

(19) 日本国特許庁(JP)

## (12) 特許公報(B2)

(11) 特許番号

特許第5336590号  
(P5336590)

(45) 発行日 平成25年11月6日(2013.11.6)

(24) 登録日 平成25年8月9日(2013.8.9)

(51) Int.Cl.

F 1

G 1 O L 15/32 (2013.01)  
G 0 6 F 9/50 (2006.01)G 1 O L 15/28 3 6 O Z  
G 1 O L 15/28 3 6 O B  
G 1 O L 15/28 3 6 O C  
G 0 6 F 9/46 4 6 5 Z

請求項の数 21 (全 29 頁)

(21) 出願番号 特願2011-516887 (P2011-516887)  
 (86) (22) 出願日 平成21年7月2日 (2009.7.2)  
 (65) 公表番号 特表2011-527030 (P2011-527030A)  
 (43) 公表日 平成23年10月20日 (2011.10.20)  
 (86) 国際出願番号 PCT/US2009/049604  
 (87) 国際公開番号 WO2010/003109  
 (87) 国際公開日 平成22年1月7日 (2010.1.7)  
 審査請求日 平成24年6月29日 (2012.6.29)  
 (31) 優先権主張番号 12/166,822  
 (32) 優先日 平成20年7月2日 (2008.7.2)  
 (33) 優先権主張国 米国(US)

(73) 特許権者 507103802  
 グーグル・インコーポレーテッド  
 アメリカ合衆国・カリフォルニア・940  
 43・マウンテン・ビュー・アンフィシア  
 ター・パークウェイ・1600  
 (74) 代理人 100108453  
 弁理士 村山 靖彦  
 (74) 代理人 100064908  
 弁理士 志賀 正武  
 (74) 代理人 100089037  
 弁理士 渡邊 隆  
 (74) 代理人 100110364  
 弁理士 実広 信哉

最終頁に続く

(54) 【発明の名称】並列認識タスクを用いた音声認識

## (57) 【特許請求の範囲】

## 【請求項 1】

コンピュータで実施される方法であって、  
 音声信号を受け取る段階と、

コンピュータシステムにより、複数の信号を送る段階と、を含み、前記複数の信号のうちの各信号は、複数の音声認識システム(SRS)のうちのあるSRSによる音声認識タスクの実行を開始するために構成され、各音声認識タスクは、(1)前記音声信号に含まれると予想される音声を特定する認識結果と、(2)当該SRSによって生成された認識結果が正しい可能性を示す信頼値との生成を含み、

前記方法は、

複数の音声認識タスクの一部が完了したことの検出に応じて、前記コンピュータシステムにより、(1)前記完了した音声認識タスクのそれぞれに関する1つまたは複数の認識結果と、(2)前記1つまたは複数の認識結果のそれぞれに関する1つまたは複数の信頼値とを生成する段階と、

前記音声認識タスクの完了した部分に関する前記1つまたは複数の信頼値のうちの各信頼値が閾値を満たすかどうか判定する段階と、

前記音声認識タスクの完了した部分に関する1つまたは複数の信頼値が、前記閾値を満たすとの判定に応じて、前記複数のSRSのうち少なくとも1つのSRSに、未完了でかつ信頼値を生成していない音声認識タスクを中断させるように構成された信号を送る段階と、

10

20

前記生成された 1 つまたは複数の認識結果のうち少なくとも 1 つに少なくとも部分的に基づいて、最終的な認識結果を出力する段階とを含む、方法。

【請求項 2】

音声認識タスクが、前記音声認識タスクの完了のために許可された最大時間枠を指定する閾値を超過した場合、前記複数の SRS のうち少なくとも 1 つの SRS に、前記音声認識タスクを中断させるように構成された信号を送る段階をさらに含む、請求項 1 に記載の方法。

【請求項 3】

完了した音声認識タスクが、生成された認識結果中の低い信頼値と関連している場合、前記複数の SRS のうち少なくとも 1 つの SRS に、未完了の音声認識タスクを中断させるように構成された信号を送る段階をさらに含む、請求項 1 に記載の方法。 10

【請求項 4】

前記複数の SRS のうち少なくとも 1 つが、複数の認識結果を生成する、請求項 1 に記載の方法。

【請求項 5】

前記複数の認識結果のうち上位 N 個の認識結果と関連する信頼値に基づいて、前記上位 N 個の認識結果を順位付けするステップをさらに含み、ここで、N は正の整数を表す、請求項 4 に記載の方法。

【請求項 6】

前記複数の SRS のうち異なる SRS によって生成された信頼値に基づいて、前記上位 N 個の認識結果と関連する信頼値の移動平均を求める段階をさらに含む、請求項 5 に記載の方法。 20

【請求項 7】

いずれの認識結果が前記信頼値の中から最も高い信頼値と関連するかに基づいて、前記上位 N 個の結果から前記最終的な認識結果を選択する段階をさらに含む、請求項 6 に記載の方法。

【請求項 8】

いずれの認識結果が最も大きい信頼値と関連するかに基づいて、前記最終的な認識結果を選択する段階をさらに含む、請求項 1 に記載の方法。

【請求項 9】

前記複数の SRS のうち、前記最終的な認識結果および前記最終的な認識結果と関連する少なくとも 1 つの信頼値を生成した SRS の数に基づいて、前記最終的な認識結果を選択する段階をさらに含む、請求項 1 に記載の方法。 30

【請求項 10】

最大時間枠が超過した場合、前記複数の SRS のうち少なくとも 1 つの SRS に、複数の音声認識タスクのそれぞれを中断させるように構成された信号を送る段階をさらに含む、請求項 1 に記載の方法。

【請求項 11】

発話を繰り返すようにユーザに促す段階をさらに含む、請求項 10 に記載の方法。

【請求項 12】

人間のオペレータが前記音声信号を受ける、との要求を送る段階をさらに含む、請求項 10 に記載の方法。 40

【請求項 13】

正規化された尺度での前記信頼値の分布に基づいて、前記最終的な認識結果を選択する段階をさらに含む、請求項 1 に記載の方法。

【請求項 14】

前記 1 つまたは複数の認識結果のうち 1 つの認識結果が高い信頼値に関連し、かつ前記 1 つまたは複数の認識結果のうち他の認識結果のうち少なくとも一部が低い信頼値の比較的小さい範囲内で互いに集中する場合、前記 1 つの認識結果を前記最終的な認識結果として選択する段階をさらに含む、請求項 13 に記載の方法。 50

**【請求項 15】**

前記最終的な認識結果の選択において、前記1つまたは複数の認識結果のうち特定の認識結果を生成した、前記複数のSRSのうち少なくとも2つのSRS間の相関に基づいて、前記特定の認識結果を重み付けする段階をさらに含む、請求項1に記載の方法。

**【請求項 16】**

前記複数のSRSのうち少なくとも一部が、異なる言語モデルまたは音響モデルを含む、請求項1に記載の方法。

**【請求項 17】**

コンピュータで実施される方法であって、  
音声信号を受け取る段階と、

10

複数の音声認識システム(SRS)による音声認識タスクを開始する段階とを含み、各SRSが、前記音声信号に含まれると予想される音声を特定する認識結果と、前記認識結果の正確さの信頼度を示す信頼値とを生成するように構成され、

前記方法は、

1つまたは複数の認識結果、および前記1つまたは複数の認識結果に関する1つまたは複数の信頼値を生成することを含む前記音声認識タスクの一部を完了する段階と、

前記1つまたは複数の信頼値が信頼閾値を満たすかどうか判定する段階と、

前記1つまたは複数の信頼値が信頼閾値を満たす場合、未完了でかつ信頼値を生成していないSRSに関する音声認識タスクの残りの部分を停止する段階と、

前記生成された1つまたは複数の認識結果の少なくとも1つに基づいて、最終的な認識結果を出力する段階とを含む、方法。

20

**【請求項 18】**

前記最終的な認識結果が正しくないと指示を受け取った場合、前記音声認識タスクの停止された残りの部分を再開する段階をさらに含む、請求項17に記載の方法。

**【請求項 19】**

受け取った音声信号内に符号化されたと予想される音声を識別するための複数の音声認識タスクの実行を開始する複数の音声認識システム(SRS)を具備するシステムであって、各音声認識タスクは、(1)前記受け取った音声信号に含まれると予想される音声を特定する認識結果と、(2)当該SRSによって生成された認識結果が正しい可能性を示す信頼値との生成を含み、

30

前記システムは、認識管理モジュールを具備し、前記認識管理モジュールは、

完了した複数の音声認識タスクの一部に関して、1つまたは複数の認識結果と、前記1つまたは複数の認識結果と関連する1つまたは複数の信頼値とを受け取り、

前記音声認識タスクの完了した部分に対する1つまたは複数の信頼値のうちの各信頼値が閾値を満たすかどうか判定するとともに、

前記音声認識タスクの完了した部分に関する1つまたは複数の信頼値が閾値を満たすとの判定に応じて、前記複数のSRSのうち少なくとも1つのSRSに、未完了でかつ信頼値を生成していない音声認識タスクを中断させるように構成された信号を送り、

前記システムは、前記1つまたは複数の認識結果に関連する1つまたは複数の信頼値に基づいて、選択された最終的な認識結果を送るためのインターフェースを具備する、システム。

40

**【請求項 20】**

受け取った音声信号内に符号化されたと予想される音声を識別するための複数の音声認識タスクの実行を開始する複数の音声認識システム(SRS)を具備するシステムであって、各音声認識タスクは、(1)前記受け取った音声信号に含まれると予想される音声を特定する認識結果と、(2)前記SRSによって生成された認識結果が正しい可能性を示す信頼値との生成を含み、

前記システムは、

完了した複数の音声認識タスクの一部に関して、1つまたは複数の認識結果と、前記1つまたは複数の認識結果と関連する1つまたは複数の信頼値とを受け取り、

50

前記音声認識タスクの完了した部分に関する 1 つまたは複数の信頼値のうちの各信頼値が閾値を満たすかどうか判定するとともに、

前記音声認識タスクの完了した部分に関する 1 つまたは複数の信頼値が閾値を満たすとの判定に応じて、前記複数の S R S のうち少なくとも 1 つの S R S に、未完了でかつ信頼値を生成していない音声認識タスクを中断させるように構成された信号を送る手段と、

前記 1 つまたは複数の認識結果に関連する 1 つまたは複数の信頼値に基づいて、選択された最終的な認識結果を送るためのインターフェースと、を具備し、前記最終的な認識結果は、前記受け取った音声信号内の予想される音声を表す、システム。

#### 【請求項 2 1】

10

前記最終的な認識結果が正しいとの指示を受け取る段階と、

前記指示の受け取りに応じて、未完了の S R S に関する前記音声認識タスクの残りの部分を中断する段階とをさらに含む、請求項 1 7 に記載の方法。

#### 【発明の詳細な説明】

##### 【技術分野】

##### 【0 0 0 1】

本明細書は、音声認識に関する。

##### 【背景技術】

##### 【0 0 0 2】

多くのアプリケーションが、ボイスコマンドまたはボイス照会の形の入力を受け取ることから恩恵を得る。これは特に、ユーザの指で操作される小型のキーパッドや他の装置を介して入力を入力することが、装置の縮小されたサイズのために難しいことのある、携帯電話などのモバイル装置上で動作中のアプリケーションに関して当てはまる。同様に、モバイル装置が別の装置上で動作中のアプリケーション(例えば、Eメールサーバ、マップ/方角サーバ、または電話帳サーバ)にアクセスするのに使用される場合、小型のキーパッドなどを介してコマンドを入力することなどの代わりに、アプリケーションにボイスコマンドを送ることは、ユーザにとってずっと容易であることがある。

20

##### 【0 0 0 3】

アプリケーションが音声入力に従って動作するために、音声認識器が、音声入力をアプリケーションで使用される記号表現に変換することができる。現在のいくつかの音声認識器は、音声入力内の予想される音声を識別しようと試みる単一認識システムを使用することがある。単一音声認識システムの使用により、音声識別の精度が、単一音声認識システムに関連する 精度に限定されることがある。別の現在の音声認識器は、連続音声認識を使用することがあり、連続音声認識では、音声入力に対して 2 つ以上のバスが実施され、どの音声が音声入力によって表される可能性が最も高いかが決定される。複数のバスの使用により、最終的音声認識結果を生成するのに必要な時間が増大することがある。

30

##### 【0 0 0 4】

さらに別の現在の音声認識器では、複数の音声認識システムがそれぞれ、音声入力を完全に処理し、次いで結果を出力することができる。この複数の音声認識システムの使用でも、処理時間の長さが、最も遅い音声認識システム(および/または音声認識アルゴリズムを実行する最も遅いコンピュータ)によって決定されるので、最終的結果を生成するのに必要な時間が増大することがある。

40

##### 【発明の概要】

##### 【課題を解決するための手段】

##### 【0 0 0 5】

一般には、本文書は、複数の音声認識システム(SRS)を並列に使用して音声を認識するが、生成した認識結果が所望の閾値を満たす場合、完了前に一部を停止することを説明する。例えば、各SRSは、音声認識タスクを実施する際に異なる待ち時間および精度を有することがある。待ち時間の短いSRSが、音声認識結果と、結果の高い信頼度を表す信頼値とを出力する場合、残りのSRSによって実行される音声認識タスクを停止することができ

50

る。信頼値が信頼閾値に比べて低過ぎる場合、別のSRSに結果を生成することを許可することができる。こうした結果が信頼閾値を満たす場合、音声認識タスクをまだ完了していないSRSを停止することができ、以下同様である。

#### 【0006】

第1の一般的な態様では、コンピュータで実施される方法が説明される。この方法は、音声信号を受け取り、複数の音声認識システム(SRS)による音声認識タスクを開始することを含む。各SRSは、音声信号に含まれると予想される音声を指定する認識結果と、音声結果の正確さの信頼度を示す信頼値とを生成するように構成される。この方法はまた、1つまたは複数の認識結果、および1つまたは複数の認識結果に関する1つまたは複数の信頼値を生成すること、1つまたは複数の信頼値が信頼閾値を満たすかどうかを判定すること、認識結果を生成していないSRSに関する音声認識タスクの残りの部分を停止すること、および生成した1つまたは複数の音声結果のうちの少なくとも1つに基づいて最終的な認識結果を出力することを含む音声認識タスクの一部を完了することも含む。10

#### 【0007】

第2の一般的な態様では、音声認識タスクを開始して、受け取った音声信号内に符号化された予想される音声を識別する複数の音声認識システムを含むシステムが説明される。各音声認識システム(SRS)は、認識結果と、認識結果の正確さの信頼度を示す信頼値とを生成するように構成される。このシステムはまた、認識結果がSRSによって生成されたときに認識結果を受け取り、生成された認識結果に関連する信頼値を受け取る認識管理モジュールをも含む。受け取った信頼値のうちの1つまたは複数が信頼閾値を満たす場合、認識管理モジュールは、認識結果を生成していないSRSによる未完了の音声認識タスクを停止する。このシステムは、生成された認識結果の信頼値に基づいて選択された最終的な認識結果を送るインターフェースを含む。20

#### 【0008】

第3の一般的な態様では、受け取った音声信号に関する音声認識タスクを開始する複数の音声認識システムを含むシステムが説明され、各音声認識システム(SRS)は、音声信号内の予想される音声を識別する認識結果と、認識結果の正確さの信頼度を示す信頼値とを生成するように構成される。このシステムは、1つまたは複数の認識結果がSRSによって生成されたときに、1つまたは複数の認識結果および1つまたは複数の対応する信頼値を各SRSから受け取り、受け取った信頼値のうちの1つまたは複数が信頼閾値を満たす場合、認識結果を生成していないSRSによる未完了の音声認識タスクを停止し、生成された認識結果の信頼値に基づいて最終的な認識結果を選択する手段を含む。このシステムはまた、音声信号内の予想される音声を表す最終的な認識結果を送るインターフェースをも含む。30

#### 【0009】

ここで説明されるシステムおよび技法は、以下の利点のうちの1つまたは複数を実現することができる。まず、複数の音声認識システムを使用して音声を並列に復号化するシステムは、満足の行く結果を受け取った場合に、終了していない認識タスクを停止することができる。さらに、複数の認識システムを使用するシステムは、拒否率を改善する(すなわち、拒否率を低減する)ことができる。このシステムはまた、複数の認識システムによって出力された認識結果を比較することによって精度を向上させることができる。認識性能の向上を実現するために使用される計算資源量のスケーリング(例えば向上)のためのフレームワークも提供することができる。40

#### 【0010】

1つまたは複数の実施形態の詳細が、添付の図面および以下の説明で述べられる。この説明および図面ならびに特許請求の範囲から、他の特徴および利点が明らかとなるであろう。

#### 【図面の簡単な説明】

#### 【0011】

【図1】音声を認識する例示的システムの図である。50

【図2】音声伝送に埋め込まれた音声を復号化する例示的システムのより詳細な図である。

【図3】並列復号化を使用して音声を認識する例示的方法のフローチャートである。

【図4A】例示的音声認識タスクの実行を示す図である。

【図4B】例示的音声認識タスクの実行を示す図である。

【図4C】例示的音声認識タスクの実行を示す図である。

【図5A】SRSによって生成される例示的認識結果および信頼値、ならびに最終的な認識結果を選択する方法の図である。

【図5B】SRSによって生成される例示的認識結果および信頼値、ならびに最終的な認識結果を選択する方法の図である。 10

【図5C】SRSによって生成される例示的認識結果および信頼値、ならびに最終的な認識結果を選択する方法の図である。

【図6】最終的な認識結果の選択で使用される値を重み付けするのに使用される信頼値の分布の例示的グラフである。

【図7A】認識結果を重み付けするのに使用することのできる、SRSによって出力される例示的認識結果集合と、集合間の相関とを示すベン図である。

【図7B】認識結果を重み付けするのに使用することのできる、SRSによって出力される例示的認識結果集合と、集合間の相関とを示すベン図である。

【図7C】認識結果を重み付けするのに使用することのできる、SRSによって出力される例示的認識結果集合と、集合間の相関とを示すベン図である。 20

【図7D】認識結果を重み付けするのに使用することのできる、SRSによって出力される例示的認識結果集合と、集合間の相関とを示すベン図である。

【図7E】認識結果を重み付けするのに使用することのできる、SRSによって出力される例示的認識結果集合と、集合間の相関とを示すベン図である。

【図8A】音声復号化システムの実行時動作中にSRS間の共通部分がどのように適応または変化することができるかを示すベン図である。

【図8B】音声復号化システムの実行時動作中にSRS間の共通部分がどのように適応または変化することができるかを示すベン図である。

【図9】SRS誤り率と認識結果に関連する重みとの間の例示的相関を示すグラフである。

【図10】本文書に記載のシステムおよび方法を実装するのに使用することのできるコンピューティング装置のブロック図である。 30

#### 【発明を実施するための形態】

##### 【0012】

様々な図面内の同様の参照記号は同様の要素を示す。

##### 【0013】

本文書は、複数の音声認識システム(SRS)を使用して発話を復号化するシステムおよび技法を説明する。ある実装では、各SRSは、精度、待ち時間、辞書などの異なる特性を有し、その結果、複数のSRSの一部が他のSRSの前に認識結果を出力する。出力認識結果が一定の要件を満たす(例えば、生成された結果のうちの1つまたは複数が、閾値信頼度を満たし、またはそれを超える指定の信頼値に関連付けられる)場合、音声復号化システムが、残りのSRSが音声認識タスクを完了する前に残りのSRSを停止することができる。 40

##### 【0014】

図1は、音声を認識する例示的システム100の図である。一般には、システム100は、例えば携帯電話から受け取った音声信号を処理する複数のSRSを含む。この例では、ユーザが音声対応電話帳サービスを呼び出し、音声対応電話帳サービスは、ユーザの音声を含む音声信号を複数のSRSを有する音声認識器に転送する。

##### 【0015】

複数のSRSは、音声信号を並列に処理することができるが、一部のSRSは、他のSRSの前に認識結果を生成することができる。認識結果を生成するSRSが、こうした結果で十分に高い信頼度を表す場合、残りの未完了の音声認識タスクを停止することができ、SRSのす 50

べてがその音声認識タスクを完了するのを待機するのではなく、現在生成される認識結果に基づいて最終的な認識結果を決定することができる。

#### 【0016】

例示的システム100は、音声入力を音声信号104の形で音声対応電話帳情報サーバ106に送る携帯電話102を含み、音声対応電話帳情報サーバ106は、携帯電話ユーザが電話帳情報を口頭で要求することを可能にし、要求された情報で応答する。

#### 【0017】

図1の例では、情報サーバ106は、音声信号104に埋め込まれた音声を復号化するために音声認識器108に音声信号104を送る。あるアプリケーションでは、音声認識器108は、並列に動作して音声信号104内の音声を復号化する複数のSRS<sub>A-E</sub>を含む。

10

#### 【0018】

音声認識システム(SRS)管理モジュール110は、SRSのいずれかが認識結果を生成したかどうかを監視し、そうした結果に関連する信頼値を収集する。監視は、SRSの並列実行を示すダイアグラム112で示されている。ダイアグラム112は、SRS<sub>A</sub>が最初に信頼値0.3で認識結果を生成することを示す。次に、SRS<sub>E</sub>が、信頼値0.6で認識結果を生成する。その後に、SRS<sub>B</sub>が、信頼値0.8で認識結果を生成する。SRS管理モジュール110は、認識結果がSRS<sub>B</sub>で生成された後、SRS<sub>C,D</sub>で実施される残りの音声認識タスクを停止することができる。この実装では、SRS<sub>B</sub>で生成された認識結果が、所定の信頼閾値を満たす信頼値を有するからである。

#### 【0019】

ある実装では、SRS管理モジュール110内の最終結果選択モジュール113が、最終的な認識結果114を選択して、音声対応電話帳情報サーバ106に出力することができる。例えば、SRS管理モジュール110は、音声認識タスクを完了したSRS<sub>A,B,E</sub>によって出力される、生成された1組の認識結果および関連する信頼値116に基づいて、最終的な認識結果を選択することができる。ある実装では、最終的な認識結果114は、音声信号104から復号化された推定音声の記号表現である。例えば、電話帳情報サーバ106が、都市名と州名を言うようにユーザに促した可能性がある。発話された都市名と州名は、音声信号104として符号化することができ、音声信号104が携帯電話102のユーザから受信され、音声認識器108で復号化される。ある実装では、最終的結果は、音声信号104によって表される可能性が最も高いとデコーダが判定した音声である。

20

#### 【0020】

SRSによって出力される信頼値116は、いくつかの方法に従って生成することができる。ある実装では、第1SRSは、発話の識別に関して、いくつかの仮説、または推測を生成することができる。生成された最上の仮説を、SRSで生成された競合する仮説と比較することができ、信頼値は、生成された最上の仮説と、競合する仮説との間の認識スコアの差に基づいて変化することができる。

30

#### 【0021】

さらに別の例では、第1認識結果に関する信頼値は、認識結果の生成、またはフロントエンド信号解析の計算で使用される信号または特性に基づくことができる。例えば、信号は、探索で使用されるいくつかの競合する仮説、または調査される探索グラフの密度を含むことができ、フロントエンドの例は、ノイズ特性に対する推定信号、または推定チャネルタイプ(例えば、ハンズフリーとセルラと陸線)の既存のモデルに対する合致に基づくこうしたチャネルタイプを含むことができる。こうした信号の組合せは、提供されたデータに基づいて条件付きで最適化することができる。

40

#### 【0022】

信頼度は、音響モデルおよび言語モデルからのスコアの非線形結合として推定することができる。例えば、最上の仮説が与えられると、システムは、別々の言語モデルスコア(例えば、音声が処理される前の、認識結果に関する事前推定)、および音響モデルスコア(例えば、この発話が、最上の結果に関連する音響単位とどれほど良く合致したか)を抽出することができる。次いで、システムは、提供されたデータにわたって条件付きで最適化

50

されたこうした2つのスコアの非線形結合として合計信頼度結果を推定することができる。

**【 0 0 2 3 】**

スコアに加えて、信頼度を最適するのに使用することのできる別の信号が、言語モデルを介して伝えられる経路のタイプに基づく。例えば、nグラム言語モデルでは、システムが特定の3ワードシーケンスに遭遇しなかったとき、システムは「バックオフ」することができ、すなわちシステムが遭遇した2ワードシーケンスから3ワードシーケンスに関するプライヤ(prior)を推定することができる。探索が所与の仮説に関するバックオフ推定を経なければならない回数をカウントすることにより、所与の発話に関する信頼度を条件付きで推定するときに使用することのできる別の信号が与えられる。

10

**【 0 0 2 4 】**

さらに別の実装では、信頼値は、認識結果が正しいという事後確率でよい。ある実装では、事後確率は、格子密度計算に基づいて計算することができる。別の例では、事後確率は、モノフォンループ、主音響モデルよりも少ないパラメータでトレーニングされたオールスピーチ(all-speech)gmmなどの音声の具体性の低い音響モデルを使用する、すべての音声のより汎用的なモデルと、最上の仮説を比較することによって求めることができる。こうした信頼度に関する事後推定の方法のどちらも当技術分野で周知であり、こうした推定の組合せも同様である。

**【 0 0 2 5 】**

ある実装では、信頼値が、複数の技法を使用して計算される。例えば、信頼値は、事後確率、ならびに音声認識タスク中に生成された他の認識結果と比べた結果の類似性に基づく。

20

**【 0 0 2 6 】**

選択モジュール113は、最終的な認識結果114をインターフェース118に送ることができ、インターフェース118は、最終的な認識結果114を音声対応電話帳情報サーバ106に送ることができる。ある実装では、インターフェース118は、情報サーバ106で実行されるソフトウェアとインターフェースする1組のAPIを使用する。例えば、情報サーバ106は、都市、州、および商号のテキスト表現を入力するための共通の方法を有するソフトウェアを実行することができる。別の実装では、インターフェース118は、ネットワークを介して情報サーバ106に情報を送信するネットワーキングプロトコル(例えばTCP/IP)を含むことができる。

30

**【 0 0 2 7 】**

図1は、別々のコンピューティング装置上の電話帳情報サーバ106および音声認識器を示すが、これは必要であるわけではない。ある実装では、どちらのシステムも単一のコンピューティング装置上に実装することができる。同様に、各システムは、いくつかのコンピューティング装置を使用して実装することができる。例えば、各SRSは、図2に示されるように1つまたは複数のコンピュータを使用して実装することができる。

**【 0 0 2 8 】**

図2は、音声伝送に埋め込まれた音声を復号化する例示的システム200の図である。例示のために、システム200が、音声伝送セグメント202および音声認識器セグメント204という2つのセグメントに分割される。音声伝送セグメント202は、ユーザからアプリケーションサーバに音声信号を送信する例示的アーキテクチャを示す。音声認識器セグメント204は、音声伝送で表される音声を解釈または復号化する例示的アーキテクチャを示す。この実装では、復号化がアプリケーションサーバの代わりに実施され、復号化音声が、ユーザの要求を処理する際に使用するためにアプリケーションサーバに送り戻される。

40

**【 0 0 2 9 】**

ある実装では、システム200、音声伝送セグメント202は、携帯電話206などの電話装置を含み、電話装置は、ネットワーク(例えば、POTS、セルラ、インターネットなど)を介して音声信号208を電話サーバ210に送信する。電話サーバは、音声信号を、ソフトウェアアプリケーションサーバ212などの別のコンピューティング装置に送信することができ、ま

50

たは後で説明する音声認識システムに直接送信することができる。

**【0030】**

ソフトウェアアプリケーションサーバ212は、ユーザが口頭で対話しているソフトウェアアプリケーションを含むことができる。例えば、ソフトウェアアプリケーションサーバはカレンダーアプリケーションでよい。ユーザは、カレンダーアプリケーションを呼び出し、カレンダーアプリケーションが2012年5月16日の午後1:30から午後2:00までのイベントを作成することを要求することができる。ソフトウェアアプリケーションサーバ212は、受け取った、イベント作成を要求する音声入力を復号化のために音声認識器セグメント204に転送することができる。

**【0031】**

ある実装では、音声認識器セグメント204は、音声認識システム、SRS<sub>A-D</sub>、および音声認識システム(SRS)管理モジュールを含み、SRS管理モジュールは、音声信号208によって表される可能性が最も高いのはどの発話かを決定する際に使用するためにSRSを調整することができる。

**【0032】**

各SRSは、1つまたは複数の方式の点で異なる可能性がある。ある実装では、SRSは、その基礎となる音響モデルによって異なることがある。例えば、異なる音響モデルは、特定の条件、例えばユーザの性別、アクセント、年齢の範囲、または特定の背景雑音条件および前景雑音条件(foreground noise condition)、または特定の伝送チャネルを目標とすることができる。音響モデルは、そのアーキテクチャおよびサイズによって異なることもあります、例えばパラメータの少ない小型のモデルは、より高速な認識を生成することができ、パラメータの多い大型のモデルは、より正確な結果を生成することができる。他の例では、音響モデルは、そのトレーニング手順によって異なることがある(例えば、異なる確率化トレーニングセットを使用してモデルをトレーニングすることができ、または異なるトレーニングアルゴリズムを使用することができる)。

**【0033】**

ある実装では、SRSはその言語モデルによって異なることがある。例えば、モデルは、異なるタイプのデータ、例えば異なる領域特有の言語モデル、異なる細粒度、または異なる地理的位置を目標とすることができます。別の例では、モデルは、そのアーキテクチャ、サイズ、トレーニング手順などによって異なることがある。

**【0034】**

さらに別の実装では、SRSは、エンドポインタ、フロントエンド、辞書、信頼度推定アルゴリズム、探索構成などの別の構成要素によって異なる可能性がある。

**【0035】**

例示のために、SRS<sub>D</sub>に関する言語モデル252、音響モデル254、および音声認識アルゴリズム256が図2に示されている。

**【0036】**

ある実装では、SRS管理モジュール250が音声信号208を受信したとき、SRS管理モジュール250はプロセスを開始し、そのプロセスによって、着信発話がSRSのうちの2つ以上を使用して並列に認識される。例えば、4つの音声認識タスクが、音声信号208で表される同じ発話を認識しようと試みて、4つのSRS(SRS<sub>A-D</sub>)によって並列に実行される。

**【0037】**

ある実装では、各SRSは特定の待ち時間を有することがある。待ち時間は、SRSアーキテクチャ(例えば、音響モデル、言語モデル、または他の構成要素)に依存することがあるが、音声認識タスクの特定の具体化に基づいて変化することもある。例えば、発話が一定のグループのワード(例えば、yes、no、nope、yeah、肯定的、否定的、no way、yipperなど)内に含まれることを示す情報をSRSが有する場合、特定のモデルに関する待ち時間は、発話が発せられた制限された文脈を示す情報をSRSが有さないとき(例えば、発話が一般疑問文の文脈ではなかった)よりもずっと短いことがある。

**【0038】**

10

20

30

40

50

ある実装では、各SRSは、その音声認識タスクの完了時に、認識結果(例えば、着信発話が何を言ったかをSRSが決定したもの)と、SRSがその結果の正確さにどれほど自信があるかの尺度とを出力する。

#### 【 0 0 3 9 】

ある実装では、SRS管理モジュール250は、SRSによって生成される認識結果262を追跡する認識結果モニタ258を有する。結果モニタ258はまた、認識結果262に関連する信頼値264、またはスコアも監視することができる。

#### 【 0 0 4 0 】

ある実装では、結果モニタ258は、各認識結果に関する組合せ信頼度スコアを計算することができる。例えば、 $SRS_A$ および $SRS_B$ は、着信発話に関する認識結果「5月16日」を生成することができる。 $SRS_A$ は、その結果を信頼値0.8と関連付けることがあり、 $SRS_B$ は、その結果を信頼値0.7と関連付けることがある。結果モニタ258は、結果「5月16日」に関する現在の移動平均(running average)が0.75(すなわち、 $(0.8+0.7)/2$ )であることを計算することができる。認識結果(および対応する信頼値)がSRSによって生成されるごとに、組合せ信頼値を更新することができる。10

#### 【 0 0 4 1 】

SRS管理モジュール250はまた、各SRSについての待ち時間(例えば、音声認識タスクの完了までの実際の時間または推定時間)を追跡する待ち時間モニタ260をも含むことができる。例えば、待ち時間モニタ260は、SRSが認識結果を生成するのに特定の音声認識タスクがどれほどの時間を要するかを追跡することができる。20

#### 【 0 0 4 2 】

待ち時間モニタ260はまた、(例えば、SRSが認識結果を出力したかどうかを監視することによって)SRSが音声認識タスクを完了したか否かを監視することもできる。さらに、待ち時間モニタ260は、SRSが音声認識タスクを完了するための予測待ち時間を推定することができる。例えば、待ち時間モニタ260は、SRSが類似の音声認識タスクを完了するのにどれほどの時間がかったか(例えば、促される特定の質問に対する回答などの類似の文脈で話された発話をSRSが復号化するのに以前にどれほどの時間がかったか)を示す、SRSに関して収集された経験的情報にアクセスすることができる。

#### 【 0 0 4 3 】

待ち時間モニタ260はまた、SRSが音声認識タスクを完了するための推定時間を求めるのに、基礎となるモデルの特性についての情報にもアクセスすることができる(例えば、待ち時間モニタ260は、認識結果を識別するために探索しなければならないワードの辞書が大きいために、SRSが音声認識を完了するのにかかる時間が長くなる可能性があるかどうかを判定することができる)。30

#### 【 0 0 4 4 】

SRS管理モジュールは、待ち時間モニタ260および/または認識結果モニタ258と通信して、受信した音声信号208の復号化をまだ完了していないSRSについて停止コマンド266を送るかどうかを判定するSRSアボータ(aborter)270を含むことができる。例えば、SRSアボータ270は、追跡される信頼値および/または待ち時間が動作点または動作曲線を満たすかどうかを判定することができる。そうである場合、残りのすべての音声認識を停止することができる。40

#### 【 0 0 4 5 】

単純な例では、SRS管理モジュール250は、任意の未完了音声認識タスクを停止する前に、認識結果に関する信頼閾値0.75に達するべきであると決定することができる。ある実装では、信頼閾値は、異なる音声認識タスクについて変化することができる。例えば、特定の音声認識タスクに関連する文脈情報が、認識結果が比較的少数の発話を限定される(例えば、認識タスクが、ユーザに提示された一般疑問文に対する回答の文脈である)ことを示す場合、SRSアボータ270は、信頼値が比較的高い(例えば、0.8、0.9、0.95)はずであると判定することができる。

#### 【 0 0 4 6 】

50

認識結果が多数の予想される発話のうちのいずれか1つを含む可能性があることを文脈が示す場合(例えば、ユーザが「今日何をしたいですか」などの自由回答式の質問を尋ねられる)、SRSアボータ270は、認識結果に関する信頼閾値が比較的低い可能性があり(例えば、0.49、0.55、0.61)、それでもなおSRS管理モジュール250にとって受け入れられると判定することができる。

#### 【0047】

ある実装では、しきい信頼点(例えば0.75)が認識結果のうちの1つによって満たされる場合、アボータ270は、終了していない任意のSRSに停止コマンド266を送ることができる。例えば、 $SRS_{A,B}$ の組合せ信頼値が0.75以上である場合、アボータ270は、認識結果をまだ生成していないSRSに停止コマンド266を送ることができる。

10

#### 【0048】

別の例では、認識結果のうちの1つが0.75以上の信頼値を有する場合、アボータ270は停止コマンド266を送ることができる。この場合、信頼値は、組合せ信頼値ではないことがあり、その代わりに、単一のSRSによって生成された認識結果に関連する単一信頼値でよい。

#### 【0049】

別の実装では、SRSアボータ270は、SRSに関する実際の待ち時間または推定待ち時間に基づいて停止コマンドを送ることができる。例えば、 $SRS_A$ および $SRS_B$ が認識結果を生成し、認識結果が非常に低い信頼値に関連付けられる場合、SRSアボータ270は、残りの終了していない音声認識タスクを停止することができる。この場合、他のSRSが高信頼値を有する認識結果を生成していないので、認識タスクを完了していないSRSがそのような認識結果を生成しないという仮定の下で、残りの音声認識タスクを取り消すことができる。

20

#### 【0050】

残りのSRSが終了するのを待機する代わりに、ある実装では、SRSアボータ270は、停止コマンド266を送り、ユーザが発話274を繰り返すように要求を受けるプロセスを開始することができる。次いで、SRSは、新しい発話を復号化することを試みることができる。

#### 【0051】

別の実装では、認識結果が信頼値または他のパラメータに基づいて満足の行くものではない場合、SRS管理モジュール250は、人間のオペレータに切り換えるように要求が行われるプロセスを開始することができる。例えば、音声信号をコールセンタ278に送ることを示す矢印276で示されるように、人間のオペレータは、ユーザに応答して音声信号を傍受することができる。人間のオペレータは、ユーザによって伝達される要求または情報に対処することができる。

30

#### 【0052】

ある実装では、SRSアボータ270はまず、待ち時間モニタ260に照会して、終了していないSRSに停止コマンド266を送るかどうかを決定することができる。例えば、SRSアボータ270が待ち時間モニタ260に照会して、終了していないSRSのうちの1つが比較的短い時間量で完了する可能性が高いと判定した場合、SRSアボータ270は延期して、ほぼ完了しているSRSから結果を得ることができる。結果が生成された後、SRSアボータ270は停止コマンド266を送り、残りの終了していないSRSが音声認識を完了するのを停止することができる。

40

#### 【0053】

ある実装では、追加の認識結果および関連情報が、ほぼ終了しているSRSが完了するまで停止コマンドを送ることを遅延するのに足りるほど価値があることがある。例えば、ほぼ終了しているSRSに関連する特性が、その認識結果が以前に完了したSRSの結果よりもしばしば正確であることを示す場合、アボータ270は、残りの終了していない音声認識タスクを停止する前に、ほぼ終了しているSRSが認識結果を生成するまで待機することができる。

#### 【0054】

ある実装では、1つまたは複数の変数を有する関数を使用して信頼閾値が求められる。例えば、信頼度関数は、信頼値および待ち時間を含む変数を有することができます。生成さ

50

れた信頼値および観測された待ち時間が信頼度関数を満たす場合、アボータ270は、終了していない任意の音声認識タスクを取り消すことができる。例えば、短い待ち時間内で、信頼度関数は、信頼度関数が満たされるために信頼値が非常に高くなるべきであることを指示することがある。このことは、アボータ270が停止コマンドを迅速に発行する場合、他の潜在的認識結果が生成されないので、生成される結果が正しい確率が非常に高いはずであるという仮定に部分的に基づくことができる。処理するのに長くかかる音声認識タスクはより難しい可能性があり、したがって結果の信頼度が低くなる可能性が高いという仮定の下で、待ち時間が増大するにつれて、必要な信頼値は低下することがある。

#### 【0055】

SRS管理モジュール250はまた、最終結果選択モジュール280をも含むことができ、ある実装では、最終結果選択モジュール280は、生成された認識結果から最終的な認識結果を選択する。例えば、選択モジュール280は、最高の組合せ信頼値を有する認識結果、または最高の単一信頼値を有する認識結果を選択することができる。

10

#### 【0056】

ある実装では、最終的な認識結果の選択は、どのSRSが認識結果を生成したかに基づいて影響を受ける可能性がある。例えば、比較的異なる構成要素(例えば、言語モデル、音響モデル、音声認識アルゴリズムなど)を有するSRSおよび/または異なる認識結果を通常は生成するSRSによって認識結果が生成される場合、選択モジュール280は、認識結果の選択を重み付けし、または支持することができる。

#### 【0057】

ある実装では、SRS関連モニタ282が、SRSに関する出力認識結果間の相関を追跡することができる。2つのSRSの相関が高くないことを出力認識結果が示すが、特定の認識タスクではその2つのSRSが同一の認識結果を生成する場合、最終的な認識結果の選択で、その結果をより強く重み付けし、または支持することができる。あるいは、認識結果を生成するSRSの相関が高い場合、最終的な認識結果の選択で結果が必ずしも支持されないように、認識結果を割り引き、または重み付けしないことができる。

20

#### 【0058】

相関モニタ282は、SRSで生成された認識結果に基づいて、さらに2つのSRS間の相関を指定する相関値を動的に更新することができる。例えば、低い相関値に関連する2つのSRSは、類似の認識結果の生成を開始することができる。相関モニタ282は、相関値を更新して(例えば増加させて)、SRS間の認識結果の重複の増大を反映させることができる。

30

#### 【0059】

最終結果選択モジュール280が最終的結果を識別した後、SRS管理モデルは、音声信号の復号化を要求したソフトウェアアプリケーションサーバに結果を送り戻すことができる。ソフトウェアアプリケーションサーバは、復号化音声信号を使用して、ユーザの要求を処理することができる。

#### 【0060】

図3は、並列復号化を使用して音声を認識する例示的方法300のフローチャートである。方法300は、例えば、システム100および200などのシステムで実施することができ、提示がわかりやすいように、以下の説明では、この方法を説明する一例に関する基礎としてシステム100および200を用いる。しかし、別のシステムまたはシステムの組合せを使用して方法300を実施することもできる。

40

#### 【0061】

ステップ302では、音声信号が受け取られる。例えば、音声認識器108が音声信号104を受け取ることができる。ある実装では、音声認識器108に送る前に、処理のために音声信号104が既にサンプリングされ、デジタルフレームにセグメント化されている。別の実装では、音声認識器108がこうした機能を実施することもできる。

#### 【0062】

ある実装では、音声信号を前処理して、信号のどの部分が音声を含み、どの部分が雑音であると判定されるかを識別することができる。受け取られる音声信号104は、音声を有

50

すると判定される部分のみを含むことができ、次いでその部分を音声認識器108によって以下のステップで復号化することができる。

#### 【0063】

ステップ304A～Nでは、音声認識タスク(SRT)が開始される。ある実装では、SRTがほぼ同時に開始され、音声信号104で表される音声の復号化が開始される。図2のSRS<sub>A-D</sub>などのSRSは、音声信号を処理する際に異なる待ち時間を有することがあり、その結果、SRTは、完了するのに必要な時間量が異なることがある。

#### 【0064】

ステップ306では、SRTの進行が監視される。例えば、待ち時間モニタ260が、各SRSに関連する待ち時間(実際の待ち時間と推定待ち時間の両方)を追跡することができる。

10

#### 【0065】

ステップ308では、SRT<sub>1-N</sub>のうちのいずれかが認識結果を生成したかどうかが判定される。例えば、SRSは、認識結果(または結果が存在するという指示)が生成された後、それを認識結果モニタ258に出力することができる。SRSのいずれも認識結果を生成していない場合、方法300はステップ306に戻り、SRTの進行を引き続き監視することができる。SRSが1つまたは複数の認識結果を生成した場合、方法はステップ310に進むことができる。

#### 【0066】

ステップ310では、生成された認識結果に関する何らかの信頼値が信頼閾値を満たすかどうかの判定が行われる。例えば、SRSアボータ270は、前述のように、認識結果に関する信頼値(単一信頼値または組合せ信頼値)を信頼点または信頼度関数と比較することができる。現信頼値が信頼閾値を満たさない場合、方法300はステップ306に戻ることができ、SRTの進行が監視される。信頼閾値が満たされた場合、方法300はステップ312に進むことができる。

20

#### 【0067】

ステップ312では、終了していないSRTが停止される。例えば、並列に実行中の10個のSRTがあり、4つが完了した場合、残りの6つを取り消し、または停止することができる。ある実装では、SRSアボータ270は、適切なSRSが音声認識タスクを中止するように、そのSRSに停止コマンド266を送ることができる。

#### 【0068】

ある実装では、音声認識タスクのうちの1つまたは複数が停止されず、単に「休止」される(例えば、処理タスクの状態を保管し、後で再始動することができる)。例えば、認識結果が不正確であることがわかった(例えば、ソフトウェアアプリケーションサーバが、音声が正しく復号化されたことを確認するようにユーザに促したときに、ユーザが否定表現で応答した)場合、「休止された」音声認識タスクを再始動することができる。

30

#### 【0069】

ある実装では、例えば、SRTを実行中のSRSの精度に基づいて、SRTを選択的に休止することができる。例えば、認識結果が、信頼閾値をかろうじて満たすだけの信頼値に関連付けられる場合、アボータ270は、より正確なSRSのSRTを選択的に休止すると共に、残りのSRTを停止することができる。認識結果が不正確であることがわかった場合、より正確なSRSの休止したSRTを再始動することができる。

40

#### 【0070】

ある実装では、以前に完了したSRT、および以前に停止されたSRTは、「休止解除された」SRTと同時に開始することができる。これにより、SRTが完全に再始動された場合よりも、より正確なSRTに、完了するための時間をより多く与えることができる。さらに別の実装では、ユーザの不正確な認識の確認に基づいて推論または決定された情報は、休止解除されたSRTならびに再始動されたタスクと一体化することができる。例えば、音声復号化の新しいラウンドでは、誤りのある発話を考慮から取り除くことができる。さらに、認識処理の第2ラウンドでは、誤りのある結果を求めるのに使用されるいくつかの音声、ワードなどを割り引くことができ、または考慮から除外することができる。

#### 【0071】

50

ステップ314では、最終的な認識結果が、生成された結果に基づいて選択される。例えば、最終結果選択モジュール280は、最高の平均信頼度スコアに関連付けられる認識結果を識別することができる。ある実装では、選択は、結果を生成するSRSの精度に基づいて重み付けすることもでき、通常は正確なSRSからの結果が、正確性が低いSRSに優先して支持される。さらに別の実装では、選択は、結果または結果に関連する出現頻度を生成するマシン間の相関に基づくこともできる。選択された結果は、音声信号の復号化を要求したアプリケーションに出力することができる。次に、方法は終了することができる。

#### 【0072】

図4A～Cは、例示的音声認識タスクの実行を示す図を示す。図4Aは、4つのSRSによる4つのSRTの実行を示す。図示される実装では、SRTが並列に開始され、 $SRS_A$ が認識結果を最初に生成する。 $SRS_A$ は、その認識結果について信頼値0.7を求める。ある実装では、SRS管理モジュール110が、信頼値を信頼閾値と比較することができる。信頼値が閾値を満たさない場合、残りのタスクが実行を許可される。例えば、信頼閾値が固定定数0.9であった場合、初期認識結果0.7は閾値を満たさないので、SRS管理モジュールは、残りのSRSが続行することを許可する。

#### 【0073】

次に、 $SRS_B$ が認識結果および関連する値0.85を生成する。この信頼値も0.9の信頼閾値を満たさず、したがって残りのタスクが続行を許可される。

#### 【0074】

さらに、SRS管理システムは、各SRSに関連する待ち時間を追跡することもでき、許可された待ち時間閾値とこうした待ち時間を比較することができる。図4Aに示されるように、ある実装では、SRS(例えば、 $SRS_C$ および $SRS_D$ )が待ち時間閾値の前に認識結果を生成しない場合、SRS管理モジュール110は、SRSに停止コマンドを送ることができる。

#### 【0075】

ある実装では、信頼閾値を満たす認識結果が生成される前にSRTが停止される場合、SRS管理モジュール110は、信頼閾値を満たさなかった場合であっても、最高の信頼値を有する結果を選択することができる。ある実装では、次に高い信頼値が、選択されるためには、求めた信頼閾値の範囲(例えば10%)以内になければならないことがある。さらに別の実装では、認識結果が選択されない場合、SRS管理モジュール110は、音声入力を繰り返す要求を送ることができる。

#### 【0076】

図4Bは、SRSが信頼閾値を満たす信頼値を有する認識結果を生成した後に、終了していないSRTを停止することを示す図である。この例では、信頼閾値は0.9である。 $SRS_A$ が最初に認識結果を生成するが、 $SRS_A$ は、信頼閾値よりも低い信頼値0.7を結果に割り当てる。したがって、SRS管理モジュール110は、 $SRS_{B-D}$ が実行を実行することを許可する。

#### 【0077】

$SRS_B$ が次に認識結果を生成し、それに信頼値0.9を割り当てる。SRS管理モジュール110は、この信頼値を信頼閾値と比較し、閾値が満たされると判定する。次いで、SRS管理モジュールは、 $SRS_C$ および $SRS_D$ に停止コマンドを送出することができ、 $SRS_C$ および $SRS_D$ は、認識結果を生成することなくそれぞれのSRTを停止する。

#### 【0078】

図4Cは、生成された認識結果の信頼値が低いことに基づいて、終了していないSRTを停止することを示す図である。この例では、信頼閾値を固定小数点0.9に設定することができる。 $SRS_A$ および $SRS_B$ が認識結果を生成するが、どちらの結果も、それぞれ比較的低い信頼値0.3および0.25に関連付けられる。信頼値が共に比較的低いことを考えると、SRS管理モジュール110は、前のSRSが信頼閾値よりも著しく低い信頼値を有する認識結果を生成したので、 $SRS_C$ および $SRS_D$ が信頼閾値を満たす信頼値を有する認識結果を生成する可能性が低いという仮定の下で、こうしたSRSに停止しコマンドを送ることができる。

#### 【0079】

図4Cに示されるある実装では、SRS管理モジュール110は、以前に生成された認識結果の

10

20

30

40

50

信頼値が低いことに基づいて、停止コマンドを送る前に、求めた時間量の間、待機することができる。ある実装では、SRS管理モジュール110は、最後の認識結果が生成されたときに基づいて時間枠を開始する。求めた時間枠は、別のSRSがそのSRTを完了することを可能にすることができますが、許可された時間枠中に結果が生成されない場合、終了していない任意のSRTを停止するコマンドを送ることができる。

#### 【0080】

ある実装では、待機する時間枠の決定は、認識結果を生成していないSRSのうちの1つまたは複数の推定待ち時間に基づくことができる。例えば、SRS管理モジュール110は、SRS<sub>C</sub>が残りのSRSのうちの最短の推定待ち時間を有すると判定することができる。例えば、SRS<sub>C</sub>は、典型的待ち時間0.5秒を有することがある。SRS<sub>B</sub>が0.4秒後に認識結果を生成する場合、音声認識管理モジュール100は、0.1秒遅延して、停止コマンドを送る前にSRS<sub>C</sub>が認識結果を生成するかどうかを判定することができる。10

#### 【0081】

別の実装では、停止コマンドを直ちに送ることができる。例えば、SRS管理モジュール110は、低信頼値にやはり関連付けられる、求めた数のSRSが認識結果を生成した後に、停止コマンドを送ることができます。図4Cに示されるケースでは、SRSの半分が低信頼値に関連付けられる認識結果を返すとすぐに、停止コマンドが送られる。

#### 【0082】

ある実装では、信頼値が低い場合、システムが複合信頼値(例えば合計/累積信頼値)がある閾値よりも上であることを確認するまで、システムはより多くの認識結果を引き続き受け取る。ある認識タスクでは、確認は決して行われず、システムは発話を拒否することによって認識プロセスを終了させることができる。したがって、ある実装では、3つのタイプの信頼度があり、第1は、各認識プロセスからの元の信頼度、第2は、各認識プロセスからの元の信頼度から求められる累積合計信頼度、第3は、システムがより多くの認識イベントを待機するにつれて合計信頼度が変化する(例えば増加する)という予想である。20

#### 【0083】

ある場合には、システムは、相関解除された認識器にわたって十分な数の一貫して低い信頼度の結果を受け取り、すべての認識タスクを停止して発話を拒否するように促す。拒否が生じる場合、システムは、ユーザに発話を繰り返すように促すことができる。拒否のケースは、例えば、個々の元の信頼値が一貫して低く、累積合計信頼度が低く、より多くの認識と共に合計信頼度が変化する可能性があるという予想も低いときに生じことがある。30

#### 【0084】

ある実装では、特定の1組の信頼値が与えられたときの推定予想信頼度変化に関するトレーニングが、部分的認識信頼度のトレーニング例が与えられたときの最終認識信頼度の分布をカウントすることによって前もって実施される(例えば、最初の20個の認識器から0.1未満の20個の信頼値を確認した後、システムは、20個を越える認識器を用いる組合せ信頼度によって合計信頼度が0.5より上に増加する例には決して直面しない。したがって、システムは、この状況が生じるときに発話を拒否するようにトレーニングされる)。40

#### 【0085】

ある実装では、最終的な認識結果に関連する組合せ信頼度は、個々のSRSからの個々の信頼値の関数でよい。互いとも合致する、多くの認識器からの高い信頼値を伴う結果には、高い組合せ信頼度が与えられる。各認識器の個々の寄与の重み付けは、トレーニングプロセス中のテストデータの認識の実験的最適化に基づくことができる。

#### 【0086】

図5A～Cは、SRSによって生成される例示的認識結果および信頼値、ならびに最終的な認識結果を選択する様々な方法の図である。具体的には、図5A～Cは、SRS<sub>A</sub>からのSRS<sub>A</sub>出力502、SRS<sub>B</sub>からのSRS<sub>B</sub>出力504、およびSRS<sub>C</sub>からのSRS<sub>C</sub>出力506を示す。この例では、各SRSが「carry」というワードを表す音声信号を復号化しようと試みることに応答して、出力が生成される。各SRSは異なることがあるので、各SRSによって生成される認識結果は、図50

5A～Cで示されるように異なることがある。

#### 【0087】

ある実装では、SRS出力は、上位N個の認識結果を含み(ただしNは、任意の正の整数または0を表すことができる)、N個の認識結果は、どの認識結果が最高の信頼値に関連するかに基づいて選択される。例えば、SRS<sub>A</sub>出力502は、SRS<sub>A</sub>に関する上位4つの認識結果および関連する信頼値Result=carry、Confidence=0.75;Result=Cory、Confidence=0.72;Result=query、Confidence=0.6;およびResult=hoary、Confidence=0.25を含む。

#### 【0088】

SRS<sub>B</sub>出力504は、Result=quarry、Confidence=0.64;Result=Cory、Confidence=0.59;Result=hoary、Confidence=0.4;およびResult=Terry、Confidence=0.39を含む。

10

#### 【0089】

SRS<sub>C</sub>出力506は、Result=tarry、Confidence=0.58;Result=Terry、Confidence=0.57;Result=Cory、Confidence=0.55;およびResult=carry、Confidence=0.2を含む。

#### 【0090】

図5Aは、最高の信頼値に関連する認識結果を選択する例示的選択アルゴリズムを示す。例えば、最終結果選択モジュール113は、認識結果のすべてを比較して、最高の信頼値に関連する認識結果を選択することができる。この例では、結果「carry」がすべての信頼値のうちの最高の信頼値0.75に関連付けられるので、それが最終的な認識結果として選択される。次いで、選択モジュールは、音声復号化を要求したアプリケーションでさらに処理するために、認識結果「carry」を出力することができる。

20

#### 【0091】

図5Bは、どの結果が最高の組合せ信頼値を有するかに基づいて認識結果を選択する例示的選択アルゴリズムを示す。例えば、複数のSRSが、同一の認識結果を生成することがあるが、結果に対して異なる信頼値を割り当てることがある。ある実装では、同一の結果に関する複数の信頼度スコアを平均化して(あるいは組み合わせて)、組合せ信頼度スコアを生成することができる。例えば、認識結果「carry」がSRS<sub>A</sub>とSRS<sub>C</sub>の両方で生成されるが、SRS<sub>A</sub>は結果に0.75信頼値を割り当て、SRS<sub>C</sub>は結果に0.2信頼値を割り当てる。これらの信頼値の平均は0.475である。

#### 【0092】

同様に、認識結果「Cory」に関する平均組合せ信頼度スコアは0.61であり、「quarry」に関する組合せ信頼度スコアは0.62である。この例では、選択モジュール113は、「quarry」の組合せ信頼値が他の結果の組合せ信頼値よりも高いので、「quarry」を最終的な認識結果として選択することができる。選択が認識結果の同じプールから行われたにも関わらず、この選択アルゴリズムは、図5Aに記載のアルゴリズムとは異なる最終的結果を生成することに留意されたい。

30

#### 【0093】

図5Cは、認識結果の選択で重み因子を考慮に入れる例示的選択アルゴリズムを示す。ある実装では、重みは、認識結果の出現頻度に基づくことができる。例えば、表550は、新しい重み付け信頼度スコアを生成するために、先に議論した組合せ信頼度スコアに乗じることのできる3つの重みを列挙する。

40

#### 【0094】

この例では、認識結果が単一のSRSで生成される場合(例えば、結果が頻度「1」で生じる場合)、組合せ信頼度スコアに重み「1」を乗じることができる。したがって、認識結果が1度だけ生じる場合、認識結果は、重み付けから恩恵を受けることはない。認識結果が2度生じる場合、1度だけ生じる別の認識結果より幾分優先してこの認識結果を支持する因子1.02を用いて、認識結果を重み付けすることができる。認識結果が3度生じる場合、認識結果を因子1.04で重み付けすることができる。

#### 【0095】

図5Cの例では、認識結果「Cory」に関する組合せ信頼値が因子1.04に対して重み付けされ、その結果、重み付け値0.6344が得られる。認識結果「quarry」に関する組合せ信頼値

50

が因子1.02に対して重み付けされ、その結果、重み付け値0.6324が得られる。この場合、結果「Cory」の非重み付け組合せ信頼度スコアが結果「quarry」の非重み付け組合せ信頼度スコアより低くても、結果「Cory」の重み付け組合せ信頼度スコアが結果「quarry」よりも高いので、選択モジュール113は、結果「quarry」に優先して結果「Cory」を選択することができる。

#### 【0096】

最終的な認識結果を選択するのに使用される値は、限定はしないが、SRSで生成された信頼度スコアの分布、認識結果を生成したSRSの特性(例えば、全体的精度、特定の文脈での精度、定義済みの時間枠にわたる精度など)、ならびに同一の認識結果を生成するSRS間の類似性を含むいくつかの基準に基づいて重み付けすることができる。

10

#### 【0097】

別の実装では、認識信頼値と、認識器および最終的複合認識器に関する認識誤りとの相関を使用して、最終的な認識結果を重み付けすることができる。例えば、トレーニング中、システムは、特定の認識器が信頼値0.3で応答する回数をカウントすることができ、こうした「0.3信頼度認識結果」がその認識器に関する誤りである頻度はどれほどであるか、および最終的組合せ認識も認識誤りである頻度はどれほどであるかもカウントすることができる。システムは、類似の認識結果を組み合わせるとき、同一の正規化カウンティングを使用することができる。組合せ信頼度は、認識器が同一の結果を得た回数(所与の信頼値を有する)、および共通の結果が正しかった回数から推定することができる。

#### 【0098】

20

図6は、最終的な認識結果の選択で使用される値を重み付けするのに使用される信頼値の分布の例示的グラフ600である。グラフのy軸は、正規化目盛(0.0から1.0)に沿って、特定の信頼値がどこに位置するかを示す。グラフのx軸は、どの特定のSRSがどの認識結果を生成するかを表す。この例では、 $SRS_A$ が5つの認識結果を生成し、そのうちの4つが、中程度から低い範囲の予想される信頼値に向かって互いに比較的接近して集中する。単一の認識結果quarryが、他の認識結果からはかなり離れて位置し、比較的高い信頼値を有する。このことは、互いにより代替可能である他の結果よりも結果「quarry」が著しく良好であることに $SRS_A$ がより高い信頼を持つことを示すことがある。

#### 【0099】

ある実装では、外れ値、すなわち離れている高い信頼値を重み付けし、関連する認識結果の選択を支持することができる。例えば、選択モジュール113は、結果「quarry」に関する信頼値0.9を定数1.05で重み付けすることができる。その場合、「quarry」に関する得られる信頼値は、0.945に増加する。

30

#### 【0100】

あるいは、より均一な間隔に配置された信頼値が、追加の重み付けを受けないことがある(またはより低い重み付けを受けることがある)。例えば、 $SRS_B$ で生成された認識結果に関する信頼値が、著しい外れ値のない、より均一な間隔に配置される。この場合、最上位の認識結果「quarry」が正しい可能性が低い(例えば、「quarry」は、より低い信頼値を有する結果のクラスタの中の最上の結果から離れて突出しない)ので、選択モジュール113は、「quarry」に関する信頼値に重みを加えないことがある。

40

#### 【0101】

図7A～Eは、認識結果を重み付けするのに使用することのできる、SRSによって出力される例示的認識結果集合と、集合間の相関とを示すベン図である。図7Aは、 $SRS_A$ で生成される認識結果<sub>A</sub>、 $SRS_B$ で生成される認識結果<sub>B</sub>、および $SRS_C$ で生成される認識結果<sub>C</sub>という3つの認識結果集合を含むベン図700である。

#### 【0102】

ベン図700で示されるように、結果<sub>A</sub>、結果<sub>B</sub>、および結果<sub>C</sub>は部分的に重なる。この例では、結果<sub>A</sub>および結果<sub>B</sub>は、結果<sub>A</sub>と結果<sub>C</sub>の重なり、または結果<sub>B</sub>と結果<sub>C</sub>の重なりと比べて、重なる結果をより多く有する。このことは、 $SRS_A$ と $SRS_B$ がしばしば同一の認識結果を生成するのに対して、 $SRS_C$ の結果は、それほど頻繁には $SRS_A$ または $SRS_B$ の結果と対応しない

50

ことを示すことがある。

#### 【0103】

ある実装では、結果の共通部分は、特定の音声認識タスクに応答してどのSRSが同一の認識結果を生成するかに基づく。例えば、2つのSRSが特定のタスクに関する最上位の認識結果を生成する場合、この結果を共通部分集合に追加することができる。

#### 【0104】

別の例では、第1SRSが認識結果「Cory」をその最上の結果として生成し、第2SRSが認識結果「Cory」を(生成した5個の結果のうちの)その第4順位の結果として生成する場合、結果「Cory」が共通部分集合に追加される。ある実装では、どちらも最上位に関連付けられない結果を共通部分結果に追加することができるが、それらが順位付けの点で異なることを示す割引き因子に関連付けることもできる。例えば、2つの順位付けの間の差を使用して、共通部分に関連する重み因子を割り引くことができる(例えば、順位付けのそれぞれの差を割引き因子に関連付けることができる)。例えば、順位付けが1および4である場合、差の絶対値は3であり、これを割引き因子0.3と関連付けることができ、割引き因子が、共通部分に関連する重みに乘じられる。例えば、重みが1.03であり、割引き因子が0.3である場合、全重みを、重み1.03の「ブースト」因子に対して、すなわち0.03に対して乗じることができる。この結果、より小さい新しいブースト因子0.01が得られ、したがって新しい全重み値は1.01である。

#### 【0105】

ある実装では、SRS間の認識結果の重なりを使用して認識結果を重み付けすることができ、その結果、認識結果が最終的な認識結果の選択で支持され、または不支持となる。例えば、認識結果がしばしば一致する2つのSRSによって生成される場合、あまり一致しない2つのSRSによって生成される認識結果に比べて、認識結果の重み付けを低くする(または不支持とする)ことができる。図7B～Eは、このことをより詳細に示す。

#### 【0106】

図7Bは、図7Aのベン図700からの結果<sub>A</sub>および結果<sub>B</sub>のみを含むベン図710である。前述のように、SRS<sub>A</sub>およびSRS<sub>B</sub>は、その認識結果の類似性に基づいて、ある程度同様に分類することができる。ある実装では、2つ(または3つ以上)のSRS間の重なりの中にある認識結果に重み因子を割り当てることができる。例えば、この集合の中にある認識結果に重み因子0.01を関連付けることができる。

#### 【0107】

ある実装では、この重み因子は、重なりが大きいときに小さくなり、重み因子は、重なりが小さいときに大きくなる。このことは、あまり一致しないSRSによって生成される結果は正しい可能性がより高いので、こうした重なる結果を支持すべきであるという仮定を反映することができる。例えば、異なる結果を生成するSRSは、基礎となる異なるアーキテクチャを有することがあり、異なるタイプの認識誤りの影響を受けやすいことがある。

#### 【0108】

図7Cは、図7Aのベン図700からの結果<sub>A</sub>および結果<sub>C</sub>のみを含むベン図720を示す。この例では、結果間の重なりは、図7Bに示される重なりよりも小さい。したがって、この実装では、重み因子0.6は、重なりの中にある結果について、図7Bに示される共通部分内の結果と比べて大きい。

#### 【0109】

同様に、図7Dは、結果<sub>B</sub>および結果<sub>C</sub>を含むベン図730を示す。これらの結果の共通部分は、図7Bおよび7Cの共通部分の間のサイズである。したがって、この実装では、重み因子も、図7Bおよび7Cの共通部分に関連する重み因子の間のサイズ(例えば0.03)である。

#### 【0110】

図7Eは、図7Aにも示されるベン図700を示すが、すべてのSRS<sub>A-C</sub>の結果の間の共通部分が強調表示されている。共通部分は、各SRSによって生成された認識結果の集合を反映する。3つのSRS間の一一致が(この例では)比較的まれであるとすると、この集合内にある認識結果は、他の重みよりも高い重み、すなわち0.1に関連付けることができる。

10

20

30

40

50

**【 0 1 1 1 】**

図8Aおよび8Bは、SRS間の共通部分がシステムの実行時動作中にどのように適応または変化することができるかを示すベン図800および810を示す。ある実装では、認識結果の共通部分が変化するとき、共通部分に関連する重みも変化することがある。

**【 0 1 1 2 】**

図8Aは、 $SRS_A$ および $SRS_B$ によって生成された認識結果の例示的第1共通部分を示す。第1共通部分は、重み0.01に関連付けられる。ある実装では、音声認識器108が、追加の音声復号化を実施して、追加の認識結果を生成する。 $SRS_{相関モニタ}282$ が、結果を監視して、様々なSRS間の結果の共通部分を識別することができる。

**【 0 1 1 3 】**

$SRS_{相関モニタ}282$ は、より多くの結果が生成されるときに、共通部分計算を動的に更新することができる。このことが図8Bに示されており、図8Bは、共通部分が変化したことについて、図8Aと同じ $SRS_A$ および $SRS_B$ を示す。この例では、特定の音声認識タスクについてSRSが一致した回数が、SRSが実施したタスク数と比べて増加したので、共通部分が増大している。

**【 0 1 1 4 】**

共通部分集合の増大に応答して、重みも低下させることができる。例えば、ベン図810の共通部分結果集合に、より小さい重み0.001を関連付けることができる。ある実装では、重み値の変化は、共通部分結果集合のサイズの変化に線形に関連付けることができる。例えば、システムは、認識器が別の認識器と類似するときに、認識器からの結果の重み付けまたは支持を低くすることができます。図8Aおよび8Bでは、2つの認識器に関する認識結果の類似性が、2つの認識器間の共通部分として表され、共通部分が大きいと、両方の認識器が同一の結果を生成するときにシステムが認識結果に結び付けることのできる重みが小さくなる。一方、2つの認識器が非常に異なる(例えば、音声認識アルゴリズムなどが異なるために、異なる認識結果を一般に生成した)とき、結果の共通部分が小さくなることがある。次いでこうした2つの異なる認識器が発話に関して一致したとき、一致は、結果が正しい可能性がより高いことを示すので、システムは、その結果がシステムでより重視されるようにその結果を重み付けすることができる。

**【 0 1 1 5 】**

図9は、SRS誤り率と認識結果に関連する重みとの間の例示的相関を示すグラフ900である。ある実装では、最終的な認識結果の選択で、誤り率が低いSRSによって生成される認識結果をより大きく重み付けすることができる。例えば、SRSが高い誤り率を有する場合、非常に正確なSRSによって生成された認識結果と比べて、その認識結果を割り引く(または大きく重み付けしない)ことができる。

**【 0 1 1 6 】**

グラフ900は、特定のSRSに重みを割り当てるための例示的関数、またはアルゴリズムを示す。グラフ900のy軸は、SRSに関連する誤り率を示し、x軸は、SRSに関連する重みを示す。この例では、求めた閾値より高い誤り率を有するSRS(例えば、 $SRS_A$ 、 $SRS_E$ 、 $SRS_C$ )を重み付けするのに、割引き重み(例えば、0.9、0.95、0.8)が使用される。閾値より低い誤り率を有するSRS(例えば $SRS_B$ )を重み付けするのに、ブースト重み(例えば、1.01、1.04、1.1)が使用される。この例では、誤り閾値上にあるSRS(例えば $SRS_D$ )を重み付けするのに、中立(neutral)重み(例えば1)が使用される。

**【 0 1 1 7 】**

ある実装では、各SRSに関連する誤り率は、認識結果が不正確であるという確認に基づいて更新することができる(例えば、結果が最終的な認識結果として選択され、ユーザによって拒否され、第1結果が最終的な認識結果として選択され、ユーザの受諾に基づいて正しいと判定され、したがって選択されない結果が、誤りのある結果として記録されるなど)。選択モジュール113は、各SRSに関連する更新後誤り率に基づいて、重みを動的に変更することができる。

**【 0 1 1 8 】**

10

20

30

40

50

図10は、クライアント、またはサーバ、または複数のサーバとして、本文書に記載のシステムおよび方法を実装するのに使用することのできるコンピューティング装置1000、1050のブロック図である。コンピューティング装置1000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、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ものとする。コンピューティング装置1050は、携帯情報端末、携帯電話、スマートフォン、他の類似のコンピューティングなどの様々な形態のモバイル装置を表すものとする。さらに、コンピューティング装置1000または1050は、ユニバーサルシリアルバス(USB)フラッシュドライブを含むことができる。USBフラッシュドライブは、オペレーティングシステムおよび他のアプリケーションを格納することができる。USBフラッシュドライブは、ワイヤレス送信機や、別のコンピューティング装置のUSBポートに挿入することのできるUSBコネクタなどの入力/出力構成要素を含むことができる。ここで示す構成要素、その接続および関係、ならびにその機能は、例示的なものに過ぎないものとし、本文書で説明され、かつ/または特許請求される本発明の実装を限定しないものとする。

#### 【0119】

コンピューティング装置1000は、プロセッサ1010、メモリ1020、記憶装置1030、メモリ1020および高速拡張ポートに接続される高速インターフェース、ならびに低速バスおよび記憶装置1030に接続される低速インターフェースを含む。各構成要素1010、1020、および1030は、様々なバス1050を使用して相互接続され、共通マザーボード上に取り付けることができ、または他の方式で適宜取り付けることができる。プロセッサ1010は、高速インターフェースに結合されたディスプレイなどの外部入力/出力装置1040上のGUI用のグラフィカル情報を表示するためにメモリ1020内または記憶装置1030上に格納された命令を含む、コンピューティング装置1000内で実行するための命令を処理することができる。別の実装では、複数のプロセッサおよび/または複数のバスを複数のメモリおよび複数のタイプのメモリと共に適宜使用することができる。さらに、複数のコンピューティング装置1000を接続することができ、各装置が必要な動作の各部分を実現する(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)。

#### 【0120】

メモリ1020は、コンピューティング装置1000内の情報を格納する。一実装では、メモリ1020は揮発性メモリユニットである。別の実装では、メモリ1020は不揮発性メモリユニットである。メモリ1020はまた、磁気ディスクや光ディスクなどの別の形態のコンピュータ可読媒体でもよい。

#### 【0121】

記憶装置1030は、コンピューティング装置1000用の大容量ストレージを提供することができる。一実装では、記憶装置1030は、フロッピィ(登録商標)ディスク装置、ハードディスク装置、光ディスク装置、テープ装置、フラッシュメモリまたは他の類似の固体メモリ装置、ストレージエリアネットワーク内の装置または他の構成を含む装置のアレイなどのコンピュータ可読媒体でよく、またはコンピュータ可読媒体を含むことができる。コンピュータプログラム製品は、情報キャリアとして有形に実施することができる。コンピュータプログラム製品はまた、実行されたときに上述のような1つまたは複数の方法を実施する命令を含むことができる。情報キャリアは、メモリ1020、記憶装置1030、プロセッサ1010上のメモリなどのコンピュータ可読媒体または機械可読媒体である。

#### 【0122】

高速コントローラは、コンピューティング装置1000に関する帯域幅を多用する動作を管理し、低速コントローラは、帯域幅を多用しない動作を管理する。そのような機能の割振りは例示的なものに過ぎない。一実装では、高速コントローラがメモリ1020、ディスプレイ(例えば、グラフィックスプロセッサまたはアクセラレータを介して)、および様々な拡張カード(図示せず)を受け入れることのできる高速拡張ポートに結合される。この実装では、低速コントローラが、記憶装置1030および低速拡張ポートに結合される。様々な通信ポート(例えば、USB、Bluetooth(登録商標)、イーサネット(登録商標)、ワイヤレスイー

10

20

30

40

50

サネット(登録商標))を含むことができる低速拡張ポートは、キーボード、ポインティングデバイス、スキャナなどの1つまたは複数の入力/出力装置1040、あるいは交換機やルータなどのネットワーキング装置に、例えばネットワークアダプタを介して結合することができる。

#### 【0123】

コンピューティング装置1000は、図に示されるように、いくつかの異なる形式で実装することができる。例えば、コンピューティング装置1000は、標準サーバとして実装することができ、またはそのようなサーバのグループとして複数回実装することができる。コンピューティング装置1000は、ラックサーバシステムの一部として実装することができる。さらに、コンピューティング装置1000は、ラップトップコンピュータなどのパーソナルコンピュータとして実装することができる。あるいは、コンピューティング装置1000の構成要素を、モバイル装置(図示せず)内の他の構成要素と組み合わせることができる。そのような装置のそれぞれは、1つまたは複数のコンピューティング装置1000を含むことができ、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装置1000から構成することができる。

10

#### 【0124】

別の実施形態において、装置1000は、他の構成要素の中でもとりわけ、プロセッサ1010と、メモリ1020と、ディスプレイ、通信インターフェース、トランシーバなどの入力/出力装置1040とを含む。装置1000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や他の装置などの記憶装置1030を備えることもできる。各構成要素1000、1010、1020、1030、および1040は、様々なバス1050を使用して相互接続され、構成要素のうちのいくつかは、共通マザーボード上に取り付けることができ、または他の方式で適宜取り付けることができる。

20

#### 【0125】

プロセッサ1010は、メモリ1020内に格納された命令を含む、コンピューティング装置1000内の命令を実行することができる。プロセッサ1010は、別々の複数のアナログプロセッサ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することができる。さらに、プロセッサ1010は、いくつかのアーキテクチャのうちのいずれかを使用して実装することができる。例えば、プロセッサ1010は、CISC(Complex Instruction Set Computer)プロセッサ、RISC(Reduced Instruction Set Computer)プロセッサ、またはMISC(Minimal Instruction Set Computer)プロセッサでよい。プロセッサ1010は例えば、ユーザインターフェースの制御など、装置1000の別の構成要素と協調するために、装置1000で実行されるアプリケーション、および装置1000によるワイヤレス通信を実現することができる。

30

#### 【0126】

プロセッサ1010は、制御インターフェース、およびディスプレイに結合されたディスプレイインターフェースを介してユーザと通信することができる。ディスプレイは、例えばTFT(薄膜トランジスタ液晶ディスプレイ)ディスプレイまたは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よい。ディスプレイインターフェースは、ディスプレイを駆動して、グラフィカル情報または他の情報をユーザに提示するための適切な回路を備えることができる。制御インターフェースは、ユーザからコマンドを受け取り、それをプロセッサ1010にサブミットするために変換することができる。さらに、外部インターフェースは、装置1000の他の装置との間の近領域通信を可能にするよう、プロセッサ1010との通信を実現することができる。外部インターフェースは例えば、ある実装では有線通信を実現することができ、または別の実装ではワイヤレス通信を実現することができ、複数のインターフェースを使用することもできる。

40

#### 【0127】

メモリ1020は、コンピューティング装置1000内の情報を格納する。メモリ1020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。拡張メモリを設け、拡張インターフェースを介

50

して装置1000に接続することもでき、拡張メモリは、例えばSIMM(Single In Line Memory Module)カードインターフェースを含むことができる。そのような拡張メモリは、装置1000用の追加の記憶空間を提供することができ、または装置1000用のアプリケーションまたは他の情報を格納することもできる。具体的には、拡張メモリは、上述のプロセスを実施または補足する命令を含むことができ、セキュア情報も含むことができる。したがって、例えば、拡張メモリは、装置1000用のセキュリティモジュールとして提供することができ、装置1000の安全な使用を許可する命令と共にプログラムすることができる。さらに、ハッキング不可能な方式でSIMMカード上に識別情報を配置することなど、SIMMカードを介してセキュアアプリケーションを追加の情報と共に提供することができる。

## 【0128】

10

メモリ1020は、例えば、上記で議論したように、フラッシュメモリおよび/またはNVRAMメモリを含むことができる。一実装では、コンピュータプログラム製品は、情報キャリアとして有形に実施することができる。コンピュータプログラム製品は、実行されたときに上述のような1つまたは複数の方法を実施する命令を含む。情報キャリアは、メモリ1020、拡張メモリ、プロセッサ1010上のメモリなどのコンピュータ可読媒体または機械可読媒体である。

## 【0129】

20

装置1000は、通信インターフェースを介してワイヤレスに通信することができ、通信インターフェースは、必要ならデジタル信号処理回路を含むことができる。通信インターフェースは、数ある中でも、GSM(登録商標)音声呼出し、SMS、EMS、またはMMSメッセージング、CDMA、TDMA、PDC、WCDMA(登録商標)、CDMA2000、またはGPRSなどの様々なモードまたはプロトコル下の通信を実現することができる。そのような通信は、例えば無線周波数トランシーバを介して行うことができる。さらに、Bluetooth(登録商標)、WiFi、またはそのような他のトランシーバ(図示せず)を使用するような近距離通信を行うことができる。さらに、GPS(全地球測位システム)受信機モジュールが、追加のナビゲーション関連ワイヤレスデータおよび位置関連ワイヤレスデータを装置1000に提供することができ、そのデータは、装置1000上で実行中のアプリケーションによって適宜使用することができる。

## 【0130】

30

装置1000はまた、オーディオコーデックを使用して可聴に通信することもでき、オーディオコーデック1060は、ユーザから発話情報を受け取って、それを使用可能なデジタル情報に変換することができる。オーディオコーデックは同様に、スピーカ、例えば装置1000のハンドセット内のスピーカなどを介して、ユーザに対する可聴音を生成することができる。そのような音は、音声電話呼出しからの音を含むことができ、記録音(例えば、ボイスメッセージ、音楽ファイルなど)を含むことができ、装置1000上で動作するアプリケーションによって生成された音も含むことができる。

## 【0131】

40

コンピューティング装置1000は、図に示されるように、いくつかの異なる形態で実装することができる。例えば、コンピューティング装置1000は、携帯電話として実装することができる。コンピューティング装置1000はまた、スマートフォン、携帯情報端末、または他の類似のモバイル装置の一部として実装することもできる。

## 【0132】

ここで説明したシステムおよび技法の様々な実装は、デジタル電子回路、集積回路、特に設計された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として実現することができる。こうした様々な実装は、記憶システム、少なくとも1つの入力装置、および少なくとも1つの出力装置からデータおよび命令を受け取り、それらにデータおよび命令を送るように結合された、専用または汎用でよい少なくとも1つのプログラム可能プロセッサを含むプログラム可能システム上で実行可能および/または解釈可能である1つまたは複数のコンピュータプログラムとしての実装を含むことができる。

## 【0133】

50

こうした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、またはコードとも呼ばれる)は、プログラム可能プロセッサ用の機械語命令を含み、高水準手続型および/またはオブジェクト指向プログラミング言語、および/またはアセンブリ/機械語として実装することができる。本明細書では、「機械可読媒体」、「コンピュータ可読媒体」という用語は、機械語命令を機械可読信号として受け取る機械可読媒体を含む、プログラム可能プロセッサに機械語命令および/またはデータを提供するのに使用される任意のコンピュータプログラム製品、機器、および/または装置(例えば、磁気ディスク、光ディスク、メモリ、プログラム可能論理装置(PLD))を指す。「機械可読信号」という用語は、機械語命令および/またはデータをプログラム可能プロセッサに提供するのに使用される任意の信号を指す。

10

#### 【0134】

ユーザとの対話を実現するために、ここで説明するシステムおよび技法は、ユーザに情報を表示するディスプレイ装置(例えば、CRT(陰極線管)モニタまたはLCD(液晶ディスプレイ)モニタ)と、ユーザがコンピュータに入力を与え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で実装することができる。他の種類の装置を使用して、ユーザとの対話を実現することもでき、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、視覚フィードバック、聴覚フィードバック、または触覚フィードバック)でよく、ユーザからの入力は、音響入力、音声入力、または触覚入力を含む任意の形態で受け取ることができる。

20

#### 【0135】

ここで説明するシステムおよび技法は、(例えばデータサーバとしての)バックエンド構成要素を含み、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)を含み、またはフロントエンド構成要素(例えば、ユーザがここで説明するシステムおよび技法の一実装と対話することの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)を含み、あるい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、また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として実装することができる。システムの構成要素は、任意の形態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」)、広域ネットワーク(「WAN」)、(アドホックメンバまたは静的メンバを有する)ピアツーピアネットワーク、グリッドコンピューティングインフラストラクチャ、およびインターネットを含む。

30

#### 【0136】

コンピューティングシステムは、クライアントおよびサーバを含むことができる。クライアントとサーバは、一般には互いに離れており、通常は通信ネットワークを介して対話する。クライアントとサーバの関係は、それぞれのコンピュータ上で実行中であり、互いにクライアント-サーバ関係を有するコンピュータプログラムによって生じる。

#### 【0137】

本発明のいくつかの実施形態を説明した。それでも、本発明の精神および範囲から逸脱することなく様々な修正を行えることを理解されよう。例えば、複数のSRSの組合せスコア、結合スコア、信頼度スコアが、仮説の整合性、発話の識別に関する推測などの特徴を含むことができる。例えば、信頼度0.8の第1結果を出力する3つのSRSは、信頼度0.9の第2結果を出力する1つのSRSよりも信頼度が高いことがある。

40

#### 【0138】

ある実装では、待ち時間または他の因子に基づいて、所定の1組のSRSを使用のために選択することができる。例えば、一般疑問文に対する回答を求めてユーザを促すことに応答して音声が受信される場合、すべての利用可能なSRSが回答を処理することを許可するのではなく、最速の2つのSRSを選択して回答を処理することができる。

#### 【0139】

さらに、ある実装では、SRSによって生成された個々の認識結果が一致しないとき、最

50

終的な認識結果での全体的信頼度が低下することがある。認識結果が全く重複しないときに「最良」の現結果を選択するための1つの例示的アルゴリズムは、個々の最高の信頼度を有する認識結果を選択することである。この例では、組合せ信頼度は、システムが重複のない類似の条件および類似の与えられた信頼値を有したときにトレーニング中にカウントされた正しい認識結果の予想数となる。与えられた量の認識結果の部分的重複について、類似のカウントおよび統計を推定することができる。したがって、重複の程度がトレーニング中の全認識誤りの低下に相関するので/相関する場合、システム全体は、部分的に重複する認識結果の組合せにより高い信頼値を割り当てることができる。

#### 【0140】

例えば、上記で示した様々な形態のフローは、ステップを再順序付けし、追加し、または除去して使用することができる。さらに、音声復号化で複数の音声認識システムを使用するいくつかの応用例および方法を説明したが、他の多数の応用例も企図されることを理解されたい。したがって、他の実施形態は、以下の特許請求の範囲内にある。

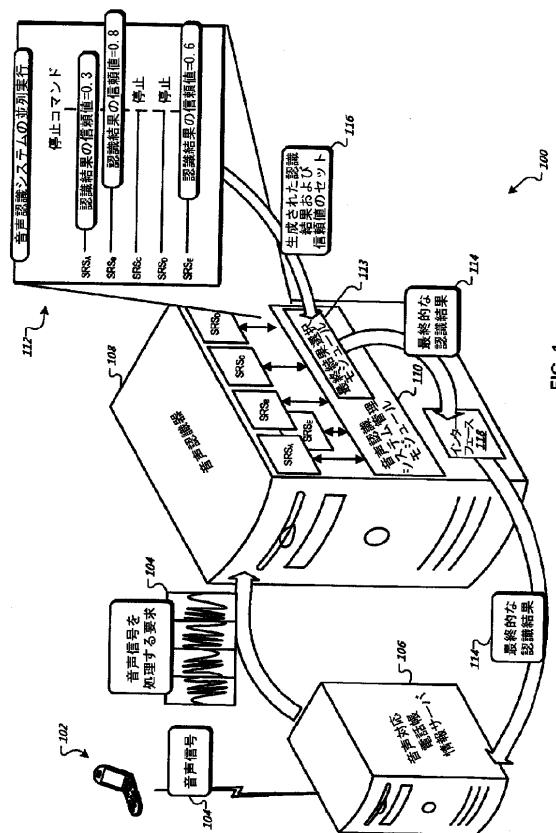
10

#### 【符号の説明】

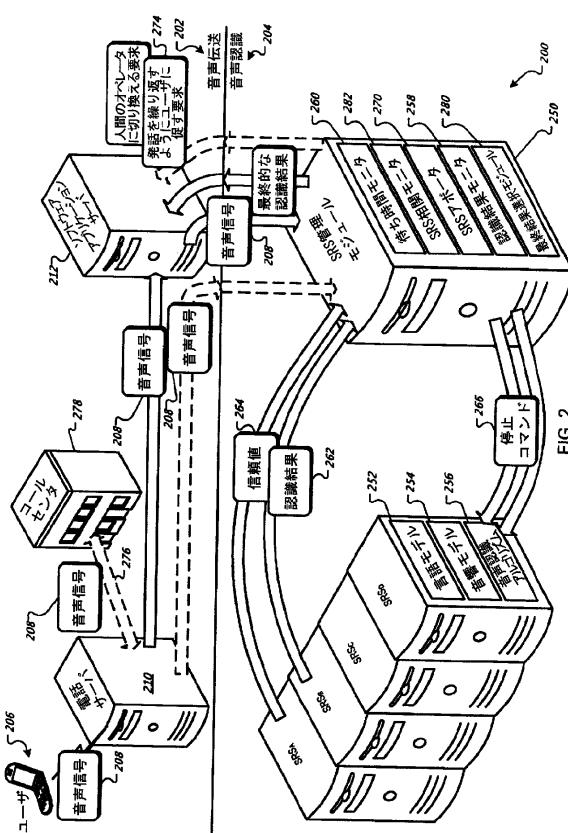
#### 【0141】

100、200	例示的システム	
102、206	携帯電話	
104、208	音声信号	
106	音声対応電話帳情報サーバ	
108	音声認識器	20
110、250	SRS管理モジュール	
113、280	最終結果選択モジュール	
114	最終的な認識結果	
116、264	信頼値	
202	音声伝送セグメント	
204	音声認識器セグメント	
212	電話サーバ	
214	ソフトウェアアプリケーションサーバ	
252	言語モデル	
254	音響モデル	30
256	音声認識アルゴリズム	
258	認識結果モニタ	
262	認識結果	
266	停止コマンド	
268	待ち時間モニタ	
270	SRSアボータ	
282	S R S 相関モニタ	

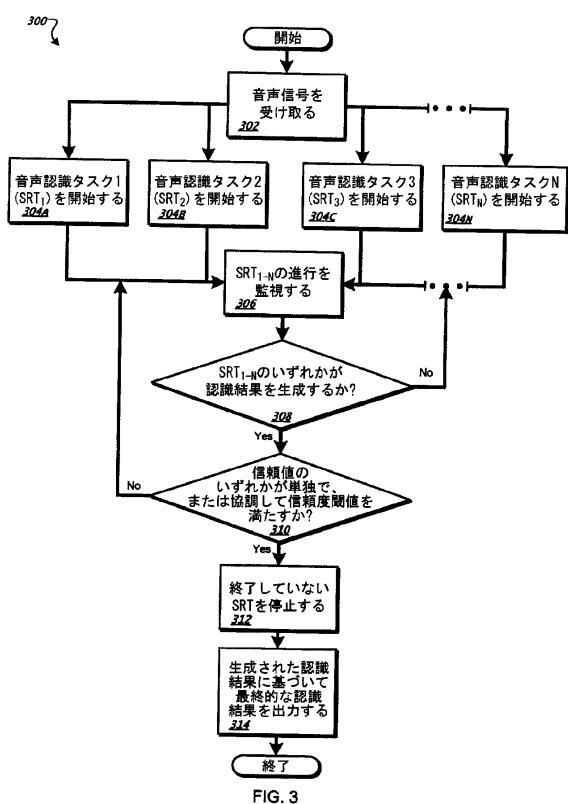
【図1】



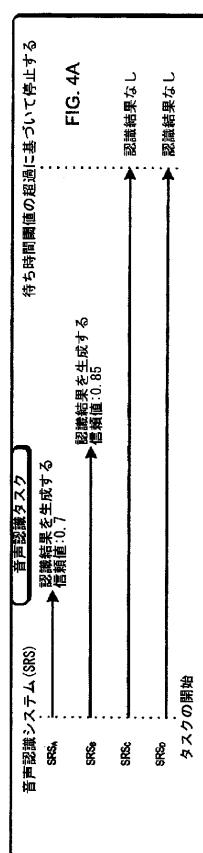
【図2】



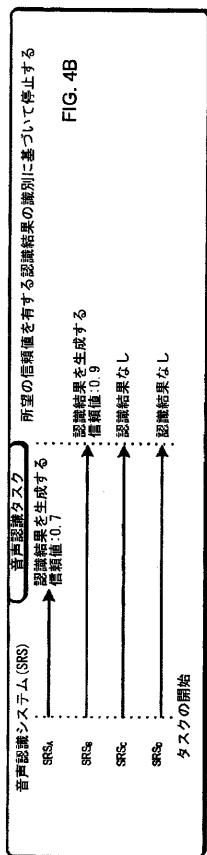
【図3】



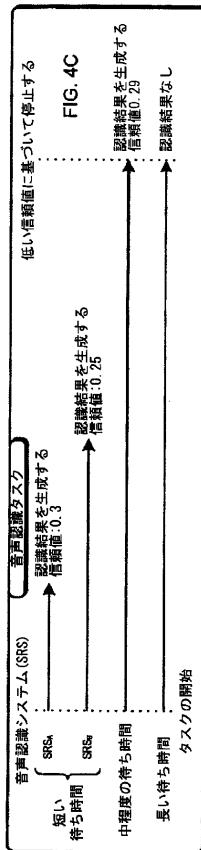
【図4A】



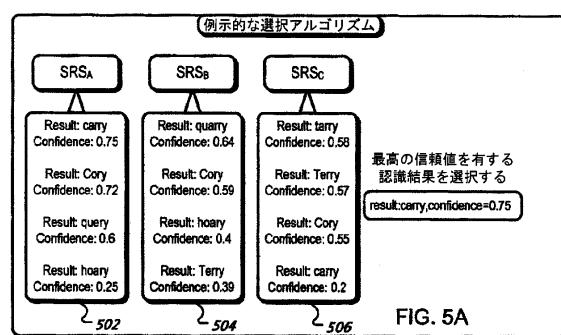
【図 4 B】



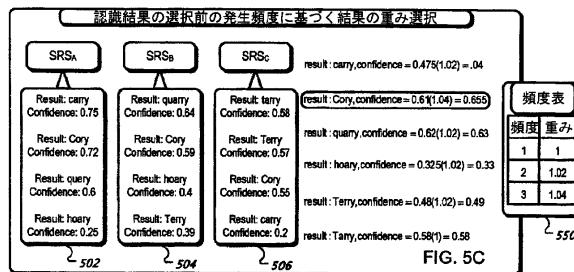
【図 4 C】



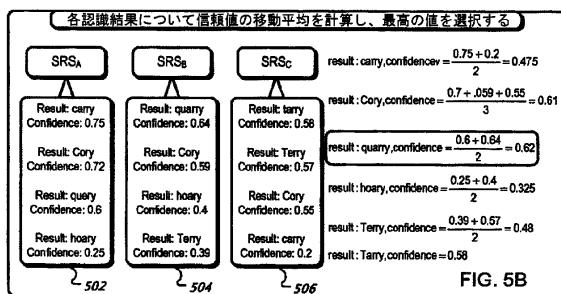
【図 5 A】



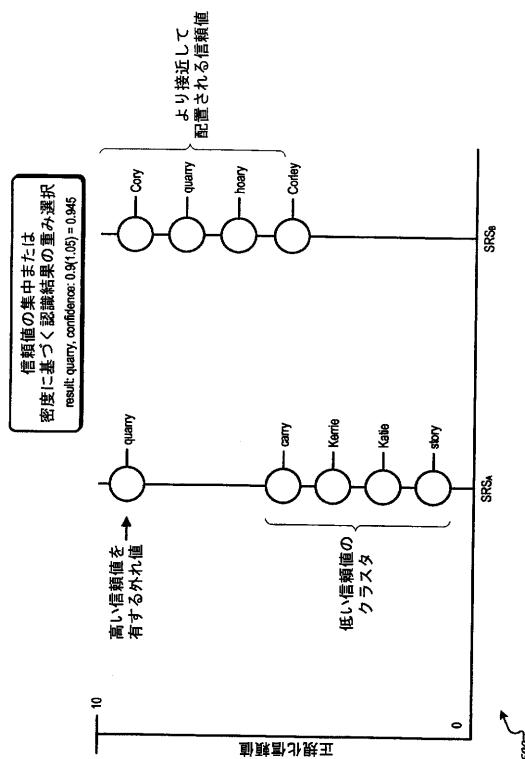
【図 5 C】



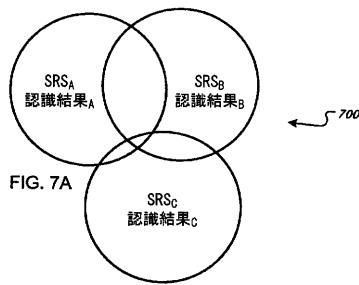
【図 5 B】



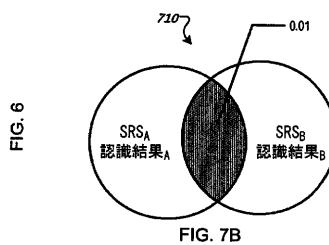
【図 6】



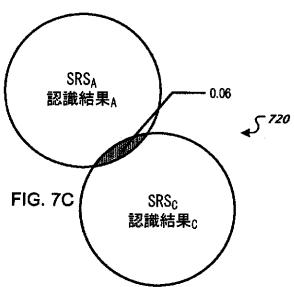
【図 7 A】



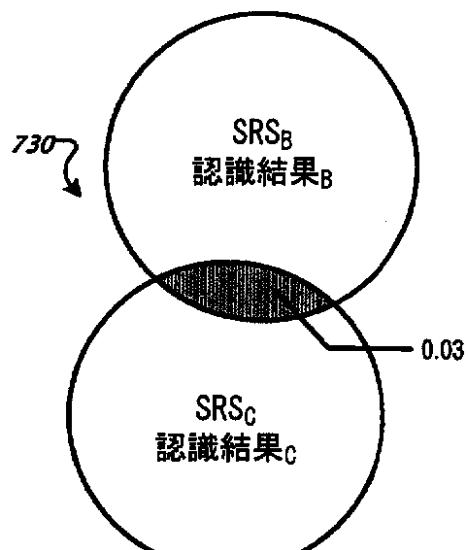
【図 7 B】



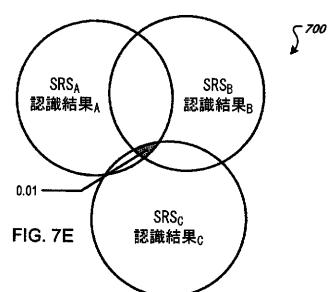
【図 7 C】



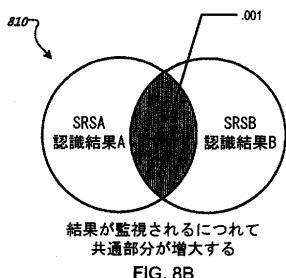
【図 7 D】



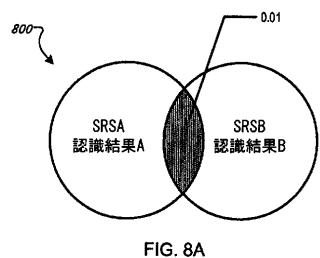
【図 7 E】



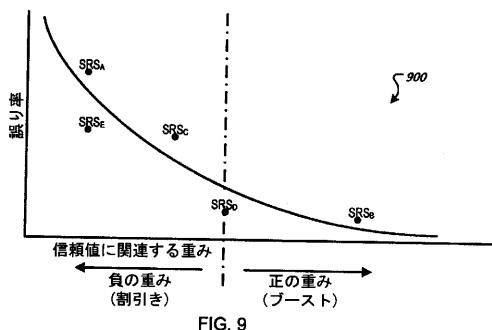
【図 8 B】



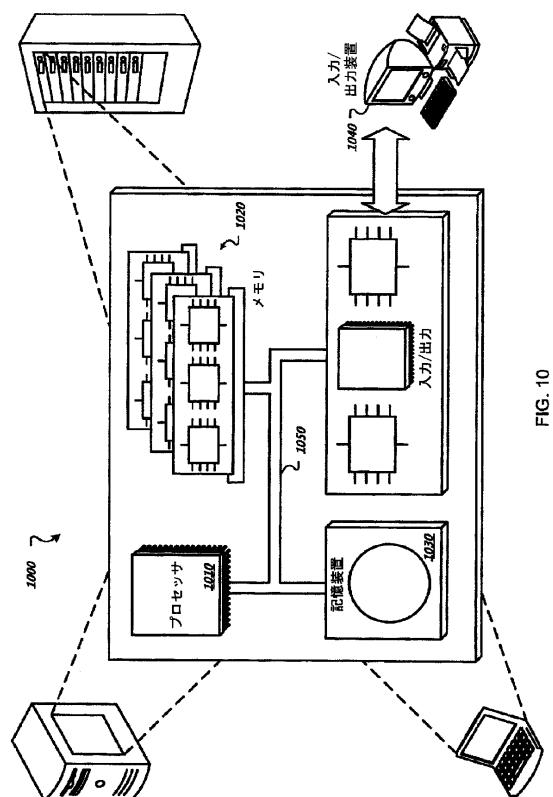
【図 8 A】



【図 9】



【図 10】



---

フロントページの続き

(72)発明者 ブライアン・ストローブ

アメリカ合衆国・カリフォルニア・94303・パロ・アルト・ドリフトウッド・ドライブ・86  
8

(72)発明者 フランソワーズ・ボーフェイ

アメリカ合衆国・カリフォルニア・94040・マウンテン・ビュー・キャシー・ウェイ・113  
9

(72)発明者 オリヴィエ・ショアン

アメリカ合衆国・ニュージャージー・07452・グレン・ロック・ウォルドロン・アヴェニュー  
・88

審査官 五貫 昭一

(56)参考文献 特開2009-265307(JP,A)

特開2004-198597(JP,A)

特開2004-198656(JP,A)

特開2005-221727(JP,A)

特開2006-107108(JP,A)

特開2006-228136(JP,A)

特開2007-206501(JP,A)

(58)調査した分野(Int.Cl., DB名)

G10L 15/32

G06F 9/50