(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第3587048号
(P3587048)
(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 13/08　G10L 3/00　H
G10L 13/04　G10L 5/02　J

請求項の数 8　(全 10 頁)

(21)出願番号　特願平10-49161
(22)出願日　平成10年3月2日(1998.3.2)
(65)公開番号　特開平11-249677
(43)公開日　平成11年9月17日(1999.9.17)
審査請求日　平成14年7月26日(2002.7.26)

(73)特許権者　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地
(74)代理人　100075096
弁理士　作田　康夫
(72)発明者　額賀　信尾
東京都国分寺市東恋ケ窪一丁目２８０番地
株式会社日立製作所中央研究所内
(72)発明者　北原　義典
東京都国分寺市東恋ケ窪一丁目２８０番地
株式会社日立製作所中央研究所内

最終頁に続く

(54)【発明の名称】韻律制御方法及び音声合成装置

(57)【特許請求の範囲】
【請求項１】
テキストデータの韻律パラメータを計算し、上記計算された韻律パラメータより合成波形を生成する音声合成装置における韻律制御方法において、
上記音声合成装置は、複数のテキストデータについて、各テキストデータの人間の発声に基づく複数種の韻律パラメータを記憶する音声コーパスを備え、
上記入力テキストデータと上記音声コーパスに記憶されたテキストデータとを順次比較し、少なくともその一部が一致するテキストデータを上記音声コーパスより選択し、
上記選択されたテキストデータと上記入力テキストデータとの不一致部分についての上記複数種の韻律パラメータを求め、
上記選択されたテキストデータと上記入力テキストデータとの一致部分について記憶される上記各韻律パラメータと上記不一致部分について求められた上記各韻律パラメータとを結合し、上記入力テキストデータの韻律を求めることを特徴とする韻律制御方法。
【請求項２】
請求項１記載の韻律制御方法において、
上記複数の韻律パラメータとして、基本周波数パターンと継続時間長とを含むことを特徴とする韻律制御方法。
【請求項３】
請求項１記載の韻律制御方法において、
上記入力テキストデータについて形態素分割を行い、各形態素Ｄｉ（ｉ＝１～ｎ）につい

てその品詞とアクセント型を求め、
上記音声コーパスに記憶されたテキストデータのうち、その形態素数が一致するテキストデータの各形態素Ｄ’ｊ（ｊ＝１～ｎ）についてそれぞれ上記Ｄｉと品詞及びアクセント型を比較することにより、上記音声コーパスに記憶されたテキストデータから少なくとも１つのテキストデータを選択することを特徴とする韻律制御方法。
【請求項４】
請求項３の韻律制御方法において、
上記音声合成装置は、単語のモーラ数及びアクセント型の組み合わせに対して１つ以上の基本周波数パタンを記憶する単語基本周波数パタンテーブルを備え、
上記選択されたテキストデータと上記入力テキストデータとの不一致部分について、上記単語基本周波数パタンテーブルを索表することにより上記不一致部分の基本周波数パタンを計算することを特徴とする韻律制御方法。
【請求項５】
複数のテキストデータと該テキストデータの人間の発声に基づく複数種の韻律パラメータを記憶する音声コーパスメモリと、
入力されたテキストデータを解析し、上記音声コーパスメモリから上記テキストデータと韻律が類似するテキストデータを検索する音声コーパス検索部と、
上記検索されたテキストデータと上記入力されたテキストデータの不一致部分の韻律パラメータを計算し、上記検索されたテキストデータと上記入力されたテキストデータの一致部分について上記音声コーパスメモリに記憶される上記韻律パラメータと上記不一致部分について計算された韻律パラメータとを統合する韻律パラメータ加工部と、
上記韻律パラメータ加工部で求められた韻律パラメータを利用して合成波形データを生成する波形合成部と、を有することを特徴とする音声合成装置。
【請求項６】
上記音声コーパス検索部は、上記入力されたテキストデータを単語に分割し、該単語毎に表記、読み、品詞、アクセント情報から構成される構造化データ列を得る形態素分割処理を行うことを特徴とする請求項５記載の音声合成装置。
【請求項７】
上記音声コーパス検索部は、上記形態素分割処理の結果と１の上記テキストデータのセットの構造化データの数、品詞及びアクセント型について比較して類似度を計算する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】
上記韻律パラメータの加工部は、上記入力されたテキストデータから生成した発音記号列を用いて、上記検索された韻律パラメータを加工することを特徴とする請求項５乃至７の何れかに記載の音声合成装置。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】
本発明は、テキストから音声を合成するテキスト音声合成技術に関し、特に、文の抑揚や時間長を制御する韻律制御方法に関する。
【０００２】
【従来の技術】
テキスト音声合成は、入力されたテキストに対して言語処理を行い、中間的な発音記号列を生成した後、発音記号から基本周波数パタンや音素継続時間長などの韻律パラメータと、同時に発音記号から各音素の声道パラメータを求め、これらのパラメータから音声を合成する方法が一般的である。例えば、文献「音声合成技術：広瀬啓吉（音声処理技術とその応用、情報処理、ｐｐ．９８４－９９１、１９９７年１１月）」にはこの方式を利用した音声合成装置が記載されている。
【０００３】
上記の方法を利用した場合、韻律パラメータは音声の抑揚やリズム、なめらかさといった

全体にかかわる自然性としての特徴を決め、声道パラメータは単語や文を構成する個々の音節の明瞭性としての特徴を決定する。

【０００４】

韻律パラメータのうち、基本周波数パタンのパラメータ生成方法の代表的なものとして「重畳型モデル」が挙げられる。この基本周波数パラメータの生成モデルは、句などに対応して基本周波数が緩やかに下降するフレーズ成分に、文節単位のアクセント型に対応する基本周波数の上昇・下降であるアクセント成分を重畳するものである。重畳型モデルは、人間の発声機構を模擬したモデルであるため、直感的に理解しやすく、現実の音声現象とも良く合致しているが、モデルの性能を発揮するためには高度な言語処理が必要であるという問題があった。

【０００５】

韻律パラメータのうち、音節（音素）継続時間長は、音節（音素）が置かれるコンテキストをはじめとする種々の要因によって変化する。継続時間長に影響を与える要因として、調音上の制約、タイミング、単語の重要度、発話区分境界の明示、発話区分内のテンポ、全体のテンポ、構文意味内容等の言語的制約がある。継続時間長制御においては、実際に観測される継続時間長データに対して、上に挙げた要因に関する影響度を統計的に分析し、その結果得られる規則を利用する方法が一般的である。そのため、多様なコンテキストにおける継続時間長モデルを構築しようとすると大規模なデータベースが必要となるという問題がある。

【０００６】

このような韻律パラメータの他、パワーに関するパラメータなど各種制御モデルが考案されているが、いずれのモデルも各韻律パラメータ独立のモデルであり、これら単独の制御モデルの性能向上だけでは、自然性向上に限界がある。また、文音声発声には、規則によるモデル化が困難な韻律現象があることも指摘されている。

【０００７】

そのため、肉声から抽出した韻律パラメータをデータベースとして記憶し、データベースを利用して韻律パラメータを生成する韻律パラメータ計算モデルが提案されている。例えば、片江ら「文型－韻律データベースを用いた定型文音声合成システム、音講論、ｐｐ．２７５－２７６、平成８年３月」や、斎藤ら「付属語連鎖単位を音声合成単位とする規則合成音声、音講論、ｐｐ．３１７－３１８、平成６年１０月」である。しかしながら、これらは基本周波数パターンのみを韻律パラメータとするものであり、なお文音声発声の自然性を高めるには不十分なものであった。

【０００８】

【発明が解決しようとする課題】

本発明では、肉声の各種の韻律パラメータ（基本周波数パターン、音節継続時間長等）を相互に保存したまま編集・加工することにより、さらに自然性の高い音声を合成する音声合成装置を実現するものである。

【０００９】

そこで本発明では、文音声、その韻律パラメータ、形態素・構文解析データなどからなる音声コーパスを備え、音声コーパスを検索して入力文との類似度が最も大きいデータを抽出し、抽出されたデータに対して韻律パラメータの加工・修正を行うことにより韻律パラメータを生成することで、従来の方法と比較して自然性の高い音声を合成するテキスト音声合成装置を提供する。

【００１０】

【発明の実施の形態】

図１は、本発明の音声合成装置の一実施態様の構成を示すブロック図である。６は汎用メモリ、４は韻律パラメータ計算部、５は波形合成部である。韻律パラメータ計算部４は、音声コーパスメモリ１、音声コーパス検索部２、韻律パラメータ加工部３を含む。

【００１１】

図２に汎用メモリ６の内容を示す。汎用メモリ６には、テキストデータ１０、韻律パラメ

ータ１１、合成波形データ１２、音声コーパス検索部検索結果１３、及び計算処理途中のデータ１４が記憶される。

【００１２】

図３に、本発明の音声合成処理の流れを示す。音声コーパス検索部２は、入力されたテキストデータ２１の解析を行って、音声コーパスメモリ１より韻律が類似しているデータを検索する。検索結果２４は韻律パラメータ加工部３に入力される。なお、音声コーパスメモリ１に、テキストデータ２１と韻律が類似するデータが存在しない場合には、韻律の類似するデータが存在していないことを検索結果２４として出力する。韻律パラメータ加工部３では、検索結果２４とテキストデータ２１から、韻律パラメータ２６を計算する。波形合成部５では、韻律パラメータ２６を利用して合成波形データ２８を生成する。

【００１３】

以下、「渋谷まで渋滞しています。」というテキストデータを例として、本発明によるテキストデータから合成波形への変換方法を図を用いて説明する。

【００１４】

図４は、音声コーパス検索部２における処理の流れである。まず、テキストデータ３１「渋谷まで渋滞しています。」を、汎用メモリ６のテキストデータ１０より読み出す（ステップ１０１）。読み出されたテキストデータ３１は、単語に分割され、単語毎に表記、読み、品詞、及びアクセント情報から構成される構造化データ列３３に変換する（形態素分割処理１０２）。構造化データ列３３は、汎用メモリ６の計算処理途中データ１４として記憶する。テキストデータを単語に分割する方法としては、清水ら「隣接単語間の結合関係に着目したテキスト音声変換用形態素解析処理、日本音響学会誌、５１巻、１号、ｐｐ．３－１３、１９９５」に記載の処理が利用できる。単語毎の読み及びアクセント情報は、辞書に登録されているデータを設定する。

【００１５】

構造化データ列について図６を参照して説明する。４０は１つの単語の構造化データであり、単語の表記４２、読み４３、品詞４４、アクセント情報４５を含む。テキストデータ３１は例えば、「渋谷／まで／渋滞／し／て／い／ます／。」と単語分割されるので、形態素分析結果は、各単語の構造化データの列３３のとおりとなる。

【００１６】

次に、音声コーパスメモリ１から一のデータセットを読み出す（ステップ１０３）。図７に音声コーパスメモリ１のデータ構造を示す。音声コーパスメモリ１は、複数のデータセット４０１、４０２・・・から構成される。各データセットは、図８に示すように、文字表記データ５０１、文字表記データ５０１を発声した音声波形データ５０２、音声波形データ５０２の基本周波数パタンデータ５０３、音声波形データ５０２の継続時間長データ５０４から構成される。なお、データセットには、パワーなどの韻律パラメータやケプストラムなどの音響パラメータさらに、文字表記データ５０１の形態素分割結果を含んでも良い。

【００１７】

音声コーパスメモリのデータセットについて、図９を参照して文字表記データ６０１が「新宿まで運転しています。」という文であった場合を例に説明する。基本周波数パタンデータ６０３は、音節の始終端における基本周波数の列として記憶されている。例えば、文字表記データ６０１の先頭の音節「し」の始端周波数６０５は「２１４」であり、終端周波数６０６は「１９０」である。また、先頭の音節「し」の継続時間長データ６０４は、音節のミリ秒の単位で記憶され、子音の時間長６０７は「１０１」であり、母音の時間長６０８は「７５」である。

【００１８】

次に、音声コーパス検索部２では、一のデータセット６００の文字表記データ６０１を読み出し（ステップ１０４）、図１０に示すように、形態素分割を行って形態素分割結果７０３を汎用メモリ６の計算途中のデータ１４に記憶する（ステップ１０５）。なお、データセット６００に形態素分割結果が含まれている場合には、本処理は不要である。

【００１９】
次に、汎用メモリ６より、テキストデータ３１から求めた形態素解析結果３３と、一のデータセット６００の文字表記データ６０１の形態素分割結果７０３を読み出し、類似度の計算を行う（ステップ１０６）。
【００２０】
以下、図１１を用いて類似度計算の例を示す。最初に、形態素解析結果３３と形態素分割結果７０３との構造化データ数を比較する（ステップ８００）。ここでは、３３と７０３の構造化データ数はともに「８」であり、一致する。構造化データ数が一致すれば次の判定に進み（ステップ８０１）、一致しなければ類似度「０」と判定し、類似度計算１０６の処理を終了する（ステップ８０２、８０３）。
【００２１】
次に構造化データ数が一致した場合には、構造化データのそれぞれについて品詞及びアクセント型の比較を行う（ステップ８０４）。形態素解析結果３３から各構造化データＤｉ（ｉ＝１～ｎ）と、形態素解析結果７０３から各構造化データＤ’ｉ（ｉ＝１～ｎ）とを比較する。例えば、Ｄ１「渋谷」とＤ’１「新宿」とでは、ともに品詞は「地名名詞」、アクセント型は「平板型」であり一致する。このように、全てのＤｉとＤ’ｉについて品詞とアクセント型の比較を行って、全て一致すれば、類似度を「１」として類似度計算を終了する（ステップ８０８）。また、１つでも一致しない構造化データがあれば、類似度を「０」として類似度計算を終了する（ステップ８０７）。出力された類似度は、汎用メモリ６の計算途中のデータ１４に記憶される。
【００２２】
次に、汎用メモリ６の計算途中のデータ１４から類似度を読み出し、予め設定したしきい値（基準類似度）との比較を行い（ステップ１０７）、検索結果を出力する（ステップ１０８）。図１１に記載の類似計算の場合では、予め設定したしきい値を「１」とし、類似度計算１０６により計算された類似度が１であれば、比較結果として「一致」を出力し、類似度が０であれば「不一致」を出力する。検索結果の出力１０８では、類似度比較の結果が「不一致」であれば（１０９）、音声コーパスメモリ１に記憶されているデータセットを順次読み出し（ステップ１０３）、データセットがなくなるまで類似度の計算を行う。類似度比較の結果が「一致」であれば、一致したデータセット６００を汎用メモリ６の音声コーパス検索部結果１３に出力する。
【００２３】
もし、音声コーパスメモリ１の全てのデータセットに対して上記の類似度計算処理を行った結果、基準類似度を満たすデータセットが存在しない場合には、その旨を示すデータフラグ（「類似データなしフラグ」と呼ぶ）を、汎用メモリ６の音声コーパス検索結果１３に出力する。上記の処理を経て、１つ以上の類似するデータセット、あるいは、類似データなしフラグが汎用メモリ６の音声コーパス検索部１３に出力される。
【００２４】
図１２は、韻律パラメータ加工部３における処理の流れである。まず、テキストデータ３１を読み出し、発音記号列を生成する（ステップ１００１）。テキストデータから発音記号列へ変換する方法は、匂坂ら「日本語単語連鎖のアクセント規則、電子情報通信学会論文誌、Ｊ６６－Ｄ、Ｎｏ．７、ｐｐ．８４９－８５６、１９８３」を用いることができる。図１３に発音記号列の例を示す。発音記号列９０１には、読みの情報のほかに、文節の区切れ９０４、文末記号９０５、母音の無声化記号９０３、アクセント記号９０２などが含まれている。なお、発音記号列９０１の生成には、汎用メモリ６に記憶されているテキストデータ３１の形態素分割結果３３を利用することができる。
【００２５】
次に、汎用メモリ６の音声コーパス検索部検索結果１３から、類似度計算処理の結果を読み出す（ステップ１００２）。類似度計算処理の結果は、（１）１つ以上のデータセットである場合（１００３）と、（２）類似データなしフラグである場合（１００４）のいずれかである。

【００２６】
類似するデータセットが存在すれば、１つのデータセットを選択する。このデータセットを「選択データセット」とよぶ。テキストデータ３１「渋谷まで渋滞しています」に対して、データセット６００「新宿まで運転しています」（図９）が選択データセットの例となる。選択データセットは、テキストデータ３１と韻律が近いデータである。なぜなら、図５及び図１０に示すように両者の形態素分割結果は、「渋谷」３２及び「新宿」７０２に対応する構造化データと、「渋滞」３８及び「運転」７０８に対応する構造化データ以外は、全て同一の構造化データあり、また、異なる構造化データに関しても、品詞及びアクセント型が等しいからである。
【００２７】
そこで、テキストデータ３１の韻律パラメータを計算するために、選択データセット６００の韻律パラメータである基本周波数パタンデータ６０３及び継続時間長データ６０４を利用すれば、肉声の韻律パラメータと近い韻律パラメータが得られ、自然性が格段に向上する。韻律パラメータ計算の方法を、図１４を用いて説明する。
【００２８】
まず、テキストデータの形態素分割結果１１０１（先の例では形態素分割結果３３）と選択データセット１１０２（先の例では形態素分割結果７０１）に対して、一致部と不一致部の分離（ステップ１１０３）を行う。１１０４は分離した結果を模式的に表したものであり、構造化データ１１０５及び１１０６は一致した構造化データを、構造化データ１１０７及び１１０８は一致しなかった構造化データを示す（先の例では、構造化データ３２及び７０２、または構造化データ３８及び７０８である）。
【００２９】
次に、テキストデータ３１「渋谷まで渋滞しています。」の音節数から成るデータ列を生成し、分離結果１１０４に基づき、一致部の韻律パラメータをコピーする（ステップ１１０９）。一致部の韻律パラメータは、選択データセット１１０２（先の例ではデータセット６００）の韻律パラメータをそのまま用いる。基本周波数パタンデータ１１１０、継続時間長データ１１１１は、テキストデータ３１の韻律パラメータであり、網掛けの部分は一致した構造化データに対応する音節の韻律パラメータが、空白の部分が不一致の構造化データに対応する音節の韻律パラメータが格納されている。
【００３０】
次に、不一致部の音節について韻律パラメータを計算する（ステップ１１１２）。基本周波数パタンは、単語のモーラ数とアクセント型で１つの基本周波数パタンデータを記憶する単語基本周波数パタンテーブルを用意して、この単語基本周波数パタンテーブルから検索することで、単語基本周波数パタンを求めることができる。また、単語継続時間長は、匂坂ら「規則による音声合成のための音韻時間長制御、信学論、Ｖｏｌ．Ｊ６７－Ａ、Ｎｏ．７、ｐｐ．６２９－６３６、１９８４」を利用して求めることができる。これら公知の方法により、不一致部の単語基本周波数パタンデータ１１１３（１１１５）、不一致部の継続時間長データ１１１４（１１１６）を求めることができる。
【００３１】
最後に、不一致部の韻律パラメータを、一致部の韻律パラメータと滑らかに接続するように、不一致部の韻律パラメータを変形・統合する（ステップ１１１７）。基本周波数パタンデータに関しては、始端周波数１１２０及び終端周波数１１２１の基本周波数が、選択データセット１１０２（先の例ではデータセット６００）における相当する基本周波数の値と一致するように、単語基本周波数パタンデータを線形に変形する。また、継続時間長に関しては、選択データセット１１０２における相当する単語の継続時間長をモーラ数で除した値（１モーラあたりの時間長Ｌ）を利用して、１１１４及び１１１６における１モーラあたりの時間長がＬと等しくなるように、１１１４及び１１１６の継続時間長を線形伸縮する。このようにして、テキストデータ３１の韻律パラメータとして、基本周波数パタンデータ１１１８及び継続時間長データ１１１９が計算される。
【００３２】

なお、音声コーパス検索部２から類似データなしフラグが出力された場合には、音声コーパスの韻律データを利用して韻律パラメータを計算することができない。従って、発音記号列９０１より、従来の方法に従って韻律パラメータを計算する（ステップ１００６）。なお、この場合は音声コーパスを利用して合成した音声より自然性が劣るため、任意文を合成するために音声コーパスは大容量の記憶媒体に記憶させることが望ましく、例えば、磁気記憶媒体、光記憶媒体、光磁気記憶媒体、フラッシュメモリなどに記憶させることができる。また、音声コーパスと通信線を介した接続することで、複数の音声合成装置で共有することも有効である。

【００３３】

求められた韻律パラメータ１００７は、汎用メモリ６の韻律パラメータ１１に記憶する。

【００３４】

このように、韻律パラメータ計算部３により計算された基本周波数パタン及び該継続時間長を汎用メモリ６の韻律パラメータ１１より読み出し、波形合成部５において、出力音声波形を合成する。合成された波形データは、汎用メモリ６の合成波形データ１２に記憶される。

【００３５】

【発明の効果】

本発明により、抑揚やリズムが自然で肉声に近い合成音を生成することができる。

【図面の簡単な説明】

【図１】本発明の音声合成装置のブロック図である。

【図２】音声合成装置の汎用メモリに記憶される内容の図である。

【図３】本発明の音声合成装置における処理の流れ図である。

【図４】音声コーパス検索部における処理例である。

【図５】テキストデータから形態素分割結果への変換例である。

【図６】形態素分割結果の構造化データ列である。

【図７】音声コーパスのデータ構造例である。

【図８】音声コーパスのデータセットのデータ構造例である。

【図９】データセット例である。

【図１０】データセットの文字表記データから形態素分割結果への変換例である。

【図１１】類似度計算処理例である。

【図１２】韻律パラメータ計算部における処理例である。

【図１３】発音記号列の例である。

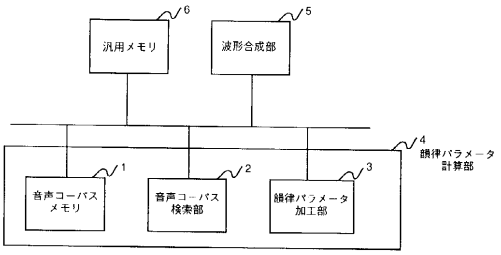
【図１４】音声コーパスのデータを利用した韻律パラメータ計算例である。

【符号の説明】

１…音声コーパスメモリ、２…音声コーパス検索部、３…韻律パラメータ加工部、４…韻律パラメータ計算部、５…波形合成部、６…汎用メモリ。

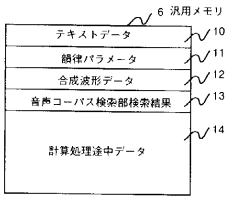
【図１】

図1



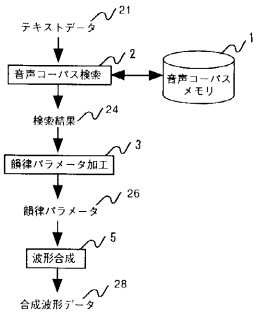
【図２】

図2



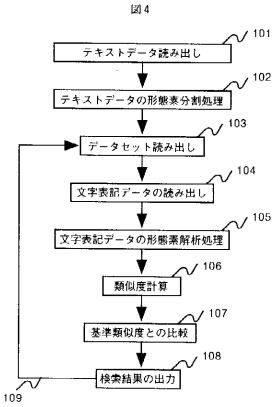
【図３】

図3



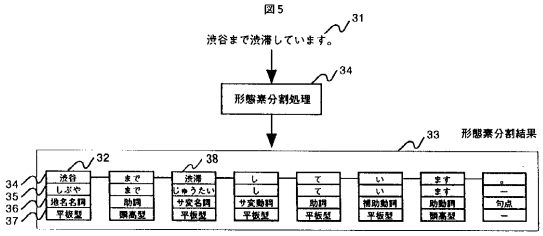
【図４】

図4



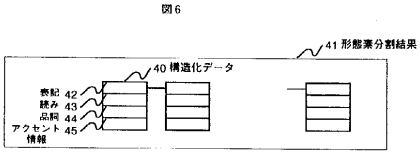
【図５】

図5



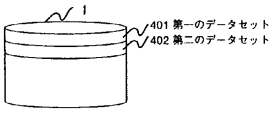
【図６】

図6



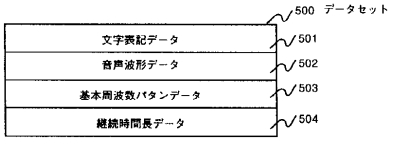
【図７】

図7



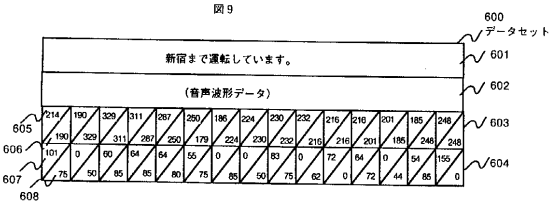
【図８】

図8



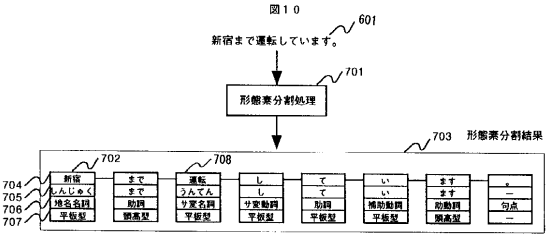
【図９】

図9



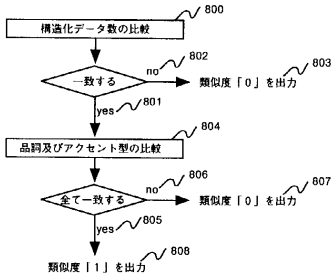
【図１０】

図10



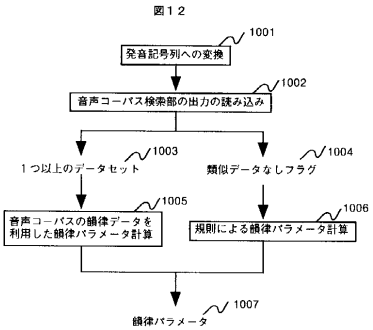
【図１１】

図１１



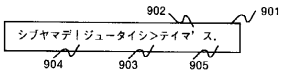
【図１２】

図１２



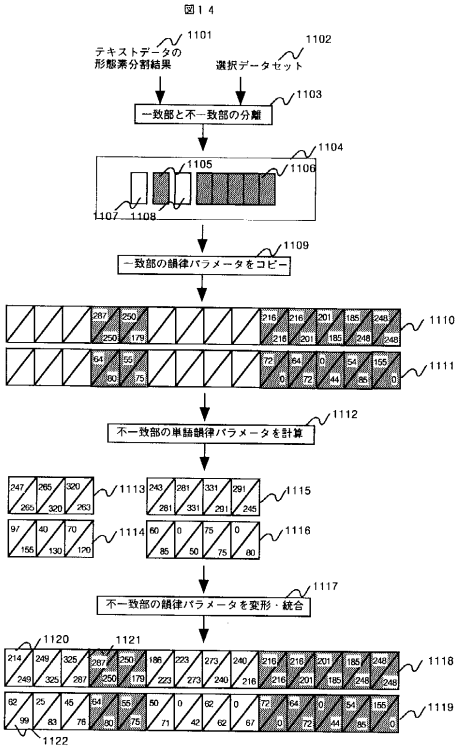
【図１３】

図１３



【図１４】

図１４



フロントページの続き

(72)発明者 藤田 啓子
東京都国分寺市東恋ケ窪一丁目２８０番地 株式会社日立製作所中央研究所内
(72)発明者 安藤 ハル
東京都国分寺市東恋ケ窪一丁目２８０番地 株式会社日立製作所中央研究所内
(72)発明者 矢島 俊一
東京都小平市上水本町五丁目２０番１号 株式会社日立製作所半導体事業部内

審査官 山下 剛史

(56)参考文献 特開平０７－０４４１８６（ＪＰ，Ａ）
特開昭６０－２０５５９６（ＪＰ，Ａ）
特開平０８－０８７２９７（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.^{7}，ＤＢ名)
G10L 13/00-13/08