(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第3881970号
(P3881970)
(45)発行日　平成19年2月14日(2007.2.14)　(24)登録日　平成18年11月17日(2006.11.17)

(51)Int.Cl.　　F I
G10L 13/06 (2006.01)　　G10L 13/06　24OC
G10L 13/04 (2006.01)　　G10L 13/04　Z

請求項の数 7　(全 13 頁)

(21)出願番号　特願2003-280402(P2003-280402)
(22)出願日　平成15年7月25日(2003.7.25)
(65)公開番号　特開2005-43828(P2005-43828A)
(43)公開日　平成17年2月17日(2005.2.17)
審査請求日　平成16年6月24日(2004.6.24)

（出願人による申告）国等の委託研究の成果に係る特許出願（平成１５年度通信・放送機構、研究テーマ「大規模コーパスベース音声対話翻訳技術の研究開発」に関する委託研究、産業活力再生特別措置法第３０条の適用を受けるもの）

(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　戸田　智基
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
(72)発明者　津崎　実
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

最終頁に続く

(54)【発明の名称】知覚試験用音声データセット作成装置、コンピュータプログラム、音声合成用サブコスト関数の最適化装置、及び音声合成装置

(57)【特許請求の範囲】
【請求項１】
それぞれ単位波形素片に分離可能な複数の発話音声データを含む音声コーパスから、発話音声データの、所定の複数種類の特徴量の各々について算出されるサブコストを統合して得られるコスト計算によって選択した発話単位を接続して音声を合成する際の、前記複数種類の特徴量のうちの予め定める第１の種類の特徴量と、前記所定の複数種類の特徴量を用いたコスト計算により合成された音声の自然さに対する人間による知覚評価との間のマッピングを行なう際に使用される知覚試験用音声データセットを作成するための知覚試験用音声データセット作成装置であって、
前記音声コーパスに含まれる発話音声データの各々について、任意の単位波形素片を、前記音声コーパスに含まれる発話音声データが持つ、前記任意の単位波形素片と同じ音素を含む単位波形素片で置換する事により、単位波形素片が置換された置換後の発話音声データの集合を作成するための手段と、
前記置換後の発話音声データの各々について、前記複数種類の特徴量を算出するための特徴量算出手段と、
前記特徴量算出手段により算出された前記複数種類の特徴量に基づき、前記第１の種類の特徴量の変動があるしきい値以上であり、かつ前記複数種類の特徴量のうち、前記第１の種類の特徴量以外の特徴量の変動がいずれもあるしきい値未満であるような発話音声データの集合を、前記置換後の発話音声データの集合から抽出するための手段とを含む、知覚試験用音声データセット作成装置。

【請求項２】
前記置換後の発話音声データの集合を作成するための手段は、
前記音声コーパスに含まれる発話音声データの各々について、
当該発話音声データに含まれる任意の単位波形素片を選択するための手段と、
前記選択するための手段により選択された単位波形素片と同じ音素を含む単位波形素片を含む、別の発話音声データを前記音声コーパスの中で特定するための手段と、
前記特定するための手段により特定された前記別の発話音声データに含まれる、前記選択された単位波形素片と同じ音素を含む単位波形素片で、前記選択された単位波形素片を置換するための手段と、
前記置換するための手段による置換が行なわれた発話音声データを予め定める記憶媒体に記憶させるための手段とを含み、
前記記憶媒体に記憶される発話音声データにより前記知覚試験用音声データセットが形成される、請求項１に記載の知覚試験用音声データセット作成装置。
【請求項３】
コンピュータにより実行されると、当該コンピュータを請求項１又は請求項２に記載の知覚試験用音声データセット作成装置として動作させる、コンピュータプログラム。
【請求項４】
請求項１又は請求項２に記載の知覚試験用音声データセット作成装置により作成される知覚試験用音声データセットに含まれる発話音声データによる音声と、音声コーパス中の、前記発話音声データを作成する基になった発話音声データによる音声とを対比して被験者に提示し、前記知覚試験用音声データセットに含まれる発話音声データによる音声の自然さに関する被験者による評価点の入力を受けるための手段と、
前記評価点を受けるための手段により得られた評価点を、前記知覚試験用音声データセットに含まれる発話音声データから算出される前記第１の種類の特徴量に対してプロットして得られた点列との間の自乗誤差の和を最小化する様に、前記第１の種類の特徴量から音声の自然さに対する知覚評価へのマッピング関数を最適化するための手段とを含む、音声合成用サブコスト関数の最適化装置。
【請求項５】
コンピュータにより実行されると、請求項４に記載の音声合成用サブコスト関数の最適化装置として当該コンピュータを動作させる、コンピュータプログラム。
【請求項６】
請求項４に記載の音声合成用サブコスト関数の最適化装置と、
前記サブコスト関数の最適化装置により最適化されるサブコスト関数を含んで定義されるコスト関数を用いて、入力音声テキストの音素に対する波形を音声コーパスから選択し接続する事により出力音声波形を合成するための音声合成手段とを含む、音声合成装置。
【請求項７】
コンピュータにより実行されると、請求項６に記載の音声合成装置として当該コンピュータを動作させる、コンピュータプログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
この発明は音声合成技術に関し、特に、音声コーパスから選択された単位波形素片を接続する事により自然な音声を合成するための波形接続型音声合成技術に関する。
【背景技術】
【０００２】
近年、人間と機械との間のコミュニケーションを実現するための技術の重要性が増している。それらの中でも、音声によるコミュニケーションのための音声認識及び音声合成の技術の進展が著しい。音声認識では、話者を特定する事なくかなりの精度で音声認識を行なう技術が開発されている。音声合成の実用化はさらに進んでおり、かなり自然な発音でテキストを音声に変換する技術が普及しつつある。

【０００３】
音声合成技術、特にテキスト音声合成（ＴＴＳ：Ｔｅｘｔ－Ｔｏ－Ｓｐｅｅｃｈ）で近年主流となっているのは、音声コーパスを用いたコーパスベースのものである。図８に、コーパスベースの音声合成システムのブロック図を示す。図８を参照して、コーパスベースの音声合成システムでは、予め人間による自然な発話における音声の波形素片を音声コーパス３２としてコーパス化しておく。そして、入力テキスト３０が音声合成装置３４に与えられると、入力テキスト３０を構成する音声に対応する音声波形素片を何らかの基準によりこの音声コーパス３２から抽出し、それらを接続して出力音声波形３６を合成する（波形接続型音声合成）。

【０００４】
音声コーパスを用いた音声合成では、実際に人間による発声を用いているので、合成された音声に「機械音らしさ」を感じる事はあまりない。しかし、別々の発話を構成していた音声波形素片を接続するため、接続時の不自然さが残るという問題がある。この不自然さのため、合成音声の品質はまだ十分とはいえない。従って、音声波形素片を接続する際の不自然さを解消する技術が望まれている。

【０００５】
こうした不自然さを解消するために、合成に用いる音声波形素片をどの様にして選択するかが問題となる。通常、各音声波形素片に関連する何らかの音響特徴量を算出し、所定の条件に合致する音声波形素片が選択される。不自然さを小さくするためには、知覚特性に一致した尺度（コスト）を用いて素片選択を行なう事が重要である。

【０００６】
後掲の非特許文献２では、知覚特性を反映した「コスト関数」と呼ばれる関数を用いて候補の音声素片についてコストを算出し、その算出されたコストが最小となる波形素片を選択する。このようなコスト関数を用いて波形素片を選択する事で、より自然性の高い音声を合成できると期待される。

【０００７】
しかし、どのような物理尺度を用いれば、波形接続時の不自然さが解消されるかについての、物理尺度と合成音声の自然さとの対応関係は明らかでない。そのため非特許文献２では、コスト関数を様々な要因に対応する複数のサブコスト関数に分けている。

【０００８】
図９に、コスト関数とサブコスト関数との概念を示す。図９を参照して、コスト関数２００は、複数個のサブコスト関数２２０Ａ～２２０Ｎからなる。サブコスト関数２２０Ａ～２２０Ｎは、それぞれ対応の物理量（観測可能なもの）が与えられる事により、その関数としてサブコストを出力する。これらサブコストに重み（w_1～w_N）２２２Ａ～２２２Ｎを乗算し、加算（２２４）する事によりコスト２１０が算出される。

【０００９】
非特許文献２では、韻律に関するサブコスト関数、Ｆ０（フォルマント）の不連続に関するサブコスト関数、音素環境代替におけるサブコスト関数、スペクトルの不自然に関するサブコスト関数、音素の適合性に関するサブコスト関数を用いている。そして、これらサブコスト関数のうち、特に知覚評価との関係が比較的分かりやすい要因である音素環境代替に関しては、知覚評価と物理量との間のマッピングを行なっている。しかしその他の要因については知覚評価を用いていない。

【００１０】
【非特許文献１】河井　恒、津崎　実、枡田　剛志、岩澤　秀紀、「波形素片接続時の音素環境代替による自然性劣化の知覚評価」、電子情報通信学会技術研究報告、Vol. 2001-16, pp. 9-16, 2001.
【非特許文献２】戸田　智基、河井　恒、津崎　実、鹿野　清宏、「素片接続型日本語テキスト音声合成における音素単位とダイフォン単位に基づく素片選択」、電子情報通信学会論文誌、Vol. J85-D-II., No. 12, pp. 1760-1770, Dec. 2002.

【発明の開示】

【発明が解決しようとする課題】
【００１１】
非特許文献２に記載技術では、音素環境代替による自然性劣化を知覚評価により評価し、その結果をサブコスト関数に反映している。しかし、合成音声の自然性劣化に関する他の要因については非特許文献２では考慮されていない。これは、種々の物理的尺度と知覚評価との間の対応関係が不明であるか、それを特定するのが極めて難しいためである。
【００１２】
また、非特許文献２に記載されたものにおける知覚実験では、実験に用いられる刺激音声が文章ではなく極めて短い音素連鎖である。そのため、実際の波形接続型音声合成の動作時における条件（実際にコスト関数が使用される環境）とは条件が大きく異なる。その結果、サブコスト関数が実際の動作時に正しく物理尺度とコストとのマッピングをとる事ができるか否かについて問題がある。そのため、マッピングが正確にできる様にする技術が望まれている。
【００１３】
それゆえに本発明の目的は、任意の物理量が与えられたときに、その物理量と知覚評価との間の対応関係を特定する事（マッピング）を可能とする事である。
【００１４】
本発明の他の目的は、任意の物理量が与えられたときに、その物理量と知覚評価との間のサブコスト関数を最適化可能とする事である。
【００１５】
本発明のさらに他の目的は、任意の物理量が与えられたときに、その物理量と知覚評価との間のマッピングを可能とするような音声データセットを容易に作成できる様にする事である。
【００１６】
本発明のさらに他の目的は、音声波形素片接続型音声合成において、知覚評価を反映した形で自然に波形接続が可能な音声合成装置を提供する事である。
【００１７】
本発明のさらに他の目的は、知覚評価とのマッピングに基づいて定められたサブコスト関数から構成されるコスト関数を容易に定める事ができる様にする事である。
【課題を解決するための手段】
【００１８】
本発明の第１の局面に係る知覚試験用音声データセットの作成装置は、それぞれ単位波形素片に分離可能な複数の発話音声データを含む音声コーパスから、発話音声データの予め定める第１の種類の特徴量と人間による知覚評価との間のマッピングを行なう際に使用される知覚試験用音声データセットを作成するための装置である。この装置は、音声コーパスに含まれる発話音声データの各々について、任意の単位波形素片を、音声コーパスに含まれる発話音声データが持つ、任意の単位波形素片に対し所定の関係を充足する単位波形素片で置換する事により、単位波形素片が置換された置換後の発話音声データの集合を作成するための手段と、置換後の発話音声データの各々について、第１の種類の特徴量を含む複数種類の特徴量を算出するための特徴量算出手段と、特徴量算出手段により算出された複数種類の特徴量に基づき、第１の種類の特徴量の変動が所定の第１の条件を充足し、かつ複数種類の特徴量のうち、第１の種類の特徴量以外の特徴量の変動が所定の第２の条件を充足するような発話音声データの集合を、置換後の発話音声データの集合から抽出するための手段とを含む。
【００１９】
好ましくは、置換後の発話音声データの集合を作成するための手段は、音声コーパスに含まれる発話音声データの各々について、任意の単位波形素片を、音声コーパスに含まれる発話音声データが持つ、任意の単位波形素片と同じ音素を含む単位波形素片で置換する事により、置換後の発話音声データの集合を作成するための手段を含む。
【００２０】

例えば、第１の条件は、第１の種類の特徴量の変動が所定のしきい値以上であるという条件であり、第２の条件は、複数種類の特徴量のうち、第１の種類の特徴量以外の特徴量の変動がそれぞれ所定のしきい値以下であるという条件である。

【００２１】

さらに好ましくは、知覚試験用音声データセットの作成装置は、特徴量算出手段により算出された複数種類の特徴量に基づき、複数種類の特徴量のうち、第１の種類の特徴量と異なる第２の種類の特徴量の変動が所定の値以上で、かつ複数種類の特徴量のうち、第１の種類及び第２の種類の特徴量以外の特徴量の変動が所定の値以下となるような発話音声データの集合を、置換後の発話音声データの集合から抽出するための手段をさらに含む。

【００２２】

置換後の発話音声データの集合を作成するための手段は、音声コーパスに含まれる発話音声データの各々について、当該発話音声データに含まれる任意の単位波形素片を選択するための手段と、選択するための手段により選択された単位波形素片と同じ音素を含む単位波形素片を含む、別の発話音声データを音声コーパスの中で特定するための手段と、特定するための手段により特定された別の発話音声データに含まれる、選択された単位波形素片と同じ音素を含む単位波形素片で、選択された単位波形素片を置換するための手段と、置換するための手段による置換が行なわれた発話音声データを予め定める記憶媒体に記憶させるための手段とを含んでもよい。この記憶媒体に記憶される発話音声データにより知覚試験用音声データセットが形成される。

【００２３】

本発明の第２の局面に係るコンピュータプログラムは、コンピュータにより実行されると、当該コンピュータを上記したいずれかの知覚試験用音声データセットの作成装置として動作させるものである。

【００２４】

本発明の第３の局面に係る音声合成用サブコスト関数の最適化装置は、上記したいずれかの知覚試験用音声データセットのうち、第１の種類の特徴量に対応する知覚試験用音声データセットに含まれる発話音声データにより生成された音声の自然性に関する知覚試験の評価を取得するための手段と、評価を取得するための手段により得られた評価と、知覚試験用音声データセットのうち、第１の種類の特徴量に対応するものに含まれる発話音声データに対して算出された第１の種類の特徴量との間の対応関係を表す様に、予め想定された関数を最適化するための手段とを含む。

【００２５】

好ましくは、最適化するための手段は、知覚試験用音声データセットのうち、第１の種類の特徴量に対応するものに含まれる発話音声データに対して算出された第１の種類の特徴量に対して関数により計算される値と、評価を取得するための手段により得られた評価との間の自乗誤差を最小化する様に関数を最適化するための手段を含む。

【００２６】

本発明の第４の局面に係るコンピュータプログラムは、コンピュータにより実行されると、上記したサブコスト関数の関数最適化装置として当該コンピュータを動作させるものである。

【００２７】

本発明の第５の局面に係る音声合成装置は、上記したいずれかのサブコスト関数の最適化装置と、このサブコスト関数の最適化装置により最適化されるサブコスト関数を含んで定義されるコスト関数を用いて、入力音声テキストの音素に対する波形を音声コーパスから選択し接続する事により出力音声波形を合成するための音声合成手段とを含む。

【００２８】

本発明の第６の局面に係る音声合成装置は、コンピュータにより実行されると、上記した音声合成装置として当該コンピュータを動作させる。

【発明を実施するための最良の形態】

【００２９】

＜第1の実施の形態＞

－構成－

以下、本発明の一実施の形態について図を参照して説明する。図１は、本実施の形態に係る音声合成システムの全体構成を示す。図１を参照して、このシステムは、音声コーパス２０と、音声コーパス２０に含まれる発話音声データと知覚評価とに基づいて、コスト関数２４を構成する複数のサブコスト関数と知覚評価とをマッピングし、コスト関数２４を決定するためのサブコスト関数決定部２２と、サブコスト関数決定部２２により決定されたコスト関数２４を用いて入力テキスト３０に対して音声コーパス２０から音素波形素片を選択し接続する事により出力音声波形３６を合成するための音声合成装置３４とを含む。

【００３０】

音声合成装置３４及び音声コーパス２０は図８に示すものを使用する事ができる。ただし、音声合成装置３４が使用するコスト関数は図８の場合と異なる。

【００３１】

図２に、サブコスト関数決定部２２の詳細な構成をブロック図形式で示す。図２を参照して、サブコスト関数決定部２２は、音声コーパス２０に含まれる発話音声データの各々について、その中の任意の一つの単位素片を同じ音素を含む別の単位素片で置換する事により、置換後の発話音声データを作成するための単位素片置換部４０と、単位素片置換部４０により生成された、一部の単位素片が置換された発話音声データからなる置換後音声コーパス４２とを含む。単位素片置換部４０が置換の際に用いる単位素片は、後述する様に音声コーパス２０に含まれる別の発話音声データから選択される。

【００３２】

サブコスト関数決定部２２はさらに、単位素片置換部４０から出力される置換後の発話音声データの各々について、コスト関数で考慮される全ての特徴量及びその統計を算出するための特徴量・特徴量統計算出部４４と、特徴量・特徴量統計算出部４４により算出された特徴量及び特徴量の統計を記憶するための記憶部４６とを含む。

【００３３】

サブコスト関数決定部２２はこれに加えて、記憶部４６に記憶された特徴量及びその統計に基づいて置換後音声コーパス４２に記憶された音声データから自然性劣化の要因に対応する複数の刺激音声データセットを生成し、その刺激音声データを使用して行なわれる知覚評価の結果に基づいてそれぞれのサブコスト関数を導出するための、複数のサブコスト関数導出部４８Ａ～４８Ｎを含む。これらサブコスト関数導出部４８Ａ～４８Ｎにより導出されるサブコスト関数５０Ａ～５０Ｎに、それぞれ重みW_1～W_Nをかけて加算する事によりコスト関数２４が得られる。

【００３４】

単位素片置換部４０による置換後音声コーパス４２の作成処理について説明する。図４に、単位素片の置換の概念を示す。図４を参照して単位素片置換部４０は、音声コーパス２０に含まれる発話音声データのうちの一つを、ターゲット１００として選ぶ。このターゲット１００の発話音声データのうち、任意の単位素片１０２の部分を別の音素を含む単位素片で置換する。この単位素片としては、他の発話音声データ（例え発話音声データ１１０）のうち、この単位素片１０２と同じ音素を含む単位素片（例えば単位素片１１２）を用いる。

【００３５】

全ての発話データ１２０、…、１３０等について、ターゲット１００の単位素片１０２と同じ音素を含む単位素片１１２、１２２、…、１３２等を探す。これら単位素片１１２、１２２、…、１３２を用いて、ターゲット１００の単位素片１０２を置換する。これにより、単位素片を置換した多数の発話データが作成される。この作業を、音声コーパス２０に含まれる全ての発話データをターゲットとし、かつ各ターゲットに含まれる全ての音素に対して行なう事により、置換後音声コーパス４２を作成する。

【００３６】

なお、図５に示す様に、ターゲット１００の単位素片１０２と一致する単位素片を、別の発話音声データが２つ以上含んでいる場合がある。図５に示す例では、発話音声データ１４０はそうした単位素片を３つ（単位素片１４２、１４４，１４６）含んでいる。この場合、ターゲット１００の単位素片１０２をこれら単位素片１４２、１４４、１４６の各々で置換する事により、３つの発話データ１６０、１６２、１６４が生成される事になる。

【００３７】

特徴量・特徴量統計算出部４４は、単位素片置換部４０により単位素片が置換された発話音声データの各々と、元の音声コーパス２０に含まれる発話音声データの各々とに対し、予めサブコスト関数５０Ａ～５０Ｎに対応して定められている特徴データを全て算出する機能を持つ。特徴量・特徴量統計算出部４４はまた、この様にして算出された特徴データについて、特徴データの種類ごとに平均、分散、変動などの統計量を算出する機能も持つ。算出された値は、記憶部４６に記憶される。

【００３８】

サブコスト関数導出部４８Ａ～４８Ｎはいずれも同じ構成を有している。以下、サブコスト関数導出部４８Ａについて説明する。

【００３９】

図３は、サブコスト関数導出部４８Ａの詳細をブロック図形式で示す。図３を参照して、サブコスト関数導出部４８Ａは、記憶部４６に記憶された特徴量及び統計量に基づいて、特定の特徴量については変動量が大きく、他の要因については変動量が所定範囲内であるような音声データを置換後音声コーパス４２から抽出し、前記した特定の要因に関する知覚実験のための刺激音声セット７２を作成するための刺激音声セット抽出部７０と、この刺激音声セット７２を用い、自然性劣化に関する、被験者による知覚試験を行なってその評価を－３～＋３までの７段階で取得する作業を行なうための知覚試験処理部７４とを含む。刺激音声セット７２は、このサブコスト関数導出部４８Ａに対応するサブコスト関数を最適化するためのものである。

【００４０】

刺激音声セット７２は何らかの記憶媒体、例えばハードディスク等に記憶させることができる。この刺激音声セット７２を記憶した記憶媒体を一旦作成すれば、この刺激音声セット７２を用いた知覚試験を別の装置で実行することもできる。本実施の形態では、刺激音声セット７２を作成したものと同じ装置を用いて知覚試験以下の作業を実行するものとする。

【００４１】

サブコスト関数導出部４８Ａは知覚試験の評価を取得するために、刺激音声セット７２に含まれる刺激音声を再生するための音声再生部７６と、被験者が知覚評価を入力するための操作盤７８とをさらに含む。

【００４２】

変動量が大きいか小さいかを判定するためには、通常はしきい値を用いる。このしきい値は、各特徴量の種類によって異なり、また使用された音声コーパス２０に含まれる発話音声データの内容によっても異なる。特徴量・特徴量統計算出部４４による特徴量及び統計量の算出が終了した時点で、このしきい値を何らかの方法により定めることが望ましい。

【００４３】

サブコスト関数導出部４８Ａはまた、知覚試験処理部７４により取得された知覚試験の評価に基づき、刺激音声セット抽出部７０によって刺激音声セット７２を抽出する際に変動量が大きくなる様に設定された特定の要因と、知覚評価との間のマッピングをサブコスト関数５０Ａの形で決定するためのサブコスト関数決定部８０とを含む。

【００４４】

サブコスト関数決定部８０は、次の原理に従ってこのサブコスト関数導出部４８Ａに対応するサブコスト関数を最適化する。すなわち、刺激音声セット７２に含まれる単位素片

置換後の発話音声データについて、知覚試験処理部７４による評点を、このサブコスト関数導出部４８Ａに対応する特徴量の値に対してプロットする。プロットの例を図６に示す。そして、図７に示す様に、この様にプロットされた点と、サブコスト関数を表す曲線１８０との間の自乗誤差の和が最小となる様にサブコスト関数を最適化する。

【００４５】

この様にして、特徴量毎に、対応するサブコスト関数により算出される値が知覚評価をよく反映したものとなる。全てのサブコスト関数に対して知覚特性を考慮にいれた最適化が行なわれる。その結果、これらサブコスト関数により構成されるコスト関数２４を用いて音声波形素片を選択して接続して音声を合成する事により、合成音声の自然性が大きく改善される事が期待される。

【００４６】

－動作－

以上の構成を持つシステムは以下の様に動作する。予め、図１及び図２に示す音声コーパス２０は準備されているものとする。図２を参照して、単位素片置換部４０は次の様にして置換後音声コーパス４２を作成する。すなわち単位素片置換部４０は、音声コーパス２０の中の任意の一つの発話音声データを選択し、ターゲットとする。ターゲットに含まれる全ての単位素片について、音声コーパス２０中の他の発話音声データに含まれる同じ音素を含む単位素片で置換する事により、単位素片置換後の１又は複数の発話音声データを作成し、置換後音声コーパス４２に記憶させる。また、それらの単位素片置換後の発話音声データを特徴量・特徴量統計算出部４４にも与える。

【００４７】

単位素片置換部４０は、この動作を、音声コーパス２０に含まれる全ての発話音声データをターゲットにして行なう。その結果、置換後音声コーパス４２には、音声コーパス２０に含まれていた発話音声データの各々について、その中の一つの単位素片データのみが他の発話音声データの単位素片データで置換されたものが多数含まれる事になる。

【００４８】

特徴量・特徴量統計算出部４４は、単位素片置換部４０により生成される、単位素片置換後の発話音声データの各々について、サブコスト関数にそれぞれ対応する複数種類の特徴量を算出し、各発話音声データに関連付けて記憶部４６に記憶させる。特徴量・特徴量統計算出部４４はまた、算出された特徴量とデータ数とに基づいて、特徴量の各々に関する予め定められた統計量も算出する。算出された統計量も記憶部４６に記憶される。

【００４９】

複数のサブコスト関数導出部４８Ａ～４８Ｎの各々は、以下の様に動作する。以下の説明では代表としてサブコスト関数導出部４８Ａについてのみその動作を説明する。

【００５０】

図３を参照して、刺激音声セット抽出部７０は、記憶部４６に記憶されている特徴量及びその統計量に基づいて、このサブコスト関数導出部４８Ａに対応する特徴量については大きな変動範囲を示し、他の特徴量については小さな変動範囲しか示さない音声波形データの集合を抽出する。この結果、このサブコスト関数導出部４８Ａに対応するサブコスト関数を最適化するための刺激音声セット７２が作成される。

【００５１】

この際には、抽出する音声波形データの数を一定としてもよいし、抽出後の音声波形データの集合が上記した条件を充足する限り、できる限り多くの音声波形データを抽出する様にしてもよい。また、このサブコスト関数導出部４８Ａに対応する特徴量の分布に偏りが生じないよう、上記した条件を充足する音声波形データのうちでも一部のみを抽出する様にしてもよい。分布を考慮する際には、線形軸だけでなく、対数軸などの上での分布を考慮する様にしてもよい。

【００５２】

知覚試験処理部７４は、音声再生部７６を用いて、刺激音声セット７２中の各発話音声データを、元の発話音声データと対比する形で被験者に提示する。被験者は、両者を対比

して単位素片置換後の発話音声の自然度を－３～＋３の７段階で評価する。評価結果は操作盤７８を用いて知覚試験処理部７４に入力される。知覚試験処理部７４は、この評価結果をその単位素片置換後の発話音声と関連付けて記憶する。

【００５３】

サブコスト関数決定部８０は、知覚試験処理部７４により取得された評価結果を用い、このサブコスト関数導出部４８Ａに対応する特徴量によるサブコスト関数を、知覚試験の評価結果との間の自乗誤差が最小となる様に最適化する。

【００５４】

以上の処理を、サブコスト関数導出部４８Ａ～４８Ｎの全てにおいて行なう。これにより、考慮の対象となっている全ての特徴量（物理量）と、知覚試験との間のマッピングを、それぞれサブコスト関数の形で定式化できる。それらサブコスト関数を加重加算する事により、コスト関数を得る事ができる。このコスト関数は、知覚試験の結果を反映したサブコスト関数の結果を総合したものである。図１に示す音声合成装置３４は、このコスト関数により計算されるコストが最も小さくなる様に音声波形素片を音声コーパス２０から選択し、接続する事で音声合成を行なう。

【００５５】

コスト関数は、知覚試験の結果を反映したサブコスト関数の結果を総合したものであるから、その値もまた知覚試験の結果を反映したものとなる。その結果、このコスト関数に基づいて音声波形素片を選択し接続する事により得られる合成音声は、人間が聞いたときに自然な発話として聞こえるものとなる事が期待できる。

【００５６】

また、知覚試験処理部７４による知覚試験においては、刺激音声として一発話の全体を用いる。そのため、実際の波形接続型音声合成が行なわれる場合に即した条件の下での知覚評価を行なう事ができる。サブコスト関数はその知覚評価の結果を反映する様に最適化されるため、最終的に得られるコスト関数もまた実際の音声合成の場面で自然な音声合成を実現する事ができる。

【００５７】

以上ブロック図形式で説明した各機能部は、いずれもコンピュータ及び当該コンピュータ上で実行されるプログラムにより実現することができる。このコンピュータとしては、音声を扱う設備を持ったものであれば、汎用のハードウェアを有するものを用いることができる。また、上で説明した装置の各機能ブロックは、この明細書の記載に基づき、当業者であればプログラムで実現することができる。そうしたプログラムもまた一つのデータであり、記憶媒体に記憶させて流通させることができる。

【００５８】

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】

【００５９】

【図１】本発明の一実施の形態に係るコスト関数算出システム及び音声合成システムを示すブロック図である。

【図２】サブコスト関数決定部のブロック図である。

【図３】サブコスト関数導出部のブロック図である。

【図４】ターゲットの単位素片の置換の概念を模式的に示す図である。

【図５】ターゲットの単位素片の置換を説明するための模式図である。

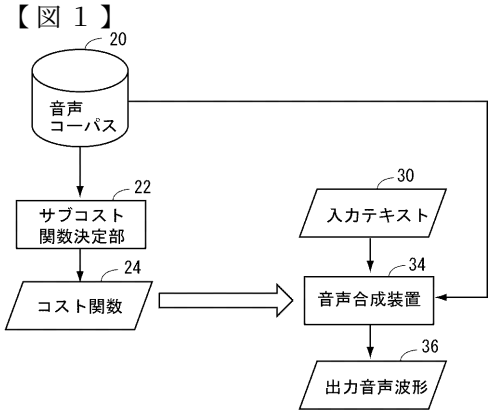
【図６】置換後の合成音声に対する知覚評価を、その特徴量に対してプロットした例を示すグラフである。

【図７】サブコスト関数の最適化の概念を模式的に示すグラフである。

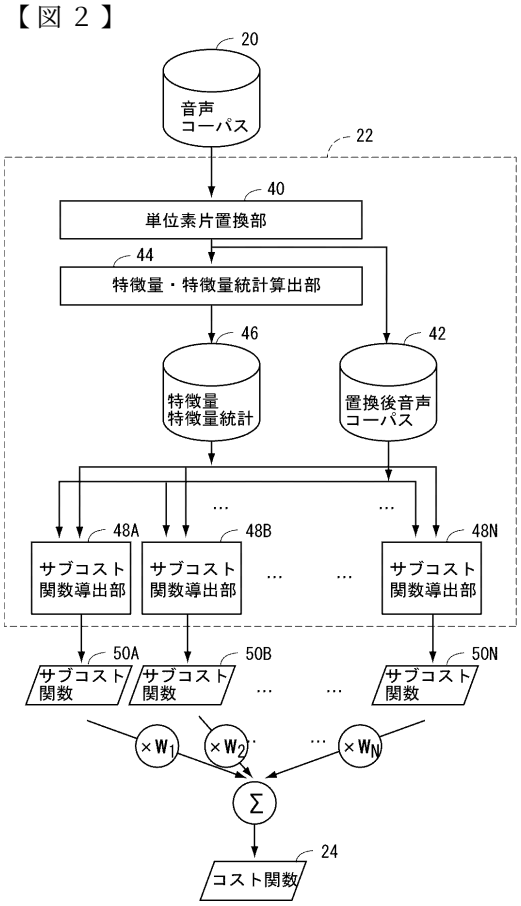
【図８】波形接続型音声合成の概念を示すブロック図である。

【図９】コスト関数及びサブコスト関数の関係を示す図である。
【符号の説明】
【００６０】
２０、３２　音声コーパス、２２　サブコスト関数決定部、２４　コスト関数、　３０　入力テキスト、３４　音声合成装置、３６　出力音声波形、４０　単位素片置換部、４２　置換後音声コーパス、４４　特徴量・特徴量統計算出部、４６　記憶部、４８Ａ～４８Ｎ　サブコスト関数導出部、５０Ａ～５０Ｎ　サブコスト関数、７０　刺激音声セット抽出部、７２　刺激音声セット、７４　知覚試験処理部、８０　サブコスト関数決定部

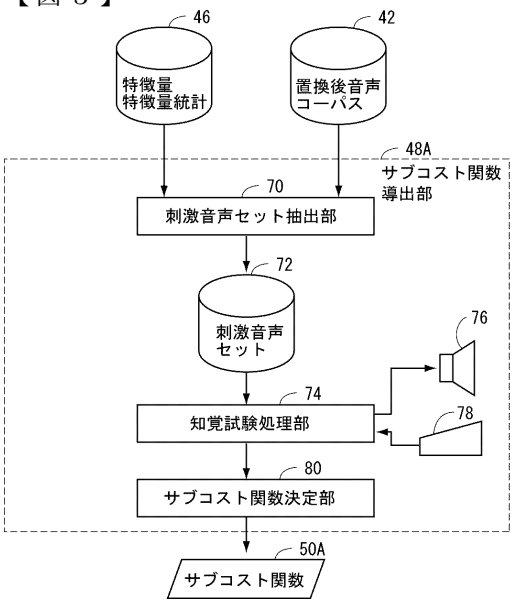
【図１】



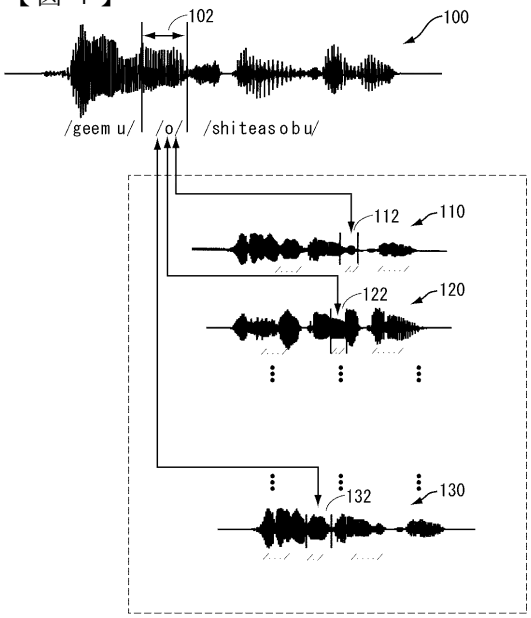
【図２】



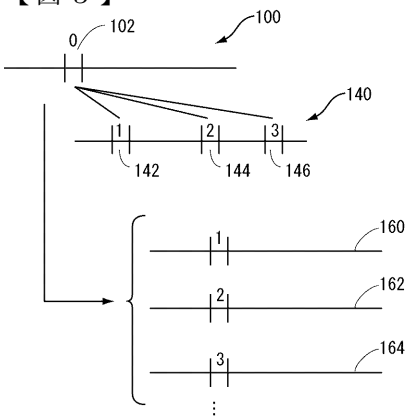
【図３】



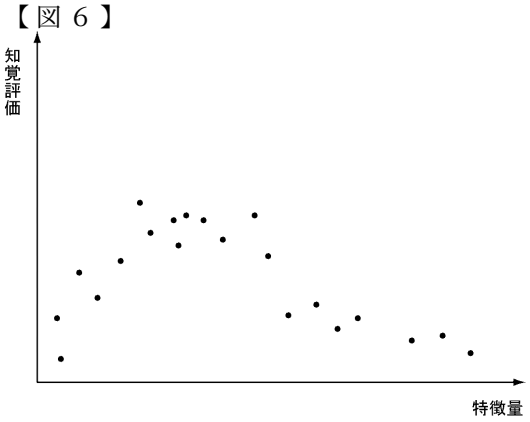
【図４】



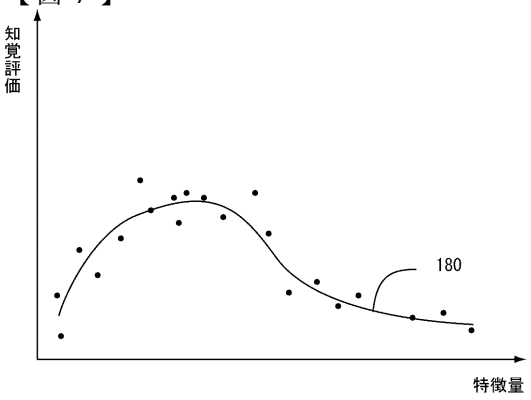
【図５】



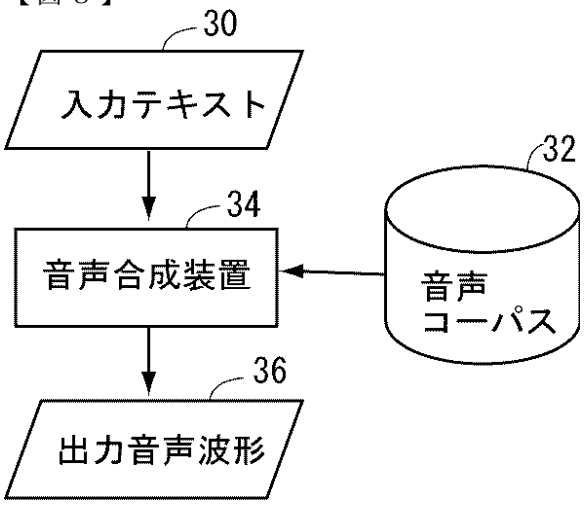
【図６】



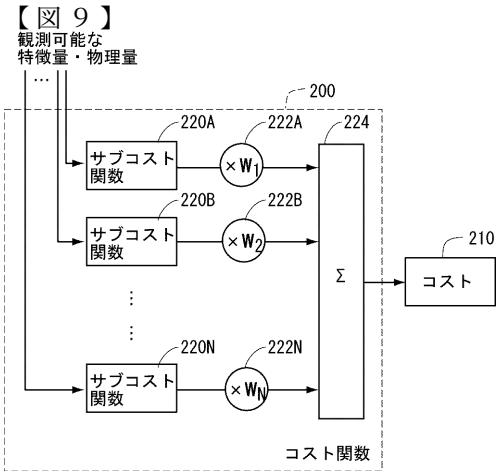
【図７】



【図８】



【図９】


観測可能な
特徴量・物理量
…
200
220A
222A
224
サブコスト
関数
×W1
220B
222B
サブコスト
関数
×W2
210
Σ
コスト
220N
222N
サブコスト
関数
×WN
コスト関数

フロントページの続き

審査官　荏原　雄一

(56)参考文献　特開平１０－２５４４７１（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ１０Ｌ　１３／００－１３／０８