

(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2009-3394

(P2009-3394A)

(43) 公開日 平成21年1月8日(2009.1.8)

|                                |  |               |         |             |
|--------------------------------|--|---------------|---------|-------------|
| (51) Int.Cl.                   |  | F I           |         | テーマコード (参考) |
| <b>G 1 0 L 13/08 (2006.01)</b> |  | G 1 0 L 13/08 | 1 3 1 Z |             |
| <b>G 1 0 L 13/00 (2006.01)</b> |  | G 1 0 L 13/00 | 1 0 0 K |             |
|                                |  | G 1 0 L 13/08 | 1 2 7 F |             |
|                                |  | G 1 0 L 13/08 | 1 2 7 D |             |

審査請求 未請求 請求項の数 10 O L (全 37 頁)

|           |                              |          |                                |
|-----------|------------------------------|----------|--------------------------------|
| (21) 出願番号 | 特願2007-167018 (P2007-167018) | (71) 出願人 | 000005223                      |
| (22) 出願日  | 平成19年6月25日 (2007. 6. 25)     |          | 富士通株式会社                        |
|           |                              |          | 神奈川県川崎市中原区上小田中4丁目1番1号          |
|           |                              | (74) 代理人 | 100083725                      |
|           |                              |          | 弁理士 畝本 正一                      |
|           |                              | (72) 発明者 | 西池 理香                          |
|           |                              |          | 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内 |
|           |                              | (72) 発明者 | 佐々木 均                          |
|           |                              |          | 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内 |
|           |                              | (72) 発明者 | 片江 伸之                          |
|           |                              |          | 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内 |

最終頁に続く

(54) 【発明の名称】 音声読み上げのための装置、プログラム及び方法

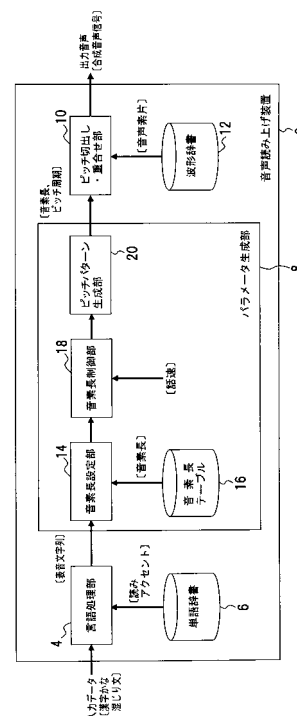
(57) 【要約】

【課題】 文字データの音声読み上げに関し、音素長の調整によって音声読み上げの認識性を高める。

【解決手段】 文字データを音声に変換して読み上げる装置（音声読み上げ装置2）、プログラム及び方法に関し、文字データからポーズの存在を認識し、ポーズ直後の音素やその他の音素の音素長を制御する。この音素長の制御は例えば、読み上げ速度に応じて実行し、前記読み上げ速度が高速の場合にはポーズ直後の音素が持つ音素長を伸長させ、また、特定の音素についてはその音素が持つ音素長を短縮させ又は基準速度と同一の音素長にする構成である。

【選択図】 図1

第1の実施形態の形態に係る音声読み上げ装置の構成例を示すブロック図



## 【特許請求の範囲】

## 【請求項 1】

文字データを音声に変換して読み上げる音声読み上げ装置であって、  
前記文字データから音素の種類を判定する音素判定部と、  
音素に読み上げ速度に応じた音素長を設定し、音素が前記文字データのポーズ直後の音素である場合に前記音素判定部の判定結果に基づき、ポーズ直後の前記音素の音素長を調整する音素長調整部と、  
を備えることを特徴とする音声読み上げ装置。

## 【請求項 2】

請求項 1 の音声読み上げ装置において、  
音素の読み上げ速度を判定する速度判定部を備え、前記音素長調整部は、前記読み上げ速度の判定結果に基づき、前記読み上げ速度が高速の場合にはポーズ直後の音素が持つ音素長を伸長させることを特徴とする音声読み上げ装置。 10

## 【請求項 3】

請求項 1 の音声読み上げ装置において、  
前記音素が摩擦音である場合に、前記音素長調整部は、前記音素判定部の判定結果に基づいて前記摩擦音の音素を伸長させることを特徴とする音声読み上げ装置。

## 【請求項 4】

請求項 1 の音声読み上げ装置において、  
呼気段落の長さを演算する呼気段落演算部を備え、前記音素長調整部は、前記音素長の調整分を前記呼気段落演算部の演算結果に基づき、前記呼気段落の各音素長を比例配分して増減させることを特徴とする音声読み上げ装置。 20

## 【請求項 5】

請求項 1 の音声読み上げ装置において、  
読み上げ文章の長さを演算する文章演算部を備え、前記音素長調整部は、前記音素長の調整分を前記文章演算部の演算結果に基づき、前記文章の各音素長を比例配分して増減させることを特徴とする音声読み上げ装置。

## 【請求項 6】

請求項 1 の音声読み上げ装置において、  
前記音素長調整部は、読み上げ速度が高速の場合、前記文字データ中の一部又は全部のポーズが持つポーズ長を前記読み上げ速度に応じた長さより短縮することを特徴とする音声読み上げ装置。 30

## 【請求項 7】

請求項 1 の音声読み上げ装置において、  
前記音素長調整部は、読み上げ速度が高速の場合、前記文字データ中の一部又は全部のポーズを削除することを特徴とする音声読み上げ装置。

## 【請求項 8】

請求項 2 の音声読み上げ装置において、  
前記音素長調整部は、前記音素長の伸長に対応し、ポーズ長を含む他の音素長を短縮することを特徴とする音声読み上げ装置。 40

## 【請求項 9】

文字データを音声に変換して読み上げる手順をコンピュータに実行させる音声読み上げプログラムであって、  
前記文字データから音素の種類を判定する手順と、  
音素に読み上げ速度に応じた音素長を設定する手順と、  
音素が前記文字データのポーズ直後の音素である場合に前記判定の結果に基づき、ポーズ直後の前記音素の音素長を調整する手順と、  
を前記コンピュータに実行させることを特徴とする音声読み上げプログラム。

## 【請求項 10】

文字データを音声に変換して読み上げる音声読み上げ方法であって、 50

前記文字データから音素の種類を判定する工程と、  
音素に読み上げ速度に応じた音素長を設定する工程と、  
音素が前記文字データのポーズ直後の音素である場合に前記判定の結果に基づき、ポーズ直後の前記音素の音素長を調整する工程と、  
を備えることを特徴とする音声読み上げ方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、文書等の表音文字を含む文字データを音声に変換して出力する音声読み上げのための装置、プログラム及び方法に関し、音素長を読み上げ速度に応じて制御し、特に、高速読み上げ等、読み上げ速度に応じて特定の音素長等を伸縮させる音声読み上げのための装置、プログラム及び方法に関する。

10

【背景技術】

【0002】

表音文字を含む文字データを解析し、その文字データから音声合成法により音声合成を行い、文字データを音声として出力する所謂音声読み上げの技術が知られている。携帯電話機等の携帯端末装置にあっては、メール等の自由文章を読み上げる音声合成機能が普及し始めている。また、パーソナルコンピュータ（PC）にあっては、スクリーンリーダーと呼ばれるソフトウェアが普及し始めている。文章の内容を音声により理解する場合には、聴覚に作用する母音、子音、ポーズ等を表す音素の長さが認識性を高める上で重要なファクターとなる。

20

【0003】

このような音声読み上げに関し、特許文献1には、発話速度情報が予め定められた値未満のときは、発話速度をその情報に基づいて標準より速くするため、モーラ長を最小限に設定し、発話速度情報に応じた短いフレーム周期を設定し、発話速度情報が予め定められた値以上のときは、発話速度をその情報に基づいて標準より遅くするため、発話速度情報に応じた長いモーラ長を設定し、フレーム周期を最大値に設定する音声合成が開示されている。

【特許文献1】特開平6-149283号公報（要約及び図1等）

【発明の開示】

30

【発明が解決しようとする課題】

【0004】

ところで、読み上げ速度（話速）を設定可能にした場合、話速に反比例して各音素長が設定されるものとする。例えば、話速を2倍の速さとするれば、その音素長は $1/2$ となり、話速を $1/2$ に遅くすれば、音素長は2倍となる。このように話速と音素長との関係を単純な関係に設定し、即ち、話速と音素長とを単純に反比例とすれば、通常の話速では自然な場合（聴き取り易い場合）であっても、高速読みや低速読みの場合には、聞き取り難く、違和感があり、認識性を低下させる場合がある。

【0005】

斯かる要求や課題について、特許文献1にはその開示や示唆はなく、それを解決する構成等についての開示や示唆はない。

40

【0006】

そこで、本発明の目的は、文字データの音声読み上げに関し、音素長の調整によって音声読み上げの認識性を高めることにある。

【0007】

斯かる目的は、文字データのポーズ直後の音素又はその他の音素等に対する認識性が読み上げ速度に応じて影響を受けるとの知見に基づくものである。

【0008】

斯かる目的を具体的に述べれば、文字データの音声読み上げに関し、聴感上、読み上げられる音声の聞き取り易さを向上させることにある。

50

【課題を解決するための手段】

【0009】

上記目的を達成するため、本発明は、文字データを音声に変換して読み上げる装置、プログラム及び方法に関し、文字データからポーズの存在を認識し、ポーズ直後の前記音素が持つ音素長を制御する。この音素長の制御は例えば、読み上げ速度に応じて実行し、前記読み上げ速度が高速の場合にはポーズ直後の音素が持つ音素長を伸長させ、また、特定の音素についてはその音素が持つ音素長を短縮させ又は基準速度と同一の音素長にする構成である。斯かる構成により、聴感上、読み上げられる音声の聞き取り易さが向上し、音声読み上げの認識性を高めている。

【0010】

そこで、上記目的を達成するため、本発明の第1の側面は、文字データを音声に変換して読み上げる音声読み上げ装置であって、前記文字データから音素の種類を判定する音素判定部と、音素に読み上げ速度に応じた音素長を設定し、音素が前記文字データのポーズ直後の音素である場合に前記音素判定部の判定結果に基づき、ポーズ直後の前記音素の音素長を調整する音素長調整部とを備える構成である。

【0011】

斯かる構成によれば、文字データ、表音文字列から音素の種類を判別し、読み上げ速度に応じた音素長を設定するとともに、ポーズ直後の音素の音素長を調整するので、読み上げ速度が高速になっても、聞き取り難さがなく、また、音切れ等の違和感を生じさせることがなく、音声の認識性が高められる。

【0012】

上記目的を達成するためには、上記音声読み上げ装置において、好ましくは、音素の読み上げ速度を判定する速度判定部を備え、前記音素長調整部は、前記読み上げ速度の判定結果に基づき、前記読み上げ速度が高速の場合にはポーズ直後の音素が持つ音素長を伸長させる構成としてもよい。斯かる構成によれば、ポーズ直後の音素長を伸長させるので、既述したように、読み上げ速度が高速になっても、聞き取り難さがなく、また、音切れ等の違和感を生じさせることがなく、音声の認識性が高められる。

【0013】

上記目的を達成するためには、上記音声読み上げ装置において、好ましくは、前記音素が摩擦音である場合に、前記音素長調整部は、前記音素判定部の判定結果に基づいて前記摩擦音の音素を伸長させる構成としてもよい。斯かる構成によれば、表音文字列から摩擦音を選択し、その摩擦音の音素長を伸長させるので、聞き取り難さがなく、また、音切れ等の違和感を生じさせることがなく、音声の認識性が高められる。

【0014】

上記目的を達成するためには、上記音声読み上げ装置において、好ましくは、呼気段落の長さを演算する呼気段落演算部を備え、前記音素長調整部は、前記音素長の調整分を前記呼気段落演算部の演算結果に基づき、前記呼気段落の各音素長を比例配分して増減させる構成としてもよい。斯かる構成によれば、音素長の調整分を呼気段落を単位に他の音素長の調整で増減して補完するので、読み上げ時間の間延びを防止できる。

【0015】

上記目的を達成するためには、上記音声読み上げ装置において、好ましくは、読み上げ文章の長さを演算する文章演算部を備え、前記音素長調整部は、前記音素長の調整分を前記文章演算部の演算結果に基づき、前記文章の各音素長を比例配分して増減させる構成としてもよい。斯かる構成によれば、音素長の調整分を文章を単位に他の音素長の調整で増減して補完するので、読み上げ時間の間延びや再生時間の延びを防止できる。

【0016】

上記目的を達成するためには、上記音声読み上げ装置において、好ましくは、前記音素長調整部は、読み上げ速度が高速の場合、前記文字データ中の一部又は全部のポーズが持つポーズ長を前記読み上げ速度に応じた長さより短縮する構成としてもよい。斯かる構成によれば、ポーズが存在するための間延び感を抑制でき、再生時間が延びることを防止で

10

20

30

40

50

きる。

【0017】

上記目的を達成するためには、上記音声読み上げ装置において、好ましくは、前記音素長調整部は、読み上げ速度が高速の場合、前記文字データ中の一部又は全部のポーズを削除する構成としてもよい。斯かる構成によれば、ポーズの削除により、聞き取り易さを損なうことなく、再生時間の短縮を図ることができる。

【0018】

上記目的を達成するためには、上記音声読み上げ装置において、好ましくは、前記音素長調整部は、前記音素長の伸長に対応し、ポーズ長を含む他の音素長を短縮する構成としてもよい。斯かる構成によれば、音素長の伸長に対応してポーズ長を含む他の音素長を短縮するので、聞き取り易さを損なうことなく、再生時間の短縮を図ることができる。

10

【0019】

上記目的を達成するため、本発明の第2の側面は、文字データを音声に変換して読み上げる手順をコンピュータに実行させる音声読み上げプログラムであって、前記文字データから音素の種類を判定する手順と、音素に読み上げ速度に応じた音素長を設定する手順と、音素が前記文字データのポーズ直後の音素である場合に前記判定の結果に基づき、ポーズ直後の前記音素の音素長を調整する手順とを前記コンピュータに実行させる構成である。斯かる構成によれば、第1の側面で述べた通り、上記目的を達成することができる。

【0020】

上記目的を達成するため、本発明の第3の側面は、文字データを音声に変換して読み上げる音声読み上げ方法であって、前記文字データから音素の種類を判定する工程と、音素に読み上げ速度に応じた音素長を設定する工程と、音素が前記文字データのポーズ直後の音素である場合に前記判定の結果に基づき、ポーズ直後の前記音素の音素長を調整する工程とを備える構成である。斯かる構成によれば、第1の側面で述べた通り、上記目的を達成することができる。

20

【発明の効果】

【0021】

本発明によれば、次のような効果が得られる。

【0022】

(1) 文字データを音声に変換して読み上げる音素について、ポーズ直後の話頭の音素長を伸長させるので、聞き取り易くでき、認識性を高めることができる。

30

【0023】

(2) 話頭の音素長を伸長すれば、一定比率で音素長を削減する場合に比較して聞き取り易くできる。

【0024】

(3) 摩擦音の音素長を伸長すれば、聞き取り易くでき、認識性を高めることができる。

【0025】

(4) 特定の音素長を伸長した場合、その音素長の伸長に対し、その伸長分に応じて他の音素の音素長を短縮すれば、読み上げ速度を高速化しても、聞き取り易さが損なわれることがなく、再生時間も短縮できる。

40

【0026】

(5) 特定の音素長を伸長した場合、その音素長の伸長に対し、その伸長分に応じて一部又は全部のポーズ長を短縮し、又は、削除すれば、読み上げ速度を高速化しても、聞き取り易さが損なわれることがなく、再生時間も短縮できる。

【0027】

そして、本発明の他の目的、特徴及び利点は、添付図面及び各実施の形態を参照することにより、一層明確になるであろう。

【発明を実施するための最良の形態】

【0028】

〔第1の実施の形態〕

50

## 【0029】

本発明の第1の実施の形態について、図1及び図2を参照する。図1は、音声読み上げ装置の構成例を示すブロック図、図2は、音声読み上げ装置の音素長制御部の構成例を示すブロック図である。

## 【0030】

この音声読み上げ装置2は、本発明の音声読み上げのための装置、プログラム及び方法の構成例であって、コンピュータで構成され、例えば、テキスト文章（日本語では漢字かな混じり文）等、文字データを音声に変換して読み上げる音声合成装置で構成され、文字データ中のポーズ直後の音素が持つ音素長を話速（読み上げ速度）に応じて制御することにより、文字データから得た出力音声の聞き易さを高め、合成音声（読み上げ出力）の認識性を向上させたものである。ここで、文字データは音声読み上げの対象であって、表音文字、その文字列、ポーズを含むデータであり、表音文字又はその表音文字列は、音声合成で使用する韻律記号付き発音記号からなる中間言語であって、韻律記号付きの発音記号（読み仮名）である。文字データに含まれるポーズは、音声に変換されない区間等の無音区間であって、ポーズ直後の音素が持つ音素長の制御情報としては破裂音の直前の休止期間や促音は除かれる。例えば、「卒業して、信用金庫に・・・」（ローマ字表記：so t s u g y o u s h i t e, s h i n y o u k i n k o n i・・・）といった日本語文章では、「卒業して」と、「信用金庫」との間に、無音区間となる句点「、」が存在し、これがポーズの一例である。また、このポーズと呼気段落との関係について、呼気段落は人間が一息に発声する単位であり、この呼気段落の前後の息継ぎには既述のポーズが入ることになる。

## 【0031】

斯かる機能を実現するため、この音声読み上げ装置2では、図1に示すように、言語処理部4と、単語辞書6と、パラメータ生成部8と、ピッチ切出し・重合せ部10と、波形辞書12とが備えられている。

## 【0032】

言語処理部4は、漢字かな混じり文を入力して、単語辞書6を参照して単語を解析し、読み、アクセント、イントネーションを決定し、表音文字列（中間言語）を出力する言語処理手段である。また、単語辞書6には、単語の種類（品詞等）と読みや、アクセントの位置等が格納されている。

## 【0033】

アクセントと、イントネーションとは、物理的にはピッチ周波数の時間的変化パターンと密接に関係している。具体的には、アクセント位置でピッチ周波数は高くなり、イントネーションが上がると、ピッチ周波数が高くなる。そこで、言語処理部4では、入力テキストでの句読点や、単語解析によって抽出された文節に基づいて、既述の呼気段落に分割する。

## 【0034】

パラメータ生成部8は、音素継続時間、ポーズ継続時間やピッチ周波数パターンの設定を行うパラメータ生成手段である。このパラメータ生成部8では、話速に応じて音素長の制御を行っている。

## 【0035】

このパラメータ生成部8には、音素長設定部14と、音素長テーブル16と、音素長制御部18と、ピッチパターン生成部20とが備えられている。

## 【0036】

言語処理部4で生成された表音文字列の段階において、どの音素を音声合成するかが決定される。音素長設定部14では、各音素についての音素長設定手段であって、標準的な話速における音素長が設定される。音素長テーブル16は、当該音素と前後の音素に応じた標準的な話速における音素長を格納する手段である。そこで、音素長の設定例としては、当該音素と前後の音素に応じた標準的な話速における音素長（データベースから抽出した値）を音素長テーブル16に格納しておき、この値を参照して音素長が設定される。こ

の音素長は、他のパラメータ要素を以て修正する構成としてもよい。

#### 【0037】

音素長制御部18は、音素長設定部14で設定された標準的な話速での音素長を話速に応じて制御する音素長制御手段である。話速は、図示しない読み上げ速度（ユーザ設定等）の調整手段等から制御情報として音素長制御部18に加えられる。

#### 【0038】

音素長制御部18には、図2に示すように、音素長調整部24と、話速判定部26と、音素判定部28とが含まれる。音素長調整部24は、話速判定部26及び音素判定部28の各判定出力を受け、音素の長さやポーズの長さを調整する。話速判定部26は、入力された話速を判定し、その話速が標準速度、高速又は低速の何れであるかを判定し、その判定出力を音素長調整部24に加える。この場合、話速判定部26の出力する判定出力には、標準速度、高速又は低速を表す出力、その話速レベルを表す出力が含まれる。また、音素判定部28は、音素長設定部14（図1）で設定された音素長を持つ音素やポーズ等を判定し、その判定出力を音素長調整部24に加える。

#### 【0039】

そこで、このような音素長制御部18によれば、例えば、標準的な話速に対する所定の話速に反比例した音素長とし、具体的な数値を例示すれば、標準的な話速を毎秒7モーラを目安とした場合、毎秒14モーラの話速が設定されていれば、各音素長を半分にし、毎秒6モーラの話速が設定されていれば、7/6の音素長とする。ここで、モーラとは、拍を表し、概ね仮名書きしたときの一文字相当の単位であって、拗音（小さい「ゃ」「ゅ」「ょ」）「きゃ」等は1モーラである。日本語の場合、一文字のモーラが類似の長さを持つ言語である。

#### 【0040】

ピッチパターン生成部20は、表音文字列でのアクセント情報等を加味して、各音素でのピッチ周期を設定するパターン生成手段である。

#### 【0041】

ピッチ切出し・重合せ部10は例えば、PSOLA法（Pitch-Synchronous Overlap-add：波形の加算重畳によるピッチ変換方法）を使用するピッチ切出し・重合せ手段である。また、波形辞書12には音声波形と、どの部分がどの音素かを示す音素ラベル、有声音についてピッチ周期を示すピッチマークが格納されている。そこで、ピッチ切出し・重合せ部10では、パラメータ生成部8で生成されたパラメータを元に波形辞書12から2周期分の音声波形を切り出し、窓関数（例えばハニング窓）を掛け、必要に応じて振幅調整のゲインを掛ける処理を実行し、波形辞書12でのピッチ周波数と所望のピッチ周波数が違えばピッチ変換し、切り出された波形をオーバーラップさせて加算することにより、合成音声信号が出力される。

#### 【0042】

この音声読み上げ装置のハードウェアについて、図3、図4及び図5を参照する。図3は、音声読み上げ装置を搭載した携帯端末装置の一例を示すブロック図、図4は、携帯端末装置の構成例を示す図、図5は、画面表示例を示す図である。

#### 【0043】

この携帯端末装置200は、既述の音声読み上げ装置2が適用された一例であって、斯かる構成に本発明の音声読み上げのための装置、方法又はプログラムが限定されるものではない。この携帯端末装置200では、通信機能や、メール文等のテキスト文章（日本語では漢字かな混じり文）等、文字データを音声に変換して出力する機能を有する。そこで、この携帯端末装置200には、図3に示すように、プロセッサ202と、記憶部204と、無線部206と、入力部208と、表示部210と、音声入力部212と、音声出力部214とが備えられている。

#### 【0044】

プロセッサ202は、電話通信や、音声合成等の音声読み上げの実行、その他の制御を行う制御手段であって、CPU（Central Processing Unit）又はMPU（Micro Proces

10

20

30

40

50

sor Unit)で構成され、記憶部204にあるOS(Operating System)やアプリケーションプログラムを実行する。このアプリケーションプログラムには音声読み上げの処理手順を実行するプログラム等が含まれる。

#### 【0045】

記憶部204はプロセッサ202で実行されるプログラムや、その実行に用いる各種データを格納するとともに、処理エリアを形成する記録媒体であって、プログラム記憶部216、データ記憶部218、RAM(Random-Access Memory)220で構成されている。プログラム記憶部216にはOSやアプリケーションプログラムが格納され、データ記憶部218には単語辞書6、波形辞書12及び音素長テーブル16(図1)が形成され、既述のデータが格納されている。RAM220は、ワークエリアを構成する。

10

#### 【0046】

無線部206は基地局と無線によって音声信号電波やパケット信号電波等の送受信を行うための無線通信手段であって、プロセッサ202によって制御される。

#### 【0047】

入力部208はユーザの操作により制御データや表示部210に展開されるダイアログに対する応答を入力するための手段であって、キーボードやタッチパネル等で構成される。

#### 【0048】

表示部210はプロセッサ202によって制御され、文字や図形等を表示する表示手段であって、例えば、LCD(Liquid Crystal Display)素子で構成される。この表示部210には音声読み上げのテキスト文章等が表示される。

20

#### 【0049】

音声入力部212はプロセッサ202で制御される音声入力手段であって、マイクロフォン222を備える。入力音声はマイクロフォン222で音声信号に変換され、その音声信号がデジタル信号に変換されてプロセッサ202に取り込まれる。

#### 【0050】

音声出力部214はプロセッサ202で制御される音声出力手段であって、音声変換手段としてレシーバ224と、スピーカ226R、226Lとを備えている。音声読み上げの合成音声は、これらレシーバ224、スピーカ226R、226Lから再生される。

#### 【0051】

この携帯端末装置200において、既述の音声読み上げ装置2は例えば、プロセッサ202、記憶部204、表示部210、音声出力部214等で構成される。

30

#### 【0052】

そして、この携帯端末装置200は、図4に示すように、一例として筐体228に第1の筐体部230と、第2の筐体部232とが含まれ、これら筐体部230、232はヒンジ部234で連結して折畳み可能に構成され、筐体部230には入力部208、マイクロフォン222が配置され、筐体部232には表示部210、レシーバ224、スピーカ226R、226Lが設置されている。入力部208には文字等の入力に用いる複数の記号キー236、カーソルキー238、決定キー240等が配置されている。

#### 【0053】

そこで、この携帯端末装置200による音声読み上げでは、メール文や小説文等の各種のテキスト文が対象となり、表示部210の画面上に展開される文章等が音声合成されてレシーバ224やスピーカ226R、226Lから再生される。その場合、図5に示すように、表示部210に展開されたメール文表示画面242にはメール文が表示され、このメール文が音声として出力される。この例では、メール文表示画面242に「山梨県の高校を卒業して、信用金庫に入って4年目です。」と表示されており、これが音声として再生される。

40

#### 【0054】

次に、音素長の制御について、図6を参照する。図6は、第1の実施の形態に係る音素長制御の処理手順の一例を示すフローチャートである。

50

## 【0055】

この処理手順は、音声読み上げのためのプログラム又は方法の一例であって、第1の実施の形態では、高速読み上げの際にポーズの直後の音素即ち、話頭か否かを判断し、話頭の音素であれば、その音素長を制御する手順又は工程として話頭の音素を伸長させる手順又は工程を含んでいる。この処理手順は、音声読み上げ装置2（図1）の音素長制御部18（図2）で実行される。この実施の形態では、話頭（呼気段落単位での先頭音素）は、話速に応じての修正後、音素長を他の音素長の例えば、1.5〔倍〕にすることで、聞き取り易さを高めている。

## 【0056】

そこで、この処理手順は、図6に示すように、言語処理（ステップS101）、音素長設定処理（ステップS102）を実行する。言語処理（ステップS101）は、言語処理部4において実行され、入力データから表音文字列を生成し、この段階でどの音素を音声合成するかが決定される。次に、音素長設定処理（ステップS102）は、音素長設定部14において実行され、各音素について、標準的な話速における音素長が設定される。この場合、音素長は、当該音素と前後の音素に応じた標準的な話速における音素長が音素長テーブル16を参照して設定される。

## 【0057】

このような音素長の設定処理の後、呼気段落内の音素についての処理として、音素番号nを初期化（ $n=1$ ）し（ステップS103）、話速に応じた音素長の制御を行う（ステップS104～S110）。この音素長の制御は呼気段落を単位として実行され、ステップS105～S109が呼気段落の音素処理のループである。この音素長の制御には、制御対象である音素の判定処理、その判定結果に対応した音素長の調整処理が含まれる。

## 【0058】

音素長制御部18では、入力された話速情報が認識に基づき、その話速に応じて音素長が制御され、この場合、固定倍の音素長が設定され（ステップS104）、設定された話速が高速読み、かつ先頭音素（ $n=1$ ）であるか否かが判定される（ステップS105）。即ち、この判定処理では、ポーズ直後（話頭）の音素が持つ音素長が調整対象として特定されることになる。

## 【0059】

話速が高速読みかつ先頭音素（ $n=1$ ）であれば（ステップS105のYes）、その音素長が所定倍として例えば、1.5〔倍〕に設定、即ち、調整され（ステップS106）、また、話速が高速読みかつ先頭音素（ $n=1$ ）でなければ（ステップS105のNo）、その音素長を調整しない。このような調整又は無調整の後、音素番号nの更新（ $n=n+1$ ）を行い（ステップS107）、呼気段落内の音素は終了したか、即ち、呼気段落内の音素番号nが音素数nに到達したか否かが判定され（ステップS108）、呼気段落内の全ての音素についての処理が実行される。

## 【0060】

呼気段落内の音素の処理が行われ、呼気段落の終端のポーズに到達した場合には、そのポーズ長を話速に応じて固定倍とし（ステップS109）、終了判定が行われる（ステップS110）。この終了判定では、入力データの全データの処理が完了したか否かが判定され（ステップS110）、全てのデータの処理が完了するまで、ステップS103からステップS110の処理が繰り返される。この終了判定の後、音声合成が実行され（ステップS111）、音声が出力される。

## 【0061】

このように、呼気段落単位での先頭音素が話速に応じて修正され、高速読み上げの際にポーズ直後の音素の音素長を既述のように、その一例として1.5〔倍〕に調整することにより、高速読み上げによる不明瞭感が解消されて聞き取り易くなり、音声に変換された読み上げ文の認識性を向上させることができる。

## 【0062】

〔第2の実施の形態〕

10

20

30

40

50

## 【0063】

次に、第2の実施の形態について、図7を参照する。図7は、第2の実施の形態に係る音素長制御の処理手順の一例を示すフローチャートである。

## 【0064】

この処理手順は、音声読み上げのプログラム又は方法の一例であって、既述の音声読み上げ装置2（図1）及び音素長制御部18（図2）を用いて実行されるが、この第2の実施の形態では、第1の実施の形態の音素長の調整に加え、音素が摩擦音であるか否かの判定を行い、高速読み上げの際に、判定された摩擦音の音素長の調整として音素長を伸長させ、音声読み上げのトータル再生時間を極端に延長させることなく、聞き取り易さを高めている。

10

## 【0065】

この第2の実施の形態では、音素長を伸張すべき音素を特定するため、音素判定部28（図2）において、摩擦音であるか否かを判定し、その判定に基づき、摩擦音の音素長の伸長処理を実行している。

## 【0066】

そこで、この処理手順では、図7に示すように、言語処理（ステップS201）、音素長設定処理（ステップS202）を実行する。これら言語処理（ステップS201）及び音素長設定処理（ステップS202）の後、呼気段落内の音素の処理として、音素番号nを初期化（ $n=1$ ）し（ステップS203）、話速に応じた音素長の制御を行う（ステップS204～S214）。この音素長の制御が呼気段落を単位とするのは第1の実施の形態と同様である。

20

## 【0067】

音素長制御部18では、入力された話速情報の認識に基づき、その話速に応じて音素長が制御され、この場合、固定倍の音素長が設定され（ステップS204）、高速読みかつ先頭音素（ $n=1$ ）であるか否かが判定される（ステップS205）。即ち、この判定処理では、ポーズ直後（話頭）の音素が持つ音素長が調整対象として特定されることになる。

## 【0068】

話速が高速読みかつ先頭音素（ $n=1$ ）であれば（ステップS205のYes）、音素が摩擦音か否かが判定され（ステップS206）、話速が高速読みかつ先頭音素（ $n=1$ ）かつ摩擦音であれば（ステップS206のYes）、その音素長が所定倍即ち、 $\alpha$ 倍として例えば、 $\alpha=1.7$ 〔倍〕に設定ないし調整され（ステップS207）、また、先頭音素（ $n=1$ ）でも摩擦音でもなければ（ステップS208のNo）、その音素長の調整はない。即ち、この場合、ステップS204で固定倍とされたそのままの状態が維持される。

30

## 【0069】

高速読みかつ先頭音素の場合には（ステップS206のNo）、その音素長が所定倍即ち、 $\beta$ 倍として例えば、 $\beta=1.5$ 〔倍〕に設定ないし調整され（ステップS209）、高速読みかつ摩擦音の場合には（ステップS208のYes）、その音素長が所定倍即ち、 $\gamma$ 倍として例えば、 $\gamma=1.4$ 〔倍〕に設定ないし調整される（ステップS210）。

40

## 【0070】

従って、高速読みかつ先頭音素かつ摩擦音の場合、高速読みかつ先頭音素の場合、高速読みかつ摩擦音の場合、先頭音素でも摩擦音でもない場合について、音素長の調整又は無調整は次の表1の通りとなる。

## 【0071】

【表 1】

|               | 高速読みかつ<br>先頭音素            | 左記以外                      |
|---------------|---------------------------|---------------------------|
| 高速読みかつ<br>摩擦音 | $\alpha$ 倍<br>(例: 1. 7 倍) | $\gamma$ 倍<br>(例: 1. 4 倍) |
| 上記以外          | $\beta$ 倍<br>(例: 1. 5 倍)  | そのまま                      |

10

## 【0072】

そして、このような処理の後、音素番号  $n$  の更新 ( $n = n + 1$ ) を行い (ステップ S 2 1 1)、呼気段落内の音素は終了したか否かが判定され (ステップ S 2 1 2)、呼気段落内の全ての音素についての処理が実行される。

## 【0073】

呼気段落の音素の処理が行われ、呼気段落の終端のポーズに到達した場合には、そのポーズ長を話速に応じて固定倍とし (ステップ S 2 1 3)、終了判定が行われる (ステップ S 2 1 4)。全てのデータの処理が完了するまで、ステップ S 2 0 3 からステップ S 2 1 4 の処理が繰り返される。この終了判定の後、音声合成が実行され (ステップ S 2 1 5)、

20

## 【0074】

このように、呼気段落単位での先頭音素及び摩擦音が話速に応じて修正され、ポーズ直後の音素の音素長や摩擦音の音素長の場合、又は何れでもない場合には、既述のように、音素長の伸長を異ならせた設定とするので、合成音声の聞き取り易さが高められ、音声に変換された読み上げ文の認識性が改善される。

## 【0075】

〔第 3 の実施の形態〕

## 【0076】

次に、第 3 の実施の形態について、図 8 を参照する。図 8 は、第 3 の実施の形態に係る音素長制御の処理手順の一例を示すフローチャートである。

30

## 【0077】

この処理手順は、音声読み上げのプログラム又は方法の一例であって、既述の音声読み上げ装置 2 (図 1) 及び音素長制御部 1 8 (図 2) を用いて実行されるが、この第 3 の実施の形態では、第 1 の実施の形態の音素長の調整に加え、即ち、その音素長の伸長に対し、その他の音素の音素長を短縮することにより、読み上げ文の音声変換時間を伸長させることなく、聞き取り易さを高めている。この実施の形態では、その他の音素として、母音の音素長を短縮している。

## 【0078】

この第 3 の実施の形態では、音素長を調整すべき音素を特定するため、音素判定部 2 8 (図 2) において、音素が母音であるか否かを判定し、その判定に基づき、母音の音素長に対する短縮処理を実行している。

40

## 【0079】

そこで、この処理手順では、図 8 に示すように、言語処理 (ステップ S 3 0 1)、音素長設定処理 (ステップ S 3 0 2)、呼気段落内の音素の処理として、音素番号  $n$  の初期化 ( $n = 1$ ) (ステップ S 3 0 3)、話速に応じた音素長の制御を行う (ステップ S 3 0 4 ~ S 3 1 2)。この音素長の制御が呼気段落を単位とするのは第 1 の実施の形態と同様である。

## 【0080】

音素長制御部 1 8 では、入力された話速情報が認識に基づき、その話速に応じて音素長

50

が制御され、この場合、固定倍の音素長が設定され（ステップS304）、高速読みかつ先頭音素（ $n = 1$ ）であるか否かが判定される（ステップS305）。即ち、この判定処理では、ポーズ直後（話頭）の音素が持つ音素長が調整対象として特定されることになる。

#### 【0081】

話速が高速読みかつ先頭音素（ $n = 1$ ）であれば（ステップS305のYes）、その音素長が所定倍として例えば、1.5〔倍〕に設定、即ち、調整され（ステップS306）、また、話速が高速読みかつ先頭音素（ $n = 1$ ）でなければ（ステップS305のNo）、その音素長を調整しない。

#### 【0082】

このような処理の後、高速読みかつ音素が母音であるか否かが判定され（ステップS307）、話速が高速読みかつ母音であれば（ステップS307のYes）、その音素長が所定倍として例えば、0.9〔倍〕に設定、即ち、調整され（ステップS308）、また、母音でなければ（ステップS307のNo）、その音素長を調整しない。

#### 【0083】

そして、このような調整又は無調整の後、音素番号 $n$ の更新（ $n = n + 1$ ）を行い（ステップS309）、呼気段落内の音素は終了したか否かが判定され（ステップS310）、呼気段落内の全ての音素についての処理が実行された後、呼気段落の終端のポーズに到達した場合には、そのポーズ長を話速に応じて固定倍とし（ステップS311）、終了判定が行われる（ステップS312）。全てのデータの処理が完了するまで、ステップS303からステップS312の処理が繰り返される。この終了判定の後、音声合成が実行され（ステップS313）、音声が出力される。

#### 【0084】

このように、呼気段落単位での先頭音素及び母音が話速に応じて修正され、ポーズ直後の音素の音素長を既述のように、その一例として1.5〔倍〕にするのに対し、母音の音素長を既述のように短縮、その一例として0.9〔倍〕に調整することにより、音素長の伸長時間が母音の音素長の短縮によって補完されるので、音声出力の全再生時間の伸長を来すことなく、概ね全体の長さを同一に保ちつつ、合成音声の聞き取り易さが高められ、音声に変換された読み上げ文の認識性が改善される。

#### 【0085】

〔第4の実施の形態〕

#### 【0086】

次に、第4の実施の形態について、図9及び図10を参照する。図9は、第4の実施の形態に係る音素長制御部を示すブロック図、図10は、第4の実施の形態に係る音素長制御の処理手順の一例を示すフローチャートである。図9において、図2と同一部分には同一符号を付してある。

#### 【0087】

この処理手順は、音声読み上げのプログラム又は方法の一例であって、既述の音声読み上げ装置2（図1）及び音素長制御部18（図2）を用いて実行されるが、この第4の実施の形態では、第1の実施の形態の音素長の調整に加え、即ち、その音素長の伸長に対し、話頭音素の音素長の伸長分を呼気段落内の音素に比例配分して短縮することにより、呼気段落の長さを保ちつつ、即ち、読み上げ文の音声変換時間を伸長させることなく、聞き取り易さを高める。

#### 【0088】

この第4の実施の形態では、音声読み上げ装置2（図1）の音素長制御部18（図2）に関し、図9に示すように、呼気段落長演算部30が設置され、この呼気段落長演算部30は、音素長調整部24の出力から呼気段落の全体の長さを演算する。その演算結果は制御情報として音素長調整部24に加えられ、音素長調整部24は、特定の音素の音素長、この場合、先頭音素の音素長の伸長分を呼気段落内の全音素に比例配分して全音素の各音素長を短縮し、呼気段落の読み上げ時間の長さが所定の長さになるように制御する機能を

10

20

30

40

50

備えている。

#### 【0089】

そこで、この処理手順では、図10に示すように、言語処理（ステップS401）、音素長設定処理（ステップS402）、呼気段落内の音素の処理として、音素番号nの初期化（ $n=1$ ）（ステップS403）、話速に応じた音素長の制御を行う（ステップS404～S412）。この音素長の制御が呼気段落を単位とするのは第1の実施の形態と同様である。

#### 【0090】

音素長制御部18では、入力された話速情報が認識に基づき、その話速に応じて音素長が制御され、この場合、固定倍の音素長が設定され（ステップS404）、高速読みかつ先頭音素（ $n=1$ ）であるか否かが判定される（ステップS405）。即ち、この判定処理では、ポーズ直後（話頭）の音素が持つ音素長が調整対象として特定されることになる。

10

#### 【0091】

話速が高速読みかつ先頭音素（ $n=1$ ）であれば（ステップS405のYes）、その音素長が所定倍として例えば、1.5〔倍〕に設定、即ち、調整され（ステップS406）、また、話速が高速読みかつ先頭音素（ $n=1$ ）でなければ（ステップS405のNo）、その音素長を調整しない。

#### 【0092】

このような調整又は無調整の後、音素番号nの更新（ $n=n+1$ ）を行い（ステップS407）、呼気段落内の音素は終了したか否かが判定され（ステップS408）、呼気段落内の全ての音素についての処理が実行された後、呼気段落の終端のポーズに到達した場合には、そのポーズ長を話速に応じて固定倍とする（ステップS409）。

20

#### 【0093】

この設定の後、呼気段落全体の長さを計算し（ステップS410）、呼気段落の長さが所定の長さ例えば、音素長を伸長しない場合の長さと同様又は同等程度の長さになるように全音素の音素長を比例配分して調整し（ステップS411）、終了判定が行われる（ステップS412）。全てのデータの処理が完了するまで、ステップS403からステップS412の処理が繰り返される。この終了判定の後、音声合成が実行され（ステップS413）、音声が出力される。

30

#### 【0094】

このように、呼気段落単位での先頭音素が話速に応じて修正され、ポーズ直後の音素の音素長を既述のように、その一例として1.5〔倍〕にするのに対し、話頭音素の音素長の伸長分を呼気段落内の音素に比例配分して短縮することにより、呼気段落の長さが保たれるとともに、合成音声の聞き取り易さが高められ、音声に変換された読み上げ文の認識性が改善される。

#### 【0095】

〔第5の実施の形態〕

#### 【0096】

次に、第5の実施の形態について、図11及び図12を参照する。図11は、第5の実施の形態に係る音素長制御部を示すブロック図、図12は、第5の実施の形態に係る音素長制御の処理手順の一例を示すフローチャートである。図11において、図2と同一部分には同一符号を付してある。

40

#### 【0097】

この処理手順は、音声読み上げのプログラム又は方法の一例であって、既述の音声読み上げ装置2（図1）及び音素長制御部18（図2）を用いて実行されるが、この第5の実施の形態では、第1の実施の形態の音素長の調整に加え、即ち、その音素長の伸長に対し、話頭音素の音素長の伸長分を文章全体の音素に比例配分して短くすることで、文章全体の長さを保ちつつ、即ち、読み上げ文の音声変換時間を伸長させることなく、聞き取り易さを高めている。

50

## 【0098】

この第5の実施の形態では、音声読み上げ装置2（図1）の音素長制御部18（図2）に関し、図11に示すように、文章全体長演算部32が設置され、この文章全体長演算部32は、音素長調整部24の出力から文章全体の長さを演算する。その演算結果は制御情報として音素長調整部24に加えられ、音素長調整部24は、特定の音素の音素長、この場合、先頭音素の音素長の伸長分を文章全体の全音素に比例配分して全音素の各音素長を短縮し、文章の読み上げ時間の長さが所定の長さになるように制御する機能を備えている。

## 【0099】

そこで、この処理手順では、図12に示すように、言語処理（ステップS501）、音素長設定処理（ステップS502）、呼気段落内の音素の処理として、音素番号nの初期化（ $n=1$ ）（ステップS503）、話速に応じた音素長の制御を行う（ステップS503～S512）。この音素長の制御が呼気段落を単位とするのは第1の実施の形態と同様である。

## 【0100】

音素長制御部18では、入力された話速情報が認識に基づき、その話速に応じて音素長が制御され、この場合、固定倍の音素長が設定され（ステップS504）、高速読みかつ先頭音素（ $n=1$ ）であるか否かが判定される（ステップS505）。即ち、この判定処理では、ポーズ直後（話頭）の音素が持つ音素長が調整対象として特定されることになる。

## 【0101】

話速が高速読みかつ先頭音素（ $n=1$ ）であれば（ステップS505のYes）、その音素長が所定倍として例えば、1.5〔倍〕に設定、即ち、調整され（ステップS506）、また、話速が高速読みかつ先頭音素（ $n=1$ ）でなければ（ステップS505のNo）、その音素長を調整しない。

## 【0102】

このような調整又は無調整の後、音素番号nの更新（ $n=n+1$ ）を行い（ステップS507）、呼気段落内の音素は終了したか否かが判定され（ステップS508）、呼気段落内の全ての音素についての処理が実行された後、呼気段落の終端のポーズに到達した場合には、そのポーズ長を話速に応じて固定倍とし（ステップS509）、終了判定が行われる（ステップS510）。全てのデータの処理が完了するまで、ステップS503からステップS510の処理が繰り返される。

## 【0103】

全データの処理が終了した後、文章全体の長さを計算し（ステップS511）、文章全体の長さ、即ち、読み上げ時間が所定の長さ例えば、音素長を伸長しない場合の長さと同等又は同等程度の長さになるように文章全体の全音素の音素長を比例配分して調整し（ステップS512）、この処理の終了後、音声合成が実行され（ステップS513）、音声が出力される。

## 【0104】

このように、呼気段落単位での先頭音素が話速に応じて修正され、ポーズ直後の音素の音素長を既述のように、その一例として1.5〔倍〕にするのに対し、話頭音素の音素長の伸長分を文章全体の全音素に比例配分して短縮することにより、全文章の読み上げの長さが保たれるとともに、合成音声の聞き取り易さが高められ、音声に変換された読み上げ文の認識性が改善される。

## 【0105】

〔第6の実施の形態〕

## 【0106】

次に、第6の実施の形態について、図13を参照する。図13は、第6の実施の形態に係る音素長制御の処理手順の一例を示すフローチャートである。

## 【0107】

この処理手順は、音声読み上げのプログラム又は方法の一例であって、既述の音声読み上げ装置 2（図 1）及び音素長制御部 18（図 2）を用いて実行されるが、この第 6 の実施の形態では、第 2 の実施の形態（図 7）の音素長の調整と、第 3 の実施の形態（図 8）の音素長の調整とを併用することにより、その話頭の音素、摩擦音の音素長の伸長に対し、その他の音素の音素長として例えば、母音の音素長を短縮しており、読み上げ文の音声変換時間を伸長させることなく、聞き取り易さを高めている。

#### 【0108】

そこで、この処理手順では、図 13 に示すように、言語処理（ステップ S 6 0 1）、音素長設定処理（ステップ S 6 0 2）、呼気段落内の音素の処理として、音素番号  $n$  の初期化（ $n = 1$ ）（ステップ S 6 0 3）、話速に応じた音素長の制御を行う（ステップ S 6 0 3 ~ S 6 1 6）。この音素長の制御が呼気段落を単位とするのは第 2 の実施の形態（図 7）と同様である。

10

#### 【0109】

この第 6 の実施の形態においても、話速に応じた固定倍の音素長の設定（ステップ S 6 0 4）、高速読みかつ先頭音素（ $n = 1$ ）であるか否かの判定を行い（ステップ S 6 0 5）、話速が高速読みかつ先頭音素（ $n = 1$ ）であれば（ステップ S 6 0 5 の Yes）、音素が摩擦音か否かの判定（ステップ S 6 0 6）、話速が高速読みかつ先頭音素（ $n = 1$ ）かつ摩擦音であれば（ステップ S 6 0 6 の Yes）、その音素長が所定倍即ち、 $\alpha$  倍として例えば、 $\alpha = 1.7$  [倍] に設定ないし調整（ステップ S 6 0 7）、また、先頭音素（ $n = 1$ ）でも摩擦音でもなければ（ステップ S 6 0 8 の No）、その音素長の調整はない。即ち、この場合、ステップ S 6 0 4 で固定倍とされたそのままの状態が維持される。

20

#### 【0110】

高速読みかつ先頭音素の場合には（ステップ S 6 0 6 の No）、その音素長が所定倍即ち、 $\beta$  倍として例えば、 $\beta = 1.5$  [倍] に設定ないし調整され（ステップ S 6 0 9）、高速読みかつ摩擦音の場合には（ステップ S 6 0 8 の Yes）、その音素長が所定倍即ち、 $\gamma$  倍として例えば、 $\gamma = 1.4$  [倍] に設定ないし調整される（ステップ S 6 1 0）。

#### 【0111】

従って、高速読みかつ先頭音素かつ摩擦音の場合、高速読みかつ先頭音素の場合、高速読みかつ摩擦音の場合、先頭音素でも摩擦音でもない場合について、音素長の調整又は無調整は既述の表 1 の通りとなる。

30

#### 【0112】

このような処理の後、高速読みかつ音素が母音であるか否かが判定される（ステップ S 6 1 1）、話速が高速読みかつ母音であれば（ステップ S 6 1 1 の Yes）、その音素長が所定倍として例えば、 $0.9$  [倍] に設定、即ち、調整され（ステップ S 6 1 2）、また、母音でなければ（ステップ S 6 1 1 の No）、その音素長を調整しない。

#### 【0113】

そして、既述の通り、音素番号  $n$  の更新（ $n = n + 1$ ）（ステップ S 6 1 3）、呼気段落内の音素の終了判定（ステップ S 6 1 4）、呼気段落の終端のポーズに到達した場合のポーズ長を話速に応じた固定倍の設定（ステップ S 6 1 5）、終了判定（ステップ S 6 1 6）、音声合成（ステップ S 6 1 7）が実行される。

40

#### 【0114】

このように、呼気段落単位での先頭音素及び摩擦音が話速に応じて修正され、ポーズ直後の音素の音素長や摩擦音の音素長の場合、又は何れでもない場合には、既述のように、音素長の伸長を異ならせた設定とし、母音の場合にはその音素長を既述のように短縮することにより、ポーズ後の音素及び摩擦音の音素長の伸長時間が母音の音素長の短縮分だけ補完されるので、音声出力の全再生時間の伸長を来すことなく、概ね全体の長さを同一に保ちつつ、合成音声の聞き取り易さが高められ、音声に変換された読み上げ文の認識性が改善される。

#### 【0115】

50

〔第 7 の実施の形態〕

【0116】

次に、第 7 の実施の形態について、図 14 を参照する。図 14 は、第 7 の実施の形態に係る音素長制御の処理手順の一例を示すフローチャートである。

【0117】

この処理手順は、音声読み上げのプログラム又は方法の一例であって、既述の音声読み上げ装置 2（図 1）及び音素長制御部 18（図 2）を用いて実行されるが、この実施の形態では、第 2 の実施の形態（図 7）の音素長の調整に加え、即ち、その話頭の音素、摩擦音の音素長の伸長に対し、当該音素長を長めに確保した音素長分、ポーズ等、他の音素長を確保しない又は短縮する構成として、話頭及び摩擦音の音素長の伸長分を呼気段落内の音素に比例配分して短縮することにより、呼気段落の長さを保ちつつ、即ち、読み上げ文の音声変換時間を伸長させることなく、聞き取り易さを高めている。

10

【0118】

この第 7 の実施の形態では、第 4 の実施の形態と同様に（図 9）、音素長制御部 18 の音素長調整部 24 に呼気段落長演算部 30 を設置して音素長調整部 24 の出力から呼気段落の全体の長さを演算し、その演算結果は制御情報として音素長調整部 24 に加える。音素長調整部 24 は、特定の音素の音素長、この場合、先頭音素及び摩擦音の音素長の伸長分を呼気段落内の全音素に比例配分して全音素の各音素長を短縮し、呼気段落の読み上げ時間の長さが所定の長さになるように制御する機能を備えている。

20

【0119】

そこで、この処理手順では、図 14 に示すように、言語処理（ステップ S701）、音素長設定処理（ステップ S702）、呼気段落内の音素の処理として、音素番号  $n$  の初期化（ $n=1$ ）（ステップ S703）、話速に応じた音素長の制御を行う（ステップ S703～S716）。この音素長の制御が呼気段落を単位とするのは第 2 の実施の形態（図 7）と同様である。

【0120】

この第 7 の実施の形態においても、話速に応じた固定倍の音素長の設定（ステップ S704）、高速読みかつ先頭音素（ $n=1$ ）であるか否かの判定を行い（ステップ S705）、話速が高速読みかつ先頭音素（ $n=1$ ）であれば（ステップ S705 の Yes）、音素が摩擦音か否かの判定（ステップ S706）、話速が高速読みかつ先頭音素（ $n=1$ ）かつ摩擦音であれば（ステップ S706 の Yes）、その音素長が所定倍即ち、 $\alpha$  倍として例えば、 $\alpha=1.7$ 〔倍〕に設定ないし調整（ステップ S707）、また、先頭音素（ $n=1$ ）でも摩擦音でもなければ（ステップ S708 の No）、その音素長の調整はない。即ち、この場合、ステップ S704 で固定倍とされたそのままの状態が維持される。

30

【0121】

高速読みかつ先頭音素の場合には（ステップ S706 の No）、その音素長が所定倍即ち、 $\beta$  倍として例えば、 $\beta=1.5$ 〔倍〕に設定ないし調整され（ステップ S709）、高速読みかつ摩擦音の場合には（ステップ S708 の Yes）、その音素長が所定倍即ち、 $\gamma$  倍として例えば、 $\gamma=1.4$ 〔倍〕に設定ないし調整される（ステップ S710）。

40

【0122】

従って、高速読みかつ先頭音素かつ摩擦音の場合、高速読みかつ先頭音素の場合、高速読みかつ摩擦音の場合、先頭音素でも摩擦音でもない場合について、音素長の調整又は無調整は既述の表 1 の通りとなる。

【0123】

このような処理の後、音素番号  $n$  の更新（ $n=n+1$ ）（ステップ S711）、呼気段落内の音素の終了判定（ステップ S712）、呼気段落の終端のポーズに到達した場合のポーズ長を話速に応じた固定倍の設定（ステップ S713）の後、呼気段落全体の長さを計算し（ステップ S714）、呼気段落の長さが所定の長さ例えば、音素長を伸長しない場合の長さと同様又は同等程度の長さになるように全音素の音素長を比例配分して調整し

50

(ステップ S 7 1 5)、終了判定が行われる(ステップ S 7 1 6)。全てのデータの処理が完了するまで、ステップ S 7 0 3 からステップ S 7 1 6 の処理が繰り返される。この終了判定の後、音声合成が実行され(ステップ S 7 1 7)、音声が出力される。

#### 【0124】

このように、呼気段落単位での先頭音素及び摩擦音が話速に応じて修正され、ポーズ直後の音素の音素長や摩擦音の音素長の場合、又は何れでもない場合には、既述のように、音素長の伸長を異ならせた設定とし、これら音素の音素長の伸長分を呼気段落内の音素に比例配分して短縮することにより、呼気段落の長さが保たれるとともに、合成音声の聞き取り易さが高められ、音声に変換された読み上げ文の認識性が改善される。

#### 【0125】

10

〔第 8 の実施の形態〕

#### 【0126】

次に、第 8 の実施の形態について、図 1 5 を参照する。図 1 5 は、第 8 の実施の形態に係る音素長制御の処理手順の一例を示すフローチャートである。

#### 【0127】

この処理手順は、音声読み上げのプログラム又は方法の一例であって、既述の音声読み上げ装置 2 (図 1) を用いて実行されるが、この実施の形態では、第 2 の実施の形態(図 7)の音素長の調整に加え、即ち、その話頭の音素、摩擦音の音素の音素長の伸長に対し、その伸長分を文章全体の音素に比例配分で短くすることで、文章全体の長さを保ちつつ、即ち、読み上げ文の音声変換時間を伸長させることなく、聞き取り易さを高めている。

20

#### 【0128】

この第 8 の実施の形態では、第 5 の実施の形態と同様に(図 1 1)、音声読み上げ装置 2 (図 1)の音素長制御部 1 8 に文章全体長演算部 3 2 が併設され、この文章全体長演算部 3 2 は、音素長調整部 2 4 の出力から文章全体の長さを演算し、その演算結果が制御情報として音素長調整部 2 4 に加えられる。音素長調整部 2 4 は、特定の音素の音素長、この場合、先頭音素及び摩擦音の音素長の伸長分を文章全体の全音素に比例配分して全音素の各音素長を短縮し、文章の読み上げ時間の長さが所定の長さになるように制御する機能を備えている。

#### 【0129】

そこで、この処理手順では、図 1 5 に示すように、言語処理(ステップ S 8 0 1)、音素長設定処理(ステップ S 8 0 2)、呼気段落内の音素の処理として、音素番号  $n$  の初期化( $n = 1$ )(ステップ S 8 0 3)、話速に応じた音素長の制御を行う(ステップ S 8 0 3 ~ S 8 1 6)。この音素長の制御が呼気段落を単位とするのは第 2 の実施の形態(図 7)と同様である。

30

#### 【0130】

この第 8 の実施の形態においても、話速に応じた固定倍の音素長の設定(ステップ S 8 0 4)、高速読みかつ先頭音素( $n = 1$ )であるか否かの判定を行い(ステップ S 8 0 5)、話速が高速読みかつ先頭音素( $n = 1$ )であれば(ステップ S 8 0 5 の Yes)、音素が摩擦音か否かの判定(ステップ S 8 0 6)、話速が高速読みかつ先頭音素( $n = 1$ )かつ摩擦音であれば(ステップ S 8 0 6 の Yes)、その音素長が所定倍即ち、 $\alpha$  倍即ち、として例えば、 $\alpha = 1.7$  [倍] に設定(ステップ S 8 0 7)、また、先頭音素( $n = 1$ )でも摩擦音でもなければ(ステップ S 8 0 8 の No)、その音素長の調整はない。即ち、この場合、ステップ S 8 0 4 で固定倍とされたそのままの状態が維持される。

40

#### 【0131】

高速読みかつ先頭音素の場合には(ステップ S 8 0 6 の No)、その音素長が所定倍即ち、 $\beta$  倍として例えば、 $\beta = 1.5$  [倍] に設定され(ステップ S 8 0 9)、高速読みかつ摩擦音の場合には(ステップ S 8 0 8 の Yes)、その音素長が所定倍即ち、 $\gamma$  倍として例えば、 $\gamma = 1.4$  [倍] に設定される(ステップ S 8 1 0)。

#### 【0132】

50

従って、高速読みかつ先頭音素かつ摩擦音の場合、高速読みかつ先頭音素の場合、高速読みかつ摩擦音の場合、先頭音素でも摩擦音でもない場合について、音素長の調整又は無調整は既述の表 1 の通りとなる。

#### 【0133】

このような処理の後、音素番号  $n$  の更新 ( $n = n + 1$ ) (ステップ S 8 1 1)、呼気段落内の音素の終了判定 (ステップ S 8 1 2)、呼気段落の終端のポーズに到達した場合のポーズ長を話速に応じた固定倍の設定 (ステップ S 8 1 3)、終了判定 (ステップ S 8 1 4) を行う。

#### 【0134】

全データの処理が終了した後、文章全体の長さを計算し (ステップ S 8 1 5)、文章全体の長さ、即ち、読み上げ時間が所定の長さ例えば、音素長を伸長しない場合の長さと同等又は同等程度の長さになるように文章全体の全音素の音素長を比例配分して調整し (ステップ S 8 1 6)、この処理の終了後、音声合成が実行され (ステップ S 8 1 7)、音声

10

#### 【0135】

このように、呼気段落単位での先頭音素及び摩擦音が話速に応じて修正され、ポーズ直後の音素の音素長や摩擦音の音素長の場合、又は何れでもない場合には、既述のように、音素長の伸長を異ならせた設定とし、これら音素長の伸長分を全文章の全音素に比例配分して短縮することにより、全文章の読み上げの長さが保たれるとともに、合成音声の聞き取り易さが高められ、音声に変換された読み上げ文の認識性が改善される。

20

#### 【0136】

〔第 9 の実施の形態〕

#### 【0137】

次に、第 9 の実施の形態について、図 1 6 を参照する。図 1 6 は、第 9 の実施の形態に係る音素長制御の処理手順の一例を示すフローチャートである。

#### 【0138】

この処理手順は、音声読み上げのプログラム又は方法の一例であって、既述の音声読み上げ装置 2 (図 1) 及び音素長制御部 1 8 (図 2) を用いて実行されるが、この実施の形態では、高速読みの場合にポーズの長さを短縮することにより、聞き取り易さは同等であって、読み上げ時間の長さを短縮したものである。話速を例えば、3 倍速とすれば、話速の反比例によってポーズ長をその半分にすれば、標準話速に対してのポーズ長は 6 分の 1 となり、ポーズ長の短縮によって読み上げ時間の長さを短くできる。

30

#### 【0139】

そこで、この処理手順では、図 1 6 に示すように、言語処理 (ステップ S 9 0 1)、音素長設定処理 (ステップ S 9 0 2)、呼気段落内の音素の処理として、音素番号  $n$  の初期化 ( $n = 1$ ) (ステップ S 9 0 3)、話速に応じた音素長の制御を行う (ステップ S 9 0 3 ~ S 9 1 0)。この音素長の制御が呼気段落を単位とするのは第 1 の実施の形態 (図 5) と同様である。

#### 【0140】

この第 9 の実施の形態においても、話速に応じた固定倍の音素長の設定 (ステップ S 9 0 4)、音素番号  $n$  の更新 ( $n = n + 1$ ) (ステップ S 9 0 5)、呼気段落内の音素の終了判定 (ステップ S 9 0 6) を行う。

40

#### 【0141】

この場合、高速読みか否かの判定を行い (ステップ S 9 0 7)、高速読みであれば (ステップ S 9 0 7 の Yes)、終端のポーズ長を話速に応じた固定倍に対する所定の倍率としての例えば、2 分の 1 に設定する (ステップ S 9 0 8)。

#### 【0142】

高速読みでなければ (ステップ S 9 0 7 の No)、呼気段落の終端のポーズに到達した場合のポーズ長を話速に応じた固定倍に設定し (ステップ S 9 0 9)、全データが終了したか否かの終了判定を行い (ステップ S 9 1 0)、全データの終了後、音声合成が実行さ

50

れ（ステップ S 9 1 1）、音声が出力される。

【0 1 4 3】

このように、高速読みの場合に、呼気段落の終端のポーズ長を短縮することにより、全文章の読み上げの長さが保たれるとともに、合成音声の聞き取り易さが高められ、音声に変換された読み上げ文の認識性が改善される。

【0 1 4 4】

〔第 1 0 の実施の形態〕

【0 1 4 5】

次に、第 1 0 の実施の形態について、図 1 7 及び図 1 8 を参照する。図 1 7 は、第 1 0 の実施の形態に係る音声読み上げ装置のパラメータ生成部の構成例を示すブロック図、図 1 8 は、第 1 0 の実施の形態に係る音素長制御の処理手順の一例を示すフローチャートである。図 1 7 において、図 1 と同一部分には同一符号を付してある。

10

【0 1 4 6】

この第 1 0 の実施の形態では、パラメータ生成部 8 に区切り変更部 3 4 が音素長設定部 1 4 の前段に設置されており、この区切り変更部 3 4 では、言語処理部 4（図 1）で生成された表音文字列の呼気段落の区切りのポーズ長が変更される構成である。このような区切り変更部 3 4 を備えたことにより、各音素長を確保した状況で、読み上げを行う文章の全体の再生時間を短縮することができる。

【0 1 4 7】

この場合、言語処理結果の表音文字列を、

20

「ヤマナシ' ケンノ コーコーオ ソツギョーシテ、シンヨーキ' シンコニ ハネイッテ・ヨネンメ' デス。」

とすれば、この表音文字列に対し、区切り変更部 3 4 では、呼気段落区切りの長さの段階を 1 段階短くする。具体的には、ポーズ長が小の midpoint 「・」はアクセント区切りの空白（ポーズはない）にし、ポーズ長が中の読点「、」はポーズ長が小の midpoint 「・」にし、ポーズ長が大の句点「。」はポーズ長が中の読点「、」にする。

【0 1 4 8】

即ち、表音文字列は、

「ヤマナシ' ケンノ コーコーオ ソツギョーシテ・シンヨーキ' シンコニ ハネイッテ  
ヨネンメ' デス、」

30

に変更され、読み上げ文の全体の再生時間を削減できる。

【0 1 4 9】

そこで、この処理手順では、図 1 8 に示すように、言語処理（ステップ S 1 0 0 1）、音素長設定処理（ステップ S 1 0 0 2）、呼気段落内の音素の処理として、音素番号 n の初期化（n = 1）（ステップ S 1 0 0 3）、話速に応じた音素長の制御を行う（ステップ S 1 0 0 3 ~ S 1 0 1 4）。この音素長の制御が呼気段落を単位とするのは第 1 の実施の形態（図 6）と同様である。

【0 1 5 0】

この第 1 0 の実施の形態では、話速に応じた固定倍の音素長を設定し（ステップ S 1 0 0 4）、この音素長の設定の後、文字が句点「。」であるか否かを判定し（ステップ S 1 0 0 5）、句点「。」であれば、当該文字を読点「、」に置き換え（ステップ S 1 0 0 6）、ステップ S 1 0 1 1 に移行する。

40

【0 1 5 1】

また、文字が句点「。」でなければ（ステップ S 1 0 0 5 の No）、文字が読点「、」であるか否かを判定し（ステップ S 1 0 0 7）、読点「、」であれば、当該文字を midpoint 「・」に置き換え（ステップ S 1 0 0 8）、ステップ S 1 0 1 1 に移行する。

【0 1 5 2】

また、読点「、」でなければ（ステップ S 1 0 0 7 の No）、文字が midpoint 「・」であるか否かを判定し（ステップ S 1 0 0 9）、midpoint 「・」であれば、当該文字を空白「 」に置き換え（ステップ S 1 0 1 0）、ステップ S 1 0 1 1 に移行する。

50

## 【0153】

このような処理の後、ステップS1011では、音素音素番号nの更新( $n = n + 1$ )を行い、呼気段落内の音素の終了判定(ステップS1012)、呼気段落の終端のポーズに到達した場合のポーズ長を話速に応じた固定倍の設定(ステップS1013)を行い、終了判定(ステップS1014)を行う。この処理の終了後、音声合成が実行され(ステップS1015)、音声が出力される。

## 【0154】

この処理手順によれば、呼気段落区切りを示す文字を置換することで、区切り長さの段階を1段階だけ短くする。具体的には、

- ・ ポーズ長が小の中間点「・」(例えば、標準話速で0.1秒)はアクセント区切りの空白(ポーズはない)に、

- ・ ポーズ長が中の読点「、」(例えば、標準話速で0.3秒)はポーズ長小の中間点「・」に、

- ・ ポーズ長が大の句点「。」(例えば、標準話速で0.8秒)はポーズ長中の読点「、」にする。即ち、表音文字列は、

「ヤマナシ' ケンノ コーコーオ ソツギョーシテ・シンヨーキ' ンコニ ハネイッテ ヨネンメ' デス、」

となり、このような変更により、全体の再生時間を削減できる。

## 【0155】

従って、呼気段落単位での各音素長を確保した状況で、読み上げを行う文章の全体の再生時間を短縮することができる。

## 【0156】

〔その他の実施の形態〕

## 【0157】

(1) 音素長制御部18に入力する話速情報について、図19を参照する。図19は、話速調整部を備えるパラメータ生成部を示すブロック図である。上記実施の形態では、音素長制御部18に話速情報を入力しているが、図19に示すように、外部から話速を調整可能な話速調整部22をパラメータ生成部8に設置し、外部から任意の話速設定を可能な構成としてもよい。

## 【0158】

(2) 上記実施の形態では、ポーズ直後の音素長を伸長する場合について説明したが、本発明は、短縮する場合にも適用することができる。

## 【0159】

(3) 第1の実施の形態に携帯端末装置200(図3、図4)を例示したが、本発明は、携帯情報端末機(PDA: Personal Digital Assistant)や、パーソナルコンピュータ等、コンピュータを搭載して音声を出力する電子機器や、電子機器ユニットを搭載する各種機器に適用できるものであり、本発明は、上記実施の形態に限定されるものではない。

## 【0160】

(4) 上記実施の形態において、読み上げ速度が高速である場合、文字データ中の一部又は全部のポーズを削除する構成としてもよく、ポーズの削除により、聞き取り易さを損なうことなく、再生時間の短縮を図ることができる。

## 【0161】

(5) 読み上げ速度が低速の場合にはポーズ直後の音素が持つ音素長を短縮させ又は基準速度と同一の音素長にする構成としてもよい。

## 【0162】

(6) 上記第6の実施の形態(図13)では、読み上げ速度が高速である場合、先頭音素の音素長や摩擦音の伸長に対し、他の音素として母音の音素長を短縮しているが、特定のポーズや音素長の伸長に対応し、他の音素長を短縮する構成としてもよく、斯かる構成とすれば、同様に、読み上げ時間の伸長を抑制できる。

## 【0163】

10

20

30

40

50

(7) 上記第10の実施の形態(図18)では、呼気段落を単位とする処理を示しているが、呼気段落以外の文章単位でもよく、特定の文章の段落内処理であってもよい。

#### 【0164】

(8) 上記第2、第6、第7、第8の実施の形態では、特定の音素として摩擦音を例示し、その音素長を伸長させているが、摩擦音の伸長を省略してもよいし、摩擦音に代え、他の音素長であってもよい。

#### 【実施例】

#### 【0165】

#### 〔実施例1〕

#### 【0166】

実施例1について、図20及び図21を参照する。図20は、図6のフローチャートに対応する比較例であるフローチャート、図21は、言語処理結果を示す図である。

#### 【0167】

この音声読み上げ装置2(図1)では、話速に応じて各音素の音素長を同様に伸長する場合には、図20に示すフローチャートの処理となる。この場合、図6に示すフローチャートと同一ステップには同一符号を付し、ポーズ後の話頭の音素長を調整しない場合の処理を示している。即ち、図20のフローチャートは、図6のフローチャートからステップS105、S106の処理がない場合であって、高速読み上げで先頭音素についての音素長を伸長しない場合の処理であり、高速読み上げに反比例して音素長を固定倍している。

#### 【0168】

このような処理において、入力テキストの文例が例えば、

「山梨県の高校を卒業して、信用金庫に入って4年目です。」(図5)

であるとすれば、その単語解析結果は図21に示すように、入力テキスト、品詞、表音文字列で表すことができる。

#### 【0169】

この文例の「山梨県の高校を卒業して、信用金庫に入って4年目です。」において、「山梨」は名詞であり、その表音文字列は「ヤマナシ」となり、「県」は名詞であり、その表音文字列は「ケン」となり、「の」は助詞であり、その表音文字列は「ノ」となり、この「の」の後部分はアクセント句境界で空白となり、「高校」は名詞であり、その表音文字列は「コーコー」となり、「を」は助詞であり、その表音文字列は「オ」となり、その後の部分はアクセント句境界で空白となり、「卒業し」は動詞(連用形)であり、その表音文字列は「ソツギョーシ」となり、「て」は助詞であり、その表音文字列は「テ」となり、「、」は呼気段落境界(ポーズ長は中)であり、その表音文字列は「、」となり、「信用」は名詞であり、その表音文字列は「シンヨー」となり、「金庫」は名詞であり、その表音文字列は「キ'ンコ」となり、「に」は助詞であり、その表音文字列は「ニ」となり、その後の部分はアクセント句境界で空白となり、「入っ」は動詞(連用形、促音便)であり、その表音文字列は「ハ\*イッ」となり、「て」は助詞であり、その表音文字列は「テ」となり、その後の部分は呼気段落境界(ポーズ長は小)となり、その表音文字列は「・」となり、「4」は数詞であり、その表音文字列は「ヨ」となり、「年」は助数詞であり、その表音文字列は「ネン」となり、「目」は助数詞の後置詞であり、その表音文字列は「メ'」となり、「です」は助動詞であり、その表音文字列は「デス」となり、「。」は呼気段落境界(ポーズ長は大)であり、その表音文字列は「。」となる。従って、上記文例の表音文字列は、

「ヤマナシ'ケンノ コーコーオ ソツギョーシテ、シンヨーキ'ンコニ ハ\*イッテ・ヨネンメ'デス。」

となる。

#### 【0170】

この表音文字列中の「シンヨー」の部分の音素長作成と話速による音素長の修正について、図22を参照する。図22は、この場合の音素長の生成例を示す図である。

#### 【0171】

10

20

30

40

50

この例では、概ね毎秒 7 モーラを 1 倍速として、3 倍速（目安として毎秒 21 モーラ）を生成する場合では、1 倍速での音素長を音素長テーブル 16（図 1）から読み出し、話速に反比例して音素長を修正している。この修正の後、アクセント等の情報を基にピッチパターンが生成され、音声波形が合成される。

#### 【0172】

これに対し、第 1 の実施の形態（図 6）の処理結果について、図 23 を参照する。図 23 は、第 1 の実施の形態（図 6）の音素長生成例を示す図である。

#### 【0173】

この場合、3 倍速での音素長を生成する場合には、ポーズ後の話頭である「sh」の音素長が単純な反比例での長さの 1.5〔倍〕に設定される。この結果、図 23 に示すように、1 倍速での音素長が 117〔ms〕であるのに対し、3 倍速での音素長は 59〔ms〕となっている。これらの音素長を他の音素「I」、「N」、「y」、「O」、「O」と比較すると、1 倍速の音素「sh」の音素長 117〔ms〕は、他の音素「I」= 60〔ms〕、「N」= 60〔ms〕、「y」= 65〔ms〕、「O」= 80〔ms〕、「O」= 105〔ms〕であって、顕著な差異がないのに対し、3 倍速の音素「sh」の音素長 59〔ms〕は、他の音素「I」= 20〔ms〕、「N」= 20〔ms〕、「y」= 22〔ms〕、「O」= 27〔ms〕、「O」= 35〔ms〕であり、顕著な差異が生じている。この結果、聴感上の聞き取り易さを向上させることができ、認識性が高められる。

#### 【0174】

これらの処理結果である音声合成波形について、図 24 を参照する。図 24 において、A は、通常速度で「卒業して、信用金庫に」と読み上げた場合の音声合成波形であり、図 20 に示すフローチャートの処理を以て読み上げられた場合である。また、図 24 の B は、同じ読み上げ文章を、一律、高速読み上げした場合の波形であって、同様に、図 20 に示すフローチャートの処理を以て読み上げられた場合である。即ち、ポーズ直後の話頭の音素長を伸長しない場合である。また、図 24 の C は、高速読み上げにおいて、第 1 の実施の形態（図 6 に示すフローチャート）の処理を適用した場合であり、話頭の音素長が伸長された場合の音声合成波形である。図 24 の A、B、C において、図 24 の A の読み上げ時間を  $T_o$  とすれば、図 24 の B、C の読み上げ時間は、3 倍の話速を設定しているので、 $T_o/3$  に短縮されたものをほぼ同一の尺度で記載している。

#### 【0175】

図 24 の A の破線包囲部 a はポーズ直後の話頭の音素であり、図 24 の B の破線包囲部 b は同様の音素を示しているが、b の音素長が話速が 3 倍になっている分だけ縮んでいることが理解できよう。このような読み上げ音を聞いた場合、音切れのように感じられ、話頭が聞き取りづらくなることが確認された。これに対し、図 24 の C の破線包囲部 c は、3 倍の話速に対し、話頭の音素の音素長を伸長させているため、話速を高めて読み上げ音を聞いた場合にも、音切れが発生することがなく、聞き取り易さが高められる。

#### 【0176】

##### 〔実施例 2〕

#### 【0177】

実施例 2 の処理結果を示す波形について、図 25 及び図 26 を参照する。図 25 は比較例である音声合成波形を示す図、図 26 は、実施例 2 に係る音声合成波形を示す図である。図 25 において、A は、標準速度の場合の波形であり、B は、高速読み上げの場合の波形である。A の標準速度の読み上げに対し、B の高速読み上げの場合には、d のポーズ直後の音素長が話速比例通り短縮し、この例では、15〔msec〕に短縮される。

#### 【0178】

これに対し、図 26 において、A は、第 1 の実施の形態（図 6 のフローチャート）の処理の標準速度の場合の波形、B は、高速読み上げに対応してポーズ直後の話頭の音素長を伸長した場合の波形である。

#### 【0179】

図 25 の B の d と図 26 の B の e とを対比すれば、ポーズ直後の話頭の音素が持つ音素

長を話速比例よりも伸長（確保）する場合、即ち、この例（図 2 6 の B の e）では、3 5 [m s e c] に伸長するので、音素長は 2. 3 倍程度に伸長されるので、音切れが発生することがなく、聞き取り易さが高められる。

【0 1 8 0】

〔実施例 3〕

【0 1 8 1】

実施例 3 の処理結果を示す波形について、図 2 7 及び図 2 8 を参照する。図 2 7 は比較例である音声合成波形を示す図、図 2 8 は、実施例 3 に係る音声合成波形を示す図である。実施例 1、2 が日本語であるのに対し、実施例 3 は英語文「h a p p y, s h o c k, s h o o t」を読み上げた場合である。

10

【0 1 8 2】

図 2 7 において、A は、標準速度の場合の波形であり、B は、高速読み上げの場合の波形である。A の標準速度の読み上げに対し、B の高速読み上げの場合には、f、g のポーズ直後の音素長が話速比例通り短縮し、この例では、f の箇所では 1 9 [m s e c]、g の箇所では 1 4 [m s e c] に短縮される。

【0 1 8 3】

これに対し、図 2 8 において、A は、第 1 の実施の形態（図 6 のフローチャート）の処理の標準速度の場合の波形、B は、高速読み上げに対応してポーズ直後の話頭の音素長を伸長した場合の波形である。

【0 1 8 4】

20

図 2 7 の B の f、g と図 2 8 の B の h、i とを対比すれば、ポーズ直後の話頭の音素が持つ音素長を話速比例よりも伸長（確保）する場合、即ち、この例（図 2 8 の B の h、i）では、h が 2 7 [m s e c]、i が 2 5 [m s e c] に伸長するので、音素長は 2 倍程度に伸長されるので、音切れが発生することがなく、聞き取り易さが高められる。

【0 1 8 5】

〔実施例 4〕

【0 1 8 6】

実施例 4 の処理結果を示す波形について、図 2 9 及び図 3 0 を参照する。図 2 9 は比較例である音声合成波形を示す図、図 3 0 は、実施例 4 に係る音声合成波形を示す図である。図 2 9 において、A は、標準速度の場合の波形であり、B は、高速読み上げの場合の波形である。A の標準速度の読み上げの場合のポーズ区間 j は、B の高速読み上げの場合には、ポーズ区間 k に変化し、ポーズ区間の長さが話速に応じて短縮されている。

30

【0 1 8 7】

これに対し、図 3 0 において、A は、第 9 の実施の形態（図 1 6 のフローチャート）の処理の標準速度の場合の波形であり、l はその場合のポーズ区間、B は、高速読み上げに対応してポーズ長を話速による短縮よりも更に短縮した場合の波形であり、m がその場合のポーズ区間である。

【0 1 8 8】

図 2 9 の B のポーズ区間 k と図 3 0 の B のポーズ区間 m とを対比すれば、ポーズ区間が話速比例のポーズ区間よりも短縮され、音切れを発生させることなく、即ち、聞き取り易さを損なうことなく、読み上げ時間の短縮が図られる。

40

【0 1 8 9】

〔実施例 5〕

【0 1 9 0】

実施例 5 の処理結果を示す波形について、図 3 1 を参照する。実施例 1、2、4 が日本語であるのに対し、実施例 5 は実施例 3 と同様に、英語文「h a p p y s h o c k s h o o t」を読み上げた場合である。

【0 1 9 1】

図 3 1 において、A は、第 9 の実施の形態（図 1 6 のフローチャート）の処理の標準速度の場合の波形であり、n、o はその場合のポーズ区間、B は、高速読み上げに対応して

50

ポーズを話速による短縮よりも更に短縮した場合の波形であり、p、qがその場合のポーズ区間である。

【0192】

図31のAのポーズ区間n、oとBのポーズ区間p、qとを対比すれば、ポーズ区間が話速比例のポーズ区間よりも短縮され、音切れを発生させることなく、即ち、聞き取り易さを損なうことなく、読み上げ時間の短縮が図られる。

【0193】

次に、以上述べた本発明の実施の形態から抽出される技術的思想を請求項の記載形式に準じて付記として列挙する。本発明に係る技術的思想は上位概念から下位概念まで、様々なレベルやバリエーションにより把握できるものであり、以下の付記に本発明が限定されるものではない。

10

【0194】

(付記1) 文字データを音声に変換して読み上げる音声読み上げ装置であって、前記文字データから音素の種類を判定する音素判定部と、

音素に読み上げ速度に応じた音素長を設定し、音素が前記文字データのポーズ直後の音素である場合に前記音素判定部の判定結果に基づき、ポーズ直後の前記音素の音素長を調整する音素長調整部と、

を備えることを特徴とする音声読み上げ装置。

【0195】

(付記2) 付記1の音声読み上げ装置において、

20

音素の読み上げ速度を判定する速度判定部を備え、前記音素長調整部は、前記読み上げ速度の判定結果に基づき、前記読み上げ速度が高速の場合にはポーズ直後の音素が持つ音素長を伸長させることを特徴とする音声読み上げ装置。

【0196】

(付記3) 付記1の音声読み上げ装置において、

前記音素が摩擦音である場合に、前記音素長調整部は、前記音素判定部の判定結果に基づいて前記摩擦音の音素を伸長させることを特徴とする音声読み上げ装置。

【0197】

(付記4) 付記1の音声読み上げ装置において、

呼気段落の長さを演算する呼気段落演算部を備え、前記音素長調整部は、前記音素長の調整分を前記呼気段落演算部の演算結果に基づき、前記呼気段落の各音素長を比例配分して増減させることを特徴とする音声読み上げ装置。

30

【0198】

(付記5) 付記1の音声読み上げ装置において、

読み上げ文章の長さを演算する文章演算部を備え、前記音素長調整部は、前記音素長の調整分を前記文章演算部の演算結果に基づき、前記文章の各音素長を比例配分して増減させることを特徴とする音声読み上げ装置。

【0199】

(付記6) 付記1の音声読み上げ装置において、

前記音素長調整部は、読み上げ速度が高速の場合、前記文字データ中の一部又は全部のポーズが持つポーズ長を前記読み上げ速度に応じた長さより短縮することを特徴とする音声読み上げ装置。

40

【0200】

(付記7) 付記1の音声読み上げ装置において、

前記音素長調整部は、読み上げ速度が高速の場合、前記文字データ中の一部又は全部のポーズを削除することを特徴とする音声読み上げ装置。

【0201】

(付記8) 付記2の音声読み上げ装置において、

前記音素長調整部は、前記音素長の伸長に対応し、ポーズ長を含む他の音素長を短縮することを特徴とする音声読み上げ装置。他の音素長は、母音、子音、促音等の音素長であ

50

る。

【0202】

(付記9) 付記3の音声読み上げ装置において、

前記音素長調整部は、前記音素長の伸長に対応し、ポーズ長を含む他の音素長を短縮することを特徴とする音声読み上げ装置。他の音素長は、母音、子音、促音等の音素長である。

【0203】

(付記10) 文字データを音声に変換して読み上げる手順をコンピュータに実行させる音声読み上げプログラムであって、

前記文字データから音素の種類を判定する手順と、

音素に読み上げ速度に応じた音素長を設定する手順と、

音素が前記文字データのポーズ直後の音素である場合に前記判定の結果に基づき、ポーズ直後の前記音素の音素長を調整する手順と、

を前記コンピュータに実行させることを特徴とする音声読み上げプログラム。

10

【0204】

(付記11) 付記10の音声読み上げプログラムにおいて、

音素の読み上げ速度を判定する手順と、

前記読み上げ速度の判定結果に基づき、前記読み上げ速度が高速の場合にはポーズ直後の音素が持つ音素長を伸長させる手順と、

を含むことを特徴とする音声読み上げプログラム。

20

【0205】

(付記12) 付記10の音声読み上げプログラムにおいて、

前記音素が摩擦音であるか否かを判定する手順と、

前記判定の結果に基づき、前記摩擦音の音素を伸長させる手順と、

を含むことを特徴とする音声読み上げプログラム。

【0206】

(付記13) 付記10の音声読み上げプログラムにおいて、

呼気段落の長さを演算する手順と、

前記音素長の調整分を前記演算の結果に基づき、前記呼気段落の各音素長を比例配分して増減させる手順と、

を含むことを特徴とする音声読み上げプログラム。

30

【0207】

(付記14) 付記10の音声読み上げプログラムにおいて、

読み上げ文章の長さを演算する手順と、

前記音素長の調整分を前記演算の結果に基づき、前記文章の各音素長を比例配分して増減させる手順と、

を含むことを特徴とする音声読み上げプログラム。

【0208】

(付記15) 付記10の音声読み上げプログラムにおいて、

読み上げ速度が高速の場合、前記文字データ中の一部又は全部のポーズが持つポーズ長を前記読み上げ速度に応じた長さより短縮する手順と、

を含むことを特徴とする音声読み上げプログラム。

40

【0209】

(付記16) 付記10の音声読み上げプログラムにおいて、

読み上げ速度が高速の場合、前記文字データ中の一部又は全部のポーズを削除する手順と、

を含むことを特徴とする音声読み上げプログラム。

【0210】

(付記17) 付記11の音声読み上げプログラムにおいて、

前記音素長の伸長に対応し、ポーズ長を含む他の音素長を短縮する手順と、

50

を含むことを特徴とする音声読み上げプログラム。

【0211】

(付記18) 付記12の音声読み上げプログラムにおいて、  
前記音素長の伸長に対応し、ポーズ長を含む他の音素長を短縮する手順と、  
を含むことを特徴とする音声読み上げプログラム。

【0212】

(付記19) 文字データを音声に変換して読み上げる音声読み上げ方法であって、  
前記文字データから音素の種類を判定する工程と、  
音素に読み上げ速度に応じた音素長を設定する工程と、  
音素が前記文字データのポーズ直後の音素である場合に前記判定の結果に基づき、ポーズ直後の前記音素の音素長を調整する工程と、  
を備えることを特徴とする音声読み上げ方法。 10

【0213】

(付記20) 付記19の音声読み上げ方法において、  
音素の読み上げ速度を判定する工程と、  
前記読み上げ速度の判定結果に基づき、前記読み上げ速度が高速の場合にはポーズ直後の音素が持つ音素長を伸長させる工程と、  
を含むことを特徴とする音声読み上げ方法。

【0214】

(付記21) 付記19の音声読み上げ方法において、 20  
前記音素が摩擦音であるか否かを判定する工程と、  
前記判定の結果に基づき、前記摩擦音の音素を伸長させる工程と、  
を含むことを特徴とする音声読み上げ方法。

【0215】

(付記22) 付記19の音声読み上げ方法において、  
呼気段落の長さを演算する工程と、  
前記音素長の調整分を前記演算の結果に基づき、前記呼気段落の各音素長を比例配分して増減させる工程と、  
を含むことを特徴とする音声読み上げ方法。

【0216】

(付記23) 付記19の音声読み上げ方法において、 30  
読み上げ文章の長さを演算する工程と、  
前記音素長の調整分を前記演算の結果に基づき、前記文章の各音素長を比例配分して増減させる工程と、  
を含むことを特徴とする音声読み上げ方法。

【0217】

(付記24) 付記19の音声読み上げ方法において、  
読み上げ速度が高速の場合、前記文字データ中の一部又は全部のポーズが持つポーズ長を前記読み上げ速度に応じた長さより短縮する工程と、  
を含むことを特徴とする音声読み上げ方法。 40

【0218】

(付記25) 付記19の音声読み上げ方法において、  
読み上げ速度が高速の場合、前記文字データ中の一部又は全部のポーズを削除する工程と、  
を含むことを特徴とする音声読み上げ方法。

【0219】

(付記26) 付記20の音声読み上げ方法において、  
前記音素長の伸長に対応し、ポーズ長を含む他の音素長を短縮する工程と、  
を含むことを特徴とする音声読み上げ方法。

【0220】

(付記 27) 付記 21 の音声読み上げ方法において、  
前記音素長の伸長に対応し、ポーズ長を含む他の音素長を短縮する工程と、  
を含むことを特徴とする音声読み上げ方法。

【0221】

(付記 28) 文字データを音声に変換して読み上げる音声読み上げ装置であって、  
読み上げ速度を判定する読み上げ速度判定部と、  
前記読み上げ判定部の判定に基づき、読み上げ速度が高速の場合、前記文字データ中の  
一部又は全部のポーズが持つポーズ長を前記読み上げ速度に応じて調整する調整部と、  
を備えることを特徴とする音声読み上げ装置。

【0222】

(付記 29) 文字データを音声に変換して読み上げる音声読み上げ装置であって、  
読み上げ速度を判定する読み上げ速度判定部と、  
前記読み上げ判定部の判定に基づき、読み上げ速度が高速の場合、前記文字データ中の  
一部又は全部のポーズを削除する調整部と、  
を備えることを特徴とする音声読み上げ装置。

【0223】

(付記 30) 文字データを音声に変換して読み上げる音声読み上げ装置であって、  
文字データから音素の種類を判定する音素判定部と、  
前記音素判定部の判定に基づき、特定の音素の音素長を伸長させ、この音素長の伸長に  
対応し、ポーズ長を含む他の音素長を短縮する音素長調整部と、  
を備えることを特徴とする音声読み上げ装置。

【0224】

以上説明したように、本発明の最も好ましい実施の形態等について説明したが、本発明  
は、上記記載に限定されるものではなく、特許請求の範囲に記載され、又は明細書に開示  
された発明の要旨に基づき、当業者において様々な変形や変更が可能であることは勿論で  
あり、斯かる変形や変更が、本発明の範囲に含まれることは言うまでもない。

【産業上の利用可能性】

【0225】

本発明は、文字データを音声に変換して読み上げる装置、プログラム及び方法に関し、  
文字データからポーズの存在を認識し、ポーズ直後の音素又は他の音素の音素長やポーズ  
長を制御し、読み上げ速度を高速化しても、合成音声の聞き取り易さを高めることができ  
、認識性の向上が図られるので、音声合成等の処理に有用である。

【図面の簡単な説明】

【0226】

【図 1】第 1 の実施の形態に係る音声読み上げ装置の構成例を示すブロック図である。

【図 2】音声読み上げ装置の音素長制御部の構成例を示すブロック図である。

【図 3】音声読み上げ装置を搭載した携帯端末装置の一例を示すブロック図である。

【図 4】携帯端末装置の構成例を示す図である。

【図 5】画面表示例を示す図である。

【図 6】第 1 の実施の形態に係る音素長制御の処理手順の一例を示すフローチャートであ  
る。

【図 7】第 2 の実施の形態に係る音素長制御の処理手順の一例を示すフローチャートであ  
る。

【図 8】第 3 の実施の形態に係る音素長制御の処理手順の一例を示すフローチャートであ  
る。

【図 9】第 4 の実施の形態に係る音素長制御部を示すブロック図である。

【図 10】第 4 の実施の形態に係る音素長制御の処理手順の一例を示すフローチャートで  
ある。

【図 11】第 5 の実施の形態に係る音素長制御部を示すブロック図である。

【図 12】第 5 の実施の形態に係る音素長制御の処理手順の一例を示すフローチャートで

10

20

30

40

50

ある。

【図 1 3】第 6 の実施の形態に係る音素長制御の処理手順の一例を示すフローチャートである。

【図 1 4】第 7 の実施の形態に係る音素長制御の処理手順の一例を示すフローチャートである。

【図 1 5】第 8 の実施の形態に係る音素長制御の処理手順の一例を示すフローチャートである。

【図 1 6】第 9 の実施の形態に係る音素長制御の処理手順の一例を示すフローチャートである。

【図 1 7】第 1 0 の実施の形態に係る音声読み上げ装置のパラメータ生成部の構成例を示すブロック図である。 10

【図 1 8】第 1 0 の実施の形態に係る音素長制御の処理手順の一例を示すフローチャートである。

【図 1 9】話速調整部を備えるパラメータ生成部を示すブロック図である。

【図 2 0】音素長制御の処理手順の一例を示すフローチャートである。

【図 2 1】言語処理結果を示す図である。

【図 2 2】音素長の生成例を示す図である。

【図 2 3】音素長の生成例を示す図である。

【図 2 4】音声合成波形を示す図である。

【図 2 5】音声合成波形を示す図である。 20

【図 2 6】音声合成波形を示す図である。

【図 2 7】音声合成波形を示す図である。

【図 2 8】音声合成波形を示す図である。

【図 2 9】音声合成波形を示す図である。

【図 3 0】音声合成波形を示す図である。

【図 3 1】音声合成波形を示す図である。

【符号の説明】

【0 2 2 7】

2 音声読み上げ装置

2 4 音素長調整部 30

2 6 話速判定部

2 8 音素判定部

3 0 呼気段落長演算部

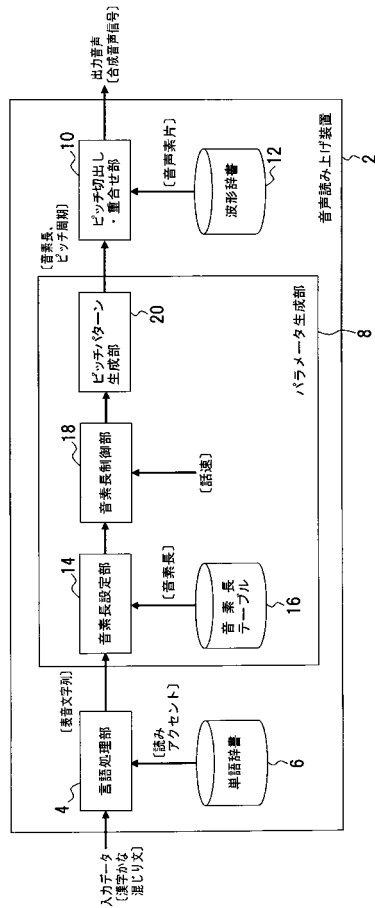
3 2 文章全体長演算部

3 4 区切り変更部

2 0 0 携帯端末装置

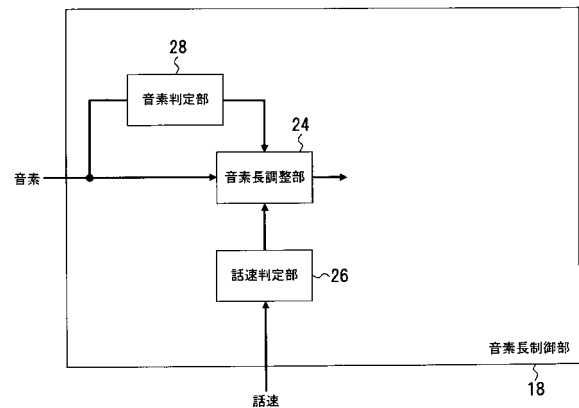
【図 1】

第1の実施の形態に係る音声読み上げ装置の構成例を示すブロック図



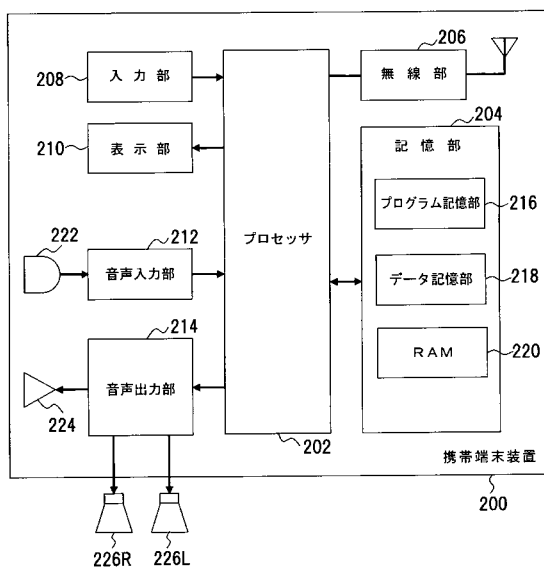
【図 2】

音声読み上げ装置の音楽長制御部の構成例を示すブロック図



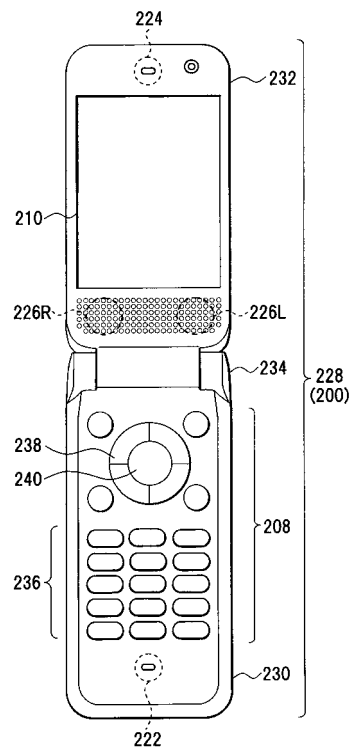
【図 3】

音声読み上げ装置を搭載した携帯端末装置の一例を示すブロック図



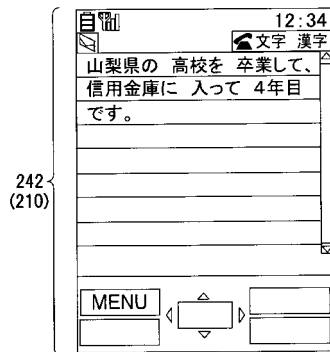
【図 4】

携帯端末装置の構成例を示す図



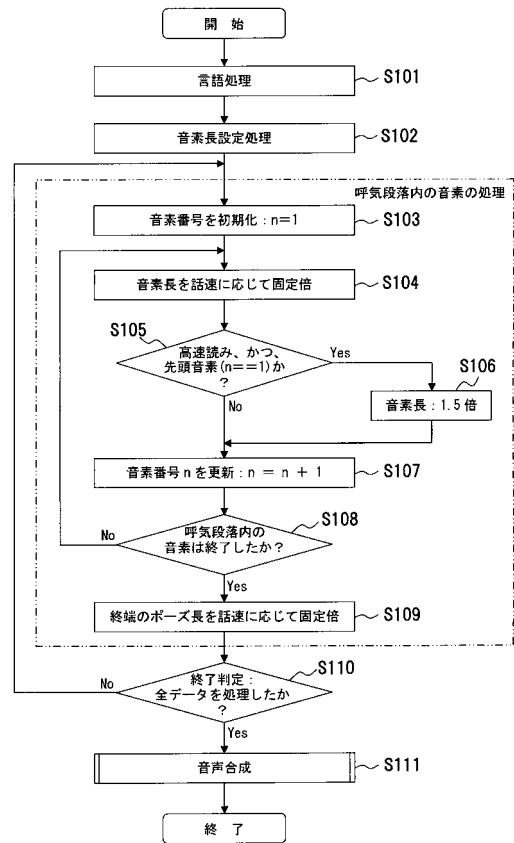
【図 5】

画面表示例を示す図



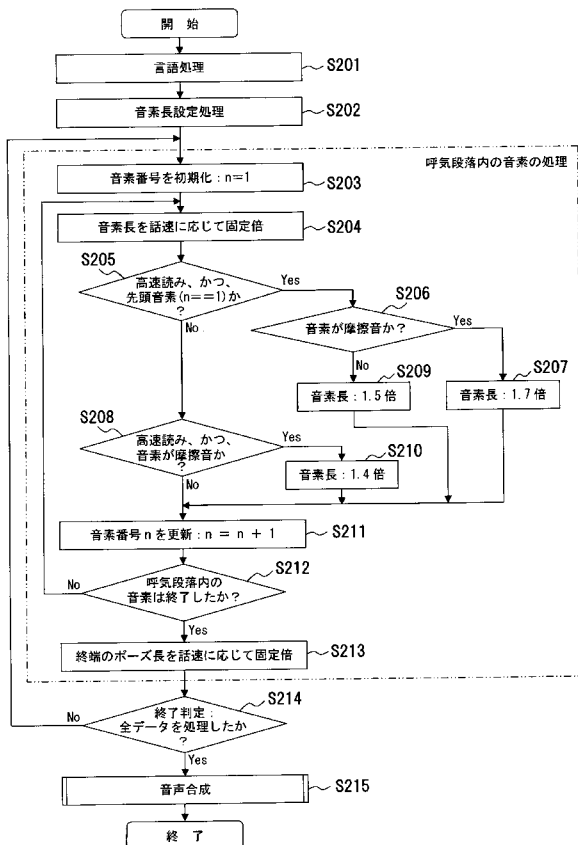
【図 6】

第 1 の実施の形態に係る音素長制御の処理手順の一例を示すフローチャート



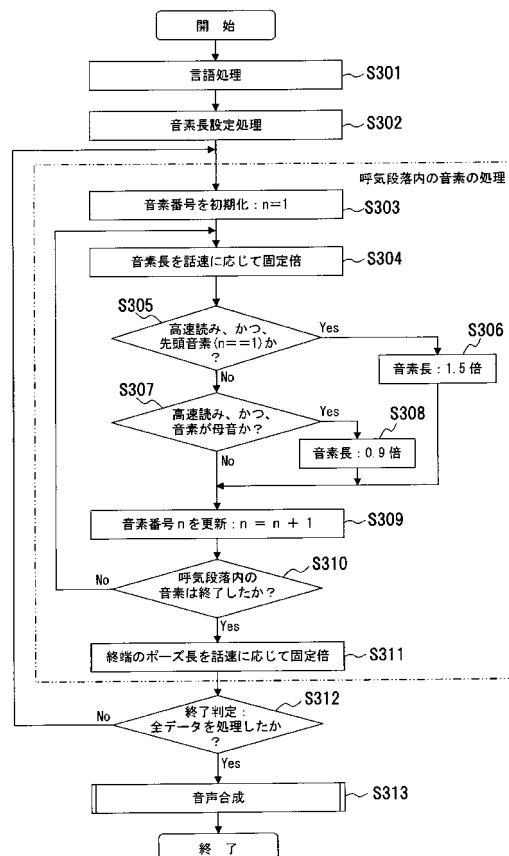
【図 7】

第 2 の実施の形態に係る音素長制御の処理手順の一例を示すフローチャート



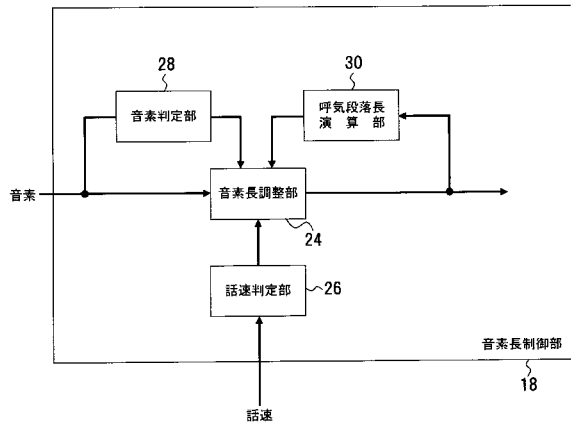
【図 8】

第 3 の実施の形態に係る音素長制御の処理手順の一例を示すフローチャート



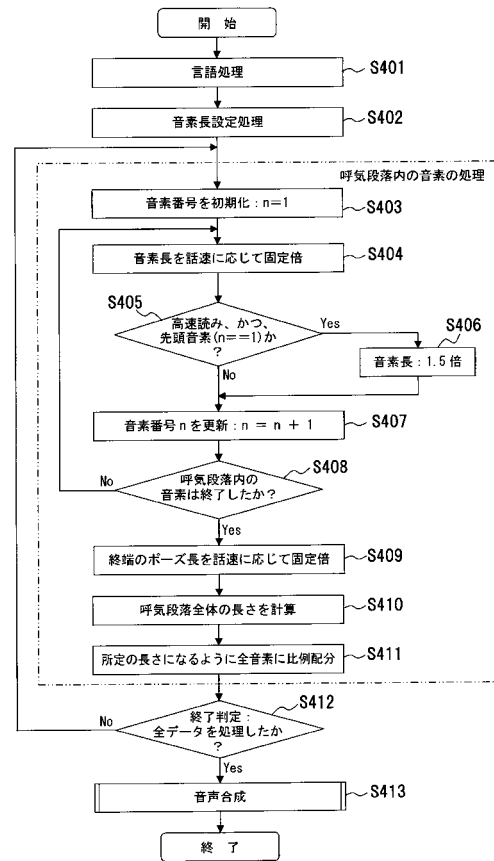
【図 9】

第4の実施の形態に係る音素長制御部を示すブロック図



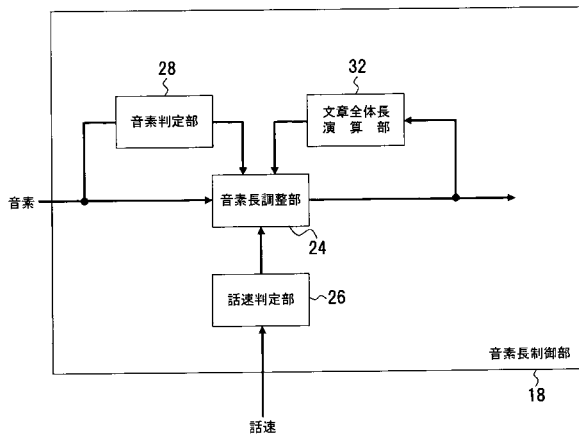
【図 10】

第4の実施の形態に係る音素長制御の処理手順の一例を示すフローチャート



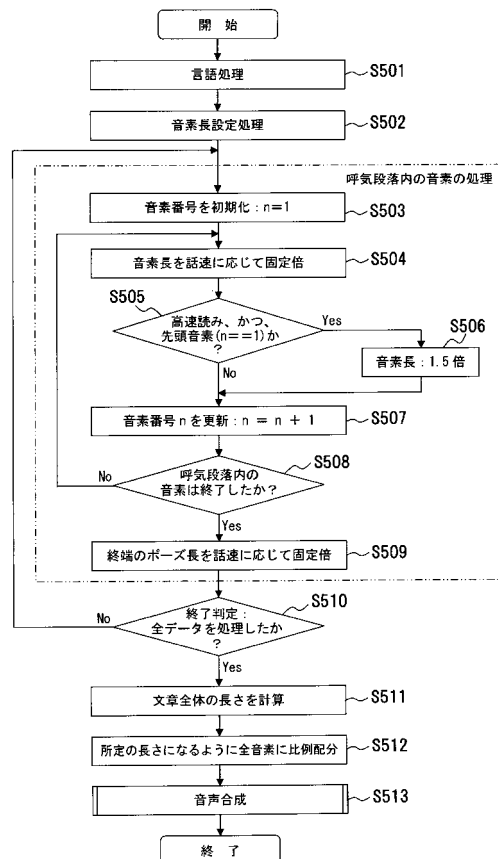
【図 11】

第5の実施の形態に係る音素長制御部を示すブロック図



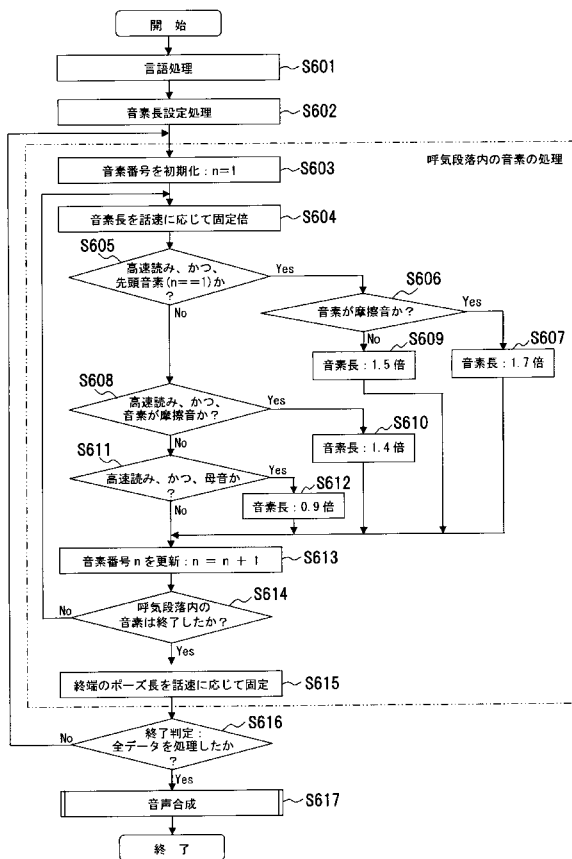
【図 12】

第5の実施の形態に係る音素長制御の処理手順の一例を示すフローチャート



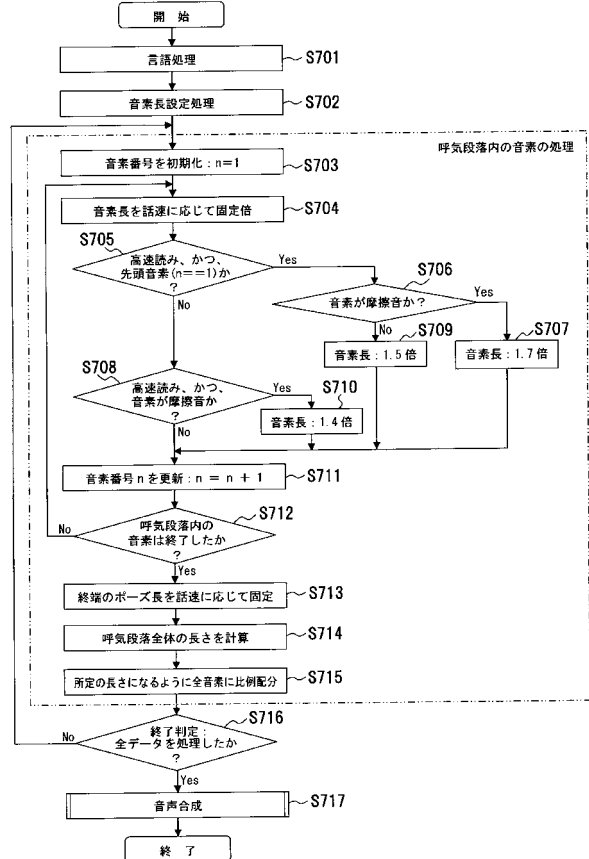
【図 13】

第6の実施の形態に係る音素長制御の処理手順の一例を示すフローチャート



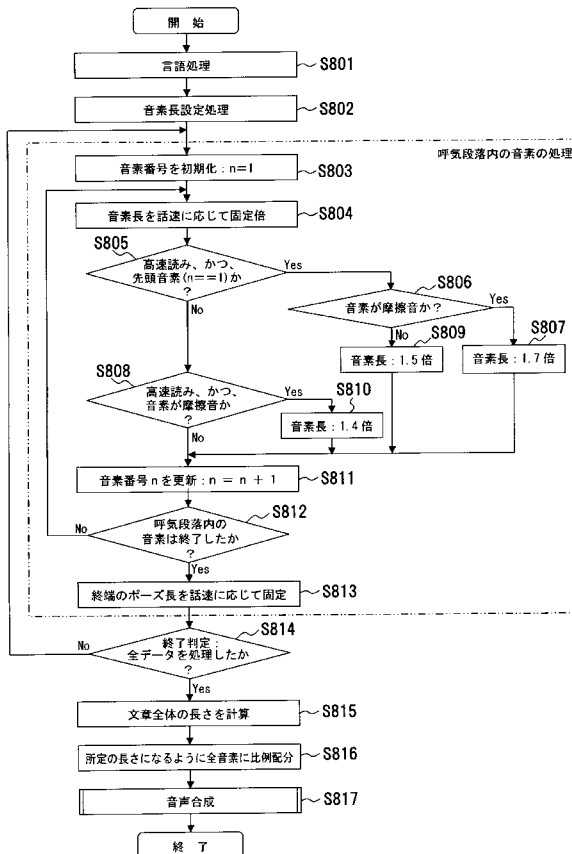
【図 14】

第7の実施の形態に係る音素長制御の処理手順の一例を示すフローチャート



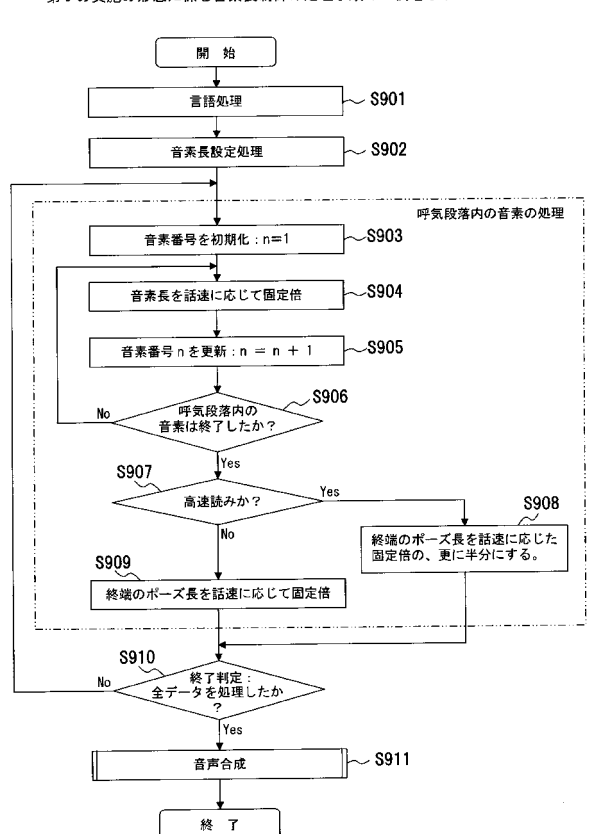
【図 15】

第8の実施の形態に係る音素長制御の処理手順の一例を示すフローチャート



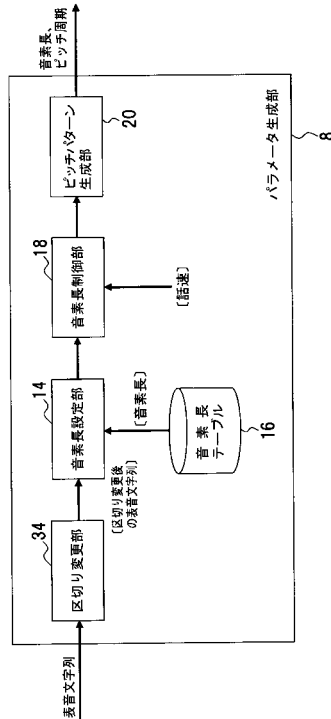
【図 16】

第9の実施の形態に係る音素長制御の処理手順の一例を示すフローチャート



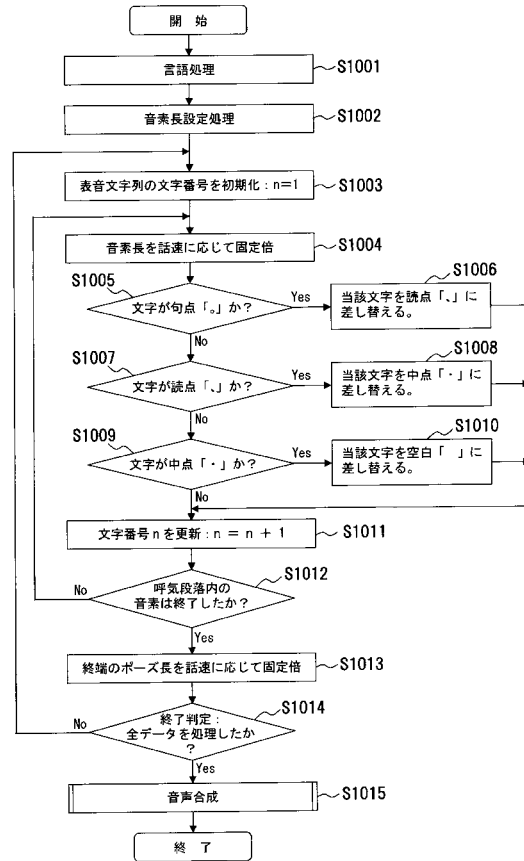
【図 17】

第10の実施の形態に係る音声読み上げ装置のパラメータ生成部の構成例を示すブロック図



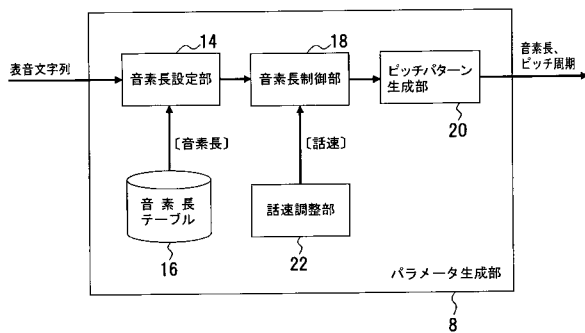
【図 18】

第10の実施の形態に係る音楽長制御の処理手順の一例を示すフローチャート



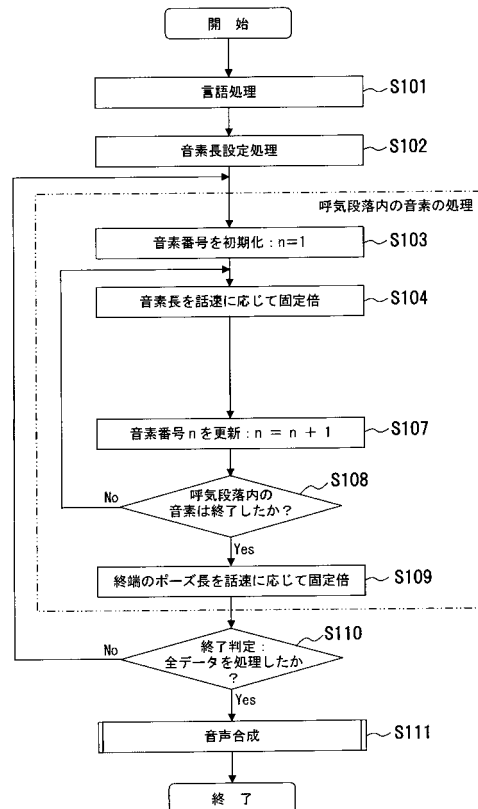
【図 19】

話速調整部を備えるパラメータ生成部を示すブロック図



【図 20】

音楽長制御の処理手順の一例を示すフローチャート



【図 2 1】

言語処理結果を示す図

| 入力テキスト | 品 詞             | 表音文字列  |
|--------|-----------------|--------|
| 山梨     | 名詞              | ヤマナシ   |
| 県      | 名詞              | ケン     |
| の      | 助詞              | ノ      |
|        | アクセント句境界        | (空白)   |
| 高校     | 名詞              | コウコウ   |
| を      | 助詞              | オ      |
|        | アクセント句境界        | (空白)   |
| 卒業し    | 動詞 (連用形)        | ソツギョーシ |
| て      | 助詞              | テ      |
| 、      | 呼吸段落境界 (ポーズ長は中) | 、      |
| 信用     | 名詞              | シンヨウ   |
| 金庫     | 名詞              | キンコ    |
| に      | 助詞              | ニ      |
|        | アクセント句境界        | (空白)   |
| 入っ     | 動詞 (連用形、促音便)    | ハ*イツ   |
| て      | 助詞              | テ      |
|        | 呼吸段落境界 (ポーズ長は小) | ・      |
| 4      | 数詞              | ヨ      |
| 年      | 助数詞             | ネン     |
| 目      | 助数詞の後置詞         | メ'     |
| です     | 助動詞             | デス     |
| 。      | 呼吸段落境界 (ポーズ長は大) | 。      |

【図 2 2】

音素長の生成例を示す図

| 音素 | 呼吸段落一音素番号 | 1 倍速での音素長 [ms] | 3 倍速での音素長 [ms] |
|----|-----------|----------------|----------------|
| sh | 2-1       | 117            | 39             |
| I  | 2-2       | 60             | 20             |
| N  | 2-3       | 60             | 20             |
| y  | 2-4       | 65             | 22             |
| o  | 2-5       | 80             | 27             |
| o  | 2-6       | 105            | 35             |

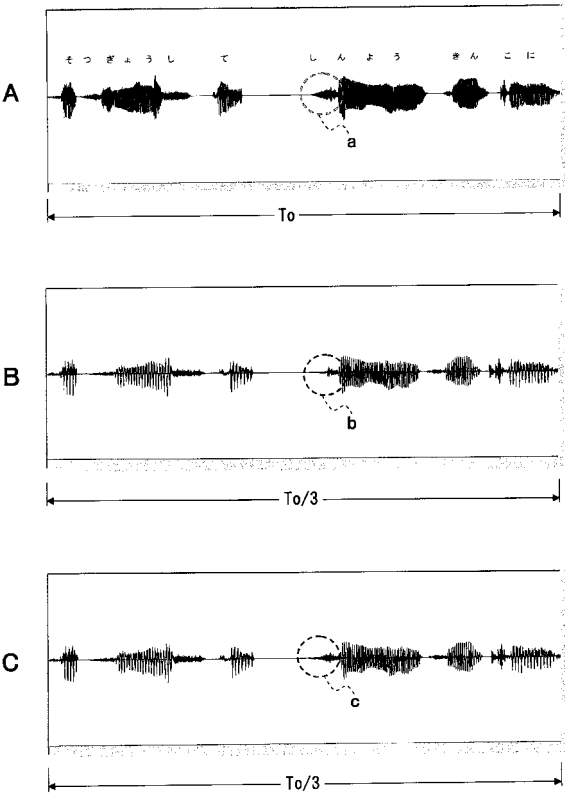
【図 2 3】

音素長の生成例を示す図

| 音素 | 呼吸段落一音素番号 | 1 倍速での音素長 [ms] | 3 倍速での音素長 [ms] |
|----|-----------|----------------|----------------|
| sh | 2-1       | 117            | 59             |
| I  | 2-2       | 60             | 20             |
| N  | 2-3       | 60             | 20             |
| y  | 2-4       | 65             | 22             |
| o  | 2-5       | 80             | 27             |
| o  | 2-6       | 105            | 35             |

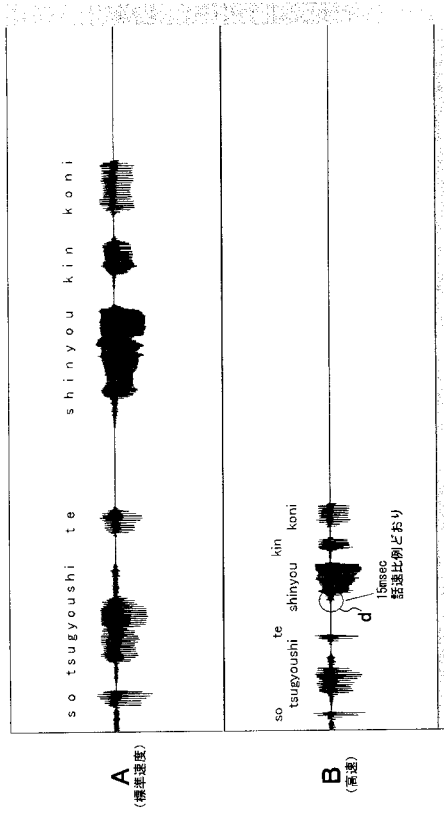
【図 2 4】

音声合成波形を示す図



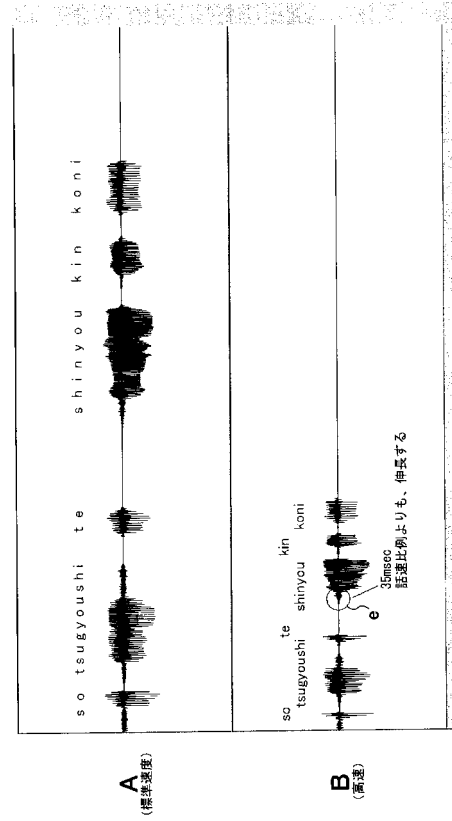
【図 25】

音声合成波形を示す図



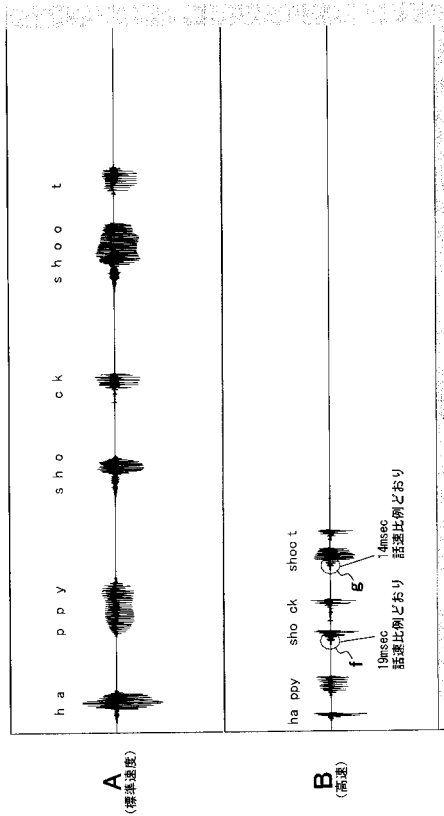
【図 26】

音声合成波形を示す図



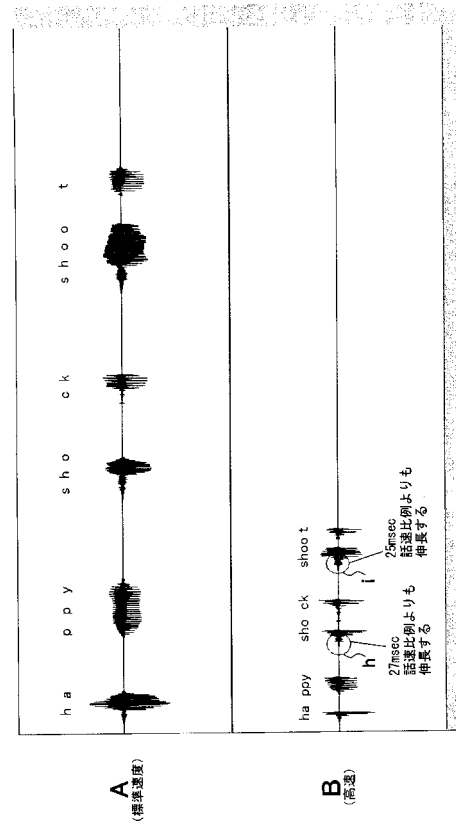
【図 27】

音声合成波形を示す図



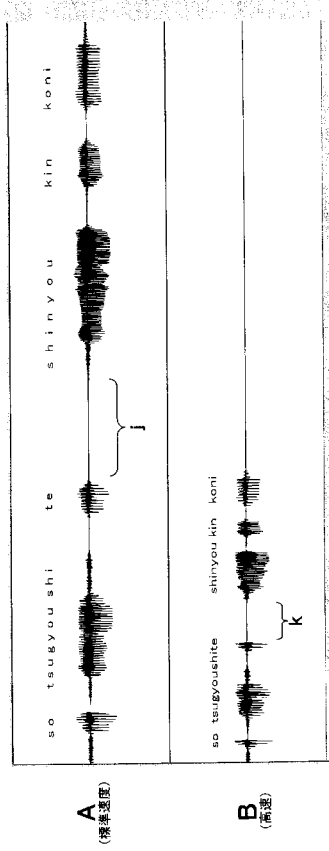
【図 28】

音声合成波形を示す図



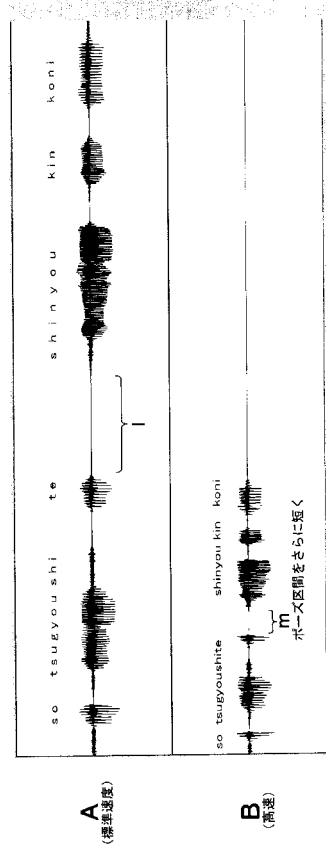
【図 29】

音声合成波形を示す図



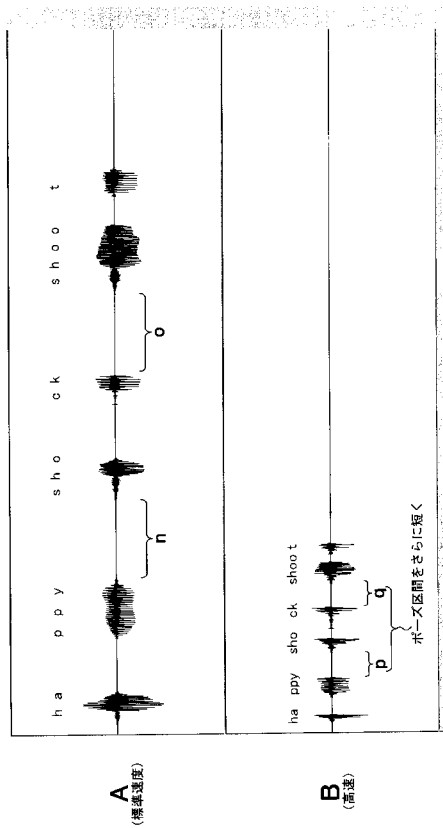
【図 30】

音声合成波形を示す図



【図 31】

音声合成波形を示す図



---

フロントページの続き

(72)発明者 村瀬 健太郎

神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

(72)発明者 野田 拓也

神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内