(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第3622990号
(P3622990)
(45)発行日　平成17年2月23日(2005.2.23)　(24)登録日　平成16年12月3日(2004.12.3)

(51)Int.Cl.[7]　F I
G10L 13/08　G10L 3/00　H
G10L 13/06　G10L 5/04　F

請求項の数 12　(全 17 頁)

(21)出願番号　特願平5-228156
(22)出願日　平成5年8月19日(1993.8.19)
(65)公開番号　特開平7-56590
(43)公開日　平成7年3月3日(1995.3.3)
審査請求日　平成12年8月21日(2000.8.21)
審判番号　不服2003-3417(P2003-3417/J1)
審判請求日　平成15年3月3日(2003.3.3)

(73)特許権者　000002185
ソニー株式会社
東京都品川区北品川６丁目７番３５号
(74)代理人　100082740
弁理士　田辺　恵基
(72)発明者　山田　敬一
東京都品川区北品川６丁目７番３５号ソニー株式会社内
(72)発明者　及川　芳明
東京都品川区北品川６丁目７番３５号ソニー株式会社内

最終頁に続く

(54)【発明の名称】音声合成装置及び方法

(57)【特許請求の範囲】
【請求項１】
音韻記号と韻律記号とに基づいて所定の音韻規則及び韻律規則によつて韻律情報を設定する音声合成規則部と、音声波形データを必要フレーム数貯えた音声単位及び上記韻律情報に基づいて合成音を生成する音声合成部とを有する波形編集型の音声合成装置において、同じ音素の接続フレームを持つ複数の音声単位について、接続フレームにおける代表的なスペクトル包絡特性を求め、上記接続フレームにおけるスペクトル包絡特性が上記代表的なスペクトル包絡特性になるように、上記同じ音素の接続フレームを持つ音声単位のスペクトル包絡軌道を正規化して求めた音声単位を記憶する音声単位記憶部
を具えることを特徴とする音声合成装置。
【請求項２】
上記音声単位の上記スペクトル包絡軌道の正規化を、上記音声単位の前端及び又は後端の任意数のフレームに対して行うようにした
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】
上記音声単位の上記スペクトル包絡軌道の正規化を、上記音声単位内の有声部分に対して行うようにした
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】
上記音声単位の上記スペクトル包絡軌道の正規化を、上記音声単位内の前端及び又は後端

の接続フレームを含む音素に対して行うようにした
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５】
上記音声単位記憶部に貯えられている上記音声単位内の有声部分について、各フレームの上記音声波形データの位相特性を最小位相特性に整えるようにした
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６】
上記音声単位記憶部に貯えられている上記音声単位内の有声部分について、各フレームの上記音声波形データの位相特性を最大位相特性に整えるようにした
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７】
音韻記号と韻律記号とに基づいて所定の音韻規則及び韻律規則によつて韻律情報を設定すると共に、音声波形データを必要フレーム数貯えた音声単位及び上記韻律情報に基づいて合成音を生成する波形編集型の音声合成方法において、
同じ音素の接続フレームを持つ複数の音声単位について、接続フレームにおける代表的なスペクトル包絡特性を求め、上記接続フレームにおけるスペクトル包絡特性が上記代表的なスペクトル包絡特性になるように、上記同じ音素の接続フレームを持つ音声単位のスペクトル包絡軌道を正規化して求めた音声単位を記憶する
ことを特徴とする音声合成方法。
【請求項８】
上記音声単位の上記スペクトル包絡軌道の正規化を、上記音声単位の前端及び又は後端の任意数のフレームに対して行うようにした
ことを特徴とする請求項７に記載の音声合成方法。
【請求項９】
上記音声単位の上記スペクトル包絡軌道の正規化を、上記音声単位内の有声部分に対して行うようにした
ことを特徴とする請求項７に記載の音声合成方法。
【請求項１０】
上記音声単位の上記スペクトル包絡軌道の正規化を、上記音声単位内の前端及び又は後端の接続フレームを含む音素に対して行うようにした
ことを特徴とする請求項７に記載の音声合成方法。
【請求項１１】
上記音声単位内の有声部分において、各フレームの上記音声波形データの位相特性を最小位相特性に整えるようにした
ことを特徴とする請求項７に記載の音声合成方法。
【請求項１２】
上記音声単位内の有声部分において、各フレームの上記音声波形データの位相特性を最大位相特性に整えるようにした
ことを特徴とする請求項７に記載の音声合成方法。
【発明の詳細な説明】
【０００１】
【目次】
以下の順序で本発明を説明する。
産業上の利用分野
従来の技術
発明が解決しようとする課題
課題を解決するための手段（図１、図２、図６）
作用（図１、図２、図６）
実施例（図１～図９）
（１）第１実施例（図１～図５）

（２）第２実施例（図１、図５～図９）
（３）他の実施例
発明の効果
【０００２】
【産業上の利用分野】
本発明は音声合成装置及び方法に関し、特に単音節又はそれ以上の音節数からなる音声単位波形を同一音素内で編集合成する波形編集型の音声合成装置に適用して好適なものである。
【０００３】
【従来の技術】
従来、規則合成方式による音声合成装置においては、入力された文字の系列を解析した後、所定の規則に従つてパラメータを合成することにより、いかなる言葉でも音声合成し得るようになされている。すなわち規則合成方式による音声合成装置は、入力された文字の系列を解析した後、所定の規則に従つて各文節ごとにアクセントを検出し、各文節の並びから文字系列全体としての抑揚、ポーズ等を表現するピツチパラメータを合成する。
【０００４】
さらに音声合成装置は、同様に所定の規則に従つて各文節を例えばＣＶ／ＶＣ単位のような音声単位に分割した後、そのスペクトラムを表現する合成パラメータを生成する。これによりピツチパラメータ及び合成パラメータに基づいて合成音を発声するようになされている。
【０００５】
またより高品質な音声を合成するために、音声単位として周期性を有する有声部分については実音声を分析処理し、その１周期にあたる音声波形データをそれぞれ保持し、また周期性を有さない無声部分については実音声をそのまま音声波形データとして保持し、合成時にはこれらの音声波形データをピツチパラメータに基づいて波形編集して合成音を生成する。
【０００６】
【発明が解決しようとする課題】
ところでこのような波形編集の手法では、ピツチパラメータに基づいて音声単位内の音声波形データを繰り返したり、あるいは間引くことによつてフレーム数を調整して音声を合成している。ここで用いられる個々の音声単位は、それが抽出された実音声内での前後の音韻環境の影響を受けており、その影響が合成音声内に表れてくる。
【０００７】
すなわちある音声単位では合成時における音韻環境と、抽出された実音声内での音韻環境とが異なる場合が生じ、合成音声の各音声単位を接続した場合には、実音声と比べて不自然な音声波形が生成され、周波数領域での不連続性が原因となつて異聴等が発生する。またＣＶ／ＶＣ単位による音声合成のように音声単位を同一音素内で接続する場合には、周波数領域での不連続性が聴感上特に感知されやすく、合成音声の品質が劣化しやすいという問題がある。
【０００８】
この問題を解決するために、従来の音声合成装置では音声単位間の接続部分で補間処理を行うことが一般的である。補間処理の一つとして、音声波形データの時間領域における線形補間があるが、補間波形はスペクトル的に補間されたものとは対応せず、合成された音声のスペクトル特性は自然音声からかけ離れたものとなることがある。
【００１０】
また一般的に人間の聴覚は音の周波数特性を聞き分けるが、位相特性に対しては周波数特性ほどの弁別能力はないと言われている。しかし上述したような音声単位のスペクトル包絡軌道の正規化手法を用いた波形編集による合成音声では、音声単位間の接続部分での接続歪みがほとんどないために、合成音声の有声部における接続部分での位相特性の不連続性が、実音声に比べて聴感的に違和感を生じさせてる問題がある。

【００１１】
この問題の解決策として、音声単位内の有声部の位相特性を零位相特性に整える方法がある。この方法によつて得られる音声波形データは、波形データの中心のサンプル点において左右対称な波形となる。従つて得られる音声波形データの半分のみを音声単位記憶部に貯えればよく、この点で記憶データ容量の削減が可能である。しかし音声合成時には半分のデータから対称な音声波形データを生成する処理を必要とする。
【００１２】
本発明は以上の点を考慮してなされたもので、実際の人間の音声に比して違和感のない合成音を発声することができる音声合成装置、音声合成方法及び記録媒体を提案しようとするものである。
【００１３】
【課題を解決するための手段】
かかる課題を解決するために本発明においては、音韻記号と韻律記号とに基づいて所定の音韻規則及び韻律規則によつて韻律情報を設定する音声合成規則部４と、音声波形データを必要フレーム数貯えた音声単位及び韻律情報に基づいて合成音を生成する音声合成部５とを有する波形編集型の音声合成装置１において、同じ音素の接続フレームを持つ複数の音声単位について、接続フレームにおける代表的なスペクトル包絡特性を求め、接続フレームにおけるスペクトル包絡特性が代表的なスペクトル包絡特性になるように、同じ音素の接続フレームを持つ音声単位のスペクトル包絡軌道を正規化して求めた音声単位を記憶する音声単位記憶部２を設けるようにした。
【００１４】
また本発明においては、これに加えて、音声単位記憶部２に貯えられている音声単位内の有声部分において、各フレームの音声波形データの位相特性を最小位相特性に整えるようにした。
【００１５】
また本発明においては、音韻記号と韻律記号とに基づいて所定の音韻規則及び韻律規則によつて韻律情報を設定すると共に、音声波形データを必要フレーム数貯えた音声単位及び韻律情報に基づいて合成音を生成する波形編集型の音声合成方法において、同じ音素の接続フレームを持つ複数の音声単位について、接続フレームにおける代表的なスペクトル包絡特性を求め、接続フレームにおけるスペクトル包絡特性が代表的なスペクトル包絡特性になるように、同じ音素の接続フレームを持つ音声単位のスペクトル包絡軌道を正規化して求めた音声単位を記憶するようにした。
【００１６】
また本発明においては、これに加えて、音声単位内の有声部分において、各フレームの音声波形データの位相特性を最小位相特性に整えるようにした。
【００１７】
【作用】
音声単位記憶部２におけるスペクトル包絡軌道が正規化された音声単位データセツトを用いることによつて、時間軸上での単純な波形補間による合成方式に比べ、補間処理を行うことなしに接続部をよりなめらかに接続していくことができ、人間の音声に近い高品質な任意の合成音を得ることができる。
【００１８】
また音声波形データのスペクトル包絡軌道が正規化された音声単位による波形編集では、合成音声の有声部における音声単位間の接続部分での位相特性の不連続性から、実音声に比べて聴感的に違和感が生じ品質の劣化が問題となるが、音声波形データの位相特性を全て最小位相特性に整えることによつて、聴感上の違和感を解消することが出来、より高品質な任意の合成音声を得ることができる。
【００１９】
また、最小位相特性ということからその音声波形データは、波形データの中心のサンプル点を原点とした場合に、負の時間領域では波形振幅値が全て０であるので、音声単位記憶

部２には音声波形データの正の時間領域だけのデータを貯えれば良い。従つて零位相特性の音声波形データと同様のデータ削減が可能である上に、音声合成時には、零位相特性の音声波形データでは必要な対称な音声波形を生成する処理が不用となり、合成アルゴリズムを簡略化し得る。

【００２０】

【実施例】

以下図面について、本発明の一実施例を詳述する。

【００２１】

（１）第１実施例

図１において、１は全体として演算処理装置構成の音声合成装置を示し、音声単位記憶部２、文章解析部３、音声合成規則部４及び音声合成部５に分割される。文章解析部３は、所定の入力装置から入力されたテキスト入力（文字の系列で表された文章等でなる）を所定の辞書を基準にして解析し、仮名文字列に変換した後、単語、文節毎に分解する。

【００２２】

すなわち日本語においては、英語のように単語が分かち書きされていないことから、例えば「米国産業界」のような言葉は、「米国／産業・界」、「米／国産／業界」のように２種類以上に区分化し得る。このため文章解析部３は、辞書を参考にしながら、言葉の連続関係及び単語の統計的性質を利用して、テキスト入力を単語、文節毎に分解するようになされ、これにより単語、文節の境界を検出するようになされている。さらに文章解析部３は、各単語毎に基本アクセントを検出した後、音声合成規則部４に出力する。

【００２３】

音声合成規則部４は、日本語の特徴に基づいて設定された所定の音韻規則に従つて、文章解析部３の検出結果及びテキスト入力を処理するようになされている。すなわち、日本語の自然な音声は、言語学的特性に基づいて区別すると、約１００程度の発声の単位に区分することができる。例えば、「さくら」という単語を発声の単位に区分すると、「ｓａ」＋「ａｋ」＋「ｋｕ」＋「ｕｒ」＋「ｒａ」の５つのＣＶ／ＶＣ単位に分割することができる。

【００２４】

さらに日本語は、単語が連続する場合、連なつた後ろの語の語頭音節が濁音化したり（すなわち続濁でなる）、語頭以外のガ行音が鼻音化したりして、単語単体の場合と発声が変化する特徴がある。従つて音声合成規則部４は、これら日本語の特徴に従つて音韻規則が設定されるようになされ、その規則に従つてテキスト入力を音韻記号列（すなわち上述の「ｓａ」＋「ａｋ」＋「ｋｕ」＋「ｕｒ」＋「ｒａ」等の連続する列でなる）に変換するようになされている。さらに音声合成規則部４は、この音韻記号列に基づいて、音声単位記憶部２から各音声単位データをロードする。

【００２５】

ここで音声合成装置１は、波形編集の手法を用いて合成音を発声するようになされ、音声単位記憶部２からロードされるデータは、各ＣＶ／ＶＣ単位で表される合成音を生成する際に用いられる波形データでなる。この波形合成に用いる音声単位データは次のように構成される。

【００２６】

すなわち音声単位データの有声部に関しては、実音声の有声部分において１ピツチに対応する音声波形データを必要なフレーム数だけ貯えたものからなり、また音声単位データの無声部に関しては、実音声の無声部分の波形を切り出してそのまま貯えたものからなる。例えば音声単位データがＣＶ／ＶＣ単位である場合には、ある１つの音声単位ＣＶは、その子音部Ｃが無声子音である時には無声部分の切り出し波形と、１ピツチの音声波形の複数フレームからなる音声波形データから構成され、また子音部Ｃが有声子音である時には１ピツチの音声波形からなる複数フレームの音声波形データから構成される。

【００２７】

またこの音声単位データは、音声単位記憶部２に貯えられている全ての音声単位データの

集まりである音声単位データセツト内において、図２に示すようなブロツクによつて、音声単位データ内の音声波形データのスペクトル包絡軌道が正規化される。この音声波形データのスペクトル包絡軌道の正規化処理の具体例を以下に示す。

【００２８】

すなわちまず音声単位データセツトに含まれる少なくとも一つの音素に対して、音声単位間を接続する場合の接続フレームにおける代表的なスペクトル包絡特性を設定する。これはＣＶ／ＶＣ単位による音声単位データセツトについて、音素／ａ／　に対する代表的なスペクトル包絡特性を設定する場合には、／ａｋ／、／ａｓ／、／ｋａ／、／ｓａ／のように音素／ａ／　を含む音声単位データセツト内の当該音声単位データ全てについて、音素／ａ／　が音声単位データの前方音素となる場合にはその音声単位データ内の前端フレームを対象の接続フレームとし、また音素／ａ／　が音声単位データの後方音素となる場合には、その音声単位データ内の後端フレームを対象の接続フレームとして、対象の接続フレームの音声波形データのスペクトル包絡特性を求める。

【００２９】

このようにして求められた音声単位データ全てにおけるスペクトル包絡特性から、そのスペクトル包絡特性の空間内での重心であるセントロイドを求め、これを音素／ａ／　における代表的なスペクトル包絡特性とする。あるいはスペクトル包絡特性の空間内において求められたセントロイドに最も近い位置にあるスペクトル包絡特性を代表的なスペクトル包絡特性としても良い。同様にして、スペクトル包絡軌道の正規化を行う他の音素に対しても、その代表的なスペクトル包絡特性を設定する。

【００３０】

次に音素に対して設定された代表的なスペクトル包絡特性を用いて、各音声単位データ内の音声波形データのスペクトル包絡軌道の正規化を行う。この具体的な方法は、音声単位データ／ａｍ／の場合では次のようになる。すなわち音声単位データ／ａｍ／内の各フレームにおける音声波形データを周波数分析し、それぞれの音声波形データのスペクトル包絡特性とスペクトル微細構造特性を抽出する。

【００３１】

次に音素／ａ／　の代表的なスペクトル包絡特性と、抽出された音声単位データ／ａｍ／内の前端フレームにおけるスペクトル包絡特性との差分を計算して、これを前端フレームにおけるスペクトル包絡特性のギヤツプとし、また音素／ｍ／　の代表的なスペクトル包絡特性と、抽出された音声単位データ／ａｍ／内の後端フレームにおけるスペクトル包絡特性との差分を計算して、これを後端フレームにおけるスペクトル包絡特性のギヤツプとする。

【００３２】

音声単位データ／ａｍ／内の音素／ａ／　と音素／ｍ／　との境界となるフレームを中心として、求められた両端のフレームにおけるスペクトル包絡特性のギヤツプを打ち消すように、音声単位データ／ａｍ／に対するスペクトル包絡軌道の正規化関数を設定する。図３はスペクトル包絡軌道の正規化関数である。この正規化関数は、音声単位データ内の音素境界に接するフレームでスペクトル包絡特性の補正量が０となるように、音声単位データの両端のスペクトル包絡特性のギヤツプを直線補間する関数である。

【００３３】

また図４はスペクトル包絡軌道の正規化処理である。設定された正規化関数を抽出された音声単位データ／ａｍ／の各フレームのスペクトル包絡特性に適用することで、両端のフレームにおけるスペクトル包絡特性はそれぞれ音素／ａ／　と音素／ｍ／　との代表的なスペクトル包絡特性となり、しかも音声単位データ内では滑らかなスペクトル包絡軌道が実現できる。

【００３４】

最後に音声単位データ／ａｍ／内の各フレームにおいて、正規化されたスペクトル包絡情報と、周波数分析によつて得られた原音声波形データの微細構造情報とを用いて時間波形を求め、この時間波形を各フレームの音声波形データとする。このような手法による音声

単位データのスペクトル包絡軌道の正規化を、該当する音声単位データ全てに対して行う。

【００３５】

音声合成規則部４は、音声単位記憶部２からロードされた音声単位データをテキスト入力に応じた順序（以下このデータを合成波形データと呼ぶ）で合成し、かくして抑揚のない状態で、テキスト入力を読み上げた合成音声波形を得ることができる。さらに音声合成規則部４は所定の韻律規則に基づいて、テキスト入力を適当な長さで分割して、切れ目すなわちポーズを検出する。かくして図５に示すように、例えばテキスト入力として文章「きれいな花を山田さんからもらいました」が入力された場合は（図５（Ａ））、当該テキスト入力は「きれいな」、「はなを」、「やまださんから」、「もらいました」に分解された後、「はなを」及び「やまださんから」の間にポーズが検出される（図５（Ｂ））。

【００３６】

さらに音声合成規則部４は、韻律規則及び各単語の基本アクセントに基づいて、各文節のアクセントを検出する。すなわち日本語の文節単体のアクセントは、感覚的に仮名文字を単位として（以下モーラと呼ぶ）、高低の２レベルで表現することができる。このとき文節の内容等に応じて、文節のアクセント位置を区別することができる。例えば、端、箸、橋は、２モーラの単語で、それぞれアクセントのない０型、アクセントの位置が先頭のモーラにある１型、アクセントの位置が２モーラ目にある２型に分類することができる。かくして、この実施例において音声合成規則部４は、テキスト入力の各文節を、それぞれ１型、２型、０型、４型と分類し（図５（Ｃ））、これにより文節単位でアクセント及びポーズを検出する。

【００３７】

さらに音声合成規則部４は、アクセント及びポーズの検出結果に基づいて、テキスト入力全体の抑揚を表す基本ピツチパターンを生成する。すなわち日本語において文節のアクセントは、感覚的に２レベルで表し得るのに対し、実際の抑揚は、アクセントの位置から徐々に低下する特徴がある（図５（Ｄ））。さらに日本語においては、文節が連続して１つの文章になると、ポーズから続くポーズに向かつて、抑揚が徐々に低下する特徴がある（図５（Ｅ））。

【００３８】

従つて音声合成規則部４は、かかる日本語の特徴に基づいて、テキスト入力全体の抑揚を表すパラメータを各モーラ毎に生成した後、人間が発声した場合と同様に抑揚が滑らかに変化するように、モーラ間の補間によりパラメータを設定する。かくして音声合成規則部４は、テキスト入力に応じた順序で、各モーラのパラメータ及び補間したパラメータを合成し（以下ピツチパターンと呼ぶ）、かくしてテキスト入力を読み上げた音声の抑揚を表すピツチパターン（図５（Ｆ））を得ることができる。

【００３９】

音声合成部５は、合成波形データ及びピツチパターンに基づいて波形合成処理を行ない合成音を生成する。この波形合成処理は次のように行う。すなわち合成音声の有声部においては、合成波形データ内の１ピツチに対応した波形データをピツチパターンに基づいて並べ重畳していく。また合成音声の無声部分においては、合成波形データ内の切り出し波形をそのまま所望の合成音声の波形とする。これによりピツチパターンの変化に追従して抑揚の変化する合成音を得ることができる。

【００４０】

このように音声を合成するために用いる音声単位データのスペクトル包絡軌道を正規化することによつて、高品質な合成音声が得られる波形重畳方式の音声合成において、同一音素内における音声単位接続部での接続歪みがほとんど解消され、音声合成時における補間処理を行うことなしに、音声単位データが滑らかに接続された人間の音声に近い高品質な任意の合成音が得られる。

【００４１】

以上の構成において、所定の入力装置から入力されたテキスト入力は、文章解析部３で、

所定の辞書を基準にして解析され、単語、文節の境界及び基本アクセントが検出される。単語、文節の境界及び基本アクセントの検出結果は、音声合成規則部４で、所定の音韻規則に従つて処理され、抑揚のない状態でテキスト入力を読み上げた音声を表す合成波形データが生成される。

【００４２】

さらに単語、文節の境界及び基本アクセントの検出結果は、音声合成規則部４で、所定の韻律規則に従つて処理され、テキスト入力全体の抑揚を表すピツチパターンが生成される。ピツチパターンは合成波形データと共に音声合成部５に出力され、ここでピツチパターン及び合成波形データに基づいて合成音が生成される。

【００４３】

以上の構成によれば、高品質な合成音声が得られる波形重畳方式の音声合成において、合成時における音声単位間の補間処理を行うことなしに滑らかに音声単位が接続され、人間の音声に近い高品質な合成音声を任意に生成することができる。

【００４４】

（２）第２実施例

この第２実施例の場合、音声単位データは音声単位記憶部２に貯えられている全ての音声単位データの集まりである音声単位データセツト内において、図６に示すようなブロツクによつて、音声単位データ内の音声波形データのスペクトル包絡軌道が正規化されている。この音声波形データのスペクトル包絡軌道の正規化処理の具体例を以下に示す。

【００４５】

すなわちまず図２の場合と同様にして、音声単位データセツトに含まれる少なくとも一つの音素に対して、音声単位間を接続する場合の接続フレームにおける代表的なスペクトル包絡特性を設定する。これはＣＶ／ＶＣ単位による音声単位データセツトについて、音素／ａ／　に対する代表的なスペクトル包絡特性を設定する場合には、／ａｋ／、／ａｓ／、／ｋａ／、／ｓａ／のように音素／ａ／　を含む該当音声単位データ全てについて、音素／ａ／　が音声単位データの前方音素となる場合にはその音声単位データ内の前端フレームを対象の接続フレームとし、また音素／ａ／　が音声単位データの後方音素となる場合には、その音声単位データ内の後端フレームを対象の接続フレームとして、対象の接続フレームの音声波形データのスペクトル包絡特性を求める。

【００４６】

このようにして求められた音声単位データ全てにおけるスペクトル包絡特性から、そのスペクトル包絡特性の空間内での重心であるセントロイドを求め、これを音素／ａ／における代表的なスペクトル包絡特性とする。あるいは、スペクトル包絡特性の空間内において求められたセントロイドに最も近い位置にあるスペクトル包絡特性を代表的なスペクトル包絡特性としても良い。同様にして、スペクトル包絡軌道の正規化を行う他の音素に対しても、その代表的なスペクトル包絡特性を設定する。

【００４７】

次に音素に対して設定された代表的なスペクトル包絡特性を用いて、各音声単位データ内の音声波形データのスペクトル包絡軌道の正規化を行う。この具体的な方法は、音声単位データ／ａｍ／の場合では次のようになる。すなわち音声単位データ／ａｍ／内の各フレームにおける音声波形データ（図７（Ａ））をケプストラム分析し、該当する音声波形データのケプストラムｃ（ｎ）（図７（Ｂ））を求める（ここでｎ：－（Ｎ／２）≦　ｎ＜（Ｎ／２），Ｎ：ケプストラム分析のフレーム長）。この結果求められたケプストラムｃ（ｎ）に対して任意の次数ｋ　（０　＜　ｋ＜（Ｎ／２）　、通常３２次程度）を設定し、次式

【数１】

$$w1(n) = \begin{cases} 1 & (|n| < k) \\ 0 & (|n| \geqq k) \end{cases} \quad \cdots\cdots (1)$$

$$w2(n) = \begin{cases} 0 & (|n| < k) \\ 1 & (|n| \geqq k) \end{cases} \quad \cdots\cdots (2)$$

$$c1(n) = c(n) \times w1(n) \quad \cdots\cdots (3)$$

$$c2(n) = c(n) \times w2(n) \quad \cdots\cdots (4)$$

ようなケプストラム窓ｗ１、ｗ２によつてリフタリングを行う。

【００４８】

この（３）式によつて、ｋ　次よりも低次のケプストラム係数、つまり低ケフレンシー部ｃ１（ｎ）　（図７（Ｃ））が抽出される。この低ケフレンシー部は、周波数領域では分析対象の音声波形データのスペクトル包絡特性を表現する。また（４）式によつて、ｋ次以上のケプストラム係数、つまり高ケフレンシー部ｃ２（ｎ）　（図７（Ｄ））が抽出される。この高ケフレンシー部は、周波数領域では分析対象の音声波形データの微細構造特性を表現する。このようにして、分析対象の音声波形データのスペクトル包絡特性と微細構造特性をケプストラムによつて分離しそれぞれを抽出する。

【００４９】

次に音素／ａ／　の代表的なスペクトル包絡特性のケプストラムを求め、（１）式のケプストラム窓によつて、低ケフレンシー部を取り出す。この音素／ａ／　の代表的なスペクトル包絡特性を表すケプストラムと、抽出された音声単位データ／ａｍ／内の前端フレームにおけるスペクトル包絡特性を表すケプストラムｃ１（ｎ）　との差分を計算して、これを前端フレームにおけるスペクトル包絡特性のケプストラム次元でのギヤツプとする。同様に音素／ｍ／　の代表的なスペクトル包絡特性を表すケプストラムを求め、抽出された音声単位データ／ａｍ／内の後端フレームにおけるスペクトル包絡特性を表すケプストラムｃ１（ｎ）　との差分を計算し、これを後端フレームにおけるスペクトル包絡特性のケプストラム次元でのギヤツプとする。

【００５０】

音声単位データ／ａｍ／内の音素／ａ／　と音素／ｍ／　との境界となるフレームを中心として、求められた両端のフレームにおけるスペクトル包絡特性のケプストラム次元でのギヤツプを打ち消すように、音声単位データ　／ａｍ／　に対するスペクトル包絡軌道の正規化関数を設定する。図８はスペクトル包絡軌道の正規化関数を周波数領域で表現したものである。この正規化関数は、次式

【数２】

$$c1'(n) = f(c1(n)) \quad \cdots\cdots (5)$$

（ｆ :正規化関数　）

で表され、音声単位データ内の音素境界に接するフレームでスペクトル包絡特性の補正量が０となるように、音声単位データの両端のフレームにおける低次のケプストラム係数の差、つまりスペクトル包絡特性のケプストラム次元でのギヤツプを直線補間する関数である。

【００５１】
図９はスペクトル包絡軌道の正規化処理である。設定された正規化関数を、抽出された音声単位データ／ａｍ／の各フレームのスペクトル包絡特性を表す低ケフレンシー部に適用することで、両端のフレームにおけるスペクトル包絡特性は、それぞれ音素／ａ／　と音素／ｍ／　との代表的なスペクトル包絡特性となり、しかも音声単位データ内では滑らかなスペクトル包絡軌道が実現できる。
【００５２】
次に音声単位データの／ａｍ／内の各フレームにおいて、正規化されたスペクトル包絡情報を表す低ケフレンシー部ｃ１’（ｎ）（図７（Ｅ））と、ケプストラム分析及び（２）式のケプストラム窓によつて抽出されている原音声波形データの微細構造情報を表す高ケフレンシー部ｃ２（ｎ）　とを次式
【数３】

$$c'(n) = c1'(n) + c2(n) \qquad \cdots\cdots (6)$$

のようにして加算する。
【００５３】
この（６）式によつて得られるケプストラムｃ’（ｎ）　は、分析対象の音声波形データの微細構造特性と、正規化されたスペクトル包絡特性を合わせ持つ周波数特性を表現する。求められたケプストラムｃ’（ｎ）　に対して、ケプストラム分析における逆変換を行つて、スペクトル包絡特性が正規化された音声波形データを得るが、この時点では音声波形データが最小位相信号となるために次のような処理を行う。
【００５４】
すなわち、ケプストラムｃ’（ｎ）　は零位相特性であるが、次式
【数４】

$$c''(n) = \begin{cases} 0 & (n < 0) \\ c'(n) & (n = 0) \\ 2 \times c'(n) & (n > 0) \end{cases} \qquad \cdots\cdots (7)$$

によつて、ｃ’（ｎ）　と同じ周波数特性であり、なおかつ最小位相特性である複素ケプストラムｃ’’（ｎ）（図７（Ｆ））が得られることが知られている（Ａ．Ｖ．Ｏｐｐｅｎｈｅｉｍ，”Ａ Ｓｐｅｅｃｈ Ａｎａｌｙｓｉｓ－Ｓｙｎｔｈｅｓｉｓ Ｓｙｓｔｅｍ Ｂａｓｅｄ ｏｎ Ｈｏｍｏｍｏｒｐｈｉｃ Ｆｉｌｔｅｒｉｎｇ”，Ｊ．Ａｃｏｕｓｔ．Ｓｏｃ．Ａｍ．，Ｖｏｌ．４５，ｐｐ．４５８－４６５，Ｆｅｂ １９６９ ）。
【００５５】
最後に（７）式によつて求められる最小位相の複素ケプストラム　ｃ’’（ｎ）　を逆変換することによつて、正規化されたスペクトル包絡特性を持ち、なおかつ最小位相信号である時間波形（図７（Ｇ））が得られ、この時間波形を各フレームの音声波形データとする。上述した手法による音声単位データのスペクトル包絡軌道の正規化を、該当する音声単位データ全てに対して行う。
【００５６】
音声合成規則部４は、音声単位記憶部２からロードされた音声単位データをテキスト入力に応じた順序（以下このデータを合成波形データと呼ぶ）で合成し、かくして抑揚のない状態で、テキスト入力を読み上げた合成音声波形を得ることができる。さらに音声合成規

則部４は所定の韻律規則に基づいて、テキスト入力を適当な長さで分割して、切れ目すなわちポーズを検出する。かくして図５に示すように、例えばテキスト入力として文章「きれいな花を山田さんからもらいました」が入力された場合は（図５（Ａ））、当該テキスト入力は「きれいな」、「はなを」、「やまださんから」、「もらいました」に分解された後、「はなを」及び「やまださんから」の間にポーズが検出される（図５（Ｂ））。

【００５７】

さらに音声合成規則部４は、韻律規則及び各単語の基本アクセントに基づいて、各文節のアクセントを検出する。すなわち日本語の文節単体のアクセントは、感覚的に仮名文字を単位として（以下モーラと呼ぶ）、高低の２レベルで表現することができる。このとき文節の内容等に応じて、文節のアクセント位置を区別することができる。例えば、端、箸、橋は、２モーラの単語で、それぞれアクセントのない０型、アクセントの位置が先頭のモーラにある１型、アクセントの位置が２モーラ目にある２型に分類することができる。かくして、この実施例において音声合成規則部４は、テキスト入力の各文節を、それぞれ１型、２型、０型、４型と分類し（図５（Ｃ））、これにより文節単位でアクセント及びポーズを検出する。

【００５８】

さらに音声合成規則部４は、アクセント及びポーズの検出結果に基づいて、テキスト入力全体の抑揚を表す基本ピツチパターンを生成する。すなわち日本語において文節のアクセントは、感覚的に２レベルで表し得るのに対し、実際の抑揚は、アクセントの位置から徐々に低下する特徴がある（図５（Ｄ））。さらに日本語においては、文節が連続して１つの文章になると、ポーズから続くポーズに向かつて、抑揚が徐々に低下する特徴がある（図５（Ｅ））。

【００５９】

従つて音声合成規則部４は、かかる日本語の特徴に基づいて、テキスト入力全体の抑揚を表すパラメータを各モーラ毎に生成した後、人間が発声した場合と同様に抑揚が滑らかに変化するように、モーラ間の補間によりパラメータを設定する。かくして音声合成規則部４は、テキスト入力に応じた順序で、各モーラのパラメータ及び補間したパラメータを合成し（以下ピツチパターンと呼ぶ）、かくしてテキスト入力を読み上げた音声の抑揚を表すピツチパターン（図５（Ｆ））を得ることができる。

【００６０】

音声合成部５は、合成波形データ及びピツチパターンに基づいて波形合成処理を行ない合成音を生成する。この波形合成処理は次のように行う。すなわち合成音声の有声部においては、合成波形データ内の１ピツチに対応した波形データをピツチパターンに基づいて並べ重畳していく。また合成音声の無声部分においては、合成波形データ内の切り出し波形をそのまま所望の合成音声の波形とする。これによりピツチパターンの変化に追従して抑揚の変化する合成音を得ることができる。

【００６１】

このように、スペクトル包絡軌道が正規化された音声単位データにおいて、その有声部分の音声波形データの位相特性を最小位相特性に整えることによつて、音声単位データ間の接続部分での位相特性が連続となり、聴感的な違和感が解消され、高品質な任意合成音が得られる波形重畳方式において、より自然で滑らかな音声単位データの接続が実現できる。

【００６２】

以上の構成において、所定の入力装置から入力されたテキスト入力は、文章解析部２で、所定の辞書を基準にして解析され、単語、文節の境界及び基本アクセントが検出される。単語、文節の境界及び基本アクセントの検出結果は、音声合成規則部４で、所定の音韻規則に従つて処理され、抑揚のない状態でテキスト入力を読み上げた音声を表す合成波形データが生成される。

【００６３】

さらに単語、文節の境界及び基本アクセントの検出結果は、音声合成規則部４で、所定の

韻律規則に従つて処理され、テキスト入力全体の抑揚を表すピツチパターンが生成される。ピツチパターンは合成波形データと共に音声合成部５に出力され、ここでピツチパターン及び合成波形データに基づいて合成音が生成される。

【００６４】

以上の構成によれば、スペクトル包絡軌道が正規化された音声単位データセツトを用いた波形重畳方式の音声合成において、合成時における音声単位データ間の位相特性のずれが解消され、人間の音声に近い高品質な合成音声を任意に生成することができる。

【００６５】

（３）他の実施例

なお上述の第２実施例においては、音声単位データ内の有声部分の音声波形データの位相特性を最小位相特性に整える場合について述べたが、最小位相特性の複素ケプストラムを求める（７）式の変わりに、次式

【数５】

$$c'''(n) = \begin{cases} 2 \times c'(n) & (n < 0) \\ c'(n) & (n = 0) \\ 0 & (n > 0) \end{cases} \qquad \cdots\cdots (8)$$

を用いて最大位相特性の複素ケプストラムを求め、これによつて音声波形データの位相特性を最大位相特性に整えるようにしても良い。

【００６６】

また上述の実施例においては、文章解析部でテキスト入力を解析したが、これに代え音声合成装置内には文章解析部を持たず、音声合成装置への直接の入力として、音韻記号と韻律記号とが与えられるようになされても上述の実施例と同様の効果を実現できる。

【００６７】

さらに上述の実施例においては、音声単位データに対するスペクトル包絡軌道の正規化処理を、音声単位データ内の音素境界を中心にして全てのフレームに対して施す場合について述べたが、本発明はこれに限らず、音声単位データの前端からの任意のフレーム数及び後端からの任意のフレーム数のみに対して正規化処理を施しても良い。

【００６８】

さらに上述の実施例においては、音声単位データに対するスペクトル包絡軌道の正規化処理を、音声単位データ全体に対して施す場合について述べたが、本発明はこれに限らず、音声単位内の有声部分に対してのみ正規化処理を施しても良い。

【００６９】

さらに上述の実施例においては、音声単位データがＣＶ／ＶＣ単位である場合について述べたが、本発明はこれに限らず、音声単位データがＶＣＶ単位やＣＶＣ単位、あるいはその両者のように、音声単位データを同一音素内で接続する音声合成方式において、音声単位データ内の音韻連鎖が任意の数であつたり、音声単位データ内の音韻連鎖のパターンが任意である場合にも、音声単位内の前端フレーム及び後端フレームを含む音素に対してのみ正規化処理を施しても良い。

【００７０】

【発明の効果】

上述のように本発明によれば、高品質な合成音声が得られる波形重畳方式の音声合成において、音声合成時の音声単位間の補間処理を行うことなく、音声単位接続部での接続歪みをほとんど解消することができ、人間の音声に近い高品質な合成音を任意に合成することができる音声合成装置及び方法を得ることができる。

【００７１】

さらに上述のように本発明によれば、スペクトル包絡軌道が正規化された音声単位データセツトを用いた波形重畳方式の音声合成において、合成時における音声単位データ間の位相特性のずれが解消され、人間の音声に近い高品質な合成音を任意に合成することができる音声合成装置及び方法を得ることができる。

【図面の簡単な説明】

【図１】本発明による音声合成装置の一実施例を示すブロツク図である。

【図２】図１の音声合成装置における音声単位データセツトの正規化処理の第１実施例を示すブロツク図である。

【図３】音声単位データのスペクトル包絡軌道の正規化関数を示す特性曲線図である。

【図４】音声単位データのスペクトル包絡軌道の正規化処理の説明に供する特性曲線図である。

【図５】本発明の一実施例の動作として基本ピツチパターンの生成の説明に供する略線図である。

【図６】図１の音声合成装置における音声単位データセツトの正規化処理の第２実施例を示すブロツク図である。

【図７】図６の正規化処理の動作の説明に供する特性曲線図である。

【図８】音声単位データのスペクトル包絡軌道の正規化関数を周波数領域で表現して示す特性曲線図である。

【図９】音声単位データのスペクトル包絡軌道の正規化処理の説明に供する特性曲線図である。

【符号の説明】

１……音声合成装置、２……音声単位記憶部、３……文章解析部、４……音声合成規則部、５……音声合成部。

【図１】

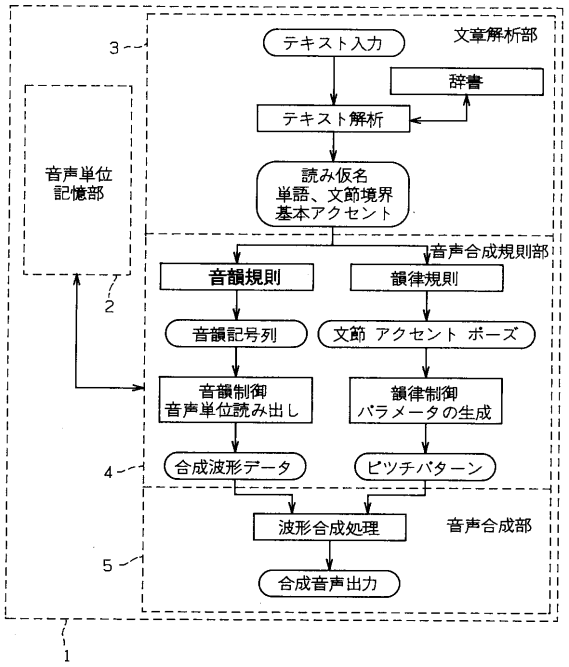

図１　実施例による音声合成装置

【図２】

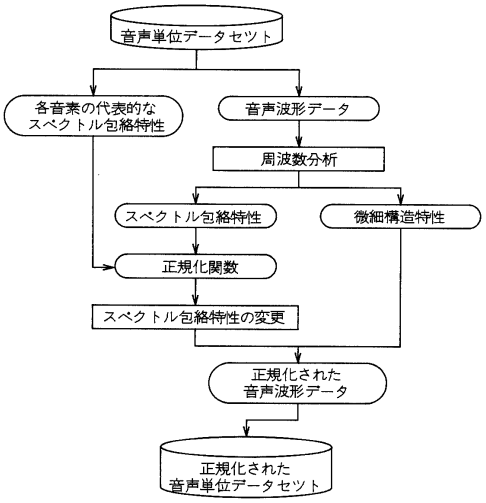

図２　第１実施例による音声単位データセットの正規化処理

【図３】

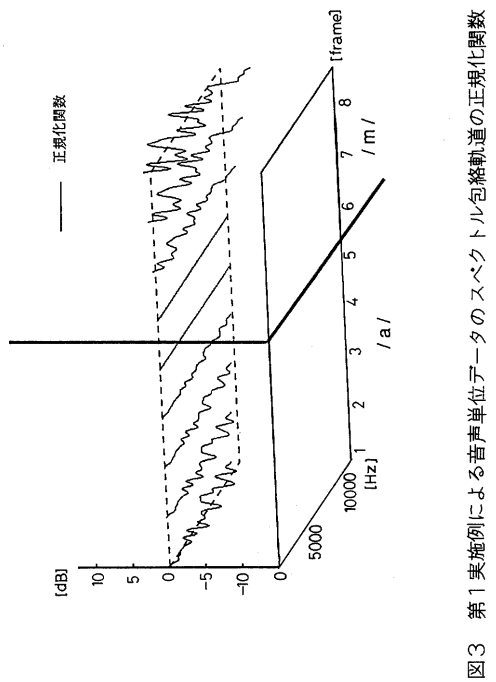

図３　第１実施例による音声単位データのスペクトル包絡軌道の正規化関数

【図４】

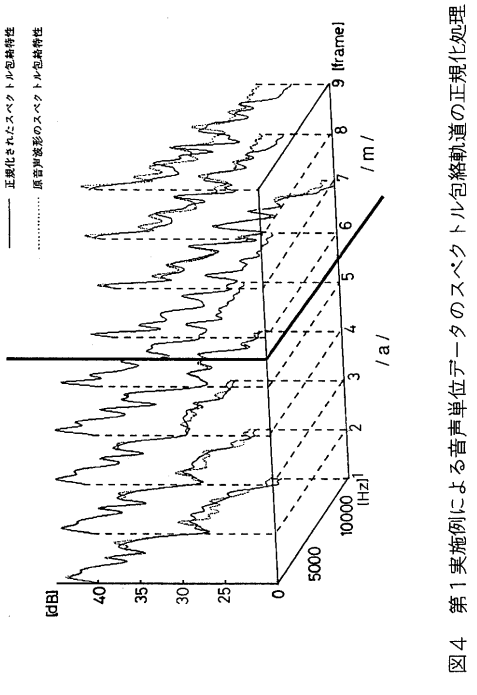

図４　第１実施例による音声単位データのスペクトル包絡軌道の正規化処理

【図５】

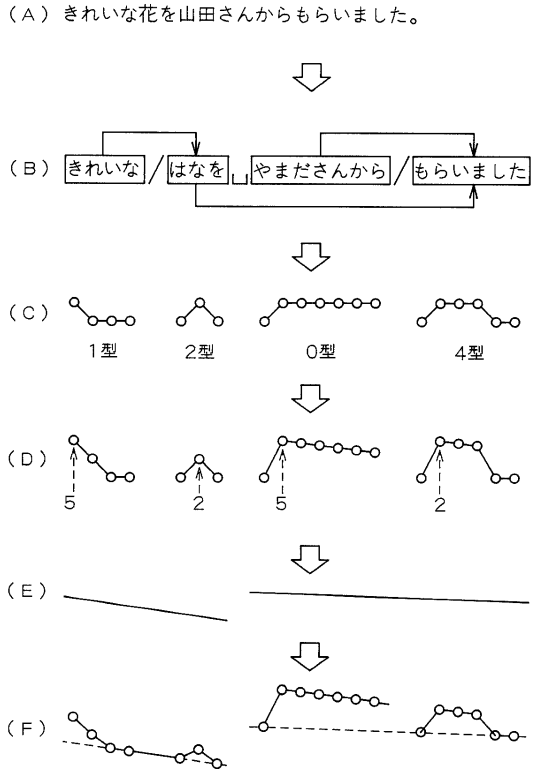

図５　基本ピッチパターンの生成

【図６】

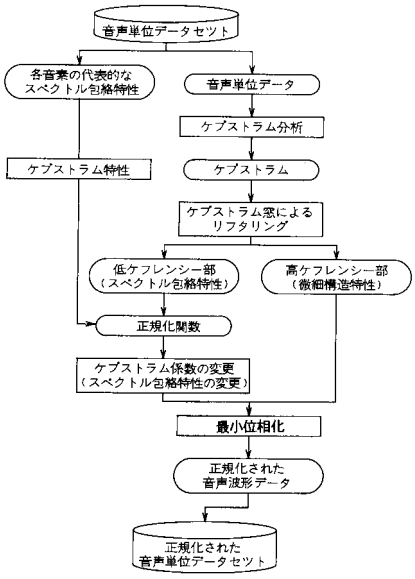

図６　第２実施例による音声単位データセットの正規化処理

【図７】

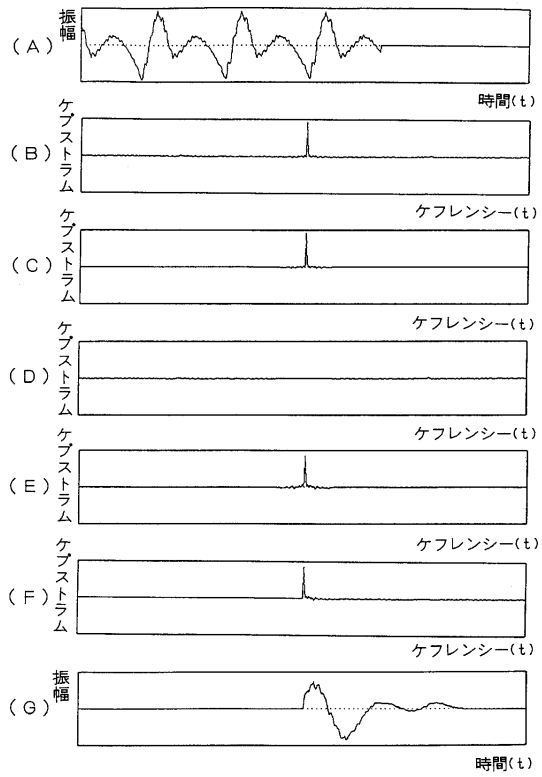

図７　第２実施例による正規化処理の動作

【図８】

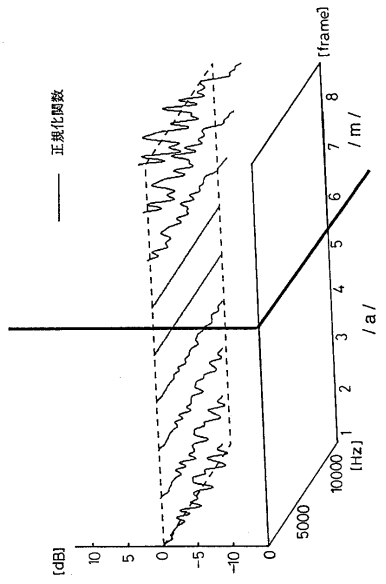

図８　第２実施例による音声単位データのスペクトル包絡軌道の正規化関数

【図９】

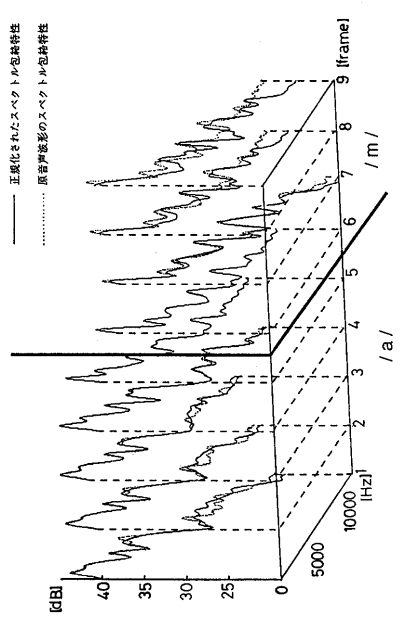

図９　第２実施例による音声単位データのスペクトル包絡軌道の正規化処理

フロントページの続き

合議体
審判長　原　光明
審判官　小松　正
審判官　橋爪　正樹

(56)参考文献　特開平２－１３５９３１（ＪＰ，Ａ）
特開昭５８－１００９００（ＪＰ，Ａ）
特開昭５７－１７１３９８（ＪＰ，Ａ）
特開昭６１－２５９３００（ＪＰ，Ａ）
特開平５－１８１４９１（ＪＰ，Ａ）
特開平３－２３３５００（ＪＰ，Ａ）
特開平３－２０８１００（ＪＰ，Ａ）
特開平３－１０３００（ＪＰ，Ａ）