(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第4586386号
(P4586386)
(45)発行日　平成22年11月24日(2010.11.24)　(24)登録日　平成22年9月17日(2010.9.17)

(51)Int.Cl.　F I
G10L 13/06 (2006.01)　G10L 13/06　240C

請求項の数 12　(全 12 頁)

(21)出願番号　特願2004-73977(P2004-73977)
(22)出願日　平成16年3月16日(2004.3.16)
(65)公開番号　特開2005-265895(P2005-265895A)
(43)公開日　平成17年9月29日(2005.9.29)
審査請求日　平成19年2月26日(2007.2.26)

(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
(74)代理人　100099933
弁理士　清水　敏
(72)発明者　西澤　信行
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
(72)発明者　河井　恒
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

審査官　井上　健一

最終頁に続く

(54)【発明の名称】素片接続型音声合成装置及び方法

(57)【特許請求の範囲】
【請求項１】
合成音声の目標と音声素片候補との間で、複数のサブコストを含むコストを算出し、当該コストに基づいて音声素片データベースから音声素片を選択し接続することにより音声合成を行なう素片接続型音声合成装置であって、
前記音声素片データベースに含まれる音声素片候補から、前記複数のサブコストのうちの一部のみを用い、２以上の段階に分けて複数の音声素片候補列を選択するための多段予備選択手段と、
前記合成音声の目標との間で、前記複数のサブコストを全て含んで算出されるコストが所定の条件を充足する一つの音声素片候補列を、前記多段予備選択手段により予備的に選択された前記複数の音声素片候補列から選択するための選択手段と、
前記選択手段により選択された音声素片候補列の音声波形を前記合成器指令に従って接続し合成音声波形を出力するための接続手段とを含む、素片接続型音声合成装置。
【請求項２】
前記多段予備選択手段は、前記音声素片データベースに含まれる音声素片候補から、前記複数のサブコストのうちの一部のみを用い、２以上の段階に分けて、かつ後段の予備選択では前段の予備選択で用いられたサブコストより多数種類のサブコストを用いて予備選択を行なって、複数の音声素片候補列を選択するための手段を含む、請求項１に記載の素片接続型音声合成装置。
【請求項３】

前記多段予備選択手段は、前記音声素片データベースに含まれる音声素片候補から、前記複数のサブコストのうちの一部のみを用い、２以上の段階に分けて、かつ後段の予備選択では、前段の予備選択で用いられたサブコストより多数種類のサブコストであって、かつ前段の予備選択で用いられたサブコストを含むサブコストを用いて予備選択を行なって、複数の音声素片候補列を選択するための手段を含む、請求項２に記載の素片接続型音声合成装置。

【請求項４】

前記多段予備選択手段は、

前記合成音声の目標と前記音声素片データベース中の各音声素片候補との間で、第１のサブコストを算出し、算出された第１のサブコストを用いて複数の音声素片候補列を選択するための第１段の予備選択手段と、

前記第１のサブコストと、前記第１のサブコストと異なる第２のサブコストとの双方を用いて、前記第１段の予備選択手段により選択された複数の音声素片候補列の中から複数の音声素片候補列を選択するための第２段の予備選択手段とを含む、請求項１に記載の素片接続型音声合成装置。

【請求項５】

前記第１段の予備選択手段は、

前記合成音声の目標と前記音声素片データベース中の音声素片候補からなる各音声素片候補列との間で、第１のサブコストを算出するための第１のサブコスト算出手段と、

前記第１のサブコスト算出手段により算出された第１のサブコストを記憶するための第１のサブコスト記憶手段と、

前記第１のサブコスト算出手段により算出された第１のサブコストが所定のしきい値よりも小さな音声素片候補列を選択するための手段とを含む、請求項４に記載の素片接続型音声合成装置。

【請求項６】

前記多段予備選択手段は、

前記合成音声の目標と前記音声素片データベース中の音声素片候補からなる各音声素片候補列との間で、ターゲットコストのみからなる第１のサブコストを算出し、算出された第１のサブコストを用いて複数の音声素片候補列を選択するための第１段の予備選択手段と、

前記第１のサブコストと、接続コストを含む第２のサブコストとの双方を用いて、前記第１段の予備選択手段により選択された複数の音声素片候補列の中から複数の音声素片候補列を選択するための第２段の予備選択手段とを含む、請求項１に記載の素片接続型音声合成装置。

【請求項７】

合成音声の目標と音声素片候補との間で、複数のサブコストを含むコストを算出し、当該コストに基づいて音声素片データベースから音声素片を選択し接続することにより音声合成を行なう素片接続型音声合成方法であって、

前記音声素片データベースに含まれる音声素片候補から、前記複数のサブコストのうちの一部のみを用い、２以上の段階に分けて複数の音声素片候補列を選択する多段予備選択ステップと、

前記合成音声の目標との間で、前記複数のサブコストを全て含んで算出されるコストが所定の条件を充足する一つの音声素片候補列を、前記多段予備選択ステップにおいて予備的に選択された前記複数の音声素片候補列から選択する選択ステップと、

前記選択ステップにおいて選択された音声素片候補列の音声波形を前記合成器指令に従って接続し合成音声波形を出力する接続ステップとを含む、素片接続型音声合成方法。

【請求項８】

前記多段予備選択ステップは、前記音声素片データベースに含まれる音声素片候補から、前記複数のサブコストのうちの一部のみを用い、２以上の段階に分けて、かつ後段の予備選択では前段の予備選択で用いられたサブコストより多数種類のサブコストを用いて予備

選択を行なって、複数の音声素片候補列を選択するステップを含む、請求項７に記載の素片接続型音声合成方法。
【請求項９】
前記多段予備選択ステップは、前記音声素片データベースに含まれる音声素片候補から、前記複数のサブコストのうちの一部のみを用い、２以上の段階に分けて、かつ後段の予備選択では、前段の予備選択で用いられたサブコストより多数種類のサブコストであって、かつ前段の予備選択で用いられたサブコストを含むサブコストを用いて予備選択を行なって、複数の音声素片候補列を選択するステップを含む、請求項８に記載の素片接続型音声合成方法。
【請求項１０】
前記多段予備選択ステップは、
前記合成音声の目標と前記音声素片データベース中の各音声素片候補との間で、第１のサブコストを算出し、算出された第１のサブコストを用いて複数の音声素候補列を選択する第１段の予備選択ステップと、
前記第１のサブコストと、前記第１のサブコストと異なる第２のサブコストとの双方を用いて、前記第１段の予備選択ステップにおいて選択された複数の音声素片候補列の中から複数の音声素片候補列を選択する第２段の予備選択ステップとを含む、請求項７に記載の素片接続型音声合成方法。
【請求項１１】
前記第１段の予備選択ステップは、
前記合成音声の目標と前記音声素片データベース中の音声素片からなる各音声素片候補列との間で、第１のサブコストを算出する第１のサブコスト算出ステップと、
前記第１のサブコスト算出ステップにおいて算出された第１のサブコストを、第１のサブコスト記憶手段に記憶させるステップと、
前記第１のサブコスト算出ステップにおいて算出された第１のサブコストが所定のしきい値よりも小さな音声素片候補列を選択するステップとを含む、請求項１０に記載の素片接続型音声合成方法。
【請求項１２】
前記多段予備選択ステップは、
前記合成音声の目標と前記音声素片データベース中の音声素片からなる各音声素片候補列との間で、ターゲットコストのみからなる第１のサブコストを算出し、算出された第１のサブコストを用いて複数の音声素片候補列を選択する第１段の予備選択ステップと、
前記第１のサブコストと、接続コストを含む第２のサブコストとの双方を用いて、前記第１段の予備選択ステップにおいて選択された複数の音声素片候補列の中から複数の音声素片候補列を選択する第２段の予備選択ステップとを含む、請求項７に記載の素片接続型音声合成方法。
【発明の詳細な説明】
【技術分野】
【０００１】
この発明は音声合成装置に関し、特に、所定のコスト関数に基づいて音声素片を選択し接続することにより合成器指令に合致した音声合成を行なう音声合成装置に関する。
【背景技術】
【０００２】
音声認識、音声合成は、人間とコンピュータを用いた諸システムとのインターフェースを実現する技術として重要である。これらと人工知能技術とを併用することにより、利用者は相手がコンピュータシステムであることを意識せずに様々なサービスを利用することができる。
【０００３】
中でも音声合成は、人間に対するシステム出力のためのインターフェースとしてその重要性は大きい。人間は、合成された音声の不自然さを敏感に感じ取る。合成された音声が

不自然であると利用者が感じると、発話にも影響を及ぼし、その結果、人間とシステムとの間の対話がうまく行かなくなるおそれもある。

【０００４】

最近の音声合成技術としては、予め人間の発話を多数集めて語・音節・音素等を単位とする音声素片を音素ラベルと関連付けてデータベース化しておき、合成時には、指定された語・音節・音素等に対応する音声素片の中から、最も適切と思われるものを選択して接続するものが知られている。これを素片接続型音声合成と呼ぶ。なお、音素ラベルとは、通常は各音素の音素記号とその開始・終了時刻を記述したものをいう。これに加えて、その区間におけるＭＦＣＣ（Ｍｅｌ－Ｆｒｅｑｕｅｎｃｙ Ｃｅｐｓｔｒｕｍ Ｃｏｅｆｆｉｃｉｅｎｔ）、基本周波数（Ｆ０）等の音響特徴量、さらに前後の素片の音素記号を含む場合もある。

【０００５】

素片接続型音声合成では、与えられた合成目標を基準として、いかにして適切な音声素片をデータベース中から取出すかが問題となる。

【０００６】

合成目標を構成するデータは、典型的には音素と、Ｆ０、持続時間、ＭＦＣＣ、及びパワー等の音声特徴量とを含む。これらを以下「合成器指令」と呼ぶ。

【０００７】

素片接続型音声合成では、合成器指令と音声素片のＦ０、持続時間、ＭＦＣＣ、パワー等とのずれ、及び接続に伴う自然劣化を表現するための「コスト」と呼ばれる評価関数を定義し、コストを最小とする音声素片を求めることにより、最適な音声素片系列を決定する。

【０００８】

本件出願の出願人は、上記した「コスト」を、それぞれある音声の特徴に対応するような「サブコスト」に分解し、それらを結合したもの（例えば線形和）により定義した素片接続型音声合成を提案している。例えば特許文献１を参照されたい。

【０００９】

サブコストには、物理量から計算されるものと、シンボリックな情報から事前に作成した規則から基づき得られるものとがある。前者は、複数のサンプル値に対する非線形演算であることも多く、その計算量は相対的に大きい。後者は、単純なテーブル参照の形であることが多く、テーブル参照で実現される場合にはサブコスト計算に必要な計算量は非常に少ない。

【００１０】

以上はあくまで一例であるが、この例に限らず、各サブコストの計算量はその種類により大きなばらつきがある場合が多い。

【００１１】

一方、上記とは別に、サブコストは、ターゲットコストに関係するものと接続コストに属するものとの二つに大別することもできる。ターゲットコストは、合成目標と素片候補との間の誤差を表す。接続コストは、合成音声において隣接する素片間の誤差（不連続性）を表す。

【００１２】

素片接続型音声合成では、コストの最小化に基づく素片選択が行なわれるが、特に素片候補数が多い場合にはコストの計算に要する計算量が問題となる。

【００１３】

最小コストとなる素片候補系列の推定において、可能な組合せのコストを全て調べることは、組合せ爆発により記憶容量・計算時間の双方において非現実的である。そこで、各時刻の素片候補を予備選択により絞り込む方法が考えられる。この際、計算量を考えて、前後の素片関係の影響を受けないターゲットコストに属するサブコストのみで予備選択を行なう方法が有力である。

【００１４】

【特許文献１】特開２００３－２０８１８８号公報（段落００１４～００４７）
【発明の開示】
【発明が解決しようとする課題】
【００１５】
しかし、合成音声に対する接続コストの影響が比較的大きいことから、予備選択段階でターゲットコストのみに基づいて絞り込むことができる候補数には限界がある。ターゲットコストのみに基づいてあまり候補を絞り込むと、高品質な音声合成を行なうことが可能な候補が捨てられてしまうおそれがあるためである。その結果、本選択時の計算量の削減にも限界がある。
【００１６】
それゆえに、本発明の目的は、高品質な音声合成が可能で、かつ選択のための計算量を削減できる素片接続型音声合成装置及び方法を提供することである。
【課題を解決するための手段】
【００１７】
本発明の第１の局面に係る素片接続型音声合成装置は、合成音声の目標と音声素片候補との間で、複数のサブコストを含むコストを算出し、当該コストに基づいて音声素片データベースから音声素片を選択し接続することにより音声合成を行なう素片接続型音声合成装置であって、音声素片データベースに含まれる音声素片候補から、複数のサブコストのうちの一部のみを用い、２以上の段階に分けて複数の音声素片候補を選択するための多段予備選択手段と、合成音声の目標との間で、複数のサブコストを全て含んで算出されるコストが所定の条件を充足する一つの音声素片候補を、予備選択手段により予備的に選択された複数の音声素片候補から選択するための選択手段と、選択手段により選択された音声素片候補の音声波形を合成器指令に従って接続し合成音声波形を出力するための接続手段とを含む。
【００１８】
好ましくは、多段予備選択手段は、音声素片データベースに含まれる音声素片候補から、複数のサブコストのうちの一部のみを用い、２以上の段階に分けて、かつ後段の予備選択では前段の予備選択で用いられたサブコストより多数種類のサブコストを用いて予備選択を行なって、複数の音声素片候補を選択するための手段を含む。
【００１９】
より好ましくは、多段予備選択手段は、音声素片データベースに含まれる音声素片候補から、複数のサブコストのうちの一部のみを用い、２以上の段階に分けて、かつ後段の予備選択では、前段の予備選択で用いられたサブコストより多数種類のサブコストであって、かつ前段の予備選択で用いられたサブコストを含むサブコストを用いて予備選択を行なって、複数の音声素片候補を選択するための手段を含む。
【００２０】
多段予備選択手段は、合成音声の目標と音声素片データベース中の各音声素片候補との間で、第１のサブコストを算出し、算出された第１のサブコストを用いて複数の音声素片を選択するための第１段の予備選択手段と、第１のサブコストと、第１のサブコストと異なる第２のサブコストとの双方を用いて、第１段の予備選択手段により選択された複数の音声素片の中から複数の音声素片を選択するための第２段の予備選択手段とを含んでもよい。
【００２１】
さらに好ましくは、第１段の予備選択手段は、合成音声の目標と音声素片データベース中の各音声素片候補との間で、第１のサブコストを算出するための第１のサブコスト算出手段と、第１のサブコスト算出手段により算出された第１のサブコストを記憶するための第１のサブコスト記憶手段と、第１のサブコスト算出手段により算出された第１のサブコストが所定のしきい値よりも小さな音声素片候補を選択するための手段とを含む。
【００２２】
好ましくは、多段予備選択手段は、合成音声の目標と音声素片データベース中の各音声

素片候補との間で、ターゲットコストのみからなる第１のサブコストを算出し、算出された第１のサブコストを用いて複数の音声素片を選択するための第１段の予備選択手段と、第１のサブコストと、接続コストを含む第２のサブコストとの双方を用いて、第１段の予備選択手段により選択された複数の音声素片からなる音声素片系列の中から複数の音声素片系列を選択するための第２段の予備選択手段とを含む。

【００２３】

本発明の第２の局面に係る素片接続型音声合成方法は、合成音声の目標と音声素片候補との間で、複数のサブコストを含むコストを算出し、当該コストに基づいて音声素片データベースから音声素片を選択し接続することにより音声合成を行なう素片接続型音声合成方法であって、音声素片データベースに含まれる音声素片候補から、複数のサブコストのうちの一部のみを用い、２以上の段階に分けて複数の音声素片候補を選択する多段予備選択ステップと、合成音声の目標との間で、複数のサブコストを全て含んで算出されるコストが所定の条件を充足する一つの音声素片候補を、予備選択ステップにおいて予備的に選択された複数の音声素片候補から選択する選択ステップと、選択ステップにおいて選択された音声素片候補の音声波形を合成器指令に従って接続し合成音声波形を出力する接続ステップとを含む。

【００２４】

好ましくは、多段予備選択ステップは、音声素片データベースに含まれる音声素片候補から、複数のサブコストのうちの一部のみを用い、２以上の段階に分けて、かつ後段の予備選択では前段の予備選択で用いられたサブコストより多数種類のサブコストを用いて予備選択を行なって、複数の音声素片候補を選択するステップを含む。

【００２５】

さらに好ましくは、多段予備選択ステップは、音声素片データベースに含まれる音声素片候補から、複数のサブコストのうちの一部のみを用い、２以上の段階に分けて、かつ後段の予備選択では、前段の予備選択で用いられたサブコストより多数種類のサブコストであって、かつ前段の予備選択で用いられたサブコストを含むサブコストを用いて予備選択を行なって、複数の音声素片候補を選択するステップを含む。

【００２６】

多段予備選択ステップは、合成音声の目標と音声素片データベース中の各音声素片候補との間で、第１のサブコストを算出し、算出された第１のサブコストを用いて複数の音声素片を選択する第１段の予備選択ステップと、第１のサブコストと、第１のサブコストと異なる第２のサブコストとの双方を用いて、第１段の予備選択ステップにおいて選択された複数の音声素片の中から複数の音声素片を選択する第２段の予備選択ステップとを含んでもよい。

【００２７】

さらに好ましくは、第１段の予備選択ステップは、合成音声の目標と音声素片データベース中の各音声素片候補との間で、第１のサブコストを算出する第１のサブコスト算出ステップと、第１のサブコスト算出ステップにおいて算出された第１のサブコストを、第１のサブコスト記憶手段に記憶させるステップと、第１のサブコスト算出ステップにおいて算出された第１のサブコストが所定のしきい値よりも小さな音声素片候補を選択するステップとを含む。

【００２８】

好ましくは、多段予備選択ステップは、合成音声の目標と音声素片データベース中の各音声素片候補との間で、ターゲットコストのみからなる第１のサブコストを算出し、算出された第１のサブコストを用いて複数の音声素片を選択する第１段の予備選択ステップと、第１のサブコストと、接続コストを含む第２のサブコストとの双方を用いて、第１段の予備選択ステップにおいて選択された複数の音声素片からなる音声素片系列の中から複数の音声素片系列を選択する第２段の予備選択ステップとを含む。

【発明を実施するための最良の形態】

【００２９】

［第１の実施の形態］

図１に、本発明の第１の実施の形態に係る音声合成システム２０のブロック図を示す。図１を参照して、この音声合成システム２０は、従来と同様の音声素片ＤＢ３４と、合成目標となるテキストを分析した結果得られる合成器指令３６を入力として受け、音声素片ＤＢ３４に含まれる拡張された音声素片から適切な音声素片を選択し接続して合成音声波形４０を出力するための音声合成装置３８とを含む。

【００３０】

音声合成装置３８は、合成器指令３６を受け、合成器指令３６により指定された音声素片のうちで、後述するように多段の予備選択を行なって予備選択候補群６２を選択するための多段予備選択部６０と、合成器指令３６を受け、予備選択候補群６２から全サブコストを用いて計算したコストの最も小さな素片を選択するための素片選択部６４と、素片選択部６４により選択された音声素片を接続して合成音声波形４０を出力するための接続部６６とを含む。なお、予備選択候補群６２は素片の選択のみに用いられるので、コスト計算に必要な特徴量のみを含み、音声素片データそのものは含まない。接続部６６は、素片選択部６４により選択された素片の音声素片データを音声素片ＤＢ３４を参照して得ることになる。

【００３１】

本実施の形態で使用されるサブコストは、基本周波数（Ｆ０）誤差、継続時間長誤差、ＭＦＣＣ誤差、Ｆ０不連続誤差、ＭＦＣＣ不連続誤差、音素環境誤差にそれぞれ対応する６種類のサブコストを含む。これらのうち、前３者はターゲットコストに属し、後３者は接続コストに属する。

【００３２】

本実施の形態に係る素片選択部６４によるコスト計算では、コストC_0は以下のようにしてサブコストから計算される。

【００３３】

【数１】

$$C_o=\left(\sum_{i1=1}^{N_1}(w_{i1}C_{i1})^{p_1}\right)^{\frac{1}{p_1}}+\left(\sum_{i2=1}^{N_2}(w_{i2}C_{i2})^{p_2}\right)^{\frac{1}{p_2}} \qquad (1)$$

ただし、C_{i1}（ｉ1＝１～３）はターゲットサブコスト、C_{i2}（ｉ2＝１～３）は接続コスト、w_{i1}（ｉ1＝１～３）はターゲットサブコスト間に定義された重み、w_{i2}（ｉ2＝１～３）は接続サブコスト間に定義された重み、p_1及びp_2はそれぞれ、ターゲットコストと接続コスト間に定義された重みである。ただし、本実施の形態では後述するように多段予備選択における計算量を削減するため、p_1及びp_2はいずれも１とする。

【００３４】

一般的に、音素環境誤差のサブコストは比較的単純なテーブル参照である。したがってその計算量は非常に小さい。それ以外については、サブコストの計算量は比較的大きい。例えばＭＦＣＣは多次元量であるため、そのサブコストの計算に要する時間は他のサブコストより大きくなる。

【００３５】

図１を参照して、多段予備選択部６０は、４つの予備選択部７０、８０、９０及び１００を含む。予備選択をどのような順番でどのサブコストに基づいて行なうかは、アプリケーション、より具体的には各サブコストに対し予想される計算量の相違により異なる。Ｆ０誤差、継続時間長誤差に関するサブコスト計算が比較的小さい場合には、図１に示すような構成が考えられる。４つの予備選択部７０、８０、９０及び１００、並びに素片選択部６４の機能は以下のとおりである。なお、接続コストに関するサブコストが予備選択コストに含まれる場合、コストは前後の素片の影響を受ける。したがって、その時点での予備選択コスト関数を最小化する解について、各時刻において独立に素片候補を予備選択するのではなく、素片候補の選択系列のＮ－ベスト解を得ておく必要がある。その後は、そ

のＮ－ベスト解について後段の予備選択関数で再度コスト計算を行なってその結果のＮ－ベスト解を得る、という処理を繰返す必要がある。

【００３６】

第１の予備選択部７０：合成器指令３６を受け、音声素片ＤＢ３４中の素片候補から各時刻におけるＦ０誤差、継続時間長誤差による予備選択をして第１の候補群７２を出力する。

【００３７】

第２の予備選択部８０：第１の候補群７２中の素片から、各時刻におけるＦ０誤差、継続時間長誤差、ＭＦＣＣ誤差による予備選択をして第２の候補群８２を出力する。

【００３８】

第３の予備選択部９０：第２の候補群８２中の素片から、各時刻におけるＦ０誤差、継続時間長誤差、ＭＦＣＣ誤差、及び音素環境誤差を考慮したＮ－ベスト探索を行ない第３の候補群（選択系列群）９２を出力する。

【００３９】

第４の予備選択部１００：第３の候補群９２中の素片候補に、Ｆ０誤差、継続時間長誤差、ＭＦＣＣ誤差、音素環境誤差、及びＦ０不連続誤差を考慮したＮ－ベスト探索を行ない、予備選択候補群（選択系列群）６２を出力する。

【００４０】

素片選択部６４：予備選択候補群６２に含まれるＮ－ベスト選択系列に対して、全てのサブコストを考慮して行なう１－ベスト探索を行ない、素片を一つ選択し接続部６６に与える。

【００４１】

なお、Ｎ－ベスト解はビームサーチ又はＮ－ベストＤＰ（Ｄｙｎａｍｉｃ　Ｐｒｏｇｒａｍｍｉｎｇ）サーチにより行なうことができる。（ここでＮ－ベストＤＰサーチとは、ＤＰ探索における各ノードでＮ－ベスト解を保持する方法のことをいう。通常のＤＰサーチは各ノードで１－ベストの解のみを保持している。）

【００４２】

ここで、ビームサーチについては、Ｎ－ベスト解で選択される候補系列の数Ｎに対して、ビーム幅が小さいほど最適解に近い解が得られる可能性が小さくなる。一方、Ｎ－ベストＤＰサーチでは、各ノードが保持するＮ－ベスト解の数が少ないほど、最適解が得られる可能性が低くなる。（ここで、各ノードにおけるＮ－ベスト解の数が、最終的に必要となるＮ－ベスト解の数と同数以上であれば、解が真のＮ－ベスト解であることは保証される。しかし、多段選択の途中におけるＮ－ベスト解の中に真の最適解が含まれている保証はなく、計算途中で真のＮ－ベスト解を得ること自体にはそれほど意味はない。）ただし、素片候補が大量に存在する場合には、仮に最終的に最適解でない解が得られたとしても実用上十分な品質が得られる可能性が高い。

【００４３】

この実施の形態では、前段の予備選択部で算出されたサブコストは、後段の予備選択部でも素片選択に使用される。したがって、サブコストが式（１）で表され、かつp_1及びp_2がいずれも１として設計した場合（すなわちコストがサブコストの線形和で表される場合）、前段の予備選択部で算出したサブコストをそのまま次の予備選択部でのコスト計算に用いることができる。そのために多段予備選択部６０は、それぞれ予備選択部７０、８０、９０及び１００で行なわれたサブコスト計算の結果を記憶するための第１～第４のコスト記憶部７４、８４、９４及び１０４をさらに含む。これら第１～第４のコスト記憶部７４、８４、９４及び１０４に記憶されたサブコストは、それぞれ予備選択部８０、９０、及び１００並びに素片選択部６４に与えられ、コスト計算に用いられる。

【００４４】

この音声合成システム２０は以下のように動作する。まず合成器指令３６が音声合成装置３８に与えられる。多段予備選択部６０の第１の予備選択部７０は、合成器指令３６に基づいて、合成器指令３６により指定された音素に対応する音声素片であってかつＦ０誤

差及び継続時間長誤差により算出されたサブコストの線形和が所定のしきい値以下であるものを音声素片ＤＢ３４から抽出し、第１の候補群７２として出力する。このときのサブコストの計算結果は第１のコスト記憶部７４に記憶される。

【００４５】

第２の予備選択部８０は、合成器指令３６に基づいて、Ｆ０誤差、継続時間長誤差及びＭＦＣＣ誤差により算出されたサブコストの線形和が所定のしきい値以下であるものを第１の候補群７２から抽出し、第２の候補群８２として出力する。このとき、第２の予備選択部８０は、第１のコスト記憶部７４に記憶されたＦ０誤差及び継続時間長誤差により算出されたサブコストをサブコスト計算に用いる。したがって実質的にはＭＦＣＣによるサブコスト計算のみが行なわれる。第２の予備選択部８０によるサブコストの計算結果は第２のコスト記憶部８４に記憶される。

【００４６】

第３の予備選択部９０は、合成器指令３６に基づいて、Ｆ０誤差、継続時間長誤差、ＭＦＣＣ誤差、及び音素環境誤差に基づいて算出されたサブコストの線形和に基づき、素片候補のＮ－ベスト解を第２の候補群８２から抽出し、第３の候補群９２として出力する。このとき、第３の予備選択部９０は、第２のコスト記憶部８４に記憶されたＦ０誤差、継続時間長誤差、及びＭＦＣＣ誤差により算出されたサブコストをサブコスト計算に用いる。したがって、実質的には第３の予備選択部９０では音素環境誤差のみに基づくサブコスト計算が行なわれる。第３の予備選択部９０によるサブコストの計算結果は第３のコスト記憶部９４に記憶される。

【００４７】

第４の予備選択部１００は、合成器指令３６に基づいて、Ｆ０誤差、継続時間長誤差、ＭＦＣＣ誤差、音素環境誤差、及びＦ０不連続誤差に基づいて算出されたサブコストの線形和に基づき、素片候補のＮ－ベスト解を第３の候補群９２から抽出し、予備選択候補群６２として出力する。このとき、第４の予備選択部１００は、第３のコスト記憶部９４に記憶されたＦ０誤差、継続時間長誤差、ＭＦＣＣ誤差、及び音素環境誤差により算出されたサブコストをサブコスト計算に用いる。したがって、実質的には第４の予備選択部１００ではＦ０不連続誤差のみに基づくサブコスト計算が行なわれる。第４の予備選択部１００によるサブコストの計算結果は第４のコスト記憶部１０４に記憶される。

【００４８】

素片選択部６４は、合成器指令３６を受け、予備選択候補群６２に含まれる音声素片のうち、式（１）により算出されるコストが最も小さなものを選択して接続部６６に与える。

【００４９】

接続部６６は、素片選択部６４により選択された音声素片に対応する音声素片データを音声素片ＤＢ３４から読出し、音声が滑らかに接続されるように変形して接続し、合成音声波形４０として出力する。

【００５０】

多段予備選択部６０により合成器指令３６に対し計算されるサブコストが小さなものを予備的に選択しておくため、素片選択部６４が式（１）にしたがって素片選択を行なう際のコスト計算の計算量は少なくて済む。多段予備選択部６０内の各予備選択部７０、８０、９０及び１００によるサブコストの算出では、それぞれ前段でのサブコスト計算の結果を用いる。したがって各予備選択部７０、８０、９０及び１００における計算量は実質的には少なくて済む。

【００５１】

また、予備選択部７０、８０、９０及び１００による予備選択では、徐々に選択の基準が細かくなっていくため、素片候補の限定は徐々に行なわれる。その結果、予備選択の段階で適切な素片候補が捨てられる危険性も低くなる。多段予備選択で得られた予備選択候補群６２の中からコスト最小の音声素片を選択して接続した場合、接続時の変形による品質低下はほとんどない。その結果、最終的に得られる合成音声波形４０にも、音声素片の

接続による品質低下はほとんどない。

【００５２】

なお、上記した実施の形態では、予備選択部７０、８０、９０及び１００による４段階の多段予備選択を行なっているが、予備選択の各段階でのサブコスト計算及び段数がこの実施の形態に限定されないことはもちろんである。アプリケーションにより、種々の形で多段予備選択を行なうことができる。

【００５３】

例えば、Ｆ０誤差、継続時間長誤差による予備選択での計算量が比較的大きいと思われる場合には、ＭＦＣＣ不連続誤差による音質への影響が比較的小さいことを考慮し、次のような多段予備選択を行なうことも考えられる。

【００５４】

（１） 音素環境誤差を考慮したＮ－ベスト探索。

【００５５】

（２） （１）により得られたＮ－ベスト解に対して、Ｆ０誤差、継続時間長誤差、及び音素環境誤差を考慮したＮ－ベスト探索。

【００５６】

（３） （２）により得られたＮ－ベスト解に対して、Ｆ０誤差、継続時間長誤差、Ｆ０不連続誤差、及び音素環境誤差を考慮したＮ－ベスト探索。

【００５７】

（４） （３）により得られたＮ－ベスト解に対して、Ｆ０誤差、継続時間長誤差、Ｆ０不連続誤差、ＭＦＣＣ不連続誤差、及び音素環境誤差を考慮したＮ－ベスト探索。

【００５８】

（５） （４）により得られたＮ－ベスト解に対して、Ｆ０誤差、継続時間長誤差、Ｆ０不連続誤差、ＭＦＣＣ不連続誤差、ＭＦＣＣ誤差、及び音素環境誤差を考慮したＮ－ベスト探索。

【００５９】

（６） （５）により得られたＮ－ベスト解に対して、全てのサブコストを考慮した１－ベスト探索による素片選択。

【００６０】

上記した実施の形態の説明では、式（１）における重みp_1及びp_2の値をいずれも１として説明した。しかし、本発明はそのような実施の形態には限定されず、重みp_1及びp_2の値のいずれか、又は双方を１以外の値としてもよい。

【００６１】

また、サブコスト関数も式（１）に示すものには限定されず、設計思想により様々なサブコスト関数を考えることができる。その場合も、多段予備選択でのサブコスト計算を考慮して、後段のサブコスト計算では前段のサブコスト計算の結果を使用できるようにすると効率がよい。

【００６２】

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および範囲内でのすべての変更を含む。

【図面の簡単な説明】

【００６３】

【図１】本発明の第１の実施の形態にかかる音声合成システム２０のブロック図である。

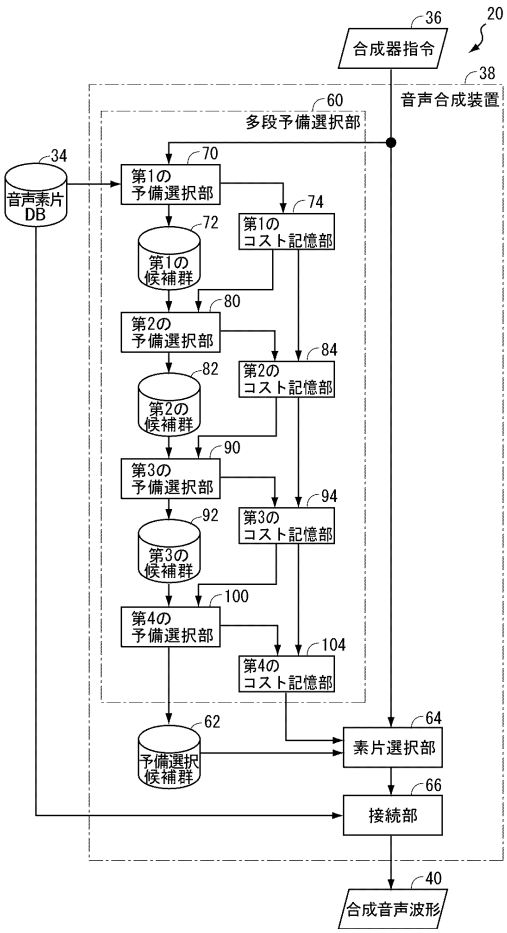
【符号の説明】

【００６４】

２０ 音声合成システム、３４ 音声素片ＤＢ、３６ 合成器指令、３８ 音声合成装置、４０ 合成音声波形、６０ 多段予備選択部、６２ 予備選択候補群、６４ 素片選択部、６６ 接続部、７０，８０，９０，１００ 予備選択部、７２，８２，９２ 候補

群、７４，８４，９４，１０４　コスト記憶部

【図１】



フロントページの続き

(56)参考文献 特開平０８－２６３０９５（ＪＰ，Ａ）
特開平０８－２４８９７２（ＪＰ，Ａ）
特開２００３－２０８１８８（ＪＰ，Ａ）
戸田 智基，波形接続型音声合成における知覚的評価に基づく素片選択サブコスト関数の最適化，電子情報通信学会技術研究報告，日本，社団法人電子情報通信学会，２００３年 8月１５日，Ｖｏｌ．１０３ Ｎｏ．２６４

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ１０Ｌ １３／０６