

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4619623号
(P4619623)

(45) 発行日 平成23年1月26日(2011.1.26)

(24) 登録日 平成22年11月5日(2010.11.5)

(51) Int. Cl. F I
 HO 4 M 3/533 (2006.01) HO 4 M 3/533
 HO 4 M 3/42 (2006.01) HO 4 M 3/42 J

請求項の数 18 外国語出願 (全 16 頁)

<p>(21) 出願番号 特願2003-68978 (P2003-68978) (22) 出願日 平成15年3月13日 (2003. 3. 13) (65) 公開番号 特開2003-289387 (P2003-289387A) (43) 公開日 平成15年10月10日 (2003.10.10) 審査請求日 平成18年1月31日 (2006. 1. 31) (31) 優先権主張番号 10/099, 606 (32) 優先日 平成14年3月15日 (2002. 3. 15) (33) 優先権主張国 米国 (US)</p>	<p>(73) 特許権者 500046438 マイクロソフト コーポレーション アメリカ合衆国 ワシントン州 9805 2-6399 レッドモンド ワン マイ クロソフト ウェイ (74) 代理人 100077481 弁理士 谷 義一 (74) 代理人 100088915 弁理士 阿部 和夫 (72) 発明者 エリック イーチャオ チャン 中華人民共和国 100022 ペキン ジャンゴ ロード 88 アpartment 1-1808 審査官 永井 啓司</p>
--	---

最終頁に続く

(54) 【発明の名称】 ボイスメッセージ処理システムおよび方法

(57) 【特許請求の範囲】

【請求項 1】

複数のボイスメッセージを示すボイスメッセージデータを格納するボイスメッセージ (VM) データ記憶と、

前記 VM データ記憶に結合され、および前記ボイスメッセージにアクセスし、有用な情報を前記ボイスメッセージから抽出して、抽出した情報から特定される所定の情報を前記 VM データ記憶に格納された VM データに付加するように構成されたボイスデータプロセッサと、

前記 VM データ記憶に結合され、および前記付加された VM データへのユーザアクセスを提供するように構成されたユーザインタフェースコンポーネントと

少なくとも 1 つの話者識別モデルを格納する話者識別モデルデータ記憶と、

該話者識別モデルデータ記憶にアクセスして、前記 VM データに対応するボイスメッセージに関連する話者の識別の指示を前記所定の情報として提供するように構成された話者識別コンポーネントと

話者の感情を表す音響特徴または発話速度を示す音響特徴を前記 VM データから抽出する音響特徴抽出器と、

前記抽出した話者の感情を表す音響特徴に基づいて該話者の感情を示す感情出力を前記所定の情報として提供する感情識別器と、

前記抽出した発話速度を示す音響特徴に基づいて該発話速度を示す速度出力を前記所定の情報として提供する速度識別器と、

10

20

ボイスメッセージの内容を表すテキストの出力を生成して前記所定の情報として提供するよう
に構成された音声テキストコンポーネントと、

を備え、

前記ボイスデータプロセッサは、前記ボイスメッセージに付加された話者の識別、話者の感情、発話速度およびボイスメッセージの内容の少なくとも1つに基づいて前記ボイスメッセージをソートする規則適用コンポーネントをさらに備えたことを特徴とするボイスメッセージ処理システム。

【請求項2】

前記ボイスデータプロセッサは、

VMデータを受信して、前記VMデータに基づく話者識別モデルを更新して前記VMデータに対応するボイスメッセージの話者を表すユーザ入力に関連付けるように構成された話者モデルトレーニングコンポーネントを備えたことを特徴とする請求項1に記載のシステム。

10

【請求項3】

前記ボイスデータプロセッサは、

前記速度出力を受信して、関連するボイスメッセージを予め定められた発話速度に標準化するように構成された速度標準化コンポーネントを備えたことを特徴とする請求項1または2に記載のシステム。

【請求項4】

前記音声テキストコンポーネントは、前記ボイスメッセージの転写を前記テキストの出力として生成するように構成されたことを特徴とする請求項1または2に記載のシステム。

20

【請求項5】

前記ボイスデータプロセッサは、

前記ボイスメッセージの要約を生成するように構成された要約コンポーネントを備えたことを特徴とする請求項1または2に記載のシステム。

【請求項6】

前記ボイスデータプロセッサは、

前記ボイスメッセージの少なくとも一部の意味解析を生成するように構成された意味解析器を備えたことを特徴とする請求項1または2に記載のシステム。

30

【請求項7】

前記規則適用コンポーネントは、前記所定の情報に基づいて警告を生成することを特徴とする請求項1または2に記載のシステム。

【請求項8】

前記ユーザインタフェースコンポーネントは、前記ボイスメッセージのユーザによる処理のためにユーザ選択可能な入力を可視出力(exposing)するユーザインタフェースを生成することを特徴とする請求項1または2に記載のシステム。

【請求項9】

前記ユーザ選択可能な入力は、

ユーザによって起動されると、ボイスメッセージに関連する発話速度が変更される速度変更入力を備えたことを特徴とする請求項8に記載のシステム。

40

【請求項10】

前記ユーザインタフェースは、ボイスメッセージの内容を示すテキストを表示することを特徴とする請求項8に記載のシステム。

【請求項11】

前記ユーザインタフェースは、ボイスメッセージの話者の個人を示す個人情報を表示することを特徴とする請求項8に記載のシステム。

【請求項12】

前記ユーザインタフェースは、ボイスメッセージの話者の感情を表す感情情報(emotion indicator)を表示することを特徴とする請求項8に記載のシステム

50

【請求項 13】

前記ユーザインタフェースは、適用される規則を表す規則情報 (a r u l e i n d i c a t o r) を表示することを特徴とする請求項 8 に記載のシステム。

【請求項 14】

ボイスメッセージを処理する方法であって、

前記ボイスメッセージデータをボイスメッセージ (V M) データ記憶に格納するステップと、

前記 V M データ記憶を断続的にアクセスして、新しいボイスメッセージが格納されたか否かを決定するステップと、

新しいボイスメッセージの各々について、プロセッサにおいて前記新しいボイスメッセージを処理して、所定の情報を提供するステップであって、前記提供するステップは、さらに、少なくとも 1 つの話者識別モデルを格納する話者識別モデルデータ記憶にアクセスして、前記 V M データに対応するボイスメッセージに関連する話者の識別の指示を前記所定の情報として提供するステップと、話者の感情を表す音響特徴を前記 V M データから抽出して該話者の感情を示す感情出力を前記所定の情報として提供し、かつ発話速度を示す音響特徴を前記 V M データから抽出して該発話速度を示す速度出力を前記所定の情報として提供するステップと、ボイスメッセージの内容を表すテキストの出力を生成して前記所定の情報として提供するステップとを含む、提供するステップと、

前記提供された前記ボイスメッセージに付加された話者の識別、話者の感情、発話速度およびボイスメッセージの内容を V M データ記憶中のデータに付加するステップと

前記ボイスメッセージに付加された話者の識別、話者の感情、発話速度およびボイスメッセージの内容の少なくとも 1 つに基づいて前記ボイスメッセージをソートするステップと

を備えたことを特徴とする方法。

【請求項 15】

前記音響特徴は、前記新しいボイスメッセージの話者の発話速度を表す発話速度情報 (s p e a k i n g r a t e i n d i c a t o r) を含み、

前記発話速度をユーザ選択の速度に標準化するステップをさらに備えたことを特徴とする請求項 14 に記載の方法。

【請求項 16】

前記話者の個人情報を取得するステップは、予め格納された話者を識別するためのモデルに基づいて話者の個人情報が不明であると決定されるとき不明出力を提供するステップを含み、

前記新しいボイスメッセージについての話者の個人情報を表すユーザ入力を受信するステップと、

話者を識別するためのモデルを更新して前記ユーザ入力に関連付けるステップと

をさらに備えたことを特徴とする請求項 14 に記載の方法。

【請求項 17】

前記新しいボイスメッセージのためのテキスト表現を意味的に構文解析するステップをさらに備えたことを特徴とする請求項 14 に記載の方法。

【請求項 18】

前記 V M データ記憶へのユーザインタフェースを生成するステップであって、該ユーザインタフェースは前記 V M データ記憶にあるボイスメッセージを処理するためのユーザ起動入力を含むステップ

をさらに備えたことを特徴とする請求項 14 に記載の方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

10

20

30

40

50

本発明は、音声処理に関し、より具体的には、分散システムによって受信されたボイスメッセージを処理するためのボイスメッセージ処理に関する。

【0002】

【従来の技術】

現在、多くの人々が広範な種類の情報源から極めて多くの種々のメッセージを受信する。例えば、人々にとって、週に数十通のボイスメールメッセージを受信することは珍しいことではない。現在ユニファイドメッセージを使用することが、この問題を悪化させている。ユニファイドメッセージシステムにおいては、広範な種類の情報源から受信されたメッセージ、例えばボイスメッセージ、電子メールメッセージ、ファックスメッセージおよびインスタントメッセージなどは、シームレスに結び付けられてアクセスすることができる。しかしながら、電子メールメッセージおよびインスタントメッセージシステムと比べ、ボイスメッセージに関連する情報の種類は非常に限定される。

10

【0003】

【発明が解決しようとする課題】

例えば、電子メールメッセージは、通常、その送信者の識別情報、件名および優先順位を含む。同様に、このような情報は、元々テキストなので、非常に容易にスキャンたり、コピーやペーストをしたりすることができる。反対に、ボイスメールメッセージは、通常、送信者のいかなる識別情報も持たない。発信者表示が組み込まれたシステムでは、入力している番号を識別することができ、その入力している番号が個人に関連付けられていれば、送信者と思われる人を識別することもできる。しかしながら、このようなシステムが追跡するのは電話器だけであって、話者ではない。通常、ボイスメールメッセージは、件名や優先順位に関する指示が含まれておらず、元々書かれたものではなく音声なので、スキャンしたり、コピーやペーストをしたりするのも困難である。

20

【0004】

ボイスメッセージに関連する情報が不足しているため、処理にはより多くの時間が費やされることとなる。例えば、件名や送信者をざっと見て多くの電子メールメッセージを除去し、必要でなければ直ちにメールボックスから削除したり、所定のホルダを構成したりすることができる。実際、特定のユーザからのメールメッセージ、あるいは特定の件名を持つメールメッセージを削除するための規則を指定して、自動的にこれらの処理をすることさえできる。

30

【0005】

一方、ユーザは単に送信者や件名などの基本的な情報を抽出するだけでもメッセージを1件1件聞かなければならないので、ボイスメールメッセージを調べるには、一般に非常に多くの時間を必要とする。現時点では、ボイスメールメッセージを（送信者、件名あるいは緊急度で整理するなど）予め構成しておくための規則を自動的に生成することも、実質上不可能である。

【0006】

【課題を解決するための手段】

分散データ記憶上に複数のボイスメッセージを示すボイスメッセージデータを格納することによって、分散システムにおいてはボイスメッセージが処理される。分散データプロセッサは、ボイスメッセージにアクセスして、そのボイスメッセージから所定の情報を抽出する。次に、データプロセッサは、ボイスメッセージデータ記憶に格納されたデータに抽出された情報を付加する。ユーザは、ユーザインタフェースコンポーネントにより、付加されたデータをもってボイスメッセージにアクセスすることができる。

40

【0007】

一実施形態において、分散ボイスデータプロセッサは、ソートや警告の生成などユーザ選択の規則をそのデータに適用する。

【0008】

ボイスデータプロセッサは、例えば（話者識別モデルを用いた）話者の識別、話者の感情や発話速度のような広範な情報を抽出する。ボイスデータプロセッサは、また、メッセー

50

ジをユーザによって選択可能な所定の発話速度に標準化することもできる。

一実施形態において、ボイスデータプロセッサは、メッセージを文字に転写 (transcribing) して要約し、ボイスメッセージに関し (意味解析のような) 自然言語処理を実行するための文字転写 (transcription) コンポーネントも含む。

【0009】

ユーザ入力によって、ユーザはボイスメッセージを扱うための広範囲のユーザ起動入力を得ることができる。このような入力は、例えばボイスメッセージの速度を上げたり、または下げるための速度変更入力、規則を設定するための入力、ボイスメッセージから抽出された種々の情報の表示、およびユーザによって選択あるいは除外された規則の表示を含む。

10

【0010】

【発明の実施の形態】

本発明は、分散処理システム上で実現され、所定の情報をボイスメッセージから抽出する。本発明は、所定の情報を抽出して、抽出された情報をもつボイスメッセージを含むボイスデータ記憶に付加する。ユーザインタフェースによって、ボイスメッセージは、追加された付加情報を付与して容易に取り扱うことができる。本明細書の分散されたとの記載は、サーバを含まないシステムを意味し、デスクトップシステム、パーソナルデジタルアシスタント (PDA)、電話、ラップトップコンピュータなどのような個人ユーザがコントロールするシステムを意味する。したがって、例えば、本明細書で分散処理プロセッサを議論する際、それは例えばサーバ上に存在するというよりも、ネットワークの一部とすることができるのであって、ユーザ個人がコントロールする装置上に存在するプロセッサを意味する。

20

【0011】

図1は、本発明を実施することができる適切なコンピューティングシステム環境100を例示する。コンピューティング環境100は適切なコンピューティング環境の一例にすぎず、本発明の使用または機能の範囲に関しいかなる限定を示唆することも企図していない。コンピューティング環境100は、例示の動作環境100に示されたいかなる各コンポーネントまたはその組み合わせに依存し、またはこれらを要求するものと解釈すべきではない。

【0012】

本発明は、多くの他の汎用のまたは専用のコンピューティングシステム環境または構成とともに動作することができる。本発明で使用するのに適する例としてよく知られたコンピューティングシステム、環境、および/または構成、例えばパーソナルコンピュータ、ハンドヘルドまたはラップトップ装置、マルチプロセッサシステム、マイクロプロセッサベースのシステム、セットトップボックス、プログラム可能なコンシューマ電子機器、ネットワークPC、ミニコンピュータ、メインフレームコンピュータ、上記のいずれかのシステムまたは装置を含む分散コンピューティング環境などが含まれるが、これに限定されない。

30

【0013】

本発明は、コンピュータによって実行されるプログラムモジュールといったコンピュータ実行可能命令の通常のコテキストで記述することができる。一般に、プログラムモジュールは、特定のタスクを実行しあるいは特定の抽象データタイプを実装するルーチン、プログラム、オブジェクト、コンポーネント、データ構造などを含む。本発明はまた、分散コンピューティング環境で実施することもでき、そこではコミュニケーションネットワークを通して接続されたリモート処理装置によってタスクが実行される。分散コンピューティング環境において、プログラムモジュールは、メモリ記憶装置を含むローカルおよびリモートの双方のコンピュータ記憶媒体に置くことができる。

40

【0014】

図1を参照すると、本発明を実施するための例示のシステムは、コンピュータ110の形の汎用コンピューティング装置を含む。コンピュータ110のコンポーネントには、処理

50

ユニット120、システムメモリ130および、システムメモリを含む種々のシステムコンポーネントを処理ユニット120に結合するシステムバス121を含むがこれに限定されない。システムバス121は、いかなる種類のバスアーキテクチャをも使用するメモリバスまたはコントローラ、ペリフェラルバスおよびローカルバスを含む数種のバス構造とすることができる。例として、このようなアーキテクチャには工業標準アーキテクチャ（ISA）バス、マイクロチャネルアーキテクチャ（MCA）バス、拡張ISA（EISA）バス、ビデオ電子標準化協会（VESA）ローカルバスおよびメザニンバスとしても知られる周辺コンポーネント相互接続（PCI）バスが含まれるがこれに限られない。

【0015】

コンピュータ110は、通常、種々の読み取り可能媒体を含む。コンピュータ読み取り可能媒体は、コンピュータ110によりアクセス可能ないかなる媒体とすることができ、揮発性および不揮発性媒体、並びに取外し可能および固定媒体の双方を含む。例として、コンピュータ読み取り可能媒体は、コンピュータ記憶媒体および通信媒体を備えることができるが、これに限られない。コンピュータ記憶媒体は、情報、例えばコンピュータ読み取り可能命令、データ構造、プログラムモジュールまたはその他のデータを記憶するためのいずれかの方法または技術で実施される揮発性および不揮発性媒体、並びに取外し可能および固定媒体の双方を含む。コンピュータ記憶媒体は、RAM、ROM、EEPROM、フラッシュメモリまたはその他のメモリ技術、CD-ROM、デジタル多用途ディスク（DVD）またはその他の光ディスク記憶、磁気カセット、磁気テープ、磁気記憶ディスクまたはその他の磁気記憶機器、あるいは所定の情報を記憶するための使用およびコンピュータ110からのアクセスが可能な他のいずれかの媒体も含むがこれに限られない。通信媒体は、一般に、コンピュータ読み取り可能命令、プログラムモジュールあるいはその他のデータから変調されたデータ信号、例えば搬送波やその他の伝送手法に具現化するものであり、いかなる配信媒体をも含む。用語「変調されたデータ信号」とは、一定の手法、例えば情報を信号に符号化して設定または変更された1つまたは複数の特性を有する信号を意味する。例として、通信媒体は、有線ネットワークやディレクトワイヤード接続などの有線媒体および音声、電波、赤外線その他のような無線媒体を含むが、これに限られない。コンピュータ読み取り可能媒体の範囲には上述のいずれの組み合わせも含まれるべきである。

【0016】

システムメモリ130は、コンピュータ記憶媒体を揮発性および/または不揮発性のメモリ、例えば読取専用メモリ（ROM）131および逐次アクセスメモリ（RAM）132の形式で備える。基本入出力システム133（BIOS）は、スタートアップの最中などにコンピュータ110内の要素間における情報の移送を支援する基本ルーチンを含み、通常ROM131に格納される。RAM132は一般に、処理ユニット120にすぐにアクセス可能であり、および/または現在処理ユニット120によってオペレートされているデータおよび/またはプログラムモジュールを含む。例として、図1はオペレーティングシステム134、アプリケーションプログラム135、その他のプログラムモジュール136およびプログラムデータ137を示しているがこれに限られない。

【0017】

コンピュータ110は、その他の取外し可能/固定で、揮発性/不揮発性のコンピュータ記憶媒体も含むことができる。単に例としてであるが、図1は、固定で不揮発性の磁気媒体の読み書きを行うハードディスクドライブ141、取外し可能で不揮発性の磁気ディスク152の読み書きを行う磁気ディスクドライブ151、およびCD-ROMやその他の光媒体のような取外し可能で不揮発性の光ディスク156の読み書きを行う光ディスクドライブ155を示す。例示の動作環境において使用することができる他の取外し可能/固定で揮発性/不揮発性のコンピュータ記憶媒体には、磁気テープカセット、フラッシュメモリカード、デジタル多用途ディスク、デジタルビデオテープ、ソリッドステートRAM、ソリッドステートROMなどが含まれるがこれに限定されない。通常、ハードディスクドライブ141は、インタフェース140のような固定メモリインタフェースを通してシ

10

20

30

40

50

システムバス 1 2 1 に接続されており、磁気ディスクドライブ 1 5 1 と光ディスクドライブ 1 5 5 とは、インタフェース 1 5 0 のような取外し可能メモリインタフェースによってシステムバス 1 2 1 に接続される。

【 0 0 1 8 】

上記で議論し図 1 に示されたドライブと関連するコンピュータ記憶媒体によって、コンピュータ読み取り可能命令、データ構造、プログラムモジュールおよびコンピュータ 1 1 0 のためのその他のデータが記憶される。例えば、図 1 において、ハードディスクドライブ 1 4 1 は、オペレーティングシステム 1 4 4、アプリケーションプログラム 1 4 5、その他プログラムモジュール 1 4 6 およびプログラムデータ 1 4 7 を格納するように示されている。これらのコンポーネントは、オペレーティングシステム 1 3 4、アプリケーション 10
プログラム 1 3 5、その他プログラムモジュール 1 3 6 およびプログラムデータ 1 3 7 と同一または異なるもののいずれかとすることができ、符号は最低限それらが異なるコピーであることを示すために付与されていることに留意されたい。

【 0 0 1 9 】

ユーザは、キーボード 1 6 2、マイクロホン 1 6 3 などの入力機器およびマウス、トラックボールまたはタッチパッドなどのポインティング機器 1 6 1 を通してコマンドおよび情報をコンピュータ 1 1 0 に入力することができる。その他の入力機器（図示せず）には、ジョイスティック、ゲームパッド、サテライトディッシュ、スキャナ、などを含むことができる。これらおよび他の入力機器はしばしば、システムバスに結合するユーザ入力インタフェース 1 6 0 を通して処理ユニット 1 2 0 に接続するが、他のインタフェースおよび 20
バス構造、例えばパラレルポート、ゲームポート、あるいはユニバーサルシリアルバス（USB）などによって接続することもできる。モニタ 1 9 1 または他の種類の表示機器もまた、ビデオインタフェース 1 9 0 のようなインタフェースを介して、システムバス 1 2 1 に接続される。モニタに加え、コンピュータは、その他の周辺出力機器、例えばスピーカ 1 9 7 およびプリンタ 1 9 6 を備えることもでき、出力周辺インタフェース 1 9 5 を通して接続することができる。

【 0 0 2 0 】

コンピュータ 1 1 0 は、1 つまたは複数のリモートコンピュータ 1 8 0 のようなリモートコンピュータへの論理接続を用い、ネットワーク化された環境で動作することができる。リモートコンピュータ 1 8 0 は、パーソナルコンピュータ、ハンドヘルド機器、サーバ、 30
ルータ、ネットワーク PC、ピア機器またはその他従来のネットワークノードとすることができ、一般にコンピュータ 1 1 0 に関する上述した要素の多くまたは全てを含む。図 1 に示される論理接続はローカルエリアネットワーク（LAN）1 7 1 および広域ネットワーク（WAN）1 7 3 を含むが、他のネットワークも含むことができる。このようなネットワーク環境は、オフィス、企業内コンピュータネットワーク、イントラネットおよびインターネットにおいては、通常のものである。

【 0 0 2 1 】

LAN ネットワーク環境で用いられる際、コンピュータ 1 1 0 はネットワークインタフェースまたはアダプタ 1 7 0 を通して LAN 1 7 1 に接続する。WAN ネットワーク環境で使用される際は、コンピュータ 1 1 0 は一般的に、インターネットのような WAN 1 7 3 40
を介して通信を確立するためのモデム 1 7 2 または他の手法を備える。モデム 1 7 2 は、内蔵または外付けとすることができるが、ユーザインタフェース 1 6 0 またはその他の適切な仕組みを介してシステムバス 1 2 1 に接続することができる。ネットワーク化された環境においては、コンピュータ 1 1 0 あるいはその部分に関連して示されたプログラムモジュールは、リモートメモリ記憶機器に格納することができる。例として、図 1 は、リモートのアプリケーションプログラム 1 8 5 をリモートコンピュータ 1 8 0 に存在するように示しているがこれに限られない。示されたネットワーク接続が例示であり、コンピュータ間における通信接続を確立する他の手法を用いることができることが理解されるであろう。

【 0 0 2 2 】

10

20

30

40

50

図2は、本発明の一実施形態に合致したボイスメッセージ処理システム200のより詳細なブロック図である。システム200は、例示するようにボイスデータ入力コンポーネント202、ボイスデータ記憶204、ユーザインタフェースコンポーネント206およびボイスデータプロセッサ208を含む。ボイスデータ入力コンポーネント202は、例示するようにボイスデータにボイスメールメッセージが含まれる場合には電話、ボイスデータが録音された講義や会話の場合にはマイクロフォンを含むことができ、および他のコンポーネント、例えばラジオ、コンパクトディスクプレーヤなどとするすることができる。

【0023】

ボイスデータ記憶204は、例示するようにWAVファイルのようなボイスデータを格納するメモリの一部である。ユーザインタフェースコンポーネント206は、例示するようにユーザによって起動され、ユーザインタフェースを生成して、ボイスデータ記憶204に格納されたボイスメッセージを操作し、および整理することができる。ボイスデータプロセッサ208は例示するように、ボイスメッセージから有用な情報を抽出する情報抽出コンポーネント210およびボイスメッセージにユーザ選択の規則を適用する規則適用コンポーネント212を含む。

10

【0024】

図3は、システム200の一般的な動作を示すフロー図である。まず、データ入力コンポーネント202からボイスメッセージを受信して、ボイスデータ記憶204に格納する。これは図3のブロック214によって示される。情報抽出コンポーネント210は、定期的あるいは断続的に、データ記憶204にアクセスして、それが情報抽出コンポーネント210によって最後にアクセスされてから新たなボイスメッセージがデータ記憶204に格納されたか否かを決定する。これは図3のブロック216および218によって示される。データ記憶204が情報抽出コンポーネント210によって最後にアクセスされてから、新たなボイスメッセージがデータ記憶204に格納されなかった場合、処理は単にブロック216に戻るだけである。

20

【0025】

しかし、情報抽出コンポーネント210は、ブロック218において処理されなかった新しいボイスメッセージに遭遇した場合、それらの新しいメッセージをボイスデータ処理に委ね、新しいメッセージから所定の情報を抽出する。これはブロック220によって示される。所定の情報の例は、より詳細に以下で議論するが、例示するようにデータ記憶204にあるボイスメッセージの処理と整理を改善するように、およびこれらのメッセージへの規則の適用を改善するように所定の情報を適合させることができる。

30

【0026】

いかなる場合においても、所定の情報が新しいメッセージから抽出されると、ボイスデータ記憶204に格納されている(新しいメッセージに対応する)情報は、情報抽出コンポーネント210によってちょうど抽出された追加情報を付加される。これは図3のブロック222によって示される。

【0027】

ボイスメールメッセージから抽出される情報の種類は、上述のように広範囲にわたり変化に富む可能性があるが、ユーザに効果的にボイスメッセージを処理させることができるように、抽出される多くの種類の情報には、話者の識別、話者の発話速度、話者の心理状態、メッセージの内容などが含まれる。図4は、これらの種類の情報を抽出するための情報抽出コンポーネント210の一実施形態を示すブロック図である。当然、その他の異なる情報も同様に抽出することができる。

40

【0028】

図4で例示されるように、情報抽出コンポーネント210は、モデルトレーニングコンポーネント300、話者識別コンポーネント302、話者識別モデル304、音響特徴抽出コンポーネント306、感情識別器(emotion identifier)308、速度標準化コンポーネント310、音声-テキストコンポーネント312および自然言語プロセッサ314を含む。一実施形態においては、新たなメッセージボイスデータ316

50

は、ボイスデータ記憶 204 から得られる。データ 316 は、例示されるように、情報抽出コンポーネント 210 によってまだ処理されたことがなかった新しいボイスメッセージを表現する WAV ファイル、その他のファイルであり、データ記憶 204 に格納されている。

【0029】

一実施形態においては、データ 316 を話者識別コンポーネント 302 に提供する。コンポーネント 302 は、話者識別モデル 304 にアクセスして、話者の識別（情報）を示す話者認識出力（話者 ID）320 を生成する。話者識別コンポーネント 302 および話者識別モデル 304 は、例示するように特定の話者に関してトレーニングされたいずれかの既知の話者識別コンポーネントおよび話者識別モデルとすることができる。話者認識出力 320 は話者のテキストにおける名前、符号化された識別情報（*identifier*）またはユーザによって割り当てられたいずれかの識別情報とすることができる。

10

【0030】

コンポーネント 302 が話者を識別できない場合（例えば、モデル 304 が、新しいメッセージの話者に関連するモデルを含んでいない場合）、コンポーネント 302 は、例示されるように話者の個人情報不明であることを示して話者認識出力 320 を提供する。この場合、ユーザは新しいメッセージを再検討し、話者 ID 320 が不知と表示されると、例示する（図 5 に関連して以下でより詳細に議論する）ようにユーザインタフェース上でユーザ入力を起動する。これによって、モデルトレーニングコンポーネント 300 は、その新しいメッセージに関連する WAV ファイル（あるいは、他のボイスデータ）を取得する。次に、モデルトレーニングコンポーネント 300 は、この話者に対応する話者識別モデルをトレーニングして、それとユーザによる話者識別の入力またはデフォルトの話者識別とを関連付ける。したがって、次にこの話者からのボイスメッセージを処理する際は、話者識別コンポーネント 302 は、その話者に関連する話者識別モデル 304 を有するので、正確な話者 ID 320 を生成する。

20

【0031】

モデルトレーニングコンポーネント 300 は、話者識別コンポーネント 302 が誤りを起こしていた場合、モデルを改善させることもできる。システムが誤りを犯す場合、ユーザは例示するように、ユーザインタフェースのウィンドウ内に正確な名前を打ち込みユーザ入力コマンドを入力して、モデルトレーニングコンポーネント 300 がその特定の話者に対する新たな話者モデル 304 を自動的にトレーニングするよう命令する。また、ユーザは、将来、システムがより多く使用されればされるほど話者識別が正確になるように、使用中にモデルを更新するよう選択することもできる。反対に、トレーニングコンポーネント 300 は、指図されることなく話者モデル 304 を徐々に更新することができる。例えば、ユーザが新しいボイスメッセージにアクセスし、それが話者の個人情報を表示していて、ユーザがそれを変更しない場合、モデルトレーニングコンポーネント 300 は、そのメッセージに関連するボイスデータにアクセスして、その話者に対応するモデルを改善させることができる。

30

【0032】

また、話者識別コンポーネント 302 は話者 ID 320 に加えて、認識された個人情報のどの程度機密であるかを示す機密度を提供することもできる。ユーザによって確認されたシステムの判断に基づいて、話者識別コンポーネント 302 は、自動的にそのパラメータを更新して、時間に関する性能を改善することができる。

40

【0033】

本発明の他の形態によると、情報抽出コンポーネント 310 は、所定の音響情報をボイスデータ 316 から抽出して、ボイスメッセージを処理する際にユーザに有用な別のデータを生成するための音響特徴抽出コンポーネント 306 を含む。例えば、所定の音響の特徴を抽出することによって、感情識別器 308 は、予め定められた話者の感情を識別して、その感情を表す話者感情 ID 322 を出力する。

【0034】

50

感情識別器 308 は、いかなる既知の感情識別器とすることもでき、2001年の論文「EMOTION DETECTION FROM SPEECH TO ENRICH MULTIMEDIA CONTENT」(F. Yu等執筆)に記載された感情識別器とすることもできる。このシステムは、感情を一般的なカテゴリ、例えば怒り、恐れおよび緊張に分類する。このような情報を用いることによって、このシステムは送信者とその心理状態に基づき緊急度に合わせて容易にメッセージを分類することができる。

【0035】

例示の一実施形態においては、音響特徴抽出コンポーネント 306 は、入ってくる音声のピッチを抽出し、ピッチ信号から引き出される複数の特徴を基本的なものとして使用する。これらの特徴は、感情識別器 308 にあるサポートベクトルマシンに入力され、各々の文章は嬉しい、悲しいまたは怒っているに分類される。サポートベクトルマシンは例示されるように、それぞれ二値のクラシファイヤである。このため、感情識別器 308 は、各文章に複数の感情が種々の重みをもって存在することを決定することができる。これは、単一の文章中に複数の感情が存在する可能性のあることに対応するものである。したがって、話者感情識別出力 322 は、これらの感情の全てを対応する重みをもって表示するか、または単に最も強い感情、またはその他の感情の任意の組み合わせを表示することができる。

10

【0036】

また、一実施形態においては、音響特徴抽出コンポーネント 306 は、例示するようにメッセージの発話速度も抽出する。これは、多くの異なる手法を用いて行うことができる。例えば、音響特徴抽出コンポーネント 306 は、ケプストラム測定を用いて、新たなボイスメッセージに関連するケプストラムパターンの変化する速度を決定することができる。これによって新たなボイスメッセージについての発話速度についての目安(例えば、1分間の単語数)が示される。

20

【0037】

一実施形態においては、速度標準化コンポーネント 310 が用いられる。この形態のとおり、ユーザは所定の発話速度を入力することができる(あるいは、予め設定されたリストから1つを選択することができる)。次に、速度標準化コンポーネント 310 は、新しいボイスメッセージに関連する発話速度を音響特徴抽出コンポーネント 306 から受信して、そのメッセージに対する発話速度をユーザによって選択された速度に標準化する。速度標準化コンポーネント 310 は、ブロック 324 によって示すように、所定の速度に標準化された速度標準化音声データファイル(例えば、WAVファイル)を出力する。このファイル 324 は、例示するようにユーザインタフェースにおいて用いられ、ユーザが新しいメッセージにアクセスすると、ボイスメッセージが標準化された速度で発生される。当然、システムは元のメッセージも同様に保持することもできる。

30

【0038】

例示の一実施形態においては、発話速度を標準化するために、速度標準化コンポーネント 310 は、新しいボイスメッセージの発話速度を見積もって、既知のタイムスケール変更(modification)アルゴリズムによって各文章の発話速度を調整する。本システムはまた、より効果的に聴取するため、無音およびポーズ間隔の長さを減らすこともできる。

40

【0039】

本発明の他の実施形態にしたがい、情報抽出コンポーネント 210 はまた、音声-テキストコンポーネント 312 をも含む。コンポーネント 312 は、例示するように、新しいメッセージに対応するボイスデータをテキストに転写する音声認識器を含み、それは随意に自然言語処理プロセッサ 314 に提供される。当然、音声-テキストコンポーネント 312 は、単にメッセージ転写 330 を出力することもでき、それはデータ 316 によって示された新しいボイスメッセージの転写全体に相当する。しかし、自然言語処理プロセッサ 314 が提供される場合は、その転写にも自然言語処理を同様に適用することができる。

【0040】

50

一実施形態において、自然言語処理プロセッサ 314 は要約コンポーネント 332 および意味解析器 (semantic parser) 334 を含む。要約コンポーネント 332 は、例示するようにテキストの入力を要約するための既知の処理サブシステムである。したがって、要約コンポーネント 332 は、ボイスメッセージの要旨に対応するメッセージ要約 336 を出力する。

【0041】

意味解析器 334 が提供される一実施形態において、音声 - テキストコンポーネント 312 によって生成されるテキスト転写は、例示されるように意味解析器 334 に入力される。意味解析器 334 はテキストの入力の意味解析を生成して、意味ラベルをテキスト入力の所定の部分に割り当て、および意味解析ツリー 338 をその出力において提供する。意味解析ツリーの一例では、ボイスメッセージが格納されている計算機システムによって実行される種々のアプリケーションスキーマに対応するボイスメッセージの部分に意味ラベルを割り当てて出力され、ボイスメッセージをそのスキーマにより容易に適合させることができる。

10

【0042】

情報抽出コンポーネント 210 がこれらの出力全てを生成すると、(図 2 に示す) 規則適用コンポーネント 212 は、ボイスメッセージの構成を強化するためにユーザ指定の規則をボイスデータ 316 および抽出された情報 (320、322、324、330、336 および 338) に基づいて実行することができる。例えば、ユーザは規則を選択して、規則適用コンポーネント 212 に、ボイスメッセージを話者でソートし、それらをフィルタにかけて異なるディレクトリに入れ、(メッセージ要約 336 のような) 件名に基づいてメッセージをソートしあるいはフィルタにかけ、または日付でソートさせることができる。規則適用コンポーネント 212 は、その他の規則、例えば話者の感情 322、話者の個人情報 320 または (メッセージ転写 330、メッセージ要約 336 あるいは意味解析 338 からの) メッセージ内容のようなメッセージの所定の属性に基づいてユーザに警告するなどのためにも用いることができる。規則適用コンポーネント 212 は、所定の人たちからのメッセージまたは受信してから所定の時間が経過したメッセージを削除するように構成することもできる。規則適用コンポーネント 212 は、予め定められた基準、例えば格納されたメッセージの数、話者の個人情報 320、話者の感情 322 などに基づいて警告を生成することもできる。当然、規則適用コンポーネント 212 も同様に、他の広範な種々の規則を適用することができる。

20

30

【0043】

図 5 は、本発明の一例によるユーザインタフェースの一実施形態を示す。他の広範な種々のユーザインタフェースが用いられるか、またはユーザインタフェースは図 5 に示されるのと同じ情報を含むが、異なる構成とすることができることは当然理解されるであろう。図 5 はユーザインタフェース 400 を示しており、これは表示部 402 およびツールバー部 404 を含む。表示部 402 は、WAV ファイル 403 を一般的に表す表示が生成して示されるか、あるいは現時点で選択されたボイスメッセージの音響的表現である。表示部 402 は、テキストの転写 405 も表示して示し、テキストの要約またはそれらもしくは他の情報アイテムとのいずれかの組み合わせを示すことができる。表示部 402 は、例示するように発信者の電話番号とともに発信者の個人情報と発信日時を表示する表示部分 406 を含む。

40

【0044】

ツールバー部 404 は、例示するように種々のユーザ起動入力を含み、ユーザはこれを起動してボイスメッセージを処理し、または構成することができる。図 5 に示される入力は、例えばメッセージを削除するための削除入力 408、またはボイスメッセージの再生の開始および停止を行うための開始ボタン 410 および停止ボタン 412 を含む。図 5 はスロー / 早送り再生ワイパー 416 を示しており、これによってユーザはボイスメッセージの再生速度を遅くしたり早くしたりすることが可能となる。図 5 のインタフェース 400 は、他のユーザ起動入力、例えばメッセージを格納および印刷するために用いられる「フ

50

ファイル」並びに「印刷」もまた含むことができ、および古いメッセージまたは新しいメッセージを検索するために用いられる「メッセージ取り込み」並びに「新しいメッセージ」を含むこともできる。インタフェース400はまた、例示されるように、メッセージを所定の速度に自動的に標準化させる起動セクタ418をも含むことができる。さらに、インタフェース400は、例示するように、読み取った感情を表示する感情表示420を含む。当然、ユーザインタフェースは広範な種々の起動入力を含み、ユーザはテキスト、音響情報または付加情報を表示し、および異なる規則を適用する等ができるようにユーザインタフェースを構成することができる。

【0045】

以上により、本発明が所定の情報を抽出し、所定の情報でボイスメッセージデータ記憶を付加するための分散型プロセッサを提供することが理解される。所定の情報は、例示されるように本来的に、ユーザがボイスメッセージを整理し、ソートしおよび見直しあるいは処理するのを支援する。

10

【0046】

本発明は所定の実施形態を参照して記述してきたが、本発明の精神と範囲から離れない限り形態や細部について変更することができることを当業者は理解するであろう。

【図面の簡単な説明】

【図1】本発明を使用することができる環境を例示するブロック図である。

【図2】本発明によるシステムを示す詳細なブロック図である。

【図3】図2に示すシステムの動作を一般的に示すフロー図である。

20

【図4】本発明の一実施形態によるボイスデータ処理システムのより詳細なブロック図である。

【図5】本発明によるユーザインタフェースの一実施形態を例示する図である。

【符号の説明】

- 110 コンピュータ
- 120 プロセッシングユニット
- 121 システムバス
- 130 システムメモリ
- 131 読取専用メモリ
- 132 ランダムアクセスメモリ
- 133 基本入出力システム
- 134 オペレーティングシステム
- 135 アプリケーションプログラム
- 136 プログラムモジュール
- 137 プログラムデータ
- 140 固定メモリインタフェース
- 141 磁気ハードディスクドライブ装置
- 144 オペレーティングシステム
- 145 アプリケーションプログラム
- 146 プログラムモジュール
- 147 プログラムデータ
- 150 取外し可能メモリインタフェース
- 151 磁気ディスクドライブ装置
- 152 取外し可能な磁気ディスク
- 155 光ディスクドライブ装置
- 156 取外し可能な光ディスク
- 160 ユーザ入力インタフェース
- 161 ポインティング装置
- 162 キーボード
- 163 マイクロホン

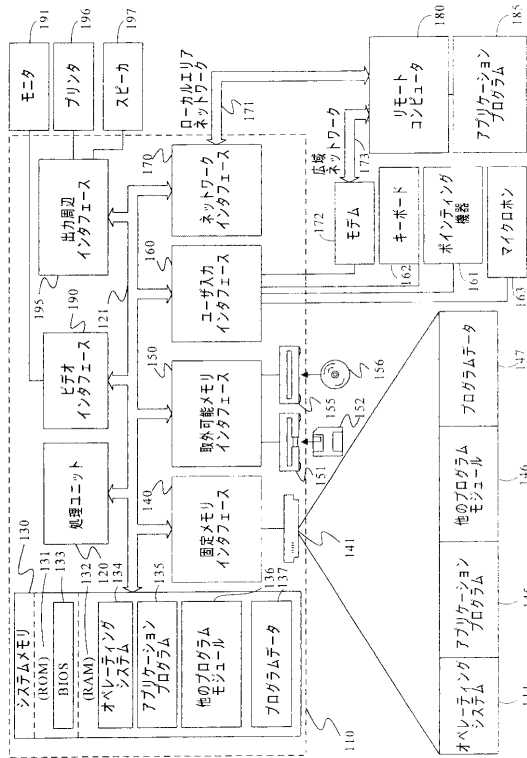
30

40

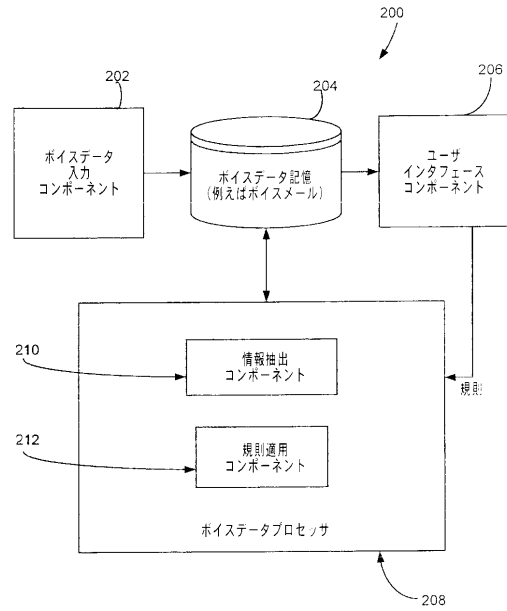
50

170	ネットワークインタフェースアダプタ	
171	ローカルネットワーク	
172	モデム	
173	広域ネットワーク	
180	リモートコンピュータ	
185	リモートアプリケーション	
190	ビデオアダプタ	
191	モニタ	
195	出力周辺インタフェース	
196	プリンタ	10
197	スピーカ	
200	ボイスメッセージ処理システム	
202	ボイスデータ入力コンポーネント	
204	ボイスデータ記憶	
206	ユーザインタフェースコンポーネント	
208	ボイスデータプロセッサ	
210	情報抽出コンポーネント	
212	規則適用コンポーネント	
300	モデルトレーニングコンポーネント	
302	話者識別コンポーネント	20
304	話者識別モデル	
306	音響特徴抽出コンポーネント	
308	感情識別器	
310	速度標準化コンポーネント	
312	音声 - テキストコンポーネント	
314	自然言語処理プロセッサ	
316	データ記憶からの新しいメッセージボイスデータ	
320	音声ID	
322	話者感情	
324	速度標準化音声データファイル	30
330	メッセージ転写	
332	要約コンポーネント	
334	意味解析器	
336	メッセージ要約	
338	意味解析	
400	ユーザインタフェース	
402	表示部	
403	WAVファイルを一般的に示す表示	
404	ツールバー	
405	テキストの転写	40
406	発信者の電話番号とともに発信者の個人情報と発信日時を表示する表示部分	
408	削除入力	
410	開始ボタン	
412	停止ボタン	
416	スロー/早送り再生ワイパー	
418	起動セレクタ	
420	感情表示	

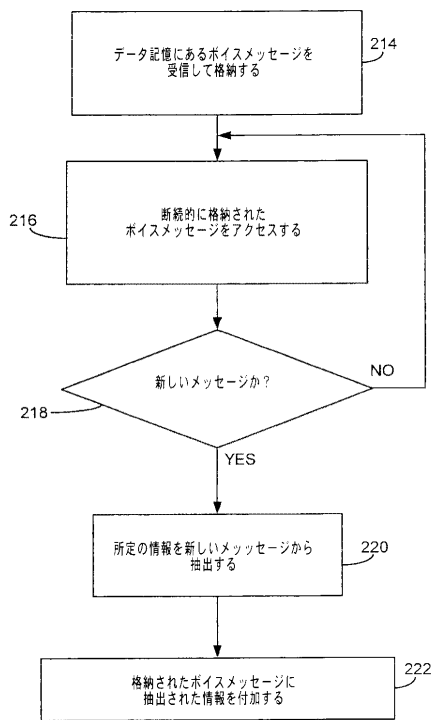
【図 1】



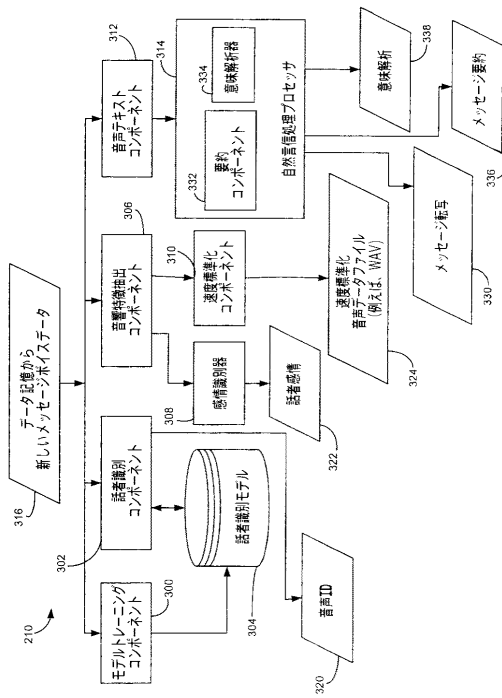
【図 2】



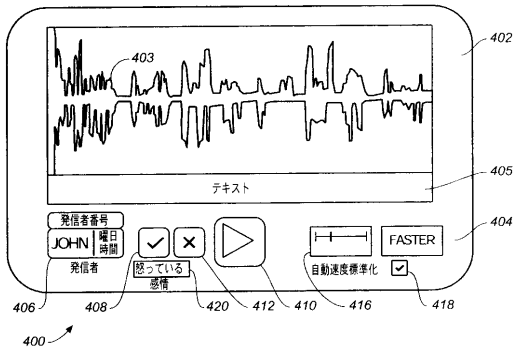
【図 3】



【図 4】



【図5】



フロントページの続き

- (56)参考文献 特開2001-024781(JP,A)
特開2001-237948(JP,A)
特開昭57-013493(JP,A)
国際公開第01/016938(WO,A1)
特開2000-242300(JP,A)
特開2000-010578(JP,A)
特開2001-101099(JP,A)
特開平07-250166(JP,A)
特開2001-292240(JP,A)
特開平09-284399(JP,A)
特開2000-209364(JP,A)
特開平10-285286(JP,A)
特開2002-032213(JP,A)

(58)調査した分野(Int.Cl., DB名)

H04M 1/00、 1/24- 1/253、
1/58- 3/00、 3/16- 3/20、
3/38- 3/58、 7/00- 7/16、
11/00-11/10、 99/00