(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2017-37658
(P2017-37658A)
(43)公開日　平成29年2月16日(2017.2.16)

(51)Int.Cl.		F I			テーマコード（参考）
G06F 3/01	(2006.01)	G06F	3/01	510	5E555
G06Q 30/02	(2012.01)	G06Q	30/02	398	5L049
G10L 15/00	(2013.01)	G06Q	30/02	470	
G10L 15/24	(2013.01)	G10L	15/00	200J	
G10L 15/22	(2006.01)	G10L	15/00	200T	

審査請求 有　請求項の数 18　OL　外国語出願　（全 24 頁）　最終頁に続く

(21)出願番号　特願2016-182671 (P2016-182671)
(22)出願日　平成28年9月20日 (2016.9.20)
(62)分割の表示　特願2015-7506 (P2015-7506)の分割
原出願日　平成22年2月22日 (2010.2.22)
(31)優先権主張番号　12/389,678
(32)優先日　平成21年2月20日 (2009.2.20)
(33)優先権主張国　米国 (US)

(71)出願人　511204429
ボイスボックス　テクノロジーズ　コーポレーション
アメリカ合衆国　ワシントン州　９８００５，ベルビュー，スイート１００，エヌイー　２４番ストリート　１１９８０
１１９８０ ＮＥ ２４ｔｈ Ｓｔｒｅｅｔ，Ｓｕｉｔｅ １００，Ｂｅｌｌｅｖｕｅ，Ｗａｓｈｉｎｇｔｏｎ ９８００５，Ｕｎｉｔｅｄ Ｓｔａｔｅｓ ｏｆ Ａｍｅｒｉｃａ
(74)代理人　100126572
弁理士　村越　智史

最終頁に続く

(54)【発明の名称】自然言語音声サービス環境においてマルチモーダル機器対話を処理するシステム及び方法

(57)【要約】　（修正有）
【課題】自然言語音声サービス環境においてマルチモーダル機器対話を処理するシステム及び方法を提供する。
【解決手段】単数又は複数のマルチモーダル機器対話は、単数又は複数の電子機器を含む自然言語音声サービス環境において受信される。このマルチモーダル機器対話は、前記電子機器の少なくとも一つ又はそれに対応付けられたアプリケーションとの非音声対話を含み、前記非音声対話に対応付けられた自然言語発話をさらに含む。前記非音声対話及び前記自然言語発話に関連するコンテクストが抽出され、前記マルチモーダル機器対話の意図を決定するために組み合わされる。また、前記マルチモーダル機器対話の前記決定された意図に基づいて、単数又は複数の前記電子機器に対して要求が発出される。
【選択図】図３

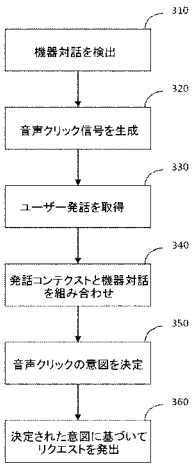

図3

【特許請求の範囲】
【請求項１】
単数又は複数の電子機器を含む自然言語音声サービス環境において単数又は複数のマルチモーダル機器対話を処理する方法であって、
前記電子機器の少なくとも一つ又は前記電子機器の少なくとも一つに対応付けられたアプリケーションとの非音声対話及び前記非音声対話に関連する少なくとも一つの自然言語発話を含む少なくとも一つのマルチモーダル機器対話を検出する工程と、
前記マルチモーダル機器対話に関連するコンテクスト情報であって前記非音声対話に関連するコンテクスト及び前記自然言語発話に関連するコンテクストを含むものを抽出する工程と、
前記非音声対話に関連するコンテクストと前記自然言語発話に関連するコンテクストとを組み合わせる工程と、
前記非音声対話及び前記自然言語発話に関連する前記組み合わされたコンテクストに基づいて、前記マルチモーダル機器対話の意図を決定する工程と、
前記マルチモーダル機器対話の前記決定された意図に基づいて、少なくとも一つのリクエストを単数又は複数の前記電子機器に発出する工程と、
を含む方法。
【請求項２】
前記電子機器の少なくとも一つが前記自然言語発話を受信するように構成された入力機器を含む請求項１の方法。
【請求項３】
前記非音声対話の検出に応答して、前記入力機器に前記自然言語発話を取得するように伝達する工程をさらに含む請求項２の方法。
【請求項４】
前記自然言語音声サービス環境において前記非音声対話を検出するように構成された単数又は複数の機器リスナーを設定する工程と、
前記機器リスナーによって検出された前記非音声対話に関連する情報及び前記入力機器によって取得された前記自然言語発話を連携させる工程と、
をさらに含む請求項３の方法
【請求項５】
前記マルチモーダル機器対話の前記決定された意図に基づいて、少なくとも一つのトランザクションリードを生成する工程と、
前記生成されたトランザクションリードに関連する少なくとも一つの追加的なマルチモーダル機器対話を受信する工程と、
前記生成されたトランザクションリードに関連する前記マルチモーダル機器対話の受信に応答して、トランザクションクリックスルーを処理する工程と、
をさらに含む請求項１の方法。
【請求項６】
前記生成されたトランザクションリードが前記マルチモーダル機器対話の前記決定された意図に関連する広告又は推薦の少なくとも一つを含む請求項５の方法。
【請求項７】
前記非音声対話が単数又は複数の前記電子機器に対応付けられた領域、項目、データ、又はアプリケーションの選択を含む請求項１の方法。
【請求項８】
前記非音声対話が単数又は複数の前記電子機器に対応付けられた焦点又は関心の焦点の特定を含む請求項１の方法。
【請求項９】
前記非音声対話が単数又は複数の前記電子機器に対応付けられた単数又は複数の固有で識別可能な対話を含む請求項１の方法。
【請求項１０】

単数又は複数の電子機器を含む自然言語音声サービス環境において単数又は複数のマルチモーダル機器対話を処理するシステムであって、当該システムは単数又は複数の処理装置を含み、当該単数又は複数の処理装置は、
前記電子機器の少なくとも一つ又は前記電子機器の少なくとも一つに対応付けられたアプリケーションとの非音声対話及び前記非音声対話に関連する少なくとも一つの自然言語発話を含む少なくとも一つのマルチモーダル機器対話を検出し、
前記マルチモーダル機器対話に関連するコンテクスト情報であって、前記非音声対話に関連するコンテクスト及び前記自然言語発話に関連するコンテクストを含むものを抽出し、
前記非音声対話に関連するコンテクストと前記自然言語発話に関連するコンテクストとを組み合わせ、
前記非音声対話及び前記自然言語発話に関連する前記組み合わされたコンテクストに基づいて、前記マルチモーダル機器対話の意図を決定し、
前記マルチモーダル機器対話の前記決定された意図に基づいて、少なくとも一つのリクエストを単数又は複数の前記電子機器に発出する、
ように構成されたシステム。

【請求項１１】
前記電子機器の少なくとも一つが前記自然言語発話を受信するように構成された入力機器を含む請求項１０のシステム。

【請求項１２】
前記非音声対話の検出に応答して、前記入力機器に前記自然言語発話を取得するように伝達する工程をさらに含む請求項１１のシステム。

【請求項１３】
前記処理装置が、さらに、
前記自然言語音声サービス環境において前記非音声対話を検出するように構成された単数又は複数の機器リスナーを設定し、
前記機器リスナーによって検出された前記非音声対話に関連する情報及び前記入力機器によって取得された前記自然言語発話を連携させるように構成された請求項１２のシステム

【請求項１４】
前記制御装置が、さらに、
前記マルチモーダル機器対話の前記決定された意図に基づいて、少なくとも一つのトランザクションリードを生成し、
前記生成されたトランザクションリードに関連する少なくとも一つの追加的なマルチモーダル機器対話を受信し、
前記生成されたトランザクションリードに関連する前記マルチモーダル機器対話の受信に応答して、トランザクションクリックスルーを処理するように構成された請求項１０のシステム。

【請求項１５】
前記生成されたトランザクションリードが前記マルチモーダル機器対話の前記決定された意図に関連する広告又は推薦の少なくとも一つを含む請求項１４のシステム。

【請求項１６】
前記非音声対話が単数又は複数の前記電子機器に対応付けられた領域、項目、データ、又はアプリケーションの選択を含む請求項１０のシステム。

【請求項１７】
前記非音声対話が単数又は複数の前記電子機器に対応付けられた焦点又は関心の焦点の特定を含む請求項１０のシステム。

【請求項１８】
前記非音声対話が単数又は複数の前記電子機器に対応付けられた単数又は複数の固有で識別可能な対話を含む請求項１０のシステム。

【発明の詳細な説明】
【技術分野】
【０００１】
本出願は、2009年2月20日出願の「SYSTEM AND METHOD FOR PROCESSING MULTI-MODAL DEVICEINTERACTIONS IN A NATURAL LANGUAGE VOICE SERVICES ENVIRONMENT」と題する米国特許出願第12/389,678号に基づく優先権を主張する。当該基礎出願の内容は参照により全体として本明細書に組み込まれる。
【０００２】
本発明は、単数又は複数の装置及び／又はアプリケーションを用いてマルチモーダル対話（multi-modalinteraction）を処理する統合自然言語音声サービス環境に関する。マルチモーダル対話は、追加的なコンテクスト（context）を提供する。この追加的なコンテクストは、マルチモーダル対話に付随する自然言語発話の解釈に協力するとともにその他の処理を行うためのものである。
【背景技術】
【０００３】
家電機器は、技術の進展にともなって、多くの人の毎日の生活のあらゆる場面に入り込むようになっている。携帯電話、ナビゲーション装置、組み込み機器等の機能や可搬性が向上した結果、これらの機器に対する要望が高まっており、かかる要望を満たすために、多くの装置においてコアアプリケーションの他にも多様な機能及び機構が提供されている。しかしながら、機能性の向上によって、学習曲線などにマイナスの影響が生じ、多くのユーザが電子機器の機能の一部しか利用できなくなっている。例えば、既存の電子機器の多くは、ユーザフレンドリーでない複雑なヒューマン・マシン・インタフェースを備えており、そのせいで多くの技術がマスマーケットで受け入れられていない。また、インタフェースさえ良ければ有用なはずの機能でも、インタフェースが操作しづらいため、その機能の発見や使用が困難になることも多い（例えば、複雑だったり操作が面倒だったりすることが原因である）。このように、多くのユーザは、自分の電子機器の潜在的な機能を使用しないことが多く、ときにはその存在すら知らないこともある。
【０００４】
市場調査によって、ユーザの多くは、機器において利用できる機能やアプリケーションのほんの一部分しか利用していないことが指摘されている。このように、電子機器の機能は、増加しているものの無駄になっていることが多い。また、無線ネットワーク通信やブロードバンドアクセスがますます普及しており、自分の電子機器からのシームレスな無線通信は消費者にとって当然の要望である。このように、電子機器を操作するためのより簡単な機構に対する要望が高まるにつれて、速く集中的な対話を妨げる操作しづらいインタフェースが重大な懸念事項となっている。しかしながら、技術を直感的に使いたいという要望は、今のところ概して満たされないままである。
【０００５】
電子機器においてユーザと機器との間の対話を単純にする一つの方法は、音声認識ソフトウェアを使用することである。ユーザは、音声認識ソフトウェアを用いることによって、今まで不慣れ、不明、又は使用困難であった機能を利用できる可能性がある。例えば、カーナビゲーションやウェブアプリケーション等の様々なアプリケーションにおいて用いられるデータを提供するNavteq Corporationが行った最近の調査では、音声認識が電子機器ユーザに最も望まれている機能の一つとして頻繁にランキングされていることが示されている。そうであったとしても、既存の音声ユーザインタフェースは、実際の動作時には依然としてユーザ側に多くの学習を要求するものである。
【０００６】
例えば、既存の音声ユーザインタフェースの多くは、特定のコマンド・コントロール・シーケンス又はコマンド・コントロール・シンタックスに従って定められたリクエストしかサポートしていない。また、既存の音声ユーザインタフェースの多くは、音声認識が不正確であるため、ユーザにフラストレーションや不満足感を引き起こすこともある。同様

に、リクエストをシステムが理解できる方法で伝達するためには、ユーザは予め定められたコマンド又はキーワードを提供しなければならないので、既存の音声ユーザインタフェースは、ユーザを生産的で協力的な対話に効果的に関与させてリクエストを決定することができず、会話を満足の行く目標に向かって進めることができない（例えば、ユーザが特定のニーズ、利用可能な情報、機器の機能等についてよく分かっていない場合）。このように、既存の音声ユーザインタフェースには、様々な欠点が存在する。例えば、ユーザを協力的な方法で及び会話形式で対話に関与させる点に重大な限界がある。

【０００７】

また、既存の音声ユーザインタフェースの多くは、異なるドメイン、機器、及びアプリケーションを横断して提供される情報を利用して自然言語音声の入力を決定することができない。このように、既存の音声ユーザインタフェースは、有限の数の専用アプリケーション又は搭載された機器に限定されるという欠点がある。技術の進展によって、ユーザは必要に応じて複数の機器を所有することが多くなっているが、ユーザは既存の音声ユーザインタフェースのせいで機器に縛り付けられたままである。例えば、ユーザは、異なるアプリケーション及び機器に関連づけられたサービスに関心を持っている場合であっても、既存の音声ユーザインタフェースによってそれらのサービスに対応するアプリケーション及び機器へのアクセスを制限されることが多い。また、ユーザは、現実には一度に有限の数の機器しか持ち歩くことができないが、ユーザの使用中の他の機器に関連づけられているコンテンツやサービスを様々な場面で使用することが望まれる。

【０００８】

したがって、ユーザは様々なニーズを持っており、異なる機器に関連づけられたコンテンツ又はサービスが様々なコンテクスト又は環境において必要とされる可能性があるが、既存の音声技術は、実質的に任意の機器又はネットワークに対応付けられるコンテンツ又はサービスがユーザによって要求可能な統合環境を提供することができないことが多い。このように、既存の音声サービス環境における情報の利用可能性及び機器対話メカニズムの制約のために、ユーザは、直感的で、自然で、及び効率的な方法で技術を利用することができない。例えば、ユーザが所定の電子機器を用いて所定の機能を実行することを望んでいるが当該機能を実行するためにどうすればよいのか必ずしも分かっていない場合において、当該機器とのマルチモーダル対話を行って単純に自然言語で語を発話しても当該機能を要求することはできない。

【０００９】

また、相対的に単純な機能は、音声認識機能を有していない電子機器を用いて実行するには退屈であることが多い。例えば、携帯電話用の新しいリングトーンを購入することは、相対的に単純な処理である場合が多いが、ユーザは、処理の完了までに複数のメニューをナビゲートし、多くの異なるボタンを押さなければならないことが多い。このように、ユーザが、埋没した機能又は使用が難しい機能を利用するために自然言語を使用する場合には、電子機器との対話がはるかに効率的であることは明らかである。既存のシステムには、上述の及びそれ以外の問題点がある。

【発明の概要】

【００１０】

本発明の一態様によれば、自然言語音声サービス環境においてマルチモーダル機器対話を処理するシステム及び方法が提供される。具体的には、単数又は複数のマルチモーダル機器対話は、単数又は複数の電子機器を含む自然言語音声サービス環境において受信される。このマルチモーダル機器対話は、ユーザが単数又は複数の前記電子機器又は当該機器に対応付けられたアプリケーションと非音声対話する一方で、前記非音声対話に関連して自然言語発話についても提供することを含む。例えば、前記非音声機器対話は、ユーザが、特定の領域、項目、データ、焦点、又は関心の焦点を選択すること、又はこれら以外の方法で前記電子機器又は前記電子機器に対応付けられたアプリケーションとの単数又は複数の固有で識別可能な対話を行うことを含む。このため、コンテクストが前記自然言語発話から抽出され、また、前記非音声機器対話が、前記自然言語発話に関する追加的なコン

テクストを提供することができる。前記発話のコンテクストと前記非音声機器対話のコンテクストとは、その後に前記マルチモーダル機器対話の意図を決定するために組み合わされる。単数又は複数の前記電子機器は、前記マルチモーダル機器対話の意図に基づいてリクエストを処理する。

【００１１】

本発明の一態様によれば、前記電子機器の少なくとも一つは、音声入力を受信するように構成された入力機器を含む。一実施態様においては、単数又は複数の電子機器又はアプリケーションとの非音声対話の検出に応答して、音声入力機器に対して自然言語発話を取得するように指示がなされる。また、前記自然言語音声サービス環境は、前記電子機器及び対応するアプリケーションに関して設定された単数又は複数のリスナーを含む。このリスナーは、前記電子機器又はアプリケーションとの非音声対話を検出するように構成される。このように、非音声対話と付随する自然言語発話に関連する情報とが連携され、前記発話及び前記非音声機器対話を協力的に処理できるようにする。

【００１２】

本発明の一態様によれば、少なくとも一つのトランザクションリードが、前記マルチモーダル機器対話の意図に基づいて生成される。例えば、最初のマルチモーダル機器対話に関して生成されたトランザクションリードに関連する追加的なマルチモーダル機器対話が受信され、この追加的なマルチモーダル機器対話に関して決定された意図に基づいて、少なくとも一つのリクエストが単数又は複数の前記電子機器に発出される。これにより、前記生成されたトランザクションリードに関連する機器対話の受信に応答してトランザクションクリックスルーを処理することができる。例えば、トランザクションリードは、元のマルチモーダル機器対話の意図に基づいて選択される広告又は推薦を含む。一方、追加的なマルチモーダル機器対話は、ユーザが広告又は推薦を選択することを含む。このように、広告又は推薦の選択は、トランザクションクリックスルーとみなされ、特定の主体（例えば、前記自然言語音声サービス環境のプロバイダー）の収益を生み出す。

【００１３】

これら以外の本発明の目的及び効果は、以下の図面及び詳細な説明により明らかになる。

【図面の簡単な説明】

【００１４】

【図１】本発明の様々な態様に従って、自然言語音声サービス環境においてマルチモーダル機器対話を処理する例示的なシステムのブロック図を示す。

【００１５】

【図２】本発明の様々な態様に従って、自然言語音声サービス環境においてマルチモーダル機器を同期させる例示的な方法のブロック図を示す。

【００１６】

【図３】本発明の様々な態様に従って、自然言語音声サービス環境においてマルチモーダル機器対話を処理する例示的な方法のフロー図を示す。

【００１７】

【図４】本発明の様々な態様に従って、自然言語音声サービス環境においてマルチモーダル機器対話を処理して単数又は複数のトランザクションリードを生成する例示的な方法のフロー図を示す。

【発明を実施するための形態】

【００１８】

図１は、本発明の様々な態様に従って、自然言語音声サービス環境においてマルチモーダル機器対話を処理する例示的なシステム１００のブロック図を示す。本明細書の説明から明らかになるように、図１に示されたシステム１００は、入力機器１０５又は入力機器１０５の組み合わせ含んでもよい。ユーザは、この入力機器１０５により、マルチモーダルな方法でシステム１００と対話することができる。特に、システム１００は、少なくとも音声クリックモジュール１０８を含む様々な自然言語処理コンポーネントを含むことが

できる。音声クリックモジュール１０８は、単数又は複数の入力機器１０５とともにユーザのマルチモーダル対話を処理することができる。例えば、一実施態様における入力機器１０５は、少なくとも一つの音声入力機器１０５ａ（例えば、マイク）と少なくとも一つの非音声入力機器１０５ｂ（例えば、マウス、タッチスクリーンディスプレイ、ホイールセレクタ等）との任意の適切な組み合わせを含むことができる。このように、入力機器１０５は、音声ベースの入力及び非音声ベースの入力（例えば、テレマティックス機器、パーソナルナビゲーション機器、携帯電話、VoIPノード、パーソナルコンピュータ、メディア機器、組み込み機器、サーバ、又はこれら以外の電子単数又は複数の機器に接続されたマイク）の両方を受信するメカニズムを有する電子機器の任意の適切な組み合わせを含むことができる。このため、ユーザは、電子機器１０５に関連づけられた単数又は複数の電子入力機器１０５又はアプリケーションを用いて、システム１００によってマルチモーダルな会話形式の対話を行うことができる。このシステム１００は、タスクを発出するため又は他の方法でリクエストを決定（resolve）するために好適な自由形式且つ協力的な方法で機器対話を処理することができる。

【００１９】

上述のように、一実施態様におけるシステムは、自由形式の発話及び／又は他の形式の機器対話をサポート可能な様々な自然言語処理コンポーネントを含んでもよく、これにより、コマンド、クエリ、又はこれら以外のリクエストの決定方法に関する制約からユーザを開放することができる。このため、ユーザは、システム１００において使用可能なコンテンツ又はサービスを要求するために、音声入力機器１０５ａに向かって話す任意の方法又はこれ以外の非音声入力機器１０５ｂと対話する任意の方法を用いて入力機器１０５と対話することができる。例えば、ユーザは、自然言語発話を音声入力機器１０５ａに提供することによって、システム１００において利用可能な任意のコンテンツ又はサービスを要求することができる。一実施態様において、発話は、2008年7月8日に発行された「System andMethod for Responding to Natural Language Speech Utterance」と題する米国特許第7,398,209号、及び、2003年6月15日に出願された「Mobile System and Method for Responding to Natural Language Speech Utterance」と題する米国特許出願第10/618,633号に記載された技術を用いて処理することができる。これらの特許及び特許出願の開示内容は、参照により全体として本明細書に組み込まれる。また、ユーザは、発話及び／又は要求されたコンテンツもしくはサービスに関連する追加的なコンテクスト等の情報を提供するために、単数又は複数の非音声入力機器１０５ｂと対話することができる。

【００２０】

一実施態様において、システム１００は、追加的なマルチモーダル機器を含む様々な他のシステムに接続されてもよい。当該他のシステムは、図１に示したものと同様の自然言語処理機能を有する。このため、システム１００は、マルチデバイス環境とのインタフェースを提供することができる。このマルチデバイス環境において、ユーザは、当該環境において様々な追加機器を通じて利用できるコンテンツ又はサービスを要求することができる。例えば、一実施態様に係るシステム１００は、当該環境における当該他のシステム及び機器を通して利用可能なコンテンツ、サービス、アプリケーション、意思決定機能、及びこれら以外の機能に関連する情報を提供する星座モデル１３０ｂを含んでもよい。例えば、一実施態様に係るシステム１００は、協力的にリクエストを決定するために、機器、アプリケーション、又は当該環境における他のシステムと対話することができる。この点は、2008年5月27日に出願された「Systemand Method for an Integrated, Multi- Modal, Multi-Device Natural Language VoiceServices Environment」と題する係属中の米国特許出願第12/127,343号において説明されている。当該米国特許出願の開示内容は、参照により全体として本明細書に組み込まれる。例えば、マルチ機器環境は、様々なシステム及び機器の間で情報を共有し、リクエストを決定するための協力的な環境を提供することができる。当該共有された情報は、機器の機能、コンテクスト、以前の対話、ドメイン情報、短期的情報（short-term knowledge）、長期的情報（long-termknowledge）、及び認知モデル等の側面に関連していてもよい。

【００２１】
上述のように、図１に示されたシステム１００は、例えば、単数又は複数のマルチモーダル機器対話をユーザから受信するインタフェース（又はインタフェースの組み合わせ）を共同で提供する単数又は複数の電子入力機器１０５を含んでもよい。この機器対話には、少なくともユーザが発した発話が含まれる。図１に示した実装態様においては、音声入力機器１０５ａと非音声入力機器１０５ｂとが別個に構成されているが、一又は複数の実装態様においては、音声入力機器１０５ａ及び非音声入力機器１０５ｂは同一の機器であってもよく別の機器であってもよい。例えば、入力機器１０５は、携帯電話に接続されたマイク（すなわち、音声入力機器１０５ａ）を含むことができ、さらに携帯電話に接続された単数又は複数のボタン、入力可能ディスプレイ、ホイールセレクタ、又はこれら以外の構成要素（すなわち、非音声入力機器１０５ｂ）を含むことができる。他の例における入力機器１０５は、テレマティックス機器に接続されたマイクの組み合わせ（すなわち、音声入力機器１０５ａ）を含むことができ、さらにテレマティックス機器と通信可能に接続されているが別体のメディアプレイヤーに接続されたボタン、タッチスクリーンディスプレイ、トラックホイール、又はこれら以外の非音声入力機器１０５ｂを含むことができる。このように、入力機器１０５は、通信可能に接続された電子機器の任意の好適な組み合わせを含むことができる。この電子機器は、自然言語発話入力を受信する少なくとも一つの入力機器、及び、マルチモーダル非音声入力を受信する少なくとも一つの入力機器を含む。

【００２２】
一実施態様において、単数又は複数の入力機器１０５に通信可能に接続された音声クリックモジュール１０８によって、音声入力機器１０５ａ及び単数又は複数の非音声入力機器１０５ｂで受信されたマルチモーダル機器対話を協力的に処理（cooperativeprocessing）することができる。例えば、音声クリックモジュール１０８は、音声入力機器１０５ａで受信された自然言語発話を処理するために用いられる情報を、非音声入力機器１０５ｂで受信された単数又は複数の非音声機器対話を考慮して、システム１００に提供することができる。このように、ユーザは、音声クリックモジュール１０８を用いることによって、直感的で自由形式の方法で様々な入力機器１０５と対話することができる。これにより、ユーザは、動作の開始を求め、情報を入手し、又はその他の方法でシステム１００で利用できるコンテンツ又はサービスを要求するときに、様々な種類の情報をシステム１００に提供することができる。

【００２３】
音声入力機器１０５ａは、自然言語発話等の発話された形式の入力を受信することができる任意の適切な機器又は適切な機器の任意の組み合わせを含むことができる。例えば、一実施態様における音声入力機器１０５ａは、指向性マイク、マイクアレイ、又はその他の符号化された音声を生成可能な機器を含むことができる。一実施態様における音声入力機器１０５ａは、符号化された音声の忠実度を最大化するように構成されてもよい。例えば、音声入力機器１０５ａは、ユーザの方向からの利得を最大化し、エコー又はヌルポイント雑音源を除去し、可変レートサンプリングを実行し、環境雑音又は背景会話をフィルターし、又はその他の方法で符号化された音声の忠実度を最大化する技術を用いるように構成されてもよい。このように、音声入力機器１０５ａは、自然言語発話を正確に解釈する際に、ノイズ等のシステム１００に干渉する要素に耐性を有するように符号化された音声を生成する。

【００２４】
非音声入力機器１０５ｂは、非音声機器対話をサポートすることができる任意の適切な機器又は適切な機器の任意の組み合わせを含むことができる。例えば、一実施態様における非音声入力機器１０５ｂは、スタイラスとタッチスクリーン又はタブレットインタフェースとの組み合わせ、BlackBerry（登録商標）、ホイールセレクタ、iPod（登録商標）、クリックホイール、マウス、キーパッド、ボタン、又はその他の任意の機器であって識別可能な非音声機器対話をサポートするものを含むことができる。このため、ユーザは、非

音声入力機器１０５ｂを用いてデータの選択を行い、又は、音声入力機器１０５ａによって提供された関連する自然言語発話に関連して処理されるべき焦点（pointsof focus）（又は関心の焦点（attention focus））を特定することができる。例えば、ユーザは、タッチスクリーンディスプレイの特定の領域をスタイラスで指し示し、マウスを用いてテキストをハイライトし、ボタンをクリックし、アプリケーションと対話し、データを選択し、又はその他の方法で焦点を特定するために任意の好適な機器対話を行う（すなわち、選択されたデータ及び／又は特定された焦点を音声起動又は「音声クリック」する）ことができる。

【００２５】

また、データ選択を行い、焦点を特定し、又はこれら以外の方法で単数又は複数の発話と関連づけて解釈されるデータを起動するために使用可能であることに加えて、ユーザはさらに、非音声入力機器１０５ｂを用いることにより、システム１００内において意味を有する専用の機器対話を行うことができる。例えば、専用の機器対話（「クリック」又は「音声クリック」と称することがある）は、所定期間続くクリック、所定期間連続して維持されるクリック、予め定められたシーケンスに従ってなされるクリック、又はその他の対話又は対話シーケンスであって入力機器１０５及び／又は音声クリックモジュール１０８が特定、検出、又はその他の方法で識別できるものを含むことができる。

【００２６】

一実施態様において、専用の機器対話は、単数又は複数の動作、クエリ、コマンド、タスク、又はその他のリクエストであってシステム１００において利用可能なアプリケーションもしくはサービスに対応付けられたものに関連づけられる。一実施態様において、専用の機器対話は、上述した2008年5月27日出願の「System andMethod for an Integrated, Multi-Modal, Multi-Device Natural Language VoiceServices Environment」と題する係属中の米国特許出願第12/127,343号において説明されているように、単数又は複数の動作、クエリ、コマンド、タスク、又はこれら以外のリクエストであってマルチデバイス環境に配置された様々な機器の任意のものと関連づけられているものをさらに含むことができる。例えば、タッチスクリーンディスプレイに表示された特定の領域又は要素をスタイラスでクリックする識別可能なシーケンスが、携帯電話で電話を開始し、ナビゲーション機器でルートを計算し、メディアプレイヤー用に音楽を購入し、又はこれら以外の種類のリクエストを行うための専用の機器対話（deviceinteraction）又は音声クリック（voice-click）として定義される。

【００２７】

このように、入力機器１０５に接続された音声クリックモジュール１０８は、少なくとも一つの非音声機器対話の発生を検出するために、ユーザと非音声入力機器１０５ｂとの対話を継続的に監視することができる。この非音声機器対話を本明細書において「音声クリック」と称することがある。このため、検出された音声クリックは、マルチモーダル機器対話を処理するための追加的なコンテクストを提供することができる。このマルチモーダル機器対話には、少なくとも一つの音声クリック及び単数又は複数の自然言語発話を含むことができ、その各々がタスク記述（taskspecification）のためのコンテクストを提供することができる。このように、音声クリックは、一般に、現在の発話又は他の音声入力が単数又は複数の機器１０５との現時点での対話とともに処理されるべきことをシステム１００に対して知らせることができる。例えば、一実施態様において、現時点の機器対話は、ユーザによる選択、ハイライト、又は他の関心の焦点の特定、目的、又は単数又は複数の機器１０５に関連づけられた他の項目を含むことができる。このように、現時点の機器対話は、付随する発話の認識、解釈、及び理解を明確にするコンテクストを提供することができ、また、現時点の発話は、付随する機器対話によって提供されるコンテクストを改善する情報を提供することができる。

【００２８】

一実施態様における音声クリックモジュール１０８は、非音声入力機器１０５ｂの特定の特徴に基づいて検出される様々な音声クリック対話を決定することができる（例えば、

音声クリック対話は、非音声入力機器１０５ｂをサポートする識別可能な対話を含むことができる）。例えば、マルチタッチディスプレイは、典型的には、表示された情報と対話するための様々な識別可能なジェスチャーをサポートするように構成されたタッチスクリーンディスプレイ機器を含む（例えば、ユーザは、特定のジェスチャー等の対話手法を用いて、マルチタッチスクリーンに表示された図形情報に対して拡大、縮小、回転、又はその他の制御を行うことができる）。このように、一例において、非音声入力機器１０５ｂはマルチタッチディスプレイを含むことができ、音声クリックモジュール１０８は、ユーザが非音声マルチタッチディスプレイ１０５ｂによってサポートされている識別可能な一又は複数のジェスチャーを行ったときに音声クリックの発生を検出するように構成される。

【００２９】

一実施態様において、ユーザは、音声クリックモジュール１０８によって検出される音声クリック対話をカスタマイズ又はそれ以外の方法で修正することができる。特に、音声クリックモジュール１０８によって検出された特定の機器対話を削除又は変更することができ、新しい機器対話を追加することもできる。このように、音声クリックモジュール１０８によって検出された音声クリック機器対話は、非音声入力機器１０５ｂ及び／又は音声クリックモジュール１０８が識別できる任意の適切な対話又は対話の任意の適切な組み合わせを含むことができる。

【００３０】

音声クリックモジュール１０８によってユーザが音声クリック機器対話を行っていることが検出されると、音声クリックモジュール１０８は、音声起動のために、当該音声クリック機器対話に関連づけられたコンテクスト情報を抽出することができる。特に、音声クリックモジュール１０８は、領域、項目、焦点、関心の焦点、又はその他のユーザによって選択されたデータに関連する情報を特定することができ、又は、これら以外にも、ユーザによって行われる特定の機器対話又は機器対話のシーケンスに関連する情報を特定することができる。このように、音声クリックモジュール１０８は、検出された音声クリックに関連して特定された情報を抽出することができる。この情報は、以前の、現在の、又は後続の単数又は複数の自然言語発話に関連づけられるコンテクスト情報として用いられる。

【００３１】

このように、音声クリックモジュール１０８における音声クリック（例えば、アイコンの選択、テキストの部分、地図表示における特定の座標、又はこれら以外の情報）の検出に応答して、音声クリックモジュール１０８は、動作、クエリ、コマンド、タスク、又はその他のリクエストであって検出された音声クリックを提供するために実行されるものを決定するために、自然言語発話音声入力（音声入力機器１０５ａによって受信される）を追加的なコンテクストとして用いるようにシステム１００に対して伝達することができる。このため、システム１００における様々な自然言語処理コンポーネントは、音声クリックと付随する自然言語発話とを組み合わせたコンテクストを用いて、音声クリック機器対話の意図を決定し、単数又は複数の動作、クエリ、コマンド、タスク、又はこれら以外のリクエストをマルチデバイス環境に配置された様々な機器の任意のものに対して適切に発出する（route）ことができる。

【００３２】

例えば、一実施態様において、マルチデバイス環境は、音声対応ナビゲーション機器を含むことができる。このように、例示的な音声クリック機器対話は、音声対応ナビゲーション機器と関連づけられたタッチスクリーンディスプレイ１０５ｂに表示された特定の交点にユーザがスタイラスを接触させることを含み、「この辺りにはどんなレストランがありますか？」等の発話をマイク１０５ａに提供することも含む。この例においては、音声クリックモジュール１０８は、音声クリックされた交点に関する情報を抽出することができ、この情報を付随する発話を処理するためのコンテクストとして用いることができる（すなわち、選択された交点は、ユーザの現在位置等の意味以外の、「この辺り」を解釈す

るためのコンテクストを提供することができる）。また、上述のように、音声入力は、タスク記述を決定するための追加的なコンテクストとして使用することができる。このように、発話は、システム１００の様々な自然言語処理コンポーネントを用いた認識及び会話的な解釈としても用いられる。この点については以下でさらに詳しく説明する。

【００３３】

一実施態様において、自動音声認識装置（ＡＳＲ）１１０は、音声入力機器１０５ａによって受信された発話の単数又は複数の予備的な解釈を生成することができる。例えば、ＡＳＲ１１０は、単数又は複数の動的に適応される認識文法を用いて、シラブル、語、句、又はその他の発話の音響的特徴を認識することができる。一実施態様における動的な認識文法は、単数又は複数の音響モデル（例えば、2005年8月5日に出願された「System and Method for Responding to Natural Language Speech Utterance」と題する係属中の米国特許出願第11/197,504に記述されている。当該出願の開示内容は参照により全体として本明細書に組み込まれる。）に基づく音声ディクテーションを用いて音素列を認識するために用いられる。

【００３４】

一実施態様において、ＡＳＲ１１０は、マルチパス音声認識を実行するように構成されてもよい。主音声認識エンジンは、発話の主トランスクリプションを生成することができ（例えば、ディクテーション文法のリストを用いる）、続いて、単数又は複数の二次音声認識エンジンからの単数又は複数の二次トランスクリプションを要求する（例えば、未知語のためのデコイ語を有する仮想的なディクテーション文法を用いる）。一実施態様における主音声認識エンジンは、主トランスクリプションの信頼度に基づいて二次トランスクリプションを要求することができる。

【００３５】

ＡＳＲ１１０において用いられる認識文法は、様々な語彙、辞書、シラブル、語、句、又はこれら以外の発話を認識するための情報を含む。一実施態様における認識文法に含まれる情報は、所定の発話について正確に認識する可能性を向上させるために動的に最適化される（例えば、語又は句の不正確な解釈があった場合には、不正確な解釈を繰り返す可能性を減らすために当該不正確な解釈を文法から除去してもよい）。また、認識文法に含まれている情報を継続的に動的に最適化するために、様々な形態の情報を用いることができる。例えば、システム１００は、環境情報（例えば、ピアツーピアの関係（affinities）、当該環境における様々な機器の機能等）、過去に用いられた情報（例えば、頻繁に行われるリクエスト、以前のコンテクスト等）、又は現在の会話的ダイアローグ又は対話に関連して短期間で共有された情報等を含む情報を有することができる。

【００３６】

一実施態様において、認識文法における情報は、コンテクスト又はアプリケーション固有のドメインに従ってさらに最適化することができる。特に、類似の発話は、当該発話に関連するコンテクストに応じて異なる内容に解釈されてもよい。このコンテクストには、ナビゲーション、音楽、映画、天気、ショッピング、ニュース、言語、時間的な又は地理的な近さ、又はこれら以外のコンテクストもしくはドメインが含まれる。例えば、「traffic」という語を含む発話は、コンテクストがナビゲーションなのか（すなわち道路状況）、音楽なのか（すなわち、1960年代のロックバンド）、又は映画なのか（すなわち、スティーブン・ソダーバーグ監督の映画）によって、異なる解釈を与えられる可能性がある。したがって、ＡＳＲ１１０は、上述の係属中の米国特許出願及び／又は2006年8月31日に出願された「Dynamic Speech Sharpening」と題する係属中の米国特許出願第11/513,269号において説明されているように、様々な技術を用いて自然言語発話の予備的な解釈を生成することができる。これらの特許出願の開示内容は、参照により全体として本明細書に組み込まれる。

【００３７】

このように、ＡＳＲ１１０は、音声クリックに含まれる自然言語発話の単数又は複数の予備的な解釈を会話言語プロセッサ１２０に提供することができる。この会話言語プロセ

ッサ１２０は、人間対人間の会話又は対話をモデルに構成された様々な自然言語処理コンポーネントを含むことができる。例えば、会話言語プロセッサ１２０は、意思決定エンジン１３０ａ、星座モデル１３０ｂ、単数又は複数のドメインエージェント１３０ｃ、コンテクストトラッキングエンジン１３０ｄ、誤認エンジン１３０ｅ、及び音声検索エンジン１３０ｆ等を含むことができる。また、会話言語プロセッサ１２０は、単数又は複数のデータレポジトリ１６０、及び、様々なコンテクスト又はドメインに関連づけられた単数又は複数のアプリケーション１５０に接続されてもよい。

【００３８】

このように、システム１００は、ユーザに協力的な会話を行わせ、ユーザの音声クリックを開始する意図に基づいて音声クリック機器対話を決定するために、会話言語プロセッサ１２０に関連づけられた様々な自然言語処理コンポーネントを用いることができる。より具体的には、意思決定エンジン１３０ａは、システム１００の機能やマルチデバイス環境における他の機器の機能に基づいて、所定のマルチモーダル機器対話の意味を定めることができる。例えば、ユーザが「この辺りにはどんなレストランがありますか？」という発話を決定するために特定の交点を音声クリックする上述の例を参照すれば、会話言語プロセッサ１２０は、音声クリックの会話の目的（conversationalgoal）を決定することができる（例えば、「どんな」は、データ取得を要求するクエリに関連する発話を示すことができる）。また、会話言語プロセッサ１２０は、コンテクストトラッキングエンジン１３０ｄを起動して、音声クリックのコンテクストを決定することができる。例えば、コンテクストトラッキングエンジン１３０ｄは、音声クリックコンテクストを決定するために、特定された焦点（すなわち、選択された交点）に関連づけられたコンテクストと、発話（すなわち、レストラン）に関連づけられたコンテクストとを組み合わせることができる。

【００３９】

その結果、音声クリックの組み合わされたコンテクスト（機器対話及び付随する発話の両方を含む）は、特定のクエリを発出するための十分な情報を提供することができる。例えば、クエリは、レストラン及び特定された交点に関連する様々なパラメータ又は基準を含むことができる。会話言語プロセッサ１２０は、処理のためにクエリが送られる特定の機器、アプリケーション、又はこれら以外のコンポーネントを選択することができる。例えば、一実施態様における会話言語プロセッサ１２０は、マルチデバイス環境における各機器の機能のモデルを含む星座モデル１３０ｂを評価することができる。一実施態様における星座モデル１３０ｂは、当該環境において各機器が利用できる情報処理リソーススス及びトレージリソースについての情報、並びに、ドメインエージェント、コンテクスト、機能、コンテンツ、サービス、又は機器の各々に関するこれら以外の情報の性質及び範囲を含むことができる。

【００４０】

このように、会話言語プロセッサ１２０は、星座モデル１３０ｂ及び／又は他の情報を用いて、起動可能な機器又は機器の組み合わせのいずれが、所定の音声クリック機器対話を処理するために好適な機能を有しているか決定することができる。例えば、上記の例を再び参照して説明すると、会話言語プロセッサ１２０は、音声クリックのコンテクストがナビゲーション機器との対話に関連することを決定し、そのため、ナビゲーションアプリケーション１５０を用いて処理するためにクエリを発出することができる。そして、クエリの結果が処理され（例えば、ベジタリアンレストランが好みであることなどのユーザに関する情報に基づいて結果を検討することができる）、出力機器１８０を介してユーザに返される。

【００４１】

図２は、本発明の様々な態様に従って、自然言語音声サービス環境においてマルチモーダル機器を同期させる例示的な方法のブロック図を示す。上述のように、マルチモーダル機器対話（又は「音声クリック」）は、一般に、ユーザが単数又は複数のマルチモーダル機器と単数又は複数の対話を行う一方、マルチモーダル機器との対話に関連する単数又は

複数の自然言語発話も提供するときに発生する。一実施態様において、マルチモーダル機器との対話に関連するコンテクスト情報は、（例えば、特定の動作、クエリ、コマンド、タスク又は他のリクエストを開始するための）音声クリックの意図を決定するために、自然言語発話に関連するコンテクスト情報と関連づけられる。

【００４２】

一実施態様において、様々な自然言語処理コンポーネントは、いつ音声クリックが発生したか決定するために、マルチモーダル機器を継続的に聴取又はその他の方法で監視するように構成されてもよい。このため、図２に示した方法は、マルチモーダル機器を継続的に聴取又はその他の方法で監視するコンポーネントを補正又はその他の方法で構成するために用いることができる。例えば、一実施態様における自然言語音声サービス環境は、異なる機能又はサービスを提供する複数のマルチモーダル機器を含み、ユーザは、任意の機器対話における様々な機器又は機能に関連するサービスを要求するために単数又は複数の音声クリックを行うことができる。

【００４３】

マルチモーダル機器対話又は音声クリックを継続的に聴取することができるように、当該環境における前記複数の機器の各々は、音声クリックに関連する情報を受信するように構成される。このように、一実施態様において、工程２１０は、当該環境における前記複数の機器の各々のために機器リスナー（devicelistener）を設定することを含む。また、工程２１０は、単数又は複数の新しい機器が当該環境に追加されたことに応答して実行されてもよい。工程２１０において設定される機器リスナーは、単数又は複数の処理装置又は他のハードウェアコンポーネントにおいて実行されるように構成された命令、ファームウェア、又はこれら以外のルーチンの任意の好適な組み合わせを含んでもよい。当該環境における機器の各々に対応付けられた機器リスナーは、機器の機能、特徴、サポートされているドメイン、又はこれら以外の機器に関する情報を決定するために当該機器と通信することができる。一実施態様における機器リスナーは、コンピュータ付属機器用に設計されたUniversalPlug and Playプロトコルを用いて当該機器と通信するように構成されてもよいが、マルチモーダル機器と通信する任意の好適なメカニズムを用いることができる。

【００４４】

当該環境における機器の各々について機器リスナーが設定されたとき（又は、当該環境に追加された新しい機器について機器リスナーが設定されたとき）に、工程２２０において、様々な機器リスナーが同期される。特に、様々な機器の各々は、異なる内部クロック又はそれ以外のタイミング機構を有していてもよく、工程２２０は、機器それぞれの内部クロック又はタイミング機構に従って様々な機器リスナーを同期させることを含んでもよい。一実施態様において、機器リスナーを同期させることは、各機器リスナーの各々が内部クロック又は関連する機器のタイミングに関する情報を公にすることを含んでも良い。

【００４５】

その後、単数又は複数の機器について単数又は複数のマルチモーダル対話又は音声クリックが発生すると、対応付けられた機器リスナーは、工程２３０において、当該音声クリックに関連する情報を検出することができる。例えば、一実施態様において、工程２１０で設定された様々な機器リスナーは、上述の図１に示された音声クリックモジュールに対応付けられる。したがって、工程２３０は、単数又は複数の機器リスナー又は音声クリックモジュールが、ユーザと単数又は複数の機器との対話（例えば、当該機器に関連するデータを選択し、当該機器に関連する焦点又は関心の焦点を特定し、又は当該機器との単数又は複数の対話又は対話シーケンスを行うこと）の発生を検出することを含むことができる。また、工程２４０は、工程２３０において検出された機器対話に関連するユーザからの発話を取得することを含んでもよい。

【００４６】

例えば、ディスプレイ装置に表示されたウェブページを見ているユーザは、当該ウェブページ上で製品名を見て当該製品の購入に関するより多くの情報の入手を望む可能性があ

る。ユーザは、当該商品名を含むウェブページからテキストを選択し（例えば、マウス又はキーボードを用いて当該テキストをハイライトすることにより）、その後、「これはAmazon.comで入手できますか？」と質問するために音声クリックを開始する。この例において、工程２３０は、ディスプレイ装置に関連づけられた機器リスナーが当該商品名に関連づけられたテキストの選択を検出することを含むことができ、一方、工程２４０は、Amazon.comにおける当該商品の在庫を問い合わせる発話を取得することを含むことができる。

【００４７】

上述のように、ユーザからの入力を受信する各機器は、内部クロック又はタイミング機構を有することができる。したがって、工程２５０において、各機器は、入力がいつ受信されたかをローカルで決定することができ、音声クリックモジュールに対して入力が受信されたことを通知することができる。特に、所定の音声クリックは、単数又は複数の他の機器との単数又は複数の追加的な対話に加えて、少なくとも音声入力機器を介して受信された自然言語発話を含むことができる。発話は、機器対話の前に、機器対話と同時に、又は機器対話の後に続いて受信される。これにより、工程２５０は、対応付けられた発話との相関のために機器対話のタイミングを決定することを含む。具体的には、工程２６０は、工程２２０を参照して説明したように同期された機器リスナー信号を用い、機器対話用の信号と発話用の信号とを連携させることを含んでもよい。機器対話及び発話信号を一致させる場合には、連携された音声コンポーネント及び非音声コンポーネントを含む音声クリック入力が生成される。音声クリック入力に対しては、その後、以下で詳細に述べる追加的な自然言語処理がなされる。

【００４８】

図３は、本発明の様々な態様に従って、自然言語音声サービス環境においてマルチモーダル機器対話を処理する例示的な方法のフロー図を示す。上述のように、マルチモーダル機器対話（又は「音声クリック」）は、一般に、ユーザが単数又は複数のマルチモーダル機器と対話する一方で当該機器対話に関連する単数又は複数の自然言語発話も提供するときに発生する。このため、一実施態様において、図３に示された方法は、単数又は複数の音声クリックがいつ発生したかを決定するために単数又は複数の自然言語処理コンポーネントが単数又は複数のマルチモーダル機器を継続して聴取し又はそれ以外の方法で監視する場合に実行することができる。

【００４９】

一実施態様において、単数又は複数の機器対話を音声クリックの開始として定義することができる。例えば、任意の所定の電子機器は、一般に、識別可能な様々の対話をサポートすることができ、所定の動作、コマンド、クエリ、又はその他の実行要求を実行することができる。このように、一実施態様において、機器対話の任意の適切な組み合わせを音声クリックとして定義することができる。この機器対話は、所定の電子機器が一意に認識可能なもの、又は、所定の電子機器が一意に認識可能な信号を生成するために用いることができるものである。ここで、音声クリックは、自然言語発話の処理を、関連する機器対話に対応付けられたコンテクストと共に行うことを示す信号を提供することができる。例えば、４方向又は５方向ナビゲーションボタンを有する機器は、特定の識別可能な対話をサポートすることができる。特定の方法でナビゲーションボタンを押すことにより、マップ表示の制御やルート計算等の特定のタスク又はそれ以外の動作が実行される。他の例において、ホイールセレクタを有するBlackBerry（登録商標）機器は、特定の焦点又は関心の焦点の上でカーソルを回転させること、特定のデータ又は所定のアプリケーションを選択するめにホイールを押下すること、又はその他の様々な対話等をサポートすることができる。機器対話を用いることにより、自然言語発話が当該機器対話と対応付けられたコンテクストとともに処理されるタイミングを示すことができるが、これに限られず、この特定の機器対話は所定の実装態様によって変化する。例えば、関連する機器対話は、タッチセンサー式スクリーン上で道具を指し示したり描いたりするジェスチャーをすること（例えば、耳の形の線を描くこと）、長いタッチやダブルタップ等の固有の対話方法を含むことができ、及び／又は、システムが上述した継続的聴取モードで動作している場合には、

予め定められたコンテクスト命令語によって、現在の機器コンテクストを当該コンテクスト命令語に続く音声入力の部分とともに処理することを示すことができる（例えば、命令語が「OK」、「Please」、「Computer」、又はその他の好適なワードである場合には、ユーザはマップ上の特定の点を選択して「Pleasezoom in」と言ったり、Ｅメールが表示されたときに「OK read it」と言ったりすることができる）。

【００５０】

このように、工程３１０は、音声クリックの開始を伝達する単数又は複数の機器対話の発生を検出するために自然言語音声サービス環境においてマルチモーダル機器対話を処理することを含んでもよい。具体的には、工程３１０において検出される機器対話は、固有の、認識可能な、又はその他のユーザに関連する識別可能な信号を電子機器に生成させる任意の好適な対話を含むことができる。この信号は、ユーザが当該機器の特定の機能に応じてデータを選択し、焦点又は関心の焦点を特定すること、アプリケーション又はタスクを起動すること、又は他の方法で機器と対話することに関連して当該電子機器によって生成される。

【００５１】

工程３１０において検出される対話は、前記ユーザ対話に応答して機器により生成される上述の特定の信号に加えて、音声クリックの開始を伝達することができ、これにより、以前の、現在の、又は後続の自然言語音声入力は、工程３１０において検出された機器対話を解釈するための追加的なコンテクストを提供することになる。例えば、自然言語処理システムは一般に、特定の機器対話が発生したとき（例えば、マイクのスイッチを入れるためにボタンを押したとき）に、音声入力を受容するように構成される。このように、図３に示された方法において、これから到着する音声入力を知らせる機器対話は、電子機器との任意の好適な対話又は電子機器との対話の任意の好適な組み合わせをさらに含んでもよい。この対話は、ユーザが機器の特定の機能に応じてデータを選択すること、ユーザが焦点又は関心の焦点を特定すること、ユーザがアプリケーション又はタスクを起動すること、又はその他の方法でユーザが機器と対話することを含む。

【００５２】

このため、音声クリック機器対話が工程３１０において検出されると、工程３２０において、自然言語音声入力を工程３２０において検出された対話と対応付けるべきことを示す音声クリック信号が生成される。続いて、工程３３０は、工程３１０において検出された対話と対応付けられるユーザ発話を取得することを含むことができる。一実施態様において、工程３１０において検出された対話は、後から音声入力が提供されることを示すことができる。ただし、一又は複数の実施態様において、工程３３０において取得される発話は、工程３１０において検出される対話の前又は後に提供されてもよい（例えば、ユーザは「このアーティストをiTunes（登録商標）で調べる」等の発話を提供した後にメディアプレイヤー上でアーティスト名を音声クリックすることができ、又は、アーティスト名を音声クリックしている間に発話を提供することができ、又は、アーティスト名を音声クリックした後に発話を提供することができる）。

【００５３】

工程３４０は、音声クリック機器対話に関連する情報及び対応する自然言語発話が受信されると、当該機器対話及び当該対応する発話のためにコンテクスト情報を抽出し組み合わせることができる。具体的には、音声クリック機器対話から抽出されたコンテクスト情報は、領域、項目、焦点、関心の焦点、もしくはユーザによって選択されたデータに関連する情報、又は、ユーザによって行われた特定の機器対話もしくは特定の機器対話のシーケンスに関連する情報を含むことができる。その後、機器対話に関する抽出されたコンテクストは、工程３３０において取得された自然言語発話について抽出されたコンテクストと組み合わされる。組み合わされたコンテクスト情報は、工程３５０において、音声クリックの意図を決定するために用いることができる。

【００５４】

例えば、例示的な音声クリック機器対話において、ユーザは、格納されている音楽をメ

ディアプレイヤーからバックアップストレージ機器に選択的に複製することができる。ユーザは、メディアプレイヤー上で音楽をブラウズしている間に、特定の楽曲を聴いて「このアルバム全部をコピー」と言うことにより当該楽曲を音声クリックすることができる（例えば、当該楽曲を強調している間にメディアプレイヤー上の特定のボタンを長時間押下することによって音声クリックする）。この例において、工程３１０は、ボタンを長時間押下するという対話を検出することを含むことができる。これにより、工程３２０において音声クリック信号が生成される。続いて、「このアルバム全部をコピー」という発話が工程３３０において取得され、当該音声クリック機器対話及び当該発話に対応するコンテクスト情報が、工程３４０において組み合わせられる。特に、機器対話のコンテクストは、選択された楽曲に関する情報等を含むことができる（例えば、当該コンテクストは、音楽ファイルのID3タグ等の当該楽曲に対応付けられたメタデータに含まれる情報をさらに含むことができる）。また、発話のコンテクストは、複製動作及び当該選択された楽曲を含むアルバムを特定する情報を含むことができる。

【００５５】

このようにして、マルチモーダル機器との音声クリック対話に関連するコンテクスト情報を、自然言語発話に関連するコンテクスト情報と組み合わせることができる。これにより、工程３５０は、音声クリック対話の意図を決定することができる。例えば、上述の例を参照すると、工程３５０において決定される意図には、強調された楽曲を含むアルバムをメディアプレイヤーからバックアップストレージ機器に複製する意図を含むことができる。このように、工程３５０において音声クリックの意図が決定されたことに応答して、単数又は複数のリクエストが工程３６０に適切に発出される。ここで説明した例では、工程３６０は、強調された楽曲を含むアルバムに対応するデータを全て特定するために、メディアプレイヤーに単数又は複数のリクエストを発出すること、及び、特定されたデータのメディアプレイヤーからバックアップストレージ機器への複製を管理することができる機器の任意の適切な組み合わせ（例えば、メディアプレイヤー及びストレージ機器の両方と接続されたパーソナルコンピュータ）に対して単数又は複数のリクエストを発出することを含むことができる。

【００５６】

図４は、本発明の様々な態様に従って、自然言語音声サービス環境においてマルチモーダル機器対話を処理して単数又は複数のトランザクションリード又は「クリックスルー」を生成する例示的な方法のフロー図を示す。具体的には、図４に示された方法は、検出された単数又は複数の音声クリック機器対話に応答して実行される単数又は複数の動作と組み合わせてトランザクションリード又はクリックスルーを生成するために用いることができる。

【００５７】

例えば、工程４１０は、ユーザから受信した単数又は複数の音声クリック機器対話を検出することを含むことができる。このとき、音声クリック機器対話は、単数又は複数の関連する自然言語発話に結びつけられた単数又は複数の機器対話の任意の好適な組み合わせを含むことができる。次に工程４２０において音声クリック機器対話を行うユーザの意図が確定された後、工程４３０において、音声クリック対話を決定するために、確定された意図に基づいて単数又は複数のリクエストが単数又は複数の処理装置に発出される。一実施態様において、工程４１０、４２０、及び４３０は、図２及び図３を参照して説明した方法と類似の方法で実行することができる。これにより、機器対話のための信号は、単数又は複数の自然言語発話のための信号と連携し、コンテクスト情報は、音声クリック機器対話の意図を決定するために当該信号から抽出される。

【００５８】

図４に示した方法は、ユーザの意図に基づいて単数又は複数のリクエストを発出することに加えて、単数又は複数のクリックスルーに繋がる可能性のある単数又は複数のトランザクションリード（transaction lead）を生成することをさらに含むことができる。例えば、クリックスルーは、一般に、ユーザが電子広告をクリックし又は他の方法で電子広告

を選択して当該広告に関連する単数又は複数のサービスにアクセスする行為を指す。多くの電子システムにおいて、クリックスルー又はクリックスルー率は、ユーザと電子広告との対話を計測するメカニズムを提供することができ、当該広告をユーザへ提供する主体に対する支払額を決定するために広告主によって用いられる様々な測定値を提供することができる。

【００５９】

このように、図４に記載された方法によって、広告又は推薦を含むトランザクションリードを生成することができる。これにより、特定の機器対話と組み合わせられたユーザの音声入力は、トランザクションリードを生成するための追加的な焦点（focus）を提供することができる。この方法において、ユーザに提供される広告又は推薦は、ユーザが対話する特定の情報との関連性が高くともよい。また、自然言語認知モデル及びユーザの嗜好に関する共有された情報を用いることにより、特定のユーザのためにカスタマイズされたターゲットトランザクションリードのための追加的なコンテクストを提供することができ、これにより、音声サービスプロバイダーに対する支払いを生み出すクリックスルーがより生じやすくなる。

【００６０】

このように、工程４４０は、音声クリック機器対話を行うユーザの意図に基づいて単数又は複数のリクエストを発出することに加えて、決定された意図に基づいて単数又は複数のトランザクションリードを生成することを含むことができる。具体的には、ターゲット広告を実行する任意の適切なシステムにおいてローカルの音声及び非音声コンテクストを状態データ（statedata）として用いるという点で、対応する自然言語発話及び組み合わされた機器対話のコンテクストに基づいて、トランザクションリードをユーザに「より親密な」方法で処理することができる。例えば、ユーザがナビゲーション機器に表示された交点を選択し「この辺りにあるレストランを探す」と言う上述の例を参照すると、工程４４０において生成されるトランザクションリードは、当該交点に近いレストランの単数又は複数の広告又は推薦を含むことができる。このレストランは、ユーザの短期間及び長期間の嗜好（例えば、好みのレストランの種類、好みの価格帯等）に関する情報に基づいて、当該ユーザに関して絞り込まれる。

【００６１】

その後、トランザクションリードを、（例えば、地図表示上の選択可能な地点として）ユーザに提示することができる。工程４５０においては、単数又は複数の追加的なマルチモーダル機器対話が発生したか否か、又は、いつ発生したかを決定するために、引き続きユーザのマルチモーダル機器対話が監視される。追加的な対話が発生していない場合は、ユーザは当該トランザクションリードに従った行動を起こさなかったという決定がなされ、処理は終了する。追加的なマルチモーダル対話が起こったときには、工程４８０において、当該マルチモーダル入力を処理することにより入力の意図が決定され、単数又は複数のリクエストがその意図に応じて発出される。また、工程４６０は、マルチモーダル入力が工程４４０において生成されたトランザクションリードに関連するか否か決定することを含んでもよい。例えば、ユーザは、当該トランザクションリードに関連する追加の動作又は情報を要求する発話、非音声機器対話、又は音声クリック機器対話を提供することによって、広告又は推薦されたレストランの一つを選択することができる。この場合、工程４７０は、工程４４０において生成されたトランザクションリードに関連してトランザクションクリックスルーを処理することをさらに含むことができる。このトランザクションクリックスルーは、特定の主体への支払い額を決定し又はその他の方法で特定の主体（例えば、音声サービスのプロバイダー又はトランザクションリード又はトランザクションクリックスルーに関連する他の主体）の収入を生み出すために用いられる。

【００６２】

本発明の様々な実施形態は、ハードウエア、ファームウェア、ソフトウェア、またはこれらの適当な組み合わせによって実現される。本実施形態は、機械読み取り可能な媒体に記憶される命令によっても実装可能である。このような命令は、単数又は複数のプロセッ

サによって読み出されて実行される。機械読み取り可能な媒体は、機械（例えば、コンピュータ装置）によって読み取り可能な形式の情報を保存し伝送する様々なメカニズムを含む。例えば、機械読み取り可能な記録媒体は、ＲＯＭ、ＲＡＭ、磁気ディスクストレージ媒体、光学ストレージ媒体、フラッシュメモリ装置等のストレージメディアを含む。また、機械読み取り可能な伝送媒体は、搬送波等の伝搬信号、赤外線信号、デジタル信号、又はこれら以外の伝送メディアを含む。さらに、ファームウェア、ソフトウェア、ルーチン、または命令は、上記開示において、所定の処理を実行する特定の例示的な態様及び実装形態及びの観点から説明されたが、そのような説明は便宜上のものに過ぎず、そのような処理は、実際には、コンピュータ装置、プロセッサ、制御装置、又はファームウェア、ソフトウェア、ルーチン、もしくは命令を実行するその他の装置によりなされるものであることは明らかである。

【００６３】

本明細書においては、自然言語音声サービス環境においてマルチモーダル機器対話を処理する技術を中心に説明しているが、本明細書において説明された特定の態様及び実装形態に関連して説明された自然言語処理機能に関しては、当該自然言語処理機能に加えて、又は自然言語処理機能に代えて、様々な追加的な自然言語処理機能を用いることができる。例えば、本明細書において説明されたシステム及び方法は、上述の係属中の米国特許出願において説明されている技術に加えて、2005年8月5日に出願された「System andMethod for Responding to Natural Language Speech Utterance」と題する係属中の米国特許出願第11/197,504号、2005年8月10日に出願された「System and Method of SupportingAdaptive Misrecognition in Conversational Speech」と題する米国特許出願第11/200,164号、2005年8月29日に出願された「Mobile System and Method of Supporting Natural LanguageHuman-Machine Interactions」と題する米国特許出願第11/212,693号、2006年10月16日に出願された「System and Method for a Cooperative Conversational VoiceUser Interface」と題する米国特許出願第11/580,926号、2007年2月6日に出願された「System andMethod for Selecting and Presenting Advertisements Based on Natural LanguageProcessing of Voice-Based Input」と題する米国特許出願第11/671,526、及び2007年12月11日に出願された「System and Method for Providing a Natural Language VoiceUser Interface in an Integrated Voice Navigation Services Environment」と題する米国特許出願第11/954,064号に記載された自然言語処理機能を用いることもできる。これらの特許出願の開示内容は、参照により全体として本明細書に組み込まれる。

【００６４】

したがって、本開示の様々な態様及び実装形態は、本明細書において特定の特性、構造、または特徴を含むように説明されるが、全ての態様や実装形態が必ずしも特定の特性、構造、または特徴を含むわけではない。さらに、特定の特性、構造、または特徴が所定の態様や実装形態に関連して説明される際には、明示的に説明されるか否かにかかわらず、そのような特性、構造または特徴は他の態様や実装形態に関連付けて含むことも可能である。このように、上記説明に対しては、本発明の範囲又は趣旨から逸脱することなく様々な変更や修正を行うことが可能である。このように、本明細書及び図面は例示に過ぎず、本発明念の範囲は、特許請求の範囲の記載によってのみ定められる。

【図１】

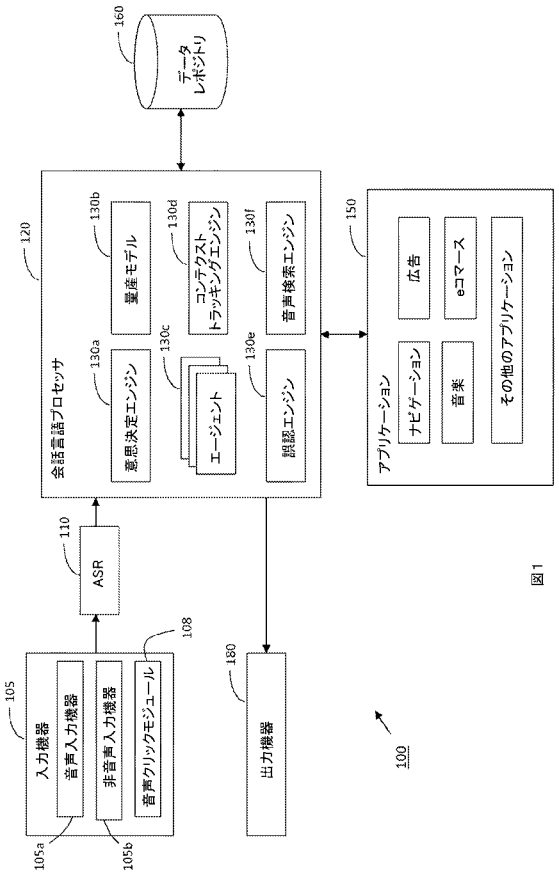

図1

【図２】

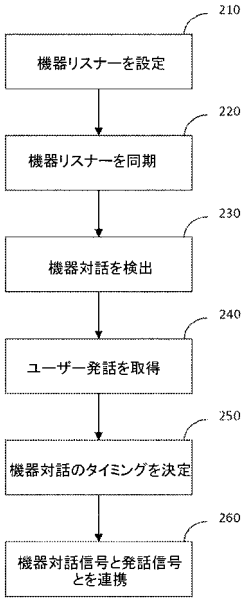

図2

【図３】

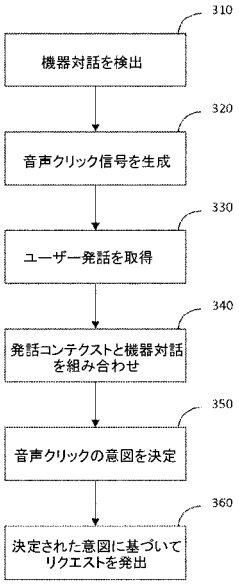

図3

【図４】

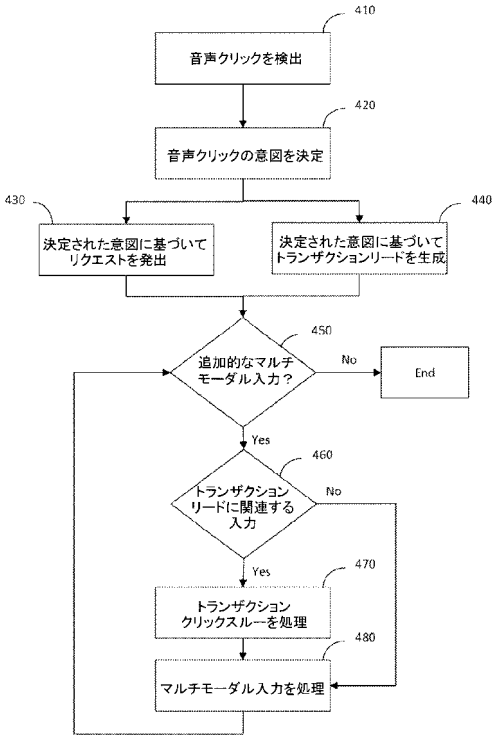

図4

【手続補正書】
【提出日】平成28年10月14日(2016.10.14)
【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】
単数又は複数の電子機器を含む自然言語音声サービス環境において単数又は複数の入力機器から受信した単数又は複数のマルチモーダル機器対話を処理する方法であって、会話言語プロセッサにおいて実装され、
前記会話言語プロセッサによって非音声対話及び音声対話を含む少なくとも一つのマルチモーダル機器対話を受信する工程であって、前記非音声対話は第１の入力機器から受信され、前記音声対話は前記第１の入力機器又は他の入力機器から受信され、前記音声対話は前記非音声対話に関連する少なくとも一つの自然言語発話を含む、工程と、
前記会話言語プロセッサによって、前記非音声対話に基づいて少なくとも第１のコンテクスト情報を決定する工程と、
前記会話言語プロセッサによって、前記音声対話に基づいて少なくとも第２のコンテクスト情報を決定する工程と、
前記会話言語プロセッサによって、前記第１のコンテクスト情報及び前記第２のコンテクスト情報に基づいて、前記マルチモーダル機器対話の意図を決定する工程と、
前記会話言語プロセッサによって、前記単数又は複数の電子機器の各々において利用可能なエージェント、コンテンツ、及び／又はサービスを記述する配列モデルにアクセスする工程と、
前記会話言語プロセッサによって、前記単数又は複数の電子機器の中から少なくとも第１の機器を特定し、前記決定された意図及び前記配列モデルに基づいて前記マルチモーダル機器対話を処理する工程と、
前記会話言語プロセッサによって、前記第１の機器による処理のための前記マルチモーダル機器対話を特定する情報を前記第１の機器に発出する工程と、
前記会話言語プロセッサによって、前記マルチモーダル機器対話の前記決定された意図に基づいてトランザクションリードを生成する工程と、
前記会話言語プロセッサによって、前記トランザクションリードを提供させる工程と、
を含む方法。
【請求項２】
前記配列モデルは前記単数又は複数の電子機器の各々に対応付けられた自然言語リソース、動的状態、及び意図決定能力をさらに記述する、請求項１の方法。
【請求項３】
前記生成されたトランザクションリードは、前記自然言語発話の前記決定された意図に関連する広告又は推薦の少なくとも一つを含む、請求項１の方法。
【請求項４】
前記会話言語プロセッサによって、前記トランザクションリードに関連するユーザ入力を受信する工程と、
前記会話言語プロセッサによって、前記トランザクションリードに関連する前記ユーザ入力の受信に応答して、トランザクションクリックスルーを処理する工程と、
をさらに含む請求項１の方法。
【請求項５】
前記単数又は複数の電子機器は前記第１の入力機器を含む、請求項１の方法。
【請求項６】
前記第１の入力機器は前記単数又は複数の電子機器とは別体である、請求項１の方法。

【請求項７】
前記第１の入力機器は機器リスナーに対応付けられ、
前記会話言語プロセッサによって、前記機器リスナーを介して前記非音声対話を検出する工程
をさらに含む請求項１の方法。
【請求項８】
前記会話言語プロセッサによって、第２の機器リスナーを第２の機器に対応付ける工程をさらに含む請求項７の方法。
【請求項９】
前記非音声対話は前記単数又は複数の電子機器のうちの単数又は複数に対応付けられた単数又は複数の固有で識別可能な対話を含む、請求項１の方法。
【請求項１０】
単数又は複数の電子機器を含む自然言語音声サービス環境において単数又は複数の入力機器から受信した単数又は複数のマルチモーダル機器対話を処理するシステムであって、
非音声対話及び音声対話を含む少なくとも一つのマルチモーダル機器対話を受信する工程であって、前記非音声対話は第１の入力機器から受信され、前記音声対話は前記第１の入力機器又は他の入力機器から受信され、前記音声対話は前記非音声対話に関連する少なくとも一つの自然言語発話を含む、工程と、
前記非音声対話に基づいて少なくとも第１のコンテクスト情報を決定する工程と、
前記音声対話に基づいて少なくとも第２のコンテクスト情報を決定する工程と、
前記第１のコンテクスト情報及び前記第２のコンテクスト情報に基づいて、前記マルチモーダル機器対話の意図を決定する工程と、
前記単数又は複数の電子機器の各々において利用可能なエージェント、コンテンツ、及び／又はサービスを記述する配列モデルにアクセスする工程と、
前記単数又は複数の電子機器の中から少なくとも第１の機器を特定し、前記決定された意図及び前記配列モデルに基づいて前記マルチモーダル機器対話を処理する工程と、
前記第１の機器による処理のための前記マルチモーダル機器対話を特定する情報を前記第１の機器に発出する工程と、
前記マルチモーダル機器対話の前記決定された意図に基づいてトランザクションリードを生成する工程と、
前記トランザクションリードを提供させる工程と、
を実行するように構成された会話言語プロセッサを含むシステム。
【請求項１１】
前記配列モデルは前記単数又は複数の電子機器の各々に対応付けられた自然言語リソース、動的状態、及び意図決定能力をさらに記述する、請求項１０のシステム。
【請求項１２】
前記生成されたトランザクションリードは、前記自然言語発話の前記決定された意図に関連する広告又は推薦の少なくとも一つを含む、請求項１０のシステム。
【請求項１３】
前記会話言語プロセッサはさらに
前記トランザクションリードに関連するユーザ入力を受信する工程と、
前記トランザクションリードに関連する前記ユーザ入力の受信に応答して、トランザクションクリックスルーを処理する工程と、
を実行するように構成された、請求項１０のシステム。
【請求項１４】
前記単数又は複数の電子機器は前記第１の入力機器を含む、請求項１０のシステム。
【請求項１５】
前記第１の入力機器は前記単数又は複数の電子機器とは別体である、請求項１０のシステム。
【請求項１６】

前記第１の入力機器は機器リスナーに対応付けられ、前記会話言語プロセッサはさらに
前記機器リスナーを介して前記非音声対話を検出する工程
を実行するように構成された、請求項１０のシステム。
【請求項１７】
前記会話言語プロセッサはさらに
第２の機器リスナーを第２の機器に対応付ける工程
を実行するように構成された、請求項１６のシステム。
【請求項１８】
前記非音声対話は前記単数又は複数の電子機器のうちの単数又は複数に対応付けられた単数又は複数の固有で識別可能な対話を含む、請求項１０のシステム。

フロントページの続き

(51)Int.Cl.	F I		テーマコード（参考）
	Ｇ１０Ｌ 15/24	Ｚ	
	Ｇ１０Ｌ 15/22	３００Ｚ	

(72)発明者 ボールドウィン，ラリー
アメリカ合衆国 ワシントン州 ９８０３８，メイプルバレー，エスイー ２４３番プレイス ２３１２５
(72)発明者 ワイダー，クリス
アメリカ合衆国 ワシントン州 ９８１３３，シアトル，グリーンウッドアベニュー ノース ＃２２０ １０７５７

Ｆターム(参考) 5E555 AA11 BA06 BA15 BB06 BB15 BC04 CA02 CA12 CA47 CB02
CB12 CB64 CC01 EA23 FA00
5L049 BB08

【外国語明細書】
2017037658000001.pdf