

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4478939号
(P4478939)

(45) 発行日 平成22年6月9日(2010.6.9)

(24) 登録日 平成22年3月26日(2010.3.26)

(51) Int.Cl.	F I
G 1 0 L 15/10 (2006.01)	G 1 0 L 15/10 5 0 0 N
G 1 0 L 15/06 (2006.01)	G 1 0 L 15/06 3 0 0 Y
G 1 0 L 13/08 (2006.01)	G 1 0 L 13/08 1 3 0 C

請求項の数 4 (全 14 頁)

(21) 出願番号 特願2004-287943 (P2004-287943)
 (22) 出願日 平成16年9月30日(2004.9.30)
 (65) 公開番号 特開2006-98993 (P2006-98993A)
 (43) 公開日 平成18年4月13日(2006.4.13)
 審査請求日 平成18年4月14日(2006.4.14)

特許権者において、実施許諾の用意がある。

(73) 特許権者 393031586
 株式会社国際電気通信基礎技術研究所
 京都府相楽郡精華町光台二丁目2番地2
 (74) 代理人 100099933
 弁理士 清水 敏
 (72) 発明者 ニック・キャンベル
 京都府相楽郡精華町光台二丁目2番地2
 株式会社国際電気通信基礎技術研究所内

審査官 毛利 太郎

最終頁に続く

(54) 【発明の名称】 音声処理装置およびそのためのコンピュータプログラム

(57) 【特許請求の範囲】

【請求項 1】

学習用音声コーパスを記憶するための学習用音声コーパス記憶手段と、
 前記学習用音声コーパスに含まれる音声の所定の発話単位ごとに音響特徴量を抽出するための特徴量抽出手段と、

前記所定の発話単位ごとに、再生時に聴者が知覚するパラ言語的情報に関する統計情報を収集するための統計収集手段と、

前記特徴量抽出手段により抽出された音響特徴量を入力データ、前記統計収集手段により収集された統計情報を正解データとして、機械学習により、音響特徴量に対して最適化された統計情報を出力する学習を行なうための学習手段とを含む、音声処理装置。

10

【請求項 2】

発話単位データに関する音響特徴量を与えられると、当該発話単位の再生時に聴者が所定の複数通りのパラ言語情報ラベルのいずれを選択するかを、前記パラ言語情報ラベルに対する確率の形で出力するパラ言語情報出力手段と、

入力音声データの発話単位から音響特徴量を抽出するための音響特徴量抽出手段と、

前記音響特徴量抽出手段が抽出した音響特徴量を前記パラ言語情報出力手段に与え、応答して前記パラ言語情報出力手段により返される前記パラ言語情報ラベルごとの確率と、前記音響特徴量とに基づいて、前記発話単位に関する発話者の発話意図を推定するための発話意図推定手段とを含む、音声処理装置。

【請求項 3】

20

前記発話意図推定手段は、前記音響特徴量と前記パラ言語情報ラベルごとの確率とに基づいて音声認識とパラ言語情報の認識とを行なう、請求項 2 に記載の音声処理装置。

【請求項 4】

コンピュータにより実行されると、請求項 1～請求項 3 のいずれかに記載の音声処理装置として当該コンピュータを動作させる、コンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

この発明は音声認識、音声合成などの音声処理技術に関し、特に、韻律以外のパラ言語情報を適切に処理できる音声処理技術に関する。

10

【背景技術】

【0002】

人間は、様々な方法により感情を表現する。音声では、情報を伝えるのと同時に、発話スタイル、音声の調子、およびイントネーションなどの変化により個人的な感情を表すことが多い。コンピュータを用いた音声処理技術では、そのような感情をいかにして表したり、理解したりするかが問題となる。

【0003】

非特許文献 1、2、3 においては、音声を自動的に分析する上で、発話を二つの主要なタイプに分類することが提案されている。すなわち、I タイプと A タイプとである。I タイプの発話は主として情報を伝達するために用いられる。A タイプの発話は主として感情を表現するために用いられる。I タイプはその発話内容をテキスト情報のみでほぼ正確に表現できるが、A タイプはその意味内容は曖昧であって、その意味を解釈しようとするれば、発話の韻律に関する知識が必要になる。

20

【0004】

例えば、非特許文献 1、4 は「(英語における) E h」という発話に注目し、会話の文脈に関する情報なしでこの間投発話のみを聞かされた者が、ほぼ例外無しに情緒的な、談話に関連した機能を持つ、というラベルをこの発話に付すことを明らかにしている。実際に選択されたラベルが厳密に一致しているわけではないが、知覚結果は概略において一致している。ある日本語の発話に対して、韓国語を母語とする話者、および米国語を母語とする話者の双方が割当てる意味がほぼ一致していることから、こうした能力は言語や文化とは無関係であるように思われる。

30

【非特許文献 1】N. キャンベル、「行間を聞く：音調により伝達されるパラ言語情報に関する研究」、言語のトーンに関する局面についての国際シンポジウム T A L 2 0 0 4 予稿集、pp. 13 - 16、2004 年 (N. Campbell, "Listening between the lines: a study of paralinguistic information carried by tone-of-voice", in Proc. International Symposium on Tonal Aspects of Languages, TAL2004, pp. 13-16, 2004)

【非特許文献 2】N. キャンベル、「事柄の本質に到達する」、言語資源および評価会議キーノートスピーチ、2004 年 (N. Campbell, "Getting to the heart of the matter", Keynote speech in Proc. Language Resources and Evaluation Conference (LREC-04), 2004, <http://feast.his.atr.jp/nick/pubs/lrec-keynote.pdf>)

40

【非特許文献 3】N. キャンベル、「意味外のプロトコル：対話合成のための入力要件」、情緒的対話システム、E. アンドレ他編、シュプリンガー・フェアラーク社刊、2004 年 (N. Campbell, "Extra-Semantic Protocols: Input Requirements for the Synthesis of Dialogue Speech" in Affective Dialogue Systems, Eds. Andre, E., Dybkjaer, L., Minker, W., & Heisterkamp, P., Springer Verlag, 2004)

【非特許文献 4】N. キャンベル他、「人は何を聞くのか？対話音声内の非言語・情緒的情報の知覚に関する研究」、日本音声学会誌、第 7 巻、第 4 号、2004 年 (N. Campbe

50

II & D. Erickson, "What do people hear? A study of the perception of non-verbal affective information in conversational speech", in Journal of the Phonetic Society of Japan, vol. 7, no. 4, 2004)

【発明の開示】

【発明が解決しようとする課題】

【0005】

しかし、たとえば発話に付随するパラ言語情報をコンピュータを用いた自然言語処理で処理しようとする、大きな困難にぶつかる。たとえばテキストとしてみると同一の発話であっても、それが使われる状況によって全く意味が異なったり、全く異なる感情を同時に表現したりすることがある。そうした場合、発話の音響的な特徴のみからパラ言語情報を取出すことは極めて困難である。

10

【0006】

そうした問題を解決する一つの手法は、聴者に発話を聞かせ、その発話からその聴者が感じ取ったパラ言語的な情報に基づき、その発話にラベルを付することである。

【0007】

しかし、発話内容の理解は人によって異なり、ある特定の話者のみによるラベル付けでは信頼性が期待できないという問題がある。

【0008】

したがって本発明の目的は、パラ言語情報を適切に処理できる音声処理装置を提供することである。

20

【0009】

本発明の他の目的は、パラ言語情報を適切に処理できるようにすることで、音声処理の適用範囲を広げることができる音声処理装置を提供することである。

【課題を解決するための手段】

【0010】

本発明の第1の局面によれば、音声処理装置は、学習用音声コーパスを記憶するための学習用音声コーパス記憶手段と、学習用音声コーパスに含まれる音声の所定の発話単位ごとに音響特徴量を抽出するための特徴量抽出手段と、所定の発話単位ごとに、再生時に聴者が知覚するパラ言語情報に関する統計情報を収集するための統計収集手段と、特徴量抽出手段により抽出された音響特徴量を入力データ、統計収集手段により収集された統計情報を正解データとして、機械学習により、音響特徴量に対して最適化された統計情報を出力する学習を行なうための学習手段とを含む。

30

【0011】

発話単位を再生したときに聴者がどのようなパラ言語情報を知覚するかに関する統計を収集する。学習手段は、収集された統計に基づいた機械学習により、音響特徴量が与えられると、学習に用いたデータを一般化して得られる、もっともらしい統計情報を出力する。音声に対してパラ言語情報を統計的情報として付することが可能になり、パラ言語情報を適切に処理することが可能になる。

【0012】

好ましくは、統計収集手段は、所定の発話単位ごとに、パラ言語情報を表す所定の複数通りのラベルの中から聴者が選択する確率をラベルごとに算出するための手段を含む。

40

【0013】

さらに好ましくは、学習手段は複数通りのラベルに対応してそれぞれ設けられた複数個のラベル統計学習手段を含み、複数個のラベル統計学習手段の各々は、特徴量抽出手段により抽出された音響特徴量を入力データ、統計収集手段により当該ラベル統計学習手段に対応するラベルに対し算出された確率を正解データとして、機械学習により、音響特徴量に対して当該ラベルが聴者により選択される確率を出力するように学習する。

【0014】

発話単位に対するパラ言語情報として、予め定められた複数通りのラベルの各々が聴者により選択される確率が得られる。種々の聴者に対する学習の結果として、聴者が知覚す

50

るパラ言語情報を数量化でき、音声処理の際のパラ言語情報の再現および解釈の精度が向上する。

【 0 0 1 5 】

本発明の第2の局面によれば、音声処理装置は、発話単位データに関する音響特徴量が与えられると、当該発話単位の再生時に聴者が所定の複数通りのパラ言語情報ラベルのいずれを選択するかを、パラ言語情報ラベルに対する確率の形で出力するパラ言語情報出力手段と、入力音声データの発話単位から音響特徴量を抽出するための音響特徴量抽出手段と、音響特徴量抽出手段が抽出した音響特徴量をパラ言語情報出力手段に与え、応答してパラ言語情報出力手段により返されるパラ言語情報ラベルごとの確率と、音響特徴量とに基づいて、発話単位に関する発話者の発話意図を推定するための発話意図推定手段とを含む。

10

【 0 0 1 6 】

入力発話に付随するパラ言語情報を、聴者により複数のパラ言語情報がそれぞれ知覚される確率として獲得することができる。それらパラ言語情報確率の集まりに基づき、発話の意味を精度よく推定できる。

【 0 0 1 7 】

本発明の第3の局面によれば、音声処理装置は、発話単位データに関する音響特徴量が与えられると、当該発話単位の再生時に聴者が所定の複数通りのパラ言語情報ラベルのいずれを選択するかを、複数通りのパラ言語情報ラベルにそれぞれ対応する複数の確率の形で出力するパラ言語情報出力手段と、入力音声データの発話単位から音響特徴量を抽出するための音響特徴量抽出手段と、所定の音声コーパスに含まれる発話単位データごとに、音響特徴量抽出手段により抽出された音響特徴量に対してパラ言語情報出力手段から出力される複数の確率をパラ言語情報ベクトルとして付することにより、パラ言語情報ベクトル付音声コーパスを生成するための手段とを含む。

20

【 0 0 1 8 】

音声コーパスに含まれる各発話単位データに対し、複数通りのパラ言語情報について聴者がそれぞれ知覚する確率という形でパラ言語情報ベクトルを作成し付することができる。このように作成されたパラ言語情報ベクトル付音声コーパスを用いることにより、音声理解、音声合成などにおいてパラ言語情報をより精度よく利用することが可能になる。

【 0 0 1 9 】

30

本発明の第4の局面によれば、音声処理装置は、パラ言語情報ベクトルが各々に付され、かつ音素ラベルを含む所定の音響特徴量が各々に付された複数の音声波形データを含む音声コーパスと、音声合成の目標となるテキストと、当該テキストの発話意図を表す発話意図情報とが与えられると、音声合成の韻律合成目標と、発話意図に対応するパラ言語情報目標ベクトルとを作成するための合成目標作成手段と、合成目標作成手段により作成された韻律合成目標およびパラ言語情報目標ベクトルに対し所定の条件を充足する音響特徴量およびパラ言語情報ベクトルを有する音声波形データを音声コーパス内に含まれる音声波形データから選択するための波形選択手段と、波形選択手段により選択された音声波形データを接続することにより、音声波形を出力するための波形接続手段とを含む。

【 0 0 2 0 】

40

この音声処理装置によれば、テキストと発話意図情報とが与えられると、テキストに合致した音響的特徴を持ち、かつ発話意図情報に合致したパラ言語情報ベクトルを持つ波形データを精度よく選択できる。その結果、テキスト内容だけではなく、パラ言語情報として発話の内容を精度よく聴者に伝達する音声を作成することができる。

【 0 0 2 1 】

本発明の第5の局面に係るコンピュータプログラムは、コンピュータにより実行されると、上記したいずれかの音声処理装置として当該コンピュータを動作させる。

【 0 0 2 2 】

本発明の第6の局面に係る記録媒体は、音声波形データを対応する音素情報とともに保持する音声コーパスを記録した記録媒体であって、音声コーパスは、複数の発話単位の各

50

々に対して音声波形データと音素情報とを含み、複数の発話単位の各々にはさらに、当該発話単位の再生時に聴者が知覚するパラ言語情報に関する統計情報が付されている。

【 0 0 2 3 】

好ましくは、パラ言語情報に関する統計情報は、予め定められた複数種類のパラ言語情報の各々について、対応する発話単位の再生時に聴者が当該パラ言語情報を知覚する確率を含んでいる。

【発明を実施するための最良の形態】

【 0 0 2 4 】

〔 概 略 〕

音声における感情に関する情報についてのラベリングでは、ラベリングをする人が異なればその結果も異なる。また、例えば疑問文があいづちを意味したり、ときには笑いが、驚きとともに聴者も話者と同じく幸せを感じていることを表したりする、ということがあることも分かっている。幸福を感じている人が、自分とは直接関係を持たない何か悲しいことを話しているときには、幸福と不幸という、一見したところ互いに矛盾する感情が音声によって表されることもある。

【 0 0 2 5 】

こうした状況を考えると、音声に対して一つのラベルに限定してラベリングをするよりも、複数のラベルを用いて音声のラベリングをすることの方が合理的である。したがって以下に説明する実施の形態では、予め複数通りのラベルを定め、音声の各発話単位に対して統計的にどれほどの割合の人がそれらラベルをそれぞれ選択したかを表す数値を要素とするベクトルによって、各音声のラベリングを行なう。このベクトルを、以下「パラ言語情報ベクトル」と呼ぶことにする。

【 0 0 2 6 】

〔 第 1 の実施の形態 〕

- 構成 -

図 1 は、本発明の第 1 の実施の形態に係る音声理解システム 20 のブロック図である。図 1 を参照して、音声理解システム 20 は、発話の音響情報が与えられると、前述したパラ言語情報ベクトルの各要素に対応するラベルが当該発話に付される確率を要素ごとに決定する決定木群 38 を用いる点に特徴がある。すなわち、決定木群 38 はパラ言語情報を構成する要素に対応する数だけの決定木を含んでいる。第 1 の決定木は 1 番目の要素のラベルが付される確率を出力し、第 2 の決定木は 2 番目の要素のラベルが付される確率を出力し、以下同様である。パラ言語情報ベクトルの各要素の値は、0 ~ 1 の範囲に正規化されているものとする。

【 0 0 2 7 】

図 1 を参照して、この音声理解システム 20 は、学習用音声コーパス 30 と、スピーカ 32 および入力装置 34 に接続され、学習用音声コーパス 30 内の音声の各音素に対し、所定数の被験者によってどのようなラベルが付されたかという統計的データを収集し、収集されたデータに基づいて決定木群 38 内の各決定木の学習を行なわせるための決定木学習部 36 とを含む。この決定木学習部 36 による学習によって、決定木群 38 の各決定木は、音響情報が与えられると、前述した所定数の被験者の中のどの程度の割合のものが各要素に対応するラベリングをするか、という確率を出力するように設定される。

【 0 0 2 8 】

音声理解システム 20 はさらに、入力音声データ 50 が与えられると、入力音声データ 50 に対する音声認識を行なうとともに、決定木群 38 を用いて入力音声データ 50 が表す感情についてまで含めた音声理解を行ない、認識テキストと入力音声データ 50 の発話者の意図を表す発話意図情報とからなる音声解釈結果 58 を出力するための音声認識装置 40 とを含む。

【 0 0 2 9 】

図 2 を参照して、決定木学習部 36 は、学習用音声コーパス 30 の音声に対し被験者が割当てたラベルを、対応する音声データとともに学習用の統計情報として収集する処理を

10

20

30

40

50

行なうためのラベル付け処理部 70 を含む。学習用音声コーパス 30 の音声はスピーカ 32 により再生される。被験者はこの音声に対してラベルを割当て、入力装置 34 を用いて決定木学習部 36 に与える。

【0030】

決定木学習部 36 はさらに、ラベル付け処理部 70 により蓄積された学習データを記憶するための学習データ記憶部 72 と、学習データ記憶部 72 に記憶された学習データの中の発話音声データに対する音響分析を行なって所定の音響特徴量を出力するための音響分析部 74 と、学習データ記憶部 72 に記憶された学習データ中で、ある音素に対しどの程度の割合の被験者がどのラベルを割当てたかを統計処理し、その結果をラベルごとに出力するための統計処理部 78 とを含む。

10

【0031】

決定木学習部 36 はさらに、音響分析部 74 から与えられた音響特徴量を学習データ、その音声に対し決定木群 38 内の各決定木に対応する特定のラベルが割当てられた確率を正解データとして決定木群 38 内の各決定木の学習を機械学習により行なうための学習処理部 76 を含む。決定木学習部 36 の学習により、決定木群 38 は、与えられた音響特徴量に対して最適化された統計情報を出力するようになる。すなわち決定木群 38 は、ある音声についての音響特徴量が与えられると、その音声について前述した各ラベルが被験者によって割当てられる確率としてもっともらしい値を推定し出力するようになる。

【0032】

決定木学習部 36 は、図では決定木群 38 に対し一つのみ示されているが、決定木群 38 に含まれる決定木ごとに、対応のラベルが聴者により選択される確率を統計情報に基づいて推定するような、ラベル統計による学習を行なうための機能部を決定木の数と等しい数だけ含んでいる。

20

【0033】

図 3 を参照して、音声認識装置 40 は、入力音声データ 50 に対し音響分析部 74 と同様の音響分析を行ない音響特徴量を出力するための音響分析部 52 と、音響分析部 52 の出力する音響特徴量を決定木群 38 の各決定木に与え、応答して各決定木から返される確率をラベルごとに所定の順序で並べることにより、入力音声データ 50 の発話者の意図を推定し、発話者の意図（発話の意味）を表す発話意図ベクトルを生成するための発話意図ベクトル生成部 54 と、発話意図ベクトル生成部 54 から与えられる発話意図ベクトルと、音響分析部 52 からの音響特徴量とを入力として音声認識とその意味的な理解とを行ない、音声解釈結果 58 を出力するための音声理解部 56 とを含む。音声理解部 56 は、予め学習用音声コーパスと、学習用音声コーパスの各発話に対するパラ言語情報ベクトルと、その発話に対する被験者の意味的理解の結果とを入力として学習しておいた意味理解モデルを用いて実現できる。

30

【0034】

- 動作 -

音声理解システム 20 の動作には二つのフェーズが存在する。第 1 のフェーズは決定木学習部 36 による決定木群 38 の学習フェーズである。第 2 のフェーズは、このようにして学習の終了した決定木群 38 を用い、音声認識装置 40 が入力音声データ 50 に対する意味理解を行なう動作フェーズとである。以下、順に説明する。

40

【0035】

・学習フェーズ

学習フェーズに先立ち、学習用音声コーパス 30 が予め準備されているものとする。所定数（例えば 100 名）の被験者が予め選抜され、学習用データとして所定数（例えば 100 個）の発話が定められているものとする。

【0036】

図 2 に示すラベル付け処理部 70 は、第 1 の被験者に対し、学習用音声コーパス 30 から第 1 の発話を取り出してスピーカ 32 を用いて再生する。被験者は、再生された音声に対し感じ取ったパラ言語的な情報を、予め定められた複数通りのラベルのうちのいずれかに

50

割当て、入力装置 34 を介してラベル付け処理部 70 に与える。ラベル付け処理部 70 は、1 番目の発話に対して 1 番目の被験者が割当てたラベルを当該音声データを特定する情報とともに学習データ記憶部 72 に蓄積する。

【0037】

ラベル付け処理部 70 はさらに、学習用音声コーパス 30 から次の発話を読み出し、上記処理と同様の処理を第 1 番目の被験者に対して行なう。以下同様である。

【0038】

第 1 番目の被験者に対し全ての学習用発話を用いて上記した処理を行なうことにより、この第 1 番目の被験者が各学習用発話に対しどのラベルを割当てたか、という情報が蓄積される。

10

【0039】

こうした処理を全ての被験者に対し繰返し行なうことにより、各学習用発話に対し、どのラベルがどれだけの回数割当てられたか、という情報が蓄積される。

【0040】

全ての被験者に対して上記した処理が終了すると、決定木群 38 の学習が以下のように行なわれる。全ての発話について、音響分析部 74 が音響分析を行ない、音響特徴量を学習処理部 76 に与える。統計処理部 78 は、全ての発話に対し、どのラベルがどの程度の確率で割当てられたか、という統計的処理を行ない、その結果を学習処理部 76 に与える。

【0041】

20

学習処理部 76 は、決定木群 38 に含まれる各決定木に対し学習を行なう。この際の学習データとしては、音響分析部 74 から与えられる各発話の音響特徴量が用いられる。正解データとしては、その発話について当該決定木に対応するラベルが割当てられた確率が用いられる。この確率は統計処理部 78 から与えられる。全ての発話についてこの学習処理が完了すると、音声認識装置 40 による音声理解が可能になる。

【0042】

・動作フェーズ

動作フェーズでは、入力音声データ 50 が与えられると音響分析部 52 がその発話に対する音響分析を行ない、音響特徴量を抽出して発話意図ベクトル生成部 54 および音声理解部 56 に与える。発話意図ベクトル生成部 54 は、決定木群 38 の各決定木に対し、音響分析部 52 から与えられた音響特徴量を与える。決定木群 38 の各決定木は、それぞれに対応するラベルがその発話に割当てられる確率を発話意図ベクトル生成部 54 に返す。

30

【0043】

発話意図ベクトル生成部 54 は、ラベルごとに受取った確率を所定の順番で要素とする発話意図ベクトルを生成し、音声理解部 56 に与える。

【0044】

音声理解部 56 は、音響分析部 52 から与えられる音響特徴量と、発話意図ベクトル生成部 54 から与えられる発話意図ベクトルとに基づいて、入力音声データ 50 の音声認識結果のテキストと、入力音声データ 50 の発話者の意図を表す発話意図情報との組合せとして確率の高い上位所定個数の音声解釈結果 58 を出力する。

40

【0045】

以上のようにして、本実施の形態に係る音声理解システム 20 によれば、単に入力音声データに対する音声認識を行なうだけではなく、入力音声データの背後にある発話者の意図まで含めた、発話の意味的な理解を行なうことが可能となる。

【0046】

なお、本実施の形態では、学習用音声コーパス 30 からの学習に決定木を用いている。しかし本発明はそのような実施の形態には限定されない。決定木に代えて、ニューラルネットワーク、隠れマルコフモデル (HMM) など、任意の機械学習手段を用いてもよい。これは、後に説明する第 2 の実施の形態以下でも同様である。

【0047】

50

〔第２の実施の形態〕

第１の実施の形態に係るシステムは、入力音声データ５０に対する意味的な理解を可能とするものであった。決定木群３８と、このシステムの動作原理を利用すると、音声コーパスに含まれる各発話に対し、意味的な情報を表す発話意図ベクトルでラベリングをすることができる。図４にそのための音声コーパスラベリング装置８０の概略構成を示す。

【００４８】

図４を参照して、音声コーパスラベリング装置８０は、第１の実施の形態で用いたものと同じ決定木群３８と、ラベリング対象となる音声コーパス９０から音声データを読み出すための音声データ読み出し部９２と、音声データ読み出し部９２により読み出された音声データに対する音響分析を行ない、音響特徴量を出力するための音響分析部９４と、音響分析部９４から与えられる音響特徴量を決定木群３８の各決定木に与え、各決定木から返される確率を所定の順番で並べて要素とする発話意図ベクトルを生成するための発話意図ベクトル生成部９６と、発話意図ベクトル生成部９６により生成された発話意図ベクトルで音声コーパス９０内の対応する発話に対するラベリングを行なうためのラベリング処理部９８とを含む。

10

【００４９】

図５に、音声コーパス９０に含まれる音声データ１１０の構成を示す。図５を参照して、音声データ１１０は、音声の波形データ１１２を含む。波形データ１１２は複数の発話波形データ１１４、１１６、１１８、…、１２０、…を含む。

【００５０】

各発話、例えば発話波形データ１１８には、韻律情報１３０が付されている。韻律情報１３０は、発話波形データ１１８の表す音韻、波形データ１１２の先頭から測定した発話波形データ１１８の開始時間および終了時間、音響特徴量等に加え、図４に示す発話意図ベクトル生成部９６により付された発話意図ベクトルがパラ言語情報ベクトルとして含まれている。

20

【００５１】

このように音声コーパス９０の各発話にパラ言語情報ベクトルを付しておくことで、音声コーパス９０はパラ言語情報ベクトル付音声コーパスとなる。パラ言語情報ベクトル付音声コーパス９０を用いることで、例えば音声合成において、単にテキストに対応し、かつ音韻的に自然な音声にとどまらず、所望の発話意図に沿ったパラ言語情報を持った音声を合成することが可能になる。

30

【００５２】

〔第３の実施の形態〕

- 構成 -

第３の実施の形態は、第２の実施の形態の音声コーパスラベリング装置８０によりラベリングされた後の音声コーパス９０と同様の音声コーパスを用いた音声合成装置に関する。図６に第３の実施の形態に係る音声合成装置１４２のブロック図を示す。この音声合成装置１４２は、発話条件情報が付された入力テキスト１４０を受け、入力テキストに応じた自然な音声であって、かつ発話条件情報に合致したパラ言語的な情報（感情）を表す出力音声波形１４４を合成する機能を持つ、いわゆる波形接続型音声合成装置である。

40

【００５３】

図６を参照して、音声合成装置１４２は、入力テキスト１４０の入力テキストから韻律合成目標を作成するための韻律合成目標作成部１５６と、入力テキスト１４０に含まれる発話条件情報から、パラ言語情報目標ベクトルを作成するためのパラ言語情報目標ベクトル作成部１５８と、音声コーパスラベリング装置８０によりパラ言語情報ベクトルが付された音声コーパス９０と同様のパラ言語情報ベクトル付音声コーパス１５０と、パラ言語情報ベクトル付音声コーパス１５０から韻律合成目標作成部１５６の出力に応じた複数の波形候補を選択し、その音響特徴量を読み出すための音響特徴量読み出し部１５２と、音響特徴量読み出し部１５２と同じ波形候補のパラ言語情報ベクトルを読み出すためのパラ言語情報読み出し部１５４とを含む。

50

【 0 0 5 4 】

音声合成装置 1 4 2 はさらに、音響特徴量読出部 1 5 2 が読出した各波形候補の音響特徴量およびパラ言語情報読出部 1 5 4 が読出した各波形候補の音響特徴量と、韻律合成目標作成部 1 5 6 の作成した韻律合成目標およびパラ言語情報目標ベクトル作成部 1 5 8 の作成したパラ言語情報目標ベクトルとの間で、韻律合成目標とどの程度異なった音声か、隣接する音声の間の接続がどの程度不連続か、および目標となるパラ言語情報ベクトルと波形候補のパラ言語情報ベクトルとがどの程度相違しているか、を示す尺度となるコストを予め定められた算出式にしたがって算出するためのコスト算出部 1 6 0 と、コスト算出部 1 6 0 が算出した各波形候補のコストに基づき最小コストとなるいくつかの波形候補を選択するための波形選択部 1 6 2 と、波形選択部 1 6 2 により選択された波形候補に対応する波形データをパラ言語情報ベクトル付音声コーパス 1 5 0 から読出して接続することにより、出力音声波形 1 4 4 を出力するための波形接続部 1 6 4 とを含む。

10

【 0 0 5 5 】

- 動作 -

この第 3 の実施の形態に係る音声合成装置 1 4 2 は以下のように動作する。入力テキスト 1 4 0 が与えられると、韻律合成目標作成部 1 5 6 が入力テキストに対するテキスト処理を行ない、韻律合成目標を作成し音響特徴量読出部 1 5 2、パラ言語情報読出部 1 5 4 およびコスト算出部 1 6 0 に与える。パラ言語情報目標ベクトル作成部 1 5 8 は、入力テキスト 1 4 0 から発話条件情報を抽出し、抽出された発話条件情報に基づいてパラ言語目標ベクトルを作成しコスト算出部 1 6 0 に与える。

20

【 0 0 5 6 】

音響特徴量読出部 1 5 2 は、韻律合成目標作成部 1 5 6 から与えられた韻律合成目標に基づき、パラ言語情報ベクトル付音声コーパス 1 5 0 から複数の波形候補を選択しコスト算出部 1 6 0 に与える。パラ言語情報読出部 1 5 4 も同様に、音響特徴量読出部 1 5 2 が読出したものと同じ波形候補のパラ言語情報ベクトルを読出し、コスト算出部 1 6 0 に与える。

【 0 0 5 7 】

コスト算出部 1 6 0 は、韻律合成目標作成部 1 5 6 からの韻律合成目標およびパラ言語情報目標ベクトル作成部 1 5 8 からのパラ言語情報目標ベクトルと、音響特徴量読出部 1 5 2 から与えられた各波形候補の音響特徴量およびパラ言語情報読出部 1 5 4 から与えられた各波形候補のパラ言語情報ベクトルとの間で所定のコスト演算を行ない、その結果を波形候補ごとに波形選択部 1 6 2 に対し出力する。

30

【 0 0 5 8 】

波形選択部 1 6 2 は、コスト算出部 1 6 0 から与えられたコストに基づき、コスト最小の所定個数の波形候補を選択し、当該波形候補のパラ言語情報ベクトル付音声コーパス 1 5 0 内の位置を表す情報を波形接続部 1 6 4 に与える。

【 0 0 5 9 】

波形接続部 1 6 4 は、波形選択部 1 6 2 から与えられた情報に基づき、パラ言語情報ベクトル付音声コーパス 1 5 0 から波形候補を読出し、直前に選択された波形の直後に接続する。複数候補が選択されているため、波形接続部 1 6 4 の処理によって出力音声波形 1 4 4 の候補が複数個作成されるが、所定のタイミングでその中で累積コストが最小のものが選択され出力音声波形 1 4 4 として出力される。

40

【 0 0 6 0 】

以上のとおり本実施の形態に係る音声合成装置 1 4 2 によれば、単に入力テキストにより指定される音韻と合致するだけでなく、入力テキスト 1 4 0 に含まれる発話条件情報に合致したパラ言語情報を伝えることができるような波形候補が選択され、出力音声波形 1 4 4 の生成に用いられる。その結果、入力テキスト 1 4 0 の発話条件情報で指定された発話条件に合致し、所望の感情に関する情報をパラ言語情報として伝達することができる。パラ言語情報ベクトル付音声コーパス 1 5 0 の各波形には、パラ言語情報としてベクトルが付されており、パラ言語情報間のコスト計算がベクトル計算として行われるため、互い

50

に相反した感情を伝達したり、入力テキストの内容とは一見無関係な情報をパラ言語情報として伝達したりすることが可能になる。

【 0 0 6 1 】

〔コンピュータによる実現〕

上述した第 1 の実施の形態に係る音声理解システム 2 0、第 2 の実施の形態に係る音声コーパスラベリング装置 8 0、および第 3 の実施の形態に係る音声合成装置 1 4 2 は、いずれもコンピュータハードウェアと、そのコンピュータハードウェアにより実行されるプログラムと、コンピュータハードウェアに格納されるデータとにより実現される。図 7 はこのコンピュータシステム 2 5 0 の外観を示す。

【 0 0 6 2 】

図 7 を参照して、このコンピュータシステム 2 5 0 は、F D (フレキシブルディスク) ドライブ 2 7 2 および C D - R O M (コンパクトディスク読出専用メモリ) ドライブ 2 7 0 を有するコンピュータ 2 6 0 と、キーボード 2 6 6 と、マウス 2 6 8 と、モニタ 2 6 2 と、スピーカ 2 7 8 およびマイクロフォン 2 6 4 とを含む。スピーカ 2 7 8 は図 1 などに示すスピーカ 3 2 として利用される。キーボード 2 6 6 およびマウス 2 6 8 は、図 1 などに示す入力装置 3 4 として利用される。

【 0 0 6 3 】

図 8 を参照して、コンピュータ 2 6 0 は、F D ドライブ 2 7 2 および C D - R O M ドライブ 2 7 0 に加えて、C P U (中央処理装置) 3 4 0 と、C P U 3 4 0、F D ドライブ 2 7 2 および C D - R O M ドライブ 2 7 0 に接続されたバス 3 4 2 と、ブートアッププログラム等を記憶する読出専用メモリ (R O M) 3 4 4 と、バス 3 4 2 に接続され、プログラム命令、システムプログラム、および作業データ等を記憶するランダムアクセスメモリ (R A M) 3 4 6 とを含む。コンピュータシステム 2 5 0 はさらに、図示しないプリンタを含んでもよい。

【 0 0 6 4 】

コンピュータ 2 6 0 はさらに、バス 3 4 2 に接続され、スピーカ 2 7 8 およびマイクロフォン 2 6 4 が接続されるサウンドボード 3 5 0 と、バス 3 4 2 に接続された大容量の外部記憶装置であるハードディスク 3 4 8 と、バス 3 4 2 を介してローカルエリアネットワーク (L A N) への接続を C P U 3 4 0 に提供するネットワークボード 3 5 2 を含む。

【 0 0 6 5 】

コンピュータシステム 2 5 0 に上記した音声理解システム 2 0 等としての動作を行なわせるためのコンピュータプログラムは、C D - R O M ドライブ 2 7 0 または F D ドライブ 2 7 2 に挿入される C D - R O M 3 6 0 または F D 3 6 2 に記憶され、さらにハードディスク 3 4 8 に転送される。または、プログラムはネットワークおよびネットワークボード 3 5 2 を通じてコンピュータ 2 6 0 に送信されハードディスク 3 4 8 に記憶されてもよい。プログラムは実行の際に R A M 3 4 6 にロードされる。C D - R O M 3 6 0 から、F D 3 6 2 から、またはネットワークを介して、直接に R A M 3 4 6 にプログラムをロードしてもよい。

【 0 0 6 6 】

このプログラムは、コンピュータ 2 6 0 に音声理解システム 2 0 等として動作を行なわせる複数の命令を含む。この動作を行なわせるのに必要な基本的機能のいくつかはコンピュータ 2 6 0 上で動作するオペレーティングシステム (O S) またはサードパーティのプログラム、もしくはコンピュータ 2 6 0 にインストールされる各種ツールキットのモジュールにより提供される。したがって、このプログラムはこの実施の形態のシステムおよび方法を実現するのに必要な機能全てを必ずしも含まなくてよい。このプログラムは、命令のうち、所望の結果が得られるように制御されたやり方で適切な機能または「ツール」を呼出すことにより、上記した音声理解システム 2 0、音声コーパスラベリング装置 8 0 または音声合成装置 1 4 2 としての動作を実行する命令のみを含んでいればよい。コンピュータシステム 2 5 0 の一般的な動作は周知であるので、ここでは繰返さない。

【 0 0 6 7 】

なお、上記した実施の形態の決定木群 38 の各決定木は、コンピュータ上で並列に動作する複数のデーモンとして実現可能である。また、プロセッサを複数個搭載したコンピュータであれば決定木群 38 の各決定木を複数のプロセッサに分散させるようにしてもよい。ネットワーク接続された複数のコンピュータを用いる場合も同様で、複数のコンピュータに 1 または複数の決定木として動作するプログラムを実行させればよい。図 6 に示す音声合成装置 142 において、コスト算出部 160 を複数のデーモンで実現したり、複数のプロセッサにより実行されるプログラムにより実現したりすることもできる。

【0068】

今回開示された実施の形態は単に例示であって、本発明が上記した実施の形態のみに制限されるわけではない。本発明の範囲は、発明の詳細な説明の記載を参酌した上で、特許請求の範囲の各請求項によって示され、そこに記載された文言と均等の意味および範囲内でのすべての変更を含む。

【図面の簡単な説明】

【0069】

【図 1】本発明の第 1 の実施の形態に係る音声理解システム 20 のブロック図である。

【図 2】図 1 に示す決定木学習部 36 のブロック図である。

【図 3】図 1 に示す音声認識装置 40 のブロック図である。

【図 4】本発明の第 2 の実施の形態に係る音声コーパスラベリング装置 80 のブロック図である。

【図 5】音声コーパス 90 内の音声データ 110 の構成を模式的に示す図である。

【図 6】本発明の第 3 の実施の形態に係る音声合成装置 142 のブロック図である。

【図 7】本発明の一実施の形態に係る音声理解システム 20 などを実現するコンピュータシステム 250 の外観図である。

【図 8】図 7 に示すコンピュータ 260 のブロック図である。

【符号の説明】

【0070】

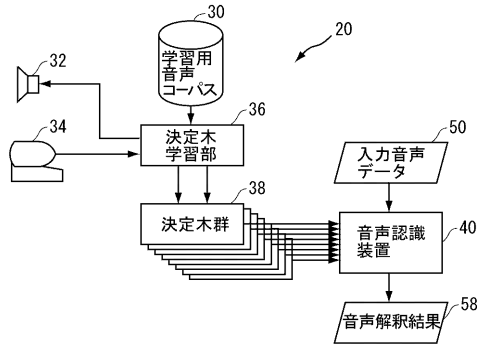
20 音声理解システム、30 学習用音声コーパス、32 スピーカ、34 入力装置、36 決定木学習部、38 決定木群、40 音声認識装置、50 入力音声データ、52 音響分析部、54 発話意図ベクトル生成部、56 音声理解部、58 音声解釈結果、70 ラベル付け処理部、72 学習データ記憶部、74 音響分析部、76 学習処理部、78 統計処理部、80 音声コーパスラベリング装置、90 音声コーパス、92 音声データ読出部、94 音響分析部、96 発話意図ベクトル生成部、140 入力テキスト、142 音声合成装置、144 出力音声波形、150 パラ言語情報ベクトル付音声コーパス、152 音響特徴量読出部、154 パラ言語情報読出部、156 韻律合成目標作成部、158 パラ言語情報目標ベクトル作成部、160 コスト算出部、162 波形選択部、164 波形接続部

10

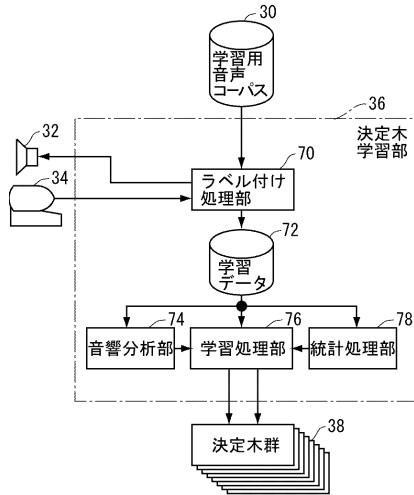
20

30

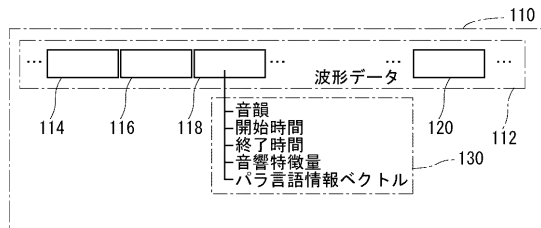
【図 1】



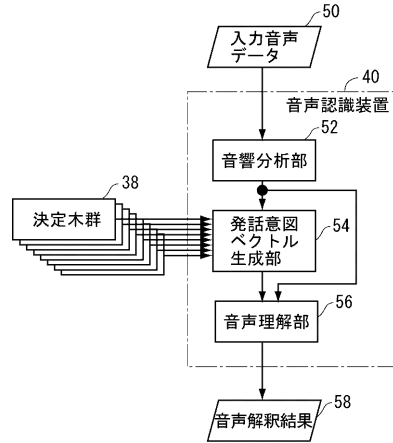
【図 2】



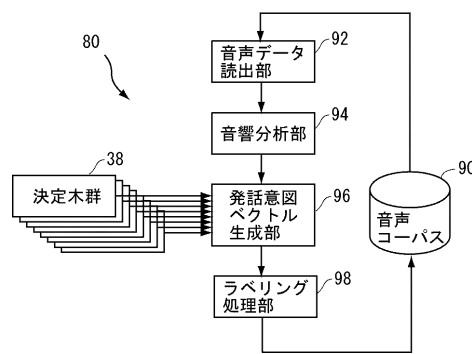
【図 5】



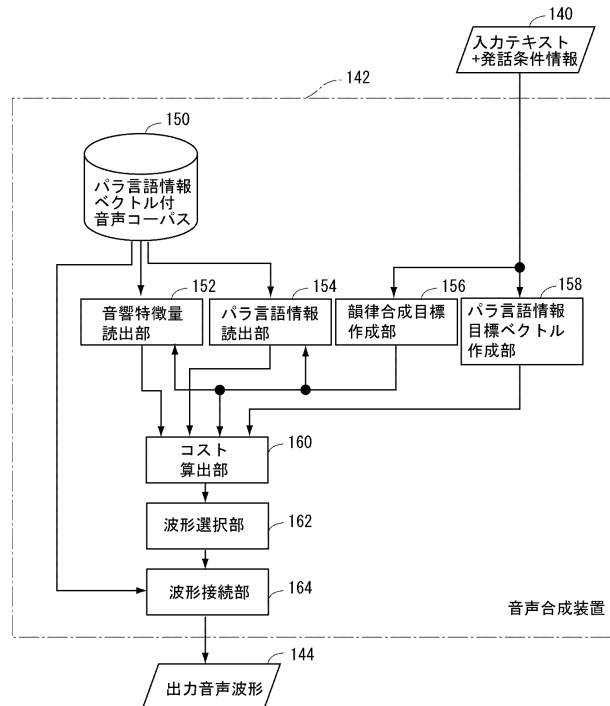
【図 3】



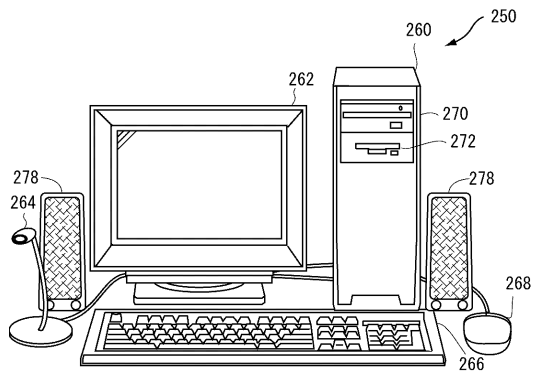
【図 4】



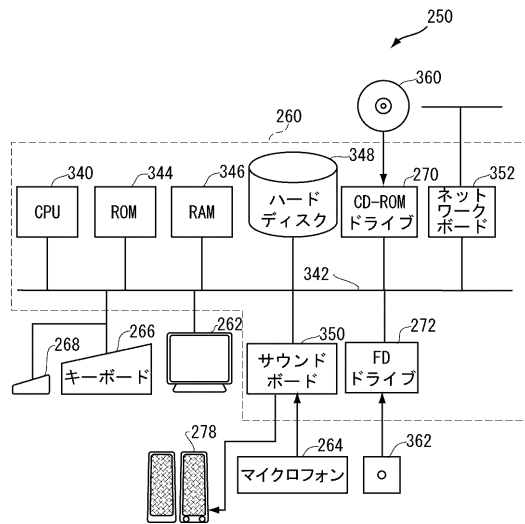
【図 6】



【図 7】



【図 8】



フロントページの続き

(56)参考文献 特表2003-508805(JP,A)

白澤 敏行 Toshiyuki Shirasawa, 音声に込められた感情の判別 Discriminating Emotions intended in Speech, 電子情報通信学会技術研究報告 Vol.96 No.499 IEICE Technical Report, 日本, 社団法人電子情報通信学会 The Institute of Electronics, Information and Communication Engineers, 1997年 1月24日, 第96巻, p.79-84

(58)調査した分野(Int.Cl., DB名)

G10L 15/00 - 15/28

G10L 13/00 - 13/08