

(19) 日本国特許庁(JP)

## (12) 特許公報(B2)

(11) 特許番号

特許第4635384号  
(P4635384)

(45) 発行日 平成23年2月23日(2011.2.23)

(24) 登録日 平成22年12月3日(2010.12.3)

(51) Int.Cl.

F 1

G 1 O L 13/02 (2006.01)  
G 1 O L 13/08 (2006.01)G 1 O L 13/02 121 B  
G 1 O L 13/08 11 O Z  
G 1 O L 13/08 127 D

請求項の数 9 (全 9 頁)

(21) 出願番号	特願2001-202693 (P2001-202693)
(22) 出願日	平成13年7月3日 (2001.7.3)
(65) 公開番号	特開2003-15680 (P2003-15680A)
(43) 公開日	平成15年1月17日 (2003.1.17)
審査請求日	平成20年5月12日 (2008.5.12)

(73) 特許権者	000004237 日本電気株式会社 東京都港区芝五丁目7番1号
(74) 代理人	100109313 弁理士 机 昌彦
(74) 代理人	100121290 弁理士 木村 明隆
(74) 代理人	100160554 弁理士 浅井 俊雄
(72) 発明者	若尾 淳 東京都港区芝五丁目7番1号 日本電気株式会社内

審査官 間宮 嘉善

最終頁に続く

(54) 【発明の名称】音声合成システム、音声合成方法および音声合成用プログラム

## (57) 【特許請求の範囲】

## 【請求項 1】

文全体の係り受け関係を取得する係り受け解析手段と、取得した前記文全体の係り受けの信頼度を求める信頼度抽出手段と、求めた前記信頼度を閾値と比較して前記文全体の係り受けを破棄するか決定する信頼度判定手段と、前記文全体の係り受けを前記信頼度判定手段が破棄した場合に、文の局所の係り受け関係を取得する局所係り受け解析手段と、ポーズを設定するポーズ挿入手段とを備え、前記ポーズ挿入手段は、前記文全体の係り受けが破棄された場合には前記文の局所の係り受けを使用し、前記文全体の係り受けが破棄されなかった場合には前記文全体の係り受けを使用することを特徴とする音声合成システム。

## 【請求項 2】

テキストを形態素解析する形態素解析手段と、形態素解析された各単語を文節単位にまとめ上げる文節生成手段と、文節列のどこにポーズを設定するか決めるポーズ設定手段と、アクセントの立ち上がりを含む韻律情報を生成する韻律設定手段と、前記韻律設定手段が生成した発音記号から音声波形を生成する音声波形生成手段とを有し、前記ポーズ設定手段は、文全体の係り受け関係を取得する係り受け解析手段と、取得した前記文全体の係り受けの信頼度を求める信頼度抽出手段と、求めた前記信頼度を閾値と比較して前記文全体の係り受けを破棄するか決定する信頼度判定手段と、前記文全体の係り受けを前記信頼度判定手段が破棄した場合に、文の局所の係り受け関係を取得する局所係り受け解析手段と、ポーズを設定するポーズ挿入手段とを備え、前記ポーズ挿入手段は、前記文全体の係り受けが破棄された場合には前記文の局所の係り受けを使用し、前記文全体の係り受けが破

10

20

棄されなかった場合には前記文全体の係り受けを使用することを特徴とする音声合成システム。

【請求項 3】

前記局所係り受け解析手段は隣接した 2 文節が係ることができるかを判定して文の局所の係り受け関係を取得することを特徴とする請求項 1 または 2 記載の音声合成システム。

【請求項 4】

文全体の係り受け関係を取得し、取得した前記文全体の係り受けの信頼度を求め、求められた前記信頼度を閾値と比較して前記文全体の係り受けを破棄するか決定し、前記文全体の係り受けを破棄した場合には文の局所の係り受け関係を取得してポーズを設定し、前記文全体の係り受けを破棄しなかった場合には前記文全体の係り受けを元にポーズを設定することを特徴とする音声合成方法。 10

【請求項 5】

テキストを形態素解析し、形態素解析された各単語を文節単位にまとめ上げ、文全体の係り受け関係を取得し、取得した前記文全体の係り受けの信頼度を求め、求められた前記信頼度を閾値と比較して前記文全体の係り受けを破棄するか決定し、前記文全体の係り受けを破棄した場合には文の局所の係り受け関係を取得してポーズを設定し、前記文全体の係り受けを破棄しなかった場合には前記文全体の係り受けを元にポーズを設定し、アクセントの立ち上がりを含む韻律情報を生成し、生成した発音記号から音声波形を生成することを特徴とする音声合成方法。

【請求項 6】 20

隣接した 2 文節が係ることができると判定して文の局所の係り受け関係を取得することを特徴とする請求項 4 または 5 記載の音声合成方法。

【請求項 7】

文全体の係り受け関係を取得する処理と、取得した前記文全体の係り受けの信頼度を求める処理と、求められた前記信頼度を閾値と比較して前記文全体の係り受けを破棄するか決定する処理と、前記文全体の係り受けを破棄した場合には文の局所の係り受け関係を取得してポーズを設定し、前記文全体の係り受けを破棄しなかった場合には前記文全体の係り受けを元にポーズを設定する処理とをコンピュータに実行させるためのプログラム。

【請求項 8】

テキストを形態素解析する処理と、形態素解析された各単語を文節単位にまとめ上げる処理と、文全体の係り受け関係を取得する処理と、取得した前記文全体の係り受けの信頼度を求める処理と、求められた前記信頼度を閾値と比較して前記文全体の係り受けを破棄するか決定する処理と、前記文全体の係り受けを破棄した場合には文の局所の係り受け関係を取得してポーズを設定し、前記文全体の係り受けを破棄しなかった場合には前記文全体の係り受けを元にポーズを設定する処理と、アクセントの立ち上がりを含む韻律情報を生成する処理と、生成した発音記号から音声波形を生成する処理とをコンピュータに実行させるためのプログラム。 30

【請求項 9】

隣接した 2 文節が係ことができると判定して文の局所の係り受け関係を取得する処理をコンピュータに実行させるための請求項 7 または 8 記載のプログラム。 40

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は音声合成システム、音声合成方法および音声合成用プログラムに関し、特に読み上げの最中に高精度で安定したポーズ挿入を行う音声合成システム、音声合成方法および音声合成用プログラムに関する。

【0002】

【従来の技術】

音声合成システムでは、漢字仮名まじりの日本語テキストに対し、辞書等を参照した形態素解析により単語（形態素）に分割し、各単語の読み、品詞等からテキストの読み、アク 50

セント、ポーズなどの読み韻律情報を生成し、この結果を元に音声波形を生成して合成音声を出力する。

【0003】

このうち、ポーズに関しては、特開平6-59695号公報、「鈴木、斎藤：日本語テキスト音声合成のためのN文節構造解析とそれに基づく韻律制御」（電子情報通信学会論文誌 Vol.J78-D2, No.2）等に記載されるように、局所的な文節間の係り受け関係を求める上で、ポーズを設定している。

【0004】

これらの手法では、文全体の正しい係り受け関係を求めることができないので、あまりよい精度は得られない。

10

【0005】

また、「箱田、佐藤：文音声合成における音調規則（電子情報通信学会論文誌 Vol.J63-D, No.9）」、「海木、匂坂：局所的な句構造によるポーズ挿入規則化の検討（電子情報通信学会論文誌 Vol.J79-D-2, No.9）」、「藤尾、匂坂、樋口：確率文脈自由文法を用いた韻律句境界とポーズ位置の予測（電子情報学会論文誌 Vol.J80-D2, No.1）」、特開平6-342297号公報等では、文全体における文節間の係り受け関係に基づいて設定している。

【0006】

これらの手法では、多量の統計データが存在すれば、精度のよいポーズ設定が可能である。

20

【0007】

【発明が解決しようとする課題】

上述した従来の音声合成システムは、ポーズ設定の精度が不十分という問題点がある。

【0008】

その理由は、特開平6-59695号公報等に記載された手法では、文全体の正しい係り受け関係を得ることができないので、十分なポーズ精度が得られないことである。また、特開平6-342297号公報等に記載された手法では、限られたデータを用いた際に、データと類似した構造のテキストに対してはよい精度が得られるが、そうでない場合には精度が不十分である。

【0009】

30

本発明の目的は、高精度なポーズ設定の可能な音声合成システム、音声合成方法および音声合成用プログラムを提供することにある。

【0010】

【課題を解決するための手段】

本発明の音声合成システムは、文全体の係り受け関係を取得する係り受け解析手段と、取得した前記文全体の係り受けの信頼度を求める信頼度抽出手段と、求めた前記信頼度を閾値と比較して前記文全体の係り受けを破棄するか決定する信頼度判定手段と、前記文全体の係り受けを前記信頼度判定手段が破棄した場合に、文の局所の係り受け関係を取得する局所係り受け解析手段と、ポーズを設定するポーズ挿入手段とを備え、前記ポーズ挿入手段は、前記文全体の係り受けが破棄された場合には前記文の局所の係り受けを使用し、前記文全体の係り受けが破棄されなかった場合には前記文全体の係り受けを使用することを特徴とする。

40

【0011】

本発明の音声合成システムは、テキストを形態素解析する形態素解析手段と、形態素解析された各単語を文節単位にまとめ上げる文節生成手段と、文節列のどこにポーズを設定するか決めるポーズ設定手段と、アクセントの立ち上がりを含む韻律情報を生成する韻律設定手段と、前記韻律設定手段が生成した発音記号から音声波形を生成する音声波形生成手段とを有し、前記ポーズ設定手段は、文全体の係り受け関係を取得する係り受け解析手段と、取得した前記文全体の係り受けの信頼度を求める信頼度抽出手段と、求めた前記信頼度を閾値と比較して前記文全体の係り受けを破棄するか決定する信頼度判定手段と、前記

50

文全体の係り受けを前記信頼度判定手段が破棄した場合に、文の局所の係り受け関係を取得する局所係り受け解析手段と、ポーズを設定するポーズ挿入手段とを備え、前記ポーズ挿入手段は、前記文全体の係り受けが破棄された場合には前記文の局所の係り受けを使用し、前記文全体の係り受けが破棄されなかった場合には前記文全体の係り受けを使用することを特徴としてもよい。

【0012】

本発明の音声合成システムは、前記局所係り受け解析手段は隣接した2文節が係ることができるかを判定して文の局所の係り受け関係を取得することを特徴としてもよい。

【0013】

本発明の音声合成方法は、文全体の係り受け関係を取得し、取得した前記文全体の係り受けの信頼度を求め、求められた前記信頼度を閾値と比較して前記文全体の係り受けを破棄するか決定し、前記文全体の係り受けを破棄した場合には文の局所の係り受け関係を取得してポーズを設定し、前記文全体の係り受けを破棄しなかった場合には前記文全体の係り受けを元にポーズを設定することを特徴とする。10

【0014】

本発明の音声合成方法は、テキストを形態素解析し、形態素解析された各単語を文節単位にまとめ上げ、文全体の係り受け関係を取得し、取得した前記文全体の係り受けの信頼度を求め、求められた前記信頼度を閾値と比較して前記文全体の係り受けを破棄するか決定し、前記文全体の係り受けを破棄した場合には文の局所の係り受け関係を取得してポーズを設定し、前記文全体の係り受けを破棄しなかった場合には前記文全体の係り受けを元にポーズを設定し、アクセントの立ち上がりを含む韻律情報を生成し、生成した発音記号から音声波形を生成することを特徴としてもよい。20

【0015】

本発明の音声合成方法は、隣接した2文節が係ることができると判定して文の局所の係り受け関係を取得することを特徴としてもよい。

【0016】

本発明のプログラムは、文全体の係り受け関係を取得する処理と、取得した前記文全体の係り受けの信頼度を求める処理と、求められた前記信頼度を閾値と比較して前記文全体の係り受けを破棄するか決定する処理と、前記文全体の係り受けを破棄した場合には文の局所の係り受け関係を取得してポーズを設定し、前記文全体の係り受けを破棄しなかった場合には前記文全体の係り受けを元にポーズを設定する処理とをコンピュータに実行させる。30

【0017】

本発明のプログラムは、テキストを形態素解析する処理と、形態素解析された各単語を文節単位にまとめ上げる処理と、文全体の係り受け関係を取得する処理と、取得した前記文全体の係り受けの信頼度を求める処理と、求められた前記信頼度を閾値と比較して前記文全体の係り受けを破棄するか決定する処理と、前記文全体の係り受けを破棄した場合には文の局所の係り受け関係を取得してポーズを設定し、前記文全体の係り受けを破棄しなかった場合には前記文全体の係り受けを元にポーズを設定する処理と、アクセントの立ち上がりを含む韻律情報を生成する処理と、生成した発音記号から音声波形を生成する処理とをコンピュータに実行させることを特徴としてもよい。40

【0018】

本発明のプログラムは、隣接した2文節が係ことができると判定して文の局所の係り受け関係を取得する処理をコンピュータに実行せることを特徴としてもよい。

【0019】

【発明の実施の形態】

次に、本発明の実施の形態について図面を参照して詳細に説明する。図1は本発明の実施の形態の構成を示すブロック図である。

【0020】

図1を参照すると、本発明の実施の形態は、テキストを形態素解析する形態素解析手段1

50

と、形態素解析された各単語を文節単位にまとめ上げる文節生成手段2と、文節列のどこにポーズを設定するか決めるポーズ設定手段3と、アクセントの立ち上がり等の韻律情報を生成する韻律設定手段4と、韻律設定手段4が生成した発音記号から音声波形を生成する音声波形生成手段5とを含む。

**【0021】**

図2はポーズ設定手段3の構成を示すブロック図である。図2を参照すると、ポーズ設定手段3は、文全体の係り受けを求める係り受け解析手段31と、係り受け解析手段31が求めた係り受けの信頼度を求める信頼度抽出手段32と、信頼度抽出手段32が求めた信頼度から文全体の係り受け関係が信頼できるか判定する信頼度判定手段33と、信頼度判定手段33によって信頼できないと判断された際に文の局所的な係り受けを求める局所係り受け解析手段34と、係り受け解析手段31または局所係り受け解析手段34が求めた係り受け結果を元にポーズを挿入するポーズ挿入手段35とを含む。

10

**【0022】**

これらの手段はそれぞれ概略つぎのように動作する。

**【0023】**

形態素解析手段1は、テキストを単語単位に分割し、各単語毎に読み、品詞等の付属情報を取得する。

**【0024】**

文節生成手段2は、形態素解析手段1で得られた単語列を各単語の品詞等を参照して文節単位にまとめ上げる。

20

**【0025】**

係り受け解析手段31は文節生成手段2で得られた文節列に対し、各文節がどの文節に係るかを文全体の情報を元に決定する。

**【0026】**

信頼度抽出手段32は、係り受け解析手段31で得られた文節間の係り受け関係が、どの程度信頼できるかを表す、信頼度を求める。

**【0027】**

信頼度判定手段33は、信頼度抽出手段32が求めた信頼度を閾値と比較し、閾値以上であれば処理をポーズ挿入手段35に移す。閾値未満の場合は、係り受け解析手段31で得られた係り受け関係を破棄して局所係り受け解析手段34に進む。

30

**【0028】**

局所係り受け解析手段34は、文節生成手段2で得られた文節列に対し、各文節がどの文節に係るかを各文節の局所的な情報を元に決定する。

**【0029】**

ポーズ挿入手段35は、係り受け解析手段31もしくは局所係り受け解析手段34で得られた係り受け関係を元に、文節列中にポーズを挿入する。

**【0030】**

韻律設定手段4は、ポーズ挿入手段35から得られたポーズの入った文節列に対し、アクセントの立ち上がりなどの韻律情報を付与し、発音記号列を生成する。

**【0031】**

音声波形生成手段5は、韻律設定手段4から得られた発音記号列を音声波形に変換して出力する。

40

**【0032】**

次に、図1及び図2を参照して本実施の形態の全体の動作について具体例をあげて詳細に説明する。

**【0033】**

まず、テキスト「彼が話す話を聞いた」が入力されると、形態素解析手段1によって、「彼 / が / 話す / 話 / を / 聞い / た」というように単語に分割される。この際同時に、「かれ / が / はな / す / はなし / お / きい / た」のような各単語の読み情報と「名詞/助詞が/動詞連体形/名詞/助詞を/動詞連用形/助動詞た終止形」のような各単語の品詞情報

50

も得る。「」はアクセントを示している。

**【0034】**

次に、このような単語列を文節生成手段2を用いて文節に組み替える。文節への組み替えの結果、「彼が／話す／話を／聞いた」「か’れが／はな’す／はなし’お／きいた」「助詞が／動詞連体形／助詞を／動詞終止系」というような文節列とその読み、品詞の情報が得られる。

**【0035】**

次に、このような文節列に対し係り受け解析手段31で文節間の係り受け関係を求める。係り受け関係は予め学習した確率文脈依存文法(Stochastic Context Free Grammar: SCFG)により決定する。SCFGによる係り受けの求め方に関しては、「藤尾、勾坂、樋口：確率文脈自由文法を用いた韻律句境界とポーズ位置の予測(電子情報学会論文誌Vol.J80-D2, No.1)」に詳述されている。この結果、各文節に関して、「彼が」「聞いた」「話す」「話を」「話を」「聞いた」という、係り受けの関係とそのような係り受けが生成される確率を得ることができる。

**【0036】**

次に、先の係り受けが生成される確率から信頼度抽出手段32で係り受けの信頼度を求める。信頼度は、確率を文節数で正規化することで得られる。

**【0037】**

次に、先の信頼度を信頼度判定手段33で閾値と比較する。信頼度が閾値を超えた場合は、先の係り受け関係(「彼が」「聞いた」「話す」「話を」「話を」「聞いた」)をそのまま用いてポーズ挿入手段35でポーズ挿入する。

**【0038】**

一方、信頼度が閾値を越えない場合、先の係り受け関係を破棄して局所係り受け解析手段34に進む。

**【0039】**

次に、局所係り受け解析手段34では先の文節生成手段2による文節列に対し文節間の係り受け関係を求める。係り受け関係は、隣接した2文節に着目し、該当文節が係ることができるか非かを判定する。

**【0040】**

この場合、「彼が」「話す」「話を」がそれぞれ直後の文節である「話す」「話を」「聞いた」に係るか非かを判定する。判定では、各文節の品詞からテーブル参照する。この結果、「彼が」「話す」「話す」「話を」「話を」「聞いた」という係り受けの関係を得ることができる。

**【0041】**

次に、ポーズ挿入手段35では先の係り受け解析手段31あるいは局所係り受け解析手段34から得られた係り受け関係から、ポーズ位置を設定する。ポーズ位置は、先の係り受け関係においてある文節が直後の文節に係らない際、その2文節の境界とする。従って、係り受け解析手段31による係り受けからポーズを求めた場合、「彼が」と「話す」の間にポーズが入る。これに対し、局所係り受け解析手段34による係り受けを用いた場合、ポーズは入らない。ここで、「彼」が「話した」のか「聞いた」のかはわからないため、音声合成システムの聞き手に「彼」が「聞いた」と思いこませるようなポーズは挿入しないほうがよい。

**【0042】**

次に、韻律設定手段4では、ポーズに加え、アクセントの立ち上がりなどを加えて発音記号を生成する。その結果、「か’れが2はな’す2はなし’お1きいた」という発音記号を得る。ここで、発音記号中の数字は、該当部分でピッチをどれだけ上げるかを示している。

**【0043】**

最後に、音声波形生成手段5で、韻律設定手段4が生成した発音記号から合成音声を生成する。

10

20

30

40

50

**【 0 0 4 4 】**

なお、本発明は、専用のシステムによらず、通常のコンピュータシステムを用いて実現可能である。音声合成用プログラムを記録媒体に格納し、記録媒体からコンピュータに読み込ませて、形態素解析手段1と文節生成手段2とポーズ設定手段3と韻律設定手段4と音声波形生成手段5の処理と同様の処理をコンピュータに実行させる。この記録媒体は、磁気ディスク、半導体メモリ、光ディスク、その他の記録媒体であってよい。音声合成用プログラムは記録媒体から直接読み込まれる代わりに、インターネット等のネットワークを介して読み込まれてもよい。

**【 0 0 4 5 】****【 発明の効果 】**

10

以上説明したように、本発明は、係り受け解析手段が信頼度の低い解析をした際、これを破棄して局所的な係り受けを求めるため、精度の高いポーズを得ることができるという効果がある。

**【 図面の簡単な説明 】**

【図1】本発明の実施の形態の構成を示すブロック図である。

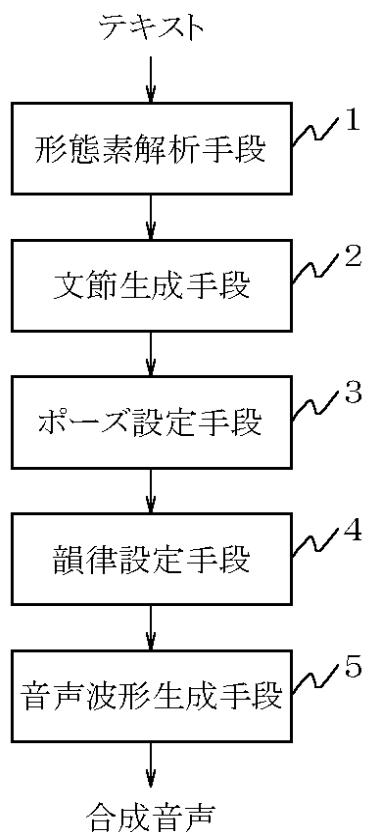
【図2】ポーズ設定手段の構成を示すブロック図である。

**【 符号の説明 】**

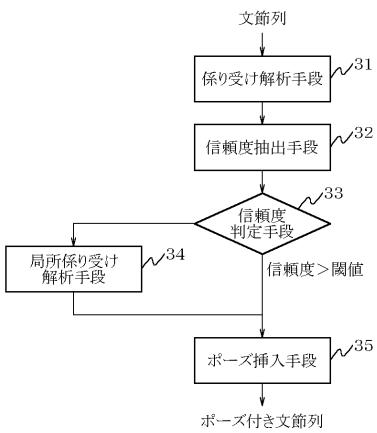
- |     |            |
|-----|------------|
| 1   | 形態素解析手段    |
| 2   | 文節生成手段     |
| 3   | ポーズ設定手段    |
| 4   | 韻律設定手段     |
| 5   | 音声波形生成手段   |
| 3 1 | 係り受け解析手段   |
| 3 2 | 信頼度抽出手段    |
| 3 3 | 信頼度判定手段    |
| 3 4 | 局所係り受け解析手段 |
| 3 5 | ポーズ挿入手段    |

20

【図1】



【図2】



---

フロントページの続き

(56)参考文献 特開平11-288292(JP,A)

特開平05-134692(JP,A)

鈴木和洋、斎藤隆, 日本語テキスト音声合成のための文節構造解析とそれに基づく韻律制御,  
電子情報通信学会論文誌. D-II, 情報・システム, II-情報処理, 日本, 電子情報通信学会, 1  
995年 2月25日, J78-D-2(2), p.177 - p.187

(58)調査した分野(Int.Cl., DB名)

G10L 13/00-13/08

CiNii