

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5874639号
(P5874639)

(45) 発行日 平成28年3月2日(2016.3.2)

(24) 登録日 平成28年1月29日(2016.1.29)

(51) Int.Cl. F I
G 1 O L 13/06 (2013.01) G 1 O L 13/06 1 3 0

請求項の数 6 (全 21 頁)

(21) 出願番号	特願2012-532854 (P2012-532854)	(73) 特許権者	000004237
(86) (22) 出願日	平成23年9月1日(2011.9.1)		日本電気株式会社
(86) 国際出願番号	PCT/JP2011/004918		東京都港区芝五丁目7番1号
(87) 国際公開番号	W02012/032748	(74) 代理人	100103090
(87) 国際公開日	平成24年3月15日(2012.3.15)		弁理士 岩壁 冬樹
審査請求日	平成26年8月4日(2014.8.4)	(74) 代理人	100124501
(31) 優先権主張番号	特願2010-199229 (P2010-199229)		弁理士 塩川 誠人
(32) 優先日	平成22年9月6日(2010.9.6)	(72) 発明者	加藤 正徳
(33) 優先権主張国	日本国(JP)		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	井上 健一

最終頁に続く

(54) 【発明の名称】 音声合成装置、音声合成方法及び音声合成プログラム

(57) 【特許請求の範囲】

【請求項1】

言語情報と韻律情報のモデルパラメータとをともに、隠れマルコフモデルにおける各状態の継続長を示す状態継続長を生成する状態継続長生成手段と、

言語情報から音声特徴量を導出し、導出された音声特徴量をもとに、前記状態継続長を補正する度合いを表す指標である継続長補正度を計算する継続長補正度計算手段と、

音韻の継続時間長を補正する補正比率を表わす音韻継続長補正パラメータと前記継続長補正度とに基づいて、前記状態継続長を補正する状態継続長補正手段とを備え、

前記継続長補正度計算手段は、前記音声特徴量を示すスペクトルまたはピッチの時間変化度を言語情報から推定し、推定した時間変化度をもとに前記継続長補正度を計算する

ことを特徴とする音声合成装置。

【請求項2】

状態継続長補正手段は、音声特徴量の時間的变化度が小さい状態における状態継続長ほど、当該状態継続長の変化度をより大きくする

請求項1記載の音声合成装置。

【請求項3】

言語情報と状態継続長生成手段が生成した状態継続長とをともに、ピッチパターンを生成するピッチパターン生成手段と、

言語情報と前記状態継続長とをともに、音声波形を表すパラメータである音声波形パラメータを生成する音声波形パラメータ生成手段とを備え、

10

20

継続長補正度計算手段は、言語情報と前記ピッチパターンと前記音声波形パラメータとに基づいて、継続長補正度を計算する

請求項 1 または 請求項 2 記載の音声合成装置。

【請求項 4】

言語情報と状態継続長補正手段が補正した状態継続長とをもとに、音声波形を表すパラメータである音声波形パラメータを生成する音声波形パラメータ生成手段と、

ピッチパターンと前記音声波形パラメータとをもとに合成音声波形を生成する波形生成手段とを備えた

請求項 1 または 請求項 2 記載の音声合成装置。

【請求項 5】

言語情報と韻律情報のモデルパラメータとをもとに、隠れマルコフモデルにおける各状態の継続長を示す状態継続長を生成し、

言語情報から音声特徴量を導出し、

導出された音声特徴量をもとに、前記状態継続長を補正する度合いを表す指標である継続長補正度を計算し、

音韻の継続時間長を補正する補正比率を表わす音韻継続長補正パラメータと前記継続長補正度とに基づいて、前記状態継続長を補正し、

継続長補正度を計算する際、前記音声特徴量を示すスペクトルまたはピッチの時間変化度を言語情報から推定し、推定した時間変化度をもとに前記継続長補正度を計算する

ことを特徴とする音声合成方法。

【請求項 6】

コンピュータに、

言語情報と韻律情報のモデルパラメータとをもとに、隠れマルコフモデルにおける各状態の継続長を示す状態継続長を生成する状態継続長生成処理、

言語情報から音声特徴量を導出し、導出された音声特徴量をもとに、前記状態継続長を補正する度合いを表す指標である継続長補正度を計算する継続長補正度計算処理、および、

音韻の継続時間長を補正する補正比率を表わす音韻継続長補正パラメータと前記継続長補正度とに基づいて、前記状態継続長を補正する状態継続長補正手処理を実行させ、

前記継続長補正度計算処理で、前記音声特徴量を示すスペクトルまたはピッチの時間変化度を言語情報から推定させ、推定された時間変化度をもとに前記継続長補正度を計算させる

ための音声合成プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、テキストから音声合成する音声合成装置、音声合成方法及び音声合成プログラムに関する。

【背景技術】

【0002】

テキスト文を解析し、その文が示す音声情報から合成音声生成する音声合成装置が知られている。近年、このような音声合成装置に対し、音声認識分野で広く普及している HMM (Hidden Markov Model: 隠れマルコフモデル) を適用する事例が注目されている。

【0003】

図 13 は、HMM を説明する説明図である。HMM は、図 13 に示すように、出力ベクトルを出力する確率分布が $b_i(o_t)$ であるような信号源 (状態) が、状態遷移確率 $a_{ij} = P(q_t = j | q_{t-1} = i)$ をもって接続されたものとして定義される。ただし、 i, j は状態番号とする。出力ベクトル o_t は、ケプストラムや線形予測係数などの音声の短時間的なスペクトルや、音声のピッチ周波数などを表現するパラメータである。す

10

20

30

40

50

なわち、HMMは、時間方向とパラメータ方向との変動を統計的にモデル化したものであるため、様々な要因で変動する音声をパラメータ系列の表現として表わすのに適していることが知られている。

【0004】

HMMに基づく音声合成装置では、まず、テキスト文の解析結果を基に合成音声の韻律情報（音の高さ（ピッチ周波数）、音の長さ（音韻継続長））を生成する。次に、テキスト解析結果と生成された韻律情報とを基に、波形生成パラメータを取得して音声波形を生成する。なお、波形生成パラメータは、メモリ（波形生成パラメータ記憶部）等に記憶されている。

【0005】

また、このような音声合成装置では、非特許文献1～3に記載されているように、韻律情報のモデルパラメータを記憶したモデルパラメータ記憶部を有している。このような音声合成装置は、音声合成を行う際、テキスト解析結果に基づいて、モデルパラメータ記憶部からHMMの状態ごとにモデルパラメータを取得して韻律情報を生成する。

【0006】

また、特許文献1には、音韻継続時間長を修正して合成音を生成する音声合成装置が記載されている。特許文献1に記載された音声合成装置では、音韻長の総和データに対する補間長の比率を個々の音韻長に乗算することにより、各音韻長への補完効果を分配した修正音韻長を算出する。この処理によって、個々の音韻長を修正する。

【0007】

なお、特許文献2には、規則音声合成装置における発声速度制御方式が記載されている。特許文献2に記載された発声速度制御方式では、各音素の継続時間長を求め、実音声を分析して得られた発声速度の変化に対する音素別の継続時間長の変化率データに基づいて発声速度を算出する。

【先行技術文献】

【特許文献】

【0008】

【特許文献1】特開2000-310996号公報

【特許文献2】特開平4-170600号公報

【非特許文献】

【0009】

【非特許文献1】益子、外3名、「動的特徴を用いたHMMに基づく音声合成」、電子情報通信学会論文誌 D-II, Vol.J79-D-II, No.12, pp.2184-2190, 1996年12月

【非特許文献2】徳田、「HMMによる音声合成の基礎」、電子情報通信学会技術研究報告、Vol.100, No.392, pp.43-50, 2000年10月

【非特許文献3】H. Zen, et. al., "A Hidden Semi-Markov Model-Based Speech Synthesis System", IEICE Trans. INF. & SYST., Vol.E90-D, No.5, pp.825-834, 2007

【発明の概要】

【発明が解決しようとする課題】

【0010】

非特許文献1や非特許文献2に記載された方法によれば、合成音声の各音素の継続時間長は、各音素に属する状態の継続長の総和で与えられる。例えば、音素の状態数が3状態であり、音素aの状態1～3までの継続長がd1, d2, d3であった場合、音素aの継続時間長は、d1+d2+d3で与えられる。各状態の継続長は、モデルパラメータである平均と分散と、文全体の時間長から定まる定数により決定される。つまり、状態1の平均がm1、分散がσ1、文全体の時間長から定まる定数をρとしたとき、状態1の状態継続長d1は、以下に示す式1で計算できる。

【0011】

$$d1 = m1 + \rho \cdot \sigma1 \quad (\text{式1})$$

【0012】

したがって、 p が平均および分散よりも著しく大きい場合、状態継続長は、分散に大きく依存することになる。すなわち、非特許文献 1～2 に記載された手法では、音韻継続時間長に相当する HMM の状態継続長は、各状態継続長のモデルパラメータである平均および分散をもとに決定されるが、分散が大きい状態における継続長は長くなりやすいという問題点がある。

【0013】

一般に、子音と母音から構成される音節の自然音声进行分析すると、子音部の時間長は、母音部よりも短いことが多い。ところが、子音に属する状態の分散が母音に属する状態の分散よりも大きいと、その音節の継続時間長は子音のほうが長くなることがある。母音よりも子音の継続時間長が長い音節が頻繁に出現すると、その合成音声の発話リズムは不自然になり、聞き取りにくい合成音声となる。このような場合、発話リズムが自然である、聞き取り易い合成音声を生成することは難しい。

10

【0014】

また、特許文献 1 に記載された音声合成装置を用いたとしても、HMM を用いたピッチパタンの生成は困難であり、発話リズムの自然性を高めた聞き取り易い合成音声を生成できるとは言い難い。

【0015】

そこで、本発明は、発話リズムの自然性が高く、聞き取り易い合成音声を生成できる音声合成装置、音声合成方法及び音声合成プログラムを提供することを目的とする。

【課題を解決するための手段】

20

【0016】

本発明による音声合成装置は、言語情報と韻律情報のモデルパラメータとをもとに、隠れマルコフモデルにおける各状態の継続長を示す状態継続長を生成する状態継続長生成手段と、言語情報から音声特徴量を導出し、導出された音声特徴量をもとに、状態継続長を補正する度合いを表す指標である継続長補正度を計算する継続長補正度計算手段と、音韻の継続時間長を補正する補正比率を表わす音韻継続長補正パラメータと継続長補正度とに基づいて、状態継続長を補正する状態継続長補正手段とを備え、継続長補正度計算手段が、音声特徴量を示すスペクトルまたはピッチの時間変化度を言語情報から推定し、推定した時間変化度をもとに継続長補正度を計算することを特徴とする。

【0017】

30

本発明による音声合成方法は、言語情報と韻律情報のモデルパラメータとをもとに、隠れマルコフモデルにおける各状態の継続長を示す状態継続長を生成し、言語情報から音声特徴量を導出し、導出された音声特徴量をもとに、状態継続長を補正する度合いを表す指標である継続長補正度を計算し、音韻の継続時間長を補正する補正比率を表わす音韻継続長補正パラメータと継続長補正度とに基づいて、状態継続長を補正し、継続長補正度を計算する際、音声特徴量を示すスペクトルまたはピッチの時間変化度を言語情報から推定し、推定した時間変化度をもとに継続長補正度を計算することを特徴とする。

【0018】

本発明による音声合成プログラムは、コンピュータに、言語情報と韻律情報のモデルパラメータとをもとに、隠れマルコフモデルにおける各状態の継続長を示す状態継続長を生成する状態継続長生成処理、言語情報から音声特徴量を導出し、導出された音声特徴量をもとに、状態継続長を補正する度合いを表す指標である継続長補正度を計算する継続長補正度計算処理、および、音韻の継続時間長を補正する補正比率を表わす音韻継続長補正パラメータと継続長補正度とに基づいて、状態継続長を補正する状態継続長補正手処理を実行させ、継続長補正度計算処理で、音声特徴量を示すスペクトルまたはピッチの時間変化度を言語情報から推定させ、推定された時間変化度をもとに継続長補正度を計算させることを特徴とする。

40

【発明の効果】

【0019】

本発明によれば、発話リズムの自然性が高く、聞き取り易い合成音声を生成できる。

50

【図面の簡単な説明】

【0020】

【図1】本発明の第1の実施形態における音声合成装置の例を示すブロック図である。

【図2】第1の実施形態における音声合成装置の動作の例を示すフローチャートである。

【図3】本発明の第2の実施形態における音声合成装置の例を示すブロック図である。

【図4】言語情報をもとに算出された各状態における補正度の例を示す説明図である。

【図5】仮ピッチパターンに基づいて計算された補正度の例を示す説明図である。

【図6】仮ピッチパターンに基づいて計算された補正度の例を示す説明図である。

【図7】音声波形パラメータに基づいて計算された補正度の例を示す説明図である。

【図8】音声波形パラメータに基づいて計算された補正度の例を示す説明図である。

10

【図9】第2の実施形態における音声合成装置の動作の例を示すフローチャートである。

【図10】本発明の第3の実施形態における音声合成装置の例を示すブロック図である。

【図11】第3の実施形態における音声合成装置の動作の例を示すフローチャートである。

。

【図12】本発明による音声合成装置の最小構成の例を示すブロック図である。

【図13】HMMを説明する説明図である。

【発明を実施するための形態】

【0021】

以下、本発明の実施形態を図面を参照して説明する。

【0022】

20

実施形態1.

図1は、本発明の第1の実施形態における音声合成装置の例を示すブロック図である。本実施形態における音声合成装置は、言語処理部1と、韻律生成部2と、素片情報記憶部12と、素片選択部4と、波形生成部5とを備えている。また、韻律生成部2は、状態継続長生成部21と、状態継続長補正部22と、音素継続長計算部23と、継続長補正度計算部24と、モデルパラメータ記憶部25と、ピッチパターン生成部3とを備えている。

【0023】

素片情報記憶部12は、音声合成単位ごとに生成された素片と、各素片の属性情報とを記憶している。素片とは、音声合成単位の音声波形を表す情報であり、波形自身や、波形から抽出されたパラメータ（例えば、スペクトル、ケプストラム、線形予測フィルタ係数）などで表わされる。より具体的には、素片は、音声合成単位毎に分割された（切り出された）音声波形、線形予測分析パラメータやケプストラム係数に代表されるような、切り出された音声波形から抽出される波形生成パラメータの時系列、などである。音素は、多くの場合、例えば、人間が発した音声（自然音声波形と言うこともある）から抽出される情報をもとに生成される。例えば、アナウンサーや声優が発した（発声した）音声を録音した情報から音素が生成される。

30

【0024】

音声合成単位は任意であり、例えば、音素、音節などであってよい。また、音声合成単位は、以下の参考文献1や参考文献2に記載されているように、音素に基づいて定められるCV単位や、VCV単位、CVC単位などであってもよい。また、音声合成単位は、COC方式に基づいて定められる単位であってもよい。ここで、Vは母音を表わし、Cは子音を表わす。

40

【0025】

<参考文献1>

Huang, Acero, Hon, “Spoken Language Processing”, Prentice Hall, pp.689-836, 2001.

<参考文献2>

阿部 外2名, “音声合成のための合成単位の基礎” 電子情報通信学会技術研究報告, Vol.100, No.392, pp.35-42, 2000.

【0026】

50

言語処理部 1 は、入力されたテキスト（文字列情報）に対して、形態素解析、構文解析、読み付け等の分析を行い、言語情報を生成する。言語処理部 1 が生成する言語情報には、少なくとも、音節記号や音素記号などの「読み」を表す情報が含まれる。また、言語処理部 1 は、上記「読み」を表わす情報に加え、形態素の品詞、活用などのいわゆる「日本語文法」を表す情報、アクセント型、アクセント位置、アクセント句区切り等を表す「アクセント情報」を含む言語情報を生成してもよい。そして、言語処理部 1 は、生成した言語情報を状態継続長生成部 2 1、ピッチパターン生成部 3 および素片選択部 4 に入力する。

【0027】

なお、後述する状態継続長生成部 2 1、ピッチパターン生成部 3 および素片選択部 4 が言語情報を利用する実施形態に応じ、言語情報に含まれるアクセント情報や形態素情報の内容はそれぞれ異なる。

10

【0028】

モデルパラメータ記憶部 2 5 は、韻律情報のモデルパラメータを記憶する。具体的には、モデルパラメータ記憶部 2 5 は、状態継続長のモデルパラメータを記憶する。また、モデルパラメータ記憶部 2 5 は、ピッチ周波数のモデルパラメータを記憶してもよい。モデルパラメータ記憶部 2 5 は、韻律情報に応じたモデルパラメータを予め記憶する。なお、モデルパラメータには、例えば、HMM によって予め韻律情報をモデル化したモデルパラメータが用いられる。

【0029】

状態継続長生成部 2 1 は、言語処理部 1 から入力された言語情報と、モデルパラメータ記憶部 2 5 に記憶されたモデルパラメータとをもとに、状態継続長を生成する。ここで、ある音素に属する各状態の継続長は、その音素（以下、該当音素と記す。）の前後に存在する音素（先行音素、後続音素と呼ぶこともある。）や、該当音素のアクセント句内でのモーラ位置、先行音素、該当音素および後続音素が属するアクセント句のモーラ長やアクセント型、該当音素が属するアクセント句の位置などの「コンテキスト」と呼ばれる情報に基づいて一意に決定される。つまり、ある任意のコンテキスト情報に対してモデルパラメータが一意に決定される。具体的には、モデルパラメータは、平均および分散である。

20

【0030】

そこで、状態継続長生成部 2 1 は、非特許文献 1～3 に記載されているように、入力されたテキストの解析結果をもとにモデルパラメータ記憶部 2 5 からモデルパラメータを選択し、選択したモデルパラメータに基づいて状態継続長を生成する。そして、状態継続長生成部 2 1 は、生成した状態継続長を状態継続長補正部 2 2 へ入力する。この状態継続長とは、HMM における各状態が継続する時間長である。

30

【0031】

モデルパラメータ記憶部 2 5 が記憶する状態継続長のモデルパラメータは、HMM の状態継続確率を特徴づけるパラメータに相当する。HMM の状態継続確率は、非特許文献 1～3 にも記載されているように、ある状態が継続する（すなわち、自己遷移する）回数の確率のことであり、ガウス分布で定義されることが多い。ガウス分布は、平均と分散の二種類の統計量で特徴づけられる。そこで、本実施形態では、状態継続長のモデルパラメータをガウス分布の平均および分散と仮定する。ここで、HMM の状態継続長の平均 ξ_j と分散 σ^2_j とは、以下に示す式 2 で算出される。このとき、生成される状態継続長は、非特許文献 3 に記載されているように、モデルパラメータの平均に一致する。

40

【0032】

【数 1】

$$\xi_j = \frac{\sum_{t_0=1}^T \sum_{t_1=t_0}^T x_{t_0, t_1}(j) \cdot (t_1 - t_0 + 1)}{\sum_{t_0=1}^T \sum_{t_1=t_0}^T x_{t_0, t_1}(j)}$$

(式 2)

10

$$\sigma_j^2 = \frac{\sum_{t_0=1}^T \sum_{t_1=t_0}^T x_{t_0, t_1}(j) \cdot (t_1 - t_0 + 1)^2}{\sum_{t_0=1}^T \sum_{t_1=t_0}^T x_{t_0, t_1}(j)} - \xi_j^2$$

【0033】

なお、状態継続長のモデルパラメータは、ガウス分布の平均および分散に限定されない。状態継続長のモデルパラメータは、例えば、非特許文献 2 の 2. 2 節に記載されているように、HMM の状態遷移確率 $a_{ij} = P(q_t = j \mid q_{t-1} = i)$ と、出力確率分布 $b_i(o_t)$ を利用して、EM アルゴリズムに基づいて推定されるものであってもよい。

20

【0034】

状態継続長のモデルパラメータに限らず、HMM のパラメータは、学習処理により求められる。学習には、音声データとその音素ラベルおよび言語情報が利用される。状態継続長のモデルパラメータの学習方法は、公知の技術であるため、詳細な説明は省略する。

【0035】

なお、状態継続長生成部 21 は、文全体の時間長を定めてから、各状態の継続長を算出してもよい（非特許文献 1 ～ 2 参照）。ただし、モデルパラメータの平均に一致する状態継続長を算出することにより、標準的な話速を実現する状態継続長を算出できるため、より好ましい。

30

【0036】

継続長補正度計算部 24 は、言語処理部 1 から入力された言語情報に基づいて、継続長補正度（以下、単に補正度と記すこともある。）を計算し、状態継続長補正部 22 へ入力する。具体的には、継続長補正度計算部 24 は、言語処理部 1 から入力された言語情報から音声特徴量を算出し、その音声特徴量をもとに継続長補正度を計算する。ここで、継続長補正度は、後述する状態継続長補正部 22 が、HMM の状態の継続長をどの程度補正するかを示す指標である。補正度が大きくなるほど、状態継続長補正部 22 が状態継続長を補正する補正量は大きくなる。なお、継続補正度は、状態ごとに算出される。

【0037】

補正度は、上述の通り、スペクトルやピッチなどの音声特徴量、及びその時間変化度に関連した値になる。なお、ここで示す音声特徴量には、時間の長さを示す情報（以下、時間長情報と記す。）は含まれない。例えば、音声特徴量の時間変化度が小さいと推測される箇所では、継続長補正度計算部 24 は、補正度を大きくする。また、音声特徴量の絶対値が大きいと推測される箇所においても、継続長補正度計算部 24 は、補正度を大きくする。

40

【0038】

本実施形態では、継続長補正度計算部 24 が言語情報から音声特徴量を示すスペクトルまたはピッチの時間変化度を推定し、推定した音声特徴量の時間変化度をもとに補正度を計算する方法を説明する。

【0039】

50

例えば、ある特定の音節に対して補正を実施する場合、子音と母音とでは一般的に母音のほうが音声特徴量の時間変化が小さいと予想される。また、母音の中でも両端部よりも中心部のほうが時間変化は小さいと推定される。したがって、継続長補正度計算部24は、母音中心、母音両端、子音の順番で小さくなるように補正度を計算する。より詳細には、継続長補正度計算部24は、子音内部では均等になるように補正度を計算する。また、継続長補正度計算部24は、母音部では補正度が中心から両端（始端および終端）にかけて小さくなるように補正度を計算する。

【0040】

音節単位で補正度を決定する場合、継続長補正度計算部24は、音節の中心から両端にかけて補正度を小さくする。また、継続長補正度計算部24は、音素種別に応じて補正度を計算してもよい。例えば、子音の中では破裂音よりも鼻音のほうが音声特徴量の時間変化度が小さいため、継続長補正度計算部24は、鼻音の補正度を破裂音よりも大きくする。

10

【0041】

また、アクセント核の位置やアクセント句区切りなどのアクセント情報が言語情報に含まれている場合、継続長補正度計算部24は、これらの情報を補正度の計算に利用してもよい。例えば、アクセント核やアクセント句区切りの付近ではピッチの変化が大きいため、継続長補正度計算部24は、この付近の補正度を小さくする。

【0042】

また、有声音と無声音とを区別して補正度を設定する方法も有効な場合がある。この区別が有効か否かは、合成音声波形を生成する処理に関係する。波形生成の方法は、有声音と無声音で大きく異なることが多い。特に、無声音波形の波形生成方法では、時間長伸縮処理に伴う音質劣化が問題になることがある。このような場合、無声音の補正度を有声音よりも小さくしたほうが望ましい。

20

【0043】

本実施形態における補正度は、最終的に状態単位で定められ、その値は、状態継続長補正部22が直接利用するものとする。具体的には、補正度は0.0よりも大きい実数とし、0.0のときに補正度が最小であるとする。また、状態継続長を大きくするような補正を行う場合、補正度は1.0よりも大きい実数とし、状態継続長を小さくするような補正の場合、補正度は1.0よりも小さく0.0よりも大きい実数とする。ただし、補正度の値は、上記値に限定されない。例えば、状態継続長を大きくするような補正を行う場合と状態継続長を小さくするような補正を行う場合のいずれも、補正度の最小を1.0としてもよい。また、補正する位置を、音節や音素の始端、終端、中心などの相対位置で表わしてもよい。

30

【0044】

また、補正度の内容は数値に限定されない。例えば、補正の度合いを表わす適当なシンボル（「大、中、小」、「a、b、c、d、e」など）で補正度を定めてもよい。この場合、実際に補正値を求める処理において、状態単位で上記シンボルを実数値に変換する処理を行えばよい。

【0045】

状態継続長補正部22は、状態継続長生成部21から入力された状態継続長と、継続長補正度計算部24から入力された継続長補正度と、ユーザ等により入力された音韻継続長補正パラメータとに基づいて、状態継続長を補正する。そして、状態継続長補正部22は、補正した状態継続長を音素継続長計算部23およびピッチパターン生成部3へ入力する。

40

【0046】

音韻継続長補正パラメータとは、生成された音韻の継続時間長を補正するための補正比率を示す値である。なお、継続時間長には、状態継続長を加算して算出した音素や音節などの時間長も含まれる。音韻継続長補正パラメータは、補正後の継続時間長を補正前の継続時間長で除算したもの、及びその近似値として定義できる。ただし、音韻継続長補正パラメータの値は、HMMの状態単位で定められるものではなく、音素などの単位で定めら

50

れる。具体的には、音韻継続長補正パラメータは、ある特定の音素または半音素に対して1つ定められていてもよく、複数の音素に対して定められていてもよい。また、複数の音素に対して定められる音韻継続長補正パラメータは、共通であってもよく、別々であってもよい。さらに、音韻継続長補正パラメータは、単語や呼気段落、文全体に1つ定められていてもよい。以上のように、音韻継続長補正パラメータは、ある特定の音素におけるある特定の状態（すなわち、音素を示す各状態）に対しては設定されないものとする。

【0047】

音韻継続長補正パラメータは、ユーザや、音声合成装置と組み合わせて使用される他の装置、音声合成装置自身が備える他の機能などによって定めた値が用いられる。例えば、ユーザが合成音声を聞き、音声合成装置にもっとゆっくり音声を出力してほしい（しゃべってほしい）と判断した場合、ユーザは、例えば、音韻継続長補正パラメータとしてより大きな値を設定してもよい。また、文中のキーワードを選択的にゆっくり出力してほしい（しゃべってほしい）場合、ユーザは、通常発話とは別にキーワード用の音韻継続長補正パラメータを設定してもよい。

10

【0048】

上述するように、継続長補正度は、音声特徴量の時間変化度が小さいと推測される箇所ほど大きくなる。そのため、状態継続長補正部22は、音声特徴量の時間的変化度が小さい状態における状態継続長ほど、その状態継続長の変化度をより大きくする。

【0049】

具体的には、状態継続長補正部22は、音韻継続長補正パラメータと、継続長補正度と、補正前の状態継続長とをもとに、各状態に対する補正量を算出する。ここで、ある音素の状態数を N 、補正前の状態継続長を $m(1)$ 、 $m(2)$ 、 $\dots$ 、 $m(N)$ 、補正度を $\alpha(1)$ 、 $\alpha(2)$ 、 $\dots$ 、 $\alpha(N)$ 、入力された音韻継続長補正パラメータを ρ とする。このとき、各状態に対する補正量 $l(1)$ 、 $l(2)$ 、 $\dots$ 、 $l(N)$ は、以下に示す式3のように与えられる。

20

【0050】

【数2】

$$l(i) = \frac{(\rho - 1) \sum_{j=1}^N m(j)}{\sum_{j=1}^N \alpha(j)} \cdot \alpha(i), \text{ for } i = 1, 2, \dots, N \quad (\text{式3})$$

30

【0051】

そして、状態継続長補正部22は、算出した補正量を補正前の状態継続長に加算して補正值を求める。上記と同様に、ある音素の状態数を N 、補正前の状態継続長を $m(1)$ 、 $m(2)$ 、 $\dots$ 、 $m(N)$ 、補正度を $\alpha(1)$ 、 $\alpha(2)$ 、 $\dots$ 、 $\alpha(N)$ 、入力された音韻継続長補正パラメータを ρ とする。このとき、補正後の状態継続長は以下に示す式4のように与えられる。

40

【0052】

【数 3】

$$n(i) = m(i) + \frac{(\rho - 1) \sum_{j=1}^N m(j)}{\sum_{j=1}^N \alpha(j)} \cdot \alpha(i), \text{ for } i = 1, 2, \dots, N \quad (\text{式 4})$$

10

【0053】

なお、複数個からなる音素列に対して一つの音韻継続長補正パラメータ値 ρ が指定されている場合、状態継続長補正部 22 は、その音素列に含まれる全ての状態に対して、上記式を用いて補正量を計算すればよい。また、状態数が総計 M の場合、状態継続長補正部 22 は、上述する式 4 において、 N の代わりに M を用いて補正量を計算すればよい。

【0054】

また、状態継続長補正部 22 は、算出した補正量を補正前の状態継続長に乗じて補正值を求めてもよい。状態継続長補正部 22 は、例えば、以下に示す式 5 を用いて補正量を計算した場合、算出した補正量を補正前の状態継続長に乗じて補正值を求めればよい。なお、補正值の算出方法は、補正量の算出方法に応じて定めればよい。

20

【0055】

【数 4】

$$l'(i) = 1 + \frac{(\rho - 1) \sum_{j=1}^N m(j)}{\sum_{j=1}^N \alpha(j)} \cdot \frac{\alpha(i)}{m(j)}, \text{ for } i = 1, 2, \dots, N \quad (\text{式 5})$$

30

【0056】

音素継続長計算部 23 は、状態継続長補正部 22 から入力された状態継続長に基づいて各音素の継続長を計算し、素片選択部 4 と波形生成部 5 に計算結果を入力する。音素継続長は、各音素に属する全ての状態の状態継続長の総和で与えられる。したがって、音素継続長計算部 23 は、全ての音素に対して、状態継続長の総和を音素毎に計算することで、各音素の継続長を計算する。

【0057】

ピッチパターン生成部 3 は、言語処理部 1 から入力された言語情報と、状態継続長補正部 22 から入力された状態継続長とをもとにピッチパターンを生成し、素片選択部 4 および波形生成部 5 に入力する。ピッチパターン生成部 3 は、例えば、非特許文献 2 に記載されているように、MSD-HMM (Multi-Space Probability Distribution-HMM) によりピッチパターンをモデル化することにより、ピッチパターンを生成してもよい。ただし、ピッチパターン生成部 3 がピッチパターンを生成する方法は、上記方法に限定されない。ピッチパターン生成部 3 は、HMM によりピッチパターンをモデル化してもよい。なお、これらの方法は、広く知られているため、詳細な説明は省略する。

40

【0058】

素片選択部 4 は、言語解析の処理結果と、音素継続長と、ピッチパターンとに基づいて、素片情報記憶部 12 に記憶されている素片の中から、音声を合成するために最適な素片を選択し、選択した素片とその属性情報とを波形生成部 5 に入力する。

50

【0059】

ここで、入力テキストから生成された継続時間長およびピッチパターンが合成音声波形に忠実に適用されるとすれば、合成音声の韻律情報と呼ぶことができる。ただし、実際には類似の韻律（すなわち、継続時間長およびピッチパターン）が適用される。そのため、生成された継続時間長およびピッチパターンは、音声合成波形を生成するときに目標となる韻律情報と言えるため、以下の説明では、生成された継続時間長およびピッチパターンを、目標韻律情報と記すこともある。

【0060】

素片選択部4は、入力された言語解析の処理結果と目標韻律情報とに基づいて、合成音声の特徴を表す情報（以下、これを「目標素片環境」と呼ぶ。）を音声合成単位毎に求める。目標素片環境とは、該当音素、先行音素、後続音素、ストレスの有無、アクセント核からの距離、音声合成単位毎のピッチ周波数、パワー、単位の継続時間長、ケプストラム、MFCC（Mel Frequency Cepstral Coefficients）、及びこれらのΔ量（単位時間あたりの変化量）などである。

10

【0061】

次に、素片選択部4は、求めた目標素片環境に含まれる特定の情報（主に該当音素）に対応（例えば、一致）する音素を有する素片を素片情報記憶部12から複数取得する。取得された素片は、音声を合成するために用いられる素片の候補である。

【0062】

そして、素片選択部4は、取得された素片に対して、音声を合成するために用いる素片としての適切度を示す指標であるコストを算出する。コストは、目標素片環境と候補素片や、隣接候補素片同士の属性情報との差異を数値化したものであり、類似度が高いほど、つまり音声を合成するための適切度が高くなるほど小さくなる値である。コストが小さい素片を用いるほど、合成された音声は、人間が発した音声と類似している程度を表す自然度が高い音声となる。そのため、素片選択部4は、算出されたコストが最も小さい素片を選択する。

20

【0063】

素片選択部4が計算するコストには、具体的には単位コストと接続コストがある。単位コストは、候補素片を目標素片環境の下で用いることにより生じる推定音質劣化度を表すもので、候補素片の素片環境と目標素片環境との類似度を基に算出される。一方、接続コストは、接続する音声素片間の素片環境が不連続であることによって生じる推定音質劣化度を表すもので、隣接候補素片同士の素片環境の親和度を基に算出される。この単位コスト及び接続コストの計算方法は、これまで各種提案されている。一般に、単位コストの計算には、目標素片環境に含まれる情報が用いられる。一方、接続コストには、素片の接続境界におけるピッチ周波数、ケプストラム、MFCC、短時間自己相関、パワー、及びこれらのΔ量などが用いられる。以上の通り、単位コスト及び接続コストは、素片に関する各種情報（ピッチ周波数、ケプストラム、パワー等）を複数用いて算出される。

30

【0064】

素片選択部4は、単位コストおよび接続コストを素片ごとに計算したのちに、接続コストと単位コストの両者が最小となる音声素片を各合成単位に対して一意に求める。なお、コスト最小化により求めた素片は、候補素片の中から音声の合成に最も適した素片として選択されたものであることから、選択素片と呼ぶこともできる。

40

【0065】

波形生成部5は、素片選択部4が選択した素片を接続して合成音声を生成する。波形生成部5は、単純に素片を接続するだけでなく、韻律生成部2から入力された目標韻律情報、素片選択部4から入力された選択素片、及び素片の属性情報をもとに、目標韻律に一致または類似する韻律を有する音声波形を生成してもよい。そして、波形生成部5は、生成した音声波形を各々接続して合成音声を生成してもよい。波形生成部5が合成音声を生成する方法として、例えば、参考文献1に記載されているPSOLA（pitch synchronous overlap-add）法が挙げられる。ただし、波形生成部5

50

が合成音声を生成する方法は、上記方法に限定されない。選択された素片から合成音声を生成する方法は広く知られているため、詳細な説明は省略する。

【0066】

素片情報記憶部12、および、モデルパラメータ記憶部25は、例えば、磁気ディスク等により実現される。また、言語処理部1と、韻律生成部2（より詳しくは、状態継続長生成部21と、状態継続長補正部22と、音素継続長計算部23と、継続長補正度計算部24と、ピッチパターン生成部3）と、素片選択部4と、波形生成部5とは、プログラム（音声合成プログラム）に従って動作するコンピュータのCPUによって実現される。例えば、プログラムは、音声合成装置の記憶部（図示せず）に記憶され、CPUは、そのプログラムを読み込み、プログラムに従って、言語処理部1、韻律生成部2（より詳しくは、状態継続長生成部21、状態継続長補正部22、音素継続長計算部23、継続長補正度計算部24、ピッチパターン生成部3）、素片選択部4および波形生成部5として動作してもよい。また、言語処理部1と、韻律生成部2（より詳しくは、状態継続長生成部21と、状態継続長補正部22と、音素継続長計算部23と、継続長補正度計算部24と、ピッチパターン生成部3）と、素片選択部4と、波形生成部5とは、それぞれが専用のハードウェアで実現されていてもよい。

10

【0067】

次に、本実施形態における音声合成装置の動作を説明する。図2は、第1の実施形態における音声合成装置の動作の例を示すフローチャートである。まず、言語処理部1は、入力されたテキストから言語情報を生成する（ステップS1）。状態継続長生成部21は、言語情報とモデルパラメータとをもとに状態継続長を生成する（ステップS2）。また、継続長補正度計算部24は、言語情報をもとに継続長補正度を計算する（ステップS3）。

20

【0068】

状態継続長補正部22は、状態継続長と継続長補正度と音韻継続長補正パラメータとをもとに、状態継続長を補正する（ステップS4）。音素継続長計算部23は、補正された状態継続長をもとに、状態継続長の総和を計算する（ステップS5）。また、ピッチパターン生成部3は、言語情報と補正された状態継続長とをもとに、ピッチパターンを生成する（ステップS6）。素片選択部4は、入力されたテキストの解析結果である言語情報と、状態継続長の総和と、ピッチパターンとをもとに、音声の合成に用いられる素片を選択する（ステップS7）。そして、波形生成部5は、選択された素片を結合して合成音声を生成する（ステップS8）。

30

【0069】

以上のように、本実施形態によれば、状態継続長生成部21が、言語情報と韻律情報のモデルパラメータとをもとに、HMMにおける各状態の状態継続長を生成する。また、継続長補正度計算部24が、言語情報から導出された音声特徴量をもとに継続長補正度を計算する。そして、状態継続長補正部22が、音韻継続長補正パラメータと継続長補正度とに基づいて状態継続長を補正する。

【0070】

すなわち、本実施形態によれば、言語情報に基づいて推定した音声特徴量、及びその変化度から補正度を求め、その補正度に基づいて音韻継続長補正パラメータに応じた状態継続長補正を行っている。この結果、一般的な音声合成装置と比較して、発話リズムの自然性が高く、聞き取り易い合成音声を生成できる。

40

【0071】

例えば、特許文献1に記載されているように、本実施形態で説明した状態継続長を補正対象にするのではなく、音素継続長を補正対象とすることも考えられる。この場合、ピッチパターンを生成し、音素継続長を生成した後で、音素継続長の補正が行われ、最終的にピッチパターンの補正が行われることになる。しかし、この場合、最後のピッチパターンの補正において、不適切な変形が行われ、音質的に問題のあるピッチパターンが生成される可能性がある。例えば、補正後の音韻継続長から状態継続長を求めるときに、音韻継続長を等間

50

隔で分割したとする。この場合、ピッチパタンの形状が不適切となり、合成音声の品質が低くなる可能性がある。補正により音韻継続長が長くなった場合には、音節中心部のピッチパタンを長くして、音節の終端や始端のピッチパタンを伸ばさない方が、ピッチパタンを全て同様に引き延ばす場合と比較しても、音質的に望ましい。これは、自然音声を観察した場合、音節の両端の方が中心部に比べてピッチの変化が大きいことが多いためである。また、他にも、継続時間長を「音節両端では短く、音節中心では長く」と単純に割り当てても考えられる。しかし、HMMでモデル化し、多量の音声データを学習して得た結果（すなわち、補正前の状態継続長）を無視して、状態継続長を新たに作り出す方法も適切とは言えない。

【0072】

一方、本実施形態では、状態継続長を補正したうえでピッチパタンを生成し、音素継続長を生成する。そのため、このような不適切な変形が行われることを抑止できる。また、本実施形態では、状態継続長を決定する際、平均および分散といったモデルパラメータだけではなく、自然音声の性質を示す音声特徴量を利用している。そのため、自然性が高い合成音声を生成できる。

【0073】

実施形態2.

図3は、本発明の第2の実施形態における音声合成装置の例を示すブロック図である。なお、第1の実施形態と同様の構成については、図1と同一の符号を付し、説明を省略する。本実施形態における音声合成装置は、言語処理部1と、韻律生成部2と、素片情報記憶部12と、素片選択部4と、波形生成部5とを備えている。また、韻律生成部2は、状態継続長生成部21と、状態継続長補正部22と、音素継続長計算部23と、継続長補正度計算部242と、仮ピッチパタン生成部28と、音声波形パラメータ生成部29と、モデルパラメータ記憶部25と、ピッチパタン生成部3とを備えている。

【0074】

すなわち、図3に例示する音声合成装置は、継続長補正度計算部24が継続長補正度計算部242に置き換わり、仮ピッチパタン生成部28と、音声波形パラメータ生成部29とを新たに備えている点において、第1の実施形態と異なる。

【0075】

仮ピッチパタン生成部28は、言語処理部1から入力された言語情報と、状態継続長生成部21から入力された状態継続長とをもとに、仮ピッチパタンを生成し、継続長補正度計算部242へ入力する。仮ピッチパタン生成部28がピッチパタンの生成方法する方法は、ピッチパタン生成部3がピッチパタンを生成する方法と同様である。

【0076】

音声波形パラメータ生成部29は、言語処理部1から入力された言語情報と、状態継続長生成部21から入力された状態継続長とをもとに、音声波形パラメータを生成し、継続長補正度計算部242へ入力する。音声波形パラメータとは、具体的にはスペクトルやケプストラム、線形予測係数など、音声波形の生成に用いられるパラメータである。音声波形パラメータ生成部29は、HMMを利用して音声波形パラメータを生成してもよい。他にも、音声波形パラメータ生成部29は、例えば、非特許文献1に記載されているように、メルケプストラムを用いて音声波形パラメータを生成してもよい。なお、これらの方法は広く知られているため、詳細な説明は省略する。

【0077】

継続長補正度計算部242は、言語処理部1から入力された言語情報と、仮ピッチパタン生成部28から入力された仮ピッチパタンと、音声波形パラメータ生成部29から入力された音声波形パラメータとに基づいて、継続長補正度を計算し、状態継続長補正部22へ入力する。第1の実施形態と同様、補正度は、スペクトルやピッチなどの音声特徴量、及びその時間変化度に関連した値になる。ただし、本実施形態では、継続長補正度計算部242が言語情報だけでなく、仮ピッチパタンや音声波形パラメータに基づいて、音声特徴量、及び音声特徴量の時間変化度を推定し、補正度に反映する点で第1の実施形態と異

10

20

30

40

50

なる。

【0078】

継続長補正度計算部242は、まず、言語情報を用いて補正度を計算する。次に、継続長補正度計算部242は、仮ピッチパターンおよび音声波形パラメータに基づいて詳細化した補正度を計算する。このように、補正度を計算することで、音声特徴量の推定に利用される情報量が増加する。そのため、第1の実施形態に比べ、より正確かつ詳細に音声特徴量を推定することが可能になる。なお、継続長補正度計算部242が言語情報を用いて最初に計算した補正度は、その後、仮ピッチパターンおよび音声波形パラメータに基づいて詳細化されることから、最初に計算された補正度は、補正度の概略とすることもできる。

【0079】

上述の通り、本実施形態では、第1の実施形態と同様に、音声特徴量の時間変化度を推定し、その推定結果を補正度に反映している。以下、継続長補正度計算部242が補正度を計算する方法を、さらに説明する。

【0080】

図4は、言語情報をもとに算出された各状態における補正度の例を示す説明図である。図4に例示する10の状態のうち、前半5つは、子音部を示す音素の状態を表し、後半5つは母音部を示す音素の状態を表す。すなわち、1つの音素あたりの状態数は5であると仮定する。また、補正度は、縦上方向に延びるほど高いことを示す。以下の説明では、図4に例示するように、言語情報を用いて求めた補正度が、子音内部では均等であり、母音部では中心から両端にかけて小さくなっているものと仮定する。

【0081】

図5は、母音部における仮ピッチパターンに基づいて計算された補正度の例を示す説明図である。母音部の仮ピッチパターンが図5における(b1)のような形状をしていた場合、全体的にピッチパターンの変化度が小さいことがわかる。そのため、継続長補正度計算部242は、母音部の補正度を全般的に大きくする。具体的には、図4に例示する補正度を、最終的には、図5における(b2)のような補正度にする。

【0082】

また、図6は、母音部における他の仮ピッチパターンに基づいて計算された補正度の例を示す説明図である。母音部の仮ピッチパターンが図6における(c1)のような形状をしていた場合、ピッチパターンの変化度は母音前半から中心にかけては小さく、母音後半は大きいことがわかる。そのため、継続長補正度計算部242は、母音前半から中心の補正度を大きく、後半は小さくする。具体的には、図4に例示する補正度を、最終的には、図6における(c2)のような補正度にする。

【0083】

図7は、母音部における音声波形パラメータに基づいて計算された補正度の例を示す説明図である。母音部の音声波形パラメータが図7における(b1)のような形状をしていた場合、全体的に音声波形パラメータの変化度が小さいことがわかる。そのため、継続長補正度計算部242は、母音部の補正度を全般的に大きくし、図4に例示する補正度を、図7における(b2)のような補正度にする。

【0084】

また、図8は、母音部における他の音声波形パラメータに基づいて計算された補正度の例を示す説明図である。母音部の音声波形パラメータが図8における(c1)のような形状をしていた場合、音声波形パラメータの変化度は母音前半から中心にかけては小さく、母音後半は大きいことがわかる。そのため、継続長補正度計算部242は、母音前半から中心の補正度を大きく、後半は小さくし、図4に例示する補正度を、図8における(c2)のような補正度にする。

【0085】

なお、図7及び図8では、音声波形パラメータを一次元で例示しているが、実際には音声波形パラメータは多次元ベクトルであることが多い。この場合、継続長補正度計算部242は、各フレームごとに平均値や総和を計算し、一次元の値に変換した値を補正に用い

10

20

30

40

50

ればよい。

【0086】

言語処理部1と、韻律生成部2（より詳しくは、状態継続長生成部21と、状態継続長補正部22と、音素継続長計算部23と、継続長補正度計算部242と、仮ピッチパターン生成部28と、音声波形パラメータ生成部29と、ピッチパターン生成部3）と、素片選択部4と、波形生成部5とは、プログラム（音声合成プログラム）に従って動作するコンピュータのCPUによって実現される。また、言語処理部1と、韻律生成部2（より詳しくは、状態継続長生成部21と、状態継続長補正部22と、音素継続長計算部23と、継続長補正度計算部242と、仮ピッチパターン生成部28と、音声波形パラメータ生成部29と、ピッチパターン生成部3）と、素片選択部4と、波形生成部5とは、それぞれが専用のハードウェアで実現されていてもよい。

10

【0087】

次に、本実施形態における音声合成装置の動作を説明する。図9は、第2の実施形態における音声合成装置の動作の例を示すフローチャートである。まず、言語処理部1は、入力されたテキストから言語情報を生成する（ステップS1）。状態継続長生成部21は、言語情報とモデルパラメータとをもとに状態継続長を生成する（ステップS2）。

【0088】

また、仮ピッチパターン生成部28は、言語情報と状態継続長とをもとに、仮ピッチパターンを生成する（ステップS11）。さらに、音声波形パラメータ生成部29は、言語情報と、状態継続長とをもとに、音声波形パラメータを生成する（ステップS12）。そして、継続長補正度計算部242は、言語情報と仮ピッチパターンと音声波形パラメータとに基づいて、継続長補正度を計算する（ステップS13）。

20

【0089】

以降、状態継続長補正部22が状態継続長を補正し、波形生成部5が合成音声を生成するまでの処理は、図2におけるステップS4～ステップS8までの処理と同様である。

【0090】

以上のように、本実施形態によれば、仮ピッチパターン生成部28が、言語情報と状態継続長とをもとに、仮ピッチパターンを生成し、音声波形パラメータ生成部29が、言語情報と、状態継続長とをもとに、音声波形パラメータを生成する。そして、継続長補正度計算部242が、言語情報と仮ピッチパターンと音声波形パラメータとに基づいて、継続長補正度を計算する。

30

【0091】

すなわち、本実施形態によれば、言語情報のほかにピッチパターンや音声波形パラメータを使用して状態長補正度の計算が行われる。そのため、第1の実施形態における音声合成装置よりも、より適切な継続長補正を計算することが可能になる。この結果、第1の実施形態における音声合成装置よりも、より発話リズムの自然性が高く、聞き取り易い合成音声を生成できる。

【0092】

実施形態3.

図10は、本発明の第3の実施形態における音声合成装置の例を示すブロック図である。なお、第1の実施形態と同様の構成については、図1と同一の符号を付し、説明を省略する。本実施形態における音声合成装置は、言語処理部1と、韻律生成部2と、音声波形パラメータ生成部42と、波形生成部52とを備えている。また、韻律生成部2は、状態継続長生成部21と、状態継続長補正部22と、継続長補正度計算部24と、モデルパラメータ記憶部25と、ピッチパターン生成部3とを備えている。

40

【0093】

すなわち、図10に例示する音声合成装置は、音素継続長計算部23が省略され、素片選択部4が音声波形パラメータ生成部42に置き換わり、波形生成部5が波形生成部52に置き換わっている点において、第1の実施形態と異なる。

【0094】

50

音声波形パラメータ生成部 4 2 は、言語処理部 1 から入力された言語情報と、状態継続長補正部 2 2 から入力された状態継続長とをともに、音声波形パラメータを生成し、波形生成部 5 2 に入力する。音声波形パラメータには、スペクトル情報が用いられる。スペクトル情報として、例えば、ケプストラムなどが挙げられる。音声波形パラメータ生成部 4 2 が音声波形パラメータを生成する方法は、音声波形パラメータ生成部 2 9 が音声波形パラメータを生成する方法と同様である。

【0095】

波形生成部 5 2 は、ピッチパターン生成部 3 から入力されたピッチパターンと、音声波形パラメータ生成部 4 2 から入力された音声波形パラメータとをともに、合成音声波形を生成する。波形生成部 5 2 は、例えば、非特許文献 1 に記載された M L S A (m e l l o g s p e c t r u m a p p r o x i m a t i o n) フィルタにより合成音声波形を生成してもよい。ただし、波形生成部 5 2 が合成音声波形を生成する方法は M L S A フィルタを用いる方法に限定されない。

10

【0096】

言語処理部 1 と、韻律生成部 2 (より詳しくは、状態継続長生成部 2 1 と、状態継続長補正部 2 2 と、継続長補正度計算部 2 4 と、ピッチパターン生成部 3) と、音声波形パラメータ生成部 4 2 と、波形生成部 5 2 とは、プログラム (音声合成プログラム) に従って動作するコンピュータの C P U によって実現される。また、言語処理部 1 と、韻律生成部 2 (より詳しくは、状態継続長生成部 2 1 と、状態継続長補正部 2 2 と、継続長補正度計算部 2 4 と、ピッチパターン生成部 3) と、音声波形パラメータ生成部 4 2 と、波形生成部 5 2 とは、それぞれが専用のハードウェアで実現されていてもよい。

20

【0097】

次に、本実施形態における音声合成装置の動作を説明する。図 1 1 は、第 3 の実施形態における音声合成装置の動作の例を示すフローチャートである。テキストが言語処理部 1 に入力され、状態継続長補正部 2 2 が状態継続長を補正するまでの処理、およびピッチパターン生成部 3 がピッチパターンを生成する処理は、図 2 におけるステップ S 1 ~ ステップ S 4、および、ステップ S 6 と同様である。音声波形パラメータ生成部 4 2 は、言語情報と補正された状態継続長とをともに、音声波形パラメータを生成する (ステップ S 2 1)。そして、波形生成部 5 2 は、ピッチパターンと音声波形パラメータとをともに、合成音声波形を生成する (ステップ S 2 2)。

30

【0098】

以上のように、本実施形態によれば、音声波形パラメータ生成部 4 2 が、言語情報と補正された状態継続長とをともに、音声波形パラメータを生成し、波形生成部 5 2 が、ピッチパターンと音声波形パラメータとをともに合成音声波形を生成する。すなわち、本実施形態では、第 1 の実施形態における音声合成装置とは異なり、音素継続長生成や素片選択を行わずに合成音声を生成している。つまり、一般的な H M M 音声合成のように、状態継続長を直接利用して音声波形パラメータを生成するような音声合成装置においても、発話リズムの自然性が高く、聞き取り易い音声合成を生成することが可能になる。

【0099】

次に、本発明による音声合成装置の最小構成の例を説明する。図 1 2 は、本発明による音声合成装置の最小構成の例を示すブロック図である。本発明による音声合成装置は、言語情報 (例えば、言語処理部 1 が入力されたテキストから解析した言語情報) と韻律情報のモデルパラメータ (例えば、状態継続長のモデルパラメータ) とをともに、隠れマルコフモデル (H M M) における各状態の継続長を示す状態継続長を生成する状態継続長生成手段 8 1 (例えば、状態継続長生成部 2 1) と、言語情報から音声特徴量 (例えば、スペクトル、ピッチ) を導出し、導出された音声特徴量をもとに、状態継続長を補正する度合いを表す指標である継続長補正度を計算する継続長補正度計算手段 8 2 (例えば、継続長補正度計算部 2 4) と、音韻の継続時間長を補正する補正比率を表わす音韻継続長補正パラメータと継続長補正度とに基づいて、状態継続長を補正する状態継続長補正手段 8 3 (例えば、状態継続長補正部 2 2) とを備えている。

40

50

【0100】

そのような構成により、発話リズムの自然性が高く、聞き取り易い合成音声を生成できる。

【0101】

また、継続長補正度計算手段82は、言語情報から導出される音声特徴量の時間変化度を推定し、推定した時間変化度をもとに継続長補正度を計算してもよい。このとき、継続長補正度計算手段82は、音声特徴量を示すスペクトルまたはピッチの時間変化度を言語情報から推定し、推定した時間変化度をもとに継続長補正度を計算してもよい。

【0102】

また、状態継続長補正手段83は、音声特徴量の時間的変化度が小さい状態における状態継続長ほど、その状態継続長の変化度をより大きくしてもよい。

10

【0103】

また、音声合成装置は、言語情報と状態継続長生成手段81が生成した状態継続長とをもとに、ピッチパターンを生成するピッチパターン生成手段（例えば、仮ピッチパターン生成部28）と、言語情報と状態継続長とをもとに、音声波形を表すパラメータである音声波形パラメータを生成する音声波形パラメータ生成手段（例えば、音声波形パラメータ生成部29）とを備えていてもよい。そして、継続長補正度計算手段82は、言語情報とピッチパターンと音声波形パラメータとに基づいて、継続長補正度を計算してもよい。そのような構成により、より発話リズムの自然性が高く、聞き取り易い合成音声を生成できる。

【0104】

20

また、言語情報と状態継続長補正手段83が補正した状態継続長とをもとに、音声波形を表すパラメータである音声波形パラメータを生成する音声波形パラメータ生成手段（音声波形パラメータ生成部42）と、ピッチパターンと音声波形パラメータとをもとに合成音声波形を生成する波形生成手段（例えば、波形生成部52）とを備えていてもよい。そのような構成により、一般的なHMM音声合成のように、状態継続長を直接利用して音声波形パラメータを生成するような音声合成装置においても、発話リズムの自然性が高く、聞き取り易い音声合成を生成することが可能になる。

【0105】

以上、実施形態及び実施例を参照して本願発明を説明したが、本願発明は、各実施形態で説明した音声合成装置及び音声合成方法に限定されるものではない。その構成および動作は、発明の趣旨を逸脱しない範囲で適宜変更することができる。

30

【0106】

この出願は、2010年9月6日に出願された日本特許出願2010-199229を基礎とする優先権を主張し、その開示の全てをここに取り込む。

【産業上の利用可能性】

【0107】

本発明は、テキストから音声を作成する音声合成装置に好適に適用される。

【符号の説明】

【0108】

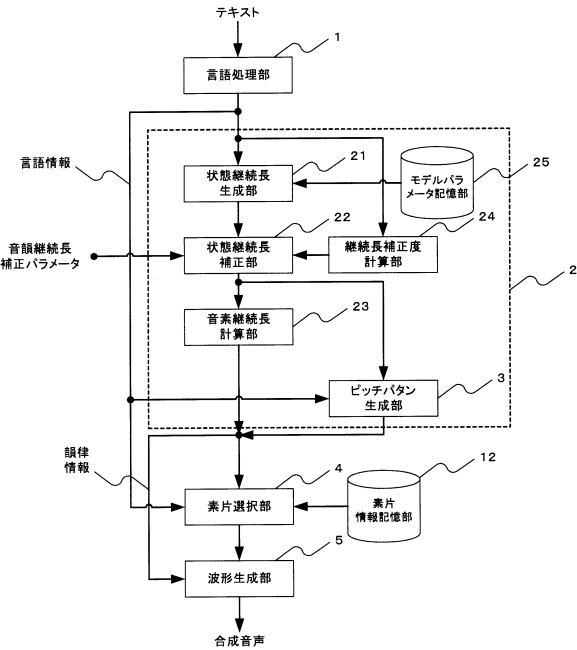
- 1 言語処理部
- 2 韻律生成部
- 3 ピッチパターン生成部
- 4 素片選択部
- 5, 52 波形生成部
- 12 素片情報記憶部
- 21 状態継続長生成部
- 22 状態継続長補正部
- 23 音素継続長計算部
- 24, 242 継続長補正度計算部
- 25 モデルパラメータ記憶部

40

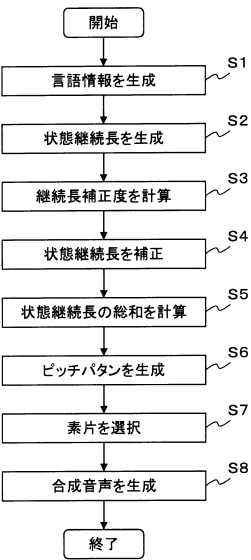
50

28 仮ピッチパターン生成部
29, 42 音声波形パラメータ生成部

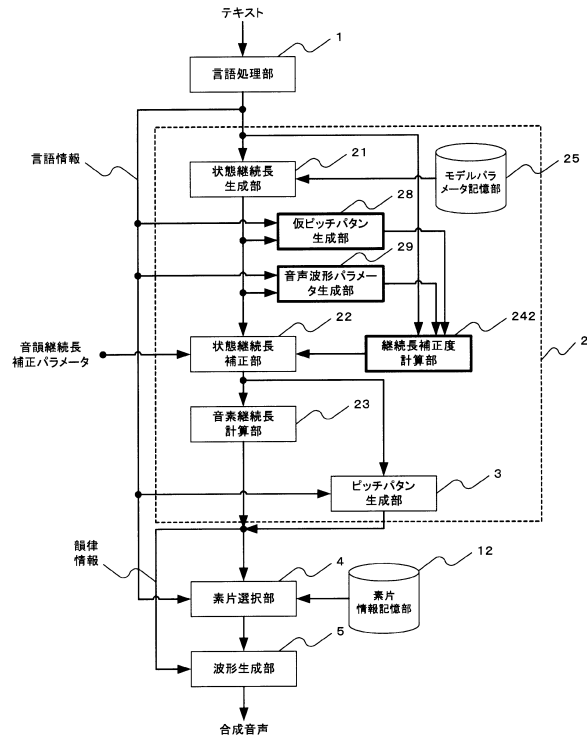
【図 1】



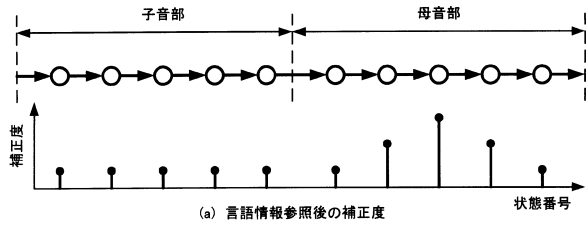
【図 2】



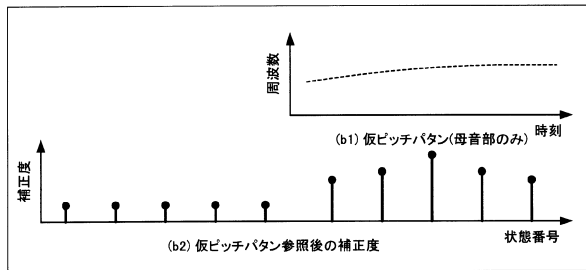
【図 3】



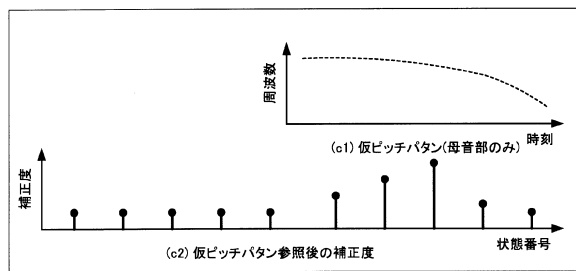
【図 4】



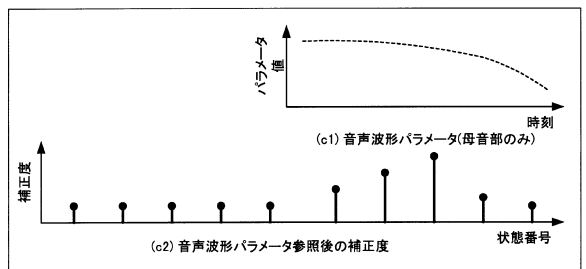
【図 5】



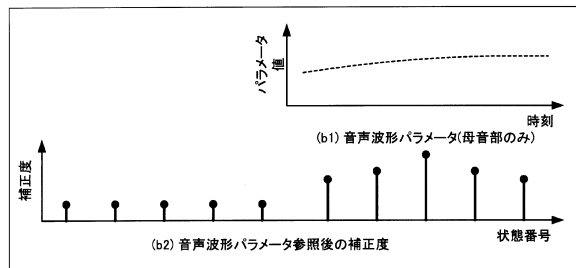
【図 6】



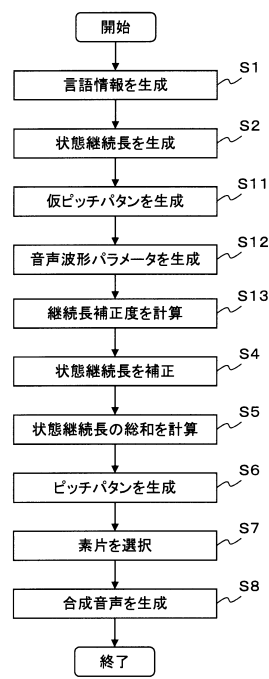
【図 8】



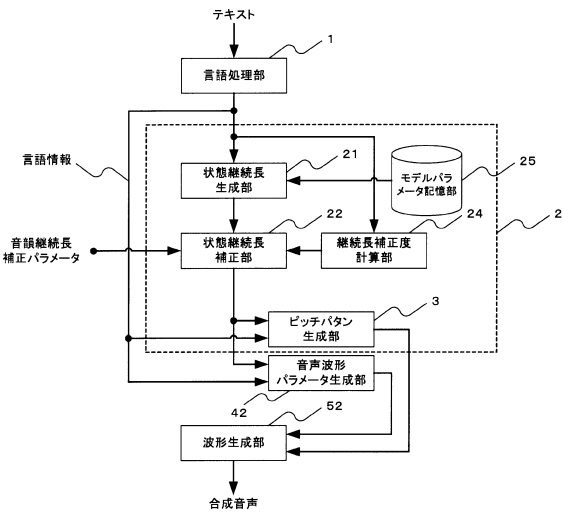
【図 7】



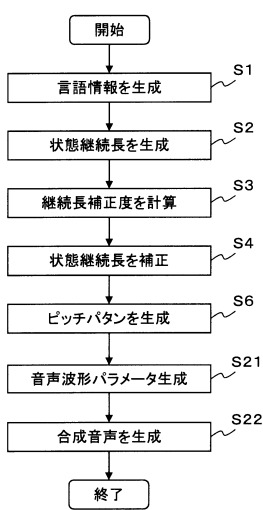
【図 9】



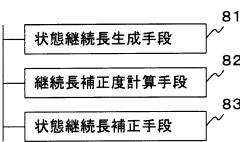
【図 10】



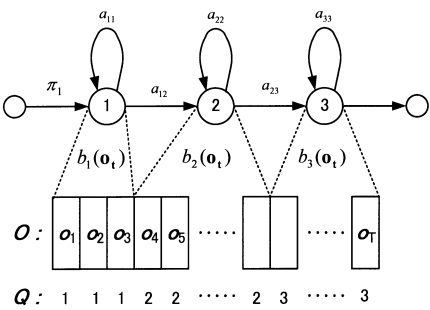
【図 11】



【図 12】



【図 13】



フロントページの続き

- (56)参考文献 特開2000-310996 (JP, A)
特開平04-170600 (JP, A)
特開2004-341259 (JP, A)
特開2002-244689 (JP, A)
国際公開第2012/063424 (WO, A1)
特開2011-28230 (JP, A)

- (58)調査した分野(Int.Cl., DB名)
G10L 13/06