(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第4680429号
(P4680429)
(45)発行日　平成23年5月11日(2011.5.11)　(24)登録日　平成23年2月10日(2011.2.10)

(51)Int.Cl.　F I
G10L 13/08 (2006.01)　G10L 13/08　132

請求項の数 9　(全 36 頁)

(21)出願番号　特願2001-192778(P2001-192778)
(22)出願日　平成13年6月26日(2001.6.26)
(65)公開番号　特開2003-5775(P2003-5775A)
(43)公開日　平成15年1月8日(2003.1.8)
審査請求日　平成20年3月3日(2008.3.3)

(73)特許権者　308033711
ＯＫＩセミコンダクタ株式会社
東京都八王子市東浅川町５５０番地１
(74)代理人　100079049
弁理士　中島　淳
(74)代理人　100084995
弁理士　加藤　和詳
(74)代理人　100099025
弁理士　福田　浩志
(72)発明者　茅原　桂一
東京都港区虎ノ門１丁目７番１２号　沖電気工業株式会社内

審査官　毛利　太郎

最終頁に続く

(54)【発明の名称】テキスト音声変換装置における高速読上げ制御方法

(57)【特許請求の範囲】

【請求項１】

入力されたテキストから音韻・韻律記号列を生成するテキスト解析手段と、前記音韻・韻律記号列に対して少なくとも音声素片・音韻継続時間・基本周波数の合成パラメータを生成するパラメータ生成手段と、音声の基本単位となる音声素片が登録された素片辞書と前記パラメータ生成手段から生成される合成パラメータに基づいて前記素片辞書を参照しながら波形重畳を行って合成波形を生成する波形生成手段とを備えたテキスト音声変換装置における高速読み上げ制御方法であって、前記パラメータ生成手段は、音韻継続時間を予め経験的に求めた継続時間規則テーブルと、音韻継続時間を統計的手法を用いて予測した継続時間予測テーブルとを併せ持ち、ユーザから指定される発声速度が閾値を超えた時には前記継続時間規則テーブルを用い、閾値を超えていない時には前記継続時間予測テーブルを用いて音韻継続時間の決定を行う音韻継続時間決定手段を有することを特徴とするテキスト音声変換装置における高速読み上げ制御方法。

【請求項２】

前記閾値は、所定の最大発声速度であることを特徴とする請求項１記載のテキスト音声変換装置における高速読み上げ制御方法。

【請求項３】

入力されたテキストから音韻・韻律記号列を生成するテキスト解析手段と、前記音韻・韻律記号列に対して少なくとも音声素片・音韻継続時間・基本周波数の合成パラメータを生成するパラメータ生成手段と、音声の基本単位となる音声素片が登録された素片辞書と

前記パラメータ生成手段から生成される合成パラメータに基づいて前記素片辞書を参照しながら波形重畳を行って合成波形を生成する波形生成手段とを備えたテキスト音声変換装置における高速読み上げ制御方法であって、前記パラメータ生成手段は、アクセント成分及びフレーズ成分を決定するために必要となるデータを、予め経験的に求めた規則テーブルと、統計的手法を用いて予測した予測テーブルとを併せ持ち、ユーザから指定される発声速度が閾値を超えた時には前記規則テーブルを用い、閾値を超えていない時には前記予測テーブルを用いてアクセント成分及びフレーズ成分を決定することによりピッチパタンを決定するピッチパタン決定手段を有することを特徴とするテキスト音声変換装置における高速読み上げ制御方法。

【請求項４】

前記閾値は、所定の最大発声速度であることを特徴とする請求項３記載のテキスト音声変換装置における高速読み上げ制御方法。

【請求項５】

入力されたテキストから音韻・韻律記号列を生成するテキスト解析手段と、前記音韻・韻律記号列に対して少なくとも音声素片・音韻継続時間・基本周波数の合成パラメータを生成するパラメータ生成手段と、音声の基本単位となる音声素片が登録された素片辞書と前記パラメータ生成手段から生成される合成パラメータに基づいて前記素片辞書を参照しながら波形重畳を行って合成波形を生成する波形生成手段とを備えたテキスト音声変換装置における高速読み上げ制御方法であって、前記パラメータ生成手段は、前記音声素片を変形させて声質を切り換えるための声質変換係数テーブルを備え、ユーザから指定される発声速度が閾値を超えたときには、声質が変化しないような係数を前記声質変換係数テーブルから選択する声質係数決定手段を有することを特徴とするテキスト音声変換装置における高速読み上げ制御方法。

【請求項６】

前記閾値は、所定の最大発声速度であることを特徴とする請求項５記載のテキスト音声変換装置における高速読み上げ制御方法。

【請求項７】

入力されたテキストから音韻・韻律記号列を生成するテキスト解析手段と、前記音韻・韻律記号列に対して少なくとも音声素片・音韻継続時間・基本周波数の合成パラメータを生成するパラメータ生成手段と、音声の基本単位となる音声素片が登録された素片辞書と前記パラメータ生成手段から生成される合成パラメータに基づいて前記素片辞書を参照しながら波形重畳を行って合成波形を生成する波形生成手段とを備えたテキスト音声変換装置における高速読み上げ制御方法であって、前記パラメータ生成手段は、ユーザが指定した抑揚レベルに応じて修正したピッチパタンを出力するピッチパタン修正手段と、ユーザが指定した発声速度に応じて前記修正したピッチパタンを基底ピッチに加算するか否かを選択する切り換え手段とを有し、前記発声速度が所定の閾値を超えた場合には前記基底ピッチを変更しないように前記切り換え手段を制御することを特徴とするテキスト音声変換装置における高速読み上げ制御方法。

【請求項８】

前記閾値は、所定の最大発声速度であることを特徴とする請求項７記載のテキスト音声変換装置における高速読み上げ制御方法。

【請求項９】

前記ピッチパタン修正手段は、ユーザが指定した前記発声速度に応じて統計的手法によりフレーズ成分を算出するか或いは当該フレーズ成分を零とする処理を入力文章中に含まれる全フレーズについて行うフレーズ成分算出処理と、ユーザが指定した前記発声速度に応じて統計的手法によりアクセント成分を算出すると共にユーザが指定した前記抑揚レベルに応じて前記算出したアクセント成分を修正するか或いは当該アクセント成分を零とする処理を入力文章中の全ての単語について行う処理とを含むピッチパタン生成処理を行うことを特徴とする請求項７記載のテキスト音声変換装置における高速読み上げ制御方法。

【発明の詳細な説明】

【０００１】
【発明の属する技術分野】
本発明は、日常読み書きしている漢字・仮名混じり文を音声として出力するテキスト音声変換技術に係わり、特に高速読上げ時の韻律制御に関するものである。
【０００２】
【従来の技術】
テキスト音声変換技術は、我々が日常読み書きしている漢字かな混じり文を入力し、それを音声に変換して出力するもので、出力語彙の制限がないことから録音・再生型の音声合成に代わる技術として種々の利用分野での応用が期待できる。
従来、この種の音声合成装置としては、図１５に示すような処理形態となっているものが代表的である。
【０００３】
日常読み書きしている漢字仮名混じり文（以下テキストと呼ぶ）を入力すると、テキスト解析部１０１は、文字情報から音韻・韻律記号列を生成する。ここで、音韻・韻律記号列とは、入力文の読みに加えて、アクセント、イントネーション等の韻律情報を文字列として記述したもの（以下中間言語と呼ぶ）である。単語辞書１０４は個々の単語の読みやアクセント等が登録された発音辞書で、テキスト解析部１０１はこの発音辞書を参照しながら、形態素解析ならびに構文解析等の言語処理を施して中間言語を生成する。
【０００４】
テキスト解析部１０１で生成された中間言語に基づいて、パラメータ生成部１０２で、音声素片（音の種類）、声質変換係数（声色の種別）、音韻継続時間（音の長さ）、音韻パワー（音の強さ）、基本周波数（声の高さ、以下ピッチと呼ぶ）等の各パタンから成る合成パラメータが決定され、波形生成部１０３に送られる。
【０００５】
ここで音声素片とは、接続して合成波形を作るための音声の基本単位で、音の種類等に応じて様々なものが用意されている。一般的に、ＣＶ、ＶＶ、ＶＣＶ、ＣＶＣ（Ｃ：子音、Ｖ：母音）といった音韻連鎖で構成されている場合が多い。
【０００６】
パラメータ生成部１０２で生成された各種パラメータに基づいて、波形生成部１０３において音声素片等を蓄積するＲＯＭ等から構成された素片辞書１０５を参照しながら、合成波形が生成され、スピーカを通して合成音声が出力される。音声合成方法としては、予め音声波形にピッチマーク（基準点）を付けておき、その位置を中心に切り出して、合成時には合成ピッチ周期に合わせて、ピッチマーク位置をずらしながら重ね合わせる方法が知られている。以上がテキスト音声変換処理の簡単な流れである。
【０００７】
次に、パラメータ生成部１０２における従来の処理を図１６を参照して詳細に説明する。
【０００８】
パラメータ生成部１０２に入力される中間言語は、アクセント位置・ポーズ位置などの韻律情報を含んだ音韻文字列であり、これより、ピッチの時間的な変化（以下ピッチパタン）、音声パワー、それぞれの音韻継続時間、素片辞書内に格納されている音声素片アドレス等の波形を生成する上でのパラメータ（以下、総称して合成パラメータと呼ぶ）を決定する。またこの時、ユーザの好みに合わせた発声様式（発声速度、声の高さ、抑揚の大きさ、声の大きさ、発声話者、声質など）を指定するための制御パラメータも入力される場合がある。
【０００９】
入力された中間言語に対して、中間言語解析部２０１で文字列の解析が行われ、中間言語上に記された呼気段落記号・単語区切り記号から単語境界を判定し、アクセント記号からアクセント核のモーラ（音節）位置を得る。呼気段落とは、一息で発声する区間の区切り単位である。アクセント核とは、アクセントが下降する位置のことで、１モーラ目にアクセント核が存在する単語を１型アクセント、ｎモーラ目にアクセント核が存在する単語を

ｎ型アクセントと呼び、総称して起伏型アクセント単語と呼ぶ。逆に、アクセント核の存在しない単語（例えば「新聞」や「パソコン」）を０型アクセントまたは平板型アクセント単語と呼ぶ。これらの韻律に関わる情報は、ピッチパタン決定部２０２、音韻継続時間決定部２０３、音韻パワー決定部２０４、音声素片決定部２０５、声質係数決定部２０６に送られる。

【００１０】

ピッチパタン決定部２０２は、中間言語上の韻律情報などからアクセント句あるいはフレーズ単位でのピッチ周波数の時間的変化パタンの算出を行う。従来では「藤崎モデル」と呼ばれる、臨界制動２次線形系で記述されるピッチ制御機構モデルが用いられてきた。声の高さの情報を与える基本周波数は、次のような過程で生成されると考えるのがピッチ制御機構モデルである。声帯振動の周波数、すなわち基本周波数は、フレーズの切り替わりごとに発せられるインパルス指令と、アクセントの上げ下げごとに発せられるステップ指令によって制御される。そのとき、生理機構の遅れ特性により、フレーズのインパルス指令は文頭から文末に向かう緩やかな下降曲線（フレーズ成分）となり、アクセントのステップ指令は局所的な起伏の激しい曲線（アクセント成分）となる。これらの二つの成分は、各指令の臨界制動２次線形系の応答としてモデル化され、対数基本周波数の時間変化パターンは、これら両成分の和（以降、抑揚成分と呼ぶ）として表現される。

【００１１】

図１８はピッチ制御機構モデルを示す。対数基本周波数ｌｎ　F_0（ｔ）（ｔは時刻）は、次式のように定式化される。

$$\ln F_0(t) = \ln F_{min} + \sum_{i=1}^{I} A_{pi} G_{pi}(t-T_{0i}) + \sum_{j=1}^{J} A_{aj}\{G_{aj}(t-T_{1j}) - G_{aj}(t-T_{2j})\} \quad \cdots (1)$$

ここで、F_{min}は最低周波数（以下、基底ピッチと呼ぶ）、Ｉは文中のフレーズ指令の数、A_{pi}は文中ｉ番目のフレーズ指令の大きさ、T_{0i}は文中ｉ番目のフレーズ指令の開始時点、Ｊは文内のアクセント指令の数、A_{aj}は文内ｊ番目のアクセント指令の大きさ、T_{1j}、T_{2j}はそれぞれｊ番目のアクセント指令の開始時点と終了時点である。

【００１２】

また、G_{pi}（ｔ）、G_{aj}（ｔ）はそれぞれ、フレーズ制御機構のインパルス応答関数、アクセント制御機構のステップ応答関数であり、次式で与えられる。

$$G_{pi}(t) = \alpha_i^2 t \exp(-\alpha_i t) \quad \cdots (2)$$

$$G_{aj}(t) = \min[1-(1+\beta_j t)\exp(-\beta_j t), \theta] \quad \cdots (3)$$

上式は、ｔ≧０の範囲での応答関数であり、ｔ＜０ではG_{pi}（ｔ）＝G_{aj}（ｔ）＝０である。式（３）の記号ｍｉｎ［ｘ，ｙ］は、ｘ，ｙのうち小さい方をとることを意味しており、実際の音声でアクセント成分が有限の時間で上限に達することに対応している。ここで、α_iはｉ番目のフレーズ指令に対するフレーズ制御機構の固有角周波数であり、例えば３．０などに選ばれる。β_jはｊ番目のアクセント指令に対するアクセント制御機構の固有角周波数であり、例えば２０．０などに選ばれる。また、θはアクセント成分の上限値であり、例えば０．９などに選ばれる。

【００１３】

なおここで、基本周波数およびピッチ制御パラメータ（A_{pi}，A_{aj}，T_{0i}，T_{1j}，T_{2j}，α_i，β_j，F_{min}）の値の単位は次のように定義される。すなわち、F_0（ｔ）およびF_{min}の単位は［Ｈｚ］、T_{0i}，T_{1j}およびT_{2j}の単位は［ｓｅｃ］、α_iおよびβ_jの単位は［ｒａｄ／ｓｅｃ］とする。またA_{pi}およびA_{aj}の値は、基本周波数およびピッチ制御パラメータの値の単位を上記のように定めたときの値を用

いる。

【００１４】

以上で述べた生成過程に基づき、ピッチパタン決定部２０２では、中間言語からピッチ制御パラメータの決定を行う。例えば、フレーズ指令の生起時点T_{0i}は中間言語上での句読点が存在する位置に設定し、アクセント指令の開始時点T_{1j}は単語境界記号直後に設定し、アクセント指令の終了時点T_{2j}はアクセント記号が存在する位置、あるいはアクセント記号がない平板型アクセント単語の場合は、次単語との単語境界記号直前に設定する。フレーズ指令の大きさを表わすA_{pi}とアクセント指令の大きさを表わすA_{aj}は、数量化Ｉ類などの統計的手法を用いて決定する場合が多い。数量化Ｉ類については公知であるのでここでは特に説明はしない。

【００１５】

図１９にピッチパタン生成に関する機能ブロック図を示す。中間言語解析部２０１からの解析結果が制御要因設定部５０１に入力される。制御要因設定部５０１では、フレーズ成分、アクセント成分の大きさを予測するために必要な制御要因の設定を行う。フレーズ成分予測には、例えば、該当するフレーズを構成しているモーラ総数、文内位置、先頭単語のアクセント型といった情報が用いられ、フレーズ成分推定部５０３に送られる。一方、アクセント成分予測には、例えば、該当するアクセント句のアクセント型、構成しているモーラ総数、品詞、フレーズ内位置といった情報が用いられ、アクセント成分推定部５０２に送られる。それぞれの成分値予測には、自然発声データを基に数量化Ｉ類などの統計的手法を用いて予め学習した予測テーブル５０６を用いて行われる。

【００１６】

予測された結果は、ピッチパタン修正部５０４に送られ、ユーザから抑揚指定があった場合は、推定された値A_{pi}、A_{aj}に対しての修正を行う。この機能は、文中のある単語を特に強調あるいは抑制したい時に用いることを想定した制御機構である。通常、抑揚指定は３～５段階に制御され、それぞれのレベルに対してあらかじめ割り当てられた定数を乗ずることにより行われる。抑揚指定がない場合は修正は行われない。

【００１７】

フレーズ・アクセント両成分値の修正が施された後、基底ピッチ加算部５０５に送られ、式（１）に従ってピッチパタンの時系列データが生成される。この時、ユーザからの声の高さ指定レベルに従って、基底ピッチテーブル５０７から指定レベルに応じたデータが基底ピッチとして呼び出され加算される。ユーザから特に指定がない場合は、予め定められたデフォルト値が呼び出され加算される。対数化基底ピッチ$\ln F_{min}$は合成音声の最低ピッチを表わしており、このパラメータが声の高さの制御に用いられている。通常$\ln F_{min}$は、５～１０段階に量子化されてテーブルとして保持されておりユーザの好みによって、全体的に声を高くしたい場合は$\ln F_{min}$を大きくし、逆に声を低くしたい場合は$\ln F_{min}$を小さくするといった処理を行う。

【００１８】

基底ピッチテーブル５０７は、男声音用と女声音用とに分けられており、ユーザから入力される話者指定によって読み出す基底ピッチを選択する。通常男性音の場合は３．０～４．０の範囲内、女性音の場合は４．０～５．０の範囲内で声の高さ指定の段階数に応じて量子化されている。以上がピッチパタン生成過程である。

【００１９】

次に音韻継続時間制御について述べる。音韻継続時間決定部２０３は、音韻文字列・韻律記号などからそれぞれの音韻の長さ、休止区間長を決定する。休止区間とは、フレーズ間、あるいは文章間でのポーズの長さである（以後ポーズ長と呼ぶ）。音韻長は通常、音節を構成している子音・母音の長さの他、破裂性を有する音韻（ｐ，ｔ，ｋなど）の直前に現れる無音長（閉鎖区間長）を、それぞれ決定する。音韻継続時間長、ポーズ長を総称して継続時間長と呼ぶことにする。音韻継続時間の決定方法は通常、目標となる音韻の前後近傍の音韻の種別あるいは、単語内・呼気段落内の音節位置などにより、数量化Ｉ類などの統計的手法が用いられる場合が多い。一方、ポーズ長は、前後隣接するフレーズのモー

ラ総数などにより同じく、数量化Ｉ類などの統計的手法が用いられる。またこの時、ユーザから発声速度を指定された場合は、それに応じて音韻継続時間の伸縮を行う。通常、発声速度指定は、５～１０段階程度に制御され、それぞれのレベルに対してあらかじめ割り当てられた定数を乗ずることにより行われる。発声速度を遅くしたい場合は音韻継続時間を長くし、発声速度を速くしたい場合は音韻継続時間を短くする。音韻継続時間制御に関しては、本発明の主題であるので後述する。

【００２０】

音韻パワー決定部２０４は、音韻文字列からそれぞれの音韻の波形振幅値の算出を行う。波形振幅値は、／ａ，ｉ，ｕ，ｅ，ｏ／などの音韻の種類・呼気段落内での音節位置などから経験的に決められる。また、音節内においても、立ち上がりの徐々に振幅値が大きくなる区間と、定常状態にある区間と、立ち下がりの徐々に振幅値が小さくなる区間のパワー遷移も同時に決定している。これらパワー制御は通常、テーブル化された係数値を用いることにより実行される。またこの時、ユーザからの声の大きさ指定があった場合は、それに応じて振幅値を増減する。通常、声の大きさ指定は、１０段階程度に制御され、それぞれのレベルに対してあらかじめ割り当てられた定数を乗ずることにより行われる。

【００２１】

音声素片決定部２０５は、音韻文字列を表現するために必要な音声素片の、素片辞書１０５内アドレスの決定を行う。素片辞書１０５は、例えば男声音と女性音といった具合に複数話者の音声素片が格納されており、ユーザからの話者指定により素片アドレスの決定を行う。素片辞書１０５に格納されている音声素片データは、ＣＶ、ＶＣＶなど前後の音韻環境に応じた形で様々な単位で構築されているため、入力テキストの音韻文字列の並びから最適な合成単位を選択する。

【００２２】

声質係数決定部２０６は、ユーザから声質変換指定があった場合に、変換パラメータの決定を行う。声質変換とは、素片辞書１０５に登録されている素片データに、信号処理等の加工を施すことにより、聴感上、別話者として取り扱えるようにした機能である。一般に、素片データを線形に伸縮する処理を施して実現する場合が多い。伸長処理は、素片データのオーバーサンプリング処理で実現され、太い声となる。逆に縮小処理は、素片データのダウンサンプリング処理で実現され、細い声となる。通常、声質変換指定は、５～１０段階程度に制御され、それぞれのレベルに対してあらかじめ割り当てられたリサンプリング・レートにより変換を行う。

【００２３】

以上の処理により生成されたピッチパタン・音韻パワー・音韻継続時間・音声素片アドレス・伸縮パラメータは合成パラメータ生成部２０７に送られ、合成パラメータが生成される。合成パラメータは、フレーム（通常８ｍｓ程度の長さ）を一つの単位とした波形生成用のパラメータであり、波形生成部１０３に送られる。

【００２４】

図１７に波形生成部の機能ブロック図を示す。素片復号部３０１では、合成パラメータのうち、素片アドレスを参照ポインタとして素片辞書１０５から素片データをロードし、必要に応じて復号処理を行う。素片辞書１０５には、音声を合成するための元となる音声素片データが格納されており、何らかの圧縮処理が施されている場合は、復号処理を施す。復号された音素片データは、振幅制御部３０２で振幅係数が乗じられてパワー制御が行われる。素片加工部３０３では、声質変換のための素片伸縮処理が施される。声質を太くする場合は素片全体を伸長し、声質を細くする場合は素片全体を縮小するといった処理が施される。重畳制御部３０４では、合成パラメータのうち、ピッチパタンや音韻継続時間といった情報から、素片データの重畳を制御し、合成波形を生成する。波形重畳が完了したデータから逐次ＤＡリングバッファ３０５に書き込み、出力サンプリング周期でＤＡコンバータに転送し、スピーカから出力する。

【００２５】

次に音韻継続時間制御について詳細に説明する。図２０に従来技術による音韻継続時間決

定部の機能ブロック図を示す。中間言語解析部２０１から解析結果が制御要因設定部６０１に入力される。制御要因設定部６０１では、例えば、音韻個々の継続時間長あるいは、単語全体での継続時間長などを予測するために必要な制御要因の設定を行う。予測には、例えば、対象となる音韻、前後の音韻の種類、構成しているフレーズのモーラ総数、文内位置といった情報が用いられ、継続時間推定部６０２に送られる。アクセント成分、フレーズ成分の各成分値予測には、自然発声データを基に数量化Ｉ類などの統計的手法を用いて予め学習した継続時間予測テーブル６０４が用いられる。予測された結果は継続時間修正部６０３に送られ、ユーザから発声速度指定があった場合は予測値の修正が施される。通常、発声速度指定は、５～１０段階程度に制御され、それぞれのレベルに対してあらかじめ割り当てられた定数を乗ずることにより行われる。発声速度を遅くしたい場合は音韻継続時間を長くし、発声速度を速くしたい場合は音韻継続時間を短くする。例えば、発声速度レベルが５段階に制御され、レベル０からレベル４まで指定可能だとする。それぞれのレベルｎに対応した定数Ｔｎを次のように定める。すなわち、

T_0＝２．０、T_1＝１．５、T_2＝１．０、T_3＝０．７５、T_4＝０．５とする。

【００２６】

先に予測された音韻継続時間のうち、母音長とポーズ長に対して、ユーザから指定されたレベルｎに対応した定数T_nが乗じられる。レベル０の場合は２．０が乗じられるので生成される波形は長くなり発声速度は遅くなる。レベル４の場合は０．５が乗じられるので生成される波形は短くなり発声速度は速くなる。上記の例では、レベル２が通常発声速度（デフォルト）となっている。

【００２７】

発声速度制御が施された合成波形の例を図２１に示す。図示したように、音韻継続時間の発声速度制御は通常、母音のみで行う。閉鎖区間長あるいは子音長は、発声速度に依らずほぼ一定と考えられるからである。発声速度を速くした（ａ）図では母音長だけが０．５倍されており、重畳される音声素片数を減じて実現している。逆に発声速度を遅くした（ｃ）図では母音長だけが１．５倍されており、重畳される音声素片数を繰り返し使うなどして実現している。また、ポーズ長に対しては母音長制御と同様に、指定レベルに応じた定数が乗じられるため、発声速度が遅くなるほどポーズ長も長くなり、発声速度が速くなるほどポーズ長も短くなる。

【００２８】

ここで発声速度が速い場合を考える。前述の例ではレベル４に当たる。テキスト音声変換システムの利用特性上、最大発声速度レベルは「早聞き機能」という意味合いが大きい。読上げ対象となるテキストの中でも、ユーザにとって、重要な部分とそうでない部分が存在するため、重要でない部分は発声速度を速くして読み飛ばし、重要な部分は通常発声速度で合成する。このような利用方法が一般的である。最近のテキスト音声変換装置では、早聞き機能用のボタンがあり、このボタンを押下すると発声速度レベルが最大に設定され最高速度で合成され、ボタンを離すと発声速度レベルが以前の設定値に復帰するといったものがある。

【００２９】

【発明が解決しようとする課題】

しかしながら上記の従来技術では、以下に述べる問題があった。

（１）早聞き機能を有効にすると、単純に音韻の継続時間長を短くする、言い換えると、生成する波形の長さを短くする処理を施しているため、波形生成部に負荷がかかるといった問題があった。波形生成部では、波形重畳が完了し、生成された波形データから逐次ＤＡリングバッファに書き込むという処理を行っているため、生成される波形長が短い場合はその分、波形生成処理に費やすことのできる時間が短くなることになる。波形データ長が半分になると、処理時間も半分で終了させなければならない。例えば、音韻継続時間長が半分になったからといって、必ずしも演算量が半分になるわけではないため、ＤＡコンバータへの転送処理に、波形生成処理が追いつかない場合は、合成音が途中で止まる「音切れ」現象が発生する場合がある。

【００３０】
（２）早聞き機能を有効にすると、単純に音韻の継続時間長を短くする処理が施されるため、ピッチパタンも基本的に線形に縮小される。つまり抑揚も時間的に速い周期で変動することになり、これは、不自然なイントネーションで非常に聞き取りにくい合成音となっていた。早聞き機能は、読上げ対象となるテキストを完全にスキップするのではなく、聞き流すという用途で用いられるため、抑揚の激しい合成音は不向きであった。従来技術において早聞き機能有効時の合成音声は、抑揚変化が激しすぎるため聞き取りにくく理解しずらいものとなっていた。
【００３１】
（３）早聞き機能を有効にすると、音韻継続時間と共に、文章間のポーズも同一比率で縮小される。そのため、文章と文章の境界がほとんどなくなり、切れ目が分かり難くなっていた。１文の合成音声を出力した直後に、さらに次の１文の合成音声が出力されるため、従来技術において早聞き機能有効時の合成音声は、テキスト内容を理解しつつ読み飛ばす用途においては不向きであった。
【００３２】
（４）早聞き機能を有効にすると、テキスト全体に渡って、発声速度が速くなるため、早聞き解除のタイミングを取ることが難しかった。通常の早聞き機能使用方法は、ある文章の中から所望の部分までを読み飛ばし、以降を通常速度で合成するというものである。従来技術によると、ユーザが欲した部分の読上げが行われ、早聞き機能解除をした時点では、所望の部分を大きく通り越してしまういった問題があった。この場合、早聞き機能を解除した後に一旦、読上げ対象区間を前にさかのぼって設定した後に通常発声速度で合成開始するといった面倒な操作をしなければいけなかった。またユーザは、必要な部分と必要でない部分とを聞き分けながら、早聞き機能の有効化・無効化の動作を行わなければならず、非常に労力を必要としていた。
【００３３】
本発明は、（Ａ）発声速度を速くした時に高負荷になって音切れが発生するという問題点と、（Ｂ）発声速度を速くした時にピッチ変動周期も速くなり、不自然なイントネーションになってしまうという問題点を解決したテキスト音声変換における高速読み上げ制御方法を提供することを目的とする。
【００３４】
【課題を解決するための手段】
この発明は、上記課題（Ａ）を解決するために、ユーザの指定する発声速度が最高速に設定された場合、すなわち早聞き機能が有効となった場合に、パラメータ生成手段における音韻継続時間決定手段において、統計的手法を用いて予測した継続時間予測テーブルに替えて、予め経験的に求めた継続時間規則テーブルを用いて音韻継続時間を決定し、また、ピッチパタン決定手段において、統計的手法により算出した予測テーブルを用いる代わりに、予め経験的に求めた規則テーブルを使用してピッチパタンを決定し、更に、声質決定手段においては声質が変化しないような声質変換係数を選択する。
【００３５】
また、この発明は、上記課題（Ｂ）を解決するために、ユーザの指定する発声速度が最高速に設定された場合に、アクセント成分及びフレーズ成分の計算を行わないようにすると共に基底ピッチを変更しないようにしている。
【００３８】
【発明の実施の形態】
第１の実施の形態
［構成］
以下、第１の実施の形態における構成を図面を参照しながら詳細に説明する。従来技術と異なる点は、発声速度が最高速に設定された場合、すなわち、早聞き機能が有効となった場合に内部演算処理の一部を簡略化、省略を行うことによって負荷軽減させた点である。
【００３９】

図１は、第１の実施の形態におけるパラメータ生成部１０２の機能ブロック図である。パラメータ生成部１０２への入力は従来と同じく、テキスト解析部１０１から出力される中間言語および、ユーザが個別に指定する韻律制御パラメータである。中間言語解析部８０１には一文毎の中間言語が入力され、以降の韻律生成処理で必要となる音韻系列・フレーズ情報・アクセント情報などといった中間言語解析結果が、それぞれピッチパタン決定部８０２、音韻継続時間決定部８０３、音韻パワー決定部８０４、音声素片決定部８０５、声質係数決定部８０６に出力される。

【００４０】

ピッチパタン決定部８０２には、前述の中間言語解析結果に加えてユーザからの抑揚指定・声の高さ指定・発声速度指定・話者指定の各パラメータが入力され、ピッチパタンが合成パラメータ生成部８０７に出力される。ピッチパタンとは基本周波数の時間的遷移のことである。

【００４１】

音韻継続時間決定部８０３には、前述の中間言語解析結果に加えてユーザからの発声速度指定のパラメータが入力され、それぞれの音韻の音韻継続時間・ポーズ長といったデータが合成パラメータ生成部８０７に出力される。

【００４２】

音韻パワー決定部８０４には、前述の中間言語解析結果に加えてユーザからの声の大きさ指定パラメータが入力され、それぞれの音韻の音韻振幅係数が合成パラメータ生成部８０７に出力される。

【００４３】

音声素片決定部８０５には、前述の中間言語解析結果に加えてユーザからの話者指定パラメータが入力され、波形重畳するための必要な音声素片アドレスが合成パラメータ生成部８０７に出力される。

【００４４】

声質係数決定部８０６には、前述の中間言語解析結果に加えてユーザからの声質指定・発声速度指定の各パラメータが入力され、声質変換パラメータが合成パラメータ生成部８０７に出力される。

【００４５】

合成パラメータ生成部８０７は、入力された各韻律パラメータ（前述したピッチパタン、音韻継続時間、ポーズ長、音韻振幅係数、音声素片アドレス、声質変換係数）から、フレーム（通常８ｍｓ程度の長さ）を一つの単位とした波形生成用のパラメータを生成し、波形生成部１０３に出力する。

【００４６】

パラメータ生成部１０２において、従来技術と比較して異なる点は、発声速度指定パラメータが音韻継続時間決定部８０３のほかに、ピッチパタン決定部８０２、声質係数決定部８０６のそれぞれに入力されている点と、ピッチパタン決定部８０２、音韻継続時間決定部８０３、声質係数決定部８０６のそれぞれの内部処理である。テキスト解析部１０１および波形生成部１０３においては、従来と同様であるため、その構成に関する説明は省略する。

【００４７】

ピッチパタン決定部８０２の構成について図２を用いて説明する。第１の実施の形態においては、アクセント成分およびフレーズ成分の決定に、数量化Ｉ類等の統計的手法を用いる場合と規則による場合との２通りの構成を有する。規則による制御の場合は、予め経験的に求められた規則テーブル９１０を用い、統計的手法による制御の場合は、自然発声データを基に数量化Ｉ類などの統計的手法を用いて予め学習した予測テーブル９０９を用いる。予測テーブル９０９のデータ出力はスイッチ９０７のａ端子に接続され、規則テーブル９１０のデータ出力はスイッチ９０７のｂ端子に接続される。いずれの端子が選択されるかは、セレクタ９０６の出力によって決定される。

【００４８】

セレクタ９０６には、ユーザから指定される発声速度レベルが入力され、スイッチ９０７を制御するための信号がスイッチ９０７に接続される。発声速度が最高レベルの場合はスイッチ９０７をｂ端子側に接続し、それ以外の場合はスイッチ９０７をａ端子側に接続する。スイッチ９０７の出力は、アクセント成分決定部９０２とフレーズ成分決定部９０３に接続される。

【００４９】

中間言語解析部８０１からの出力は制御要因設定部９０１に入力され、アクセント・フレーズ両成分の決定のための要因パラメータの解析が行われ、その出力がアクセント成分決定部９０２とフレーズ成分決定部９０３に接続される。

【００５０】

アクセント成分決定部９０２とフレーズ成分決定部９０３には、スイッチ９０７からの出力が接続されており、予測テーブル９０９もしくは規則テーブル９１０を用いてそれぞれの成分値を決定しピッチパタン修正部９０４に出力する。

【００５１】

ピッチパタン修正部９０４には、ユーザから指定される抑揚指定レベルが入力され、該レベルに応じて予め定められた定数が乗じられ、その結果が基底ピッチ加算部９０５に接続される。

【００５２】

基底ピッチ加算部９０５にはさらに、ユーザから指定される声の高さレベル・話者指定および、基底ピッチテーブル９０８が接続されている。基底ピッチテーブル９０８には、ユーザ指定された声の高さレベルと性別とに応じて予め定められた定数値が格納されており、ピッチパタン修正部９０４からの入力に加算してピッチパタン時系列データとして合成パラメータ生成部８０７に出力する。

【００５３】

音韻継続時間決定部８０３の構成について図３を用いて説明する。第１の実施の形態においては、音韻継続時間の決定に、数量化Ｉ類等の統計的手法を用いる場合と規則による場合との２通りの構成を有する。規則による制御の場合は、予め経験的に求められた継続時間規則テーブル１００７を用い、統計的手法による制御の場合は、自然発声データを基に数量化Ｉ類などの統計的手法を用いて予め学習した継続時間予測テーブル１００６を用いる。継続時間予測テーブル１００６のデータ出力はスイッチ１００５のａ端子に接続され、継続時間規則テーブル１００７のデータ出力はスイッチ１００５のｂ端子に接続される。いずれの端子が選択されるかは、セレクタ１００４の出力によって決定される。

【００５４】

セレクタ１００４には、ユーザから指定される発声速度レベルが入力され、スイッチ１００５を制御するための信号がスイッチ１００５に接続される。発声速度が最高レベルの場合はスイッチ１００５をｂ端子側に接続し、それ以外の場合はスイッチ１００５をａ端子側に接続する。スイッチ１００５の出力は、継続時間決定部１００２に接続される。

【００５５】

中間言語解析部８０１からの出力は制御要因設定部１００１に入力され、音韻継続時間決定のための要因パラメータの解析が行われ、その出力が継続時間決定部１００２に接続される。

【００５６】

継続時間決定部１００２には、スイッチ１００５からの出力が接続されており、継続時間予測テーブル１００６もしくは継続時間規則テーブル１００７を用いて音韻継続時間長を決定し継続時間修正部１００３に出力する。継続時間修正部１００３には、ユーザから指定される発声速度レベルが入力され、該レベルに応じて予め定められた定数が乗じられて修正が施され、その結果が合成パラメータ生成部８０７に出力される。

【００５７】

声質係数決定部８０６の構成について図４を用いて説明する。この例では声質変換指定レベルは５段階となっている。ユーザから指定される発声速度レベルおよび声質指定レベル

がセレクタ１１０２に入力され、スイッチ１１０３を制御するための信号がスイッチ１１０３に接続される。この時のスイッチ制御信号は、発声速度が最高レベルの場合は無条件でｃ端子有効にし、それ以外の場合は、声質指定レベルに応じた端子が有効となる。すなわち、声質レベルが０の時はａ端子、レベル１の時はｂ端子、以下同様にレベル４の時ｅ端子がそれぞれ有効となる。スイッチ１１０３のａ～ｅの各端子は、声質変換係数テーブル１１０４に接続され、それぞれに対応した声質変換係数データが呼び出され、スイッチ１１０３の出力として声質係数選択部１１０１に接続される。声質係数選択部１１０１は入力された声質変換係数を合成パラメータ生成部８０７に出力する。

【００５８】

［動作］

以上のように構成された第１の実施の形態における動作について詳細に説明する。従来技術と異なる点は、パラメータ生成に関わる処理であるので、それ以外の処理については説明を省略する。

【００５９】

テキスト解析部１０１で生成された中間言語は、パラメータ生成部１０２内部の中間言語解析部８０１に送られる。中間言語解析部８０１では、中間言語上に記述されているフレーズ区切り記号、単語区切り記号、アクセント核を示すアクセント記号、そして音韻記号列から、韻律生成に必要なデータを抽出して、ピッチパタン決定部８０２、音韻継続時間決定部８０３、音韻パワー決定部８０４、音声素片決定部８０５、声質係数決定部８０６のそれぞれの機能ブロックへ送る。

【００６０】

ピッチパタン決定部８０２では、声の高さの遷移であるイントネーションが生成され、音韻継続時間決定８０３では、音韻個々の継続時間のほか、フレーズとフレーズの切れ目あるいは、文と文との切れ目に挿入するポーズ長を決定する。また、音韻パワー決定部８０４では、音声波形の振幅値の遷移である音韻パワーが生成され、音声素片決定部８０５では合成波形を生成するために必要となる音声素片の、素片辞書１０５におけるアドレスを決定する。声質係数決定部８０６では、素片データを信号処理で加工するためのパラメータの決定が行われる。ユーザから指定される韻律制御指定のうち、抑揚指定および声の高さ指定はピッチパタン決定部８０２に、発声速度指定はピッチパタン決定部８０２と音韻継続時間決定部８０３と声質係数決定部８０６に、声の大きさ指定は音韻パワー決定部８０４に、話者指定はピッチパタン決定部８０２と音声素片決定部８０５に、声質指定は声質係数決定部８０６にそれぞれ送られている。

【００６１】

以下に、それぞれの機能ブロックごとに動作の説明を行う。

まず、図２を用いて、ピッチパタン決定部８０２の動作を詳細に説明する。中間言語解析部２０１から解析結果が制御要因設定部９０１に入力される。制御要因設定部９０１では、フレーズ成分、アクセント成分の大きさを決定するために必要な制御要因の設定を行う。フレーズ成分の大きさの決定に必要なデータとは、例えば、該当するフレーズを構成しているモーラ総数、文内での相対位置、先頭単語のアクセント型といった情報である。一方、アクセント成分の大きさの決定に必要なデータとは、例えば、該当するアクセント句のアクセント型、構成しているモーラ総数、品詞、フレーズ内での相対位置といった情報である。これらの成分値を決定するために予測テーブル９０９あるいは、規則テーブル９１０が使用される。前者は、自然発声データを基に数量化Ｉ類などの統計的手法を用いて予め学習したテーブルであり、後者は、予備実験等の実施により経験的に導き出された成分値が格納されたテーブルである。数量化Ｉ類に関しては公知であるのでここでは説明を省略する。どちらが選択されるかはスイッチ９０７により制御され、スイッチ９０７がａ端子に接続された場合は予測テーブル９０９が、ｂ端子に接続された場合は規則テーブル９１０が選択されることになる。

【００６２】

ピッチパタン決定部８０２には、ユーザから指定される発声速度レベルが入力されており

、これによりセレクタ９０６を介してスイッチ９０７が駆動されている。セレクタ９０６は、入力された発声速度レベルが最高速度であった時、スイッチ９０７をｂ端子側に接続するような制御信号を送信する。逆に、入力された発声速度レベルが最高速度ではない時、スイッチ９０７をａ端子側に接続するような制御信号を送信する。例えば、発声速度が５段階、レベル０からレベル４まで設定でき、数値が大きくなる程発声速度が速くなる仕様の場合、セレクタ９０６は、入力された発声速度レベルが４の時だけスイッチ９０７をｂ端子に接続するような制御信号を送信し、それ以外の時はａ端子に接続するような制御信号を送信する。すなわち、発声速度が最高速度の場合は規則テーブル９１０が選択され、そうでない場合は予測テーブル９０９が選択されることになる。

【００６３】

アクセント成分決定部９０２とフレーズ成分決定部９０３は、選択されたテーブルを用いてそれぞれの成分値の算出を行う。予測テーブル９０９が選択された場合は、統計的手法を用いてアクセント・フレーズ両成分の大きさを決定する。規則テーブル９１０が選択された場合は、あらかじめ決められた規則に従ってアクセント・フレーズ両成分の大きさを決定する。例えばフレーズ成分の大きさの規則化の例としては、文内の位置で決定し、文先頭フレーズは一律に０．３、文終端フレーズは一律に０．１、それ以外の文中フレーズは０．２などが考えられる。アクセント成分の大きさに関しても、アクセント型が１型の時とそれ以外の時、フレーズ内での単語位置が先頭の場合とそうでない場合といった具合に場合分けして、それぞれの条件に対して成分値を割り当てておく。このような構成にすることで、フレーズ・アクセント両成分値の決定はテーブル参照を行うだけで行える。本発明におけるピッチパタン決定部の主題は、統計的手法を用いてフレーズ・アクセント成分の大きさを決定する場合と比較して、演算量が少なく済み、処理時間の短縮が図れるモードを有する構成にすることである。したがって、規則化手順は上記に限られるものではない。

【００６４】

以上のような処理が施され決定したアクセント成分、フレーズ成分は、ピッチパタン修正部９０４で抑揚制御が行われ、基底ピッチ加算部９０５で声の高さ制御が施される。

【００６５】

ピッチパタン修正部９０４はユーザから指定される抑揚制御レベルに応じた係数を乗ずる操作が行われる。ユーザからの抑揚制御指定は例えば、３段階で与えられ、レベル１が抑揚を１．５倍に、レベル２が抑揚を１．０倍に、レベル３が抑揚を０．５倍にといった具合に定められている。

【００６６】

基底ピッチ加算部９０５では、抑揚修正されたアクセント成分、フレーズ成分に対して、ユーザから指定される声の高さレベルあるいは、話者指定（性別）に応じた定数を加算する操作が行われ、ピッチパタン時系列データとして合成パラメータ生成部８０７に送られる。例えば、声の高さレベルが５段階、レベル０からレベル４まで設定できるシステムの場合、基底ピッチテーブル９０８に格納されているデータは男声音の場合、３．０、３．２、３．４、３．６、３．８といった数値、女性音の場合は、４．０、４．２、４．４、４．６、４．８といった数値が良く用いられる。

【００６７】

次に音韻継続時間制御について図３を用いてその動作について詳細に説明する。中間言語解析部２０１から解析結果が制御要因設定部１００１に入力される。制御要因設定部１００１では、音韻継続時間（子音長・母音長・閉鎖区間長）、ポーズ長を決定するために必要な制御要因の設定を行う。音韻継続時間の決定に必要なデータとは、例えば、目標となる音韻の種別、対象音節の前後近傍の音韻の種別あるいは、単語内・呼気段落内の音節位置といった情報である。一方、ポーズ長決定に必要なデータとは、前後隣接するフレーズのモーラ総数といった情報である。これらの継続時間長を決定するために継続時間予測テーブル１００６あるいは、継続時間規則テーブル１００７が使用される。前者は、自然発声データを基に数量化Ｉ類などの統計的手法を用いて予め学習したテーブルであり、後者

は、予備実験等の実施により経験的に導き出された成分値が格納されたテーブルである。どちらが選択されるかはスイッチ１００５により制御され、スイッチ１００５がａ端子に接続された場合は継続時間予測テーブル１００６が、ｂ端子に接続された場合は継続時間規則テーブル１００７が選択されることになる。

【００６８】

音韻継続時間決定部８０３には、ユーザから指定される発声速度レベルが入力されており、これによりセレクタ１００４を介してスイッチ１００５が駆動されている。セレクタ１００４は、入力された発声速度レベルが最高速度であった時、スイッチ１００５をｂ端子側に接続するような制御信号を送信する。逆に、入力された発声速度レベルが最高速度ではない時は、スイッチ１００５をａ端子側に接続するような制御信号を送信する。例えば、発声速度が５段階、レベル０からレベル４まで設定でき、数値が大きくなる程発声速度が速くなる仕様の場合、セレクタ１００４は、入力された発声速度レベルが４の時だけスイッチ１００５をｂ端子に接続するような制御信号を送信し、それ以外の時はａ端子に接続するような制御信号を送信する。すなわち、発声速度が最高速度の場合は継続時間規則テーブル１００７が選択され、そうでない場合は継続時間予測テーブル１００６が選択されることになる。

【００６９】

継続時間決定部１００２は、選択されたテーブルを用いて音韻継続時間、ポーズ長の算出を行う。継続時間予測テーブル１００６が選択された場合は、統計的手法を用いて決定する。継続時間規則テーブル１００７が選択された場合は、あらかじめ決められた規則に従って決定する。例えば音韻継続時間の規則化の例としては、その音韻の種類、文内の位置などに応じて基本長を割り当てておく。大量の自然発声データから音韻毎に平均を算出し、これを基本長としてもよい。ポーズ長に関しては、一律に３００ｍｓを割り当てるか、あるいは、テーブル参照を行うだけで決定できるような構成が望ましい。本実施の形態における音韻継続時間決定部の主題は、統計的手法を用いて継続時間を決定する場合と比較して、演算量が少なく済み、処理時間の短縮が図れるモードを有する構成にすることである。したがって、規則化手順は上記に限られるものではない。

【００７０】

以上のような処理が施され決定した継続時間は、継続時間修正部１００３に送られる。継続時間修正部１００３には、ユーザから指定される発声速度レベルも同時に入力されており、このレベルに応じて音韻継続時間の伸縮を行う。通常、発声速度指定は、５～１０段階程度に制御され、それぞれのレベルに対してあらかじめ割り当てられた定数を母音の継続時間長あるいは、ポーズ長に対して乗ずることにより行われる。発声速度を遅くしたい場合は音韻継続時間を長くし、発声速度を速くしたい場合は音韻継続時間を短くする。

【００７１】

次に声質係数決定について図４を用いてその動作について詳細に説明する。声質係数決定部８０６には、ユーザから指定される声質変換レベルと、発声速度レベルが入力される。これらの韻律制御パラメータは、セレクタ１１０２を介してスイッチ１１０３を制御するために用いられる。セレクタ１１０２はまず、発声速度レベルの判定を行う。発声速度レベルが最高速度の場合は、スイッチ１１０３をｃ端子に接続し、最高速度以外の場合は、声質変換レベルの判定を行う。この時は、声質変換レベルに応じた端子に接続するようにスイッチ１１０３を制御する。声質指定レベルが０の時はａ端子、レベル１の時はｂ端子、以下同様にレベル４の時はｅ端子に接続する。スイッチ１１０３のａ～ｅの各端子は、声質変換係数テーブル１１０４に接続され、それぞれに対応した声質変換係数データが呼び出される機能になっている。

【００７２】

声質変換係数テーブル１１０４には、音声素片の伸縮係数が格納されており、例えば声質変換レベルｎに対応する伸縮係数をK_nを次のように定める。すなわち、

K_0＝２．０、K_1＝１．５、K_2＝１．０、K_3＝０．８、K_4＝０．５

のように設定する。これらの数値は、元となる音声素片の長さをK_n倍に伸縮した後に波

形重畳して合成音声を生成するという意味である。レベル２の時は、係数値が１．０なので声質変換のための処理は一切行われないことになる。スイッチ１１０３のａ端子に接続されている場合は、係数K_0が選択されて声質係数選択部１１０１に送られる。スイッチ１１０３のｂ端子に接続されている場合は、係数K_1が選択されて声質係数選択部１１０１に送られるといった具合である。

【００７３】

ここで、図５を参照しながら素片の線形伸縮の方法の一例について述べる。声質変換レベルｎにおける音声素片のデータの第ｍサンプル目をX_{nm}とする。このように定義すると、声質変換後のデータ系列は、変換前のデータ系列X_{2n}を用いて以下のようにして算出することができる。即ち、

レベル０では、

$$X_{00} = X_{20}$$
$$X_{01} = X_{20} \times 1/2 + X_{21} \times 1/2$$
$$X_{02} = X_{21}$$

レベル１では、

$$X_{10} = X_{20}$$
$$X_{11} = X_{20} \times 1/3 + X_{21} \times 2/3$$
$$X_{12} = X_{21} \times 2/3 + X_{22} \times 1/3$$
$$X_{13} = X_{22}$$

レベル３では、

$$X_{30} = X_{20}$$
$$X_{31} = X_{21} \times 3/4 + X_{22} \times 1/4$$
$$X_{32} = X_{22} \times 1/2 + X_{23} \times 1/2$$
$$X_{33} = X_{23} \times 1/4 + X_{24} \times 3/4$$
$$X_{34} = X_{25}$$

レベル４では、

$$X_{40} = X_{20}$$
$$X_{41} = X_{22}$$

のようになる。上記は、声質変換のための一例であって、これに限られるものではない。本実施の形態における声質係数決定部の主題は、発声速度レベルが最高速の時に声質変換指定を無効とする機能を有することにより、処理時間の短縮を図ることである。

【００７４】

以上詳細に説明したように、第１の実施の形態によれば、発声速度が既定値最大に設定された場合に、テキスト音声変換処理の中で演算負荷が大きい機能ブロックを簡略化あるいは、無効にする処理を施しているため、高負荷による音切れが発生する機会を減少させ、聞き易い合成音声を生成することが可能となる。

【００７５】

この場合、発声速度が最高レベル以外に設定された時の合成音と比較して、ピッチや継続時間などの韻律性能の若干の違い、声質変換機能が有効とならない、といったことが起きるが、最高速度での合成音出力は通常、読み飛ばしという意味合いで利用される場合がほとんどある。したがって、音声出力されるテキストの内容を把握・理解できれば良い、という程度の使用方法なので声質変換機能の有無、あるいは韻律性能低下といった点は音切れ現象と比較すると許容できるものと考えられる。

【００７６】

第２の実施の形態

［構成］

第２の実施の形態における構成を図面を参照しながら詳細に説明する。本実施の形態が従来技術と異なる点は、発声速度が最高速に設定された場合、すなわち、早聞き機能が有効となった時にピッチパタン生成処理を変更する点である。したがって、従来と異なるパラメータ生成部、ピッチパタン決定部についてのみ説明する。

【００７７】
図６は第２の実施の形態におけるパラメータ生成部の機能ブロック図を示しており、このブロック図を用いて説明する。パラメータ生成部１０２への入力は従来と同じく、テキスト解析部１０１から出力される中間言語および、ユーザが個別に指定する韻律制御パラメータである。中間言語解析部１３０１には一文毎の中間言語が入力され、以降の韻律生成処理で必要となる音韻系列・フレーズ情報・アクセント情報などといった中間言語解析結果が、それぞれピッチパタン決定部１３０２、音韻継続時間決定部１３０３、音韻パワー決定部１３０４、音声素片決定部１３０５、声質係数決定部１３０６に出力される。
【００７８】
ピッチパタン決定部１３０２には、前述の中間言語解析結果に加えてユーザからの抑揚指定・声の高さ指定・発声速度指定・話者指定の各パラメータが入力され、ピッチパタンが合成パラメータ生成部１３０７に出力される。
【００７９】
音韻継続時間決定部１３０３には、前述の中間言語解析結果に加えてユーザからの発声速度指定のパラメータが入力され、それぞれの音韻継続時間・ポーズ長といったデータが合成パラメータ生成部１３０７に出力される。
【００８０】
音韻パワー決定部１３０４には、前述の中間言語解析結果に加えてユーザからの声の大きさ指定パラメータが入力され、それぞれの音韻振幅係数が合成パラメータ生成部１３０７に出力される。
【００８１】
音声素片決定部１３０５には、前述の中間言語解析結果に加えてユーザからの話者指定パラメータが入力され、波形重畳するための必要な音声素片アドレスが合成パラメータ生成部１３０７に出力される。
【００８２】
声質係数決定部１３０６には、前述の中間言語解析結果に加えてユーザからの声質指定・発声速度指定の各パラメータが入力され、声質変換パラメータが合成パラメータ生成部１３０７に出力される。
【００８３】
合成パラメータ生成部１３０７は、入力された各韻律パラメータ（前述したピッチパタン、音韻継続時間、ポーズ長、音韻振幅係数、音声素片アドレス、声質変換係数）を、フレーム（通常８ｍｓ程度の長さ）を一つの単位とした波形生成用のパラメータに変換し、波形生成部１０３に出力する。
【００８４】
パラメータ生成部１０２において、従来技術と比較して異なる点は、発声速度指定パラメータが音韻継続時間決定部１３０３のほかに、ピッチパタン決定部１３０２に入力されている点と、ピッチパタン決定部１３０２の内部処理である。テキスト解析部１０１および波形生成部１０３においては、従来と同様であるため、その構成に関する説明は省略する。また、パラメータ生成部１０２の内部機能ブロックにおいても、ピッチパタン決定部１３０２以外は従来と同様であるため、その構成に関する説明は省略する。
【００８５】
ピッチパタン決定部１３０２の構成について図７を用いて説明する。中間言語解析部１３０１からの出力は制御要因設定部１４０１に入力され、アクセント・フレーズ両成分の決定のための要因パラメータの解析が行われ、その出力がアクセント成分決定部１４０２とフレーズ成分決定部１４０３に接続される。
【００８６】
アクセント成分決定部１４０２とフレーズ成分決定部１４０３には、予測テーブル１４０８が接続され、数量化Ｉ類等の統計的手法を用いてそれぞれの成分の大きさを予測する。予測されたアクセント成分値、フレーズ成分値はピッチパタン修正部１４０４に接続される。

【００８７】
ピッチパタン修正部１４０４にはユーザから指定される抑揚指定レベルが入力され、該レベルに応じて予め定められた定数が前述のアクセント成分、フレーズ成分に乗じられ、その結果がスイッチ１４０５のａ端子に接続される。スイッチ１４０５にはさらにｂ端子が存在し、セレクタ１４０６から出力される制御信号により、端子ａ、端子ｂのいずれかに接続されるように構成されている。
【００８８】
セレクタ１４０６には、ユーザから指定される発声速度レベルが入力され、発声速度が最高レベルの場合はスイッチ１４０５をｂ端子に接続し、それ以外の場合はスイッチ１４０５をａ端子に接続する制御信号を出力する。スイッチ１４０５のｂ端子は常にグランドに接続されており、スイッチ１４０５は、ａ端子が有効の時はピッチパタン修正部１４０４からの出力を、ｂ端子が有効の時は０を基底ピッチ加算部１４０７に出力する機能を有している。
【００８９】
基底ピッチ加算部１４０７にはさらに、ユーザから指定される声の高さレベル・話者指定および、基底ピッチテーブル１４０９が接続されている。基底ピッチテーブル１４０９には、ユーザ指定された声の高さレベルと話者の性別に応じて予め定められた定数値が格納されており、スイッチ１４０５からの入力に加算してピッチパタン時系列データとして合成パラメータ生成部１３０７に出力する。
【００９０】
［動作］
以上のように構成された本発明の第２の実施の形態における動作について詳細に説明する。
【００９１】
まず、テキスト解析部１０１で生成された中間言語は、パラメータ生成部１０２内部の中間言語解析部１３０１に送られる。中間言語解析部１３０１では、中間言語上に記述されているフレーズ区切り記号、単語区切り記号、アクセント核を示すアクセント記号、そして音韻記号列から、韻律生成に必要なデータを抽出して、ピッチパタン決定部１３０２、音韻継続時間決定部１３０３、音韻パワー決定部１３０４、音声素片決定部１３０５、声質係数決定部１３０６のそれぞれの機能ブロックへ送る。
【００９２】
ピッチパタン決定部１３０２では、声の高さの遷移であるイントネーションが生成され、音韻継続時間決定１３０３では、音韻個々の継続時間のほか、フレーズとフレーズの切れ目あるいは、文と文との切れ目に挿入するポーズ長を決定する。また、音韻パワー決定部１３０４では、音声波形の振幅値の遷移である音韻パワーが生成され、音声素片決定部１３０５では合成波形を生成するために必要となる音声素片の、素片辞書１０５におけるアドレスを決定する。声質係数決定部１３０６では、素片データを信号処理で加工するためのパラメータの決定が行われる。
【００９３】
ユーザから指定される種々の韻律制御指定のうち、抑揚指定および声の高さ指定はピッチパタン決定部１３０２に、発声速度指定はピッチパタン決定部１３０２と音韻継続時間決定部１３０３に、声の大きさ指定は音韻パワー決定部１３０４に、話者指定はピッチパタン決定部１３０２と音声素片決定部１３０５に、声質指定は声質係数決定部１３０６にそれぞれ送られている。
【００９４】
以下に図７を用いてピッチパタン決定部１３０２の動作に関して説明する。従来技術と異なる点は、ピッチパタン生成に関わる処理であるので、それ以外の処理については省略する。
【００９５】
中間言語解析部２０１から解析結果が制御要因設定部１４０１に入力される。制御要因設

定部１４０１では、フレーズ成分、アクセント成分の大きさを予測するために必要な制御要因の設定を行う。フレーズ成分の大きさの予測に必要なデータとは、例えば、該当するフレーズを構成しているモーラ総数、文内での相対位置、先頭単語のアクセント型といった情報である。一方、アクセント成分の大きさの予測に必要なデータとは、例えば、該当するアクセント句のアクセント型、構成しているモーラ総数、品詞、フレーズ内での相対位置といった情報である。これらの成分値を決定するために予測テーブル１４０８が使用される。予測テーブル１４０８は、自然発声データを基に数量化Ｉ類などの統計的手法を用いて予め学習したテーブルである。数量化Ｉ類に関しては公知であるのでここでは説明を省略する。

【００９６】

制御要因設定部１４０１で解析された予測制御要因は、アクセント成分決定部１４０２とフレーズ成分決定部１４０３に送られ、それぞれにおいてアクセント成分の大きさ、フレーズ成分の大きさが予測テーブル１４０８を用いて予測される。第１の実施の形態でも示したように、予測モデルを使わずに規則でそれぞれの成分値を決定しても構わない。算出されたアクセント成分、フレーズ成分は、ピッチパタン修正部１４０４に送られ、ユーザから指定される抑揚指定レベルに応じた係数を乗ずる操作が行われる。

【００９７】

ユーザからの抑揚制御指定は例えば、３段階で与えられ、レベル１が抑揚を１．５倍に、レベル２が抑揚を１．０倍に、レベル３が抑揚を０．５倍にといった具合に定められている。

【００９８】

修正されたアクセント、フレーズ両成分はスイッチ１４０５のａ端子に送られる。スイッチ１４０５は、ａ、ｂ、２つの端子を有しており、セレクタ１４０６からの制御信号によりどちらかの端子に接続するような機能になっている。一方のｂ端子は常に０が入力されるようになっている。

【００９９】

セレクタ１４０６にはユーザからの発声速度レベルが入力されており、これにより出力制御が行われている。セレクタ１４０６は、入力された発声速度レベルが最高速度であった時、スイッチ１４０５をｂ端子側に接続するような制御信号を送信する。逆に、入力された発声速度レベルが最高速度ではない時、スイッチ１４０５をａ端子側に接続するような制御信号を送信する。例えば、発声速度が５段階、レベル０からレベル４まで設定でき、数値が大きくなる程発声速度が速くなる仕様の場合、セレクタ１４０６は、入力された発声速度レベルが４の時だけスイッチ１４０５をｂ端子に接続するような制御信号を送信し、それ以外の時はａ端子に接続するような制御信号を送信する。すなわち、発声速度が最高速度の場合は０が選択され、そうでない場合は、ピッチパタン修正部１４０４の出力である修正されたアクセント成分値とフレーズ成分値が選択されることになる。

【０１００】

選択されたデータは基底ピッチ加算部１４０７に送られる。基底ピッチ加算部１４０７にはユーザからの声の高さ指定レベルが入力されており、基底ピッチテーブル１４０９から該レベルに対応する基底ピッチデータが読み出され、前述のスイッチ１４０５からの出力値との加算処理が施され、ピッチパタンの時系列データとして合成パラメータ生成部１３０７に出力される。

【０１０１】

例えば、声の高さレベルが５段階、レベル０からレベル４まで設定できるシステムの場合、基底ピッチテーブル１４０９に格納されているデータは男声音の場合、３．０、３．２、３．４、３．６、３．８といった数値、女性音の場合は、４．０、４．２、４．４、４．６、４．８といった数値が良く用いられる。

【０１０２】

上記の例では、ピッチパタン修正部１４０４の出力と数値０とをスイッチ１４０５で切り替える処理を行っているが、無論、発声速度指定が最高レベルの時は、制御要因設定部１

４０１からピッチパタン修正部１４０４までの処理は不要になる。

【０１０３】

図８に第２の実施の形態におけるピッチパタン生成処理のフローチャートを示す。ここで図中の記号は以下の通りとする。すなわち、入力文章中に含まれるフレーズ総数をＩ、単語総数をＪ、第ｉ番目のフレーズ成分の大きさをA_{pi}、第ｊ番目のアクセント成分の大きさをA_{aj}、第ｊ番目のアクセント句に対して指定される抑揚制御係数E_j、とする。

【０１０４】

ステップＳＴ１０１からステップＳＴ１０６にかけては、フレーズ成分の大きさA_{pi}の算出を行う。まずステップＳＴ１０１で、フレーズカウンタｉを０に初期化する。次いでステップＳＴ１０２で発声速度レベルの判定を行い、発声速度が最高速度である場合はステップＳＴ１０４に進み、そうでない場合はステップＳＴ１０３に進む。ステップＳＴ１０４では、第ｉ番目のフレーズ成分の大きさA_{pi}を０に設定してステップＳＴ１０５に進む。一方ステップＳＴ１０３では数量化Ｉ類などの統計的手法を用いて第ｉ番目のフレーズ成分の大きさA_{pi}が予測され、ステップＳＴ１０５に進む。ステップＳＴ１０５においては、フレーズカウンタｉを１インクリメントする。次いでステップＳＴ１０６で入力文章中のフレーズ総数Ｉとの比較を行い、フレーズカウンタｉが文内フレーズ総数Ｉを超えた場合、すなわち全てのフレーズに対する処理が終了した場合にフレーズ成分生成処理を終え、ステップＳＴ１０７に進む。そうでない場合は、ステップＳＴ１０２に戻り次のフレーズに対する処理を前述と同様に繰り返す。

【０１０５】

ステップＳＴ１０７からステップＳＴ１１３にかけては、アクセント成分の大きさA_{aj}の算出を行う。まずステップＳＴ１０７で、単語カウンタｊを０に初期化する。次いでステップＳＴ１０８で発声速度レベルの判定を行い、発声速度が最高速度である場合はステップＳＴ１１１に進み、そうでない場合はステップＳＴ１０９に進む。ステップＳＴ１１１では、第ｊ番目のアクセント成分の大きさA_{aj}を０に設定してステップＳＴ１１２に進む。一方ステップＳＴ１０９では数量化Ｉ類などの統計的手法を用いて第ｊ番目のアクセント成分の大きさA_{aj}が予測され、ステップＳＴ１１０に進む。ステップＳＴ１１０では、第ｊ番目のアクセント句に対して抑揚修正処理が下式により行われる。

$$A_{aj} = A_{aj} \times E_j \quad \cdots (4)$$

【０１０６】

ここでＥｊは、ユーザが指定する抑揚制御レベルに応じてあらかじめ定められている抑揚制御係数であり、先にも説明したように例えば抑揚制御レベルが３段階で与えられ、レベル０が抑揚を１．５倍に、レベル１が抑揚を１．０倍に、レベル２が抑揚を０．５倍にといった場合は以下のようになる。

レベル０（抑揚を１．５倍）　$E_j = 1.5$

レベル１（抑揚を１．０倍）　$E_j = 1.0$

レベル２（抑揚を０．５倍）　$E_j = 0.5$

【０１０７】

抑揚修正終了後ステップＳＴ１１２に進む。ステップＳＴ１１２においては、単語カウンタｊを１インクリメントする。次いでステップＳＴ１１３で入力文章中の単語総数Ｊとの比較を行い、単語カウンタｊが文内単語総数Ｊを超えた場合、すなわち全て単語に対する処理が終了した場合にアクセント成分生成処理を終え、ステップＳＴ１１４に進む。そうでない場合は、ステップＳＴ１０８に戻り次のアクセント句に対する処理を前述と同様に繰り返す。

【０１０８】

ステップＳＴ１１４では、上記の処理で決定されたフレーズ成分値A_{pi}とアクセント成分値A_{aj}、基底ピッチテーブル１４０９を参照して得られる基底ピッチ$\ln F_{min}$とから式（１）によりピッチパタンを生成する。

【０１０９】

以上詳細に説明したように本発明の第２の実施の形態によれば、発声速度が既定値最大に

設定された場合に、ピッチパタンの抑揚成分を０にしてピッチパタン生成を行うため、時間的に速い周期で抑揚が変動することがなくなり、非常に聞き取りにくい合成音となることが解消される。

【０１１０】

図９は従来技術における発声速度によるピッチパタンの違いの説明図である。上段（ａ）が通常発声速度の場合であり、下段（ｂ）が最高速度の場合である。横軸が時間であり、図中点線で示す曲線がフレーズ成分を表わし、実線で示す曲線がアクセント成分に対応している。最高速度が通常速度の２倍だとすると、生成される波形は通常時の約１／２となる。（$T_2 = T_1 / 2$）ピッチパタンの遷移も発声速度に比例して速くなるため、合成音声の抑揚は非常に速い周期での変動となることが図を見ても分かる。しかし実際の発声においては発声速度に応じて、フレーズの結合によるフレーズ境界の消失、アクセント結合によるアクセント句境界の消失といった現象が見られるため図（ｂ）のようにはならない。発声速度が速くなるにつれて、ピッチパタンの変化も相対的に緩やかになることが多い。

【０１１１】

例えば図９の例で言えば２つのフレーズで構成されているが、これが１つのフレーズとして結合するといった現象が確認されている。従来技術においては、この点を考慮に入れておらず、非常に聞きづらい合成音声となっていたが、第２の実施の形態によれば、抑揚成分を０にすることで聞き取り易い合成音声を生成することが可能となる。

【０１１２】

抑揚成分を０にすることで抑揚の全くない、平坦なロボット音声のようになってしまうが、最高速度での合成音出力は通常、読み飛ばしという意味合いで利用される場合がほとんどある。したがって、音声出力されるテキストの内容を把握・理解できれば良い、という程度の使用方法なので、抑揚のない合成音声は使用に耐え得るものである。

【０１１３】

第３の実施の形態

［構成］

発明の第３の実施の形態における構成を図面を参照しながら詳細に説明する。

本実施の形態が従来技術と異なる点は、文章間に合図音を入れることで文と文との境界を明示する点である。

【０１１４】

図１０は、第３の実施の形態におけるパラメータ生成部１０２の機能ブロック図であり、この図を用いて説明する。パラメータ生成部１０２への入力は従来と同じく、テキスト解析部１０１から出力される中間言語および、ユーザが個別に指定する韻律制御パラメータである。ユーザからの韻律制御指定には、従来技術あるいは第１、第２の実施の形態にはないパラメータとして、合図音指定入力がある。これは後述する、文章間に挿入する合図音の種類を指定するための入力である。

【０１１５】

中間言語解析部１７０１には一文毎の中間言語が入力され、以降の韻律生成処理で必要となる音韻系列・フレーズ情報・アクセント情報などといった中間言語解析結果が、それぞれピッチパタン決定部１７０２、音韻継続時間決定部１７０３、音韻パワー決定部１７０４、音声素片決定部１７０５、声質係数決定部１７０６に出力される。

【０１１６】

ピッチパタン決定部１７０２には、前述の中間言語解析結果に加えてユーザからの抑揚指定・声の高さ指定・発声速度指定・話者指定の各パラメータが入力され、ピッチパタンが合成パラメータ生成部１７０８に出力される。

【０１１７】

音韻継続時間決定部１７０３には、前述の中間言語解析結果に加えてユーザからの発声速度指定のパラメータが入力され、それぞれの音韻継続時間・ポーズ長といったデータが合成パラメータ生成部１７０８に出力される。

【０１１８】
音韻パワー決定部１７０４には、前述の中間言語解析結果に加えてユーザからの声の大きさ指定パラメータが入力され、それぞれの音韻振幅係数が合成パラメータ生成部１７０８に出力される。
【０１１９】
音声素片決定部１７０５には、前述の中間言語解析結果に加えてユーザからの話者指定パラメータが入力され、波形重畳するための必要な音声素片アドレスが合成パラメータ生成部１７０８に出力される。
【０１２０】
声質係数決定部１７０６には、前述の中間言語解析結果に加えてユーザからの声質指定パラメータが入力され、声質変換パラメータが合成パラメータ生成部１７０８に出力される。
【０１２１】
合図音決定部１７０７には、ユーザからの発声速度指定・合図音指定パラメータが入力され、合図音の種類および制御用のための合図音制御信号が波形生成部１０３に出力される。
【０１２２】
合成パラメータ生成部１７０８は、入力された各韻律パラメータ（前述したピッチパタン、音韻継続時間、ポーズ長、音韻振幅係数、音声素片アドレス、声質変換係数）から、フレーム（通常８ｍｓ程度の長さ）を一つの単位とした波形生成用のパラメータに変換し、波形生成部１０３に出力する。
【０１２３】
パラメータ生成部１０２において、従来技術と比較して異なる点は、合図音決定部１７０７が新たな機能ブロックとして存在していることと、その入力パラメータとしてユーザから合図音指定がある点および、波形生成部１０３の内部構成である。テキスト解析部１０１においては、従来と同様であるため、その構成に関する説明は省略する。
【０１２４】
はじめに合図音決定部１７０７の構成について図１１を用いて説明する。図に示すように、合図音決定部１７０７は単にスイッチの役割を果たす機能ブロックである。ユーザから指定される発声速度レベルはスイッチ１８０１の制御用端子に接続され、同じくユーザから指定される合図音コードがスイッチ１８０１のａ端子に接続される。スイッチ１８０１のｂ端子は常にグランドに接続されている。スイッチ１８０１は、発声速度レベルによって、端子ａ、端子ｂのいずかに接続されるように構成されている。発声速度が最高レベルの場合はスイッチ１８０１をａ端子に接続し、それ以外の場合はスイッチ１８０１をｂ端子に接続する。すなわちスイッチ１８０１は、発声速度が最高レベルの時には合図音コードを、それ以外の時には0を出力する構成となっている。スイッチ１８０１の出力は、合図音制御信号として波形生成部１０３に出力される。
【０１２５】
次に波形生成部１０３の構成について図１２を用いて説明する。第３の実施の形態においては、波形生成部１０３は、素片復号部１９０１と振幅制御部１９０２と素片加工部１９０３と重畳制御部１９０４と合図音制御部１９０５とＤＡリングバッファ１９０６の各機能ブロック、および合図音辞書１９０７とから構成されている。
【０１２６】
前述したパラメータ生成部１０２からの出力は、合成パラメータとして素片復号部１９０１に入力される。素片復号部１９０１には素片辞書１０５が接続されており、入力された合成パラメータのうち、素片アドレスを参照ポインタとして素片辞書１０５から素片データをロードし、必要に応じて復号処理を行い、復号素片データを振幅制御部１９０２に出力する。素片辞書１０５には、音声を合成するための元となる音声素片データが格納されており、記憶容量の節約のために何らかの圧縮処理が施されている場合がある。この時は復号処理を施し、その必要がない非圧縮素片の場合は、単に読み込んでくるだけの処理と

なる。

【０１２７】

振幅制御部１９０２には、前述の復号後の音声素片データと合成パラメータとが入力されており、合成パラメータのうち音韻振幅係数によって素片データのパワー制御が行われ、素片加工部１９０３に出力される。

【０１２８】

素片加工部１９０３には、前述の振幅制御された素片データと合成パラメータとが入力されており、合成パラメータのうち声質変換係数によって素片データの伸縮処理が施され、重畳制御部１９０４に出力される。

【０１２９】

重畳制御部１９０４には、前述の伸縮処理が施された素片データと合成パラメータとが入力されており、合成パラメータのうちピッチパタン、音韻継続時間、ポーズ長といったパラメータを用いて素片データの波形重畳処理を施す。重畳制御部１９０４で生成される波形は、逐次ＤＡリングバッファ１９０６に出力され書き込まれる。ＤＡリングバッファ１９０６に書き込まれたデータは、当該テキスト音声変換システムで設定されている出力サンプリング周期で、図示していないＤＡコンバータに送られ、合成音がスピーカなどから出力される。

【０１３０】

波形生成部１０３には、前述したパラメータ生成部１０２からの出力として合図音制御信号が合図音制御部１９０５に入力される。合図音制御部１９０５にはさらに合図音辞書１９０７が接続されており、これに格納されているデータを必要に応じて加工してＤＡリングバッファ１９０６に出力する。ただし書き込むタイミングは、重畳制御部１９０４が１文章分の合成波形を出力し終えた後あるいは、合成波形を書き込む前とする。

【０１３１】

合図音辞書１９０７には例えば、各種効果音データのＰＣＭ（Ｐｕｌｓｅ　Ｃｏｄｅ　Ｍｏｄｕｌａｔｉｏｎ）データで構築されている構成でも、基準正弦波データが格納された構成でも、どの形態でも構わない。この場合、合図音制御部１９０５は、前者の辞書構成においては合図音辞書１９０７からデータを読み出してきて、そのままＤＡリングバッファ１９０６に出力し、後者の辞書構成においては合図音辞書１９０７からデータを読み出し、それを繰り返しつなぎ合わせるなどして出力する。合図音制御部１９０５に接続されている合図音制御信号が０の場合は、ＤＡリングバッファ１９０６に出力する処理は行わない。

【０１３２】

［動作］

以上のように構成された第３の実施の形態における動作について図１０～図１２を用いて詳細に説明する。従来技術と異なる点は、ピッチパタン生成と波形生成に関わる処理であるので、それ以外の処理については省略する。

【０１３３】

まず、テキスト解析部１０１で生成された中間言語は、パラメータ生成部１０２内部の中間言語解析部１７０１に送られる。中間言語解析部１７０１では、中間言語上に記述されているフレーズ区切り記号、単語区切り記号、アクセント核を示すアクセント記号、そして音韻記号列から、韻律生成に必要なデータを抽出して、ピッチパタン決定部１７０２、音韻継続時間決定部１７０３、音韻パワー決定部１７０４、音声素片決定部１７０５、声質係数決定部１７０６のそれぞれの機能ブロックへ送る。

【０１３４】

ピッチパタン決定部１７０２では、声の高さの遷移であるイントネーションが生成され、音韻継続時間決定１７０３では、音韻個々の継続時間のほか、フレーズとフレーズの切れ目あるいは、文と文との切れ目に挿入するポーズ長を決定する。また、音韻パワー決定部１７０４では、音声波形の振幅値の遷移である音韻パワーが生成され、音声素片決定部１７０５では合成波形を生成するために必要となる音声素片の、素片辞書１０５におけるア

ドレスを決定する。声質係数決定部１７０６では、素片データを信号処理で加工するためのパラメータの決定が行われる。ユーザから指定される韻律制御指定のうち、抑揚指定および声の高さ指定はピッチパタン決定部１７０２に、発声速度指定は音韻継続時間決定部１７０３と合図音決定部１７０７に、声の大きさ指定は音韻パワー決定部１７０４に、話者指定はピッチパタン決定部１７０２と音声素片決定部１７０５に、声質指定は声質係数決定部１７０６に、合図音指定は合図音決定部１７０７に、それぞれ送られている。

【０１３５】

各機能ブロックのうち、ピッチパタン決定部１７０２、音韻継続時間決定部１７０３、音韻パワー決定部１７０４、音声素片決定部１７０５、声質係数決定部１７０６については、従来技術と同様であるのでここでは説明を省略する。

【０１３６】

第３の実施の形態におけるパラメータ生成部１０２が従来技術と異なる点は、合図音決定部１７０７が新たに加えられたことであるので、合図音決定部１７０７の動作について図１１を用いて説明する。図に示すように、合図音決定部１７０７は単にスイッチの役割を果たす機能ブロックである。スイッチ１８０１は、ユーザから指定される発声速度レベルによって制御されるような構成を有しており、これにより端子ａ、端子ｂのいずれかに接続されるようになっている。制御信号である発声速度レベルが最高速度の時は、スイッチ１８０１をａ端子に接続し、それ以外の場合はスイッチ１８０１をｂ端子に接続する。ａ端子には、ユーザから指定される合図音コードが入力されており、ｂ端子にはグランド・レベルすなわち０が入力されている。すなわちスイッチ１８０１は、発声速度が最高レベルの時には合図音コードを、それ以外の時には０を出力する構成となっている。スイッチ１８０１の出力は、合図音制御信号として波形生成部１０３に送られる。

【０１３７】

次に波形生成部１０３の動作について図１２を用いて説明する。パラメータ生成部１０２内の合成パラメータ生成部１７０８で生成された合成パラメータは、波形生成部１０３内の素片復号部１９０１と振幅制御部１９０２と素片加工部１９０３と重畳制御部１９０４に送られる。

【０１３８】

素片復号部１９０１では、合成パラメータのうち、素片アドレスを参照ポインタとして素片辞書１０５から素片データをロードし、必要に応じて復号処理を行い、復号素片データを振幅制御部１９０２に送る。素片辞書１０５には合成波形を生成するための元となる音声素片が格納されており、これをピッチパタンで示される周期で重ね合わせていくことにより音声波形を生成するしくみとなっている。

【０１３９】

ここで音声素片とは、接続して合成波形を作るための音声の基本単位で、音の種類等に応じて様々なものが用意されている。一般的に、ＣＶ、ＶＶ、ＶＣＶ、ＣＶＣ（Ｃ：子音、Ｖ：母音）といった音韻連鎖で構成されている場合が多い。上記のように、同じ音韻の素片であっても、前後の音韻環境によって様々な単位で構築されているためデータ容量は膨大となる。そのため通常は、ＡＤＰＣＭ（Ａｄａｐｔｉｖｅ　Ｄｉｆｆｅｒｅｎｔｉａｌ　ＰＣＭ）符号化や、周波数パラメータと駆動音源データの対で構成するといった、圧縮技術を施す場合が多い。無論、圧縮を行わずＰＣＭデータとして構築されている場合もある。素片復号部１９０１によって復元された音声素片データは、振幅制御部１９０２に送られパワー制御が施される。

【０１４０】

振幅制御部１９０２には、合成パラメータのうち振幅係数が入力されており、先の音声素片データに乗じられて振幅制御が施される。振幅係数は、ユーザから指定される声の大きさレベル、音韻の種類、呼気段落内での音節位置、該音韻内での位置（立ち上がり区間・定常区間・立ち下がり区間）など、様々な情報から経験的に決定されている。振幅制御された音声素片は、素片加工部１９０３に送られる。

【０１４１】

素片加工部１９０３では、ユーザから指定された声質変換レベルに応じて素片データの伸縮処理（リサンプリング）が施される。声質変換とは、素片辞書１０５に登録されている素片データに、信号処理等の加工を施すことにより、聴感上、別話者として取り扱えるようにした機能である。一般に、素片データを線形に伸縮する処理を施して実現する場合が多い。伸長処理は、素片データのオーバーサンプリング処理で実現され、太い声となる。逆に縮小処理は、素片データのダウンサンプリング処理で実現され、細い声となる。同一データで別話者を実現するための機能であるため、声質変換処理は上記の手法に限るものではない。また、ユーザからの声質変換指定がない場合は当然のことながら、素片加工部１９０３での処理は一切行われない。

【０１４２】

以上の処理によって生成された音声素片は、重畳制御部１９０４で波形重畳処理が施される。一般的に、ピッチパタンで示されたピッチ周期で素片データをずらしながら重ね合わせて加算するという手法が用いられる。

【０１４３】

このようにして生成された合成波形は、逐次ＤＡリングバッファ１９０６に書き込まれ、当該テキスト音声変換システムで設定されている出力サンプリング周期で、図示していないＤＡコンバータに送られ、合成音がスピーカなどから出力される。

【０１４４】

波形生成部１０３にはさらに、パラメータ生成部１０２内の合図音決定部１７０７から送られる合図音制御信号が入力されている。合図音制御信号は、合図音制御部１９０５を介して合図音辞書１９０７に登録されているデータをＤＡリングバッファ１９０６に書き込むための信号である。合図音制御信号が０の場合、すなわち前述したように、ユーザから指定される発声速度が最高速度レベルではない時は、合図音制御部１９０５は一切の処理を行わない。０以外の場合、すなわち前述したように、ユーザから指定される発声速度が最高速度レベルの時は、合図音制御信号を合図音の種類とみなして合図音辞書１９０７からのデータロードを行う。

【０１４５】

例えば、合図音の種類を３種類設ける。合図音辞書１９０７には、例えば、５００Ｈｚの正弦波データ、１ＫＨｚの正弦波データ、２ＫＨｚの正弦波データがそれぞれ１周期分格納されており、それらを複数回繰り返し接続することにより「ピッ」という合図音を生成することとする。合図音制御信号の取り得る値は、０、１、２、３の４種類となり、０の時は一切の処理を行わず、１の時は合図音辞書１９０７から５００Ｈｚの正弦波データを読み出してきて、それらを既定回繰り返し接続してＤＡリングバッファ１９０６に書き込む。１の時は合図音辞書１９０７から１ＫＨｚの正弦波データを読み出してきて、それらを既定回繰り返し接続してＤＡリングバッファ１９０６に書き込む。２の時は合図音辞書１９０７から２ＫＨｚの正弦波データを読み出してきて、それらを既定回繰り返し接続してＤＡリングバッファ１９０６に書き込む。ただし書き込むタイミングは、重畳制御部１９０４が１文章分の合成波形を出力し終えた後あるいは、合成波形を書き込む前である。したがって、合図音が出力されるのは文章間ということになる。出力される正弦波データは、１００ｍｓ～２００ｍｓ程度が適当と思われる。

【０１４６】

また、正弦波データではなく、出力されるべき合図音を直接ＰＣＭデータとして合図音辞書１９０７に格納しておくという構成でも構わない。この場合、合図音辞書１９０７からデータを読み出してきて、そのままＤＡリングバッファ１９０６に出力する処理が施されることになる。

【０１４７】

以上詳細に説明したように、第３の実施の形態によれば、発声速度が既定値最大に設定された場合に、文章と文章の間に合図音を挿入する機能を有しているため、早聞き機能有効時での従来技術での問題点である、文境界が把握しにくく、読上げテキストの内容理解が困難であるといったことが解消される。

【０１４８】
例えば、以下の文言をテキスト合成する場合を考える。
「出席予定者：開発部　山田部長。企画室　斉藤室長。営業１部　渡辺部長。」処理単位、すなわち１文章の区切り記号は句点「。」とすると、上記の文言は以下の３文章からなる。
（１）「出席予定者：開発部　山田部長。」
（２）「企画室　斉藤室長。」
（３）「営業１部　渡辺部長。」
従来技術によれば、発声速度が速くなるとそれぞれの文終端におけるポーズ長も短くなるため、文章（１）の最後の「山田部長」という合成音声と、文章（２）の先頭の「企画室」という合成音声がほぼ連続して出力されるため、「山田部長」＝「企画室」というような誤った認識を受ける場合も発生する。
【０１４９】
しかしながら、第３の実施の形態によれば、「山田部長」という合成音声と、「企画室」という合成音声の間に、例えば「ピッ」という合図音が挿入されるため、上記のような誤認識は発生しない。
【０１５０】
第４の実施の形態
［構成］
本発明の第４の実施の形態における構成を図１３を参照しながら詳細に説明する。この実施の形態が従来技術と異なる点は、早聞き機能有効時の音韻継続時間の伸縮率決定の際に、現在処理中のテキストが文内における先頭単語あるいは先頭フレーズであるかを判定して、その結果により伸縮係数を決定する点である。したがって、従来と異なる音韻継続時間決定部についてのみ説明し、それ以外の機能ブロックすなわち、テキスト解析部、波形生成部、音韻継続時間決定部以外のパラメータ生成部内部モジュールについては説明を省略する。
【０１５１】
音韻継続時間決定部２０３への入力は従来と同じく、中間言語解析部２０１からの音韻・韻律情報を含んだ解析結果および、ユーザからの指定される発声速度レベルである。１文章に対する中間言語解析結果は制御要因設定部２００１と単語カウンタ２００５とに接続されている。制御要因設定部２００１では、音韻継続時間決定のために必要な制御要因パラメータの解析が行われ、その出力が継続時間推定部２００２に接続される。継続時間の決定には数量化Ｉ類等の統計的手法を用いており、例えば、音韻長は通常、目標となる音韻の前後近傍の音韻の種別あるいは、単語内・呼気段落内の音節位置などにより予測され、ポーズ長は、前後隣接するフレーズのモーラ総数などといった情報から予測が行われる場合が多い。制御要因設定部２００１はこれら予測に必要な情報の抽出を行っている。
【０１５２】
継続時間推定部２００２には、継続時間予測テーブル２００４が接続されており、これを用いて継続時間の予測が行われ、継続時間修正部２００３に出力される。継続時間予測テーブル２００４は、大量の自然発声データを基に数量化Ｉ類などの統計的手法を用いて予め学習されたデータである。
【０１５３】
一方、単語カウンタ２００５では、現在解析中の音韻が、文章内のおける先頭単語あるいは先頭フレーズに含まれているのか、そうでないのかの判定を行い、その結果を伸縮係数決定部２００６に出力する。
【０１５４】
伸縮係数決定部２００６にはさらに、ユーザから指定される発声速度レベルが入力されており、現在処理中の音韻に対する音韻継続時間長の修正係数を決定する機能を有しており、これを継続時間修正部２００３に接続している。
【０１５５】

継続時間修正部２００３では、継続時間推定部２００２で予測された音韻継続時間に対して、伸縮係数決定部２００６で決定された伸縮係数を乗じることにより、音韻継続時間の修正を行い合成パラメータ生成部に出力する。

【０１５６】

［動作］

以上のように構成された本発明の第４の実施の形態における動作について図１３～図１４を用いて詳細に説明する。従来技術と異なる点は、音韻継続時間決定に関わる処理であるので、それ以外の処理については省略する。

【０１５７】

中間言語解析部２０１から１文章に対応する解析結果が制御要因設定部２００１と単語カウンタ２００５に入力される。制御要因設定部２００１では、音韻継続時間（子音長・母音長・閉鎖区間長）、ポーズ長を決定するために必要な制御要因の設定を行う。音韻継続時間の決定に必要なデータとは、例えば、目標となる音韻の種別、対象音節の前後近傍の音韻の種別あるいは、単語内・呼気段落内の音節位置といった情報である。一方、ポーズ長決定に必要なデータとは、前後隣接するフレーズのモーラ総数といった情報である。これらの継続時間長を決定するために継続時間予測テーブル２００４が使用される。

【０１５８】

継続時間予測テーブル２００４は、自然発声データを基に数量化Ｉ類などの統計的手法を用いて予め学習したテーブルである。継続時間推定部２００２は、このテーブルを参照しながら音韻継続時間、ポーズ長の予測を行う。継続時間推定部２００２で算出される個々の音韻継続時間長は、通常発声速度の場合のものである。これらは、継続時間修正部２００３において、ユーザから指定された発声速度に応じて修正が施される構成となっている。通常、発声速度指定は、５～１０段階程度に制御され、それぞれのレベルに対してあらかじめ割り当てられた定数を乗ずることにより行われる。発声速度を遅くしたい場合は音韻継続時間を長くし、発声速度を速くしたい場合は音韻継続時間を短くする。

【０１５９】

一方、単語カウンタ２００５にも、中間言語解析部２０１から１文章に対応する解析結果が入力されており、現在解析中の音韻が、文章内のおける先頭単語あるいは先頭フレーズに含まれているのか、そうでないのかの判定が行われる。本実施の形態では、文章内における先頭単語であるか否かの判定を行う機能として説明を行う。単語カウンタ２００５から送られる判定結果は、該音韻が文内先頭単語に含まれている場合にＴＲＵＥ、そうでない場合にＦＡＬＳＥを出力することとする。単語カウンタ２００５での判定結果は伸縮係数決定部２００６に送られる。

【０１６０】

伸縮係数決定部２００６には前述の単語カウンタ２００５からの判定結果に加えて、ユーザから指定される発声速度レベルが入力されており、これら２つのパラメータから該音韻の伸縮係数の算出を行う。例えば、発声速度レベルが５段階に制御され、発声速度が遅い方からレベル０、レベル１、レベル２、レベル３、レベル４まで指定可能だとする。それぞれのレベルｎに対応した定数T_nを次のように定める。すなわち、

T_0＝２．０、T_1＝１．５、T_2＝１．０、T_3＝０．７５、T_4＝０．５とする。通常発声速度はレベル２となり、早聞き機能が有効とされると発声速度はレベル４に設定されることになる。単語カウンタ２００５からの信号がＴＲＵＥの場合、発声速度レベルが０～３まで範囲であれば上記T_nをそのまま継続時間修正部２００３に出力する。発声速度レベルが４であれば、通常発声時のＴ２の数値を出力する。単語カウンタ２００５からの信号がＦＡＬＳＥの場合は、発声速度レベルに関わらず上記T_nをそのまま継続時間修正部２００３に出力する。

【０１６１】

継続時間修正部２００３では、継続時間推定部２００２から送られる音韻継続時間長に対して、伸縮係数決定部２００６からの伸縮係数を乗じて修正を施す。ただし修正を行うのは通常、母音長のみである。発声速度レベルに応じた修正が施された音韻継続時間は合成

パラメータ生成部へ送られる。

【０１６２】

さらに詳細に説明するために図１４に継続時間決定処理のフローチャートを示す。ここで図中の記号は以下の通りとする。すなわち、入力文章中に含まれる単語総数をＩ、第ｉ番目の単語を構成する音韻に対する継続時間修正係数を$ＴＣ_{i}$、ユーザから指定される発声速度レベルをｌｅｖ（ただし範囲は０～４までの５段階とし、数値が多いほど速度が速いこととする）、発声速度がレベルｎの時の伸縮係数をＴ（ｎ）、第ｉ番目の単語の第ｊ番目の母音長を$Ｔ_{ｉｊ}$、単語を構成する音節数はそれぞれの単語によって変わるがここでは簡単化のために一律Ｊとする。

【０１６３】

まずステップＳＴ２０１で単語数カウンタｉを０に初期化する。次いでステップＳＴ２０２で単語数と発声速度レベルの判定が行われる。現在処理中の単語数カウンタが０でかつ、発声速度レベルが４の時、これはすなわち、現在処理している音節が文内先頭単語に属しており、かつ発声速度が最高レベルの時であるが、この時はステップＳＴ２０４に進み、そうでないときはステップＳＴ２０３に進む。ステップＳＴ２０４では発声速度レベル２の値が修正係数として選択され、ステップＳＴ２０５に進む。すなわち、

$ＴＣ_{i}$　＝　Ｔ（２）　　…（５）

となる。

【０１６４】

ステップＳＴ２０３では、ユーザから指定されたレベル通りの修正係数が選択され、ステップＳＴ２０５に進む。すなわち、

$ＴＣ_{i}$　＝　Ｔ（ｌｅｖ）　　…（６）

となる。

【０１６５】

ステップＳＴ２０５では、音節カウンタｊが０に初期化されステップＳＴ２０６に進む。ステップＳＴ２０６では第ｉ番目の単語の第ｊ番目の母音の継続時間$Ｔ_{ｉｊ}$が、先に求められた修正係数$ＴＣ_{i}$によって下式を用いて行われる。

$Ｔ_{ｉｊ}$　＝　$Ｔ_{ｉｊ}$　×　$ＴＣ_{i}$　　…（７）

【０１６６】

次いでステップＳＴ２０７で音節カウンタｊが１インクリメントされステップＳＴ２０８に進む。ステップＳＴ２０８では、音節カウンタｊと該単語の音節総数Ｊとの比較を行い、音節カウンタｊが音節総数Ｊを超えた場合、すなわち該単語の全ての音節に対する処理が終了した場合にステップＳＴ２０９に進む。そうでない場合は、ステップＳＴ２０６に戻り次の音節に対する処理を前述と同様に繰り返す。

【０１６７】

ステップＳＴ２０９では単語数カウンタｉが１インクリメントされ、次のステップＳＴ２１０に進む。

【０１６８】

ステップＳＴ２１０では、単語数カウンタｉと単語総数Ｉとの比較を行い、単語数カウンタｉが単語総数Ｉを超えた場合、すなわち入力文章中の全て単語に対する処理が終了した場合は処理を終了し、そうでない場合は、ステップＳＴ２０２に戻り次の単語に対する処理を前述と同様に繰り返す。

【０１６９】

上記の処理により、ユーザから指定される発声速度レベルが最高速度となっても、文章先頭単語だけは通常の発声速度での合成音が生成されることになる。

【０１７０】

以上詳細に説明したように、第４の実施の形態によれば、発声速度が既定値最大に設定された場合に、文先頭の単語に対して音韻継続時間制御を通常の発声速度として処理するため、ユーザが早聞き機能解除のタイミングを計りやすいという効果がある。例えば、ソフトウェア仕様書などのマニュアル類には、「第３章」あるいは「４．１．３」などの項目

番号が付与されている場合がほとんどある。こういったマニュアル類をテキスト音声変換で読上げを行う際に、第３章から聞きたい、あるいは４．１．３節から聞きたいといった場合に、従来技術においては、早聞き機能を有効にした後ユーザが、高速で出力される合成音声の中から「ダイサンショー」あるいは「ヨンテンイッテンサン」といったキーワードを聞き分け、早聞き機能を解除するといった面倒な操作が必要であった。第４の実施の形態によれば、ユーザに負担をかけずに早聞き機能の有効化・無効化を実現することが可能となる。

【０１７１】

尚、本発明は前述の実施の形態に限定されるものではなく、本発明の趣旨に基づいて種々変形させることが可能である。例えば、第１の実施の形態において、発声速度が既定値最大に設定された場合に、テキスト音声変換処理の中で演算負荷が大きい機能ブロックを簡略化あるいは、無効にする処理を施しているが、この処理は最大発声速度に限らない。つまり、ある閾値を設けて、その閾値を超えたときに前述の処理を施す構成でも構わない。また、高負荷処理として数量化Ｉ類による韻律パラメータの予測処理、声質変換のための素片データ加工処理を挙げているが、これに限るものではない。他に高負荷処理機能（例えばエコーや高域強調などの音響処理など）を有している場合は当然のことながら、これを無効化あるいは簡略化といった処理形態にすることが望ましい。また、声質変換処理として波形そのものを線形伸縮しているが、非線形伸縮でも、あるいは周波数パラメータに対して規定の変換関数に通して変形するといった方法でも構わない。また、音韻継続時間決定規則、ピッチパタン決定規則を挙げているが、本発明では演算量が少なく済み、処理時間の短縮が図れるモードを有する構成にすること目的としているため、規則化手順は上記に限られるものではない。逆に、通常発声速度の時には、統計的手法を用いた韻律パラメータの予測を行っているが、規則化手順よりも演算負荷がかかる処理であればこれに限るものではない。また、その予測に用いる制御要因を幾つか挙げているがこれはあくまでも一例である。

【０１７２】

第２の実施の形態において、発声速度が既定値最大に設定された場合に、ピッチパタンの抑揚成分を０にしてピッチパタン生成を行っているが、この処理は最大発声速度に限らない。即ち、ある閾値を設けて、その閾値を超えたときに前述の処理を施す構成でも構わない。また、抑揚成分を完全に０にしているが、通常時に比べて抑揚成分を弱めるといった方法でも構わない。例えば、発声速度が既定値最大に設定された時は、抑揚指定レベルを強制的に最低レベルに設定し、ピッチパタン修正部において抑揚成分を縮小するといった構成でも構わない。ただこの時の抑揚指定レベルは、高速合成時においても聞き易いイントネーションとなる必要がある。また、ピッチパタンのアクセント成分、フレーズ成分を数量化Ｉ類によって決定しているが規則によって決定しても無論構わない。また、予測を行う際にその制御要因を幾つか挙げているがこれはあくまでも一例である。

【０１７３】

第３の実施の形態において、発声速度が既定値最大に設定された場合に、文章と文章の間に合図音を挿入しているが、この処理は最大発声速度に限らない。即ち、ある閾値を設けて、その閾値を超えたときに前述の処理を施す構成でも構わない。また、実施例では基準正弦波の繰り返しにより合図音を生成しているが、ユーザの注意を引けるものであればこれに限らない。録音された効果音をそのまま出力する構成でも構わない。無論、実施例で示したような合図音辞書を持たずに、内部回路あるいはプログラムでその都度生成するような構成でも構わない。またこの実施の形態では１文の合成波形直後に合図音を挿入する構成となっているが、逆に合成波形直前でも構わない。発声速度が既定値最大に設定された時に、ユーザに対して文章境界が明示できればそれでよい。また、この実施の形態ではパラメータ生成部に合図音の種類を指定するための入力が存在するが、ハードウェア規模、ソフトウェア規模の制限などから、これを省略してもよい。しかしながら、ユーザの好みによって合図音を変えることのできる構成の方が好ましい。

【０１７４】

第４の実施の形態において、発声速度が既定値最大に設定された場合に、文先頭の単語に対して音韻継続時間制御を通常（デフォルト）の発声速度として処理しているが、この処理は最大発声速度に限らない。即ち、ある閾値を設けて、その閾値を超えたときに前述の処理を施す構成でも構わない。また、通常発声速度で処理する単位を文先頭の１単語としているが、先頭２単語あるいは先頭フレーズという構成でも構わない。また、通常の発声速度ではなく、レベルを１段階落とすといった方法も十分考えられる。

【０１７５】

【発明の効果】

以上詳細に説明したように、請求項１に係る発明によれば、入力されたテキストから音韻・韻律記号列を生成するテキスト解析手段と、前記音韻・韻律記号列に対して少なくとも音声素片・音韻継続時間・基本周波数の合成パラメータを生成するパラメータ生成手段と、音声の基本単位となる音声素片が登録された素片辞書と前記パラメータ生成手段から生成される合成パラメータに基づいて前記素片辞書を参照しながら波形重畳を行って合成波形を生成する波形生成手段とを備えたテキスト音声変換装置における高速読み上げ制御方法であって、前記パラメータ生成手段は、音韻継続時間を予め経験的に求めた継続時間規則テーブルと、音韻継続時間を統計的手法を用いて予測した継続時間予測テーブルとを併せ持ち、ユーザから指定される発声速度が閾値を超えた時には前記継続時間規則テーブルを用い、閾値を超えていない時には前記継続時間予測テーブルを用いて音韻継続時間の決定を行う音韻継続時間決定手段を有する構成としたことにより、また、請求項３に係る発明によれば、前記パラメータ生成手段は、アクセント成分及びフレーズ成分を決定するために必要となるデータを、予め経験的に求めた規則テーブルと、統計的手法を用いて予測した予測テーブルとを併せ持ち、ユーザから指定される発声速度が閾値を超えた時には前記規則テーブルを用い、閾値を超えていない時には前記予測テーブルを用いてアクセント成分及びフレーズ成分を決定することによりピッチパタンを決定するピッチパタン決定手段を有する構成としたことにより、更に、請求項５に係る発明によれば、前記パラメータ生成手段は、前記音声素片を変形させて声質を切り換えるための声質変換係数テーブルを備え、ユーザから指定される発声速度が閾値を超えたときには、声質が変化しないような係数を前記声質変換係数テーブルから選択する声質係数決定手段を有する構成としたので、発声速度が既定値最大に設定された場合に、テキスト音声変換処理の中で演算負荷が大きい機能ブロックを簡略化あるいは、無効にする処理を施しているため、高負荷による音切れが発生する機会を減少させ、聞き易い合成音声を生成することが可能となる。

【０１７６】

また、請求項７に係る発明によれば、前記パラメータ生成手段は、ユーザが指定した抑揚レベルに応じて修正したピッチパタンを出力するするピッチパタン修正手段と、ユーザが指定した発声速度に応じて前記修正したピッチパタンを基底ピッチに加算するか否かを選択する切り換え手段とを有し、前記発声速度が所定の閾値を超えた場合には前記基底ピッチを変更しないように前記切り換え手段を制御する構成としたので、発声速度が既定値最大に設定された場合に、ピッチパタンの抑揚成分を０にしてピッチパタン生成を行うため、時間的に速い周期で抑揚が変動することがなくなり、非常に聞き取りにくい合成音となることが解消される。

【図面の簡単な説明】

【図１】本発明の第１の実施の形態におけるパラメータ生成部の機能ブロック図である。

【図２】本発明の第１の実施の形態におけるピッチパタン決定部の機能ブロック図である。

【図３】本発明の第１の実施の形態における音韻継続時間決定部の機能ブロック図である。

【図４】本発明の第１の実施の形態における声質係数決定部の機能ブロック図である。

【図５】声質変換のためのデータのリサンプリング周期の説明図である。

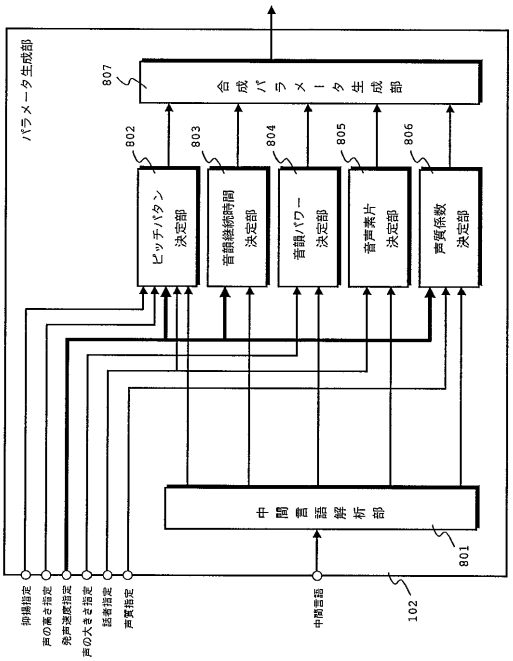
【図６】本発明の第２の実施の形態におけるパラメータ生成部の機能ブロック図である。

【図７】本発明の第２の実施の形態におけるピッチパタン決定部の機能ブロック図である

。
【図８】本発明の第２の実施の形態におけるピッチパタン生成フローチャートである。
【図９】発声速度によるピッチパタンの違いの説明図である。
【図１０】本発明の第３の実施の形態におけるパラメータ生成部の機能ブロック図である。
【図１１】本発明の第３の実施の形態における合図音決定部の機能ブロック図である。
【図１２】本発明の第３の実施の形態における波形生成部の機能ブロック図である。
【図１３】本発明の第４の実施の形態における音韻継続時間決定部の機能ブロック図である。
【図１４】本発明の第４の実施の形態における継続時間決定フローチャートである。
【図１５】一般的なテキスト音声変換処理の機能ブロック図である。
【図１６】従来技術によるパラメータ生成部の機能ブロック図である。
【図１７】従来技術による波形生成部の機能ブロック図である。
【図１８】ピッチパタン生成過程モデルの説明図である。
【図１９】従来技術によるピッチパタン決定部の機能ブロック図である。
【図２０】従来技術による音韻継続時間決定部の機能ブロック図である。
【図２１】発声速度の違いによる波形伸縮の説明図である。
【符号の説明】

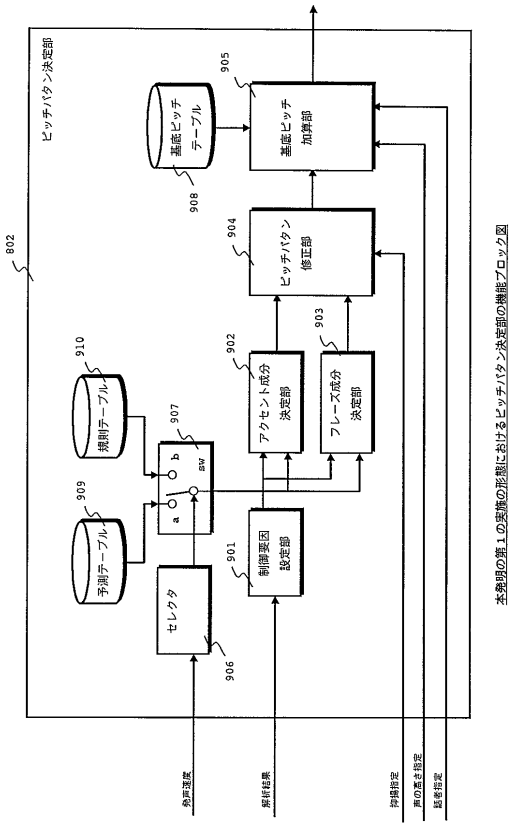
１０１	テキスト解析部
１０２	パラメータ生成部
１０３	波形生成部
１０４	単語辞書
１０５	素片辞書
８０１，１３０１，１７０１，	中間言語解析部
８０２，１３０２，１７０２，	ピッチパタン決定部
８０３，１３０３，１７０３	音韻継続時間決定部
８０４，１３０４，１７０４	音韻パワー決定部
８０５，１３０５，１７０５	音声素片決定部
８０６，１３０６，１７０６	声質係数決定部
１７０７	合図音決定部
８０７，１３０７，１７０８	合成パラメータ生成部

【図１】



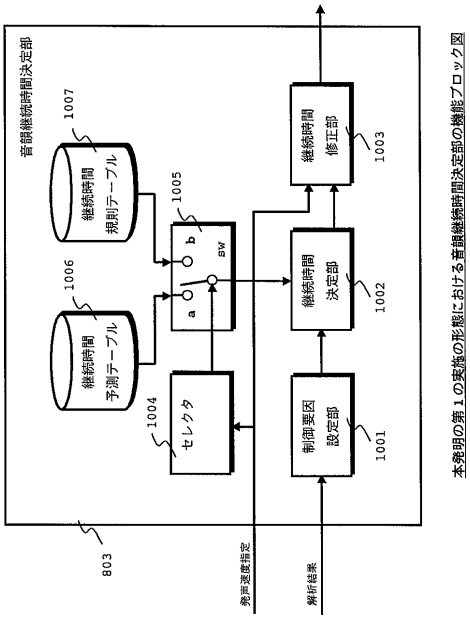
本発明の第１の実施の形態におけるパラメータ生成部の機能ブロック図

【図２】



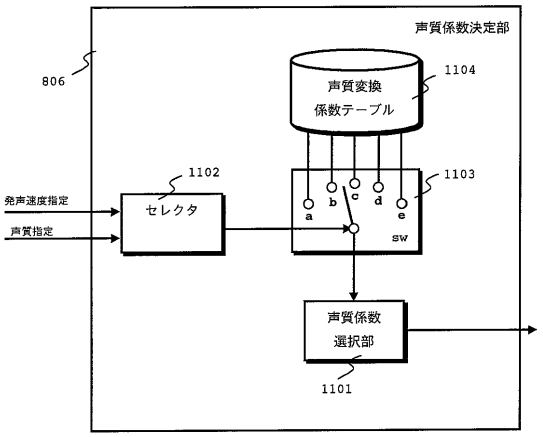
本発明の第１の実施の形態におけるピッチパタン決定部の機能ブロック図

【図３】



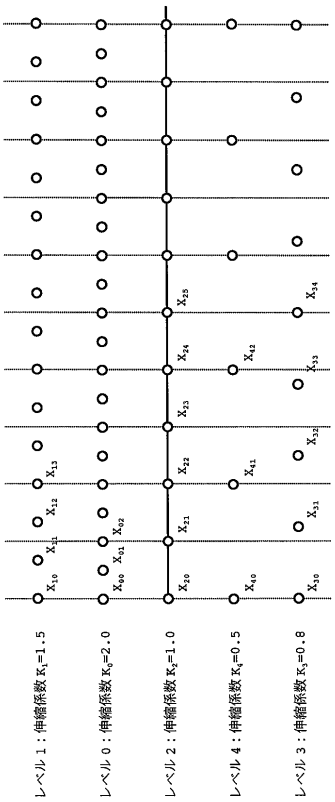
本発明の第１の実施の形態における音韻継続時間決定部の機能ブロック図

【図４】



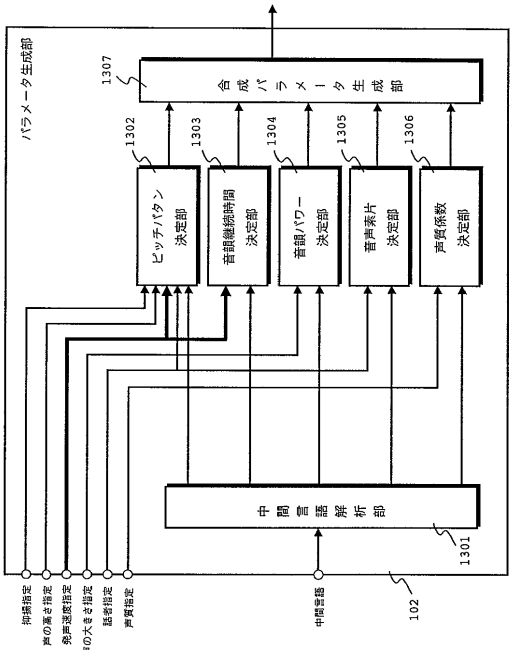
本発明の第１の実施の形態における声質係数決定部の機能ブロック図

【図５】



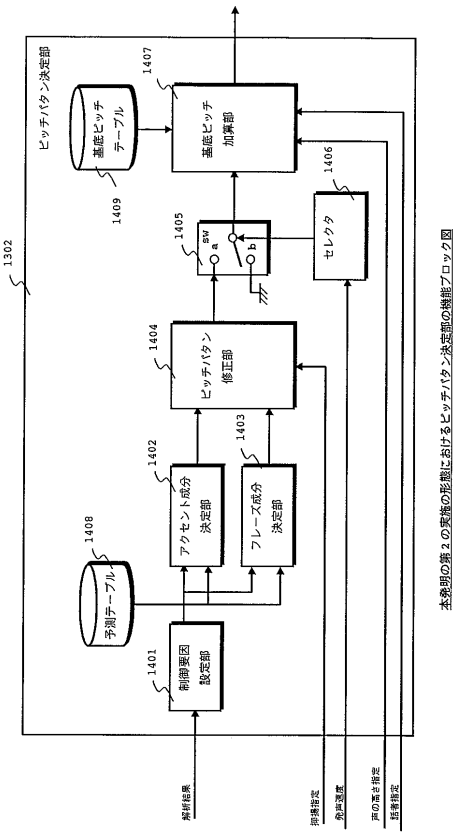
声質変換のためのデータのリサンプリング周期の説明図

【図６】



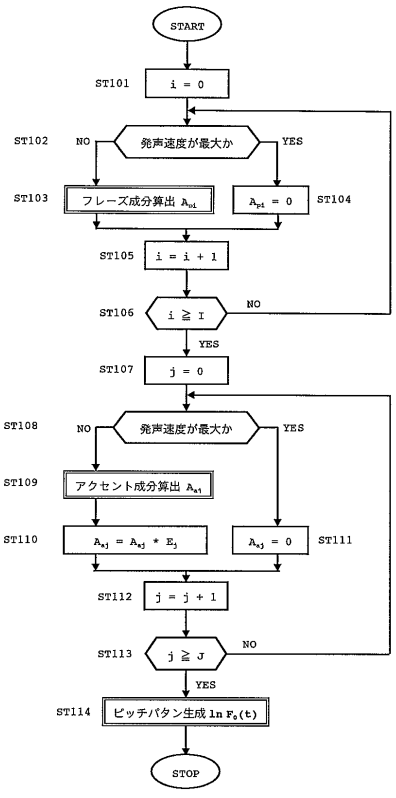
本発明の第２の実施の形態におけるパラメータ生成部の機能ブロック図

【図７】



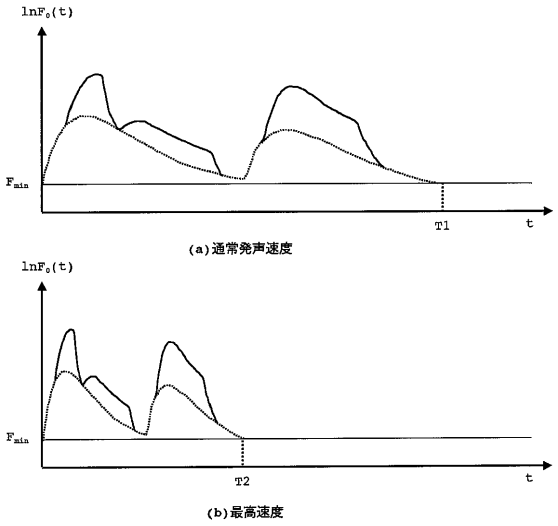
本発明の第２の実施の形態におけるピッチパタン決定部の機能ブロック図

【図８】



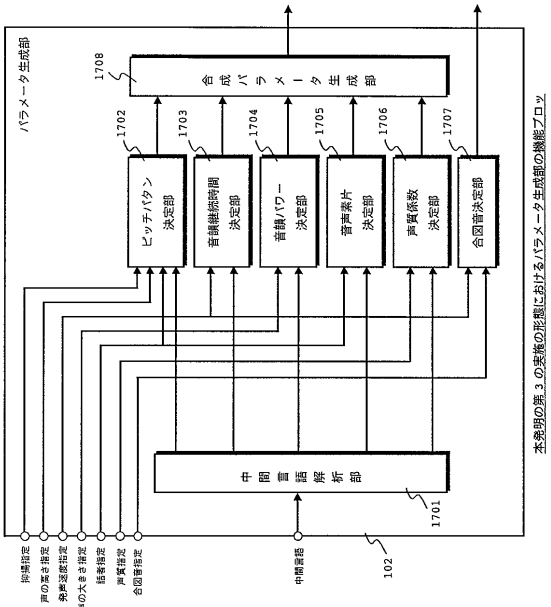
本発明の第２の実施の形態におけるピッチパタン生成フローチャ

【図９】

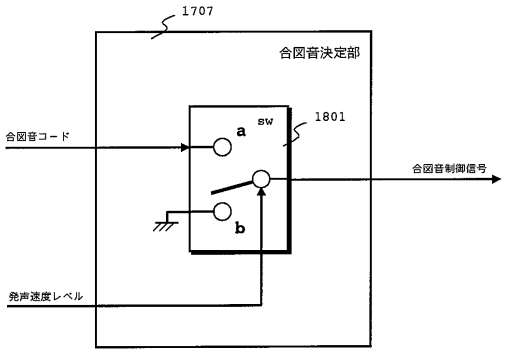


発声速度によるピッチパタンの違いの説明図

【図１０】

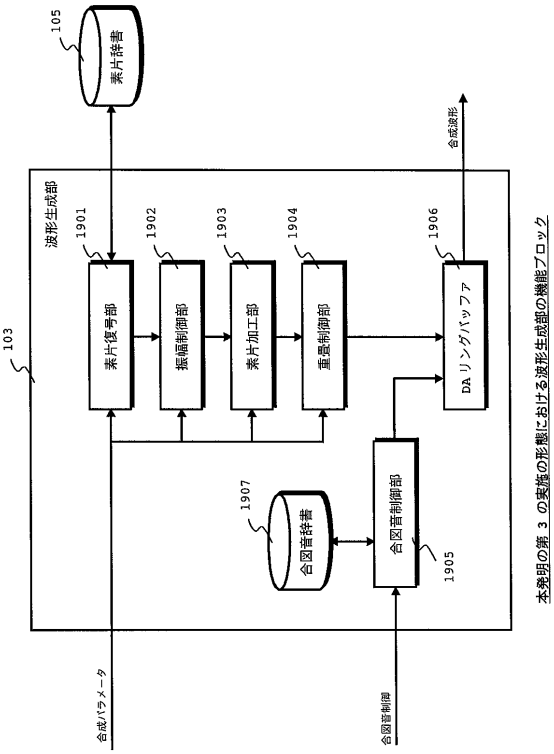


【図１１】

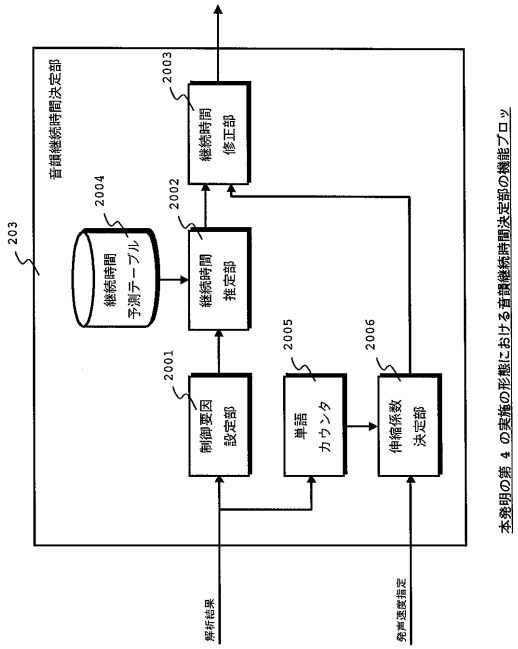


本発明の第３の実施の形態における合図音決定部の機能ブロック図

【図１２】

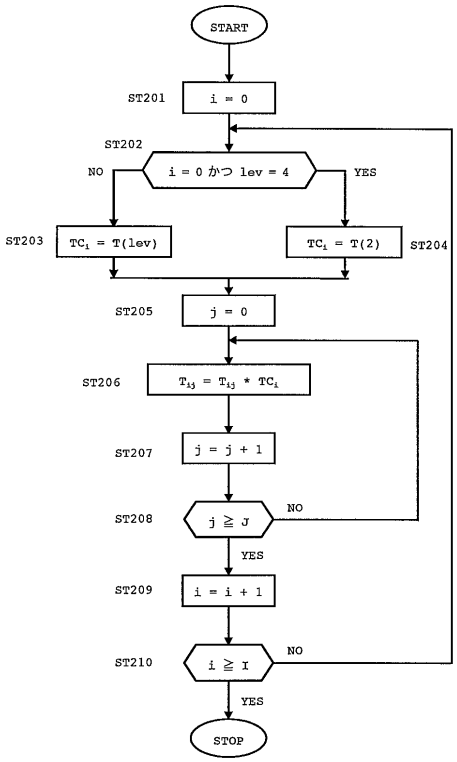


【図１３】



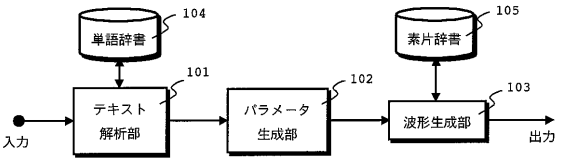
本発明の第４の実施の形態における音韻継続時間決定部の機能ブロック

【図１４】



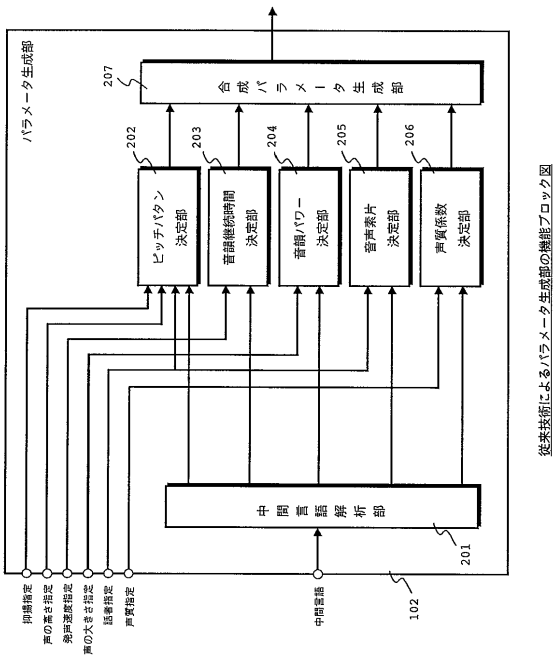
本発明の第４の実施の形態における継続時間決定フローチャート

【図１５】



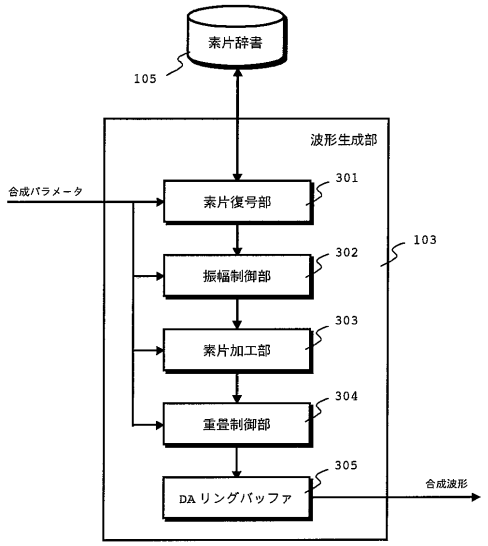
代表的なテキスト音声変換処理の機能ブロック図

【図１６】



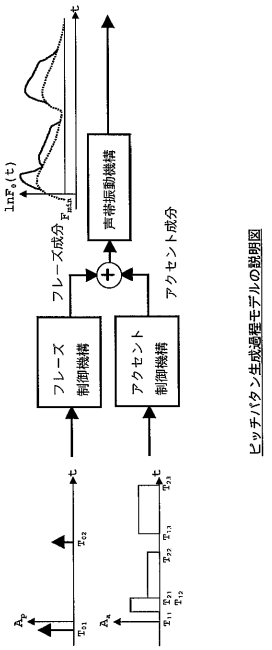
従来技術によるパラメータ生成部の機能ブロック図

【図１７】



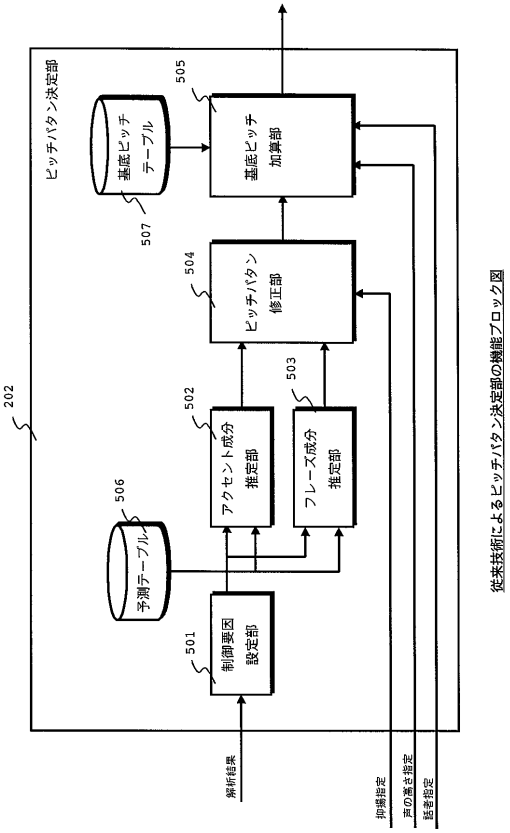
従来技術による波形生成部の機能ブロック図

【図１８】



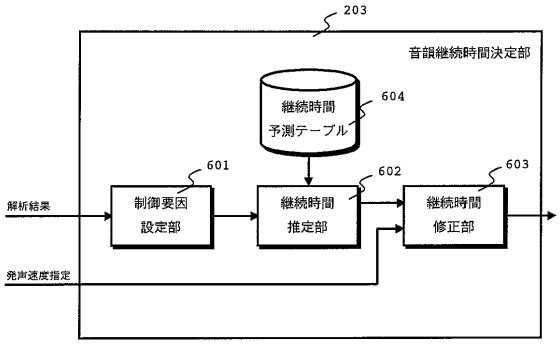
ピッチパタン生成過程モデルの説明図

【図１９】



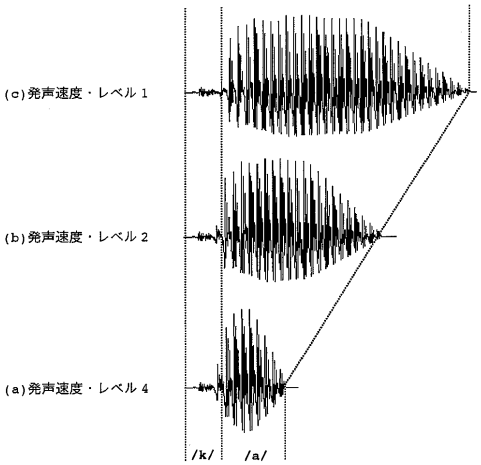
従来技術によるピッチパタン決定部の機能ブロック図

【図２０】



従来技術による音韻継続時間決定部の機能ブロック図

【図２１】



発声速度の違いによる波形伸縮の説明図

フロントページの続き

(56)参考文献 特開平１１－１６７３９８（ＪＰ，Ａ）
特開平１１－０７３２９８（ＪＰ，Ａ）
特開平０２－１９５３９７（ＪＰ，Ａ）
特開平０６－１４９２８４（ＪＰ，Ａ）
特開２０００－３０５５８２（ＪＰ，Ａ）
特開平０９－１７９５７７（ＪＰ，Ａ）
特開平０８－３３５０９６（ＪＰ，Ａ）
実開昭５９－１６０３４８（ＪＰ，Ｕ）
特開２０００－３０５５８５（ＪＰ，Ａ）

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