

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4677548号  
(P4677548)

(45) 発行日 平成23年4月27日(2011.4.27)

(24) 登録日 平成23年2月10日(2011.2.10)

(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 11/00 (2006.01) G 1 0 L 11/00 4 0 2 H

請求項の数 6 (全 15 頁)

<p>(21) 出願番号 特願2005-269699 (P2005-269699)</p> <p>(22) 出願日 平成17年9月16日 (2005.9.16)</p> <p>(65) 公開番号 特開2007-79363 (P2007-79363A)</p> <p>(43) 公開日 平成19年3月29日 (2007.3.29)</p> <p>審査請求日 平成20年3月27日 (2008.3.27)</p> <p>(出願人による申告) 平成17年度独立行政法人情報通信研究機構、研究テーマ「ネットワーク・ヒューマン・インターフェースに関する総合的な研究開発」に関する委託研究、産業活力再生特別措置法第30条の適用を受ける特許出願</p> <p>特許権者において、実施許諾の用意がある。</p>	<p>(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</p> <p>(74) 代理人 100099933 弁理士 清水 敏</p> <p>(72) 発明者 イシイ・カルロス・トシノリ 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</p> <p>(72) 発明者 石黒 浩 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</p> <p>(72) 発明者 萩田 紀博 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</p> <p style="text-align: right;">最終頁に続く</p>
---	--

(54) 【発明の名称】 パラ言語情報検出装置及びコンピュータプログラム

(57) 【特許請求の範囲】

【請求項1】

人間の発話音声信号から、発話内容に依存しないパラ言語情報を検出するためのパラ言語情報検出装置であって、

前記発話音声信号の韻律に関する情報を処理するための第1の音声処理手段と、

前記発話音声信号の声質に関する情報を処理するための第2の音声処理手段と、

前記韻律に関する情報と前記声質に関する情報とから発話音声に関するパラ言語情報を抽出するためのパラ言語情報抽出手段とを含み、

前記第2の音声処理手段は、前記発話音声信号の発話区間中にボーカル・フライ区間が占める割合を算出するためのボーカル・フライ割合算出手段を含む、パラ言語情報検出装置。

【請求項2】

前記第2の音声処理手段は、さらに、前記発話音声信号の発話区間中に非周期性ノダブル周期性区間が占める割合を算出するための非周期性ノダブル周期性割合算出手段を含む、請求項1に記載のパラ言語情報検出装置。

【請求項3】

人間の発話音声信号から、発話内容に依存しないパラ言語情報を検出するためのパラ言語情報検出装置であって、

前記発話音声信号の韻律に関する情報を処理するための第1の音声処理手段と、

前記発話音声信号の声質に関する情報を処理するための第2の音声処理手段と、

前記韻律に関する情報と前記声質に関する情報とから発話音声に関するパラ言語情報を抽出するためのパラ言語情報抽出手段とを含み、

前記第2の音声処理手段は、前記発話音声信号の発話区間中に非周期性/ダブル周期性区間が占める割合を算出するための非周期性/ダブル周期性割合算出手段を含む、パラ言語情報検出装置。

【請求項4】

前記第2の音声処理手段は、さらに、前記発話音声信号の発話区間中に気息性区間が占める割合を算出するための気息性割合算出手段を含む、請求項1~請求項3のいずれかに記載のパラ言語情報検出装置。

【請求項5】

人間の発話音声信号から、発話内容に依存しないパラ言語情報を検出するためのパラ言語情報検出装置であって、

前記発話音声信号の韻律に関する情報を処理するための第1の音声処理手段と、

前記発話音声信号の声質に関する情報を処理するための第2の音声処理手段と、

前記韻律に関する情報と前記声質に関する情報とから発話音声に関するパラ言語情報を抽出するためのパラ言語情報抽出手段とを含み、

前記第2の音声処理手段は、前記発話音声信号の発話区間中に気息性区間が占める割合を算出するための気息性割合算出手段を含む、パラ言語情報検出装置。

【請求項6】

コンピュータにより実行されると、当該コンピュータを請求項1~請求項5のいずれかに記載のパラ言語情報検出装置として動作させる、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、人間の発話音声から、発話内容に依存しないパラ言語情報を検出するための装置に関し、特に、人間の発話音声に含まれる韻律に関する情報と声質に関する情報とから、パラ言語情報を検出するためのパラ言語情報検出装置に関する。

【背景技術】

【0002】

近年の技術進歩により、人語を発する様々な装置が生産されるようになってきている。この様な装置としては、一例として、カーナビゲーションシステムが挙げられる。カーナビゲーションシステムは、機械が人間に対して一方通行の発話を行なうものであるが、人間との対話が必要とされる装置も存在する。例えば、ロボット等がこれにあたる。

【0003】

ロボットのような装置は、カーナビゲーションシステムよりもさらに人間の生活に密着する可能性が高い。従ってそうした装置で円滑に人間と対話を行なうためには、人間の発話内容だけでなく、感情まで考慮する必要性がある。

【0004】

発話に伴う発話者の感情を推定する場合、発話内容だけでなく、発話内容に依存しない情報である発話意図、態度及び感情等のパラ言語情報をさらに考慮する事が合理的である。つまり、予想されるすべての発話内容に対応する人間の感情を予め学習させるよりは、発話内容と、発話内容に付随するパラ言語情報とを用いて人間の感情を推定する方が合理的でかつ正確であると言える。

【0005】

パラ言語情報の抽出に関する従来の研究は、非特許文献1に開示される様に、韻律特徴を重視していた。

【0006】

図1に、韻律特徴を使用した従来のパラ言語情報抽出装置30の機能ブロック図を示す。図1を参照して、このパラ言語情報抽出装置30は、韻律に基づいて発話音声信号を処

10

20

30

40

50

理し、句末トーン情報と呼ばれるパラメータを出力するための韻律による音声処理部 40 と、予め学習用データを用いて学習した、句末トーン情報とパラ言語情報との関係の確率分布を用いる事により、韻律による音声処理部 40 から得られた句末トーン情報からパラ言語情報を抽出して出力するためのパラ言語情報抽出部 42 とを含む。この従来技術では、句末トーン情報として F 0 m o v e と呼ばれる音程の変化を表すパラメータを用いている。

#### 【0007】

ユーザが発話をする、その発話音声は図示しないマイクによって、発話音声信号に変換される。この発話音声信号は、音声処理部 40 に与えられる。音声処理部 40 での処理によって句末トーン情報が得られる。韻律による音声処理部 40 での処理によって得られた句末トーン情報を使用して、パラ言語情報抽出部 42 でパラ言語情報が抽出される。

【非特許文献 1】「自然発話における、知覚に関連した句末の音響的韻律特徴」、カルロス・トシノリ・イシイ、パーハム・モクタリ、ニック・キャンベル、ユーロスピーチ：p p . 4 0 5 - 4 0 8 , 2 0 0 3 ( " Perceptually-related Acoustic-Prosodic Features of Phrase Finals in Spontaneous Speech ", Carlos Toshinori Ishi, Parham Mokhtari, Nick Campbell, Eurospeech 2003: 405-408, 2003)

#### 【発明の開示】

#### 【発明が解決しようとする課題】

#### 【0008】

この様な韻律特徴のみを使用したパラ言語情報検出装置 30 においては、互いに異なった感情を表わしている発話から抽出したパラ言語情報が、互いに重なってしまう場合がある。

#### 【0009】

図 2 を参照して、この重なりについて説明する。ここでは韻律特徴として、音程の変化 F 0 m o v e と発話持続時間とを使用している。

#### 【0010】

グラフの縦軸は発話持続時間を示し、横軸は音程の変化を表わす。凡例 50 に示す様に、グラフ中にプロットされた記号はそれぞれ、発話者の感情を表わしている。このグラフに見られる様に、韻律情報のみを使用すると、異なったパラ言語情報が同じ韻律情報と発話持続時間とで表わされている。つまり、ある発話持続時間とある音程の変化とをもつパラ言語情報が「聞返し」であるのか「驚き・意外」であるのかがはっきりしないという結果になる。それゆえ、パラ言語情報検出の精度が下がる。

#### 【0011】

さらに、表現豊かな発話音声では、息漏れを含む音声である氣息性の音声の様に、音程を抽出する事が難しいものも含まれている。

#### 【0012】

そこで、本発明の目的は、これらの問題を解決し、パラ言語情報検出の際に、韻律特徴だけを用いる場合より明確にパラ言語情報を区別できる、精度の高いパラ言語情報を提供する事である。

#### 【0013】

本発明の他の目的は、パラ言語情報を、韻律情報だけでなく声質情報も用いて抽出する事により、精度の高いパラ言語情報検出装置を提供する事である。

#### 【課題を解決するための手段】

#### 【0014】

本発明の第 1 の局面に係るパラ言語情報検出装置は、人間の発話音声信号から、発話内容に依存しないパラ言語情報を検出するためのパラ言語情報検出装置であって、発話音声信号の韻律に関する情報を処理するための第 1 の音声処理手段と、発話音声信号の声質に関する情報を処理するための第 2 の音声処理手段と、韻律に関する情報と声質に関する情報とから発話音声に関するパラ言語情報を抽出するためのパラ言語情報抽出手段とを含む。

## 【 0 0 1 5 】

好ましくは、第2の音声処理手段は、発話音声信号の発話区間中にボーカル・フライ区間が占める割合を算出するためのボーカル・フライ割合算出手段を含む。

## 【 0 0 1 6 】

より好ましくは、第2の音声処理手段は、さらに、発話音声信号の発話区間中に非周期性/ダブル周期性区間が占める割合を算出するための非周期性/ダブル周期性割合算出手段を含む。

## 【 0 0 1 7 】

さらに好ましくは、第2の音声処理手段は、発話音声信号の発話区間中に非周期性/ダブル周期性区間が占める割合を算出するための非周期性/ダブル周期性割合算出手段を含む。

10

## 【 0 0 1 8 】

より好ましくは、第2の音声処理手段は、さらに、発話音声信号の発話区間中に気息性区間が占める割合を算出するための気息性割合算出手段を含む。

## 【 0 0 1 9 】

さらに好ましくは、第2の音声処理手段は、発話音声信号の発話区間中に気息性区間が占める割合を算出するための気息性割合算出手段を含む。

## 【 0 0 2 0 】

本発明の第2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、当該コンピュータを上記したいずれかのパラ言語情報検出装置として動作させる。

20

## 【 0 0 2 1 】

このパラ言語情報検出装置によると、情報検出の際に韻律に関する情報のみならず、声質に関する情報も使用できる。それゆえ、パラ言語情報検出の精度を上げる事ができる。従って、より精度の高いパラ言語情報検出装置を提供する事ができる。

## 【 発明を実施するための最良の形態 】

## 【 0 0 2 2 】

以下、図面を参照し、本発明の一実施の形態を説明する。本実施の形態は、発話音声信号から韻律による音声処理と声質による音声処理とを行ない、パラ言語情報を抽出するためのパラ言語情報検出装置に関するものである。

## 【 0 0 2 3 】

< 構成 >

30

図3に、本実施の形態に係るパラ言語情報検出装置60の機能ブロック図を示す。図3を参照して、このパラ言語情報検出装置60は、韻律に基づいて発話音声信号を処理してパラ言語情報の抽出に使用するパラメータを出力するための韻律による音声処理部70と、声質に基づいて発話音声信号を処理してパラ言語情報の抽出に使用するパラメータを出力するための声質による音声処理部72と、韻律による音声処理部70と声質による音声処理部72とから得られたパラメータから、予め学習用データを用いて学習した、パラメータとパラ言語情報との関係を示す確率分布に従ってパラ言語情報を抽出して出力するためのパラ言語情報抽出部74とを含む。

## 【 0 0 2 4 】

図4に、韻律による音声処理部70の詳細を機能ブロック図で示す。図4を参照して、韻律による音声処理部70は、発話音声信号をピッチの動き、つまり音程の変化を表わすパラメータであるF0 moveに変換する処理を行なうための韻律特徴処理部80と、発話持続時間に関する情報を抽出するための発話持続時間抽出部84と、韻律特徴処理部80で得られたF0 moveと発話持続時間抽出部84で得られた発話時間情報とからトーンパラメータを抽出するためのトーンパラメータ抽出部82とを含む。

40

## 【 0 0 2 5 】

図5を参照してトーンパラメータについて説明する。ここでは、日本語「ないね」を例にとる。トーンパラメータとは、言葉の中に含まれる音程の上下をパラメータ化したものである。例えば、トーンパラメータ1a(100)においては「ないね」という言葉の「

50

な」と「いね」との間で音程の変化が起こる。そしてその変化は、「な」から「いね」に移る際に、音程が下がるというものである。

【 0 0 2 6 】

図 5 に示された記号 は音程が下降する事、記号 は下降した音程が元の音程に戻る事、右上がりの矢印記号は音程が上昇する事を示す。

【 0 0 2 7 】

図 5 に示されたトーンパラメータは 7 種類であるが、本実施の形態では、1 a ( 1 0 0 )、2 a ( 1 0 2 )、2 b ( 1 0 4 )、2 c ( 1 0 6 )、3 ( 1 0 8 ) の 5 種類を使用する。

【 0 0 2 8 】

図 6 に、韻律特徴処理部 8 0 の詳細を機能ブロック図で示す。図 6 を参照して、韻律特徴処理部 8 0 は、発話音声信号から音程に関する情報であるパラメータ F 0 を得るための F 0 抽出部 9 0 と、パラメータ F 0 を用いてある音節内のピッチの動き ( 方向と度合い ) つまり音程の変化を半音単位で表わすパラメータである F 0 m o v e を抽出するための F 0 m o v e 抽出部 9 2 とを含む。F 0 抽出部 9 0 は、発話音声信号から音の高さに関する情報である F 0 のみを抽出し、音階で表わす様に変換する。

【 0 0 2 9 】

図 7 に、声質による音声処理部 7 2 の詳細を機能ブロック図で示す。図 7 を参照して、声質による音声処理部 7 2 は、発話音声信号からボーカル・フライを検出するためのボーカル・フライ検出部 1 2 0 と、全発話区間内に占めるボーカル・フライ区間の割合を算出するためのボーカル・フライ割合算出部 1 2 2 とを含む。ここで、ボーカル・フライとは、声道の励振がほとんど減衰した事により生じる 7 H z ~ 7 8 H z くらいの非常に低い周波数のパルス音声の事である。

【 0 0 3 0 】

声質による音声処理部 7 2 はさらに、与えられた発話音声信号のうちで、ボーカル・フライ区間以外でかつ音声波形が非周期である区間及びダブル周期性である区間の情報である非周期性区間情報及びダブル周期性区間情報を検出するための非周期性 / ダブル周期性検出部 1 2 4 と、非周期性 / ダブル周期性検出部 1 2 4 で検出された非周期性及びダブル周期性区間情報からボーカル・フライ検出部 1 2 0 で検出されたボーカル・フライ区間情報 1 3 2 を除き、これらの非周期性区間情報及びダブル周期性区間情報が全発話区間で占める割合を算出するための非周期性 / ダブル周期性割合算出部 1 2 6 とを含む。ここで、非周期性とは、音声波形が非周期的である事である。また、ダブル周期性とは、音声波形が、ピーク長及びピーク幅の異なる二つの波形からなる波形のセットが周期的に繰返された形状をもつ事をいう。

【 0 0 3 1 】

声質による音声処理部 7 2 はさらに、与えられた発話音声信号から、気息性区間情報を検出するための気息性検出部 1 2 8 と、気息性区間が全発話区間で占める割合を算出するための気息性割合算出部 1 3 0 とを含む。ここで、気息性とは、音声に含まれる息漏れの度合いの事である。気息性のある声としては例えば、ささやき声等が挙げられる。

【 0 0 3 2 】

図 8 に、ボーカル・フライ検出部 1 2 0 の詳細を機能ブロック図で示す。図 8 を参照して、ボーカル・フライ検出部 1 2 0 は、発話信号のうち 1 0 0 H z ~ 1 5 0 0 H z の周波数成分のみを通過させるためのバンドパスフィルタ 1 4 0 と、バンドパスフィルタ 1 4 0 を通過した発話信号 1 5 4 を超短期フレーム長でフレーム化し、各フレームについて、その前後 2 フレームと比較してパワーが大きく、かつその差が予め定められたパワーしきい値よりも大きいフレームのフレーム位置を示す情報 1 5 0 をパワーピーク候補の位置情報として出力するための超短期ピーク検出処理部 1 4 2 と、発話信号 1 5 4 を短期フレーム長でフレーム化したものについてフレーム内周期性 ( Intra-frame periodicity : I F P 値 ) に関する値を算出し、フレーム内周期性が所定個数以上存在するフレーム以外のフレームの I F P 値をヌルに設定するための短期周期性検出部 1 4 4 と、超短期ピーク検出

10

20

30

40

50

処理部 142 から与えられたピーク位置情報 150 のうち、短期周期性検出部 144 から与えられた短期周期性情報 152 により、フレーム値がヌルとなっている部分の情報 156 のみを類似性検査部 148 に与えるための周期性検査部 146 と、情報 156 によって特定されるパワーピーク候補の付近の波形とその前のパワーピーク付近の波形との間のパルス間類似性 (inter-pulse similarity : IPS 値) に関する値が所定のしきい値以上であるもののピーク位置情報を検出し、このピーク位置情報に基づき、隣接するパルス間で IPS 値の高いもの間のフレームからボカール・フライ区間情報を検出し、ボカール・フライ割合算出部 122 と非周期性 / ダブル周期性割合算出部 126 とに与えるための類似性検査部 148 とを含む。

#### 【 0033 】

図 9 に、非周期性 / ダブル周期性検出部 124 の詳細を機能ブロック図で示す。図 9 を参照して、非周期性 / ダブル周期性検出部 124 は、発話音声信号をフィルタリング処理して音声波形のピークを検出する事によって、正規化自己相関関数を算出するための正規化自己相関関数算出部 160 と、正規化自己相関関数算出部 160 で算出された正規化自己相関関数に基づいた正規化自己相関関数の波形から、ピーク値やピーク位置の関係等で表わされる正規化自己相関関数パラメータを算出するための正規化自己相関関数パラメータ算出部 162 と、算出された正規化自己相関関数パラメータの値から、非周期性及びダブル周期性区間情報を検出するための非周期性 / ダブル周期性区間情報検出部 164 とを含む。

#### 【 0034 】

正規化自己相関関数パラメータ算出部 162 では、正規化自己相関関数算出部 160 で得られた正規化自己相関関数より最初の 2 ピーク (P1 及び P2) を検出する。ただし、ピーク値は 0.2 を超えるもののみピークとみなす。

#### 【 0035 】

これらのピークの正規化自己相関値を NAC (P1)、NAC (P2) 及び、正規化自己相関位置を TL (P1)、TL (P2) と呼び、正規化自己相関関数パラメータとして扱う。

#### 【 0036 】

図 10 に、正規化自己相関関数算出部 160 の詳細を機能ブロック図で示す。図 10 を参照して、正規化自己相関関数算出部 160 は、発話信号のうち 60 Hz 以上の周波数成分のみを通すためのハイパスフィルタ 170 と、ハイパスフィルタ 170 の出力する音声信号の高域部分を強調する処理を行なうための高域強調部 172 と、高域強調部 172 の出力する音声信号に線型予測分析を行ない、声道パラメータ抽出部 174 で声道パラメータを抽出し、逆フィルタ 176 で、ハイパスフィルタ 170 の出力する音声信号に声道パラメータ抽出部 174 で抽出された声道パラメータを使用して、逆フィルタを行ない、声帯音源波形に対応する残差信号が得られると、後の処理に必要なピーク検出を容易にするために 2 kHz 以下の音声信号のみを通すためのローパスフィルタ 178 と、ローパスフィルタ 178 を通った音声信号が与えられるとウィンドウの大きさを 80 ms にし、そのウィンドウに含まれる音声信号から自己相関関数を算出するための自己相関関数算出部 180 と、自己相関関数算出部 180 で算出された自己相関関数の波形から、各々のフレームに含まれた最大のピークを検出するためのピーク検出部 182 と、ピーク検出部 182 で検出された最大ピークとその直前もしくは直後の最大ピークとの間の時間のずれを抽出し、ずれた時間の 4 倍の時間を 1 フレームとする様にフレーム長を再調節し、再調節されたフレームに含まれる自己相関関数の算出を行なうための自己相関関数再算出部 184 と、得られた自己相関関数を正規化する処理を行なうための正規化部 186 とを含む。

#### 【 0037 】

図 11 に、気息性検出部 128 の詳細を機能ブロック図で示す。図 11 を参照して、気息性検出部 128 は、発話音声信号のうちで、100 Hz ~ 1500 Hz の周波数成分のみを通過させるための F1 パスフィルタ 202 と、この F1 パスフィルタ 202 を通過した波形全体から、振幅の変化を抽出するための振幅包絡抽出部 204 と、発話信号のうち

10

20

30

40

50

、1800Hz～4000Hzの周波数成分のみを通過させるためのF3パスフィルタ200と、F3パスフィルタ200を通過した波形全体から、振幅の変化を抽出するための振幅包絡抽出部210と、振幅包絡抽出部204から得られた振幅の変化と振幅包絡抽出部210から得られた振幅の変化との間の相互相関を計算するための相互相関計算部214とを含む。ここで、F1パスフィルタ202を通過した周波数をF1波と呼び、F3パスフィルタ200を通過した周波数をF3波と呼ぶ。また、振幅包絡抽出部204で抽出された振幅の変化をF1振幅包絡と呼び、振幅包絡抽出部210で抽出された振幅の変化をF3振幅包絡と呼ぶ。

#### 【0038】

気息性検出部128はさらに、F1パスフィルタ202を通過した成分からなるF1波から、最大周波数成分を抽出するための第1の最大周波数成分抽出部206と、F3パスフィルタ200を通過した成分からなるF3波から、最大周波数成分を抽出するための第2の最大周波数成分抽出部212と、F1波中に含まれる最大周波数成分とF3波中に含まれる最大周波数成分との差であるスペクトル傾斜A1-A3値を算出するためのスペクトル傾斜算出部216とを含む。

10

#### 【0039】

気息性検出部128はさらに、相互相関計算部214から得られたF1F3相関値があるしきい値未満であり、かつ、スペクトル傾斜算出部216から得られたスペクトル傾斜A1-A3値があるしきい値未満であるか否かにより気息性区間か否かを判定し、気息性区間情報を入力するための気息性判定部218を含む。

20

#### 【0040】

##### <動作>

図3を参照して、まず、ユーザが発話をする、その発話音声が入力されないマイクにより発話音声信号に変換される。マイクによって変換された発話音声信号は、韻律による音声処理部70と声質による音声処理部72とに与えられる。この韻律による音声処理部70での処理によって句末トーン情報が得られる。声質による音声処理部72での処理によって発話全体に占めるボーカル・フライの割合、非周期性及びダブル周期性の割合、及び気息性の割合に関する情報が得られる。韻律による音声処理部70及び声質による音声処理部72での処理の詳細については後述する。

30

#### 【0041】

図4を参照して、韻律による音声処理部70の動作の詳細について述べる。発話音声信号を受信すると、韻律特徴処理部80では、まず、その発話音声信号をピッチの動きつまり音程の変化を表わすパラメータであるF0moveに変換する処理が行なわれる。F0moveは、音程に関する情報であるF0から得られる。

#### 【0042】

図6を参照して、韻律特徴処理部80での動作の詳細について述べる。発話音声信号を受信すると、F0抽出部90では発話音声信号から音の高さに関する情報のみを抽出し、音階情報に変換してパラメータF0を得る。

#### 【0043】

パラメータF0を用いてF0move抽出部92で、ある音節内のピッチの動き(方向と度合い)つまり音程の変化を半音単位で表わすパラメータであるF0moveが抽出される。F0moveは、複数のF0の差から求める事が可能である。

40

#### 【0044】

図4を参照して、発話持続時間抽出部84で、発話音声信号から発話持続時間に関する情報が抽出される。

#### 【0045】

韻律特徴処理部80で抽出されたF0moveと発話持続時間抽出部84で抽出された発話持続時間に関する情報とを用いて、トーンパラメータ抽出部82でトーンパラメータが抽出される。抽出されたトーンパラメータは後のパラ言語情報抽出部74での処理に使用される。

50

## 【 0 0 4 6 】

図 7 を参照して、声質による音声処理部 7 2 は以下の様に動作する。まず、発話音声信号から、ボーカル・フライ検出部 1 2 0 でボーカル・フライ区間情報が検出される。

## 【 0 0 4 7 】

図 8 を参照して、ボーカル・フライ検出部 1 2 0 は以下の様に動作する。バンドパスフィルタ 1 4 0 は、発話信号のうち 1 0 0 H z ~ 1 5 0 0 H z の周波数成分のみを通過させる。バンドパスフィルタ 1 4 0 を通過した発話信号 1 5 4 は、超短期ピーク検出処理部 1 4 2、短期周期性検出部 1 4 4 及び類似性検査部 1 4 8 に与えられる。超短期ピーク検出処理部 1 4 2 は、発話信号 1 5 4 を超短期フレーム化し、各フレームに対し超短期パワーを算出する。そして、各フレームについて、その前後 2 フレームと比較してパワーの差が

10

## 【 0 0 4 8 】

短期周期性検出部 1 4 4 は、発話信号 1 5 4 をフレーム化し、その各フレームについて I F P 値を算出する。算出された I F P 値としきい値とを比較し、しきい値未満であれば、そのフレームの I F P 値をヌルに設定する。ヌルではないフレームが少なくとも 3 フレームだけ連続していなければ、それらのフレームの I F P 値をヌルに補正する。そして補正された I F P 値が周期性検査部 1 4 6 に与えられる。

## 【 0 0 4 9 】

周期性検査部 1 4 6 は、超短期ピーク検出処理部 1 4 2 から与えられたピーク位置情報 1 5 0 のうち、短期周期性検出部 1 4 4 から与えられた短期周期性情報 1 5 2 により、フレーム I F P 値がヌルとなっている部分の情報 1 5 6 のみを類似性検査部 1 4 8 に与える。

20

## 【 0 0 5 0 】

類似性検査部 1 4 8 は、情報 1 5 6 によって特定される区間に存在するパワーピーク候補の各パワーピーク付近の波形とその前のパワーピーク付近の波形との間の I P S 値を算出する。そしてその I P S 値としきい値とを比較し、しきい値以上のパワーピークのピーク位置情報を検出する。このピーク位置情報に基づき、隣接するパルス間で I P S 値の高いものの間のフレームをボーカル・フライ区間として検出し、それらを示す情報（ボーカル・フライ区間情報）を出力する。

30

## 【 0 0 5 1 】

図 7 を参照して、検出されたボーカル・フライ区間情報はボーカル・フライ割合算出部 1 2 2 に与えられる。ボーカル・フライ区間情報から、全発話区間でボーカル・フライ区間の占める割合がボーカル・フライ割合算出部 1 2 2 で算出される。この算出はボーカル・フライ区間を全発話区間で割る事によって得られる。算出されたボーカル・フライ区間割合情報は、後の処理のためにパラ言語情報抽出部 7 4 に与えられる。

## 【 0 0 5 2 】

非周期性 / ダブル周期性検出部 1 2 4 により、発話音声信号のうちで、音声波形が非周期である区間及びダブル周期である区間の情報である非周期性区間及びダブル周期性区間が検出され、それらを示す非周期性区間情報及びダブル周期性区間情報が出力される。

40

## 【 0 0 5 3 】

図 9 を参照して、非周期性 / ダブル周期性検出部 1 2 4 は以下の様に動作する。発話音声信号が与えられると、正規化自己相関関数算出部 1 6 0 は、その音声信号をフィルタリング処理した音声波形を解析する事によって自己相関関数を算出する。そしてその自己相関関数を正規化し正規化自己相関関数を算出する。この正規化自己相関関数算出部 1 6 0 での処理の詳細については以下に述べる。

## 【 0 0 5 4 】

図 1 0 を参照して、発話信号が与えられると、ハイパスフィルタ 1 7 0 によって、6 0 H z 以上の周波数成分のみが通過させられる。6 0 H z 以上の音声信号は、高域強調部 1 7 2 と逆フィルタ 1 7 6 とに与えられる。高域強調部 1 7 2 は与えられた音声信号の高域

50

部分を強調する処理を行なう。そして、声道パラメータ抽出部 174 で、声道を特徴付けるフィルタパラメータを推測する。その後、ハイパスフィルタ 170 の出力音声信号に声道パラメータ抽出部 174 で与えられた声道パラメータを用いて、声帯音源信号を求めるために逆フィルタ 176 を行なう。

【0055】

逆フィルタ 176 で処理された残差信号は、次にローパスフィルタ 178 に与えられる。このローパスフィルタ 178 は、後の処理に必要となるピーク検出を容易にするために 2 kHz 以下の周波数成分のみを通過させる。ローパスフィルタ 178 を通過した周波数成分は、自己相関関数算出部 180 と自己相関関数再算出部 184 とに与えられる。自己相関関数算出部 180 では、検出処理の際に使用するフレームの大きさを 80 ms とし、フレーム中の音声信号波形から自己相関関数を得る。そしてこの自己相関関数を出力する。

10

【0056】

ピーク検出部 182 では、自己相関関数算出部 180 で得られた自己相関関数に含まれた最大のピークを検出する処理が行なわれる。

【0057】

自己相関関数再算出部 184 では、まず、ピーク検出部 182 で検出された最大ピークの位置の 4 倍の時間を新しいフレーム長とする。この様なフレームの再調節が行なわれるのは、自己相関関数の適切な算出を行なうためである。つまり、固定のフレーム長の場合、フレームが大きすぎても小さすぎても自己相関関数の適切な算出をする事が難しいからである。そして、そのフレームから再度自己相関関数を得る。

20

【0058】

次に、正規化部 186 で得られた自己相関関数を正規化する処理を行なう。図 9 を参照して、正規化自己相関関数算出部 160 で算出された正規化自己相関関数に基づいて、正規化自己相関関数パラメータ算出部 162 での算出処理が行なわれる。そして、音波の非周期性及びダブル周期性を抽出するために正規化自己相関関数の波形から、ピーク値及びピーク位置を検出する。そしてその後それらピーク値の比率とピーク位置の比率とを算出する。ピーク値の比率は、 $1000 * NAC(P2) / NAC(P1)$  で求められる。また、ピーク位置の比率は  $2000 * TL(P2) / TL(P1)$  で求められる。

【0059】

30

さらに、算出された正規化自己相関関数パラメータを使用して、非周期性/ダブル周期性区間情報検出部 164 で当該音声信号が非周期性もしくはダブル周期性を持つ区間が検出される。この検出処理の詳細は以下の通りである。

【0060】

つまり、上述した自己相関関数パラメータがいずれも 1000 に近似した値であれば、その自己相関関数の波形で表わされる区間の発話音声波形は周期性を持つと言える。そこで、それ以外の値を取る発話区間を非周期性及びダブル周期性区間として抽出することができる。

【0061】

非周期性/ダブル周期性区間情報検出部 164 で検出された非周期性/ダブル周期性区間情報が非周期性/ダブル周期性割合算出部 126 に与えられる。

40

【0062】

図 7 を参照して、全発話区間中で非周期性区間及びダブル周期性区間の占める割合が非周期性/ダブル周期性割合算出部 126 で算出される。この算出は、非周期性区間及びダブル周期性区間を全発話区間で割る事によって行なわれる。

【0063】

この算出処理の前にまず、ポーカル・フライ検出部 120 で、ポーカル・フライ区間として検出された区間情報を非周期性/ダブル周期性区間情報から除去する処理が行なわれる。ポーカル・フライも非周期性特徴を持つが、ここでは、ポーカル・フライ以外の非周期性/ダブル周期性を対象としているからである。

50

【 0 0 6 4 】

図 1 1 を参照して、氣息性検出部 1 2 8 は以下の様に動作する。発話音声信号が与えられ、F 1 パスフィルタ 2 0 2 は、まず、その発話音声信号のうち、1 0 0 H z ~ 1 5 0 0 H z の周波数成分のみを通過させる。振幅包絡抽出部 2 0 4 では、F 1 パスフィルタ 2 0 2 を通った F 1 波の波形から、振幅包絡を抽出する。

【 0 0 6 5 】

F 3 パスフィルタ 2 0 0 でも同様に、発話音声信号のうち、1 8 0 0 H z ~ 4 0 0 0 H z の周波数成分のみを通過させる。そして振幅包絡抽出部 2 1 0 では、F 3 パスフィルタ 2 0 0 を通った F 3 波の波形から、振幅包絡を抽出する。

【 0 0 6 6 】

振幅包絡抽出部 2 0 4 から得られた F 1 振幅包絡と振幅包絡抽出部 2 1 0 から得られた F 3 振幅包絡との相互相関を相互相関計算部 2 1 4 で計算する。この処理により、F 1 振幅包絡と F 3 振幅包絡の相互の関係を示す F 1 F 3 相関値が得られる。

10

【 0 0 6 7 】

F 1 パスフィルタ 2 0 2 を通過した F 1 波からはまた、最大周波数成分抽出部 2 0 6 でこの周波数帯域中に含まれるもののうち最大の周波数成分が抽出される。そして、F 3 パスフィルタ 2 0 0 を通過した F 3 波にも、最大周波数成分抽出部 2 1 2 で同様の処理が行なわれる。F 1 波中に含まれる最大周波数成分と F 3 波中に含まれる最大周波数成分との差、つまりスペクトル傾斜を算出する処理がスペクトル傾斜算出部 2 1 6 で行なわれる。このスペクトル傾斜を A 1 - A 3 とする。

20

【 0 0 6 8 】

氣息性判定部 2 1 8 では、F 1 F 3 相関値とスペクトル傾斜 A 1 - A 3 値とを用いて氣息性であるか否かを判定して、氣息性区間情報を出力する。ここでの処理では、F 1 F 3 相関値があるしきい値未満で、かつ A 1 - A 3 値があるしきい値未満であれば、氣息性区間であると判定する。これらのしきい値は予め学習によって得られる。このしきい値と実際に得られた F 1 F 3 相関値と A 1 - A 3 とを比較参照する事により、氣息性の有無が判定できる。

【 0 0 6 9 】

氣息性区間情報は、氣息性割合算出部 1 3 0 に与えられる。図 7 を参照して、氣息性割合算出部 1 3 0 は、全発話区間中で氣息性区間の占める割合を、氣息性区間を全発話区間で割る事によって算出する。算出された氣息性区間割合は、後の処理のためにパラ言語情報抽出部 7 4 に与えられる。

30

【 0 0 7 0 】

図 3 を参照して、韻律による音声処理部 7 0 での処理によって得られた句末トーン情報、声質による音声処理部 7 2 での処理によって得られた発話全体に占めるボーカル・フライの割合に関する情報、発話全体に占める非周期性もしくはダブル周期性の割合に関する情報及び、発話全体に占める氣息性の割合に関する情報を使用して、パラ言語情報抽出部 7 4 でパラ言語情報が抽出される。

【 0 0 7 1 】

ここでの処理においては、予め句末トーン情報、発話全体に占めるボーカル・フライ区間の割合、非周期性及びダブル周期性区間の割合及び、氣息性の割合に関する情報とパラ言語情報との関係に関するデータを集積する必要がある。この集積されたデータによってさらに、どの様なパラメータが入力されれば、どの様なパラ言語情報が検出できるかというモデルを学習によって作成する事ができる。

40

【 0 0 7 2 】

このモデルに使用されるものとしては、決定木、ニューラルネットワーク及び、SVM (Support Vector Machine) 等が考えられる。

【 0 0 7 3 】

[ コンピュータによる実現 ]

この実施の形態のシステムは、コンピュータハードウェアと、そのコンピュータハード

50

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 1 2 はこのコンピュータシステム 3 3 0 の外観を示し、図 1 3 はコンピュータシステム 3 3 0 の内部構成を示す。

【 0 0 7 4 】

図 1 2 を参照して、このコンピュータシステム 3 3 0 は、F D (フレキシブルディスク) ドライブ 3 5 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 3 5 0 を有するコンピュータ 3 4 0 と、キーボード 3 4 6 と、マウス 3 4 8 と、モニタ 3 4 2 とを含む。

【 0 0 7 5 】

図 1 3 を参照して、コンピュータ 3 4 0 は、F D ドライブ 3 5 2 および C D - R O M ドライブ 3 5 0 に加えて、C P U (中央処理装置) 3 5 6 と、C P U 3 5 6、F D ドライブ 3 5 2 および C D - R O M ドライブ 3 5 0 に接続されたバス 3 6 6 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 5 8 と、バス 3 6 6 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 6 0 とを含む。コンピュータシステム 3 3 0 はさらに、プリンタ 3 4 4 を含んでいる。

10

【 0 0 7 6 】

ここでは示さないが、コンピュータ 3 4 0 はさらにローカルエリアネットワーク (L A N) への接続を提供するネットワークアダプタボードを含んでもよい。

【 0 0 7 7 】

20

コンピュータシステム 3 3 0 にパラ言語情報抽出装置 6 0 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 3 5 0 または F D ドライブ 3 5 2 に挿入される C D - R O M 3 6 2 または F D 3 6 4 に記憶され、さらにハードディスク 3 5 4 に転送される。または、プログラムは図示しないネットワークを通じてコンピュータ 3 4 0 に送信されハードディスク 3 5 4 に記憶されてもよい。プログラムは実行の際に R A M 3 6 0 にロードされる。C D - R O M 3 6 2 から、F D 3 6 4 から、またはネットワークを介して、直接に R A M 3 6 0 にプログラムをロードしてもよい。

【 0 0 7 8 】

このプログラムは、コンピュータ 3 4 0 にこの実施の形態のパラ言語情報抽出装置 6 0 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 3 4 0 上で動作するオペレーティングシステム (O S) もしくはサードパーティのプログラム、またはコンピュータ 3 4 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。従って、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、上記したパラ言語情報抽出装置 6 0 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 3 3 0 の動作は周知であるので、ここでは繰返さない。

30

【 0 0 7 9 】

以上の様に、パラ言語情報を検出する際に、韻律に関する情報のみならず、声質に関する情報も使用する事により、パラ言語情報の検出精度が高くなる。

40

【 0 0 8 0 】

今回開示された実施の形態に使用された具体的な数字は例示である。

【 0 0 8 1 】

また、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及び範囲内でのすべての変更を含む。

【 図面の簡単な説明 】

【 0 0 8 2 】

50

【図1】韻律特徴を使用したパラ言語情報抽出装置30の機能ブロック図である。

【図2】韻律特徴を使用してパラ言語情報を検出した場合のパラ言語情報の重なりを表わすグラフである。

【図3】本実施の形態に係るパラ言語情報抽出装置60についての機能ブロック図である。

【図4】韻律による音声処理部70の処理の詳細を示す機能ブロック図である。

【図5】トーンパラメータについて説明する図である。

【図6】韻律特徴処理部80の詳細を示す機能ブロック図である。

【図7】声質による音声処理部72の詳細を示す機能ブロック図である。

【図8】ボーカル・フライ検出部120の詳細を示す機能ブロック図である。

10

【図9】非周期性/ダブル周期性検出部124の詳細を示す機能ブロック図である。

【図10】正規化自己相関関数算出部160の詳細を示す機能ブロック図である。

【図11】氣息性検出部128の詳細を示す機能ブロック図である。

【図12】本発明の一実施の形態に係るパラ言語情報抽出装置30を実現するコンピュータシステムの外觀図である。

【図13】図12に示すコンピュータのブロック図である。

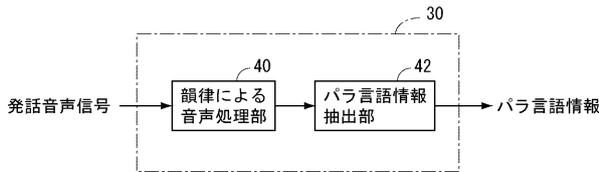
【符号の説明】

【0083】

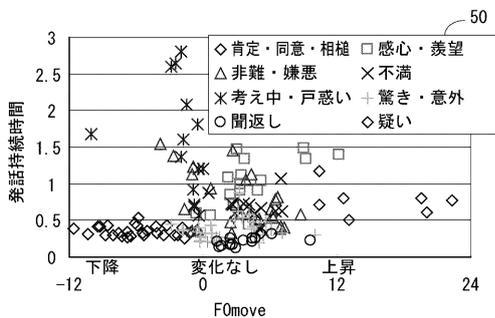
- 70 韻律による音声処理部
- 72 声質による音声処理部
- 74 パラ言語情報抽出部
- 122 ボーカル・フライ割合算出部
- 126 非周期性/ダブル周期性割合算出部
- 130 氣息性割合算出部

20

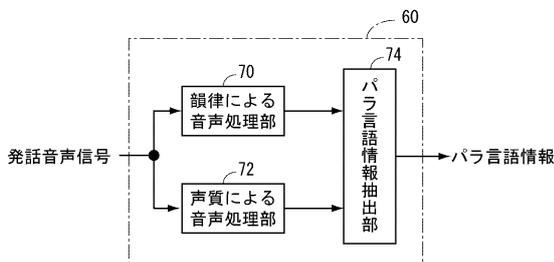
【図1】



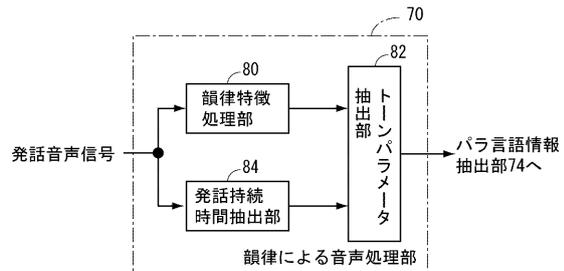
【図2】



【図3】



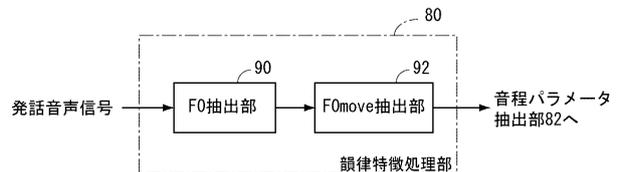
【図4】



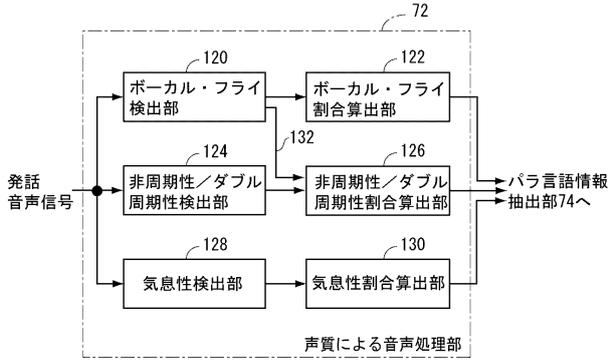
【図5】

音程パラメータ	聴覚的な特徴	例
1a	Low	naʌi ne
1b	Low + Falling tone	naʌi neʌe
2a	High	naʌi ʌne
2b	High + Lengthened	naʌi ʌnee
2c	Low + Rising tone	naʌi ne↗
3	High + Falling tone	naʌi ʌneʌe
5	High + Fall-Rise tone	naʌi ʌneʌe↗

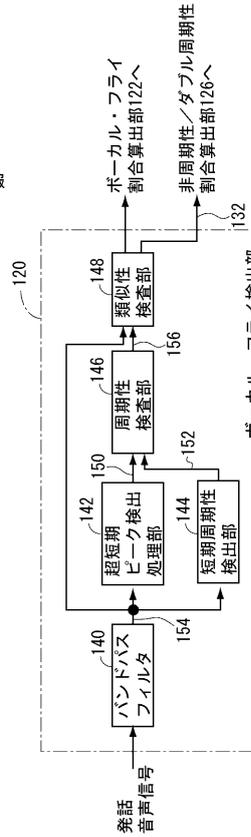
【図6】



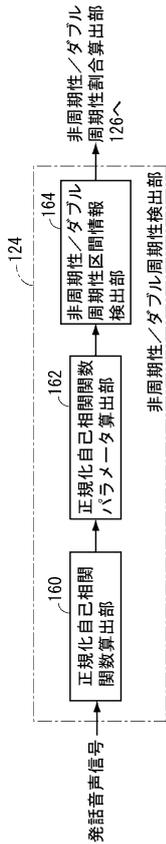
【図7】



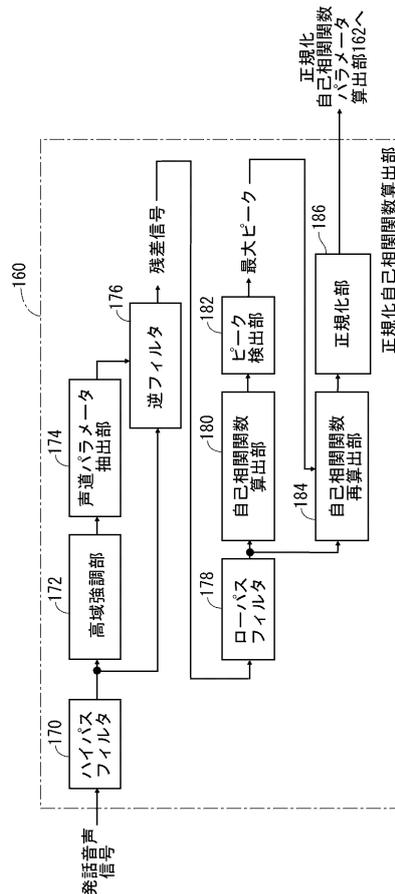
【図8】



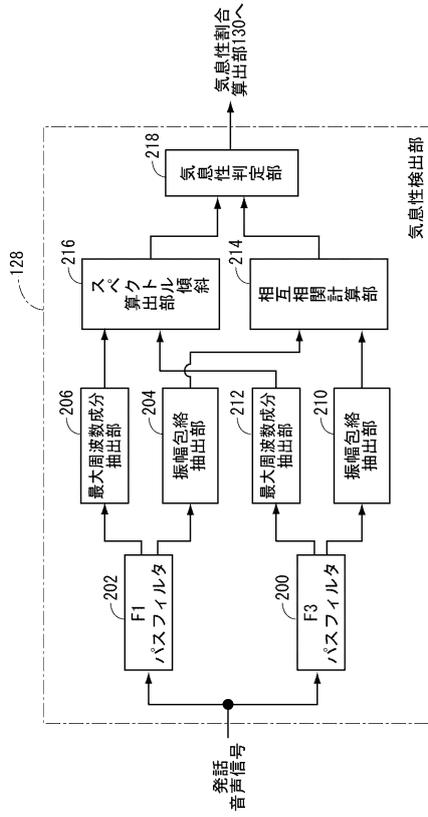
【図9】



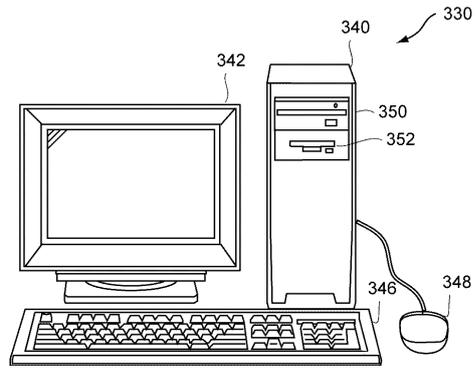
【図10】



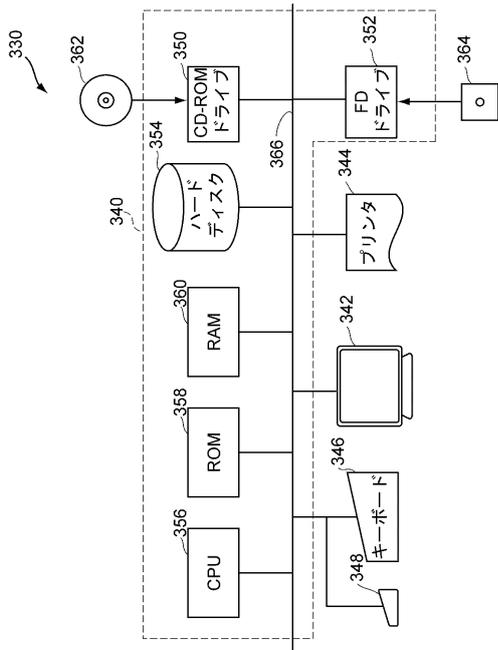
【図11】



【図12】



【図13】



---

フロントページの続き

審査官 山下 剛史

- (56)参考文献 特表2003-508805(JP,A)  
特開2003-330478(JP,A)  
前川喜久雄他, "パラ言語情報の生成と知覚:多次元尺度法による布置と音響特徴の関係", 電子情報通信学会技術研究報告, Vol.99, No.74, SP99-10(1999-05), pp.9-16  
浜野紘一他, "音声の分節的特徴に着眼したパラ・非言語情報推定に関する実験的検討", 電子情報通信学会技術研究報告, Vol.103, No.750, SP2003-197(2004-03), pp.25-30

- (58)調査した分野(Int.Cl., DB名)  
G10L 11/00, 15/10