

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4729524号
(P4729524)

(45) 発行日 平成23年7月20日 (2011. 7. 20)

(24) 登録日 平成23年4月22日 (2011. 4. 22)

(51) Int. Cl.		F I	
HO4N	5/91	(2006.01)	HO4N 5/91 Z
G1OL	15/00	(2006.01)	G1OL 15/00 200G
HO4N	5/76	(2006.01)	HO4N 5/76 B
G11B	20/10	(2006.01)	G11B 20/10 311
HO4N	5/92B	(2006.01)	HO4N 5/92 E

請求項の数 26 (全 15 頁)

(21) 出願番号	特願2007-77380 (P2007-77380)	(73) 特許権者	505473824
(22) 出願日	平成19年3月23日 (2007. 3. 23)		シュトルツ・エンドスコープ・プロダク
(65) 公開番号	特開2008-72685 (P2008-72685A)		ツィオンス・ゲーエムベーハー
(43) 公開日	平成20年3月27日 (2008. 3. 27)		ドイツ・D-78532・トゥットリンゲ
審査請求日	平成19年3月23日 (2007. 3. 23)		ン・ミッテルシュトラセ・8
(31) 優先権主張番号	11/519, 315	(74) 代理人	100108453
(32) 優先日	平成18年9月12日 (2006. 9. 12)		弁理士 村山 靖彦
(33) 優先権主張国	米国 (US)	(74) 代理人	100064908
前置審査			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】 リアルタイムの音声認識コマンドおよびコントロールシステムを備えたオーディオデータ、ビジュアルデータ、およびデバイスデータの取り込みシステム

(57) 【特許請求の範囲】

【請求項1】

オーディオデータおよびビジュアルデータの取り込みシステムであって、
複数のユーザの発声を含んだオーディオデータを記録するためのオーディオレコーダと、

ビジュアルデータを記録するための少なくとも1つのビジュアルレコーダと、
前記オーディオデータを解釈するための音声認識モジュールと、
解釈されたオーディオデータを、動的な認識文法に基づいて、転写のためのオーディオデータと、システムと通信状態にある少なくとも1つの医療用デバイスをコントロールするためのオーディオデータとに区別するための弁別器と、

前記弁別器によって転写のためのオーディオデータであると識別された前記解釈されたオーディオデータから転写データを生成するための転写モジュールと、

前記オーディオデータ、前記転写データ、および前記ビジュアルデータのそれぞれの少なくとも一部を含むデータレコードを生成するためのデータ取り込みモジュールと、

前記データレコードを保存するための少なくとも1つのストレージデバイスとを具備し、

前記認識文法は、デバイスの状態の変化とコマンドメニューの変化とのうちの少なくとも1つに基づいて更新されることを特徴とするシステム。

【請求項2】

システムと通信状態にある少なくとも1つのデバイスからデバイスデータを受信するた

めの少なくとも1つのデバイスデータレコーダをさらに具備し、

前記データレコードが、前記デバイスデータの少なくとも一部をさらに含み、

前記データ取り込みモジュールが、前記データレコード内の前記オーディオデータ、前記転写データ、前記ビジュアルデータ、および前記デバイスデータを同期化することを特徴とする請求項1に記載のシステム。

【請求項3】

前記データ取り込みモジュールが、前記データレコード内の前記オーディオデータ、前記転写データ、および前記ビジュアルデータを同期化することを特徴とする請求項1に記載のシステム。

【請求項4】

前記オーディオデータが、前記ビジュアルデータの1つまたは複数の部分に対応する1つまたは複数の音声による注釈を含むことを特徴とする請求項1に記載のシステム。

【請求項5】

前記少なくとも1つのストレージデバイスが、複数のデータレコードを含み、前記システムが、前記転写データに基づいて前記複数のデータレコードの問合せを行うためのソフトウェアを具備することを特徴とする請求項1に記載のシステム。

【請求項6】

前記オーディオレコーダが前記オーディオデータを記録している間に、前記転写モジュールが前記転写データを継続的に生成することを特徴とする請求項1に記載のシステム。

【請求項7】

前記少なくとも1つのストレージデバイスが、データベースを含み、前記データレコードが、前記転写データの少なくとも一部に基づいて前記データベースの場所に保存されることを特徴とする請求項1に記載のシステム。

【請求項8】

前記少なくとも1つのストレージデバイスが、医療記録データベースを含むことを特徴とする請求項1に記載のシステム。

【請求項9】

前記少なくとも1つのストレージデバイスが、ポータブルメディアドライブを含むことを特徴とする請求項1に記載のシステム。

【請求項10】

前記少なくとも1つのビジュアルレコーダが、動画レコーダおよび静止画像レコーダを含み、前記ビジュアルデータが、静止画像データおよび動画データの双方を含むことを特徴とする請求項1に記載のシステム。

【請求項11】

音声認識およびコントロールを備えたオーディオデータおよびビジュアルデータの取り込みシステムであって、

複数のユーザの発声を含んだオーディオデータを記録するためのオーディオレコーダと、

ビジュアルデータを記録するための少なくとも1つのビジュアルレコーダと、

前記オーディオデータを解釈するための音声認識モジュールと、

解釈されたオーディオデータを、動的な認識文法に基づいて、転写のためのオーディオデータと、システムと通信状態にある1つまたは複数の医療用デバイスをコントロールするためのオーディオデータとに区別するための弁別器と、

前記弁別器によって転写のためのオーディオデータであると識別された前記解釈されたオーディオデータから転写データを生成するための転写モジュールと、

前記弁別器によってシステムと通信状態にある1つまたは複数の医療用デバイスをコントロールするためのオーディオデータであると識別された前記解釈されたオーディオデータの少なくとも一部に基づいてコントロールコマンドを生成するためのコントローラと、

前記オーディオデータ、前記ビジュアルデータ、および前記転写データのそれぞれの少なくとも一部を含むデータレコードを生成するためのデータ取り込みモジュールと、

10

20

30

40

50

前記データレコードを保存するための少なくとも1つのストレージデバイスとを具備し、

前記認識文法は、デバイスの状態の変化とコマンドメニューの変化とのうちの少なくとも1つに基づいて更新されることを特徴とするシステム。

【請求項12】

前記データ取り込みモジュールが、前記オーディオデータ、前記ビジュアルデータ、および前記転写データを同期化することを特徴とする請求項11に記載のシステム。

【請求項13】

前記コントローラが、前記システムによって機能できるデバイスへ前記コントロールコマンドをさらに送信することを特徴とする請求項11に記載のシステム。

10

【請求項14】

前記システムによって機能できる前記デバイスが、医療用デバイスであることを特徴とする請求項13に記載のシステム。

【請求項15】

前記システムによって機能できる前記デバイスが、前記オーディオレコーダおよび前記少なくとも1つのビジュアルレコーダのうちの1つであることを特徴とする請求項13に記載のシステム。

【請求項16】

オーディオデータ、ビジュアルデータ、およびデバイスデータを処理する方法であって、

20

オーディオデータを記録するステップであって、前記オーディオデータが、1つまたは複数のユーザの発声を含むステップと、

ビジュアルデータおよびデバイスデータのうちの少なくとも1つを記録するステップと、

前記1つまたは複数のユーザの発声を解釈し、解釈されたユーザの発声を、動的な認識文法に基づいて、転写のためのユーザの発声と、少なくとも1つの医療用デバイスをコントロールするためのユーザの発声とに区別するステップと、

転写用であると識別されたユーザの発声から転写データを生成するステップと、

前記オーディオデータおよび前記転写データのそれぞれの少なくとも一部を含むデータレコードを生成するステップと、

30

前記データレコードを保存するステップとを有し、

前記認識文法は、デバイスの状態の変化とコマンドメニューの変化とのうちの少なくとも1つに基づいて更新されることを特徴とする方法。

【請求項17】

前記データレコード内の前記オーディオデータおよび前記転写データを同期化するステップをさらに有することを特徴とする請求項16に記載の方法。

【請求項18】

前記データレコードが、前記ビジュアルデータをさらに含む方法であって、前記データレコード内の前記オーディオデータ、前記転写データ、および前記ビジュアルデータを同期化するステップをさらに含むことを特徴とする請求項16に記載の方法。

40

【請求項19】

前記データレコードが、前記デバイスデータをさらに含む方法であって、前記データレコード内の前記オーディオデータ、前記転写データ、および前記デバイスデータを同期化するステップをさらに含むことを特徴とする請求項16に記載の方法。

【請求項20】

区別する前記ステップが、認識文法の問合せを行うステップを含む方法であって、デバイスの状態の変化およびコマンドメニューの変化のうちの1つに基づいて前記認識文法を更新するステップをさらに有することを特徴とする請求項16に記載の方法。

【請求項21】

コントロールコマンドとして意図されている前記ユーザの発声を実行のためにコマンド

50

コントローラへ送信するステップをさらに有することを特徴とする請求項16に記載の方法。

【請求項22】

コントロールコマンドを、前記コントロールコマンドによって機能できるデバイスへ送信するステップをさらに有することを特徴とする請求項21に記載の方法。

【請求項23】

少なくとも1つの内容に関する用語(content term)を含む検索要求を受信するステップと、

前記少なくとも1つの内容に関する用語を探して、前記データレコードの前記転写データの前記少なくとも一部に問合せを行うステップと、

前記少なくとも1つの内容に関する用語に対応する前記データレコードの一部を提供するステップとをさらに有することを特徴とする請求項16に記載の方法。

【請求項24】

前記ユーザの発声のうちの少なくとも1つが、前記ビジュアルデータの一部に対応することを特徴とする請求項16に記載の方法。

【請求項25】

前記データレコードが、前記転写データに基づいてデータベースの場所に保存されることを特徴とする請求項16に記載の方法。

【請求項26】

前記解釈されたオーディオデータに基づいてコントロールコマンドを生成して、前記コントロールコマンドを、前記コントロールコマンドによって機能できるデバイスへ送信するステップをさらに有することを特徴とする請求項16に記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、オーディオデータ、ビジュアルデータ、およびデバイスデータの取り込み(capture)に関し、より具体的には、リアルタイムのデータ転写(transcription)のための音声認識(speech recognition)を含むオーディオデータ、ビジュアルデータ、およびデバイスデータの取り込みシステムに関する。さらにいっそう具体的には、本発明は、1つのコマンドおよびコントロールシステムへと統合されたリアルタイムのデータ転写のための音声認識を含む、医療用途、技術用途、および科学用途のためのオーディオビジュアルデータおよびデバイスデータの取り込みに関する。

【背景技術】

【0002】

オーディオビジュアルデータの取り込みは、手術室コントロールシステムなど、何らかの命令およびコントロールシステムにおいて実施されてきた。例えば、Karl Storz AIDA(商標)(Advanced Image and Data Archiving System)などのデバイスによって、外科医は、外科処置中に得られたビデオストリーム信号や静止画像を取り込むことができる。画像や映像の録画には、手術を行っている外科医などのユーザによって、さらに注釈を付けることができる。一般に、音声による注釈が録音され、それらの録音は、取り込まれた静止画像や動画と共に保存される。取り込まれたオーディオデータおよびビジュアルデータは一般に、データベース内に、またはポータブルメディア(例えば、コンパクトディスク)上に保存される。

【0003】

医療分野においては、取り込まれたオーディオデータおよびビジュアルデータは、患者および/または患者に提供されている治療に関連する重要な情報を含むことができる。例えば、取り込まれたオーディオデータおよびビジュアルデータは、特定の患者に関するさらなる医療情報を入手するために、および治療に関する決定を行うために、手術後に使用することができる。オーディオデータおよびビジュアルデータは、研修の目的で、および潜在的な賠償訴訟原因に備えて医療処置を記録するために使用することもできる。しかし

10

20

30

40

50

、従来技術のオーディオデータおよびビジュアルデータの取り込みシステムには、限界がある。

【0004】

従来技術のオーディオおよびビジュアルの取り込みシステムの1つの不利な点は、生のオーディオビジュアルデータ、すなわち視覚的な像および音声の録音しか、取り込まれ保存されないという点である。したがって、データの有用性は非常に限られている。例えば、特定の主題に関連した何らかの所望のデータを見つけ出すためには、ユーザは、録画を再生し、所望の情報のすべての事例を探して耳を傾けなければならない、および/または目を凝らさなければならない。これは、所望のデータを入手する上でコスト効率が悪く、かつ不正確な手段である。したがって、非常に有用なデータを提供する改良されたオーディオおよびビジュアルの取り込みシステムを提供することが望ましい。

10

【0005】

従来技術のシステムのさらなる不利な点は、デバイスの状態に関する情報が、あらゆる関連するまたは所望のオーディオビジュアルデータと共に取り込まれず、同期化されず、保存されないという点である。このことが非常に望まれる例は、手術または処置中に、医療専門家が、異常出血やその他のそのような複雑な事態など、特定の出来事を記録したいと希望する場合である。ビジュアルデータと一緒に取り込まれ、転写され、同期化されるオーディオデータと共に、関連付けられている(1つまたは複数の)医療用デバイスによってモニターされているものなどの患者のバイタルサインを取り込んで同期化することもできる。さらに、ポンプ圧および減圧などの医療用デバイスの作動状況に関するデータ、患者が受けている投薬レベル、麻酔デバイスの設定などを、オーディオビジュアルデータと共に取り込んで同期化することができる。同様に、技術的および科学的な取り組みも、研究開発、テスト、および/またはシステムのモニタリングの取り組み中にオーディオデータ、ビジュアルデータ、およびデバイスデータをリアルタイムに取り込んで同期化することから恩恵を享受する。したがって、オーディオデータ、ビジュアルデータ、およびデバイスデータを取り込むためのシステムを提供することが望ましい。

20

【0006】

従来技術のシステムの別の不利な点は、音声認識システムを介して機器を動的にまたは同時にコントロールすることを、アドオン(add-on)のオーディオ転写サブシステム(audio transcription sub-system)を用いて実現することが難しいという点である。この難しさが生じるのは、機器をコントロールするためのコマンドとしてシステムのユーザによって意図されているユーザの発声が、リアルタイムベースで取り込まれ、同期化され、転写され、保存されることを意図されているオーディオデータであると誤って解釈される場合である。さらに、アドオンのオーディオ転写サブシステムには、会話の新しい話題を学習して、それに適合する十分な能力が欠けている。典型的なシステムやアプリケーションにおいては、言語や話題の変化は、非常に頻繁に起こり得る。変化は、時間をかけてゆっくりと、あるいはデバイスがシステムに追加される場合や、システムから取り除かれる場合のように突然に起こる可能性がある。従来技術のシステムは、これらの変化に正しく適合することができず、したがって高い初期転写エラー率を有する。

30

【0007】

したがって非常に望ましいのは、機器に対するコマンドおよびコントロールとして意図されているユーザの発声を、その他のデータと共に取り込まれ、同期化され、転写され、保存されることを意図されているオーディオデータから区別する音声認識コマンド(speech recognition command)およびコントロールシステムである。

40

【特許文献1】米国特許出願公開第2004/0172011号明細書

【発明の開示】

【発明が解決しようとする課題】

【0008】

したがって、本発明の1つの目的は、音声認識コマンドおよびコントロールシステムにおいて音声による注釈のリアルタイムな転写物を生成するためのオーディオデータおよび

50

ビジュアルデータの取り込みシステムを提供することである。

【0009】

本発明のさらなる目的は、転写物を、視覚的な像、オーディオ録音、デバイスの状態に関するデータなどを含むその他の形態のデータと同期化するシステムを提供することである。

【0010】

本発明のさらなる目的は、オーディオデータ、ビジュアルデータ、およびデバイスデータの取り込みアプリケーションを含む音声認識コマンドおよびコントロールシステムを提供することである。

【0011】

またさらに、本発明の1つの目的は、(1つまたは複数の)デバイスに対するコマンドおよびコントロールとして意図されているユーザの発声を、その他のデータと共に取り込まれ、同期化され、転写され、保存されることを意図されているオーディオデータから区別する音声認識コマンドおよびコントロールシステムを提供することである。

【0012】

またさらに、1つの目的は、会話の話題や言語における変化に適合して、それに応じて認識文法を更新するための非常に動的なコマンドおよびコントロール機能を備えた音声認識およびコマンドコントロールシステムを提供することである。

【課題を解決するための手段】

【0013】

これらおよびその他の目的は、オーディオデータを記録するためのオーディオレコーダと、ビジュアルデータを記録するための少なくとも1つのビジュアルレコーダと、オーディオデータを解釈するための音声認識モジュールと、解釈されたオーディオデータから転写データを生成するための転写モジュールと、オーディオデータ、転写データ、およびビジュアルデータのそれぞれの少なくとも一部を含むデータレコードを生成するためのデータ取り込みモジュールと、データレコードを保存するための少なくとも1つのストレージデバイスとを含む、オーディオデータおよびビジュアルデータの取り込みシステムを提供することによって、達成される。いくつかの実施形態においては、このシステムは、このシステムと通信状態にある少なくとも1つのデバイスからデバイスデータを受信するための少なくとも1つのデバイスデータレコーダをさらに含み、データレコードは、デバイスデータの少なくとも一部をさらに含む。

【0014】

さらに提供されるのは、オーディオデータを記録するためのオーディオレコーダと、ビジュアルデータを記録するための少なくとも1つのビジュアルレコーダと、オーディオデータを解釈して、転写データを生成するための音声認識モジュールと、解釈されたオーディオデータの少なくとも一部に基づいてコントロールコマンドを生成するためのコントローラと、オーディオデータ、ビジュアルデータ、および転写データのそれぞれの少なくとも一部を含むデータレコードを生成するためのデータ取り込みモジュールと、データレコードを保存するための少なくとも1つのストレージデバイスとを含む、音声認識およびコントロールを備えたオーディオデータおよびビジュアルデータの取り込みシステムである。いくつかの実施形態においては、このシステムは、コントロールコマンドとして意図されているオーディオデータ内のユーザの発声と、転写データとして意図されているユーザの発声とを区別するための弁別器をさらに含む。

【0015】

さらに提供されるのは、1つまたは複数のユーザの発声を含むオーディオデータを記録するステップと、ビジュアルデータおよびデバイスデータのうちの少なくとも1つを記録するステップと、1つまたは複数のユーザの発声を解釈するステップと、転写データを生成するステップと、オーディオデータおよび転写データのそれぞれの少なくとも一部を含むデータレコードを生成するステップと、データレコードを保存するステップとを含む、オーディオデータ、ビジュアルデータ、およびデバイスデータを処理する方法である。1

10

20

30

40

50

つまたは複数のユーザの発声を解釈するステップは、コントロールコマンドとして意図されているユーザの発声と、転写を意図されているユーザの発声とを区別するステップを任意選択で含む。

【発明を実施するための最良の形態】

【0016】

図1は、本発明によるオーディオデータ、ビジュアルデータ、およびデバイスデータの取り込みシステムを示している。このシステムは、例えば医療手術室の環境におけるオーディオデータ、ビジュアルデータ、および/またはデバイスデータの取り込みを含む任意の数の用途にとって有用なものとなることができる。

【0017】

このシステムは、1つまたは複数のビジュアルレコーダ102を含む。ビジュアルレコーダ102は、例えば、ビデオ映像を受信して録画するためのビデオレコーダとすることができる。ビジュアルレコーダ102は、静止画像レコーダとすることもできる。例えばビジュアルレコーダ102は、デジタルカメラや、静止画像または写真を録画するまたは取り込むためのその他の任意のデバイスとすることができる。ビジュアルレコーダ102は、任意の形状やサイズとすることができる。いくつかの実施形態においては、ビジュアルレコーダ102は、マイクロカメラである。ビジュアルレコーダ102は、手持ち型、取り付け型、および/またはデバイス内蔵型とすることもできる。例えばビジュアルレコーダ102は、内視鏡などの医療用デバイスの中や上に配置することができる。

【0018】

ビジュアルレコーダ102を使用して、任意の視覚的な像やイメージを取り込むことができる。例えばビジュアルレコーダ102は、外科処置のビデオ映像、および/または外科処置の静止画像を取り込むことができる。ビジュアルレコーダ102は、ビジュアルデータ104をシステムあるいはそのモジュールまたはプロセッサ110へ送信する。ビジュアルレコーダ102は、継続的に、および/またはリアルタイムで、あるいはユーザの要求やコマンドに応じて、ビジュアルデータ104を送信することができる。ビジュアルレコーダ102はさらに、任意の手段によって、例えば通信チャネルもしくはケーブル(例えば、光ファイバケーブル)および/または無線接続を介して、ビジュアルデータ104を送信することができる。

【0019】

このシステムは、1つまたは複数のオーディオレコーダ106をさらに含む。オーディオレコーダ106は、入ってくる音または音波を受信してデジタル波形および/または電流もしくは電気エネルギー(例えば、オーディオデータ108)へ変換するための任意の器具やデバイスとすることができる。例えばオーディオレコーダ106は、マイクロフォンとすることができる。レコーダ106は、通信チャネルもしくはケーブルおよび/または無線接続を介して、オーディオデータ108を含む情報をシステムへ伝達することができる。オーディオレコーダ106によって受信される入力は、任意のオーディオ入力とすることができる。例えばオーディオ入力は、話された単語や語句、あるいは単語や語句の集合など、ユーザから話された発声とすることができる。この入力はさらに、保存および/または記録してほしいとユーザがシステムに対して希望する動画や静止画像に関連付けられた話声(speech)による注釈とすることができる。いくつかの実施形態においては、オーディオレコーダ106への入力は、システムによって伝達または実施してほしいとユーザが希望する1つまたは複数の話声コマンド(speech command)を表す話された単語や語句を含むことができる。オーディオレコーダ106への入力は、その他の任意の音および/またはノイズとすることもできる。

【0020】

このシステムは、このシステムへ動作可能に接続されている1つまたは複数のデバイス132をさらに含む。デバイス132は、例えば、光源、吸入器、真空ポンプ、ビデオディスプレイ、あるいは手術台など、このシステムによってコントロールされる医療機器とすることができる。さらにデバイス132は、例えば、直接の音声認識システムのコントロール下にはない麻酔機器など、データ検索のためにこのシステムへ接続される医療機器とするこ

10

20

30

40

50

とができる。技術用途、工業用途、または科学用途に関しては、デバイス132は、例えば、テスト機器、環境制御機器、製造機器(manufacturing equipment)、あるいはエンジンの状態をコントロールする機器(engine status and control equipment)など、デバイスのコントロール、ならびに/またはデータの取り込み、同期化、および保存が望まれる特定の取り組みに必要なとされる機器とすることができる。

【0021】

デバイス132は、デバイスデータ134をシステムあるいはそのモジュールまたはプロセッサ110へ送信する。デバイスデータ134は、デバイス132からの任意のデータを含むことができ、それらのデータとしては、デバイスの状態に関する情報、デバイスの設定、および/またはデバイスの(1つまたは複数の)出力が含まれるが、これらには限定されない。デバイス132は、継続的に、および/またはリアルタイムで、あるいはユーザの要求やコマンドに応じて、デバイスデータ134を送信することができる。デバイス132はさらに、任意の手段によって、例えば通信チャネルもしくはケーブル(例えば、光ファイバーケーブル)および/または無線接続を介して、デバイスデータ134を送信することができる。

10

【0022】

このシステムは、1つまたは複数のプロセッサ110をさらに含むことができる。プロセッサ110は、少なくとも1つのオペレーションをコントロールする、あるいは1つまたは複数のソフトウェアプログラムを受信するおよび/または実行する任意のデバイス、デバイスおよび/またはコンポーネントの集合、あるいはシステムとすることができる。プロセッサ110は、例えば、デジタルシグナルプロセッサ、マイクロコントローラ、マイクロプロセッサ、またはコンピュータプログラマブルロジックデバイスのうちの1つとすることができる。プロセッサ110に関連付けられている機能は、ローカルかリモートかを問わず、集中させることも分散させることもできるという点に留意されたい。プロセッサ110は、オーディオレコーダ102およびビジュアルレコーダ106、ならびにデバイス132と通信状態にあり、それらから、ビジュアルデータ104、オーディオデータ108、および/またはデバイスデータ134を含む情報を受信することができる。プロセッサ110は、データ、例えばビジュアルデータ104、オーディオデータ108、および/またはデバイスデータ134を処理するための任意の数のアプリケーション、コンポーネント、またはモジュールを含むこと、および/またはそれらにアクセスすることができる。

20

【0023】

本発明によるシステムはまた、任意の数のストレージ、ストレージデバイス、および/またはストレージドライブを含む。例えばこのシステムは、1つまたは複数のデータベース122を含むことができる。このシステムは、1つまたは複数のポータブルメディアドライブ124と、対応するポータブルメディアとをさらに含むことができる。ポータブルメディアドライブ124は、任意のポータブルメディアまたはポータブルストレージドライブとすることができ、3.5インチフロッピー(登録商標)ディスクドライブ、ジップディスクドライブ、CD書き込みドライブ、DVD書き込みドライブ、メモ리카ードドライブ(例えば、コンパクトフラッシュ(登録商標)、セキュアデジタル、メモリスティックなど)、テープドライブ、および/または外付けハードドライブを含むが、これらには限定されない。そしてデータレコード120は、ポータブルメディアを使用してデータベースまたはリモートコンピュータへ転送することもでき、あるいはローカルエリアネットワーク(すなわちイントラネット)またはワイドエリアネットワーク(すなわちインターネット)を介してネットワークストレージデバイスへ転送することもできる。図1に示されているように、このシステムの1つまたは複数のストレージは、このシステムによって生成または更新された(1つまたは複数の)データレコード120を受信して保存することができる。

30

40

【0024】

図2は、本発明によるシステムの別の概略図を示している。図示されているように、このシステムは、プロセッサ110によって実行可能な任意の数のモジュール、コンポーネント、もしくはソフトウェアアプリケーション、および/またはプロセスを含むことができる。当業者なら理解できるであろうが、このシステムのソフトウェアは、プロセッサ110

50

のローカルに、またはリモートに、例えばデータベースおよび/または一時的なメモリ(例えばストレージ122)内に格納または配置することができる。例えば、いくつかの実施形態においては、それぞれのモジュールは、ソフトウェアを含む別個のハードウェアモジュール、ならびに、例えばプロセッサおよび/またはストレージとすることができる。

【0025】

このシステムは、オーディオデータ、ビジュアルデータ、およびデバイスデータを受信して処理するための少なくとも1つのデータ取り込みモジュール112を含む。データ取り込みモジュール112は、ハードウェア、ソフトウェア、またはそれらの組合せにおいて具体化することができる。例えばデータ取り込みモジュール112は、プロセッサ110上で実行されるデータ取り込みソフトウェアを含むことができる。データ取り込みモジュール112は、オーディオデータ108、ビジュアルデータ104、デバイスデータ134などのデータを受信して、処理する。

10

【0026】

このシステムは、音声認識モジュール114をさらに含み、この音声認識モジュール114は、オーディオデータ108を解釈して、そこから転写データ116を生成する。例えば音声認識モジュール114は、オーディオレシーバ106からオーディオデータ108を受信し、そのオーディオデータ108をテキストの単語、数字、あるいは記号のセットへと変換することができる。そして解釈されたデータすなわち転写データ116をデータ取り込みモジュール112への入力として送信することができる。いくつかの実施形態においては、このシステムは、転写データを生成するための転写モジュール(図示せず)を(音声認識モジュール114と組み合わせて、または別個のモジュールとして)さらに含む。

20

【0027】

本発明のデータ取り込みモジュール112は、オーディオデータ108、転写データ116、ビジュアルデータ104、および/またはデバイスデータ134を含むデータレコード120を生成する。ビジュアルデータ104やデバイスデータ134がまったく存在しない状況では、データレコード120は、オーディオデータ108(例えば、生のオーディオデータ)および転写データ116のみを含むことができる。データ取り込みモジュール112はさらに、データレコード120内のデータを同期化する。したがって、特定のデータレコード120にアクセスするユーザは、特定の時点に関連するすべてのデータ、例えばオーディオデータ108、転写データ116、ビジュアルデータ104、およびデバイスデータ134をレビューすることができる。

30

【0028】

データレコード120は、医療記録データベースやテスト結果データベースなどのストレージ122/124内に保存される。ストレージ122/124は、場合によっては、複数のデータレコードを含むことができる。データレコード120は、例えば転写データ116の内容に基づいて、検索することができる。上述のように、転写データ116は、オーディオデータ108、ビジュアルデータ104、デバイスデータ134などのその他の形態のデータが存在する場合には、そうしたデータと同期化される。したがってユーザは、患者または患者識別子、治療または手術中に話された特定の音声による注釈または主題、時間または期間、あるいはその他の任意の適用可能な検索パラメータに基づいて、データレコードの問合せを行うことができる。転写データ116の内容を問い合わせることによって、ユーザは、同期化されたおよび/または関連付けられたビジュアルデータ104(例えば、音声による注釈に関連付けられた動画および/または静止画像)ならびに生のオーディオデータ108および/またはデバイスデータ134にアクセスすることもできる。

40

【0029】

いくつかの実施形態においては、データレコード120は、転写データの一部に基づいてストレージまたはデータベースの場所に保存される。例えばデータレコード120は、患者識別子および/または特定の主題や話題など、転写データ116内の識別子に対応するデータベースの場所に保存することができる。データおよび/またはデータレコードはさらに、転写データ116の内容(例えば、識別子および/またはキーワード)に基づいて保存することができる。

50

【0030】

図3は、本発明によるシステムの別の概略図を示している。図示されているように、このシステムのいくつかの実施形態は、コマンドコントローラ118またはコマンドモジュール、および弁別器136を含む。コマンドコントローラ118は、弁別器136を介して、解釈されたオーディオデータ108の一部に基づいて、アクションを実行したり、あるいはシステムコマンドやデバイスコマンドを生成したりすることができる。例えばユーザは、静止画像を取り込むための、あるいはビジュアルレコーダ102を起動するためのコマンドなど、音声コマンド(voice command)を発することができる。そしてコマンドコントローラ118は、コマンドデータ130を、システムおよび/または特定のコマンドによって機能できるデバイス132またはアプリケーションへ送信することができる。デバイス132は、医療用のデバイス、器具、ツール、アプリケーションなど、任意のデバイスとすることができる。デバイス132は、オーディオレコーダ106やビジュアルレコーダ102とすることもできる。

10

【0031】

弁別器136は、オーディオデータ108のどの部分が話声コマンドとして機能できるかを判断し、そのような部分および/またはコマンドをコマンドコントローラ118へ回送する。弁別器136は、話声コマンドを識別してコマンドコントローラ118へ回送しながら、同時に同じ(1つまたは複数の)話声コマンドを転写して転写データ116内に含めるなど、システムユーザの好みに応じて別々のモードで機能することができる。このようなモードで機能する際には、このシステムは、音声認識モジュール114を介したデバイスのコントロールを可能にしながら、システムユーザによって転写を希望されているすべてのオーディオデータ108をリアルタイムに転写することになる。逆に、弁別器136は、ユーザによって転写を意図されているオーディオデータ108のみが転写データ116内に含まれるモード(すなわち、話声コマンドは転写されず、転写データ116内に含まれないモード)にすることもできる。

20

【0032】

弁別器136および/またはコマンドコントローラ118は、非常に動的であり、システムによってアクセス可能な(例えば、データベース122内の)言語モデルや認識文法を必要に応じて継続的に更新することができる。例えば認識文法は、コントロールされている機器のうちの1つがオン/オフされるときや、音声コマンドによってデバイスコントロールメニュー内の位置の変更(例えば、メニューレベルの変更)が生じるときに更新することができる。さらに、1つのデバイスが接続を解除された場合には、認識文法が直ちに更新されて、そのデバイスに関連付けられているコマンドが削除される。そのようなものとして、本発明による音声コマンドとオーディオの転写との統合は、機器やデバイスをコントロールするように意図されている音声コマンドが常に正しく認識および解釈される一方で、転写を意図されている音声による注釈がオーディオ転写システムによって吸収されるシステムを提供する。

30

【0033】

このシステムのいくつかの実施形態は、少なくとも1つのユーザインターフェースまたはグラフィカルユーザインターフェース(図示せず)をさらに含む。このユーザインターフェースは、情報やフィードバックをユーザに表示するための、もしくはその他の形で提供するための、および/または入力や情報をユーザから受信するための任意のデバイスとすることができる。例えば、このユーザインターフェースは、モニターやディスプレイ、キーパッド、キーボード、タッチスクリーン、マウス、および/またはオーディオ出力など、任意の数のコンポーネントを含むことができる。このユーザインターフェースは、このシステムの1つまたは複数のデバイスおよび/またはアプリケーションに関連付けられているコマンドを含むツリー構造のメニューなどのメニューを表示することができる。

40

【0034】

図4は、図1～図3に示されているシステムによって採用できるオーディオデータ、ビジュアルデータ、およびデバイスデータを処理する方法を示している。この方法は、オーディオ入力もしくはデータ、ビジュアル入力もしくはデータ、および/またはデバイス入力もしくはデータを記録するステップ(ステップ401)を含む。オーディオデータ、ビジュア

50

ルデータ、および/またはデバイスデータは、任意の数のビジュアルレコーダおよび/またはオーディオレコーダによって受信された像およびオーディオ入力を含むことができる。オーディオデータは、音声認識装置または音声認識モジュールへ送信される(ステップ403)。オーディオデータが解釈され、そこからリアルタイムな転写物または転写データが生成される(ステップ405)。次いで、少なくともオーディオデータ(例えば、生のオーディオデータ)と、対応する転写データを含むデータレコードが生成される(ステップ407)。ビジュアルデータおよび/またはデバイスデータが存在する場合には、さらにそれらのデータをデータレコード内に含めることができる。データレコード内のデータはまた、データレコードが生成される際に同期化される。例えば、データレコード内の転写データおよび/またはオーディオデータは、関連するビジュアルデータおよび/またはデバイスデータに同期化された音声による注釈を含むことができる。そしてデータレコードは、システムによってアクセス可能な場所(例えば、データベースやポータブルメディアデバイス)に保存される。場合によっては、データレコードは、転写データの少なくとも一部に基づいてデータベースの場所に保存される。

【0035】

図5は、図1～図3に示されているシステムによって採用できるオーディオデータ、ビジュアルデータ、および/またはデバイスデータを処理する別の方法を示している。この方法は、オーディオ入力もしくはデータ、ビジュアル入力もしくはデータ、および/またはデバイス入力もしくはデータを記録するステップ(ステップ501)と、解釈および転写データの生成のためにオーディオデータを音声認識装置または音声認識モジュールへ送信するステップ(ステップ503)とを含む。次いで、オーディオデータおよび/または転写データを、例えば弁別器によって分析して、そのオーディオデータおよび/または転写データ内に含まれているいかなる話声コマンドも識別することができる(ステップ505)。話声コマンドが見つかった場合には、その話声コマンドまたはコマンドデータは、コマンドコントローラを介して、特定のコマンドによって機能できるアプリケーションまたはデバイスへ送信される(ステップ507)。そして転写が完了する(ステップ509)。いくつかの実施形態においては、転写データは、オーディオデータのほぼすべてに対応することができる。その他のいくつかの実施形態および/またはオペレーションのモードにおいては、話声コマンドなど、オーディオデータのうちの少なくとも一部は、転写されない。そして、例えば生のオーディオデータ、転写データ、ビジュアルデータ、および/またはデバイスデータのうちの少なくとも一部を含むデータレコードを生成することができる(ステップ511)。

【0036】

実際には、本発明は、例えば医療手術室においてビジュアルデータ、オーディオデータ、および/またはデバイスデータを取り込むために使用することができる。例えば、本発明を使用して、手術のストリーミングビデオ映像を取り込み、外科医の音声による注釈をビデオ映像に記録することができる。外科医は、特定の時点において、その時点で記録されているビデオ映像に関連するコメントを作成することができる。例えば外科医は、「胆嚢除去手術を行っているときに、いくらかの疑わしい出血が見られた。この出血場所は、患者の肝臓の後ろであり、実施中の手術によって生じているものではない。さらなる検査が必要である」と述べることができる。上述のコメントは、ビデオ映像もしくはデータおよび/またはデバイスデータと共にリアルタイムに転写され、同期化され、保存される。例えば、転写データおよびビデオデータは、患者に関連付けられている医療記録データベースの場所に一緒に保存することができる。外科医、あるいはその他の任意のユーザは、例えば「疑わしい出血」などの検索用語(search term)を使用して、後からデータレコードの問合せを行うことができる。検索に基づいて、ユーザは、外科医が気づいた疑わしい出血に関連する転写データ、生のオーディオデータ、およびビジュアルデータ、あるいはビデオ映像を迅速に突き止めて、見直すことができる。

【0037】

本発明について、部品や機能などの特定の構成を参照して説明したが、これらは、すべての可能な構成や機能を網羅することを意図するものではなく、実際には、当業者なら、

10

20

30

40

50

多くの修正形態および変形形態を確かめることができるであろう。

【図面の簡単な説明】

【0038】

【図1】本発明によるシステムを示す概略図である。

【図2】図1に示されているシステムの別の概略図である。

【図3】図1に示されているシステムの別の概略図である。

【図4】図1～図3に示されているシステムによって採用できるオーディオデータ、ビジュアルデータ、およびデバイスデータを処理する方法を示す図である。

【図5】図1～図3に示されているシステムによって採用できるオーディオデータ、ビジュアルデータ、およびデバイスデータを処理する別の方法を示す図である。

10

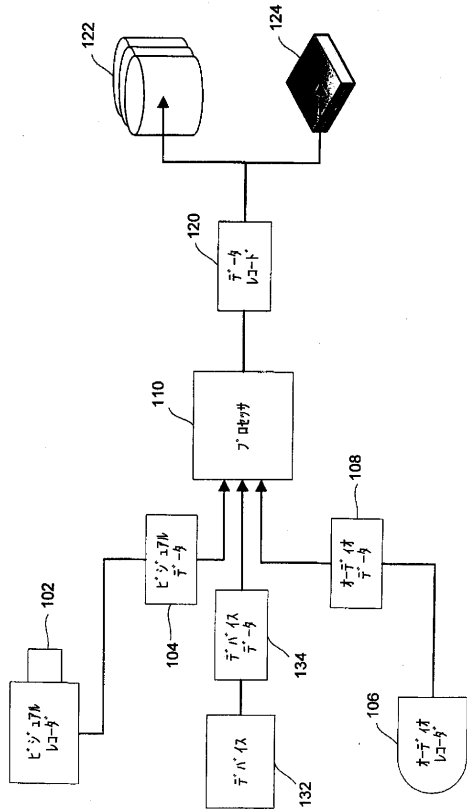
【符号の説明】

【0039】

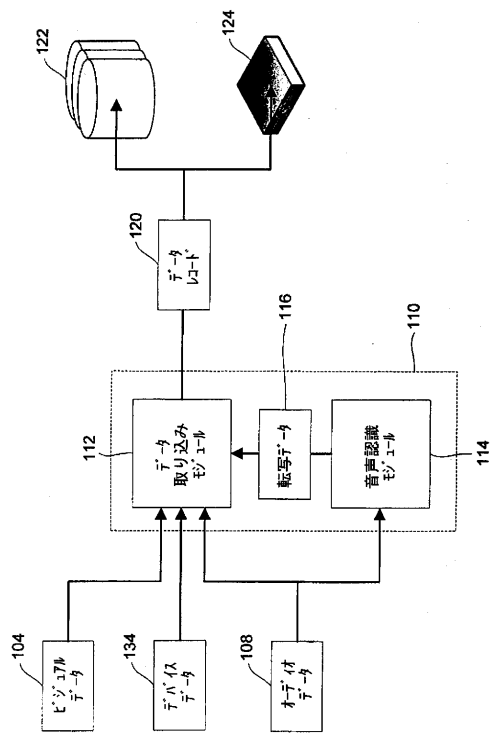
- 102 ビジュアルレコーダ
- 104 ビジュアルデータ
- 106 オーディオレコーダ
- 108 オーディオデータ
- 110 プロセッサ
- 112 データ取り込みモジュール
- 114 音声認識モジュール
- 116 転写データ
- 118 コマンドコントローラ
- 120 データレコード
- 122 データベース
- 124 ポータブルメディアドライブ
- 130 コマンドデータ
- 132 デバイス
- 134 デバイスデータ
- 136 弁別器

20

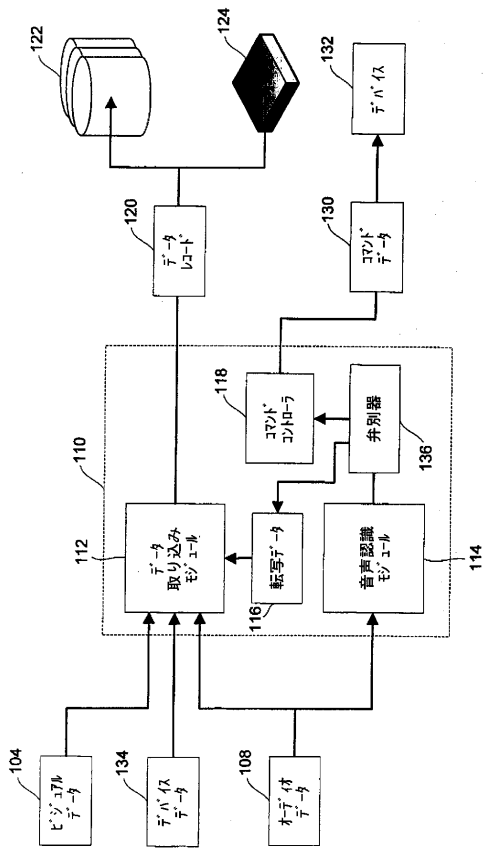
【図 1】



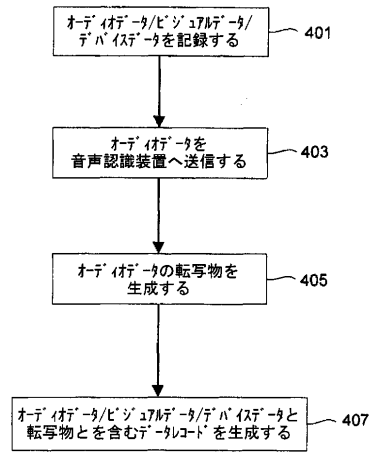
【図 2】



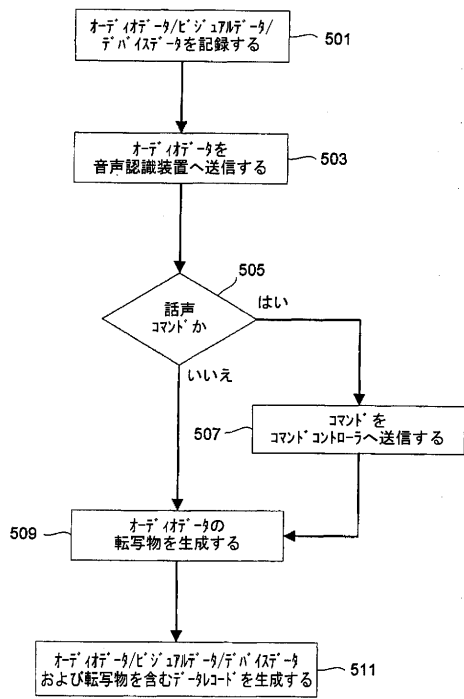
【図 3】



【図 4】



【図5】



フロントページの続き

- (72)発明者 ガン・ワン
アメリカ合衆国・カリフォルニア・93117・ゴレタ・#C・サイプレス・ウォーク・794
- (72)発明者 マッテオ・コントリニ
アメリカ合衆国・カリフォルニア・93111・サンタ・バーバラ・ペブル・ヒル・ドライブ・246
- (72)発明者 チェンギ・チェン
アメリカ合衆国・カリフォルニア・93117・ゴレタ・#6・カセドラル・オークス・ロード・7680
- (72)発明者 ヘインツ・ヴェルナー・スティラー
ドイツ・DE - 79798・イエステテン・オイレンヴェーク・1

審査官 小田 浩

- (56)参考文献 特表2004-507941(JP,A)
特開2005-135344(JP,A)
特開2003-044085(JP,A)
特開2002-336184(JP,A)
特開2002-207753(JP,A)

(58)調査した分野(Int.Cl., DB名)

H04N 5/91
G10L 15/00
G11B 20/10
H04N 5/76
H04N 5/928