(19)日本国特許庁(JP)　(12)公表特許公報(A)　(11)特許出願公表番号
特表2016-517047
(P2016-517047A)
(43)公表日　平成28年6月9日(2016.6.9)

(51)Int.Cl.		F I		テーマコード（参考）
G10L 15/14	(2006.01)	G10L	15/14	200Z
G10L 15/18	(2013.01)	G10L	15/18	300H

審査請求 有　予備審査請求 有　（全 20 頁）

(21)出願番号	特願2016-510953 (P2016-510953)
(86)(22)出願日	平成25年6月26日 (2013.6.26)
(85)翻訳文提出日	平成27年1月26日 (2015.1.26)
(86)国際出願番号	PCT/EP2013/063330
(87)国際公開番号	WO2014/177232
(87)国際公開日	平成26年11月6日 (2014.11.6)
(31)優先権主張番号	P.403724
(32)優先日	平成25年5月1日 (2013.5.1)
(33)優先権主張国	ポーランド(PL)

(71)出願人　514033448
アカデミア　ゴルニツォ－ハットニツァ　アイエム.　スタニスラワ　スタシツァ　ダブリュー　クラクフィ
ポーランド共和国、ピーエル－３０－０５９　クラクフ　エーエル.　ミツキエビツァ　３０
(74)代理人　110000877
龍華国際特許業務法人
(72)発明者　ジオルコ、バートス
ポーランド共和国、ピーエル－３０－０５９　クラクフ　エーエル.　ミツキエビツァ　３０　アカデミア　ゴルニツォ－ハットニツァ　アイエム.　スタニスラワ　スタシツァ　ダブリュー　クラクフィ内
最終頁に続く

(54)【発明の名称】音声認識システム及びダイナミックベイジアンネットワークモデルの使用方法

(57)【要約】
入力デバイス（１０２Ａ）により、音声を表す電気信号を登録し（２０１）、信号を周波数または時間周波数領域に変換するステップ（２０２）、単語（W）の仮説および観測された信号特徴（ＯＡ，ＯＶ）に基づくそれらの確率を生成するよう構成されたダイナミックベイジアンネットワーク（２０５）に基づく解析モジュールにおいて信号を分析するステップ、及び特定の単語（W）仮説及びそれらの確率に基づいて、音声を表す電気信号に対応するテキストを認識するステップ（２０９）を備える音声認識のコンピュータ実装方法。方法は、解析モジュール（２０５）に、各ラインに対して別個の時間セグメントに対する少なくとも２つの並列信号処理ライン（２０４ａ，２０４ｂ，２０４ｃ，２０４ｄ，２０１ａ）における周波数または時間周波数領域（２０２）内の信号に対して決定される観測された信号特徴（３０８－３１２）を入力すること、及び、解析モジュール（２０５）において、少なくとも２つの別個の時間セグメントに対して観測された信号特徴（３０８－３１２）の間の関係を分析することを特徴とする。

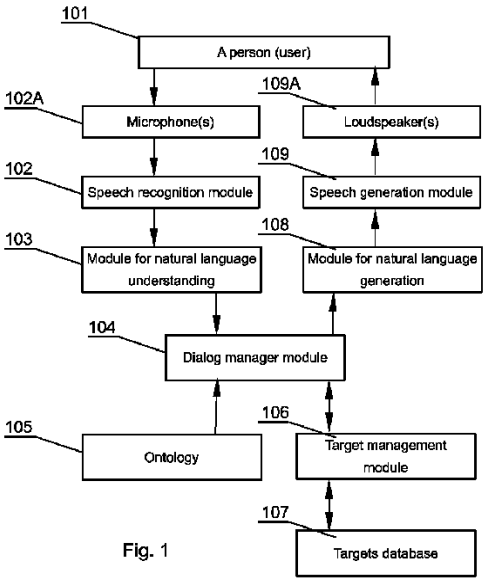


Fig. 1

【特許請求の範囲】
【請求項１】
音声認識のコンピュータ実装方法であって、
入力デバイス（１０２Ａ）により、音声を表す電気信号を登録し、前記電気信号を周波数または時間周波数領域（２０２）に変換する段階（２０１）と、
ダイナミックベイジアンネットワーク（２０５）に基づいて解析モジュール内の前記電気信号を分析する段階であり、それにより、複数の単語（Ｗ）の複数の仮説および観測された複数の信号特徴（ＯＡ，ＯＶ）に基づくそれらの確率を生成する、段階と、
特定の複数の単語（Ｗ）仮説およびそれらの確率に基づいて、音声を表す前記電気信号に対応するテキストを認識する段階（２０９）と、
前記解析モジュール（２０５）に、各ラインに対して別個の複数の時間セグメントに対する少なくとも２つの並列信号処理ライン（２０４ａ、２０４ｂ、２０４ｃ、２０４ｄ、２０１ａ）における周波数または時間周波数領域（２０２）内の前記電気信号に対して決定される観測された複数の信号特徴（３０８－３１２）を入力する段階と、
前記解析モジュール（２０５）において、少なくとも２つの別個の時間セグメントに対して観測された前記複数の信号特徴（３０８－３１２）の間の複数の関係を分析する段階と、
を備える、コンピュータ実装方法。
【請求項２】
前記複数の時間セグメントは、所定の継続時間を有する、請求項１に記載のコンピュータ実装方法。
【請求項３】
前記複数の時間セグメントは、複数の音素、複数の音節、複数の単語のような複数の音声セグメントのコンテンツに依存する、請求項１または２に記載のコンピュータ実装方法。
【請求項４】
前記解析モジュール（２０５）において、モデルを記述する複数の変数の間の複数の決定論的及び蓋然論的関係を定義する段階をさらに備え、複数の前記蓋然論的関係は、少なくとも観測された前記複数の信号特徴を現在の状態（Ｓｔｉ）にリンクするために定義される、請求項１から３のいずれか一項に記載のコンピュータ実装方法。
【請求項５】
前記複数の関係を分析する段階は、異なる観測された複数の信号特徴（ＯＡ、ＯＶ）を同時に分析する段階（２０５）を含む、請求項１から４のいずれか一項に記載のコンピュータ実装方法。
【請求項６】
音声認識のコンピュータ実装システムであって、
音声を表す電気信号を登録する入力デバイス（１０２Ａ）と、
音声を表す登録された前記電気信号を周波数または時間周波数領域に変換するモジュール（２０２）と、
音声を表す前記電気信号を分析し、複数の単語（Ｗ）の複数の仮説および観測された複数の信号特徴（ＯＡ，ＯＶ）に基づくそれらの確率を生成するダイナミックベイジアンネットワークに基づく解析モジュール（２０５）と、
複数の単語（Ｗ）の定義された前記複数の仮説及びそれらの確率に基づいて、音声を表す前記電気信号に対応するテキストを認識するモジュール（２０９）と、
各ラインに対して別個の複数の時間セグメントに対する少なくとも２つの並列信号処理ラインにおいて、少なくとも２つの観測された信号特徴（３０８－３１２）を、前記解析モジュール（２０５）に対して決定する少なくとも２つの信号パラメータ化モジュール（２０４ａ、２０４ｂ、２０４ｃ、２０４ｄ、２０１ａ）と、
を備え、前記解析モジュール（２０５）は、少なくとも２つの別個の時間セグメントに対して観測された前記複数の信号特徴（３０８－３１２）の間の複数の依存性を分析する、

コンピュータ実装システム。
【請求項７】
コンピュータ上で実行されると、請求項１から５のいずれか一項に記載のコンピュータ実装方法のすべての段階を実行するプログラムコード化手段を備えるコンピュータプログラム。
【請求項８】
コンピュータ上で実行されると、請求項１から５のいずれか一項に記載のコンピュータ実装方法のすべての段階を実行する複数のコンピュータ実行可能命令を格納するコンピュータ可読媒体。音声認識システム及びダイナミックベイジアンネットワークモデルの使用方法。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明の対象は、音声認識システム及びこの目的にためのベイジアンネットワークの使用方法である。特に、そのような自動音声認識システムは、広告及び情報提供の目的のための対話システムに適用できる。対話システムの実装は、顧客又は見物人との対話を始めて、適当なマルチメディアコンテンツを提供するインフォメーションキオスク又はブースの形をとってよい。
【背景技術】
【０００２】
音声認識システムは、日常生活において、ますます一般的になっている。例えば、それらは、公共交通機関のためのような情報コールセンタにおいて実装されている。しかし、これらのシステムは、まだ、頻繁に、音声の代わりに、入力情報のソースとしてキーパッド及びテキストにより動作している。
【０００３】
ユーザとの対話を実施可能にする様々な種類のコンピュータ化されたインタラクティブキオスクが知られている。例えば、米国特許６２５６０４６号明細書は、人の存在を示す環境内の変化を検出するために動き及び色分析を使用することにより、視覚データを処理することにより人を検出するコンピュータ化されたキオスクにおけるアクティブパブリックユーザインターフェースを開示している。相互作用空間が規定され、システムは、無生物の加算又は減算を反映し、照明変化を補償するために経時的に更新されるその環境の初期モデルを記録する。システムが移動対象物のモデルを開発し、それにより、人が相互作用空間について移動する間、彼らを追跡することが可能となる。さらに、ステレオカメラシステムが、位置及び移動を検出するシステムの性能を向上する。キオスクは、それが「見る」ものに応じて、音声及び視覚的フィードバックを提供する。
【０００４】
米国特許出願公開第２００８／０２０４４５０号明細書は、未承諾広告が自動化されたアバタにおいて具現化された仮想宇宙を提供するシステム、方法、及びプログラム製品を開示している。広告アバタを仮想宇宙に導入する登録システム、広告アバタが広告コンテンツの配信のためにユーザアバタを標的化する標的化システム、広告アバタが仮想宇宙内を移動する方法を定義する移動システム、及び広告アバタが広告コンテンツをユーザアバタに配信する方法を定義する広告配信システムを含むシステムが提供される。
【０００５】
上述のような既知の対話システムの欠点は、ユーザとの錯綜した対話を行うには不十分な音声認識性能を含む。
【０００６】
米国特許７２０３３６８号明細書は、ＨＭＭ（隠れマルコフモデル）及びＣＨＭＭ（連結隠れマルコフモデル）を用いる階層的な統計モデルを形成するパターン認識手順を開示している。階層的な統計モデルは、複数のスーパーノードを有する親レイヤ及び親レイヤの各スーパーノードに関連付けられた複数のノードを有する子レイヤをサポートする。ト

レーニングの後、階層的な統計モデルは、データセットから抽出される観測ベクトルを使用して、実質的に最適な状態シーケンスのセグメントを見つける。この処理の改良は、有利であろう。

【０００７】

ＨＭＭに基づく解より少ない制限を置くより一般的な解は、音声認識のベイジアンネットワークを使用する。ダイナミックベイジアンネットワーク（ＤＢＮ）を含むベイジアンネットワークを使用する解は、以下の刊行物に提示されている。

【０００８】

M. Wester, J. Frankel, and S. King, "Asynchronous articulatory feature recognition using dynamic Bayesian networks" (Proceedings of IEICI Beyond HMM Workshop, 2004),

J. A. Bilmes and C. Bartels, "Graphical model architectures for speech recognition", IEEE Signal Processing Magazine, vol. 22, pp. 89-100, 2005,

J. Frankel, M. Wester, and S. King, "Articulatory feature recognition using dynamic Bayesian networks", Computer Speech and Language, vol. 21, no. 4, pp. 620-640, October 2007.

【０００９】

ベイジアンネットワークを利用する音声認識方法は、特徴ベクトルに係る音の持続時間のモデリングに基づく。ＤＢＮでは、継続時間を表す変数を音を表す変数に置き換えることが可能となった。それにもかかわらず、すべての従来技術の解は、所定の時間範囲内で音声分析を行った。

【００１０】

前述の先行技術を考慮すると、人と機械との間の対話効率を改善できる音声認識システム及び方法を設計及び実装する必要がある。

【発明の概要】

【００１１】

本発明の対象は、入力デバイスにより、音声を表す電気信号を登録し、信号を周波数または時間周波数領域に変換するステップ、単語（Ｗ）の仮説及び観測された信号特徴（ＯＡ，ＯＶ）に基づくそれらの確率を生成するよう構成されたＤＢＮに基づく解析モジュールにおいて信号を分析するステップ、及び特定の単語（Ｗ）仮説及びそれらの確率に基づいて、音声を表す電気信号に対応するテキストを認識するステップを備える自動音声認識のコンピュータ実装方法である。方法は、解析モジュールに、各ラインに対して別個の時間セグメントに対する少なくとも２つの並列信号処理ラインにおける周波数又は時間周波数領域内の信号に対して決定される観測された信号特徴を入力すること、及び、解析モジュールにおいて、少なくとも２つの別個の時間セグメントに対して観測された信号特徴の間の関係を分析することを特徴とする。

【００１２】

好ましくは、時間セグメントは、所定の継続時間を有する。

【００１３】

好ましくは、時間セグメントは、音素、音節、単語のような音声セグメントのコンテンツに依存する。

【００１４】

好ましくは、方法は、さらに、解析モジュールにおいて、モデルを記述する変数の間の決定論的及び蓋然論的関係を定義する段階をさらに備え、蓋然論的関係は、少なくとも観測された信号特徴を現在の状態にリンクするために定義される。

【００１５】

好ましくは、方法は、さらに、異なる観測された信号特徴（ＯＡ，ＯＶ）を同時方法で分析する段階を備える。

【００１６】

本発明の別の対象は、音声を表す電気信号を登録する入力デバイス、音声を表す登録さ

れた電気信号を周波数または時間周波数領域に変換するモジュール、音声を表す信号を分析し、単語の仮説および観測された信号特徴（ＯＡ，ＯＶ）に基づくそれらの確率を生成するよう構成されたＤＢＮに基づく解析モジュール、及び単語の定義された仮説及びそれらの確率に基づいて、音声を表す電気信号に対応するテキストを認識するモジュールを備える音声認識のコンピュータ実装システムである。システムは、さらに、各ラインに対して別個の時間セグメントに対する少なくとも２つの並列信号処理ラインにおいて、少なくとも２つの観測された信号特徴を、解析モジュールに対して決定する少なくとも２つの信号パラメータ化モジュールを備え、解析モジュールは、少なくとも２つの別個の時間セグメントに対して観測された信号特徴の間の依存性を分析するよう構成される。

【００１７】

本発明の対象は、コンピュータ上で実行されると、本発明に係るコンピュータ実装方法のすべてのステップを実行するプログラムコード化手段を備えるコンピュータプログラムでもあるとともに、コンピュータ上で実行されると、本発明に係るコンピュータ実装方法のすべてのステップを実行するコンピュータ実行可能命令を格納するコンピュータ可読媒体でもある。

【図面の簡単な説明】

【００１８】

本発明の対象は、以下の図面内の典型的な実施形態に提示されている。

【図１】本発明に係るシステムのブロック図を示す。

【図２】自動音声認識処理のブロック図を示す。

【図３】異なる長さの並列期間上のＤＢＮを用いる音声のモデリングを示す。

【図４】単語のシーケンス（典型的な目的に対して簡素化されたバージョン）をデコードする、図３内に示される１つと同様のＤＢＮの使用例を示す。

【発明を実施するための形態】

【００１９】

図１は、本発明に係るシステムのブロック図を示す。そのようなシステムは、対話システムを提供するインタラクティブ広告又は他の情報において使用されてよい。対話は、可能な限り実際の対話に近いものでなければならない。そのような前提の実装は、パターン認識、意味分析、オントロジ知識及び音声合成に続く自然言語生成の使用のような技術の使用により可能である。

【００２０】

本発明を使用することができる対話システムは、複数の高品質のディスプレイ又はイメージプロジェクタを備え得る。好ましい実施形態では、対話システムは、ユーザ存在検出、又はより進歩的な場合では、バイオメトリック検出器、顔認識モジュール等のユーザ特性検出器を装備してもよい。対話システムは、音声のより効率的な取得のための指向性マイクを備えてもよい。

【００２１】

出力情報は、対話のコンテキストに適合され、ユーザの好みを決定する。

【００２２】

対話システムは、好ましくは、ユーザが会話をする視覚的アバタ又は人のイメージを出力する。音声認識を採用する対話システムは、インタラクティブに一人の人又は複数の人１０１と通信する。人１０１は、音声入力モジュール、例えばマイク１０２Ａに向かって話すことにより質問を入力する。マイクにより登録される音声は、音声認識モジュール１０２により処理され、続いて、自然言語を認識するためのモジュール１０３に配信される。

【００２３】

理解のためのモジュール１０３は、それらが機械に理解でき、容易且つ迅速に処理され得るような方法において、予想される応答の文脈で人１０１の陳述の認識の仮説を解釈する責任を伴う。例えば、システムが観光情報スポットで実装されている場合、それらの確率を用いる音声仮説のリストに基づく理解のためのモジュール１０３は、スピーカが、彼

がそれがどのような場所であるか探している場合、特定の場所、又はサービス、公共交通機関が運営等する時間の情報を探しているかを判断するタスクを有する。最も単純なバージョンでは、モジュールは、この目的のためにキーワードを利用するが、ここでは、D. Jurafsky, J.H. Martin, "Speech and Language Processing", Second Edition, Pearson Education, Prentice Hall, 2009に提示されるシンタックスモデル（例えば、センテンスパーサ）及び／又はセマンティックモデル（例えば、Ｗｏｒｄｎｅｔ又はセマンティックＨＭＭ）に基づくより高度な解を使用してもよい。

【００２４】

自然言語１０３を理解するためのモジュール内で処理されると、センテンス又はセンテンスの仮説は、（例えば、D. Jurafsky, J.H. Martin, "Speech and Language Processing", Second Edition, Pearson Education, Prentice Hall, 2009に記載されているように）目標管理モジュール１０６及び目標データベース１０７と協同して、適切にオントロジーモジュール１０５にクエリすることによって、ユーザクエリに提示される応答を決定する対話管理モジュール１０４に送られる。

【００２５】

オントロジーモジュール１０５は、領域についての整然とした知識、例えば、どの製品が特定の種類で入手可能か、人が選択したものと一緒に何を購入したかなど情報を備える。オントロジーモジュールは、更に、例えば、対話中の人の友達が、人が訪問等する市内にいるかどうかをチェックするソーシャルサービスからの異なる種類のデータを備えてもよい。オントロジーモジュールは、コンピュータ又は他の機械が処理できるような方法で体系化されたあらゆる他の実用的な知識を備えてもよい。

【００２６】

目標管理モジュール１０６は、コンピュータ内に、本発明に係るシステムに義務が実行される専門家（例えば、商業従業員）を導く商業、広告、交渉等の既知のルールを実装するために使用される。

【００２７】

応答のコンテンツを決定した後、自然言語の応答が、自然言語１０８を生成するためのモジュール及び続いて音声生成モジュール１０９において生成される。音声の形成において生成された応答は、スピーカ又はシステムにインストールされた他の出力デバイス１０９Ａを介して人１０１に出力される。

【００２８】

本発明において使用されるキー要素は、ベイジアンネットワークからなる分析のためのコンピュータ実装モジュールである。ベイジアンネットワークは、別個の要素が互いに依存し得る複雑な現象のモデリングを可能とする。基本モデルは、ノードがモデル（ランダム変数）の別個の要素を表す方向性非環式グラフとして生成される。ここで、エッジが、これらの要素間の依存関係を表す。

【００２９】

更に、エッジは、イベントの１つが、別のイベントが特定の値を仮定する条件の下で発生することを指定する、割り当てられた確率値を持つ。ベイズの定理を用いることにより、複雑な条件付き確率は、ベイジアンネットワークの特定のパスに対して計算され得る。これらの確率は、ネットワークの個々の要素により取られる値について推論するために使用されてもよい。

【００３０】

各ネットワーク変数は、それに接続されていない他の変数に条件付きで独立していなければならない。この方法で生成されたグラフは、イベントのコンパクトな表現、これらのイベントの発生の累積確率、及びグラフのノード間の条件付き独立性に関する前提として解釈されてよい。

【００３１】

ＤＢＮは、音声認識に採用してよい。複数のノードは、単一のランダム変数ではなく、変数のシーケンスを表す。これらは、時間の経過に応じて音声モデリングを可能にする時

間シリーズとして解釈される。従って、複数の連続する観測状態は、最終状態への明確なパスを正当化する。

【００３２】

標準的なベイジアンネットワークの使用は、音の持続時間の予測に基づいて、調音特徴のベクトルに依存する。ネットワークは、各特徴に対する単一の離散変数及び音の持続時間に対する単一の連続変数を有する。ネットワークは、特徴間の関係を記述する。特徴を表現するノードの値は、ネットワークに入る値及び任意に他の特徴に依存する。持続時間を表すノードの値は、他のノードから受信される値のみに直接依存する隠れ層（ＨＭＭにおけるように）である。

【００３３】

ＤＢＮの導入は、継続時間を表す変数を音を表す変数に置き換えることを可能にする。特徴間の関係を有するネットワーク全体は、ネットワークの１つが、時間ｔ－１で分析される信号及び時間ｔでの次の信号を表すようにコピーされる。両ネットワークは、時間的に変化する状態間の遷移の確率値を有するエッジで接続される。

【００３４】

本発明は、２つのサブネットワークを用いる場合のみに限定されるものではないことに留意すべきである。より多くのサブネットワーク、次の時間モーメントに対する各サブネットワークがあってもよい。一般的に、数１００又は数１０００のネットワークがあってもよい。そのような構造は、次の時間モーメントに何度もコピーされてもよい。更に、そのような局所ベイジアンネットワーク構造は、幾つかの場合には異なる時間の間で、それ自体を修正してもよい。

【００３５】

ＤＢＮモデルは、異なるソース、例えば音響特徴及び視覚特徴（唇の動きのような）から生じる信号についての情報を結合するために使用されてもよい。この種のシステムは、特に、異なる音響条件を有する場所での応用に有用である。低い値の信号対雑音比（ＳＮＲ）は、ストリート、空港、工場等のような場所において、唯一の音響経路に由来する情報を使用すると、得られた結果の品質の顕著な低下をもたらす。同じタイプのノイズに敏感でない別の信号タイプから得られる情報を加えることで、生じる困難を除去し、そのような場所においても音声認識システムを使用することを可能にする。

【００３６】

本発明者等は、ベイジアンネットワークが、音声分析に使用される際のＨＭＭ方法と比較して、より少ない制限を課すことに気づいた。

【００３７】

図２は、音声認識処理のブロック図を示す。次の説明は、また、異なる長さの期間に関連する時間でＤＢＮの使用とともに音声のモデリングを示す図３の幾つかの特徴を参照する。

【００３８】

図３に示されるように、ＤＢＮは、本明細書において、別個の観測が異なる持続時間を表すように音声をモデリングするために使用される。これらの異なる持続時間は、所定の長さ、例えば５ｍｓ、２０ｍｓ、６０ｍｓのセグメントであってもよく、音素、音節、単語、又は両タイプの組み合わせ、例えば５ｍｓ、２０ｍｓ、音素、単語のような音声セグメントのコンテンツに依存する。

【００３９】

提示された方法は、状態確率（図３におけるＳｔ１からＳｔ６）を評価するためにＤＢＮモデルを使用することで、異なる情報タイプの抽出及び取得した特徴の直接的融合を可能にする。

【００４０】

ＤＢＮにおける推論は、モデルを記述する変数間の２種類の関係、決定論的関係（図３に直線矢印としてマーク付けされる）及び蓋然論的関係（図３に波型矢印としてマーク付けされる）に基づく。

【００４１】
決定論的関係は、既知の事実、例えば与えられた単語Ｗｔｉを分析すると知られる位置Ｗｐｓ及び第１種の音素Ｐｔｉに基づいて定義される。そして、音素から次の音素への遷移Ｐｔｒが発生した又は発生しなかったことを知ることにより、単語内の現在の音素の位置が決定され得る。音素の遷移が起こらないと、時刻ｔ＋１でのＷｐｓは時刻ｔでのＷｐｓに等しく、上記の遷移が観察される場合、Ｗｐｓ＋１に等しい。
【００４２】
１つの単語から別の単語への遷移Ｗｔｒに関する情報も、同様に得ることができる。書き表された単語の最後の音素からの遷移の発生は、別の単語Ｗｔｉの分析の必要を意味する。
【００４３】
関係の別のタイプは、蓋然論的関係である。変数に基づいて推論するために、蓋然論的関係が存在する間で、これらのイベントが発生する確率（確率密度関数ＰＤＦ）を定義する関数を決定する必要がある。この種類の関係は、現在の状態Ｓｔｉと観測された信号の特徴をリンクするために使用される。好適なＰＤＦ機能は、ガウス混合モデルＧＭＭである。
【００４４】
幾つかの関係は、連続する単語Ｗｔｉのように決定論的及び蓋然論的の両方である。１つの単語から別の単語への遷移が発生しない場合、関係は決定論的であり、単語は時間ｔ－１でのものと同じである。遷移が発生する場合、次の単語Ｗｔｉ＋１は、言語モデルからの知識を用いる蓋然論的方法において決定される。
【００４５】
ＤＢＮにおける推論は、音響特徴の観測に基づいて影響する。しかし、あらゆる観測が測定誤差を受けやすい。同じグループ（例えば、図３内のＯＡ１１，ＯＡ２３，及びＯＡ３３，又はＯＶ１１及びＯＶ２３）に属する関係する時間－変数観測の間の蓋然論的関係の導入は、そのような誤差を減少することを可能にする。
【００４６】
状態Ｓｔｉ及び前の状態Ｓｔｉ－１は、観測が与えられた音素（図３におけるＰｔ１からＰｔ６）を話す結果である確率を評価するために使用される。
【００４７】
与えられた音素の発生は、一時的状態Ｐｔｒに確率的に関係もする。音素Ｐｔｉ、音素遷移Ｐｔｒ、単語Ｗｐｓ内の音素の位置、及び単語Ｗｔｒからの遷移は、記録された音が単語Ｗを含む仮説の正確さを評価することを可能とする。
【００４８】
音声は、特定の周波数特徴及びエネルギ特徴が短い期間内でほとんど一定である特性を有する。しかし、長い期間、それらは著しく変化する。それにもかかわらず、第１及び第２の状況が発生する特定の瞬間が定義されず、そのため、ＤＢＮモデルの使用が非常に有利である。異なるセグメント内の観測間の関係は、存在してもよいが、存在しなければならないものでもない。
【００４９】
例えば、４つの期間の構成の変形に対して、それらは、平行分析の５ｍｓ、２０ｍｓ、音素、及び単語を仮定してよい。例えば、すべての４つの範囲の間の関係があるが、５ｍｓ及び２０ｍｓの層と音素の層との間のみの関係がある、２０ｍｓの層と音素の層との間のみの関係がある、又は音素の層と単語の層との間のみの関係がある可能な異なるモデル構成がある。
【００５０】
更に、範囲のそれぞれは、音声の異なる種類の特徴に関係する幾つかの観測タイプを有する。例えば、それらの１つは周波数特徴ベクトルであり、別の１つはエネルギ及びさらに別の１つは視覚的特徴ベクトルであってもよい。これらは、同じ種類、しかし異なる方法（例えば、ＷＦＴ（ウェーブレットフーリエ変換）、ＭＦＣＣ（メル周波数ケプストラ

ム係数））を用いて得られる音響特徴であってもよいし、同じ方法を用いて、しかし異なる時間範囲に対して、例えば、２０ｍｓの移動ウィンドウに対して、５０ｍｓの移動ウィンドウに対して、１０ｍｓごとに抽出される両方に対して、得られる音響特徴であってもよい。

【００５１】

さらに、幾つかの範囲は、特定の種類の特徴の分析においてのみ発生し、他の種類では利用できない（図３、音響特徴１（３０８）の観測は最後の６０ｍｓ、音響特徴２（３１０）の観測は最後の２０ｍｓ、視覚特徴１（３０９）の観測は最後の３０ｍｓ）。

【００５２】

同時に、分析の際に使用される信号を記述するより多くのタイプの特徴、例えばピッチ周波数、フォルマント周波数、又は音の有声／無声説明があってもよい。

【００５３】

図２に提示された方法は、ステップ２０１で音声信号を取得して開始する。次のステップ２０２は、例えばＷＦＴ又は短時間フーリエ変換（ＳＴＦＴ）を用いる時間周波数変換により、信号を周波数領域に処理する。異なる時間の瞬間での異なる周波数サブバンドに含まれる情報（信号エネルギのような）の定量的記述を可能にする他の変換を適用することが可能である。

【００５４】

続いて、ステップ２０３では、時間周波数スペクトルは、例えば５ｍｓ、２０ｍｓ、６０ｍｓ等の一定のフレームに分割され、又は例えば以下に提示されるような所定のアルゴリズムに従ってセグメント化される。

P. Cardinal, G. Boulianne, and M. Comeau, "Segmentation of recordings based on partial transcriptions", Proceedings of Interspeech, pp. 3345-3348, 2005; or

K. Demuynck and T. Laureys, "A comparison of different approaches to automatic speech segmentation", Proceedings of the 5th International Conference on Text, Speech and Dialogue, pp. 277-284, 2002; or

Subramanya, J. Bilmes, and C. P. Chen, "Focused word segmentation for ASR", Proceedings of Interspeech 2005, pp. 393-396, 2005.

【００５５】

セグメント化モジュール（２０３）は、スペクトル分析の処理を独立にパラメータ化される複数のラインに分割する。

【００５６】

ラインの数は、前述の４と異なってもよい。図２の例は、５ｍｓ－２０４ａ、２０ｍｓ－２０４ｂ、音素－２０４ｃ、及び単語－２０４ｄのフレームを有する４つの別個のラインを採用する。ここで、ラインのそれぞれから、ブロック２０４ａから２０４ｄにおいて特定の時間での音声を表す特徴が抽出される。これらのパラメータ化ブロックは、ＭＦＣＣ、知覚線形予測（ＰＬＰ）又はその他以下のような処理アルゴリズムを採用してよい。

H. Misra, S. Ikbal, H. Bourlard, and H. Hermansky, "Spectral entropy based feature for robust ASR", Proceedings of ICASSP, pp. I-193-196, 2004; and/or

L. Deng, J. Wu, J. Droppo, and A. Acero, "Analysis and comparison of two speech feature extraction/compensation algorithms", IEEE Signal Processing Letters, vol. 12, no. 6, pp. 477-480, 2005; and/or

D. Zhu and K. K. Paliwal, "Product of power spectrum and group delay function for speech recognition", Proceedings of ICASSP, pp. I-125-128, 2004.

【００５７】

モジュール２０４ａから２０４ｄから得られる特徴は、信号エネルギ及び視覚的特徴ベクトルのような観測２０１ａとともにＤＢＮ２０５に通される。ビタビデコード及び／又はＢａｕｍ－Ｗｅｌｃｈのような音声認識において使用されるダイナミックプログラミングアルゴリズムを使用し、また辞書２０６のコンテンツと言語モデル２０７、例えば単語のバイグラムに基づくＢＮ、例えば変分メッセージ送信、期待プロパゲーション及び／又

はギブスサンプリングに対して近似推論のその埋め込まれたアルゴリズムを使用するＤＢＮモデルは、単語仮説を決定し、それらの確率を計算する。ほとんどの場合、ＤＢＮは同じ期間に異なる仮説を提示し得るため、仮説は部分的に重複し得る。仮説は、その後、認識音声テキスト２０９を得るために、さらなる言語モデル２０８（好ましくは、ＤＢＮで使用される第１の言語モデルより高度な）において処理されてもよい。

【００５８】

図３は、典型的なＤＢＮ構造を示す。アイテムＷ３０１は単語を意味し、Ｗｔｒ３０２は単語遷移を意味し、Ｗｐｓ３０３は特定の単語内の音素の位置を意味し、Ｐｔｒ３０４は音素遷移を意味し、Ｐｔ３０５は音素を意味し、Ｓｐｔ３０６は前の状態を意味し、Ｓ３０７は状態を意味し、ＯＡ１　３０８は６０ｍｓの時間ウィンドウにおける第１種の観測された音響特徴を意味し、ＯＶ１　３０９は３０ｍｓの時間ウィンドウにおける第１種の観測された視覚特徴を意味し、ＯＡ２　３１０は２０ｍｓの時間ウィンドウにおける第２種の観測された音響特徴を意味し、ＯＡ３　３１１は１０ｍｓの時間ウィンドウにおける第３種の観測された音響特徴を意味し、ＯＶ２　３１２は１０ｍｓの時間ウィンドウにおける第２種の観測された視覚特徴を意味する。

【００５９】

矢印は、前述の通り、変数間の関係（依存性）を表す。遷移は、トレーニングデータに基づいて、ベイジアンネットワークのトレーニング処理の間に計算される条件付き確率分布（ＣＰＤ）により定義される。

【００６０】

図４は、単語のシーケンスをデコードする、図３内に示されるＤＢＮの使用例を示す。図３の音声認識と異なり、信号の１種の音響特徴が異なる長さの２つのフレームに対して使用される。ネットワークは、フレーズ「Cat is black」、発音表記

【数１】

kæt ɪz blæk

のデコードの処理を与える。音素状態は、２種類の観測Ｏ１及びＯ２に依存する。時刻ｔでの前の状態３０６は、時刻ｔ－１での状態３０７の正確なコピーである。分析は、単語３０３内の現在位置、別の単語３０４への音素遷移の発生、音素の状態３０６及び前の状態３０７に応じて、単語３０１の次の音素に適用される。音素遷移は、遷移確率の値が０．５以上の場合に発生する。図３からベイジアンネットワークの別個のノードのシンボルは、これらの状態の値に置き換えられている。３０２及び３０４に対し、それらの値は、それぞれ、次の単語又は次の音素の間の遷移の発生又は発生なしを意味するＴ（Ｔｒｕｅ）／Ｆ（Ｆａｌｓｅ）である。単語３０３内の音素の位置に対して、それは、現在解析された音素（単語「ｃａｔ」に対して１－３、単語「ｉｓ」に対して１－２、単語「ｂｌａｃｋ」に対して１－４）のインデックスである。音素インデックスの変化は、時刻ｔ－１の前の瞬間において、音素３０４の遷移が値「Ｔ」を得たときにのみ発生する。更に、単語３０１は、特定の単語内の最後のインデックスから音素遷移３０４の瞬間で得られる単語遷移３０２の発生の場所でのみ変化する。次の単語間の関係は、そのような場合において、言語モデルを使用する結果として、決定論的から蓋然論的へ変化する。バイグラム言語モデル（単語のカップルを利用するモデル）の典型的な値は、図面の上の表に示される。更に、言語モデルにおける初期単語確率の典型的な値が提示されている。様々な持続時間及び幾つかの種類の特徴を用いてセグメントを同時に処理することにより実現される技術的効果は、音声認識の質を増大することである。なぜなら、様々な方法で話される音素の１つのタイプは、時間セグメントの１つのタイプで良く認識され、他は異なるタイプのセグメントを必要とするが、各種の音素に対して適当な分析時間ウィンドウを判断することは複雑であるからである。更に、幾つかの特徴は、よりローカルな時間セグメントでの情報の精密な抽出を可能にする定常的な特性を与えるとともに、他はよりグローバルな時間セグメントを必要とする。図３に示すような構造を使用することで、一度に両方の種類の特徴が抽出され得る。従来のシステムでは、ローカルな特徴によってのみ又はグローバルな特徴によってのみ運ばれる情報の断片が使用される。更に、例えば、視覚特徴は音

響特徴と異なる持続時間を有する、すなわち、例えば音を話すために向かい合う唇の観測は特定の音よりも長く又は短く続き得る。

【００６１】

上記の音声認識方法は、１又は複数のコンピュータプログラムにより実行され得る及び／又は制御され得ることは、当業者により容易に認識できる。そのようなコンピュータプログラムは、通常、パーソナルコンピュータ、携帯用情報端末、携帯電話、デジタルテレビの受信機及びデコーダ、インフォメーションキオスク等のような演算デバイスにおける演算リソースを利用することに実行される。アプリケーションは、不揮発性メモリ、例えばフラッシュメモリ又は揮発性メモリ、例えばＲＡＭに格納され、プロセッサにより実行される。これらのメモリは、本明細書に提示される技術的思想に従ってコンピュータ実装方法のステップのすべてを実行するコンピュータ実行可能命令を備えるコンピュータプログラムを格納する典型的な記録媒体である。

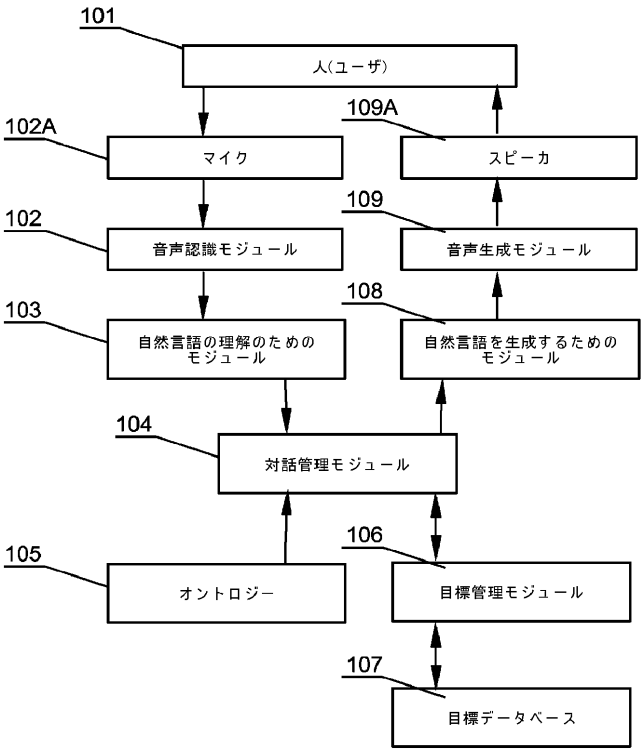
【００６２】

本明細書に提示された発明が示され、記述され、特定の好ましい実施形態を参照して定義されているが、前述の明細書におけるそのような参照及び実施例はいかなる本発明の限定を意味するものではない。しかし、様々な修正及び変更が技術的思想のより広い範囲から逸脱することなくなされ得ることは明らかである。提示された好ましい実施形態は単なる典型であり、本明細書に提示された技術的思想の範囲を網羅するものではない。

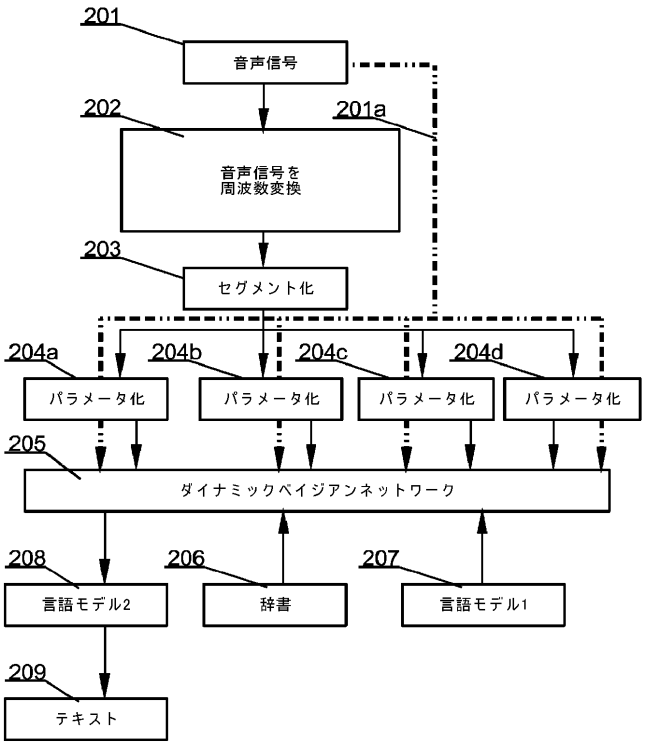
【００６３】

従って、保護の範囲は、本明細書に記載された好ましい実施形態に限定されるものではなく、続く特許請求の範囲によってのみ限定される。

【図１】



【図２】



【図３】

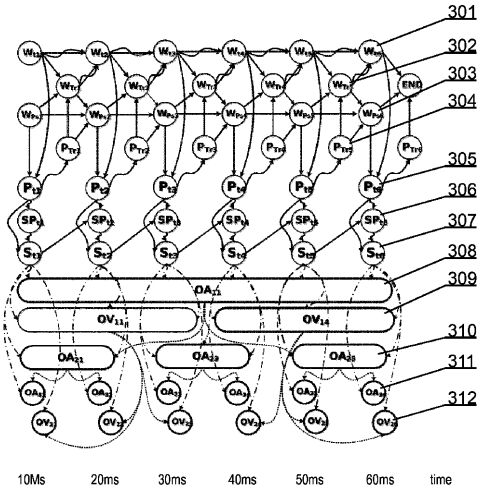


Fig. 3

【図４】

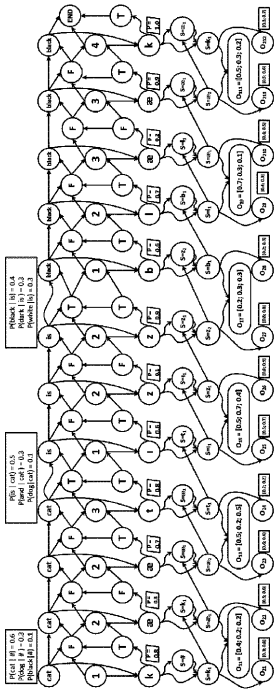


Fig. 4

【手続補正書】
【提出日】平成28年4月8日(2016.4.8)
【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】
音声認識のコンピュータ実装方法であって、
入力デバイスにより、音声を表す電気信号を登録し、前記電気信号を周波数または時間周波数領域に変換する段階と、
ダイナミックベイジアンネットワークに基づいて解析モジュール内の前記電気信号を分析する段階であり、それにより、複数の単語の複数の仮説および観測された複数の音響信号特徴または複数の音響及び視覚信号特徴に基づくそれらの確率を生成する、段階と、
特定の複数の単語仮説およびそれらの確率に基づいて、音声を表す前記電気信号に対応するテキストを認識する段階と、
前記認識する段階の間、前記解析モジュールに、少なくとも２つの並列信号処理ラインにおける周波数または時間周波数領域内の前記電気信号に対して決定される前記観測された複数の音響信号特徴または複数の音響及び視覚信号特徴を含む観測された複数の信号特徴を入力する段階であり、前記少なくとも２つの並列信号処理ラインのうちの少なくとも２つのラインは複数の音響信号特徴を決定し、少なくとも１つのラインは他のラインと異なる方法でセグメント化された信号を受信し、それにより、少なくとも１つのラインの少なくとも幾つかのフレームが前記ダイナミックベイジアンネットワークの異なる状態間で共有されることができる、段階と、

前記解析モジュールにおいて、少なくとも２つの別個の時間セグメントに対して前記観測された複数の信号特徴の間の複数の関係を分析する段階と、
を備える、コンピュータ実装方法。
【請求項２】
前記少なくとも２つの別個の時間セグメントは、所定の継続時間を有する、請求項１に記載のコンピュータ実装方法。
【請求項３】
前記少なくとも２つの別個の時間セグメントは、複数の音素、複数の音節、複数の単語のような複数の音声セグメントのコンテンツに依存する、請求項１または２に記載のコンピュータ実装方法。
【請求項４】
前記解析モジュールにおいて、モデルを記述する複数の変数の間の複数の決定論的及び蓋然論的関係を定義する段階をさらに備え、複数の前記蓋然論的関係は、少なくとも観測された前記複数の信号特徴を現在の状態にリンクするために定義される、請求項１から３のいずれか一項に記載のコンピュータ実装方法。
【請求項５】
前記複数の関係を分析する段階は、異なる観測された複数の信号特徴を同時に分析する段階を含む、請求項１から４のいずれか一項に記載のコンピュータ実装方法。
【請求項６】
音声認識のコンピュータ実装システムであって、
音声を表す電気信号を登録する入力デバイスと、
音声を表す登録された前記電気信号を周波数または時間周波数領域に変換するモジュールと、
音声を表す前記電気信号を分析し、複数の単語の複数の仮説および観測された複数の音響信号特徴または複数の音響及び視覚信号特徴に基づくそれらの確率を生成するダイナミックベイジアンネットワークに基づく解析モジュールと、
複数の単語の定義された前記複数の仮説及びそれらの確率に基づいて、音声を表す前記電気信号に対応するテキストを認識するモジュールと、
前記認識するモジュールが認識する間、前記解析モジュールに対して、少なくとも２つの並列信号処理ラインにおける少なくとも２つの観測された音響信号特徴または複数の音響及び視覚信号特徴を含む観測された複数の信号特徴を決定するための少なくとも２つの信号パラメータ化モジュールであり、前記少なくとも２つの並列信号処理ラインのうちの少なくとも２つのラインは複数の音響信号特徴を決定し、少なくとも１つのラインは他のラインと異なる方法でセグメント化された信号を受信し、それにより、少なくとも１つのラインの少なくとも幾つかのフレームが前記ダイナミックベイジアンネットワークの異なる状態間で共有されることができる、少なくとも２つの信号パラメータ化モジュールと、
を備え、前記解析モジュールは、少なくとも２つの別個の時間セグメントに対して前記観測された複数の信号特徴の間の複数の依存性を分析する、コンピュータ実装システム。
【請求項７】
コンピュータ上で実行されると、請求項１から５のいずれか一項に記載のコンピュータ実装方法のすべての段階を前記コンピュータに実行させるコンピュータプログラム。
【請求項８】
コンピュータ上で実行されると、請求項１から５のいずれか一項に記載のコンピュータ実装方法のすべての段階を前記コンピュータに実行させるコンピュータプログラムを格納するコンピュータ可読媒体。
【手続補正２】
【補正対象書類名】明細書
【補正対象項目名】００６３
【補正方法】変更
【補正の内容】

【００６３】
従って、保護の範囲は、本明細書に記載された好ましい実施形態に限定されるものではなく、続く特許請求の範囲によってのみ限定される。
本明細書によれば、以下の各項目に記載の構成もまた開示される。
［項目１］
音声認識のコンピュータ実装方法であって、
入力デバイス（１０２Ａ）により、音声を表す電気信号を登録し、前記電気信号を周波数または時間周波数領域（２０２）に変換する段階（２０１）と、
ダイナミックベイジアンネットワーク（２０５）に基づいて解析モジュール内の前記電気信号を分析する段階であり、それにより、複数の単語（Ｗ）の複数の仮説および観測された複数の信号特徴（ＯＡ，ＯＶ）に基づくそれらの確率を生成する、段階と、
特定の複数の単語（Ｗ）仮説およびそれらの確率に基づいて、音声を表す前記電気信号に対応するテキストを認識する段階（２０９）と、
前記解析モジュール（２０５）に、各ラインに対して別個の複数の時間セグメントに対する少なくとも２つの並列信号処理ライン（２０４ａ、２０４ｂ、２０４ｃ、２０４ｄ、２０１ａ）における周波数または時間周波数領域（２０２）内の前記電気信号に対して決定される観測された複数の信号特徴（３０８－３１２）を入力する段階と、
前記解析モジュール（２０５）において、少なくとも２つの別個の時間セグメントに対して観測された前記複数の信号特徴（３０８－３１２）の間の複数の関係を分析する段階と、
を備える、コンピュータ実装方法。
［項目２］
前記複数の時間セグメントは、所定の継続時間を有する、項目１に記載のコンピュータ実装方法。
［項目３］
前記複数の時間セグメントは、複数の音素、複数の音節、複数の単語のような複数の音声セグメントのコンテンツに依存する、項目１または２に記載のコンピュータ実装方法。
［項目４］
前記解析モジュール（２０５）において、モデルを記述する複数の変数の間の複数の決定論的及び蓋然論的関係を定義する段階をさらに備え、複数の前記蓋然論的関係は、少なくとも観測された前記複数の信号特徴を現在の状態（Ｓｔｉ）にリンクするために定義される、項目１から３のいずれか一項に記載のコンピュータ実装方法。
［項目５］
前記複数の関係を分析する段階は、異なる観測された複数の信号特徴（ＯＡ、ＯＶ）を同時に分析する段階（２０５）を含む、項目１から４のいずれか一項に記載のコンピュータ実装方法。
［項目６］
音声認識のコンピュータ実装システムであって、
音声を表す電気信号を登録する入力デバイス（１０２Ａ）と、
音声を表す登録された前記電気信号を周波数または時間周波数領域に変換するモジュール（２０２）と、
音声を表す前記電気信号を分析し、複数の単語（Ｗ）の複数の仮説および観測された複数の信号特徴（ＯＡ，ＯＶ）に基づくそれらの確率を生成するダイナミックベイジアンネットワークに基づく解析モジュール（２０５）と、
複数の単語（Ｗ）の定義された前記複数の仮説及びそれらの確率に基づいて、音声を表す前記電気信号に対応するテキストを認識するモジュール（２０９）と、
各ラインに対して別個の複数の時間セグメントに対する少なくとも２つの並列信号処理ラインにおいて、少なくとも２つの観測された信号特徴（３０８－３１２）を、前記解析モジュール（２０５）に対して決定する少なくとも２つの信号パラメータ化モジュール（２０４ａ、２０４ｂ、２０４ｃ、２０４ｄ、２０１ａ）と、

を備え、前記解析モジュール（２０５）は、少なくとも２つの別個の時間セグメントに対して観測された前記複数の信号特徴（３０８－３１２）の間の複数の依存性を分析する、コンピュータ実装システム。

［項目７］

コンピュータ上で実行されると、項目１から５のいずれか一項に記載のコンピュータ実装方法のすべての段階を実行するプログラムコード化手段を備えるコンピュータプログラム。

［項目８］

コンピュータ上で実行されると、項目１から５のいずれか一項に記載のコンピュータ実装方法のすべての段階を実行する複数のコンピュータ実行可能命令を格納するコンピュータ可読媒体。音声認識システム及びダイナミックベイジアンネットワークモデルの使用方法。

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2013/063330

A. CLASSIFICATION OF SUBJECT MATTER
INV. G10L15/14
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G10L G06N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, COMPENDEX, INSPEC, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	TODD A STEPHENSON HERVE BOURLARD SAMY BENGIO ANDREW C MORRIS DALLE MOLLE INSTITUTE FOR PERCEPTUAL ARTIFICIAL INTELLIGENCE (IDIAP): "AUTOMATIC SPEECH RECOGNITION USING DYNAMIC BAYESIAN NETWORKS WITH BOTH ACOUSTIC AND ARTICULATORY VARIABLES", 6TH. INTERNATIONAL CONFERENCE OF SPOKEN LANGUAGE PROCESSING. ICSLP 2000. BEIJING, CHINA, OCT. 16 - 20, 2000; [INTERNATIONAL CONFERENCE OF SPOKEN LANGUAGE PROCESSING], CENTER FOR SPOKEN LANGUAGE RESEARCH, 16 October 2000 (2000-10-16), XP007010757, ISBN: 978-7-80150-144-8	1,2,4,5
Y	the whole document	3
	----- -/--	

[X] Further documents are listed in the continuation of Box C. [X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
8 April 2014	15/04/2014

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Deltorn, Jean-Marc

1

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2013/063330

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	T.J. HAZEN: "Visual model structures and synchrony constraints for audio-visual speech recognition", IEEE TRANSACTIONS ON AUDIO, SPEECH AND LANGUAGE PROCESSING, vol. 14, no. 3, 1 May 2006 (2006-05-01), pages 1082-1089, XP055112509, ISSN: 1558-7916, DOI: 10.1109/TSA.2005.857572 the whole document	1-5
X	GOWDY J ET AL: "DBN based multi-stream models for audio-visual speech recognition", ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, 2004. PROCEEDINGS. (ICASSP ' 04). IEEE INTERNATIONAL CONFERENCE ON MONTREAL, QUEBEC, CANADA 17-21 MAY 2004, PISCATAWAY, NJ, USA,IEEE, PISCATAWAY, NJ, USA, vol. 1, 17 May 2004 (2004-05-17), pages 993-996, XP010717798, DOI: 10.1109/ICASSP.2004.1326155 ISBN: 978-0-7803-8484-2 sections 2 and 3	1-5
X	XAVIER DOMONT ET AL: "Hierarchical spectro-temporal features for robust speech recognition", ACOUSTICS, SPEECH AND SIGNAL PROCESSING, 2008. ICASSP 2008. IEEE INTERNATIONAL CONFERENCE ON, IEEE, PISCATAWAY, NJ, USA, 31 March 2008 (2008-03-31), pages 4417-4420, XP031251577, ISBN: 978-1-4244-1483-3 sections 1, 3.2 and 4.2	1-5
X	KATE SAENKO ET AL: "AN ASYNCHRONOUS DBN FOR AUDIO-VISUAL SPEECH RECOGNITION", SPOKEN LANGUAGE TECHNOLOGY WORKSHOP, 2006. IEEE, IEEE, PI, 1 December 2006 (2006-12-01), pages 154-157, XP031056424, ISBN: 978-1-4244-0872-6 sections 2 to 4	1-5
Y	STÉPHANE DUPONT ET AL: "Audio-Visual Speech Modeling for Continuous Speech Recognition", IEEE TRANSACTIONS ON MULTIMEDIA, IEEE SERVICE CENTER, PISCATAWAY, NJ, US, vol. 2, no. 3, 1 September 2000 (2000-09-01), XP011036218, ISSN: 1520-9210	3
A	section III.B	1,2,4,5
	-/--	

Form PCT/ISA/210 (continuation of second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2013/063330

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	ASTRID HAGEN ET AL: "USING MULTIPLE TIME SCALES IN THE FRAMEWORK OF MULTI-STREAM SPEECH RECOGNITION", INTERNATIONAL CONF. ON SPOKEN LANGUAGE PROCESSING, BEIJING 2000, 16 October 2000 (2000-10-16), XP007011060, the whole document -----	1-5
A	US 6 292 776 B1 (CHENGALVARAYAN RATHINAVELU [US]) 18 September 2001 (2001-09-18) claims 1-12 -----	1-5

1

Form PCT/ISA/210 (continuation of second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2013/063330

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 6292776 B1	18-09-2001	CA 2299051 A1	12-09-2000
		DE 60000074 D1	28-03-2002
		DE 60000074 T2	29-08-2002
		EP 1041540 A1	04-10-2000
		JP 3810608 B2	16-08-2006
		JP 2000267692 A	29-09-2000
		US 6292776 B1	18-09-2001

Form PCT/ISA/210 (patent family annex) (April 2005)

フロントページの続き

(81)指定国 AP(BW,GH,GM,KE,LR,LS,MW,MZ,NA,RW,SD,SL,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,RU,TJ,TM),EP(AL,AT,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR,GB,GR,HR,HU,IE,IS,IT,LT,LU,LV,MC,MK,MT,NL,NO,PL,PT,RO,RS,SE,SI,SK,SM,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,KM,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AO,AT,AU,AZ,BA,BB,BG,BH,BN,BR,BW,BY,BZ,CA,CH,CL,CN,CO,CR,CU,CZ,DE,DK,DM,DO,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,GT,HN,HR,HU,ID,IL,IN,IS,JP,KE,KG,KN,KP,KR,KZ,LA,LC,LK,LR,LS,LT,LU,LY,MA,MD,ME,MG,MK,MN,MW,MX,MY,MZ,NA,NG,NI,NO,NZ,OM,PA,PE,PG,PH,PL,PT,QA,RO,RS,RU,RW,SC,SD,SE,SG,SK,SL,SM,ST,SV,SY,TH,TJ,TM,TN,TR,TT,TZ,UA,UG,US,UZ,VC

(72)発明者 ジャドツク、トマス
ポーランド共和国、ピーエル－３０－０５９ クラクフ エーエル. ミツキエビツァ ３０ アカデミア ゴルニツォーハットニツァ アイエム. スタニスラワ スタシツァ ダブリュー クラクフィ内