれ以上）、コール受信側はおそらく応答マシンである。ステップ 306 および 308 を反対にすることもできるため、必ずしも図示される順序である必要はないことに留意されたい。同様に、指定された時間期間も調整することができる。

【0019】

コール受信側が実際の人間であるか応答マシンであるかの判断がステップ 308 の前に行なわれた場合、プロセスのフローは、可聴応答の持続時間を再び使用することができるステップ 310 に続く。具体的には、ステップ 310 では、可聴応答の持続時間を測定して、比較的短いか、例えば、2 秒または 3 秒であるかどうかを判断する。このステップの結果は、分類モジュール 123 から得られるような、音声認識機能 120 によって認識された可聴応答のコンテキストに関する情報と結合される。言い換えると、分類モジュール 123 は、音声認識機能 120 からの出力 126 を分析して、1 つまたは複数のフレーズが、実際の人間によって使用されるフレーズまたは応答マシンによって使用されるフレーズと統計的に一致しているかどうかを判断する。図 5 では、分類モジュール 123 からの情報を、ステップ 312 および 314 に示す。具体的には、ステップ 312 において、分類モジュール 123 からの情報が、可聴応答内のフレーズが実際の人間からの可能性が高いこと（高い確信度）を示す場合、コール分析モジュール 122 の出力 128 は、コール受信側が実際の人間であることを示すことになる。これに対して、ステップ 312 において、コール受信側が実際の人間であるとの高い確信度がない場合は、コール分析モジュール 122 の出力は、受信側 102 が応答マシンであることを示すことになる。同様に、ステップ 314 において、分類モジュール 123 からの情報が、可聴応答内のフレーズが応

答マシンからの可能性が高いことを示す場合、コール分析モジュール 122 の出力はコール受信側が応答マシンであることを示すことになり、一方、コール受信側が応答マシンであるとの高い確信度がない場合は、コール分析モジュール 122 の出力は、コール受信側 102 が実際の人間であることを示すことになる。この例においては、エラーの偏り (bias) は、コール受信側を応答マシンとしてではなく、実際の人間として誤って分類する側にあることに留意されたい。必要ならば、この偏りはコール受信側を応答マシンとして分類する側にすることも、異なるエラーに関連するコストに応じて全く偏りがないようにすることもできる。

【0020】

上述した統計分類機能は単なる例示であることに留意されたい。コール分析モジュール 122 は、最大エントロピー分類機能 (maximum entropy classifier) などの、他の多数の分類機能を使用することができる。さらに、コール分析モジュール 122 は、同等の精度を達成するためにサポートベクトルマシン、決定木を利用するコンポーネント、および人工ニューラルネットワークを使用することもできる。

【0021】

図 5 の例では、コールが受信側によってピックアップされたあとの沈黙の存在および可聴応答の持続時間などの非ワード特徴は、コール受信側が実際の人間であるか、応答マシンであるかを確かめるために、可聴応答のコンテンツを実際に分類する前に使用された。また、受信側がバーズインしたかどうか、または音声認識機能 120 が可聴応答を認識できたかどうかなどの、他の非ワード特徴を使用することも可能である。さらに、分類機能を、非ワード特徴に基づいてトレーニングし、可聴応答の認識されたフレーズに基づいて出力を提供する分類機能 123 とともに使用して、コール受信側が実際の人間であるか応答マシンであるかについての表示を提供することもできる。同様に、単一の分類機能を、非ワード特徴および可聴応答の認識されたフレーズに基づいてトレーニングして、コール受信側が実際の人間であるか応答マシンであるかについての表示を提供することもできる。

【0022】

一部のアプリケーションでは、例えば、コール受信側が応答マシンであると判断された場合、そのコール受信側にメッセージをプレイすることが必要になることもある。応答マシンは、挨拶が終了したこと、およびメッセージを残すべきであることを意味するトーン (tone) または沈黙を提供するが、このトーンの認識は困難である。多くの場合、トーンまたは沈黙は正確に認識されないことがあり、メッセージの少なくとも一部がこの挨拶の間にプレイされることがあるため、メッセージの始まりの部分が記録されないことがある。

【0023】

本明細書に説明される別の態様も含む、図 4 のオプションのステップ 206 は、メッセージ全体が応答マシンによって記録されたこと、必要に応じて実際の人間によって聞かれたことを確かめる手段または手法を提供する。具体的には、メッセージは、メッセージ全体が完了するまで、この音声認識機能 120 をターンオンし「バーズイン」を可能にして (挨拶の完了を検出するために) リプレイされる。例えば、応答マシンが「You have reached...Eric Moe...Please a message (こちらは...Eric Moe です...メッセージをどうぞ)」(ここで、「...」は挨拶中の沈黙を表す) という挨拶を返したと想定する。多くの現在のシステムでは、メッセージは、これらの沈黙の期間の 1 つが検出されるとプレイされ、したがって応答マシンが挨拶の最後まで到達していないのでメッセージはプレイされるが、記録されないことになる。本明細書で説明した手法を使用すると、バーズイン可能な音声認識機能 120 は、挨拶の各部分を検出し、メッセージをさらに繰り返し始めて開始させる。挨拶が明確にその最後まで到達すると、メッセージはリプレイされるが、今度は、メッセージ全体が聞かれたことまたは記録されたことを確かめるために完了するまでリプレイされる。

【0024】

10

20

30

40

50

図6は、本明細書で説明される概念を実現することができる適切なコンピューティングシステム環境400の例を図示する。具体的には、コンピューティングシステム環境400の一形態を、アウトバウンドコールシステム100またはその一部に使用することができる。やはり、コンピューティングシステム環境400も、これらのコンピュータの各々にとって適切なコンピューティング環境の単なる一例であり、以下の説明の使用または機能の範囲に対するいかなる制限も示唆するようには意図されていない。また、コンピューティング環境400は、例示的なオペレーティング環境400に図示されるコンポーネントのいずれか1つまたはその組み合わせに関してもいかなる依存関係または要件があると解釈されるべきではない。

【0025】

本明細書で提供される例に加えて、他の周知コンピューティングシステム、環境および/または構成も本明細書で説明される概念での使用に適していることがある。そのようなシステムには、パーソナルコンピュータ、サーバコンピュータ、ハンドヘルドまたはラップトップデバイス、マルチプロセッサシステム、マイクロプロセッサベースのシステム、セットトップボックス、プログラム可能な家庭用電化製品、ネットワークPC、ミニコンピュータ、メインフレームコンピュータ、上記システムまたはデバイスのいずれかを含む分散コンピューティング環境などが含まれるが、これらには限定されない。

【0026】

本明細書で説明される概念は、プログラムモジュールなどの、コンピュータ実行可能命令の一般的コンテキストにおいて具現化することができる。一般的に、プログラムモジュールには、特定のタスクを実行し、または特定の抽象データ型を実現するルーチン、プログラム、オブジェクト、コンポーネント、データ構造などが含まれる。当業者は、本明細書の記述および/または図を、以下で述べる任意の形式のコンピュータ読取可能な媒体において具現化することができるコンピュータ実行可能命令として実装することができる。

【0027】

本明細書で説明される概念を、通信ネットワークを通してリンクされるリモート処理デバイスによってタスクが実行されるような分散コンピューティング環境で実施することもできる。分散コンピューティング環境では、プログラムモジュールを、メモリストレージデバイスを含むローカルコンピュータストレージ媒体とリモートコンピュータストレージ媒体の両方に配置することができる。

【0028】

図6を参照すると、例示システムはコンピュータ410の形式の汎用コンピューティングデバイスを含んでいる。コンピュータ410のコンポーネントには、処理ユニット420、システムメモリ430、およびシステムメモリを含む様々なシステムコンポーネントを処理ユニット420に結合するシステムバス421が含まれるが、これらには限定されない。システムバス421は、様々なバスアーキテクチャのいずれかを使用するメモリバスまたはメモリコントローラ、周辺バス、およびローカルバスを含む、いくつかのタイプのバス構造のいずれかとすることができる。限定ではなく例として、そのようなアーキテクチャには、ISA(Industry Standard Architecture)バス、MCA(Micro Channel Architecture)バス、EISA(Enhanced ISA)バス、VESA(video Electronics Standards Association)ローカルバス、およびMezzanineバスとしても知られるPCI(Peripheral Component Interconnect)バスが含まれる。

【0029】

コンピュータ410は、典型的に様々なコンピュータ読取可能な媒体を備えている。コンピュータ読取可能な媒体は、コンピュータ410によってアクセス可能な任意の使用可能な媒体とすることができ、揮発性および不揮発性の媒体、取り外し可能および取り外し不能の媒体が含まれる。限定ではなく例として、コンピュータ読取可能な媒体は、コンピュータ記憶媒体を備えていることがある。コンピュータ記憶媒体には、コンピュータ読取

10

20

30

40

50

可能命令、データ構造、プログラムモジュールまたは他のデータなどの情報を記憶するための任意の方法または技術で実装される、揮発性および不揮発性、取り外し可能および取り外し不能の媒体が含まれる。コンピュータ記憶媒体には、RAM、ROM、EEPROM、フラッシュメモリまたは他のメモリ技術、CD-ROM、デジタル多用途ディスク(DVD)または他の光ディスクストレージ、磁気カセット、磁気テープ、磁気ディスクストレージまたは他の磁気ディスクストレージ、または所望の情報を格納するのに使用可能でコンピュータ400によってアクセス可能な他の任意の媒体が含まれるが、これらには限定されない。

【0030】

システムメモリ430は、ROM431やRAM432などの、揮発性および/または不揮発性メモリの形式のコンピュータ記憶媒体を含む。起動時などにコンピュータ410内のエレメント間の情報の転送を助ける基本ルーチンを含む基本入出力システム(BIOS)433は、典型的にROM431に格納されている。RAM432は、典型的に、処理ユニット420に即時にアクセス可能および/または処置ユニット420によって現在動作中のデータおよび/またはプログラムモジュールを含む。

【0031】

図6は、オペレーティングシステム434、アプリケーションプログラム435、他のプログラムモジュール436、およびプログラムデータ437を図示している。本明細書において、アプリケーションプログラム435、プログラムモジュール436、およびプログラムデータ437は、上述した概念の1つまたは複数を実装する。

【0032】

コンピュータ410は、他の取り外し可能/取り外し不能、揮発性/不揮発性のコンピュータ記憶媒体も含むことができる。単に例として図6は、取り外し不能の不揮発性磁気媒体との間で読み書きを行なうハードディスクドライブ441、取り外し可能の不揮発性磁気ディスク452との間で読み書きを行なう磁気ディスクドライブ451、およびCD-ROMまたは他の光媒体のような取り外し可能の不揮発性光ディスク456との間で読み書きを行なう光ディスクドライブ455を図示している。例示的なオペレーティング環境で使用可能な他の取り外し可能/取り外し不能、揮発性/不揮発性のコンピュータ記憶媒体には、磁気テープカセット、フラッシュメモリカード、DVD、デジタルビデオテープ、ソリッドステートRAM、ソリッドステートROMなどが含まれるが、これらには限定されない。ハードディスクドライブ441は、典型的に、インタフェース440などの取り外し不能メモリインタフェースを通してシステムバス421に接続される、磁気ディスクドライブ451および光ディスクドライブ455は、典型的に、インタフェース450などの取り外し可能メモリインタフェースによってシステムバス421に接続される。

【0033】

上述され図6に図示されるドライブおよびそれらに関連するコンピュータ記憶媒体は、コンピュータ410のコンピュータ読取可能命令、データ構造、プログラムモジュール、および他のデータのストレージを提供する。図6において、例えば、ハードディスクドライブ441は、オペレーティングシステム444、アプリケーションプログラム445、他のプログラムモジュール446、およびプログラムデータ447を格納するものとして図示されている。これらのコンポーネントは、オペレーティングシステム434、アプリケーションプログラム435、他のプログラムモジュール436、およびプログラムデータ437と同じとすることも異なるものとして行うことができることに留意されたい。オペレーティングシステム444、アプリケーションプログラム445、他のプログラムモジュール446、およびプログラムデータ447に異なる番号が与えられているのは、少なくともこれらが異なるコピーであることを示すためである。

【0034】

ユーザは、キーボード462、マイクロホン463、ならびに、マウスまたはトラックボール、タッチパッドのようなポインティングデバイス461などの入力デバイスを通して、コマンドおよび情報を入力することができる。これらおよび他の入力デバイスは、シ

システムバスに結合されるユーザ入力インタフェース 460 を通して処理ユニット 420 に接続されることが多いが、パラレルポートや USB などの他のインタフェースおよびバス構造によって接続されることもある。モニタ 491 または他のタイプのディスプレイデバイスも、ビデオインタフェース 490 などのインタフェースを介してシステムバス 421 に接続される。

【0035】

コンピュータ 410 は、リモートコンピュータ 480 などの 1 つまたは複数のリモートコンピュータに対する論理接続を使用するネットワーク化された環境で動作することがある。リモートコンピュータ 480 は、パーソナルコンピュータ、ハンドヘルドデバイス、サーバ、ルータ、ネットワーク PC、ピアデバイス、または他の共通ネットワークノードとすることができ、典型的にコンピュータ 410 に関連して上述したエレメントの多くまたはすべてを含む。図 6 に示される論理接続には、LAN 471 と WAN 473 が含まれるが、他のネットワークを含むこともできる。このようなネットワーキング環境は、オフィス、企業規模のコンピュータネットワーク、イントラネットおよびインターネットにおいて普及している。

10

【0036】

LAN ネットワーキング環境で使用されるとき、コンピュータ 410 は、ネットワークインタフェースまたはアダプタ 470 を通して LAN 471 に接続される。WAN ネットワーキング環境で使用されるとき、コンピュータ 410 は、典型的に、インターネットなどの WAN 473 上の通信を確立するためのモデム 472 や他の手段を備えている。内蔵型または外付け型とすることができるモデム 472 は、ユーザインプットインタフェース 460 または他の適切なメカニズムを通してシステムバス 421 に接続される。ネットワーク化された環境では、コンピュータ 410 に関連して示されるプログラムモジュールまたはその一部を、リモートメモリストレージデバイスに格納することができる。限定ではなく例として、図 6 は、リモートアプリケーションプログラム 485 がリモートコンピュータ 480 の上にあるものとして図示している。示されたネットワーク接続は例示であり、コンピュータ間の通信リンクを確立する他の手段が使用されることもあることは認識されよう。

20

【0037】

本明細書で説明した概念は、図 6 に関連して説明したコンピュータシステムのようなコンピュータシステムにおいて実施できることに留意されたい。しかし、他の適切なシステムは、メッセージの処理専用のサーバまたはコンピュータ、あるいは本概念の異なる部分が分散コンピューティングシステムの異なる部分において実施されるような分散システムにおけるサーバやコンピュータを含む。

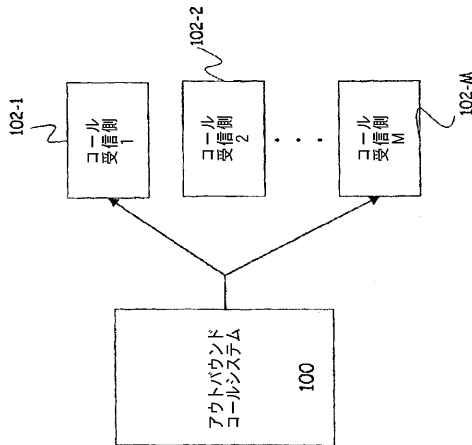
30

【0038】

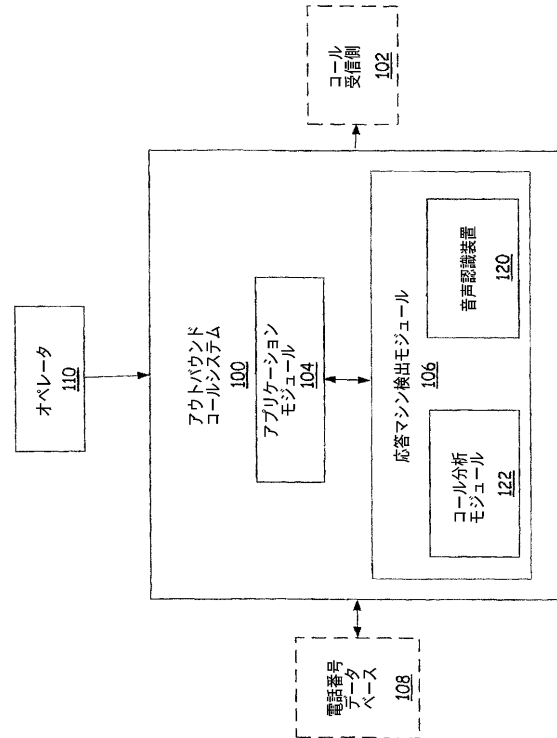
本主題を構造的な特徴および / または方法論的なアクトに特有の表現で説明してきたが、添付の特許請求の範囲に定義される主題は、裁判所によって支持されているように上述した特有の特徴またはアクトに限定されるものではないことは理解されよう。むしろ、上述した特有の特徴およびアクトは、請求項を実現する形態の例として開示されたものである。

40

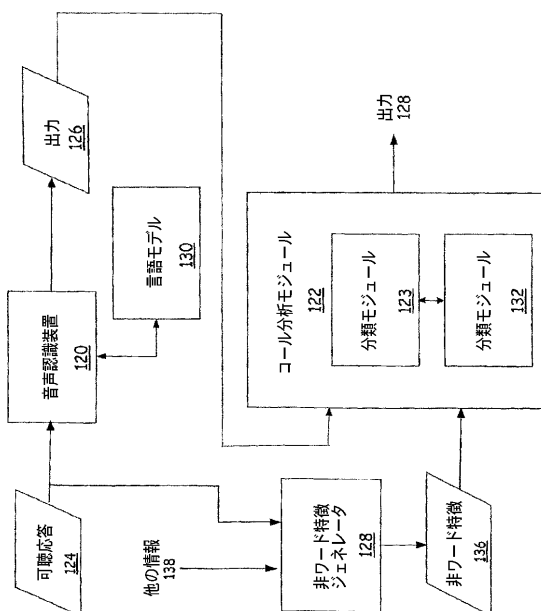
【図 1】



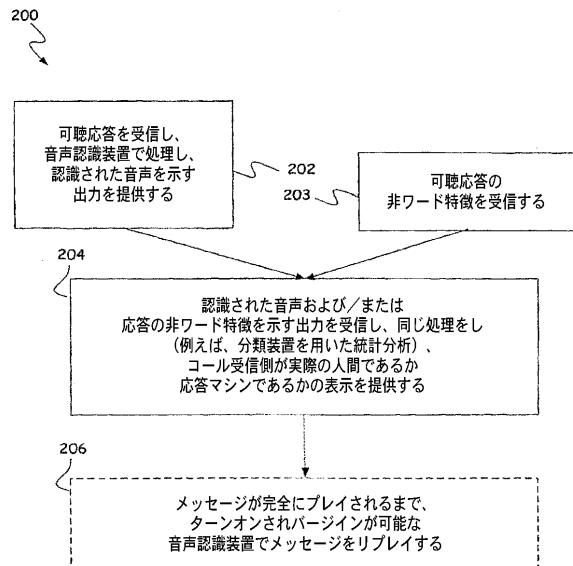
【図 2】



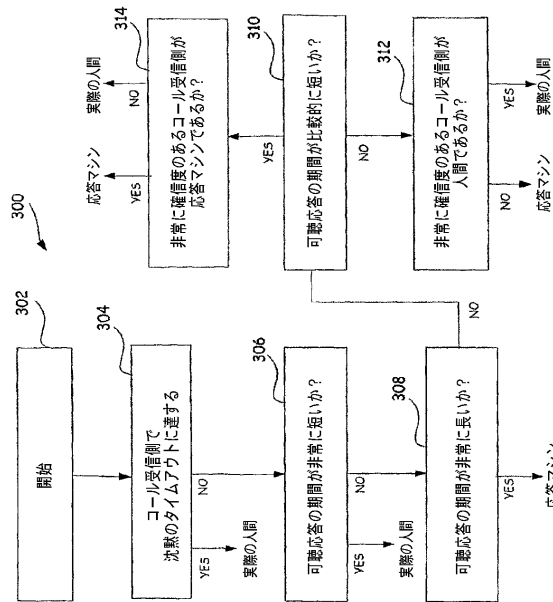
【図 3】



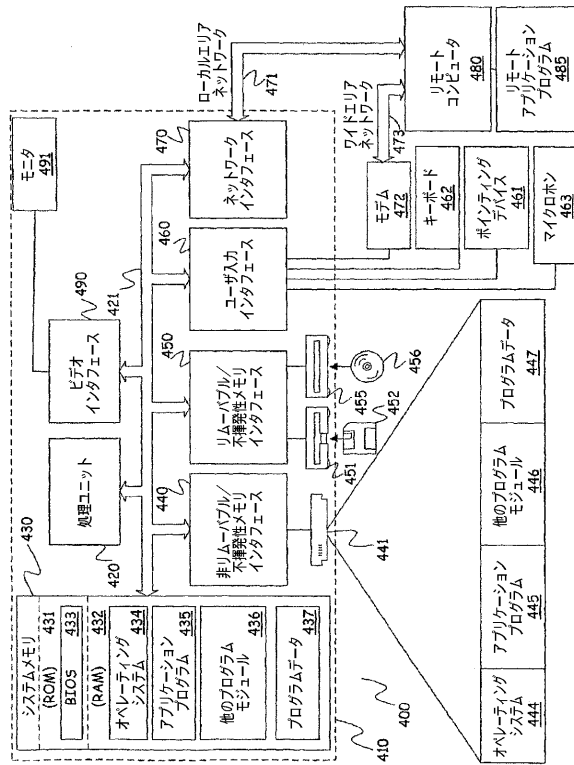
【図 4】



【 図 5 】



【 図 6 】



フロントページの続き

- (72)発明者 クレイグ エム・フィッシャー
アメリカ合衆国 98052 ワシントン州 レッドモンド ワン マイクロソフト ウェイ マ
イクロソフト コーポレーション インターナショナル パテンツ内
- (72)発明者 ドン ユー
アメリカ合衆国 98052 ワシントン州 レッドモンド ワン マイクロソフト ウェイ マ
イクロソフト コーポレーション インターナショナル パテンツ内
- (72)発明者 イエ・イ ワン
アメリカ合衆国 98052 ワシントン州 レッドモンド ワン マイクロソフト ウェイ マ
イクロソフト コーポレーション インターナショナル パテンツ内
- (72)発明者 ユ・チェン ジュイ
アメリカ合衆国 98052 ワシントン州 レッドモンド ワン マイクロソフト ウェイ マ
イクロソフト コーポレーション インターナショナル パテンツ内

審査官 田部井 和彦

- (56)参考文献 米国特許第05371787(US, A)
米国特許出願公開第2003/0086541(US, A1)
米国特許第04941168(US, A)
米国特許出願公開第2005/0276390(US, A1)
米国特許第06324262(US, B1)
米国特許第06233319(US, B1)
米国特許第05644624(US, A)
Dong Yu ; Yun Cheng Ju ; Ye-Yi Wang ; Acero, A. ; , N-Gram Based Filler Model for Robu
st Grammar Authoring, Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proce
edings. 2006 IEEE International Conference on, 2006年 5月

- (58)調査した分野(Int.Cl., D B名)
G10L 11/00-21/06
IEEE Xplore