

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5177135号
(P5177135)

(45) 発行日 平成25年4月3日(2013.4.3)

(24) 登録日 平成25年1月18日(2013.1.18)

(51) Int.Cl.		F I		
G 1 0 L 13/07	(2013.01)	G 1 0 L 13/06	2 4 0 C	
G 1 0 L 13/00	(2006.01)	G 1 0 L 13/00	1 0 0 K	

請求項の数 36 (全 22 頁)

(21) 出願番号	特願2009-514094 (P2009-514094)	(73) 特許権者	000004237
(86) (22) 出願日	平成20年4月28日(2008.4.28)		日本電気株式会社
(86) 国際出願番号	PCT/JP2008/058179		東京都港区芝五丁目7番1号
(87) 国際公開番号	W02008/139919	(74) 代理人	100080816
(87) 国際公開日	平成20年11月20日(2008.11.20)		弁理士 加藤 朝道
審査請求日	平成23年3月4日(2011.3.4)	(72) 発明者	加藤 正徳
(31) 優先権主張番号	特願2007-123422 (P2007-123422)		東京都港区芝五丁目7番1号 日本電気株式会社内
(32) 優先日	平成19年5月8日(2007.5.8)	(72) 発明者	三井 康行
(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
		(72) 発明者	近藤 玲史
			東京都港区芝五丁目7番1号 日本電気株式会社内

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】

複数の音声素片の中から、中心素片を1つ選択する中心素片選択部と、
前記中心素片を基に韻律情報を生成する韻律生成部と、
前記中心素片と前記韻律情報とを基に中心素片区間以外の素片である非中心素片を選択する非中心素片選択部と、
前記韻律情報と前記中心素片と前記非中心素片とを基に合成音声波形を生成する波形生成部と
を備えることを特徴とする音声合成装置。

【請求項2】

前記中心素片選択部は、
素片長が長い音声素片を優先的に中心素片として選択することを特徴とする請求項1に記載の音声合成装置。

【請求項3】

前記中心素片選択部は、
素片長が最長の音声素片を中心素片として選択することを特徴とする請求項1に記載の音声合成装置。

【請求項4】

複数の音声素片の中から、複数の中心素片を選択する中心素片選択部と、
前記中心素片を基に韻律情報を各中心素片に対して生成する韻律生成部と、

10

20

前記中心素片と前記韻律情報とを基に中心素片区間以外の素片である非中心素片を各中心素片に対して選択する非中心素片選択部と、

前記複数の中心素片の中から、一つの最適中心素片を選択する最適中心素片選択部と、

前記最適中心素片と、最適中心素片を基に生成された韻律情報と、最適中心素片を基に選択された非中心素片とを基に、合成音声波形を生成する波形生成部と

を備えることを特徴とする音声合成装置。

【請求項 5】

前記中心素片選択部は、

素片長が長い音声素片を優先的に中心素片として選択する

ことを特徴とする請求項 4 に記載の音声合成装置。

10

【請求項 6】

前記中心素片選択部は、

前記複数の音声素片のうち素片長が長いものから順に中心素片として選択する

ことを特徴とする請求項 4 に記載の音声合成装置。

【請求項 7】

前記中心素片選択部は、

中心素片として選択された音声素片の部分素片については、中心素片として選択しない

ことを特徴とする請求項 6 に記載の音声合成装置。

【請求項 8】

前記最適中心素片選択部は、

前記非中心素片選択部の選択結果に応じて最適中心素片を選択する

ことを特徴とする請求項 4 乃至 7 の何れか 1 項に記載の音声合成装置。

20

【請求項 9】

前記最適中心素片選択部は、

前記非中心素片選択部で各中心素片毎に計算された素片選択コストに応じて最適中心素片を選択する

ことを特徴とする請求項 4 乃至 7 の何れか 1 項に記載の音声合成装置。

【請求項 10】

前記中心素片選択部は、

入力テキストの言語処理を行う言語処理部を有し、

前記言語処理の言語処理結果との適合度が高い複数の音声素片の中から、中心素片を選択する

ことを特徴とする請求項 1 乃至 9 の何れか 1 項に記載の音声合成装置。

30

【請求項 11】

前記中心素片選択部は、

前記言語処理結果を基に韻律情報を生成する韻律生成部を有し、

前記韻律情報を基に中心素片を選択する

ことを特徴とする請求項 10 に記載の音声合成装置。

【請求項 12】

前記中心素片選択部は、

前記言語処理結果を基に入力テキストに含まれる重要表現を抽出する重要表現抽出部を更に有し、

前記重要表現を基に中心素片を選択する

ことを特徴とする請求項 10 又は 11 に記載の音声合成装置。

40

【請求項 13】

音声合成装置における音声合成方法であって、

複数の音声素片の中から、中心素片を 1 つ選択する中心素片選択ステップと、

前記中心素片を基に韻律情報を生成する韻律生成ステップと、

前記中心素片と前記韻律情報とを基に中心素片区間以外の素片である非中心素片を選択する非中心素片選択ステップと、

50

前記韻律情報と前記中心素片と前記非中心素片とを基に合成音声波形を生成する波形生成ステップと、

を含むことを特徴とする音声合成方法。

【請求項 14】

前記中心素片選択ステップは、

素片長が長い音声素片を優先的に中心素片として選択することを特徴とする請求項 13 に記載の音声合成方法。

【請求項 15】

前記中心素片選択ステップは、

素片長が最長の音声素片を中心素片として選択することを特徴とする請求項 13 に記載の音声合成方法。

10

【請求項 16】

音声合成装置における音声合成方法であって、

複数の音声素片の中から、複数の中心素片を選択する中心素片選択ステップと、

前記中心素片を基に韻律情報を各中心素片に対して生成する韻律生成ステップと、

前記中心素片と前記韻律情報とを基に中心素片区間以外の素片である非中心素片を各中心素片に対して選択する非中心素片選択ステップと、

前記複数の中心素片の中から、一つの最適中心素片を選択する最適中心素片選択ステップと、

前記最適中心素片と、最適中心素片を基に生成された韻律情報と、最適中心素片を基に選択された非中心素片とを基に、合成音声波形を生成する波形生成ステップと、

20

を備えることを特徴とする音声合成方法。

【請求項 17】

前記中心素片選択ステップは、

素片長が長い音声素片を優先的に中心素片として選択することを特徴とする請求項 16 に記載の音声合成方法。

【請求項 18】

前記中心素片選択ステップは、

前記複数の音声素片のうち素片長が長いものから順に中心素片として選択することを特徴とする請求項 16 に記載の音声合成方法。

30

【請求項 19】

前記中心素片選択ステップは、

中心素片として選択された音声素片の部分素片については、中心素片として選択しないようにする

ことを特徴とする請求項 18 に記載の音声合成方法。

【請求項 20】

前記最適中心素片選択ステップは、

前記非中心素片選択部の選択結果に応じて最適中心素片を選択する

ことを特徴とする請求項 16 乃至 19 の何れか 1 項に記載の音声合成方法。

40

【請求項 21】

前記最適中心素片選択ステップは、

前記非中心素片選択部で各中心素片毎に計算された素片選択コストに応じて最適中心素片を選択する

ことを特徴とする請求項 16 乃至 19 の何れか 1 項に記載の音声合成方法。

【請求項 22】

前記中心素片選択ステップは、

入力テキストの言語処理を行う言語処理ステップを含み、

前記言語処理の言語処理結果との適合度が高い複数の音声素片の中から、中心素片を選択する

ことを特徴とする請求項 13 乃至 21 の何れか 1 項に記載の音声合成方法。

50

【請求項 23】

前記中心素片選択ステップは、
前記言語処理結果を基に韻律情報を生成する韻律生成ステップを含み、
前記韻律情報を基に中心素片を選択する
ことを特徴とする請求項 22 に記載の音声合成方法。

【請求項 24】

前記中心素片選択ステップは、
前記言語処理結果を基に入力テキストに含まれる重要表現を抽出する重要表現抽出ステップを更に含み、
前記重要表現を基に中心素片を選択する
ことを特徴とする請求項 22 又は 23 に記載の音声合成方法。

10

【請求項 25】

音声合成装置における音声合成プログラムであって、
コンピュータを、
複数の音声素片の中から、中心素片を 1 つ選択する中心素片選択部と、
前記中心素片を基に韻律情報を生成する韻律生成部と、
前記中心素片と前記韻律情報とを基に中心素片区間以外の素片である非中心素片を選択する非中心素片選択部と、
前記韻律情報と前記中心素片と前記非中心素片とを基に合成音声波形を生成する波形生成部として機能させる
ことを特徴とする音声合成プログラム。

20

【請求項 26】

前記中心素片選択部は、
素片長が長い音声素片を優先的に中心素片として選択する
ことを特徴とする請求項 25 に記載の音声合成プログラム。

【請求項 27】

前記中心素片選択部は、
素片長が最長の音声素片を中心素片として選択する
ことを特徴とする請求項 25 に記載の音声合成プログラム。

【請求項 28】

音声合成装置における音声合成プログラムであって、
コンピュータを、
複数の音声素片の中から、複数の中心素片を選択する中心素片選択部と、
前記中心素片を基に韻律情報を各中心素片に対して生成する韻律生成部と、
前記中心素片と前記韻律情報とを基に中心素片区間以外の素片である非中心素片を各中心素片に対して選択する非中心素片選択部と、
前記複数の中心素片の中から、一つの最適中心素片を選択する最適中心素片選択部と、
前記最適中心素片と、最適中心素片を基に生成された韻律情報と、最適中心素片を基に選択された非中心素片とを基に、合成音声波形を生成する波形生成部として機能させる
ことを特徴とする音声合成プログラム。

30

40

【請求項 29】

前記中心素片選択部は、
素片長が長い音声素片を優先的に中心素片として選択する
ことを特徴とする請求項 28 に記載の音声合成プログラム。

【請求項 30】

前記中心素片選択部は、
前記複数の音声素片のうち素片長が長いものから順に中心素片として選択する
ことを特徴とする請求項 28 に記載の音声合成プログラム。

【請求項 31】

前記中心素片選択部は、

50

中心素片として選択された音声素片の部分素片については、中心素片として選択しないようにする

ことを特徴とする請求項 30 に記載の音声合成プログラム。

【請求項 32】

前記最適中心素片選択部は、

前記非中心素片選択部の選択結果に応じて最適中心素片を選択する

ことを特徴とする請求項 28 乃至 31 の何れか 1 項に記載の音声合成プログラム。

【請求項 33】

前記最適中心素片選択部は、

前記非中心素片選択部で各中心素片毎に計算された素片選択コストに応じて最適中心素片を選択する

ことを特徴とする請求項 28 乃至 31 の何れか 1 項に記載の音声合成プログラム。

【請求項 34】

前記中心素片選択部は、

入力テキストの言語処理を行う言語処理部を有し、

前記言語処理の言語処理結果との適合度が高い複数の音声素片の中から、中心素片を選択する

ことを特徴とする請求項 25 乃至 33 の何れか 1 項に記載の音声合成プログラム。

【請求項 35】

前記中心素片選択部は、

前記言語処理結果を基に韻律情報を生成する韻律生成部を有し、

前記韻律情報を基に中心素片を選択する

ことを特徴とする請求項 34 に記載の音声合成プログラム。

【請求項 36】

前記中心素片選択部は、

前記言語処理結果を基に入力テキストに含まれる重要表現を抽出する重要表現抽出部を更に有し、

前記重要表現を基に中心素片を選択する

ことを特徴とする請求項 34 又は 35 に記載の音声合成プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

(関連出願についての記載)

本願は、先の日本特許出願 2007-123422 号(2007年5月8日出願)の優先権を主張するものであり、前記先の出願の全記載内容は、本書に引用をもって繰込み記載されているものとみなされる。

本発明は、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し、特に、テキストから音声を合成するための音声合成装置、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】

従来から、テキスト文を解析し、その文が示す音声情報から規則合成により合成音声を生成する音声合成装置が、種々開発されてきた。

【0003】

図9は、従来の一般的な規則合成型の音声合成装置の構成を示したブロック図である。

【0004】

このような構成を有する音声合成装置の構成と動作の詳細については、例えば非特許文献1乃至3と、特許文献1及び2に記載されている。

【0005】

図9に示した音声合成装置は、言語処理部10と、韻律生成部11と、素片選択部16

10

20

30

40

50

と、音声素片情報記憶部 15 と、韻律制御部 18 と波形接続部 19 を有する波形生成部 17 とを備えている。

【0006】

音声素片情報記憶部 15 は、音声合成単位ごとに生成された音声素片を記憶する音声素片記憶部 152 と、各音声素片の属性情報が記憶された付属情報記憶部 151 を有している。

【0007】

ここで、音声素片とは、合成音声の波形を生成するために使われる情報で、収録された自然音声波形から抽出されることが多い。音声素片の例としては、合成単位毎に切り出された音声波形そのものや、線形予測分析パラメータ、ケプストラム係数などが挙げられる。

10

【0008】

また、音声素片の属性情報とは、各音声素片の抽出元である自然音声の音素環境や、ピッチ周波数、振幅、継続時間情報等の音韻情報や韻律情報のことである。

【0009】

従来の音声合成装置では、音声合成単位には、音素、CV、CVC、VCV(Vは母音、Cは子音)などが用いられることが多い。この音声素片の長さや合成単位の詳細については、非特許文献 1 と 3 に記述されている。

【0010】

言語処理部 10 は、入力されたテキスト文に対して形態素解析や構文解析、読み付け等の分析を行い、音素記号などの「読み」を表す記号列と、形態素の品詞、活用、アクセント型などを言語処理結果として韻律生成部 11 と素片選択部 16 に出力する。

20

【0011】

韻律生成部 11 は、言語処理部 10 から出力された言語処理結果を基に、合成音声の韻律情報(ピッチ、時間長、パワーなどに関する情報)を生成し、素片選択部 16 と韻律制御部 18 に出力する。

【0012】

素片選択部 16 は、言語処理結果と生成された韻律情報に関して適合度が高い音声素片を、音声素片情報記憶部 15 に記憶されている音声素片の中から選択し、選択した音声素片の付属情報と併せて韻律制御部 18 に出力する。

30

【0013】

韻律制御部 18 は、選択された音声素片から、韻律生成部 11 で生成した韻律に近い韻律を有する波形を生成し、波形接続部 19 に出力する。

【0014】

波形接続部 19 は、韻律制御部 18 から出力された音声素片を接続し、合成音声として出力する。

【0015】

素片選択部 16 は、入力された言語処理結果と韻律情報から、目標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。)を所定の合成単位ごとに求める。

【0016】

40

目標素片環境に含まれる情報には、該当・先行・後続の各音素名、ストレスの有無、アクセント核からの距離、合成単位のピッチ周波数やパワー、単位の継続時間長、ケプストラム、MFC C(Mel Frequency Cepstral Coefficients)、及びこれらの量(単位時間あたりの変化量)などが挙げられる。

【0017】

次に、目標素片環境が与えられると、素片選択部 16 は、音声素片情報記憶部 15 の中から目標素片環境により指定される特定の情報(主に該当音素)に合致する音声素片を複数選択する。選択された音声素片は、合成に用いる音声素片の候補となる。

【0018】

そして、選択された候補素片に対して、合成に用いる音声素片としての適切度を示す指

50

標である「コスト」を計算する。

【0019】

高音質な合成音声を生成することを目標としているため、コストが小さい、即ち適切度が高いと、合成音の音質は高くなる。

【0020】

従って、コストは、合成音声の音質の劣化度を推定するための指標であると言える。

【0021】

ここで、素片選択部16で計算されるコストには、単位コストと接続コストがある。

【0022】

単位コストは、候補素片を目標素片環境の基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。

10

【0023】

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。

【0024】

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。

【0025】

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、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、及びこれらの量などが用いられる。

20

【0026】

なお、ある2つの素片が元音声波形上で連続している場合には、その素片間の素片環境が完全に連続なので、接続コストの値はゼロになる。

【0027】

また、合成単位長の素片が元音声波形上で連続している場合には、その連続した素片を「素片長が長い素片」と表現する。

【0028】

従って、連続回数が多いほど、素片長は長くなると言える。一方、最も短い素片長は、合成単位の長さに対応する。

30

【0029】

接続コストと単位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。

【0030】

コスト最小化により求めた素片を、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ことから最適素片と呼ぶ。

【0031】

素片選択部16は、全合成単位を対象にそれぞれの最適素片を求めると、最終的に最適素片の系列(最適素片系列)を素片選択結果として韻律制御部18に出力する。

【0032】

素片選択部16では、上記のように単位コストが小さい音声素片が選択されることになっている。

40

【0033】

しかしながら、目標韻律(目標素片環境に含まれる韻律情報)に近い韻律を有する音声素片が選択されることになるが、目標韻律と同等の韻律を有する音声素片が選択されることは希である。

【0034】

従って、一般的には、素片選択後に、韻律制御部18において音声素片波形に処理を加えて、音声素片の韻律が目標韻律に一致するように補正する。

【0035】

50

音声素片の韻律を補正する方法としては、例えば特許文献4により公開されている分析手法を用いた方法が挙げられる。

【0036】

この特許文献4の分析手法によれば、元音声波形のスペクトル包絡を表す複素ケプストラムを求め、複素ケプストラムを表現するフィルタを所望のピッチ周波数に相当する時間間隔で駆動することによって、所望のピッチ周波数を持つ音声波形を再構成することができる。

【0037】

また他には、非特許文献4に記載されているPSOLA方式が挙げられる。

【0038】

但し、韻律の補正処理は、合成音声の低下を招く要因となる。特に、ピッチ周波数の変更が音質低下に与える影響は大きく、変更量が大きければ音質低下度も大きくなる。

【0039】

このため、韻律の補正処理に伴う音質劣化が十分小さくなるような基準(単位コスト重視)で単位選択を行うと、素片の接続歪みが目立つようになる。

【0040】

一方、接続歪みが小さくなるような基準(接続コスト重視)で素片選択を行うと、韻律制御に伴う音質劣化が目立つようになる。

【0041】

そこで、接続歪みと韻律制御に伴う音質劣化を同時に防ぐ方法として、様々な韻律情報を用意して単位選択を行い、音質劣化が最も小さくなるような韻律と単位選択結果の組み合わせを選択する方法が検討されている。

【0042】

例えば、特許文献3では、生成されたピッチパタンの周波数方向への平行移動と、平行移動後のピッチパターンを目標とした単位選択コストの計算を繰り返し、単位選択コストが最小となる平行移動量と単位選択結果を求める方法が提案されている。

【0043】

また、非特許文献5では、接続歪が小さくなる素片の組み合わせをまず先に求め、その中から目標韻律に最も適した単位を選択する方法が提案されている。

【0044】

また、非特許文献6では、目標韻律との類似度最大化と接続歪最小化を基準に素片を選択し、韻律制御を行わずに合成音声を生成することにより、韻律制御に伴う音質劣化を防止しつつ、接続歪みを低減する方法が提案されている。

【特許文献1】特開2005-91551号公報

【特許文献2】特開2006-84854号公報

【特許文献3】特開2004-138728号公報

【特許文献4】特許第2812184号公報

【非特許文献1】Huang, Acero, Hon: "Spoken Language Processing", Prentice Hall, pp. 689-836, 2001.

【非特許文献2】石川: "音声合成のための韻律制御の基礎", 電子情報通信学会技術研究報告, Vol. 100, No. 392, pp. 27-34, 2000.

【非特許文献3】阿部: "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.

【非特許文献4】Moulines, Charapentier: "Pitch-Synchronous Waveform Processing Techniques For Text-To-Speech Synthesis Using Diphones", Speech Communication 9, pp. 435-467, 1990.

【非特許文献5】Segi, Takagi, Ito: "A CONCATENATIVE SPEECH SYNTHESIS METHOD USING CONTEXT DEPENDENT PHONEME SEQUENCES WITH VARIABLE LENGTH AS SEARCH UNITS", Proceedings of 5th ISCA Speech Synthesis Workshop, pp. 115-120, 2004

【非特許文献6】Kawai, Toda, Ni, Tsuzaki, Tokuda: "XIMERA: A NEW TTS FROM ATR B

10

20

30

40

50

ASED ON CORPUS-BASED TECHNOLOGIES ”, Proceedings of 5 th ISCA Speech Synthesis Workshop, pp. 179 - 184, 2004

【発明の開示】

【発明が解決しようとする課題】

【0045】

以上の特許文献1～4及び非特許文献1～6の開示事項は、本書に引用をもって繰り込み記載されているものとする。以下に本発明による関連技術の分析を与える。

しかしながら、上記した特許文献及び非特許文献等に記載された従来の音声合成装置に関するものでは、下記のような問題点を有している。

【0046】

まず、特許文献3に記載された方法では、目標韻律のバリエーションが限定されているため、接続歪みが十分小さくなる素片の組み合わせを選ぶことが難しく、音質の大きな向上が望めない、という問題があった。

【0047】

また、非特許文献5の方法では、接続歪みの低減には有効な方法であるが、候補数不足のため韻律的に十分適した素片を見つけられず、韻律制御に伴う音質劣化が大きくなる、という問題があった。

【0048】

また、非特許文献6の方法では、接続歪みと韻律制御に伴う音質劣化の双方を十分に低減するには有効な方法だが、韻律制御を全く行わないため、合成音声の韻律が乱れやすい、という問題があった。

【0049】

従って、特許文献及び非特許文献に記載された音声合成装置に関するものでは、接続歪みと韻律制御に伴う音質劣化の双方を十分に低減しつつ、韻律の乱れが小さい合成音声を生成することが困難である、という問題があった。

【0050】

本発明は、上記問題点に鑑みてなされたものであり、接続歪みと韻律制御に伴う音質劣化の双方を十分に低減しつつ、韻律の乱れが小さい合成音声を生成することが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実現することを目的とする。

【課題を解決するための手段】

【0051】

本発明に係る音声合成装置は、複数の音声素片の中から、中心素片を1つ選択する中心素片選択部と、前記中心素片を基に韻律情報を生成する韻律生成部と、前記中心素片と前記韻律情報とを基に中心素片区間以外の素片である非中心素片を選択する非中心素片選択部と、前記韻律情報と前記中心素片と前記非中心素片を基に合成音声波形を生成する波形生成部とを備えることを特徴とする。

【0052】

本発明に係る音声合成方法は、音声合成装置における音声合成方法であって、複数の音声素片の中から、中心素片を1つ選択する中心素片選択ステップと、前記中心素片を基に韻律情報を生成する韻律生成ステップと、前記中心素片と前記韻律情報とを基に中心素片区間以外の素片である非中心素片を選択する非中心素片選択ステップと、前記韻律情報と前記中心素片と前記非中心素片を基に合成音声波形を生成する波形生成ステップと、を含むことを特徴とする。

【0053】

本発明に係る音声合成プログラムは、音声合成装置における音声合成プログラムであって、コンピュータを、複数の音声素片の中から、中心素片を1つ選択する中心素片選択部と、前記中心素片を基に韻律情報を生成する韻律生成部と、前記中心素片と前記韻律情報とを基に中心素片区間以外の素片である非中心素片を選択する非中心素片選択部と、前記韻律情報と前記中心素片と前記非中心素片を基に合成音声波形を生成する波形生成部として機能させることを特徴とする。

10

20

30

40

50

【発明の効果】

【0054】

本発明によれば、音声合成装置は、任意の区間における韻律生成の元となる中心素片を先ず先に選択し、その中心素片を基に韻律情報を生成するので、中心素片の区間では接続歪みと韻律制御に伴う音質劣化の双方を十分に低減することができる。

【0055】

また、中心素片が選ばれなかった区間においては、中心素片に基づいて生成された韻律情報に適した素片を選択するため、合成音声の韻律の乱れを小さくできる。

【0056】

かくして、接続歪みと韻律制御に伴う音質劣化の双方を十分に低減しつつ、韻律の乱れが小さい合成音声を生成することが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実現できる。

10

【図面の簡単な説明】

【0057】

【図1】本発明に係る第1の実施の形態による音声合成装置の構成を示すブロック図である。

【図2】本発明に係る第1の実施の形態による動作を説明するためのフローチャートである。

【図3】本発明に係る第2の実施の形態による音声合成装置の構成を示すブロック図である。

20

【図4】本発明に係る第2の実施の形態による動作を説明するためのフローチャートである。

【図5】本発明に係る第3の実施の形態による音声合成装置の構成を示すブロック図である。

【図6】本発明に係る第3の実施の形態による動作を説明するためのフローチャートである。

【図7】本発明に係る第4の実施の形態による音声合成装置の構成を示すブロック図である。

【図8】本発明に係る第4の実施の形態による動作を説明するためのフローチャートである。

30

【図9】従来一般的な規則合成型の音声合成装置の一例を示した構成図である。

【符号の説明】

【0058】

- 10 言語処理部
- 11、21、511、512・・・51M 韻律生成部
- 15 音声素片情報記憶部
- 16 素片選択部
- 17 波形生成部
- 18、28 韻律制御部
- 19 波形接続部
- 22、32、42 中心素片選択部
- 26、561、562・・・56N 非中心素片選択部
- 27 波形生成部
- 31 第二韻律生成部
- 40 重要表現抽出部
- 52 中心素片候補選択部
- 151 付属情報記憶部
- 152 音声素片記憶部
- 500 中心素片グループ生成部
- 501 最適素片選択部

40

50

【発明を実施するための最良の形態】

【0059】

次に、本発明の実施の形態の構成について図面を参照して詳細に説明する。

【0060】

(1) 第1の実施の形態による音声合成装置

(1-1) 第1の実施の形態による音声合成装置の構成

図1は、本発明の第1の実施の形態による構成を示すブロック図である。

【0061】

図1に示す本実施の形態による構成では、先行技術で説明した従来のブロック図である図9と対比して、韻律生成部11、素片選択部16及び韻律制御部18の代わりに、韻律生成部21、中心素片選択部22、非中心素片選択部26及び韻律制御部28を備えていることを特徴としている。

10

【0062】

以下、この相違点を中心に、図1のブロック図を参照しながら、第1の実施の形態による音声合成装置の詳細な動作を説明する。

【0063】

(1-2) 第1の実施の形態による音声合成装置の動作

図2は、本発明の第1の実施の形態の動作を説明するためのフローチャートである。

【0064】

図2のフローチャートを参照すると、中心素片選択部22は、言語処理部10から供給された言語処理結果のうち任意の区間（例えばアクセント句や呼気段落など）について、音声素片情報記憶部15から供給される音声素片情報を基に中心素片を選択し、選択した中心素片の情報を韻律生成部21、非中心素片選択部26及び韻律制御部28に伝達する（ステップA1）。

20

【0065】

ここで、後段の韻律生成部21において韻律情報を生成するために用いられる素片を示すことから、一般的な素片と区別して該当区間における中心素片と呼ぶ。読みが同一の素片であれば、任意の素片を中心素片に用いることが可能であるが、高音質を達成する上で望ましい条件がいくつか存在する。

【0066】

そこで、本実施の形態では、言語処理結果との適合度が高いものの中から、最も長い素片を中心素片として選択する例について説明する。

30

【0067】

言語処理結果との適合度は、入力テキストと音声素片情報記憶部15に記憶されている音波形の発声内容のそれぞれの言語情報の一致度として定義される。

【0068】

具体的には、読みの他に、先行・後続の音素、アクセント句内における位置、アクセント核からの相対位置、品詞情報などの一致度が、適合度の指標となる。

【0069】

例えば、「走る」の「は」と「箸」の「は」を比較すると、後続の音素は一致しているが、「走る」と「箸」のアクセント型が異なっているので、アクセント核からの相対位置が異なる。

40

【0070】

一方、「走る」の「は」と「橋」の「は」を比較すると、後続音素とアクセント核からの相対位置とも一致している。

【0071】

従って、この例では、「走る」の「は」に対して、「橋」の「は」の方が「箸」の「は」よりも言語情報的に一致度が高いと言える。

【0072】

また、本実施の形態で説明する「素片の長さ」は、音声素片情報記憶部15に記憶され

50

ている元音声波形上における合成単位長の素片の連続数で定義される。

【 0 0 7 3 】

一般的には、音声素片は合成単位ごとに用いられるが、元音声波形上で連続した素片が使われることもある。

【 0 0 7 4 】

例えば、合成単位を音節として、「母」と「獅子」と「橋」という発声内容の元音声波形が音声素片情報記憶部 15 に記憶されているときに、「走る」という入力テキストが与えられた場合を想定する。

【 0 0 7 5 】

「走る」の「はし」を構成するには、「母」の「は」と「獅子」の「し」のそれぞれの素片を用いることが可能であるが、一方で「橋」の「は」と「し」の素片、すなわち「はし」の素片を用いることも可能である。素片の長さの定義から、「は」や「し」という素片の長さは1であり、「はし」という素片の長さは2である。

【 0 0 7 6 】

従って、素片の長さのみを中心素片の選択基準とした場合には、「走る」の入力テキストに対して「橋」の素片を選択することになる。

【 0 0 7 7 】

言語処理結果との適合度が高くなるほど、長い素片を選択することが困難になるので、言語処理結果との適合度は、中心素片の選択においては制約条件として利用することが望ましい。

【 0 0 7 8 】

また、最長であっても素片の長さが短い場合には、韻律生成の際に悪影響を与える(ピッチパターンならば著しく不自然なパターンの生成をもたらす)ことがある。

【 0 0 7 9 】

従って、長さ基準を満たす素片が存在しなかった場合には、中心素片を選択せずに、中心素片を選択しなかったことを韻律生成部 2 1 と非中心素片選択部 2 6 に通知する。

【 0 0 8 0 】

一方、最長の素片が複数出現した場合には、言語的に重要な箇所を含むものを選択する方法が有効である。

【 0 0 8 1 】

例えば、アクセント核の位置が同じものや、文末や呼気段落末ならばこれらを含むものを優先して選択する。

【 0 0 8 2 】

韻律生成部 2 1 は、言語処理部 1 0 から供給された言語処理結果と、中心素片選択部 2 2 から供給された中心素片を基に、韻律情報を生成し、非中心素片選択部 2 6 と韻律制御部 2 8 に伝達する(ステップ A 2)。

【 0 0 8 3 】

ここでは、中心素片が存在する区間(以下、これを中心素片区間と呼ぶ。)では、中心素片の韻律に類似した韻律が実現されるような韻律情報が生成される。韻律制御に伴う音質劣化を最小にするためにも、生成される韻律と中心素片の韻律が完全に一致していることが最も望ましい。

【 0 0 8 4 】

但し、文全体の韻律のバランスに配慮すると、完全一致が不適切な場合もある。

【 0 0 8 5 】

従って、韻律制御に伴う音質劣化が目立たない範囲で、ピッチパターンならば周波数方向の平行移動、時間長であれば時間方向の伸縮などの韻律変形を行う方法も有効である。

【 0 0 8 6 】

また、中心素片の韻律に類似したものを生成する目的が、中心素片の韻律制御に伴う音質劣化を小さくすることであるため、中心素片の韻律と大きく異なる韻律情報を生成することは避けたほうが良い。

10

20

30

40

50

【 0 0 8 7 】

一方、中心素片が存在しない区間(以下、これを非中心素片区間と呼ぶ。)では、言語処理結果を基に韻律情報を生成する。

【 0 0 8 8 】

この際に、中心素片区間と非中心素片区間のそれぞれの韻律が大きく異なると韻律の自然性が大きく損なわれるため、予め生成した中心素片区間の韻律に合わせて非中心素片区間の韻律情報を生成する必要がある。

【 0 0 8 9 】

非中心素片区間の韻律情報を生成する方法の例としては、先ず言語処理結果から中心素片区間を含めた韻律情報を生成し、次に中心素片区間の韻律情報を中心素片のものに差し替える方法が挙げられる。

10

【 0 0 9 0 】

単純な差し替えでは韻律的な不整合が生じるので、差し替えた後には調整処理が必要となる。

【 0 0 9 1 】

調整処理の例として、ピッチパタンの場合には、パタンが滑らかになるようにピッチパタンを変形する方法が挙げられる。

【 0 0 9 2 】

言語処理結果から韻律情報を生成する方法については、非特許文献 1 や 3 に記載されているような従来から広く一般的に用いられている方法を採用すれば良い。

20

【 0 0 9 3 】

もし、中心素片選択部 2 2 から中心素片を選択しなかったことを通知された場合には、図 9 の韻律生成部 1 1 と同様に言語処理結果のみから韻律情報を生成する。

【 0 0 9 4 】

以上の中心素片の選択と韻律情報の生成は、任意の区間ごとに行われる。本実施の形態では、この区間としてアクセント句を用いる例について説明する。

【 0 0 9 5 】

従って、非中心素片の選択に移行する前に、全てのアクセント句に対して、中心素片の選択(ステップ A 1)と韻律情報の生成(ステップ A 2)が完了したことを確認する(ステップ A 3)。

30

【 0 0 9 6 】

非中心素片選択部 2 6 は、言語処理部 1 0 から供給された言語処理結果と、韻律生成部 2 1 から供給された韻律情報と、中心素片選択部 2 2 から供給された中心素片情報を基に、非中心部区間の素片を選択し、韻律制御部 2 1 に伝達する(ステップ A 4)。

【 0 0 9 7 】

非中心素片の選択では、従来方式と同様に、単位コストと接続コストを計算し、両者が最小となる素片列を選択する。

【 0 0 9 8 】

単位コストの計算は、非中心素片区間、接続コストの計算は非中心素片区間内、及び中心素片区間と非中心素片区間の境界において実施する。

40

【 0 0 9 9 】

中心素片区間は、既に素片選択が済んでいるので、単位コスト及び接続コストの計算は不要である。

【 0 1 0 0 】

もし、中心素片選択部 2 2 から中心素片を選択しなかったことを通知された場合には、中心素片が選択されなかったアクセント句は、非中心素片区間のみで構成されると同等なので、該当区間すべてにおいて単位コスト及び接続コストを計算する。

【 0 1 0 1 】

韻律制御部 2 8 は、韻律生成部 2 1 から供給された韻律情報と、中心素片選択部 2 2 から供給された中心素片情報、及び非中心素片選択部 2 6 から供給された非中心素片情報を

50

基に、各素片の韻律を制御し、韻律が目標韻律に修正された素片を波形接続部 19 に供給する（ステップ A5）。

【0102】

韻律の制御は、中心素片・非中心素片の区別無く、従来方式と同様の方法で実施すれば良い。

【0103】

(1-3) 第1の実施の形態による音声合成装置の効果

本実施の形態によれば、音声合成装置は、素片長が長い素片を韻律生成の元となる中心素片として選択し、選択された中心素片に基づき韻律情報を生成する。

【0104】

そして、生成された韻律情報に適した素片を選択する。

【0105】

この結果、中心素片が選ばれた区間では、その素片を基に韻律情報を生成するため、韻律制御に伴う音質劣化が十分小さくなり、接続歪みも殆ど生じない。

【0106】

特に、この音声合成装置は、素片長が長くなるほど、接続歪みと音律制御に伴う音質劣化を飛躍的に低減することができる。

【0107】

一方、それ以外の区間、すなわち非中心素片区間では、中心素片に基づいて生成された韻律情報に適した素片を選択するため、合成音声の韻律の乱れを回避することができる。

【0108】

(2) 第2の実施の形態による音声合成装置

(2-1) 第2の実施の形態による音声合成装置の構成

図3は、本発明の第2の実施の形態の構成を示すブロック図である。

【0109】

図3に示す第2の実施の形態の構成は、図1で示した第1の実施の形態の中心素片選択部22が中心素片選択部32に置換されると共に、第二韻律生成部31を更に備えている。

【0110】

以下、この相違点を中心に、図3のブロック図を参照しながら、第2の実施の形態による音声合成装置の詳細な動作を説明する。

【0111】

(2-2) 第2の実施の形態による音声合成装置の動作

図4は、本発明の第2の実施の形態の動作を説明するためのフローチャートである。

【0112】

図4のフローチャートを参照すると、第二韻律生成部31は、言語処理部10から供給された言語処理結果を基に韻律情報を生成し、中心素片選択部32に伝達する(ステップB1)。

【0113】

第二韻律生成部31で生成される韻律情報は、中心素片の選択に利用されるため、韻律生成部21で生成される韻律情報と一致している必要は無い。

【0114】

最も基本的な生成方法は、図9の韻律生成部11と同様に韻律情報を生成し、そこから中心素片選択に用いる特徴量を抽出する方法である。

【0115】

例えば、ピッチパターンを生成した場合には、各アクセント句におけるピッチ周波数、及びアクセント句内の最高ピッチ周波数などを生成したピッチパターンから算出し、これらの特徴量との類似度を中心素片の選択基準に用いる方法が挙げられる。

【0116】

また、時間長を生成した場合には、平均発話速度を選択基準に用いる方法が挙げられる

10

20

30

40

50

。

【 0 1 1 7 】

中心素片選択部 3 2 は、言語処理部 1 0 から供給された言語処理結果と、音声素片情報記憶部 1 5 から供給された音声素片情報と、第二韻律生成部 3 1 から供給された韻律情報を基に中心素片を選択し、選択した中心素片の情報を韻律生成部 2 1、非中心素片選択部 2 6 及び韻律制御部 2 8 に伝達する（ステップ B 2）。

【 0 1 1 8 】

図 1 の中心素片選択部 2 2 とは異なり、中心素片選択部 3 2 は、言語処理結果との適合度と素片の長さの他に、韻律情報を用いて中心素片を選択する。

【 0 1 1 9 】

例えば、先ず、言語処理結果との適合度と素片の長さから、中心素片の候補となる素片を複数用意し、各候補の韻律情報との類似度を選択基準として、最適な中心素片を選択する。

【 0 1 2 0 】

そして、候補素片の最高ピッチ周波数と第二韻律生成部 3 1 から供給された最高ピッチ周波数の比率を選択基準の指標とする方法が挙げられる。

【 0 1 2 1 】

また、候補素片にアクセント句の始端が含まれる場合には、候補素片の始端のピッチ周波数と第二韻律生成部 3 1 から供給された始端ピッチ周波数の比率を選択基準の指標とする方法も有効である。

【 0 1 2 2 】

同様に、候補素片の平均時間長と第二韻律生成部 3 1 から供給された平均時間長の比率や差分を指標とすることも可能である。

【 0 1 2 3 】

また、韻律情報を選択基準に用いる場合には、言語処理結果との適合度と同様に、中心素片の選択においては制約条件として利用することが望ましい。

【 0 1 2 4 】

(2 - 3) 第 2 の実施の形態による音声合成装置の効果

本実施の形態によれば、音声合成装置は、言語処理結果と素片長の他に、韻律情報も中心素片の選択に利用している。

【 0 1 2 5 】

その結果、第 1 の実施の形態と比較して、韻律生成部で生成される韻律情報の品質が向上することになり、合成音声の韻律の乱れを低減することができる。

【 0 1 2 6 】

(3) 第 3 の実施の形態による音声合成装置

(3 - 1) 第 3 の実施の形態による音声合成装置の構成

図 5 は、本発明の第 3 の実施の形態の構成を示すブロック図である。

【 0 1 2 7 】

図 5 に示す第 3 の実施の形態の構成は、図 1 で示した第 1 の実施の形態の中心素片選択部 2 2 が中心素片選択部 4 2 に置換されると共に、重要表現抽出部 4 0 を更に備えている

。

【 0 1 2 8 】

以下、この相違点を中心に、図 5 のブロック図を参照しながら、第 3 の実施の形態による音声合成装置の詳細な動作を説明する。

【 0 1 2 9 】

(3 - 2) 第 3 の実施の形態による音声合成装置の動作

図 6 は、本発明の第 3 の実施の形態の動作を説明するためのフローチャートである。

【 0 1 3 0 】

図 6 のフローチャートを参照すると、重要表現抽出部 4 0 は、言語処理部 1 0 から供給された言語処理結果を基に、入力されたテキストの中でキーワードや入力テキストの印象

10

20

30

40

50

や意味を特徴づける表現を抽出し、中心素片選択部 4 2 に伝達する(ステップ C 1)。

【 0 1 3 1 】

そして、テキスト中に含まれる重要語や、テキストの内容を特徴付ける表現などを、言語処理結果から抽出する。

【 0 1 3 2 】

また、入力テキストを直接解析して、重要表現の抽出に活用することも有効である。

【 0 1 3 3 】

重要表現は、入力テキストの内容に応じて異なることが多い。

【 0 1 3 4 】

例えば、天気予報の内容ならば、「晴れ・曇り・雨」という天気を表す単語や、降水確率の値などが重要表現として挙げられる。

【 0 1 3 5 】

従って、入力テキストの意図や内容の推定を重要表現抽出部 4 0 において行えば、重要表現の抽出精度が向上する。

【 0 1 3 6 】

中心素片選択部 4 2 は、言語処理部 1 0 から供給された言語処理結果と、音声素片情報記憶部 1 5 から供給された音声素片情報と、重要表現抽出部 4 0 から供給された重要表現情報を基に中心素片を選択し、選択した中心素片の情報を韻律生成部 2 1、非中心素片選択部 2 6 及び韻律制御部 2 8 に伝達する(ステップ C 2)。

【 0 1 3 7 】

ここで、中心素片を検索する際に、重要表現に一致する素片が存在すれば、素片長が短くても中心素片として優先的に選択する。特に、合成音声の内容理解度を改善するためには、重要表現を優先的に中心素片にすることが有効である。

【 0 1 3 8 】

(3 - 3) 第 3 の実施の形態による音声合成装置の効果

本実施の形態によれば、音声合成装置は、言語処理結果と素片長の他に、入力テキスト中から抽出した重要表現を中心素片の選択に利用している。

【 0 1 3 9 】

その結果、第 1 の実施の形態と比較して、合成音声の発声内容の中で重要な単語や表現の箇所の音質が向上し、合成音声の内容理解度を向上させることができる。

【 0 1 4 0 】

(4) 第 4 の実施の形態の音声合成装置

(4 - 1) 第 4 の実施の形態による音声合成装置の構成

図 7 は、本発明の第 4 の実施の形態の構成を示すブロック図である。

【 0 1 4 1 】

図 7 に示す第 4 の実施の形態の構成は、図 1 で示した第 1 の実施の形態の中心素片選択部 2 2、韻律生成部 2 1、及び非中心素片選択部 2 6 が、中心素片候補選択部 5 2、韻律生成部 5 1 1、5 1 2・・・5 1 M、及び非中心素片選択部 5 6 1、5 6 2・・・5 6 N に置換されると共に、中心素片グループ生成部 5 0 0 と最適素片選択部 5 0 1 とを更に備えている。

【 0 1 4 2 】

以下、この相違点を中心に、図 7 のブロック図を参照しながら、第 4 の実施の形態による音声合成装置の詳細な動作を説明する。

【 0 1 4 3 】

(4 - 2) 第 4 の実施の形態による音声合成装置の動作

図 8 は、本発明の第 4 の実施の形態の動作を説明するためのフローチャートである。

【 0 1 4 4 】

図 8 のフローチャートを参照すると、中心素片候補選択部 5 2 は、言語処理部 1 0 から供給された言語処理結果と、音声素片情報記憶部 1 5 から供給される音声素片情報を基に、中心素片になりうる候補素片を複数選択し、韻律生成部 5 1 1、5 1 2・・・5 1 M に

10

20

30

40

50

伝達する(ステップD 1)。

【0145】

ここで、第1の実施の形態では、語処理結果との適合度を制約条件として、最長の素片を中心素片として選択していたが、本実施の形態では素片の長さを選択基準にしつつも、複数の中心素片候補を選択する。

【0146】

その際、候補数が事前に定めた値(本実施例ではM)を満足するまで、素片長が長いほうから順番に候補素片として選択する。

【0147】

但し、単純に長いほうから順番に候補素片を選び出すと、ある特定の素片の部分素片が候補の大多数を占めることになる場合がある。

【0148】

例えば、長さLの素片から、長さL-1の素片は2種類、長さL-2の素片は3種類を選択することができる。

【0149】

ここで、長さL-1の素片や長さL-2の素片のことを、長さLの素片の部分素片と呼ぶ。

【0150】

ある素片の部分素片の韻律(先ほどの例の場合では、長さL-1の素片や長さL-2の素片)は、どれも類似のものになる可能性が高いため、もし韻律的に好ましくない素片から数多くの部分素片が候補素片に採用された場合には、合成音声の品質に悪影響を与える可能性が高い。

【0151】

従って、韻律の異なる様々な素片を中心素片の候補とするためにも、部分素片の種類数は、ある程度制限することが望ましい。

【0152】

本実施の形態では、候補数がMに設定されているが、必ずしもMになるまで候補素片を選択する必要は無い。つまり、素片の長さが短すぎて中心素片としての基準を満たさない素片は、候補から除外する。

【0153】

韻律生成部511、512・・・51Mは、言語処理部10から供給された言語処理結果、中心素片候補選択部52から供給された中心素片を基に韻律情報を生成し、中心素片と韻律情報を中心素片グループ生成部500に伝達する(ステップD2)。

【0154】

韻律生成部511、512・・・51Mでは、各中心素片候補に対して、韻律情報をそれぞれ生成する。韻律情報の生成方法は、図1の韻律生成部21と同様である。

【0155】

以上の中心素片の選択と韻律情報の生成は、任意の区間ごとに行われる。本実施の形態では、この区間としてアクセント句を用いる例について説明する。

【0156】

従って、中心素片グループの生成に移行する前に、全てのアクセント句に対して、中心素片候補の選択(ステップD1)と韻律情報の生成(ステップD2)が完了したことを確認する(ステップD4)。

【0157】

中心素片グループ生成部500は、韻律生成部511、512・・・51Mから供給された韻律情報及び中心素片を基に中心素片グループを生成し、生成した各グループの韻律情報と中心素片を非中心素片選択部561、562・・・56Nに伝達する(ステップD5)。

【0158】

ここで、本実施の形態では、単位コストと接続コストの計算を呼気段落毎に行う例につ

10

20

30

40

50

いて説明する。この場合、非中心素片の選択は、単位コストと接続コストを計算する必要性から、アクセント句単位ではなく呼気段落単位で行われる。

【0159】

従って、本実施の形態のように、各アクセント句に複数の中心素片が候補に挙げられている場合には、呼気段落を形成する上で考えられうる中心素片の組み合わせが複数存在することになる。

【0160】

例えば、2つのアクセント句で構成される呼気段落において、第一アクセント句で中心素片候補が3つ、第二アクセント句で中心素片候補が2つ存在する場合には、中心素片候補の組み合わせ数は6通りになる。

【0161】

全ての中心素片候補の組み合わせ(この例では6通り)において素片選択を実施するために、中心素片グループ生成部500では、全ての中心素片の組み合わせを生成し、各組み合わせに対してグループ番号を付与して、韻律情報と中心素片と共に各非中心素片選択部に伝達する。

【0162】

Nの値は、全ての中心素片候補の組み合わせ数に相当し、呼気段落に含まれるアクセント句数と各アクセント句の中心素片候補数の値に応じて変化する。

【0163】

非中心素片選択部561、562・・・56Nは、言語処理部10から供給された言語処理結果、音声素片情報記憶部15から供給される音声素片情報、中心素片グループ生成部500から供給された各中心素片グループの韻律情報と中心素片を基に、非中心素片を選択し、各グループの韻律情報、中心素片、非中心素片、そして非中心素片を選択した際に得られた素片選択コストを、最適素片選択部501に伝達する(ステップD6)。

【0164】

コストの計算方法及び非中心区間の素片の選択方法は、図1の非中心素片選択部26と同じである。

【0165】

最適素片選択部501は、非中心素片選択部561、562・・・56Nから供給された各グループの素片選択コストを基に、最適な中心素片と非中心素片の組み合わせを選択し、韻律情報と共に韻律制御部28に伝達する(ステップD8)。

【0166】

素片選択コストが小さいほど合成音声の品質が高くなると考えられるので、素片選択コストが最小のグループの中心素片と非中心素片を、最適素片として選択する。

【0167】

(4-3)第4の実施の形態による音声合成装置の効果

本実施の形態によれば、音声合成装置は、中心素片の候補を複数選択して、それぞれの候補に対して、韻律情報を生成し非中心素片の選択を行う。そして、非中心素片の選択コストを基に、最適な中心素片及び非中心素片を選択する。

【0168】

すなわち、非中心素片の選択コストを中心素片の選択に利用しているという特徴がある。

【0169】

その結果、第1の実施の形態の場合と比較して、非中心素片区間の品質改善につながる中心素片を選択することが可能になり、合成音声全体の品質が向上する。

【0170】

(5)他の実施の形態について

本発明による実施の形態は、第1の実施の形態から第4の実施の形態で説明した音声合成装置に限定されるものではなく、その構成および動作は、発明の趣旨を逸脱しない範囲で、適宜に変更することができる。

10

20

30

40

50

【0171】

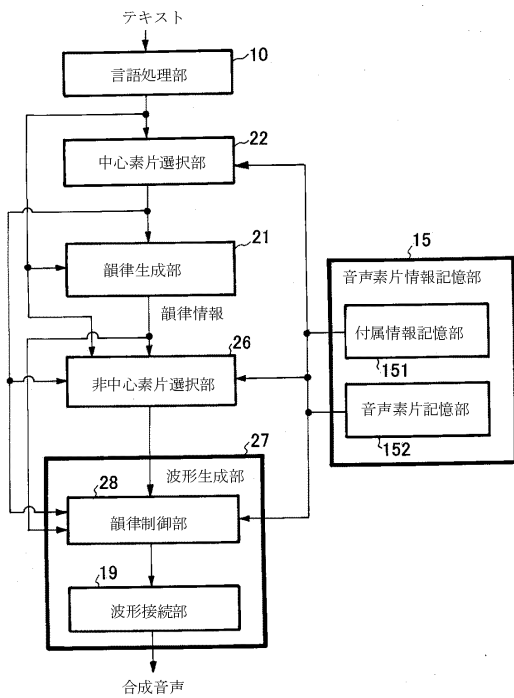
また、本発明による実施の形態は、発明の構成及び動作を中心に説明したが、本発明による実施の形態の機能または手順を、コンピュータが読み取り可能なプログラムによって実現・実行させるようにしても良い。

【0172】

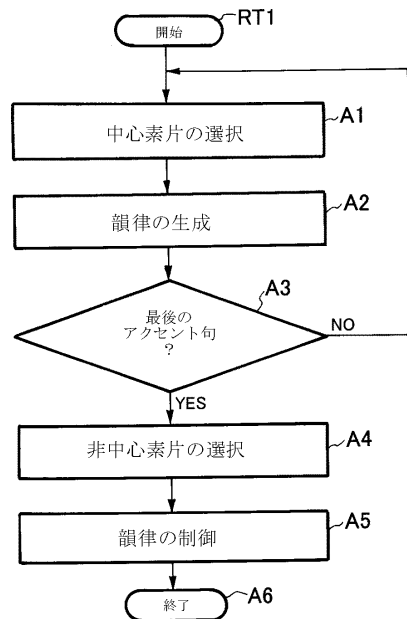
以上本発明を上記実施例に即して説明したが、本発明は、上記実施例にのみ限定されるものではなく、本願特許請求の範囲の各請求項の発明の範囲内で当業者であればなし得るであろう各種変形、修正を含むことは勿論である。

本発明の全開示（請求の範囲を含む）の枠内において、さらにその基本的技術思想に基づいて、実施形態ないし実施例の変更・調整が可能である。また、本発明の請求の範囲の枠内において種々の開示要素の多様な組み合わせないし選択が可能である。

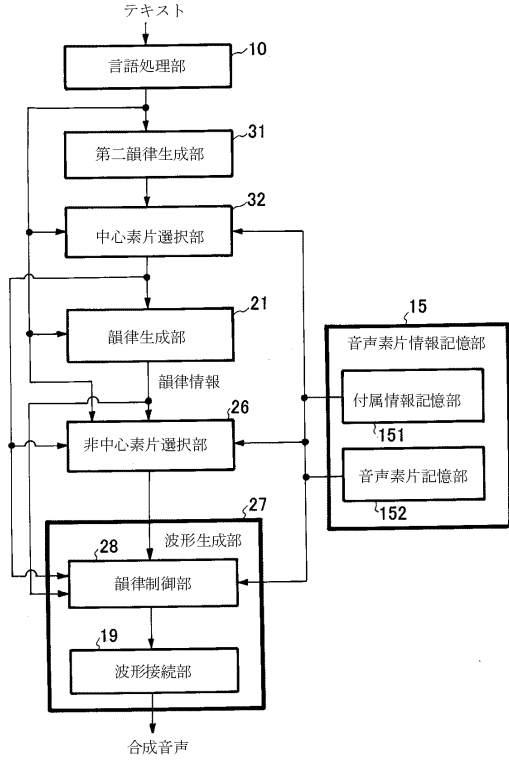
【図1】



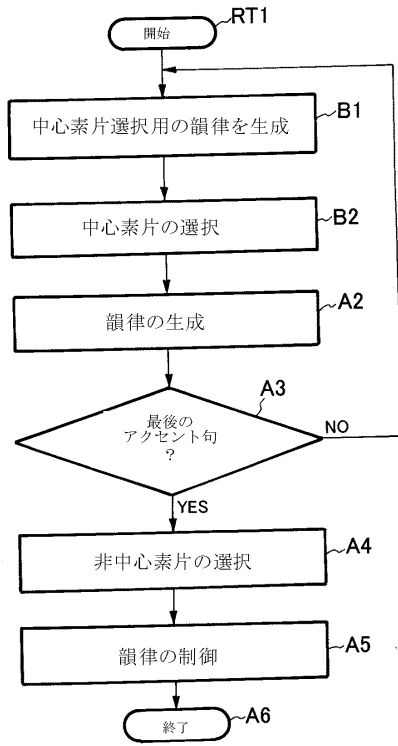
【図2】



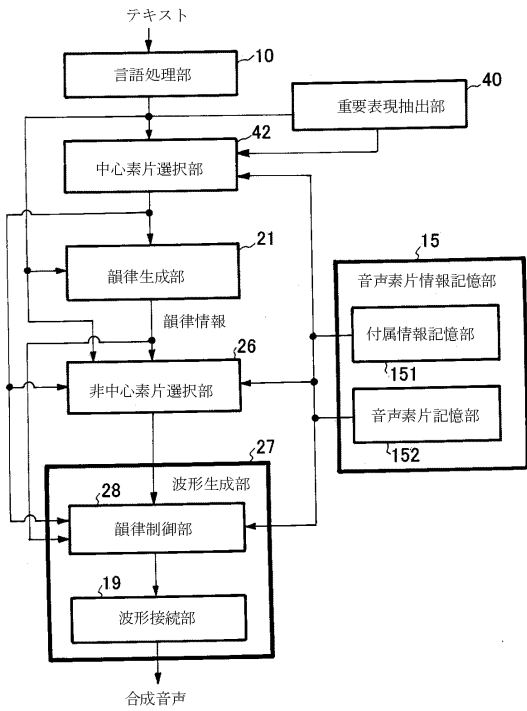
【図3】



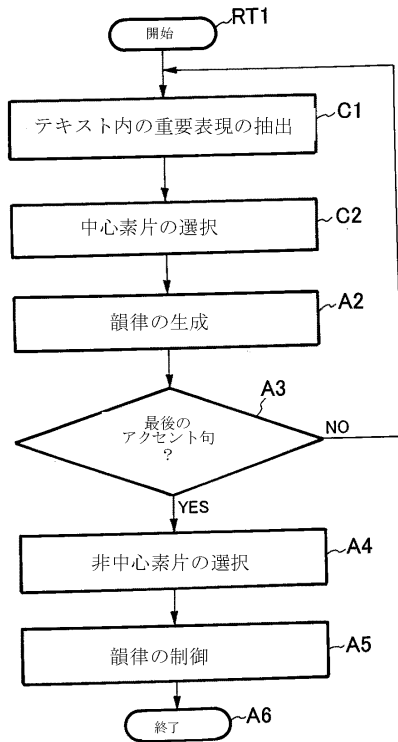
【図4】



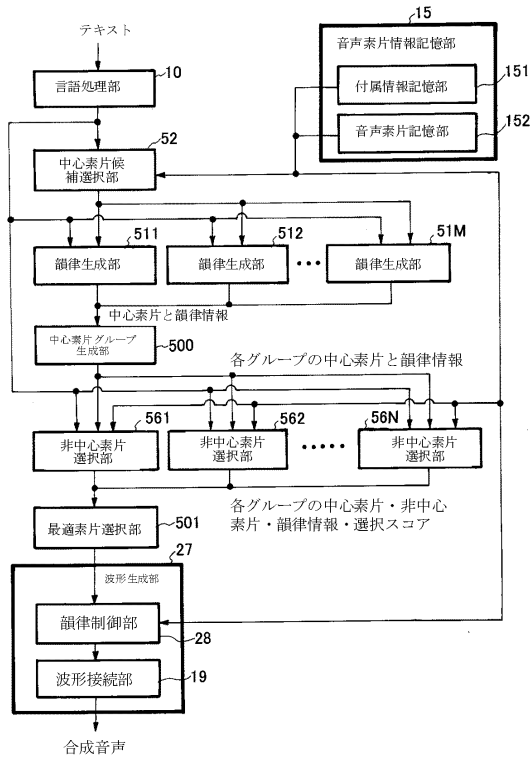
【図5】



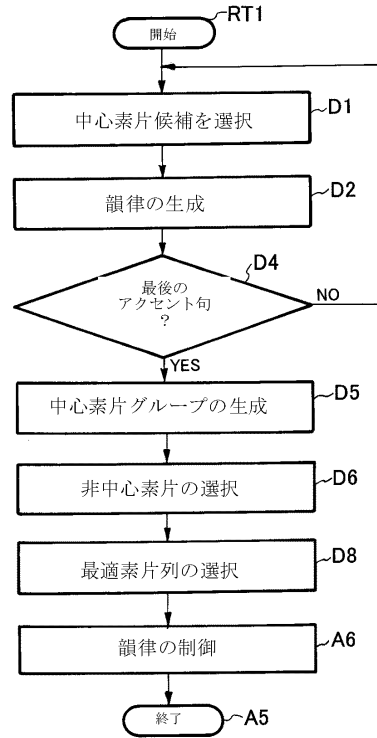
【図6】



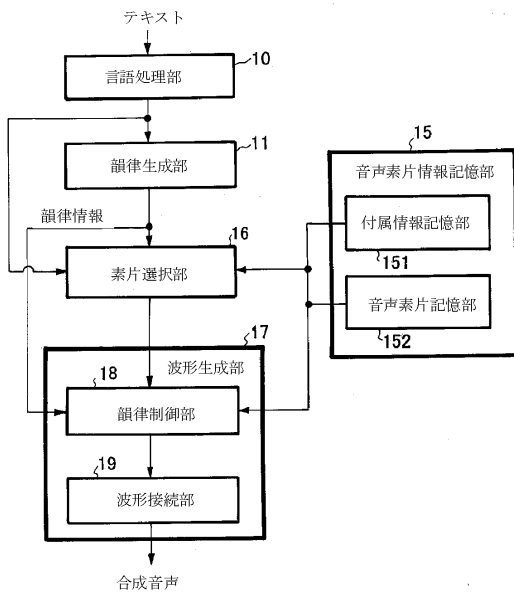
【図7】



【図8】



【図9】



フロントページの続き

審査官 五貫 昭一

- (56)参考文献 特開2001-117577(JP,A)
特開2003-271172(JP,A)
特開2005-265874(JP,A)
特開2006-84854(JP,A)
特開2006-276528(JP,A)
特開2006-337476(JP,A)
特開2008-15424(JP,A)
特開2008-20840(JP,A)
特開2008-33133(JP,A)
国際公開第2008/102710(WO,A1)

(58)調査した分野(Int.Cl., DB名)

G10L 13/06

G10L 13/00