(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第3588302号
(P3588302)
(45)発行日　平成16年11月10日(2004.11.10)　(24)登録日　平成16年8月20日(2004.8.20)

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

請求項の数 13　(全 11 頁)

(21)出願番号　特願2000-65106(P2000-65106)
(22)出願日　平成12年3月9日(2000.3.9)
(65)公開番号　特開2000-310997(P2000-310997A)
(43)公開日　平成12年11月7日(2000.11.7)
審査請求日　平成12年6月7日(2000.6.7)
(31)優先権主張番号　09/264981
(32)優先日　平成11年3月9日(1999.3.9)
(33)優先権主張国　米国(US)

(73)特許権者　000005821
松下電器産業株式会社
大阪府門真市大字門真１００６番地
(74)代理人　100062144
弁理士　青山　葆
(74)代理人　100086405
弁理士　河宮　治
(74)代理人　100098280
弁理士　石野　正弘
(72)発明者　ニコラス・キブレ
アメリカ合衆国９３４３６カリフォルニア州ロンポック、ウエスト・パイン４０１番、ナンバー３１

最終頁に続く

(54)【発明の名称】連結型音声合成のための単位重複領域の識別方法および連結型音声合成方法

(57)【特許請求の範囲】
【請求項１】
音声の時変特性を表す統計モデルを画定するステップと、
同じ母音を含む異なる音声単位に対応する複数の時系列データを提供するステップと、
前記時系列データから音声信号パラメータを抽出し、前記音声信号パラメータを用いて前記統計モデルを学習するステップと、
学習させた前記統計モデルを用いて前記時系列データ内の繰り返しシーケンスを識別し、前記繰り返しシーケンスを前記母音の中心の核をなす状態遷移部と関連付けるステップと、
前記繰り返しシーケンスを用いて、前記音声単位の少なくとも１つに対する連結型音声合成のための単位重複領域を定めるステップとを含み、前記単位重複領域は、前記繰り返しシーケンスの直前の時系列データ又は直後の時系列データであることを特徴とする、連結型音声合成のための単位重複領域の識別方法。
【請求項２】
前記統計モデルは隠れマルコフモデルである、請求項１に記載の方法。
【請求項３】
前記統計モデルはリカレントニューラルネットワークである、請求項１に記載の方法。
【請求項４】
前記音声信号パラメータは音声フォルマントを含む、請求項１に記載の方法。
【請求項５】

前記統計モデルは、前記母音の中心の核をなす状態遷移部と、前記中心の核をなす状態遷移部の周囲の遷移部とを別々にモデル化するデータ構造を有する、請求項１に記載の方法。

【請求項６】

前記統計モデルは、前記母音の中心の核をなす状態遷移部と、前記中心の核をなす状態遷移部に先行する第１の遷移部と、前記中心の核をなす状態遷移部に後続する第２の遷移部とを別々にモデル化するデータ構造を有し、

前記データ構造を用いて、前記第１の遷移部および前記第２の遷移部の１つに対応する前記時系列データの１部分を破棄するステップを含む、請求項１に記載の方法。

【請求項７】

音声の時変特性を表す統計モデルを画定するステップと、

同じ母音を含む異なる音声単位に対応する複数の時系列データを提供するステップと、

前記時系列データから音声信号パラメータを抽出し、前記音声信号パラメータを用いて前記統計モデルを学習するステップと、

学習させた前記統計モデルを用いて前記時系列データ内の繰り返しシーケンスを識別し、

前記繰り返しシーケンスを前記母音の中心の核をなす状態遷移部と関連付けるステップと、

前記繰り返しシーケンスを用いて、連結型音声合成のための単位重複領域を定めるステップとを含み、前記単位重複領域は、前記繰り返しシーケンスの直前の時系列データ又は直後の時系列データであり、

前記音声単位の各単位重複領域に基づいて、２つの異なる前記音声単位からの前記時系列データを重複させ、マージすることにより、新たな音声単位を連結して合成するステップとを含むことを特徴とする、連結型音声合成方法。

【請求項８】

前記合成するステップを行う前に、前記単位重複領域の少なくとも１つの継続時間を選択的に変化させて、前記単位重複領域の他方の継続時間に一致させるステップをさらに含む、請求項７に記載の方法。

【請求項９】

前記統計モデルは隠れマルコフモデルである、請求項７に記載の方法。

【請求項１０】

前記統計モデルはリカレントニューラルネットワークである、請求項７に記載の方法。

【請求項１１】

前記音声信号パラメータは音声フォルマントを含む、請求項７に記載の方法。

【請求項１２】

前記統計モデルは、前記母音の中心の核をなす状態遷移部と、前記中心の核をなす状態遷移部の周囲の遷移部とを別々にモデル化するデータ構造を有する、請求項７に記載の方法。

【請求項１３】

前記統計モデルは、前記母音の中心の核をなす状態遷移部と、前記中心の核をなす状態遷移部に先行する第１の遷移部と、前記中心の核をなす状態遷移部に後続する第２の遷移部とを別々にモデル化するデータ構造を有し、

前記データ構造を用いて、前記第１の遷移部および前記第２の遷移部の１つに対応する前記時系列データの１部分を破棄するステップを含む、請求項７に記載の方法。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】

本発明は、本発明は連結型（ｃｏｎｃａｔｅｎａｔｉｖｅ）音声を合成するシステムに関する。より詳しくは、本発明は、連結した音声単位（音声ユニット：ｓｐｅｅｃｈ　ｕｎｉｔ）について適切なエッジ境界領域を識別するためのシステムおよび方法に関する。システムは、音声単位モデルを用いて設けられた音声単位データベースを利用する。

【０００２】
【従来の技術】
連結型音声の合成は、今日、数多くの様々な形態で世の中に存在しており、それは、どのように連結音声単位が格納され、処理されるかに依存している。これらの形態は、時間領域波形表現や、（例えば、フォルマント線形予測コーディングＬＰＣ表現などの）周波数領域表現、またはこれらの組み合わせを含む。
【０００３】
音声単位の形態にかかわらず、連結型音声の合成は、各単位（ユニット：ｕｎｉｔ）のエッジで適切な境界領域を識別することにより行われる。ここで、単位は滑らかに重複され、それにより語や句を含む新たな音声単位に合成される。連結型音声合成システムにおける音声単位は、典型的には２音（ｄｉｐｈｏｎｅｓ）または半音節（ｄｅｍｉｓｙｌｌａｂｌｅｓ）である。この場合には、境界重複領域は音素内にある（ｐｈｏｎｅｍｅ－ｍｅｄｉａｌ）。したがって、例えば、「ｔｏｏｌ」という語は、「ｔｏｏｔｈ」および「ｆｏｏｌ」という語から導き出された単位「ｔｕ」および「ｕｌ」により組み立てられる。決定すべきは、どの程度の量のソース語が音声単位にセーブされるかであり、また一緒に置かれたときにどの程度重複するべきかである。
【０００４】
連結型テキスト－音声（ｔｅｘｔ－ｔｏ－ｓｐｅｅｃｈ　：　ＴＴＳ）システムに関する従来の研究では、重複領域を判定するのに多くの方法が利用されてきた。このようなシステムを設計するに際しては、３つの因子が考慮される。すなわち、
・シームレスな連結：音声単位の重複により、ある単位とテキストとの間は十分滑らかに遷移し、急激な変化は聞こえないようにすべきである。リスナーには、音声片から組み立てられた音声を聞いているとはわからないようする必要がある。
【０００５】
・歪みのない遷移：音声単位の重複により、それ自身の歪みを生じてはならない。単位は、非重複音声との識別ができないように混在する必要がある。
【０００６】
・最小のシステム負荷：音声合成部における計算に必要な要件および／または記憶容量の要件は、できるだけ小さくする必要がある。
【０００７】
【発明が解決しようとする課題】
現在のシステムではこれらの３つの目標の間にはトレードオフが存在し、３つのすべてに関して最適なシステムは存在していない。現在のアプローチは、一般的に３つの目標のバランスをとった、２つの選択に基づいてグループ化できる。第１の選択は、短い重複領域を用いるか、長い重複領域を用いるかである。短い重複領域を用いると、単一の声門パルスと同じ程度に早くできる。一方、長い重複領域を用いると、全音素の大部分を含むことができる。第２の選択は、重複領域は前後関係が整合しているか、または変化してもよいかである。前者の場合には、各音声単位の対応する部分は、先行する単位および後続の単位がどのような単位であるかにかかわらず重複している。後者の場合には、その単位が用いられる度に、隣接する単位に依存して、用いられる部分が変化する。
【０００８】
重複が長いと、単位間の遷移がよりシームレスになるという利点がある。その理由は、それらの間の微妙な相違が取り除かれる機会が多いからである。しかし、重複が長いと歪みを生じやすい。信号と異なり、混合すると歪みが生じる。
【０００９】
重複が短いと、歪みを最小にできるという利点がある。重複を短くすると、重複部分を十分に一致させることが簡単かつ確実にできる。短い重複領域は、（動的変化状態とは異なり）ほぼその瞬間の状態の特徴を表すと考えられる。しかし重複を短くすると、重複が長いシステムで実現できるシームレスな連結が犠牲になる。
【００１０】

重複が長い場合でシームレスが実現できることが望ましく、重複が短い場合に歪みを少なくできることが望ましいが、現在までのところ、これを達成できるシステムは存在しない。最新のシステムの中には、重複が長い場合の利点を保持しながら歪みを最小にするという目的で、可変重複領域を用いる実験が行われているものがある。しかし、このようなシステムは、計算負荷が高い処理に非常に大きく頼っているために、多くの用途には非実用的である。

【００１１】

本発明の目的は、シームレスで、かつ歪みのない重複を与える音声単位の領域を識別する方法、および連結型音声を合成する方法を提供することである。

【００１２】

【課題を解決するための手段】

本発明の連結型音声合成のための単位重複領域の識別方法は、音声の時変特性を表す統計モデルを画定するステップと、同じ母音を含む異なる音声単位に対応する複数の時系列データを提供するステップと、前記時系列データから音声信号パラメータを抽出し、前記音声信号パラメータを用いて前記統計モデルを学習するステップと、学習させた前記統計モデルを用いて前記時系列データ内の繰り返しシーケンスを識別し、前記繰り返しシーケンスを前記母音の中心の核をなす状態遷移部と関連付けるステップと、前記繰り返しシーケンスを用いて、連結型音声合成のための単位重複領域を定めるステップとからなり、それにより上記目的が達成される。

【００１３】

前記統計モデルは隠れマルコフモデルであってもよい。

【００１４】

前記統計モデルはリカレントニューラルネットワークであってもよい。

【００１５】

前記音声信号パラメータは音声フォルマントを含んでいてもよい。

【００１６】

前記統計モデルは、前記母音の中心の核をなす状態遷移部と、前記中心の核をなす状態遷移部の周囲の遷移部とを別々にモデル化するデータ構造を有していてもよい。

【００１７】

統計モデルを学習する前記ステップは、埋め込み再評価により行われ、前記時系列データによって表される全データセットにわたって整列のために収束したモデルを生成してもよい。

【００１８】

前記統計モデルは、前記母音の中心の核をなす状態遷移部と、前記中心の核をなす状態遷移部に先行する第１の遷移部と、前記中心軌線領域に後続する第２の遷移部とを別々にモデル化するデータ構造を有し、前記データ構造を用いて、前記第１の遷移部および前記第２の遷移部の１つに対応する前記時系列データの１部分を破棄するステップを含んでいてもよい。

【００１９】

本発明による連結型音声合成方法は、音声の時変特性を表す統計モデルを画定するステップと、同じ母音を含む異なる音声単位に対応する複数の時系列データを提供するステップと、前記時系列データから音声信号パラメータを抽出し、前記音声信号パラメータを用いて前記統計モデルを学習するステップと、学習させた前記統計モデルを用いて前記時系列データ内の繰り返しシーケンスを識別し、前記繰り返しシーケンスを前記母音の中心の核をなす状態遷移部と関連付けるステップと、前記繰り返しシーケンスを用いて、連結型音声合成のための単位重複領域を定めるステップと、前記音声単位の各単位重複領域に基づいて、２つの異なる前記音声単位からの前記時系列データを重複させ、マージすることにより、新たな音声単位を連結して合成するステップとからなり、それにより上記目的が達成される。

【００２０】

前記合成するステップを行う前に、前記単位重複領域の少なくとも１つの継続時間を選択的に変化させて、前記単位重複領域の他方の継続時間に一致させるステップをさらに含んでいてもよい。

【００２１】

前記統計モデルは隠れマルコフモデルであってもよい。

【００２２】

前記統計モデルはリカレントニューラルネットワークであってもよい。

【００２３】

前記音声信号パラメータは音声フォルマントを含んでいてもよい。

【００２４】

前記統計モデルは、前記母音の中心の核をなす状態遷移部と、前記中心の核をなす状態遷移部の周囲の遷移部とを別々にモデル化するデータ構造を有していてもよい。

【００２５】

統計モデルを学習する前記ステップは、埋め込み再評価により行われ、前記時系列データによって表される全データセットにわたって整列のために収束したモデルを生成してもよい。

【００２６】

前記統計モデルは、前記母音の中心の核をなす状態遷移部と、前記中心の核をなす状態遷移部に先行する第１の遷移部と、前記中心の核をなす状態遷移部に後続する第２の遷移部とを別々にモデル化するデータ構造を有し、前記データ構造を用いて、前記第１の遷移部および前記第２の遷移部の１つに対応する前記時系列データの１部分を破棄するステップを含んでいてもよい。

【００２７】

本発明は統計的モデル化技術を利用することにより、音声単位内で中心軌跡領域を識別する。これらの領域は最適な重複境界を識別するのに用いられる。好ましい本実施の形態では、時系列データが、隠れマルコフモデルを用いて統計的にモデル化される。隠れマルコフモデルは、各音声単位の音素領域上に構築され、学習または埋め込み（ｅｍｂｅｄｄｅｄ）再評価を経て整列（ａｌｉｇｎ）される。

【００２８】

好ましい実施の形態では、各音声単位の最初と最後の音素は３要素からなると考えられる。すなわち中心の核をなす状態遷移部（中心軌跡：ｎｕｃｌｅａｒ　ｔｒａｊｅｃｔｏｒｙ）、中心の核をなす状態遷移部に先行する遷移部および中心の核をなす状態遷移部に後続する遷移部である。モデル化プロセスはこれらの３要素を最適に識別し、それにより中心の核をなす状態遷移部は問題となる音素のすべてのインスタンスに対して、相対的な整合を維持する。

【００２９】

識別された中心の核をなす状態遷移部を用いると、中心の核をなす状態遷移部の先頭境界および終端境界は重複領域を画定する。重複領域はその後、連結合成に用いられる。

【００３０】

好ましい本実施の形態では、母音の中心の核をなす状態遷移部、中心の核をなす状態遷移部に先行する第１の遷移部、および中心の核をなす状態遷移部に後続する第２の遷移部を別個にモデル化するためのデータ構造を有する統計的モデルを利用する。データ構造は、音声単位データの一部分を破棄にするのに用いられる。音声単位データの一部分のデータは、連結プロセスの間には用いられない音声単位の部分に対応する。

【００３１】

本発明には多数の利点および使用法が存在するが、本発明は、連結型音声合成システムに用いられる音声単位データベースの自動構築の基礎として用いることができる。自動化技術は、導き出された合成音声の品質を向上し、データベース収集プロセスにおける労力を大幅に削減することができる。

【００３２】

音声信号パラメータは、同じ母音を含む、異なる音声単位に対応する時系列データから抽出される。抽出されたパラメータは、隠れマルコフモデルといった統計的モデルを学習するのに用いられる。統計的モデルは、母音の中心の核をなす状態遷移部と、その周りの遷移部とを別々にモデル化するデータ構造を有する。このモデルは、埋め込み再評価を経て学習され、中心の核をなす状態遷移部を識別する最適に整列されたモデルを決定する。中心の核をなす状態遷移部の境界は、後の音声単位との連結のために重複領域を定めるよう機能する。

【００３３】

【発明の実施の形態】

本発明は、以下の添付の図面を参照して説明される。

【００３４】

本発明により利用される技術をもっともよく理解するためには、連結合成の基本的な理解が必要である。図１は、例を通した連結合成プロセスを示す。この例では、異なる２つの語からの音声単位（この場合は音節）が連結され、第３の語を形成する。より具体的には、「ｓｕｆｆｉｃｅ」および「ｔｉｇｈｔ」という語からの音声単位が組み合わされ、新たな「ｆｉｇｈｔ」という語が合成される。

【００３５】

図１を参照して、「ｓｕｆｆｉｃｅ」および「ｔｉｇｈｔ」という語からの時系列データが、好ましくは音節の境界で抽出され、音声単位１０、１２を規定する。この場合、音声単位１０は１４においてさらに細分割され、連結に必要な関連部分を分離する。

【００３６】

その後、音声単位は１６で整列され、それにより各部分１８および２０により規定される重複領域が作られる。整列後、時系列データがマージされ、新たな語２２が合成される。

【００３７】

本発明は特に、重複領域１６と最適部分１８、２０に関連し、ある音声単位から別の音声単位までの遷移をシームレスで、かつ歪みがないようにする。

【００３８】

本発明は、自動化された手順を経てこの最適な重複を実現する。この手順では、母音内で中心の核をなす（中心軌跡：ｎｕｃｌｅａｒ　ｔｒａｊｅｃｔｏｒｙ）領域が探し出される（なお、「中心軌跡」の「軌跡」とは、本明細書において、目標周波数に向かって変化する概念を表すのに用いられる）。ここで母音内で「中心の核をなす」領域とは、母音の中心にある、安定した領域をいう。音声波形は、それを構成するフォーマット周波数によって表すことができる。これらの周波数は、ある音節が次の音節に融和して発音されると一定の変化を生じる。伝統的には、発声は、安定した目標周波数に向かって変化するこれらのフォーマット周波数を利用して、典型的には母音を利用してなされている。このとき周波数の波形は、直ちにより安定した波形になる。本明細書で母音内で「中心の核をなす」とは、母音によって占められる、中心にある安定した領域をいう。音声信号は、動的ではあるが同じ音素の異なる例に対しては相対的に変化がない動的パターンに続く。母音の境界領域は、隣接する子音によって影響を受けるが、中心にある安定した領域は強く影響を受けない。

【００３９】

これらの最適な重複領域を改良するための手順が、図２に示される。まず、音声単位のデータベース３０が提供されている。データベース３０は時系列データを含んでおり、時系列データは、連結合成システムを構成する異なる音声単位に対応する。好ましい本実施の形態では、音声単位は発声された語の例の中から抽出される。発声された語の例は、後に音節境界でさらに分割される。図２では、図解的に音声単位３２，３４が描かれている。音声単位３２は「ｔｉｇｈｔ」という語から抽出され、音声単位３４は「ｓｕｆｆｉｃｅ」という語から抽出されている。

【００４０】

データベース３０に格納されている時系列データはまず、３６においてパラメータ化され

る。概して、音声単位は任意の方法論を用いてパラメータ化できる。好ましい本実施の形態では、各音声単位内で音素領域をフォルマント解析してパラメータ化を行う。フォルマント解析は、必然的に音声フォルマント周波数の抽出を伴う。本実施の形態ではフォルマント周波数Ｆ１、Ｆ２およびＦ３が抽出される。必要であれば、ＲＭＳ信号レベルもまたパラメータ化できる。

【００４１】

現在のところはフォルマント解析が好ましいが、パラメータ化の他の形態もまた利用できる。例えば、音声の特徴抽出は線形予測コーディング（Ｌｉｎｅａｒ　Ｐｒｅｄｉｃｔｉｖｅ　Ｃｏｄｉｎｇ：ＬＰＣ）などの手順を用いて行い、適切な特徴パラメータを識別し、抽出できる。

【００４２】

適切なパラメータが抽出され、各音声単位の音素領域が表されると、３８で示されるようにモデルが構築され、各単位の音素領域が表される。好ましい本実施の形態はこの目的のために隠れマルコフモデルを用いる。しかし、概して時変または動的挙動を表す、適切な任意の統計的モデルを用いることができる。例えば、リカレントニューラルネットワークモデルを利用できる。

【００４３】

好ましい本実施の形態は、音素領域を３つの異なる中間領域に分割してモデル化する。これらの領域は４０で示されており、中心の核をなす状態遷移部（中心の核をなす領域）４２と、中心の核をなす状態遷移部４２に先行する状態遷移部（先行状態遷移領域）４４と、中心の核をなす状態遷移部４２に後続する状態遷移部（後続状態遷移領域）４６とを含む。好ましい実施の形態では、これらの３領域の各々について別々の隠れマルコフモデルを用いる。先行および後続の状態遷移部４４、４６には、３状態モデルが用いられる。一方、中心の核をなす状態遷移部４２には４または５状態モデルが用いられる。図２には５状態モデルが示されている。より大きな状態数を中心の核をなす状態遷移部４２に用いると、後の手順は、整合のある非ヌル中心軌線に収束する。

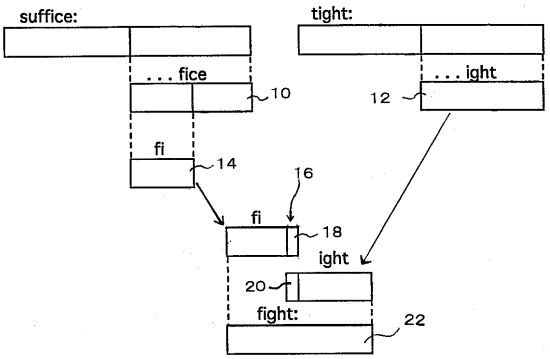
【００４４】

まず、音声モデル４０が平均的な初期値で設けられる。その後、４８で示されたこれらのモデルに関して、埋め込み（ｅｍｂｅｄｄｅｄ）再評価が行われる。再評価とは、実質的には学習プロセスを継続することである。学習プロセスによりモデルは最適化されて、時系列データ内でもっともよい繰り返しシーケンスを表す。繰り返しシーケンスとは、母音内で中心にある安定した領域に関連する時系列データが呈する、より規則的な反復パターンのシーケンスをいう。これは、音声データが時系列データとして表されたときに、子音に対応する音声部分が規則性をもって反復しない非常に無秩序なパターンを呈しやすいこととは対照的である。したがって、母音が発生される度に繰り返して生じやすい時系列データ内のパターンは、母音領域内で識別できる。時系列データの繰り返しシーケンスは、識別されて所与の母音に対応する発声部分の識別手段として用いられる。例えば、音節「ｙａ」の終端における母音音声は、音節「ａ」の統計的パターンと非常に関連のある統計的パターンを呈する。同じ統計的パターンは、例えば、音節「ｋａ」、「ｍａ」、「ｈａ」内の安定領域において見出すことができる。対照的に、安定的な母音領域に先行する音節部分では、統計的な関連がない場合が多く、したがって識別可能な繰り返しパターンも存在しない。さらなる例示のために、時系列データが統計モデルを学習するのに用いられ、各モデルがパラメータの組を規定すると仮定する。モデルを学習させた後、母音音声「ａ」はパラメータ番号のシーケンス：４－５－３．１－６に対応する。母音が存在するたびに同一の番号のパターンが発生しているとすると、そのパターンは、その母音が存在することを示すのに信頼性高く利用できる繰り返しシーケンスを構成する。本発明では、子音、または安定的な母音に融和する音声などの他の音声は、非常に繰り返しのあるシーケンスを生成することが統計的に存在しないと判断する。したがって、発せられた音声内に安定した母音領域があることを検出する手段として、非常によく反復するシーケンス（繰り返しシーケンス）を見つけ出す。

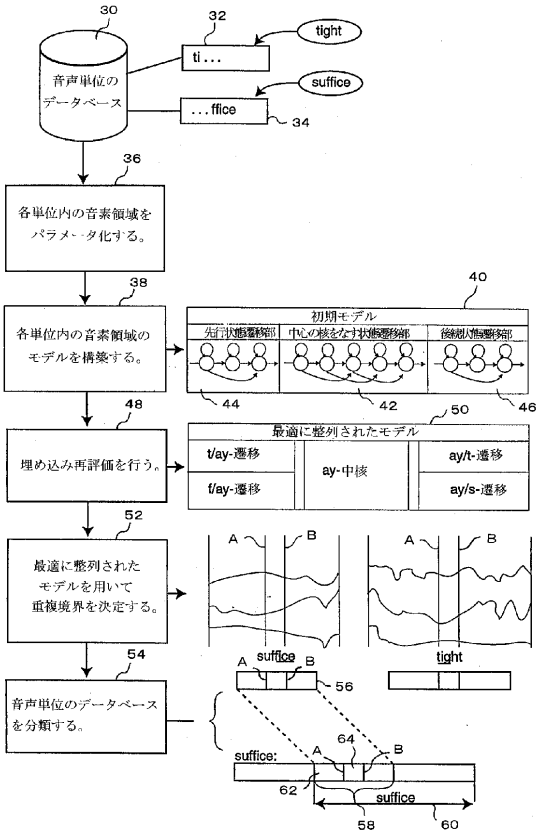
【００４５】
中心の核をなす状態遷移部４２、先行および後続の状態遷移部４４，４６は、データベース３０を介して供給される現実のデータに基づいて、学習プロセスにより各音素領域に整合するモデルが構築されるよう設計される。この点に関して、中心の核をなす部分４２は母音の核心を表し、先行および後続の状態遷移部４４，４６は、現在の音素および現在の音素に先行するおよび後続する音声に固有の母音の相を表す。例えば、「ｔｉｇｈｔ」という語から抽出された音声単位３２では、先行する遷移部は、前にある子音字「ｔ」により母音「ａｙ」の音声に与えられた音調（ｃｏｌｏｒａｔｉｏｎ）を表す。
【００４６】
整合プロセスは本来、最適な整列モデルに収束する。どのようしてそのようになるのかを理解するために、音声単位３０のデータベースが、少なくとも２つ、好ましくは多数の各母音の音声の例を含むとする。例えば図２には、「ｔｉｇｈｔ」および「ｓｕｆｆｉｃｅ」の双方に見受けられる母音の音声「ａｙ」が、音声単位３２、３４により表されている。埋め込み再評価プロセスまたは学習プロセスは、音声「ａｙ」のこのような複数のインスタンスを用いて初期音声モデル４０の学習を行い、それにより最適に整列された音声モデル５０を生成する。音声「ａｙ」の例のすべてにわたって整合のある時系列データの部分は、中核、または中心の核をなす領域を表す。５０で図示されるように、システムは、先行および後続の状態遷移部を別々に学習する。これらは、母音に先行するおよび後続する音声に依存して当然に異なっている。
【００４７】
一旦モデルが学習され、最適に整列されたモデルを生成すると、中心の核をなす領域４２の両側の境界が確定し、連結合成のための重複領域の位置が決定される。そのため、ステップ５２では最適に整列されたモデルが重複境界を決定するのに用いられる。図２は、重複境界ＡおよびＢを示す。重複境界ＡおよびＢは、「ｓｕｆｆｉｃｅ」および「ｔｉｇｈｔ」という語から導かれた音声単位に対するフォルマント周波数データに重ね合わされている。
【００４８】
パラメータデータ（この場合はフォルマント周波数データ）で識別された重複境界により、システムはステップ５４において時系列データを分類して時系列データ内の重複境界を定める。必要であれば、分類されたデータは連結型音声合成について後に使用するために、データベース３０に格納してもよい。
【００４９】
図示の関係上、オーバレイテンプレート５６として模式的に示されている重複境界領域が、「ｓｕｆｆｉｃｅ」という語の時系列データの模式的表現に重ね合わされて示されている。具体的には、テンプレート５６は、後半の音節「．．．ｆｉｃｅ」内で括弧５８によって示すように整列されている。この音声単位が連結音声に用いられると、先行領域６２は破棄され、境界ＡおよびＢにより定められている中心の核をなす領域６４は、クロスフェード領域または連結領域として働く。
【００５０】
ある実施形態では、連結合成を行うために、重複領域の継続時間を調整する必要がある。このプロセスが図３に示される。入力テキスト７０が解析され、ステップ７２に示されるようにデータベース３０から適切な音声単位が選択される。例えば、「ｆｉｇｈｔ」という語が入力テキストとして与えられると、システムは「ｔｉｇｈｔ」および「ｓｕｆｆｉｃｅ」という語から抽出した、あらかじめ格納してある音声単位を選択する。
【００５１】
各音声単位の中心の核をなす領域は必ずしも同じ時間にわたっている必要はない。そのためステップ７４では、各中心の核をなす領域の継続時間が伸張または短縮され、それにより継続時間を一致させる。図３では、中心の核をなす領域６４ａが領域６４ｂに伸張される。音声単位Ｂも同様に変更される。図３は中心の核をなす領域６４ｃが領域６４ｄに圧縮され、それにより２つの単位の各領域が同じ継続時間を持つことになる。

【００５２】
一旦継続時間が調整されて一致すると、ステップ７６において、音声単位からのデータがマージされて、７８で示される新しく連結された単語を形成する。
【００５３】
【発明の効果】
これまでの説明によれば、本発明は連結型音声合成システムに用いられる音声単位データベースを構築する自動化手段を提供することが理解される。中心の核をなす領域を分離することによって、このシステムは、シームレスで、かつ歪みのない重複を与える。有利なのは、重複領域は共通の固定サイズに伸張または圧縮され、連結プロセスを簡単化できることである。統計的モデル化プロセスを用いることで、中心の核をなす領域は音声信号の１部分を表すことができる。ここでは、音響学上の音声特性は、同じ音素の異なる例に対しては相対的に変化がない動的パターンを生じる結果となる。変化がないことにより、シームレスで、かつ歪みのない遷移が可能になる。
【００５４】
本発明の原理により生成された音声単位は、コンピュータ処理システムにかける負担を最小にして、後の抽出および連結に用いるデータベースに容易に格納できる。したがって、このシステムは、処理能力が制限されている合成音声に関する製品および応用の開発には理想的といえる。さらに、音声単位を生成する自動化プロセスは、目的が特化された音声単位データベースを構築するのに必要な時間と労力を大幅に減少させる。例えば音声単位を生成する自動化プロセスは、専門的なボキャブラリに対して、または多言語音声合成システムの開発に対して必要とされるであろう。
【００５５】
現時点での好ましい形態で本発明を説明してきたが、当業者であれば、特許請求の範囲に記載された本発明の精神から逸脱することなく本システムを修正できる。
【図面の簡単な説明】
【図１】連結型音声を合成する技術の理解に有用なブロック図である。
【図２】本発明による、音声単位が構築される手順を示すフローチャートである。
【図３】本発明の音声単位データベースを用いた、連結型音声を合成するプロセスを示すブロック図である。
【符号の説明】
４０　音声モデル
４２　中心の核をなす状態遷移部
４４　先行状態遷移部
４６　後続状態遷移部
５０　音声モデル
５６　オーバレイテンプレート
６２　先行領域
６４　中心の核をなす領域

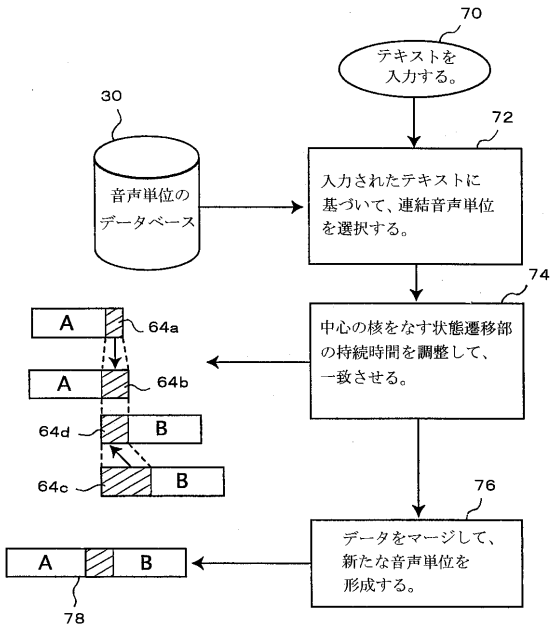
【図１】


suffice:
tight:
. . . fice
10
12
. . . ight
fi
14
16
fi
18
ight
20
fight:
22

【図２】


30
32
tight
音声単位の
データベース
ti . . .
suffice
. . . ffice
34
36
各単位内の音素領域を
パラメータ化する。
38
40
各単位内の音素領域の
モデルを構築する。
初期モデル
先行状態遷移部
中心の核をなす状態遷移部
後続状態遷移部
44
42
46
48
50
埋め込み再評価を行う。
最適に整列されたモデル
t/ay-遷移
ay-中核
ay/t-遷移
f/ay-遷移
ay/s-遷移
52
最適に整列された
モデルを用いて
重複境界を決定する。
A
B
suffice
tight
54
音声単位のデータベース
を分類する。
56
64
suffice:
62
58
60

【図３】


70
テキストを
入力する。
30
72
音声単位の
データベース
入力されたテキストに
基づいて、連結音声単位
を選択する。
74
A
64a
A
64b
64d
B
64c
B
中心の核をなす状態遷移部
の持続時間を調整して、
一致させる。
76
データをマージして、
新たな音声単位を
形成する。
A
B
78

フロントページの続き

(72)発明者　スティーブ・ピアソン
アメリカ合衆国９３１１０カリフォルニア州サンタ・バーバラ、カリェ・シタ３９０９番

審査官　渡邊　聡

(56)参考文献　特開平１０－２４７０９７（ＪＰ，Ａ）
特開平１０－０４９１９３（ＪＰ，Ａ）
特開平０７－２４４４９７（ＪＰ，Ａ）
吉村　他，HMMに基づく音声合成のための状態継続長モデルの構築，電子情報通信学会技術研究報告，電子情報通信学会，１９９８年　9月，ＳＰ９８－６４，４５－５０
益子　他，HMMを用いた音声合成における音素モデルの検討，日本音響学会講演論文集，日本音響学会，１９９６年　3月，平成8年春季Ｉ，２７３－２７４

(58)調査した分野(Int.Cl.^{7}，ＤＢ名)
G10L 13/08
G10L 13/06