

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2006-349787
(P2006-349787A)

(43) 公開日 平成18年12月28日(2006.12.28)

(51) Int. Cl.
G10L 13/08 (2006.01)

F I
G10L 13/08 110Z

テーマコード (参考)

審査請求 未請求 請求項の数 10 O L (全 11 頁)

(21) 出願番号	特願2005-173241 (P2005-173241)	(71) 出願人	000153443 株式会社日立情報制御ソリューションズ 茨城県日立市大みか町5丁目2番1号
(22) 出願日	平成17年6月14日(2005.6.14)	(74) 代理人	100093872 弁理士 高崎 芳紘
		(72) 発明者	中村 兼一 茨城県日立市幸町3丁目2番1号 日立エンジニアリング株式会社内
		(72) 発明者	大和田 潔 茨城県日立市幸町3丁目2番1号 日立エンジニアリング株式会社内

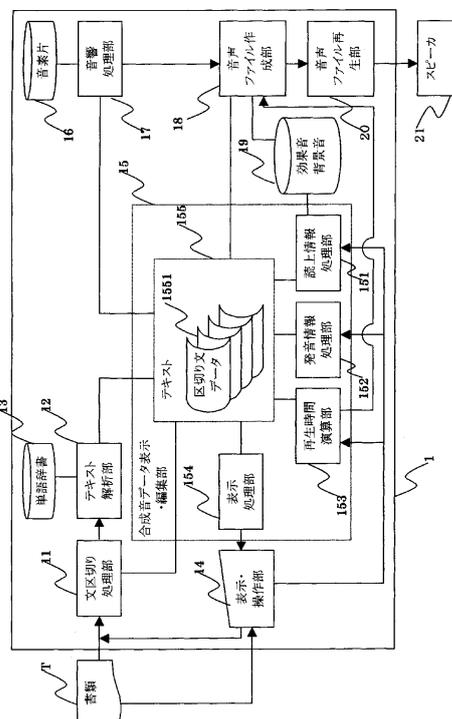
(54) 【発明の名称】 音声合成方法および装置

(57) 【要約】

【課題】 音声合成について、合成音データの編集に際し、実際の出力合成音についてのイメージをより容易につかむことができるようにする。

【解決手段】 入力されたテキスト文から合成音を生成するについて、テキスト文に設定した発音記号や読上げ形式を編集する処理を含んでいる音声合成方法について、テキスト文の文章を文区切り処理部11により所定の文区切り基準で区切ることで有限の長さの区切り文(区切り文データ1551)の集合とする文区切り処理を含み、編集処理は、文区切り処理で得られる区切り文を単位として行うようにしている。

【選択図】 図1



【特許請求の範囲】**【請求項 1】**

入力されたテキスト文から合成音を生成するについて、前記テキスト文に設定した発音記号や読上げ形式を編集する処理を含んでいる音声合成方法において、

前記テキスト文の文章を所定の文区切り基準で区切ることにより有限の長さの区切り文の集合とする文区切り処理を含み、前記編集処理は、前記文区切り処理で得られる区切り文を単位として行うようにされていることを特徴とする音声合成方法。

【請求項 2】

前記文区切り基準として、句点、疑問符、感嘆符などの終端文字を用いるようにした請求項 1 に記載の音声合成方法。

10

【請求項 3】

読上げ形式情報編集画面を表示して読上げ形式情報の編集を行う処理と発音情報編集画面を表示して発音情報の編集を行う処理を含んでいる請求項 1 または請求項 2 に記載の音声合成方法。

【請求項 4】

入力されたテキスト文から合成音を生成する音声合成装置において、

前記テキスト文の文章を所定の文区切り基準で区切ることにより有限の長さの区切り文の集合とする文区切り処理部を備えていることを特徴とする音声合成装置。

【請求項 5】

前記文区切り基準として、句点、疑問符、感嘆符などの終端文字を用いるようにされている請求項 4 に記載の音声合成装置。

20

【請求項 6】

前記テキスト文に発音情報や読上げ形式情報を設定して生成される中間合成音データについて前記読上げ形式情報を前記区切り文単位で編集するための読上情報処理部を備えている請求項 4 または請求項 5 に記載の音声合成装置。

【請求項 7】

前記中間合成音データについて前記発音情報を前記区切り文単位で編集するための発音情報処理部をさらに備えている請求項 6 に記載の音声合成装置。

【請求項 8】

入力されたテキスト文から合成音を生成する音声合成装置において、

前記テキスト文に発音情報や読上げ形式情報を設定して生成される中間合成音データを編集するための合成音データ表示・編集部を備え、前記合成音データ表示・編集部は、生成した合成音を再生して出力した際の再生に関する時間を求め、それを前記編集のための画面に表示する時間演算部を含んでいることを特徴とする音声合成装置。

30

【請求項 9】

前記テキスト文の文章を所定の文区切り基準で区切ることにより有限の長さの区切り文の集合とする文区切り処理部を備え、前記時間演算部は、前記区切り文ごとの合成音再生に要する区切り文再生時間と、前記テキスト文についての合成音再生の開始時点から前記区切り文の再生が開始されるまでの区切り文再生開始時間を求めるようにされている請求項 8 に記載の音声合成装置。

40

【請求項 10】

入力されたテキスト文から合成音を生成する音声合成装置において、

前記合成音にのせる効果音や背景音を格納する効果音・背景音格納部を備え、前記合成音に効果音・背景音をのせる編集処理を行う読上情報処理部を備えていることを特徴とする音声合成装置。

【発明の詳細な説明】**【技術分野】****【0001】**

本発明は、与えられたテキスト文から合成音を生成する音声合成技術に関する。

【背景技術】

50

【0002】

音声合成処理においては、与えられた漢字カナ混じりのテキスト文について構文解析を行った後、発音記号や韻律記号を含むカナ文字列を生成させ、さらに所望の読上げ形式を設定し、これらの合成音データに基づいて合成音を生成する。より具体的には、発音や韻律それに読上げ形式について標準設定（デフォルト）が設けられており、この標準設定によりまず上記カナ文字列を生成させるとともに読上げ形式を設定し、それからそれらを編集画面上に表示して標準設定の合成音データをマニュアル操作で変更する編集を行うという作業が音声合成処理において行われる。

【0003】

こうした音声合成処理における編集作業は、実際に出力される合成音をイメージしながら行われる。そのため実際の出力合成音のイメージを編集画面上でつかみ易いか否かが作業効率に大きく影響する。すなわち出力イメージがつかみ難いと、実際に出力させた合成音を聞いてから編集をし直すという作業を何度も繰り返すことになり、作業効率が悪くなる。

10

【0004】

このようなことから、合成音データの編集については様々な工夫がなされている。例えば特許文献1に開示の音声合成装置では、合成音データにおける各文字自体の状態表示と他の文字との相対位置関係により、その文に対する韻律を表した文字列を表示することが示されている。具体的には、音質を表示色、声の高さを文字の表示位置の上下関係、発声の速さを各文字間隔、声の大きさを文字の輝度、アクセントの大きさを文字の上方に位置するアクセント記号により、というようにして韻律などを表した文字列を編集画面に表示し、合成音の出力イメージを視覚的に捉えることができるようにしている。

20

【0005】

【特許文献1】特開平8-77152号公報

【発明の開示】

【発明が解決しようとする課題】

【0006】

合成音の出力イメージを視覚的に捉えることができるようにする特許文献1に開示の手法は、それなりの有効性を期待できる。しかし、特許文献1に開示の手法では韻律などの表示が複雑となる。そのため、音声合成の対象とするテキスト文がある程度以上の長文になると、韻律などの表示が複雑であることに起因して、逆に出力イメージをつかみ難くなる可能性が高い。すなわち編集作業における出力合成音のイメージ化の難易は、処理するテキスト文の長さによる影響をより多く受けると考えられ、このことに配慮しないと出力合成音のイメージ化の容易化には限界があるということである。

30

【0007】

以上は音声合成処理における合成音データ編集時の出力合成音のイメージ化の問題であるが、音声合成処理には、合成音によるテキスト文の読上げ速度やポーズ長の編集についての問題もある。すなわち合成音データの編集においては、合成音による読上げ速度なども編集対象となるが、与えられたテキスト文の読上げ時間に制限がある場合に、この制限時間との関係などで最適な読上げ速度などを設定するのが必ずしも容易でないという問題である。

40

【0008】

さらに音声合成処理には、それで得られる合成音の使用目的に応じて、より効果的な合成音とすることが求められる場合があり、そのような要望に応えることができるようすることも1つの課題となっている。

【0009】

本発明は、以上のような知見に基づいてなされたものであり、音声合成について、合成音データの編集に際し、実際の出力合成音についてのイメージをより容易につかむことを可能とし、それにより音声合成処理の効率を高めることができるようにすることを第1の目的とし、また音声合成について、与えられたテキスト文の読上げに時間制限などがある

50

場合の合成音データの編集をより容易に行えるようにすることを第2の目的とし、さらに音声合成について、使用目的に応じてより効果的となる合成音の生成を可能とすることを第3の目的としている。

【課題を解決するための手段】

【0010】

上記第1の目的のために本発明では、入力されたテキスト文から合成音を生成するについて、前記テキスト文に設定した発音記号や読上げ形式を編集する処理を含んでいる音声合成方法において、前記テキスト文の文章を所定の文区切り基準で区切ることにより有限の長さの区切り文の集合とする文区切り処理を含み、前記編集処理は、前記文区切り処理で得られる区切り文を単位として行うようにされていることを特徴としている。

10

【0011】

また本発明では、上記のような音声合成方法について、前記文区切り基準として、句点、疑問符、感嘆符などの終端文字を用いるようにしている。

【0012】

また本発明では、上記のような音声合成方法について、読上げ形式情報編集画面を表示して読上げ形式情報の編集を行う処理と発音情報編集画面を表示して発音情報の編集を行う処理を含ませるものとしている。

【0013】

また本発明では上記第1の目的のために、入力されたテキスト文から合成音を生成する音声合成装置において、前記テキスト文の文章を所定の文区切り基準で区切ることにより有限の長さの区切り文の集合とする文区切り処理部を備えていることを特徴としている。

20

【0014】

また本発明では、上記のような音声合成装置について、前記文区切り基準として、句点、疑問符、感嘆符などの終端文字を用いるようにしている。

【0015】

また本発明では、上記のような音声合成装置について、前記テキスト文に発音情報や読上げ形式情報を設定して生成される中間合成音データについて前記読上げ形式情報を前記区切り文単位で編集するための読上情報処理部を備えるものとしている。

【0016】

また本発明では、上記のような音声合成装置について、前記中間合成音データについて前記発音情報を前記区切り文単位で編集するための発音情報処理部をさらに備えるものとしている。

30

【0017】

また本発明では上記第2の目的のために、入力されたテキスト文から合成音を生成する音声合成装置において、前記テキスト文に発音情報や読上げ形式情報を設定して生成される中間合成音データを編集するための合成音データ表示・編集部を備え、前記合成音データ表示・編集部は、生成した合成音を再生して出力した際の再生に関する時間を求め、それを前記編集のための画面に表示する時間演算部を含んでいることを特徴としている。

【0018】

また本発明では、上記のような音声合成装置について、前記テキスト文の文章を所定の文区切り基準で区切ることにより有限の長さの区切り文の集合とする文区切り処理部を設け、前記時間演算部には、前記区切り文ごとの合成音再生に要する区切り文再生時間と、前記テキスト文についての合成音再生の開始時点から前記区切り文の再生が開始されるまでの区切り文再生開始時間を求めさせるようにしている。

40

【0019】

また本発明では上記第3の目的のために、入力されたテキスト文から合成音を生成する音声合成装置において、前記合成音にのせる効果音や背景音を格納する効果音・背景音格納部を備え、前記合成音に効果音・背景音をのせる編集処理を行う読上情報処理部を備えていることを特徴としている。

【発明の効果】

50

【0020】

本発明では、テキスト文を所定の文区切り基準で区切ることにより有限の長さの区切り文の集合とする文区切り処理を行い、発音記号や読上げ形式の編集は、区切り文を単位として行えるようにしている。このため本発明によれば、テキスト文が長文の場合でも、編集に際し、実際の出力合成音についてのイメージをより容易につかむことができるようになり、音声合成処理の効率を高めることが可能となる。

【0021】

また本発明では、生成した合成音を再生して出力した際の再生に関する時間を求め、それを編集のための画面に表示できるようにしている。このため本発明によれば、与えられたテキスト文の読上げに時間制限などがある場合の合成音データの編集をより容易に行えるようになる。

10

【0022】

また本発明では、合成音にのせる効果音や背景音を格納する効果音・背景音格納部を設け、前記合成音に効果音・背景音をのせる編集処理を行う読上情報処理部を設けるようにしている。このため本発明によれば、合成音に効果音や背景音をのせることができるようになり、使用目的に応じてより効果的となる合成音を生成することが可能となる。

【発明を実施するための最良の形態】

【0023】

以下、本発明を実施するための形態について説明する。本発明を実施するには、ハードウェア要素であるパーソナルコンピュータなどのデータ処理機器にソフトウェア要素である音声合成方法に関するコンピュータプログラムを実装して音声合成装置を構成する。図1に、一実施形態による音声合成装置のシステム構成を示す。本実施形態の音声合成装置1は、文区切り処理部11、テキスト解析部12、単語辞書13、表示・操作部14、合成音データ表示・編集部15、音素片格納部16、音響処理部17、音声ファイル作成部18、効果音・背景音格納部19、音声ファイル再生部20、およびスピーカ21を備えている。

20

【0024】

文区切り処理部11は、入力されたテキスト文Tの文章を所定の文区切り基準で区切ることにより有限の長さの区切り文の集合とする処理を行う。文区切り基準には、句点、読点、疑問符、感嘆符などの終端文字を用いるのが通常である。本実施形態では、句点、疑問符および感嘆符を文区切り基準に用いている。文区切り処理部11で複数の区切り文に区切られたテキスト文Tは、テキスト解析部12に渡されるとともに、合成音データ表示・編集部15における後述のテキスト格納部155に区切り文を単位とした区切り文データ1551の集合として格納される。

30

【0025】

テキスト解析部12は、区切り文に区切られたテキスト文Tを入力とし、単語辞書13を用いてその解析を行う。テキスト解析部12による解析は、テキスト文Tの構文解析とそれに基づく発音情報の標準設定である。発音情報の標準設定では、標準設定の合成音データ生成のために発音に関する設定（読み付与、アクセント付与、イントネーション付与、ポーズ設定など）を構文解析に基づいて行って発音記号や韻律記号付きカナ文字列を生成させる。これらの処理は音声合成分野では周知の処理である。標準的に発音記号を設定した結果は、テキスト格納部155に区切り文データ1551として格納される。ここで、テキスト解析部12による発音情報の標準設定に加えて、標準的な読上げ形式を設定することで標準設定による初期設定の合成音データが生成される。ただし、最終的な合成音データとするには発音情報と読上げ形式情報の設定を基に音素データ（音素片）を割り当てる音響変換処理が必要であり、したがって発音情報と読上げ形式情報だけが設定された時点のデータは、中間的なもので、中間合成音データである。なお、標準的な読上げ形式の設定は、後述するように合成音データ表示・編集部15における読上情報処理部151が行う。

40

【0026】

50

単語辞書 13 は、漢字カナ混じりのテキスト文 T における漢字をテキスト解析部 12 でカナに変換するのに用いられる辞書である。

【0027】

表示・操作部 14 は、音声合成処理に必要なデータの表示を行うモニタや音声合成処理に必要なデータの入力出力を行うキーボードやマウスなどの入力出力機器で構成される。

【0028】

合成音データ表示・編集部 15 は、中間合成音データの編集処理つまり中間合成音データにおける発音情報や読上げ形式情報の設定をユーザのマニュアル操作で変更する編集処理のために主に機能する。その編集処理では、標準設定などにより既に生成されている合成音データを表示・操作部 14 におけるモニタに表示し、その状態でユーザが読上げ形式情報や発音情報を変更することになる。こうした合成音データ表示・編集部 15 は、読上情報処理部 151、発音情報処理部 152、時間演算部 153、表示処理部 154、およびテキスト格納部 155 を含んでいる。

10

【0029】

読上情報処理部 151 は、3つの機能を負っている。第1の機能は、読上げ形式情報の設定をユーザのマニュアル操作で変更する編集処理のための機能である。この機能の具体的な例については後述する。

【0030】

読上情報処理部 151 の第2の機能は、合成音にベル音などの効果音や音楽などの背景音をのせるための編集処理についての機能である。この機能により合成音に効果音や背景音をのせることで、合成音をその使用目的に応じて、より効果的なものとすることができる。合成音に効果音や背景音をのせる編集処理の具体的な例については後述する。

20

【0031】

読上情報処理部 151 の第3の機能は、標準的な発音情報が設定された区切り文データに対してさらに標準的な読上げ形式を設定して標準設定の中間合成音データを生成する機能である。この機能による処理は、標準的な読上げ形式を区切り文データごとに設定することで行われ、その設定情報を区切り文ごとにその文頭に所定の記号で表示する形式で行われる。読上げ形式に関する設定情報としては、「音声種別：男性/女性」、「音声出力形式：8kHz/8bit/11kHz/8bit/16kHz/8bit/22kHz/8bit/8kHz/16bit/11kHz/16bit/16kHz/16bit/22kHz/16bit」、「読上げ方式：通常読上げ/いきいき読上げ」、「感情効果：平静/喜び/怒り/悲しみ」、「音響効果：なし/エコー/ロボット」、「文末ポーズ長：0.1秒~2.5秒」、「読上げ速度：10段階」、「声の高さ：10段階」、「音量：10段階」、「声の抑揚：10段階」、「無声化：する/しない」、「高域強調：する/しない」、「鼻濁音化：する/しない」などがある。ここで、「無声化」や「鼻濁音化」のように、後述の発音情報における項目と重なる項目も読上げ形式情報に含まれるが、読上げ形式におけるこれらの項目は、区切り文の全体に対するものであり、発音情報におけるそれらは区切り文中の個々の文字に対するものである。なお、無声化とは、無声子音に挟まれた「イ」音や「ウ」音の発声を省略することである。例えば「テキスト」を発声する場合、「ス」の母音部「ウ」をはっきりとは発声しない、というのがその例である。こうした無声化は、より自然な日本語の合成音を生成する上で重要である。

30

40

【0032】

発音情報処理部 152 は、発音情報の設定をユーザのマニュアル操作で変更する編集処理のための機能である。この機能の具体的な例については後述する。編集の対象となる発音情報としては、「ポーズの挿入/削除及びポーズ長：0.1秒~2.5秒」、「アクセント区切りの挿入/削除及びアクセントの立ち上がり具合調整」、「鼻濁音：あり/なし」、「無声化：あり/なし」、「アクセント：あり/なし」などがある。

【0033】

時間演算部 153 は、読上情報処理部 151 や発音情報処理部 152 で編集された合成音データで合成音を出力した場合における再生に関する時間を求め、それを編集画面に表

50

示するのに機能する。時間演算部 153 が求める時間は、区切り文を単位としており、区切り文ごとの合成音再生に要する時間（区切り文再生時間）と、テキスト文 T についての合成音再生の開始時点から各区切り文の再生が開始されるまでの時間（区切り文再生開始時間）である。このような時間データを求めて編集画面に表示することにより、例えば与えられたテキスト文 T の読上げ時間に制限があるような場合に、この制限時間との関係などで最適な読上げ速度やポーズ長を設定する編集作業をより効率よく行えるようになる。時間演算部 153 による時間データの算出は、音声ファイル作成部 18 で作成される音声ファイルを用いて行われる。

【0034】

表示処理部 154 は、ユーザが合成音の出力イメージを捉え易いようにして中間合成音データを編集のために表示・操作部 14 のモニタに表示する機能を負っている。 10

【0035】

テキスト格納部 155 は、区切り文に区切られたテキスト文 T ないしそのテキスト文 T に発音情報や読上げ形式情報を設定して生成される中間合成音データを区切り文単位の区切り文データ 1551 の集合として格納する。

【0036】

音素片格納部 16 は、合成音を生成させる元となる音素片（音素データ）を格納する機能部である。音素片はカナ 1 文字に対して数十～百数十用意されているのが通常である。音響処理部 17 は、中間合成音データに設定されている発音情報や読上げ形式情報を基に該当する音素片を音素片格納部 16 から抽出して中間合成音データにおける各カナ文字に割り当てて最終的な合成音データを生成する。音声ファイル作成部 18 は、音響処理部 17 で音素片を割り当てて生成された合成音データから例えば WAVE ファイルなど音声ファイルを作成する。これら音素片格納部 16、音響処理部 17、音声ファイル作成部 18 それぞれの機能は、音声合成分野で周知の機能である。 20

【0037】

効果音・背景音格納部 19 は、読上情報処理部 151 による編集で合成音にのせる効果音や背景音を格納する。

【0038】

音声ファイル再生部 20 は、音声ファイル作成部 18 で作成された音声ファイルを合成音として再生する機能部位であり、これにより再生される合成音はスピーカ 21 を通じて出力される。 30

【0039】

以下では、中間合成音データやその編集について具体的な例で説明する。図 2 に、中間合成音データの例を示す。この例の中間合成音データは、XML 文書（Extensible Markup Language、拡張可能なマークアップ言語）で記述されている。XML 文書は、“XML 宣言（省略可）”、“XML 型文書宣言（省略可）”、“XML 文書本体”から構成されており、図 2 では“XML 型文書宣言”は省略している。1 行目が XML 宣言を示しており、二行目以降が XML 文書本体である。2 行目以降の<文節>～</文節>で囲まれた記述が各区切り文データ（図 1 の区切り文データ 1551）を示している。区切り文データは、読上げ形式情報、発音情報（発音記号や韻律記号付きカナ文字列）および再生に関する時間情報を含む構成とされている。ここで、効果音や背景音を合成音にのせる場合には、それについての区切り文データを生成する。図 2 の例では効果音用の区切り文データが生成されている。なお、図 2 における「文節」は、音声合成分野で一般的に用いられるそれとは異なり、「区切り文」に対応している。 40

【0040】

図 3 に、中間合成音データの編集のために読上情報処理部 151 が表示する読上げ形式情報編集画面の例を示す。描画 51 は、効果音/背景音設定記号である。効果音や背景音は、区切り文を単位として設定することが可能で、ある区切り文に効果音あるいは背景音を設定すると、その区切り文の文頭に効果音/背景音設定記号が表示される。効果音や背景音の設定解除は、効果音/背景音設定記号をマウスクリックするなどの操作で行うこと 50

ができる。描画 5 2 は、区切り文を単位とする読上げ形式情報記号であり、区切り文の文頭に配置される。図の例においては読上げ形式情報として「感情効果」、「鼻濁音化」、「無声化」、「音量」、「読上げ速度」、「声の高さ」、「声の抑揚」、「文末ポーズ長」の各項目が記号表示されている。この表示における表示項目は変更することが可能で、通常は編集作業で設定を変更する場合の多い項目を表示するようにし、標準設定をそのまま用いることの多い項目の表示は省略する。読上げ形式を変更するには、表示されている読上げ形式情報記号をマウスクリックするなどして変更操作画面を表示させ、その画面で設定内容を変更する。描画 5 3 は、区切り文のテキスト内容を示している。描画 5 4 は、区切り文の区切りを示す文区切り記号である。テキスト文 T を区切り文に区切る処理は、文区切り処理部 1 1 による自動処理でなすのが基本である。ただ、その自動処理で用いる句点、疑問符、感嘆符などの文区切り基準で区切ると、区切り文が長くなり過ぎるような場合もあり得る。そのような場合に対応するために、この編集画面においてマニュアル操作で文区切り記号を挿入して区切り文の長さを調節できるようにすることも可能である。描画 5 5 は、時間演算部 1 5 3 で求めた時間データの表示であり、区切り文再生時間と区切り文再生開始時間が表示される。この表示は、時間データを表示させたい区切り文をマウスなどで指定する操作を行うと、当該区切り文についてなされる。

10

【 0 0 4 1 】

図 4 に、中間合成音データの編集のために発音情報処理部 1 5 2 が表示する発音情報編集画面の例を示す。発音情報編集画面では、発音情報つまり発音記号や韻律記号付きカナ文字列が区切り文単位で表示される。描画 6 1 は、真下の文字にアクセントがあることを示すアクセント記号である。描画 6 2 は、アクセントの区切りを示すアクセント区切り記号である。描画 6 3 は、ポーズの区切りを示すポーズ区切り記号である。描画 6 4 は、真上の文字が無声音であることを示す無声音記号である。描画 6 5 は、真上の文字が鼻濁音であることを示す鼻濁音記号である。発音についての設定を変更するには、これらの記号を消去したり、あるいは新たに追加したりする。

20

【 0 0 4 2 】

以上のように区切り文を単位として中間合成音データの編集を行えるようにしたことにより、入力されるテキスト文が長文の場合でも、中間合成音データの編集に際し、実際の出力合成音についてのイメージをより容易につかむことができるようになり、音声合成処理の効率を高めることが可能となる。

30

【 0 0 4 3 】

以下では、図 1 の音声合成装置で実行される音声合成処理について説明する。図 5 に、その処理についての流れを示す。処理開始後の最初のステップ 4 0 では文区切り処理部 1 1 による文区切り処理がなされる。具体的には、入力されたテキスト文 T の文章を句点、疑問符および感嘆符文で区切って区切り文の集合とする処理を行う。区切り文に区切られたテキスト文 T は、テキスト解析部 1 2 に渡されるとともに、区切り文データ 1 5 5 1 の集合としてテキスト格納部 1 5 5 に格納される。ステップ 4 1 では、テキスト解析部 1 2 によるテキスト解析がなされる。このテキスト解析は、構文解析と、それに基づく発音情報の設定（発音記号や韻律記号付きカナ文字列を生成）であることは上述のとおりである。ステップ 4 2 では、読上情報処理部 1 5 1 による読上げ形式の初期設定がなされる。この処理は、上述のように、標準的な読上げ形式を区切り文データごとに設定することで行われ、その設定情報を区切り文ごとにその文頭に記号で表示する形式で行われる。ステップ 4 3 では、ステップ 4 2 を経て得られる中間合成音データをモニタに表示する。ステップ 4 4 は、ステップ 4 3 で表示された中間合成音データに対する編集の要否を判断する。表示されている中間合成音データにおける発音設定と読上げ形式設定をそのままにして最終的な合成音データを作成する場合、つまり編集不要（判定が“ N o ”）の場合にはステップ 4 6 に進み、一方編集を必要とする（判定が“ Y e s ”）の場合にはステップ 4 5 に進む。

40

【 0 0 4 4 】

ステップ 4 5 では、中間合成音データについて読上げ情報や発音情報について編集を行

50

う。ステップ45での編集が済むと、再度ステップ43に戻って編集結果の確認を行う。ステップ46では、合成音再生の有無を判断する。すなわち継続して音響変換処理を行うか、それとも中間合成音データの段階で処理を停止するかの判断を行う。中間合成音データの段階で処理を停止して合成音再生を行わないとした場合は、判定が“ No ”となり、処理を終了する。中間合成音データに音響変換処理を施して合成音再生を行う場合は、判定が“ Yes ”となり、ステップ47へ進む。ステップ47では、音響処理部17による音響変換処理が行われる。すなわち中間合成音データに設定されている発音情報や読上げ形式情報を基に該当する音素片を音素片格納部16から抽出して中間合成音データにおける各カナ文字に割り当てて最終的な合成音データを生成する処理が行われる。ステップ48では、ステップ47で生成された最終的な合成音データから音声ファイル作成部18により音声ファイルを生成する。ステップ49では、音声ファイル作成部18で作成された音声ファイルにより音声ファイル再生部20が合成音を出力する。ユーザはこの出力合成音を試聴し、それがイメージ通りでない場合にはステップ44に戻り、イメージ通りの出力合成音となるまでステップ44以降の処理を繰り返す。

10

【産業上の利用可能性】

【0045】

本発明は、音声合成について、合成音データの編集に際し、実際の出力合成音についてのイメージをより容易につかむことを可能とし、それにより音声合成処理の効率を高めることができるようにするなどの効果をもたらすものであり、音声合成分野において広く利用することができる。

20

【図面の簡単な説明】

【0046】

【図1】一実施形態による音声合成装置の構成を示す図である。

【図2】中間合成音データの例を示す図である。

【図3】読上げ形式情報編集画面の例を示す図である。

【図4】発音情報編集画面の例を示す図である。

【図5】図1の音声合成装置で実行される音声合成処理における処理の流れを示す図である。

【符号の説明】

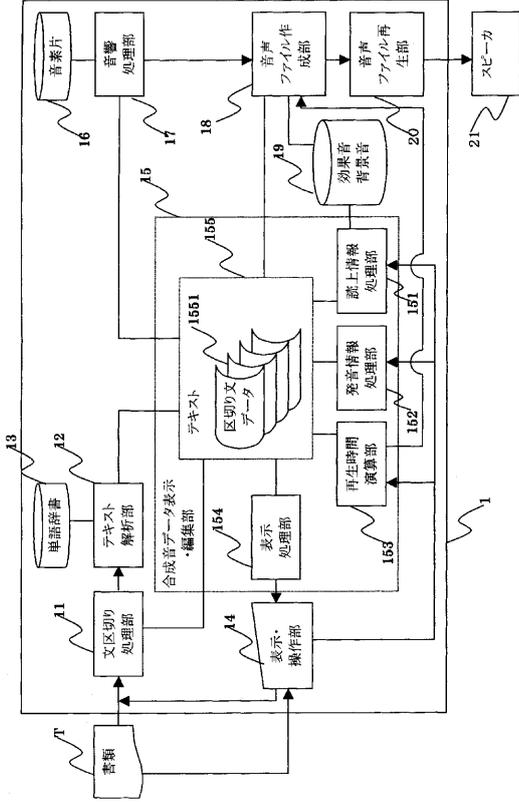
【0047】

30

- 1 音声合成装置
- 11 文区切り処理部
- 12 テキスト解析部
- 15 合成音データ表示・編集部
- 19 効果音・背景音格納部
- 151 読上情報処理部
- 152 発音情報処理部
- 153 時間演算部
- 1551 区切り文データ

40

【図1】



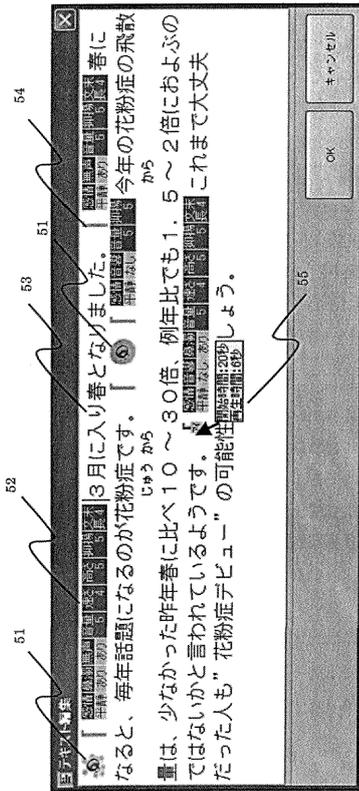
【図2】

```

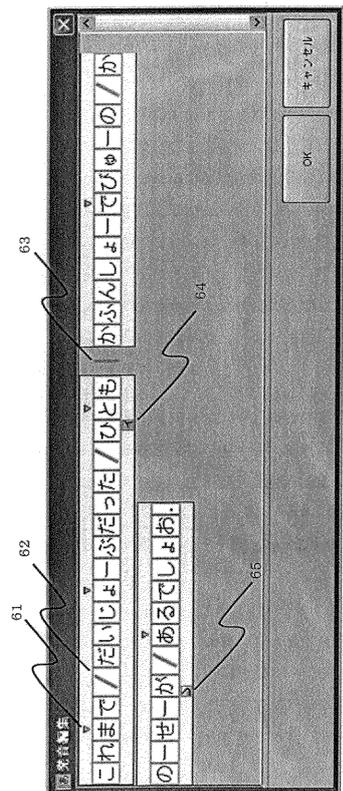
<?xml version="1.0" ?>
<データ ID="D_001">
<文題 テキスト="">
<再生時間>2.00</再生時間>
<効果音>*サウンド*効果音1.wav</効果音>
</文題>
</データ>
<文題 テキスト="">
<再生時間>3.00</再生時間>
<効果音>*サウンド*効果音1.wav</効果音>
</文題>
</データ>
<?xml version="1.0" ?>
<データ ID="D_001">
<文題 テキスト="">
<再生時間>2.00</再生時間>
<効果音>*サウンド*効果音1.wav</効果音>
</文題>
</データ>
<?xml version="1.0" ?>
<データ ID="D_001">
<文題 テキスト="">
<再生時間>2.00</再生時間>
<効果音>*サウンド*効果音1.wav</効果音>
</文題>
</データ>
<?xml version="1.0" ?>
<データ ID="D_001">
<文題 テキスト="">
<再生時間>2.00</再生時間>
<効果音>*サウンド*効果音1.wav</効果音>
</文題>
</データ>

```

【図3】



【図4】



【図5】

