

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5164041号  
(P5164041)

(45) 発行日 平成25年3月13日(2013.3.13)

(24) 登録日 平成24年12月28日(2012.12.28)

(51) Int.Cl.		F I		
<b>G 1 O L 13/10</b>	<b>(2013.01)</b>	G 1 O L 13/08	1 2 7 F	
<b>G 1 O L 13/06</b>	<b>(2013.01)</b>	G 1 O L 13/06	1 3 0	
<b>G 1 O L 13/08</b>	<b>(2013.01)</b>	G 1 O L 13/08	1 2 7 D	
		G 1 O L 13/08	1 3 2	

請求項の数 8 (全 19 頁)

(21) 出願番号	特願2008-231612 (P2008-231612)	(73) 特許権者	301022471
(22) 出願日	平成20年9月10日 (2008.9.10)		独立行政法人情報通信研究機構
(65) 公開番号	特開2010-66422 (P2010-66422A)		東京都小金井市貫井北町4-2-1
(43) 公開日	平成22年3月25日 (2010.3.25)	(74) 代理人	100115749
審査請求日	平成23年8月19日 (2011.8.19)		弁理士 谷川 英和
		(74) 代理人	100121223
			弁理士 森本 悟道
		(72) 発明者	津崎 実
			東京都小金井市貫井北町4-2-1 独立行政法人情報通信研究機構内
		(72) 発明者	坂井 信輔
			東京都小金井市貫井北町4-2-1 独立行政法人情報通信研究機構内
		審査官	毛利 太郎

最終頁に続く

(54) 【発明の名称】 音声合成装置、音声合成方法、及びプログラム

(57) 【特許請求の範囲】

【請求項1】

コーパスベースの音声合成装置であって、  
 テキスト情報と、当該テキスト情報の一部に対する速度に関する情報である速度情報とが記憶されるテキスト情報記憶部と、  
 前記テキスト情報に対応する音素列と、当該音素列の各音素の持続時間に関する情報である持続時間情報を少なくとも含む情報である韻律情報とが記憶される韻律情報記憶部と、  
 前記速度情報に応じて前記韻律情報を変更する韻律情報変更部と、  
 前記音素列と、前記韻律情報変更部による変更後の韻律情報とを用いて音声信号を生成する音声信号生成部と、  
 前記音声信号生成部が生成した音声信号を出力する音声信号出力部と、を備え、  
前記音素列には、ポーズが含まれており、  
前記韻律情報には、前記ポーズに対応する持続時間情報も含まれており、  
前記韻律情報変更部は、ポーズに対応する持続時間情報を変更しない、音声合成装置。

【請求項2】

前記音声信号生成部は、素片接続合成法により音声信号を生成する、請求項1記載の音声合成装置。

【請求項3】

前記音声信号生成部は、HMM合成法により音声信号を生成する、請求項1記載の音声合成装置。

## 【請求項 4】

前記テキスト情報から音素列、及び当該音素列に対応する韻律情報を生成し、前記韻律情報記憶部に蓄積する音素列・韻律情報生成部をさらに備えた、請求項 1 から請求項 3 のいずれか記載の音声合成装置。

## 【請求項 5】

前記韻律情報は、前記音素列の各音素の基本周波数情報を含む、請求項 1 から請求項 4 のいずれか記載の音声合成装置。

## 【請求項 6】

前記韻律情報は、前記音素列の各音素のパワースペクトル情報を含む、請求項 1 から請求項 5 のいずれか記載の音声合成装置。

10

## 【請求項 7】

コーパスベースの音声合成方法であって、

テキスト情報と当該テキスト情報の一部に対する速度に関する情報である速度情報とが記憶されるテキスト情報記憶部で記憶されている前記速度情報に応じて、前記テキスト情報に対応する音素列と当該音素列の各音素の持続時間に関する情報である持続時間情報を少なくとも含む情報である韻律情報とが記憶される韻律情報記憶部で記憶されている前記韻律情報を変更する韻律情報変更ステップと、

前記音素列と、前記韻律情報変更ステップでの変更後の韻律情報とを用いて音声信号を生成する音声信号生成ステップと、

前記音声信号生成ステップで生成した音声信号を出力する音声信号出力ステップと、を備え、

20

前記音素列には、ポーズが含まれており、

前記韻律情報には、前記ポーズに対応する持続時間情報も含まれており、

前記韻律情報変更ステップでは、ポーズに対応する持続時間情報を変更しない、音声合成方法。

## 【請求項 8】

コーパスベースの音声合成装置としてコンピュータを機能させるためのプログラムであって、

コンピュータを、

テキスト情報と当該テキスト情報の一部に対する速度に関する情報である速度情報とが記憶されるテキスト情報記憶部で記憶される前記速度情報に応じて、前記テキスト情報に対応する音素列と当該音素列の各音素の持続時間に関する情報である持続時間情報を少なくとも含む情報である韻律情報とが記憶される韻律情報記憶部で記憶される前記韻律情報を変更する韻律情報変更部、

30

前記音素列と、前記韻律情報変更部による変更後の韻律情報とを用いて音声信号を生成する音声信号生成部、

前記音声信号生成部が生成した音声信号を出力する音声信号出力部、として機能させ、

前記音素列には、ポーズが含まれており、

前記韻律情報には、前記ポーズに対応する持続時間情報も含まれており、

前記韻律情報変更部は、ポーズに対応する持続時間情報を変更しない、プログラム。

40

## 【発明の詳細な説明】

## 【技術分野】

## 【0001】

本発明は、コーパスベースの音声合成装置等に関する。

## 【背景技術】

## 【0002】

従来の音声合成装置において、発話速度を変換する技術が確立されてきている（例えば、特許文献 1 参照）。したがって、状況の切迫度などに応じて、発話速度の速い音声出力することもできる。

## 【特許文献 1】特開 2007 - 279975 号公報

50

## 【発明の開示】

## 【発明が解決しようとする課題】

## 【0003】

しかしながら、従来の音声合成装置においては、音声の全体について一様に発話速度を変更することしかできなかった。その結果、例えば、発話速度を速くしてしまうことによって、人名や地名、数詞、キーワードなどの正確に聞き取る必要のある部分の聞き取りが阻害されてしまうことがあった。これは人間の聴覚的な特性上避けられないことである。したがって、基本的には発話速度増加の要請と明瞭度の維持は一樣な発話速度変換によっては両立し得ない。

## 【0004】

一般的に言えば、従来の音声合成装置では、部分的に発話速度を変更することができず、そのため、聞き手にとって聞きやすい発話速度となる音声を出力することが困難であった。

## 【0005】

本発明は、上記問題点を解決するためになされたものであり、部分的に発話速度を変更することが可能な音声合成装置等を提供することを目的とする。

## 【課題を解決するための手段】

## 【0006】

上記目的を達成するため、本発明による音声合成装置は、コーパスベースの音声合成装置であって、テキスト情報と、当該テキスト情報の一部に対する速度に関する情報である速度情報とが記憶されるテキスト情報記憶部と、前記テキスト情報に対応する音素列と、当該音素列の各音素の持続時間に関する情報である持続時間情報を少なくとも含む情報である韻律情報とが記憶される韻律情報記憶部と、前記速度情報に応じて前記韻律情報を変更する韻律情報変更部と、前記音素列と、前記韻律情報変更部による変更後の韻律情報とを用いて音声信号を生成する音声信号生成部と、前記音声信号生成部が生成した音声信号を出力する音声信号出力部と、を備えたものである。

## 【0007】

このような構成により、速度情報を設定することによって、テキスト情報の一部の速度を他の部分に比べて速くしたり、遅くしたりすることができる。その結果、例えば、テキスト情報の重要な部分は遅いスピードで読み上げられ、重要でない部分は早いスピードで読み上げられるように設定することも可能となる。

## 【0008】

また、本発明による音声合成装置では、前記音声信号生成部は、素片接続合成法により音声信号を生成してもよく、HMM合成法により音声信号を生成してもよい。

このような構成により、素片接続合成法やHMM合成法での音声合成を行うことができる。

## 【0009】

また、本発明による音声合成装置では、前記テキスト情報から音素列、及び当該音素列に対応する韻律情報を生成し、前記韻律情報記憶部に蓄積する音素列・韻律情報生成部をさらに備えてもよい。

## 【0010】

このような構成により、テキスト情報からの音素列、韻律情報の生成、速度情報に応じた韻律情報の変更、音素列、韻律情報を用いた音声信号の生成の一連の処理を音声合成装置において行うことができるようになる。

## 【0011】

また、本発明による音声合成装置では、前記韻律情報は、前記音素列の各音素の基本周波数情報を含んでもよく、前記音素列の各音素のパワースペクトル情報を含んでもよい。

このような構成により、基本周波数情報や、パワースペクトル情報を含む韻律情報を用いて、音声合成を行うことができる。

## 【発明の効果】

10

20

30

40

50

## 【0012】

本発明による音声合成装置等によれば、部分的に発話速度を変更することができ、その結果、例えば、聞き手にとって聞きやすい発話速度の音声信号を出力することが可能となりうる。

## 【発明を実施するための最良の形態】

## 【0013】

以下、本発明による音声合成装置について、実施の形態を用いて説明する。なお、以下の実施の形態において、同じ符号を付した構成要素及びステップは同一または相当するものであり、再度の説明を省略することがある。

## 【0014】

## (実施の形態1)

本発明の実施の形態1による音声合成装置について、図面を参照しながら説明する。本実施の形態による音声合成装置は、部分的に発話速度を変更することができるものである。

## 【0015】

図1は、本実施の形態による音声合成装置1の構成を示すブロック図である。本実施の形態による音声合成装置1は、テキスト情報記憶部11と、音素列・韻律情報生成部12と、韻律情報記憶部13と、韻律情報変更部14と、音声信号生成部15と、音声信号出力部16とを備える。

## 【0016】

テキスト情報記憶部11では、テキスト情報と、そのテキスト情報の一部に対する速度に関する情報である速度情報とが記憶される。このテキスト情報は、本実施の形態による音声合成装置1での音声合成の対象となるテキスト情報である。また、速度情報は、速度に関する情報であれば、例えば、「速い」「遅い」等の情報であってもよく、「2倍」「0.5倍」等の情報であってもよく、あるテキストの読み上げ時間を示す情報（例えば、「10時30分」を5秒で読み上げる旨の情報）であってもよく、キーワードの読み上げスピードを他のキーワードでないテキストよりも遅くするというルールがある場合には、どのテキストがキーワードであるのかを示す情報であってもよい。また、速度情報は、テキスト情報全体に対する速度に関する情報を含んでいてもよい。速度情報は、テキスト情報のどの部分に対する速度に関する情報であるのかが明らかになるように記述されることが好適である。速度情報は、例えば、XML等のマークアップ言語を用いて記述されてもよく、あるいは、その他の方法によって記述されてもよい。

## 【0017】

テキスト情報記憶部11にテキスト情報や速度情報が記憶される過程は問わない。例えば、記録媒体を介してテキスト情報や速度情報がテキスト情報記憶部11で記憶されるようになってもよく、通信回線等を介して送信されたテキスト情報や速度情報がテキスト情報記憶部11で記憶されるようになってもよく、あるいは、入力デバイスを介して入力されたテキスト情報や速度情報がテキスト情報記憶部11で記憶されるようになってもよい。テキスト情報記憶部11での記憶は、RAM等における一時的な記憶でもよく、あるいは、長期的な記憶でもよい。テキスト情報記憶部11は、所定の記録媒体（例えば、半導体メモリや磁気ディスク、光ディスクなど）によって実現されうる。

## 【0018】

音素列・韻律情報生成部12は、テキスト情報記憶部11で記憶されているテキスト情報から音素列、及びその音素列に対応する韻律情報を生成し、韻律情報記憶部13に蓄積する。音素列・韻律情報生成部12は、例えば、テキスト情報に対して形態素解析や漢字カナ変換、アクセント処理等を行い、さらに、それらを元にして韻律情報を生成する。韻律情報は、テキスト情報に対応する音素列の各音素の持続時間に関する情報である持続時間情報を少なくともも含んでいる。持続時間情報は、結果として各音素の持続時間が分かるのであれば、どのような情報であってもよい。持続時間情報は、例えば、持続時間そのものを示す情報であってもよく、各音素の開始時刻を示すものであってもよく（この場合に

10

20

30

40

50

は、ある音素の開始時刻と、その音素に後続する次の音素の開始時刻とを用いることによって持続時間を算出することができる)、あるいは、各音素の開始時刻と終了時刻とを示すものであってもよい(この場合には、ある音素の開始時刻と終了時刻とを用いることによって持続時間を算出することができる)。また、韻律情報には、音素列の各音素の基本周波数情報が含まれてもよく、音素列の各音素のパワースペクトル情報が含まれてもよい。パワースペクトル情報は、例えば、音韻性を示す情報である25次元のケプストラム情報であってもよい。ケプストラム情報は、メルケプストラム情報であってもよい。韻律情報に含まれる全部の情報、あるいは、一部の情報は、フレーム単位の情報であってもよい。例えば、韻律情報に含まれる基本周波数情報と、パワースペクトル情報は、5msごとのフレーム単位の情報であってもよい。この場合であっても、先頭からのフレーム数をカ  
10  
ウントすることによって、各フレームがどの音素に対応しているのかを知ることができ  
る。音素列・韻律情報生成部12は、韻律モデルを用いて音素列や、韻律情報を生成して  
もよい。その韻律モデルを学習するために用いられる音声コーパスは、例えば、後述す  
る音声信号生成部15が素片接続合成法により音声信号を生成する際に用いる音声コーパ  
スや、音声信号生成部15がHMM(Hidden Markov Model, 隠れマル  
コフモデル)合成法により音声信号を生成する際に用いるモデルの学習で用いられた音  
声コーパスと同じものであってもよく、あるいは、異なってもよい。なお、テキスト情  
報から音素列と、韻律情報とを生成する方法はすでに公知であり、その詳細な説明を省略  
する。

#### 【0019】

韻律情報記憶部13では、音素列・韻律情報生成部12が生成した、テキスト情報に対  
10  
応する音素列と、その音素列の各音素に対応する韻律情報とが記憶される。韻律情報記憶  
部13での記憶は、RAM等における一時的な記憶でもよく、あるいは、長期的な記憶で  
もよい。韻律情報記憶部13は、所定の記録媒体(例えば、半導体メモリや磁気ディスク  
、光ディスクなど)によって実現されうる。

#### 【0020】

韻律情報変更部14は、テキスト情報記憶部11で記憶されている速度情報に応じて、  
韻律情報記憶部13で記憶されている韻律情報を変更する。この変更によって、速度情報  
で示される速度となるように韻律情報が変更されることになる。なお、例えば、テキ  
スト情報記憶部11に記憶されているテキスト情報及び速度情報が、XML等のタグ付き  
30  
テキストで記述されている場合には、韻律情報変更部14は、タグ解釈を行い、速度を  
変更しない部分と、速度を変更する部分とを特定し、また、速度を変更する箇所につ  
いてはどれぐらい速度を変更するのかを特定してもよい。韻律情報に含まれる持続時間  
情報が、各音素の開始時刻を示す場合であっても、速度情報によって一部のテキストに  
対応する速度を速くすることが示される場合には、韻律情報変更部14は、その一部  
のテキストの読み上げ速度が速くなるように、その一部のテキストに対応する各音  
素の開始時刻を変更する処理を行う。なお、その開始時刻の変更に伴って、速度が  
変更されない音素の開始時刻がずらされることもある。

#### 【0021】

また、韻律情報に含まれる基本周波数情報等がフレーム単位の情報である場合には、  
40  
各フレームの基本周波数情報等について補間処理を行い、その補間後の波形における  
速度の変更対象となる区間について速度の変更に応じた伸縮を行った後に、元のフ  
レームレートで再サンプリングを行うことによって、速度情報に応じた変更を行うこ  
とができる。この処理は、速度の変更対象となる区間についてのみ行ってもよく、あ  
るいは、全体について行ってもよい。後者の場合には、速度の変更対象となる区間  
と、そうでない区間との間で生じる誤差をなくすことができる。なお、結果として  
同様のことを行うのであれば、異なる方法によって行ってもよい。例えば、補間  
処理を行った後に、速度の変更対象となる区間については、変更後の速度に合わ  
せて変更したレートでサンプリングを行うようにしてもよい。なお、このようなサ  
ンプリングレートを変換する処理は、例えば、MATLAB(登録商標)などのソフトウ  
ェアを用いて容易に行うことができる。

10

20

30

40

50

## 【 0 0 2 2 】

また、音素列には、ポーズが含まれており、韻律情報にポーズに対応する持続時間情報も含まれている場合に、韻律情報変更部 1 4 は、ポーズに対応する持続時間情報を変更してもよく、しなくてもよい。後者の場合には、ポーズの持続時間については、あらかじめ音素列・韻律情報生成部 1 2 が生成した長さのままにすることができる。

## 【 0 0 2 3 】

音声信号生成部 1 5 は、音素列と、韻律情報変更部 1 4 による変更後の韻律情報とを用いて音声信号を生成する。なお、音声信号生成部 1 5 は、コーパスベースの音声合成を行うものである。音声信号生成部 1 5 は、素片接続合成法により音声信号を生成してもよく、あるいは、HMM合成法により音声信号を生成してもよい。なお、素片接続合成法による音声合成では、音声コーパスから抽出された音声素片であって、韻律情報に適合した音声素片をつなげることによって音声合成が行われる。音声合成の方法が素片接続合成法である場合には、韻律情報に含まれる持続時間情報がマークファイルと呼ばれ、基本周波数情報が F 0 ファイルと呼ばれ、パワースペクトル情報がケプストラムファイルと呼ばれることがある。また、HMM合成法による音声合成では、音声コーパスからモデルが学習され、その学習されたモデルを用いて音声合成が行われる。音声合成の方法がHMM合成法である場合には、韻律情報に含まれる持続時間情報がフルコンテキストラベルファイルと呼ばれることがある。なお、音素列と韻律情報を用いて音声合成を行う方法はすでに公知であり、その詳細な説明を省略する。例えば、素片接続合成法を用いた音声合成システムとしては、A T R が開発した X I M E R A が知られている。また、HMM合成法を用いた音声合成システムとしては、H T S ( <http://hts.speech.ac.jp/> ) が知られている。

## 【 0 0 2 4 】

音声信号出力部 1 6 は、音声信号生成部 1 5 が生成した音声信号を出力する。ここで、この出力は、例えば、所定の機器への通信回線を介した送信でもよく、スピーカによる音声出力でもよく、記録媒体への蓄積でもよく、他の構成要素への引き渡しでもよい。なお、音声信号出力部 1 6 は、出力を行うデバイス（例えば、通信デバイスやスピーカなど）を含んでもよく、あるいは含まなくてもよい。また、音声信号出力部 1 6 は、ハードウェアによって実現されてもよく、あるいは、それらのデバイスを駆動するドライバ等のソフトウェアによって実現されてもよい。

## 【 0 0 2 5 】

なお、テキスト情報記憶部 1 1 と、韻律情報記憶部 1 3 とは、同一の記録媒体によって実現されてもよく、あるいは、別々の記録媒体によって実現されてもよい。前者の場合には、テキスト情報等を記憶している領域がテキスト情報記憶部 1 1 となり、韻律情報等を記憶している領域が韻律情報記憶部 1 3 となる。

## 【 0 0 2 6 】

また、テキスト情報記憶部 1 1 において、テキスト情報と、速度情報とは、同一の記録媒体に記憶されてもよく、あるいは、別々の記録媒体に記憶されてもよい。また、韻律情報記憶部 1 3 において、音素列と、韻律情報とは、同一の記録媒体に記憶されてもよく、あるいは、別々の記録媒体に記憶されてもよい。

## 【 0 0 2 7 】

次に、本実施の形態による音声合成装置 1 の動作について、図 2 のフローチャートを用いて説明する。

(ステップ S 1 0 1) 音素列・韻律情報生成部 1 2 は、音声合成の一連の処理を開始するかどうか判断する。そして、音声合成の一連の処理を開始する場合には、ステップ S 1 0 2 に進み、そうでない場合には、処理を開始すると判断するまでステップ S 1 0 1 の処理を繰り返して実行する。音素列・韻律情報生成部 1 2 は、例えば、音声合成の一連の処理を開始する旨の指示を受け付けた場合に、その処理を開始すると判断してもよく、あるいは、テキスト情報記憶部 1 1 に新たなテキスト情報と速度情報とが蓄積されたタイミングで、音声合成の一連の処理を開始すると判断してもよく、その他のタイミングで、音声

10

20

30

40

50

合成の一連の処理を開始すると判断してもよい。

【0028】

(ステップS102)音素列・韻律情報生成部12は、テキスト情報記憶部11で記憶されているテキスト情報を読み出し、そのテキスト情報に対して形態素解析等を行うことにより音素列を生成し、また、その音素列の各音素に対応する韻律情報を生成する。そして、音素列・韻律情報生成部12は、生成した音素列と、韻律情報とを韻律情報記憶部13に蓄積する。

【0029】

(ステップS103)韻律情報変更部14は、テキスト情報記憶部11で記憶されている速度情報を用いて、韻律情報記憶部13で記憶されている韻律情報を変更する。この処理によって、韻律情報で示される各音素の長さが、速度情報に対応したものとなる。この処理の詳細については図3のフローチャートを用いて後述する。

10

【0030】

(ステップS104)音声信号生成部15は、韻律情報記憶部13で記憶されている音素列と韻律情報とを用いて音声信号を生成する。

【0031】

(ステップS105)音声信号出力部16は、音声信号生成部15が生成した音声信号を出力する。そして、ステップS101に戻る。

なお、図2のフローチャートにおいて、電源オフや処理終了の割り込みにより処理は終了する。

20

【0032】

図3は、図2のフローチャートにおける韻律情報の変更(ステップS103)の処理の詳細を示すフローチャートである。

【0033】

(ステップS201)韻律情報変更部14は、テキスト情報記憶部11で記憶されている速度情報を参照し、韻律情報記憶部13で記憶されている音素列のうち、速度を変更する音素を特定する。韻律情報変更部14は、例えば、速度情報を用いて速度を変更する対象となるテキストの部分特定し、そのテキストの部分に対応する音素列の部分特定することによって、速度を変更する音素の特定を行ってもよい。また、音素を特定するとは、例えば、特定対象となる音素にフラグを設定することであってもよく、特定対象となる音素を識別する情報を記録媒体に蓄積することであってもよい。ここで、例えば、テキスト情報記憶部11に記憶されているテキスト情報及び速度情報が、XML等のタグ付きテキストで記述されている場合には、韻律情報変更部14は、タグ解釈を行うことによって、速度を変更する対象となるテキストの部分特定してもよい。

30

【0034】

(ステップS202)韻律情報変更部14は、カウンタ*i*を1に設定する。

【0035】

(ステップS203)韻律情報変更部14は、*i*番目の音素が、速度を変更する音素であるかどうか判断する。すなわち、韻律情報変更部14は、*i*番目の音素がステップS201で特定された音素であるかどうか判断する。そして、速度を変更する音素である場合、すなわち、ステップS201で特定された音素である場合には、ステップS204に進み、そうでない場合には、ステップS205に進む。

40

【0036】

(ステップS204)韻律情報変更部14は、*i*番目の音素に対応する韻律情報を、その音素の速度が速度情報に対応するものとなるように変更する。例えば、韻律情報に含まれる持続時間情報が持続時間そのものを示す場合には、韻律情報変更部14は、*i*番目の音素に対応する持続時間情報を変更する。速度情報によって速度をA倍にすることが示されている場合には、韻律情報変更部14は、その持続時間情報を1/A倍にする。また、例えば、韻律情報に含まれる持続時間情報が開始時刻を示す場合には、韻律情報変更部14は、(*i*+1)番目の音素に対応する持続時間情報(開始時刻)を変更する。速度情報

50

によって速度をA倍にすることが示されている場合には、韻律情報変更部14は、i番目の音素の開始時刻と、(i+1)番目の音素の開始時刻との差が1/A倍となるように、(i+1)番目の音素の開始時刻を変更する。なお、この場合には、(i+1)番目の開始時刻の変更に伴って、後続する(i+2)番目以降の音素の開始時刻もずれることになる(これらは単なる平行移動である)。したがって、韻律情報変更部14は、後続する(i+2)番目以降の音素の開始時刻も、順次変更する。持続時間情報に終了時刻が含まれる場合にも、同様にして終了時刻の変更を行うことができる。また、速度情報が読み上げ時間を示す情報である場合には、韻律情報変更部14は、例えば、韻律情報に含まれる持続時間情報を用いて、速度の変更対象となるテキストの領域の読み上げ時間を算出し、その算出した読み上げ時間が速度情報の示す読み上げ時間となるために、持続時間を何倍にしなければならぬかを算出し、その算出した倍率を用いて、上述のようにして韻律情報を変更する処理を行ってもよい。

10

## 【0037】

(ステップS205) 韻律情報変更部14は、カウンタiを1だけインクリメントする。

## 【0038】

(ステップS206) 韻律情報変更部14は、韻律情報記憶部13でi番目の音素が記憶されているかどうか判断する。そして、記憶されている場合には、ステップS203に戻り、そうでない場合には、図2のフローチャートに戻る。

## 【0039】

なお、図3のフローチャートでは、持続時間情報の変更についてのみ説明したが、前述のように、韻律情報に所定のフレームレートでサンプリングされた基本周波数情報等が含まれる場合には、それらの情報についても順次変更する必要がある。したがって、図4を用いて、その方法について簡単に説明する。

20

## 【0040】

図4は、所定のフレームレートでサンプリングされた情報(以下の説明では、「変更対象情報」とする。変更対象情報は、例えば、基本周波数情報やパワースペクトル情報である)に対する速度情報に応じた変更処理を示すフローチャートである。

## 【0041】

(ステップS301) 韻律情報変更部14は、テキスト情報記憶部11で記憶されている速度情報を参照し、韻律情報記憶部13で記憶されている変更対象情報のうち、速度を変更するフレームを特定する。韻律情報変更部14は、例えば、ステップS201と同様にして速度を変更する音素を特定し、その音素に対応するフレームを特定することによって、変更対象となるフレームを特定してもよい。

30

## 【0042】

(ステップS302) 韻律情報変更部14は、変更対象情報の補間を行う。この補間については、線形補間やスプライン補間など各種の方法が知られており、それらの各種の補間方法を用いることができる。

## 【0043】

(ステップS303) 韻律情報変更部14は、ステップS301で特定したフレームに対応する補間後の変更対象情報の領域を、速度情報に応じて伸縮する。その結果、変更対象情報の長さが、速度情報に対応したものとなる。

40

## 【0044】

(ステップS304) 韻律情報変更部14は、伸縮後の変更対象情報をあらかじめ決められているフレームレートで再サンプリングする。そして、この再サンプリング後の情報を、変更後の変更対象情報とする。

## 【0045】

なお、変更対象情報のうち、全体の先頭のフレームと、速度を変更する対象となる先頭のフレームとの間は、ステップS302~S304の処理を行っても変更されないため、それらのフレームについては、ステップS302~S304の処理を行わなくてもよい。

50



## 【 0 0 4 6 】

また、前述のように、ステップ S 3 0 3 , S 3 0 4 の処理は、速度を変更するフレームについてのみ変更後の速度に対応したフレームレートで再サンプリングする処理と同じになる。したがって、ステップ S 3 0 3 , S 3 0 4 の処理に代えて、そのような処理を行うようにしてもよい。このように、結果としてステップ S 3 0 2 ~ S 3 0 4 と同様の処理、すなわち、フレームレートを変更する処理を行うことができるのであれば、ステップ S 3 0 2 ~ S 3 0 4 の処理に代えて、その処理を行うようにしてもよい。

## 【 0 0 4 7 】

韻律情報に所定のフレームレートでサンプリングされた基本周波数情報等が含まれる場合には、韻律情報変更部 1 4 は、図 3 で示される一連の処理と共に、図 4 で示される一連の処理を行うことによって、その基本周波数等の情報についても速度情報に応じた変更の処理を行ってもよい。なお、図 3 で示される一連の処理と、図 4 で示される一連の処理とを実行する順序は問わない。一方、韻律情報に所定のフレームレートでサンプリングされた基本周波数情報等が含まれない場合には、韻律情報変更部 1 4 は、図 3 で示される一連の処理のみを行えばよいことになる。

## 【 0 0 4 8 】

次に、本実施の形態による音声合成装置 1 の動作について、具体例を用いて説明する。

この具体例では、素片接続合成法により音声合成を行う場合について説明する。この具体例において、テキスト情報記憶部 1 1 では、図 5 で示されるテキスト情報と、速度情報とが記憶されているものとする。図 5 中のタグが速度情報である。「fast」タグによって、「次の東京行きのぞみ 1 0 0 号は 1 0 時 3 0 分に 8 番線から発車の予定です」の全体の話速を速くすることが示されている。また、「keyword」タグによって、「1 0 時 3 0 分に」の部分がキーワードであることが示されている。この具体例では、話速を速くする場合には話速を 2 倍にする、と決まっているものとする。また、この具体例では、キーワードの部分の話速を変化させないと決まっているものとする。

## 【 0 0 4 9 】

まず、ユーザがキーボードやマウス等の入力デバイスを操作することによって、音声合成の処理を開始する指示を入力したとする。すると、音素列・韻律情報生成部 1 2 は、音声合成を開始すると判断し(ステップ S 1 0 1)、音素列と、その音素列に対応した韻律情報を生成し、それらを韻律情報記憶部 1 3 に蓄積する(ステップ S 1 0 2)。図 6 は、韻律情報記憶部 1 3 に蓄積された音素列と、韻律情報に含まれる持続時間情報である開始時間との対応を示す図である。図 6 で示されるように、各音素に、持続時間情報である開始時間が対応付けられている。例えば、音素「ts」は、0.105 秒から開始されることが分かる。保護フラグについては後述する。なお、音素「sil」は、空白に対応する音素であり、音素「paу」は、ポーズに対応する音素である。また、基本周波数情報や、パワースペクトル情報であるケプストラム情報も韻律情報記憶部 1 3 で記憶されているものとする。

## 【 0 0 5 0 】

次に、韻律情報変更部 1 4 は、テキスト情報記憶部 1 1 で記憶されている速度情報を用いて、韻律情報を変更する(ステップ S 1 0 3)。具体的には、韻律情報変更部 1 4 は、図 5 のテキスト情報と速度情報とを参照し、タグ解釈を行うことによって、文の全体を 2 倍の速度にするが、「1 0 時 3 0 分に」の部分は速度を変更しない旨を解釈する。そして、その速度を変更しない「1 0 時 3 0 分に」に対応する音素列「juujisajunjuqpunnij」の各音素に対応付けて図 7 で示されるように保護フラグ「\*」を設定する(ステップ S 2 0 1)。この保護フラグは、速度を変更しないことを示すフラグである。したがって、保護フラグの設定されていない各音素が、速度を変更する音素となる。

## 【 0 0 5 1 】

その後、韻律情報変更部 1 4 は、1 番目の音素「sil」が変更対象であると判断し(ステップ S 2 0 2 , S 2 0 3)、その音素「sil」の開始時刻「0」と、2 番目の音素「ts」の開始時刻「0.105」との差「0.105」が 1/2 の「0.0525」と

10

20

30

40

50

なるように、2番目の音素「t s」の開始時刻を「0.0525」に変更する。また、それに応じて、後続する3番目の音素「u」以降の各音素の開始時刻を、「0.0525」だけ減少させる(ステップS204)。

#### 【0052】

韻律情報変更部14は、次に2番目の音素「t s」が変更対象であると判断し(ステップS205, S206, S203)、その音素「t s」の開始時刻「0.0525」と、3番目の音素「u」の開始時刻「0.1275」との差「0.075」が1/2の「0.0375」となるように、3番目の音素「u」の開始時刻を「0.09」に変更する。また、それに応じて、後続する4番目の音素「g」以降の各音素の開始時刻を、「0.0375」だけ減少させる(ステップS204)。

10

#### 【0053】

このような処理を順次繰り返すことによって、各音素に対応する開始時刻を変更する。ただし、「10時30分に」に対応する音素列「j u u j i s a N j u Q p u N n i」の各音素、すなわち、保護フラグ「\*」の設定されている各音素については、その変更の処理を行わない(ステップS203)。図8は、その処理の後の開始時刻と音素との対応を示す図である。図8において、開始時刻は、小数点以下3桁までを示している(4桁目を落としている)。

#### 【0054】

韻律情報変更部14は、また、基本周波数情報についても速度情報に応じた変更を行う。具体的には、韻律情報変更部14は、「10時30分に」の音素列「j u u j i s a N j u Q p u N n i」に対応する基本周波数情報のフレームを特定する(ステップS301)。その特定は、図7と同様に、保護フラグを各フレームに設定することによって行ってもよい。そして、韻律情報変更部14は、基本周波数情報の先頭から後端までの補間を行う(ステップS302)。その後、韻律情報変更部14は、「10時30分に」の音素列「j u u j i s a N j u Q p u N n i」に対応するフレームでないフレームに対応する領域の基本周波数情報を、時間軸方向に1/2に圧縮し(ステップS303)、元のフレームレートで再サンプリングする(ステップS304)。その結果、例えば、図9で示されるように、変更後の基本周波数情報(図9の右側)は、変更前の基本周波数情報(図9の左側)の各フレームを1個おきに間引いたものとなる。

20

#### 【0055】

なお、韻律情報変更部14は、ケプストラム情報についても、基本周波数情報と同様にして速度情報に応じた変更を行うものとする(ステップS301~S304)。

30

#### 【0056】

次に、音声信号生成部15は、音素列と、変更後の韻律情報とを用いて音声合成を行い(ステップS104)、音声信号出力部16は、その音声合成された音声信号を音声出力する(ステップS105)。その結果、図10で示されるように、ユーザは、「10時30分に」は、通常の長さであるが、「次の東京行きのぞみ100号は」と、「8番線から発車の予定です」は、通常の半分の長さに圧縮された音声を聞くことになる。したがって、全体の音声の長さを短くできると共に、重要であると考えられる「10時30分に」は、より正確に聞くことができるようにすることができる。

40

#### 【0057】

図11は、「次の東京行きのぞみ100号は10時30分に8番線から発車の予定です」に対応する、速度情報に応じた変更を行っていない通常の音声信号と、速度情報に応じた変更を行った変更後の音声信号とを比較するための波形図である。図11において、実線が通常の音声信号を示しており、破線が変更後の音声信号を示している。なお、真ん中のあたりでは、両音声信号が重なっているが、その領域が「10時30分に」に対応する領域である。

#### 【0058】

この具体例では、素片接続合成法により音声合成を行う場合について説明したが、HMM合成法により音声合成を行う場合にも、同様にして速度情報に応じた変更を行うことが

50

できる。図 1 2 は、HMM 合成法の場合における韻律情報の一例を示すものである。図 1 2 で示される韻律情報は、フルコンテキストラベルファイルと呼ばれるものの一部であり、各音素の開始時刻と、終了時刻、音素、品詞情報等が対応付けられている。音素のフィールドでは、音素が次のようにして示されている。

【 0 0 5 9 】

( 2 個前の音素 ) ^ ( 1 個前の音素 ) - ( 注目している音素 ) + ( 1 個後の音素 ) = ( 2 個後の音素 )

【 0 0 6 0 】

したがって、例えば、3 番目のレコードの音素のフィールドでは、2 個前の音素が「 s i l 」であり、1 個前の音素が「 t s 」であり、3 番目のレコードに対応する音素が「 u 」であり、1 個後の音素が「 g 」であり、2 個後の音素が「 i 」であることが示されている。

【 0 0 6 1 】

また、品詞情報等のフィールドでは、品詞情報、フレーズ内での位置、単語内での位置、単語の音節数などの情報が含まれている。このように、韻律情報には、持続時間情報や、基本周波数情報、パワースペクトル情報以外に、品詞や活用形、活用型に関する情報、先行する音素や後続する音素、音素のアクセント句内でのモーラ位置等の情報が含まれていてもよい。

【 0 0 6 2 】

韻律情報変更部 1 4 が図 1 2 で示される情報を変更する場合にも、図 3 のフローチャートを用いて説明した処理を実行することによって、開始時刻や終了時刻を変更することができ、速度情報に応じた変更を実現することができる。

【 0 0 6 3 】

以上のように、本実施の形態による音声合成装置 1 によれば、韻律情報変更部 1 4 が速度情報に応じて韻律情報を変更するため、テキスト情報の一部を、ユーザ等が指定した速度情報に対応する速度で読み上げることができるようになる。その結果、例えば、通常と同じ内容のテキストについては速いスピードで読み上げ、通常と異なる内容のテキストについては遅いスピードで読み上げるようにすることができる。そのため、通常と異なる内容について聞き逃すことを防止することができる。具体的には、音声による自動観光案内や、博物館、美術館、展示会場などにおける出展物の解説、駅や公共機関における放送などの際に、地名や人名、時刻などの情報については遅いスピードで読み上げるように設定することで、それらの情報の聞き逃しを回避することが期待できる。また、家庭用ゲーム機や家庭電化製品の使用方法を音声で説明する場合に、利用の回数に応じていつもと同じ内容については速いスピードで読み上げ、そうでない内容、すなわち、新規な内容については遅いスピードで読み上げるようにしてもよい。この場合には、例えば、出力された音声信号に対応するテキスト情報の履歴を保持しておき、あらかじめ決められた回数以上出力されている音声信号に対応するテキスト情報の部分については、他の部分に比べて速いスピードで読み上げられるように速度情報を変更する速度情報変更部 ( 図示せず ) を備えてもよい。なお、結果として、あらかじめ決められた回数以上出力されている音声信号に対応するテキスト情報の部分が、他の部分に比べて速いスピードで読み上げられるのであれば、速度情報の設定は問わない。例えば、速く読み上げるテキスト情報の部分について、速く読み上げる旨の速度情報を設定してもよく、あるいは、遅く読み上げるテキスト情報の部分について、遅く読み上げる旨の速度情報を設定してもよい。

【 0 0 6 4 】

なお、本実施の形態では、音声合成装置 1 において、テキスト情報から音素列と、韻律情報とを生成する場合について説明したが、そうでなくてもよい。例えば、音声合成装置 1 以外の装置において音素列と韻律情報とが生成され、その音素列等の情報が韻律情報記憶部 1 3 で記憶されるようになっていてもよい。その場合には、音声合成装置 1 は、音素列・韻律情報生成部 1 2 を備えていなくてもよい。また、その場合には、韻律情報記憶部 1 3 に音素列等が記憶される過程は問わない。例えば、記録媒体を介して音素列等が韻律情報

10

20

30

40

50

記憶部 13 で記憶されるようになってよく、あるいは、通信回線等を介して送信された音素列等が韻律情報記憶部 13 で記憶されるようになってよい。

【0065】

また、上記実施の形態では、音声合成装置 1 がスタンドアロンである場合について説明したが、音声合成装置 1 は、スタンドアロンの装置であってもよく、サーバ・クライアントシステムにおけるサーバ装置であってもよい。後者の場合には、出力部は、通信回線を介して情報を出力してもよい。

【0066】

また、上記実施の形態において、各処理または各機能は、単一の装置または単一のシステムによって集中処理されることによって実現されてもよく、あるいは、複数の装置または複数のシステムによって分散処理されることによって実現されてもよい。

10

【0067】

また、上記実施の形態において、各構成要素が実行する処理に係る情報、例えば、各構成要素が受け付けたり、取得したり、選択したり、生成したり、送信したり、受信したりした情報や、各構成要素が処理で用いるしきい値や数式、アドレス等の情報等は、上記説明で明記していない場合であっても、図示しない記録媒体において、一時的に、あるいは長期にわたって保持されていてもよい。また、その図示しない記録媒体への情報の蓄積を、各構成要素、あるいは、図示しない蓄積部が行ってもよい。また、その図示しない記録媒体からの情報の読み出しを、各構成要素、あるいは、図示しない読み出し部が行ってもよい。

20

【0068】

また、上記実施の形態において、各構成要素等で用いられる情報、例えば、各構成要素が処理で用いるしきい値やアドレス、各種の設定値等の情報がユーザによって変更されてもよい場合には、上記説明で明記していない場合であっても、ユーザが適宜、それらの情報を変更できるようにしてもよく、あるいは、そうでなくてもよい。それらの情報をユーザが変更可能な場合には、その変更は、例えば、ユーザからの変更指示を受け付ける図示しない受付部と、その変更指示に応じて情報を変更する図示しない変更部とによって実現されてもよい。その図示しない受付部による変更指示の受け付けは、例えば、入力デバイスからの受け付けでもよく、通信回線を介して送信された情報の受信でもよく、所定の記録媒体から読み出された情報の受け付けでもよい。

30

【0069】

また、上記実施の形態において、音声合成装置 1 に含まれる 2 以上の構成要素が通信デバイスや入力デバイス等を有する場合に、2 以上の構成要素が物理的に単一のデバイスを有してもよく、あるいは、別々のデバイスを有してもよい。

【0070】

また、上記実施の形態において、各構成要素は専用のハードウェアにより構成されてもよく、あるいは、ソフトウェアにより実現可能な構成要素については、プログラムを実行することによって実現されてもよい。例えば、ハードディスクや半導体メモリ等の記録媒体に記録されたソフトウェア・プログラムを CPU 等のプログラム実行部が読み出して実行することによって、各構成要素が実現され得る。なお、上記実施の形態における音声合成装置 1 を実現するソフトウェアは、以下のようなプログラムである。つまり、このプログラムは、コーパススペースの音声合成装置としてコンピュータを機能させるためのプログラムであって、コンピュータを、テキスト情報と当該テキスト情報の一部に対する速度に関する情報である速度情報とが記憶されるテキスト情報記憶部で記憶される前記速度情報に応じて、前記テキスト情報に対応する音素列と当該音素列の各音素の持続時間に関する情報である持続時間情報を少なくとも含む情報である韻律情報とが記憶される韻律情報記憶部で記憶される前記韻律情報を変更する韻律情報変更部、前記音素列と、前記韻律情報変更部による変更後の韻律情報とを用いて音声信号を生成する音声信号生成部、前記音声信号生成部が生成した音声信号を出力する音声信号出力部、として機能させるためのものである。

40

50

## 【 0 0 7 1 】

なお、上記プログラムにおいて、上記プログラムが実現する機能には、ハードウェアでしか実現できない機能は含まれない。例えば、情報を出力する出力部などにおけるモデムやインターフェースカードなどのハードウェアでしか実現できない機能は、上記プログラムが実現する機能には少なくとも含まれない。

## 【 0 0 7 2 】

また、このプログラムは、サーバなどからダウンロードされることによって実行されてもよく、所定の記録媒体（例えば、CD-ROMなどの光ディスクや磁気ディスク、半導体メモリなど）に記録されたプログラムが読み出されることによって実行されてもよい。また、このプログラムは、プログラムプロダクトを構成するプログラムとして用いられてもよい。

10

## 【 0 0 7 3 】

また、このプログラムを実行するコンピュータは、単数であってもよく、複数であってもよい。すなわち、集中処理を行ってもよく、あるいは分散処理を行ってもよい。

## 【 0 0 7 4 】

図13は、上記プログラムを実行して、上記実施の形態による音声合成装置1を実現するコンピュータの外観の一例を示す模式図である。上記実施の形態は、コンピュータハードウェア及びその上で実行されるコンピュータプログラムによって実現される。

## 【 0 0 7 5 】

図13において、コンピュータシステム900は、CD-ROM (Compact Disk Read Only Memory) ドライブ905、FD (Floppy (登録商標) Disk) ドライブ906を含むコンピュータ901と、キーボード902と、マウス903と、モニタ904とを備える。

20

## 【 0 0 7 6 】

図14は、コンピュータシステム900の内部構成を示す図である。図14において、コンピュータ901は、CD-ROMドライブ905、FDドライブ906に加えて、MPU (Micro Processing Unit) 911と、ブートアッププログラム等のプログラムを記憶するためのROM 912と、MPU 911に接続され、アプリケーションプログラムの命令を一時的に記憶すると共に、一時記憶空間を提供するRAM (Random Access Memory) 913と、アプリケーションプログラム、システムプログラム、及びデータを記憶するハードディスク914と、MPU 911、ROM 912等を相互に接続するバス915とを備える。なお、コンピュータ901は、LANへの接続を提供する図示しないネットワークカードを含んでいてもよい。

30

## 【 0 0 7 7 】

コンピュータシステム900に、上記実施の形態による音声合成装置1の機能を実行させるプログラムは、CD-ROM 921、またはFD 922に記憶されて、CD-ROMドライブ905、またはFDドライブ906に挿入され、ハードディスク914に転送されてもよい。これに代えて、そのプログラムは、図示しないネットワークを介してコンピュータ901に送信され、ハードディスク914に記憶されてもよい。プログラムは実行の際にRAM 913にロードされる。なお、プログラムは、CD-ROM 921やFD 922、またはネットワークから直接、ロードされてもよい。

40

## 【 0 0 7 8 】

プログラムは、コンピュータ901に、上記実施の形態による音声合成装置1の機能を実行させるオペレーティングシステム (OS)、またはサードパーティプログラム等を必ずしも含んでいなくてもよい。プログラムは、制御された態様で適切な機能 (モジュール) を呼び出し、所望の結果が得られるようにする命令の部分のみを含んでいてもよい。コンピュータシステム900がどのように動作するのかについては周知であり、詳細な説明は省略する。

## 【 0 0 7 9 】

また、本発明は、以上の実施の形態に限定されることなく、種々の変更が可能であり、

50

それらも本発明の範囲内に包含されるものであることは言うまでもない。

【産業上の利用可能性】

【0080】

以上より、本発明による音声合成装置等によれば、部分的に発話速度を変更することができるという効果が得られ、例えば、自動観光案内や、博物館、美術館、展示会場などにおける出展物の解説、駅や公共機関における放送などの音声合成システム等において有用である。

【図面の簡単な説明】

【0081】

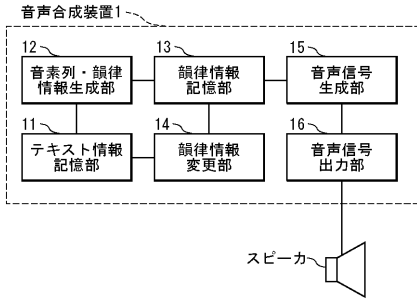
- 【図1】本発明の実施の形態1による音声合成装置の構成を示すブロック図 10
- 【図2】同実施の形態による音声合成装置の動作を示すフローチャート
- 【図3】同実施の形態による音声合成装置の動作を示すフローチャート
- 【図4】同実施の形態による音声合成装置の動作を示すフローチャート
- 【図5】同実施の形態におけるテキスト情報と速度情報との一例を示す図
- 【図6】同実施の形態における音素列と韻律情報の一例を示す図
- 【図7】同実施の形態における音素列と韻律情報の一例を示す図
- 【図8】同実施の形態における音素列と韻律情報の一例を示す図
- 【図9】同実施の形態における韻律情報の一例を示す図
- 【図10】同実施の形態における韻律情報の変更について説明するための図
- 【図11】同実施の形態における韻律情報の変更の前後の音声信号の一例を示す図 20
- 【図12】同実施の形態における音素列と韻律情報の一例を示す図
- 【図13】同実施の形態におけるコンピュータシステムの外観の一例を示す模式図
- 【図14】同実施の形態におけるコンピュータシステムの構成の一例を示す図

【符号の説明】

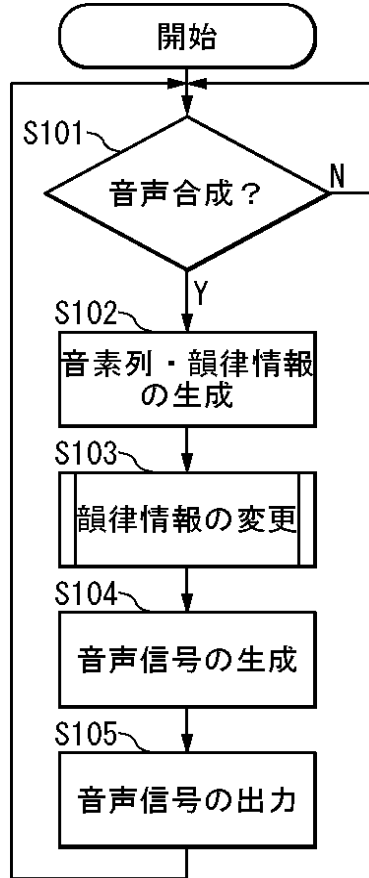
【0082】

- 1 音声合成装置
- 11 テキスト情報記憶部
- 12 韻律情報生成部
- 13 韻律情報記憶部
- 14 韻律情報変更部 30
- 15 音声信号生成部
- 16 音声信号出力部

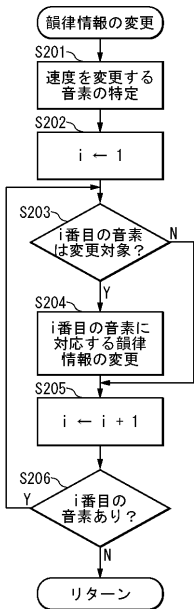
【図1】



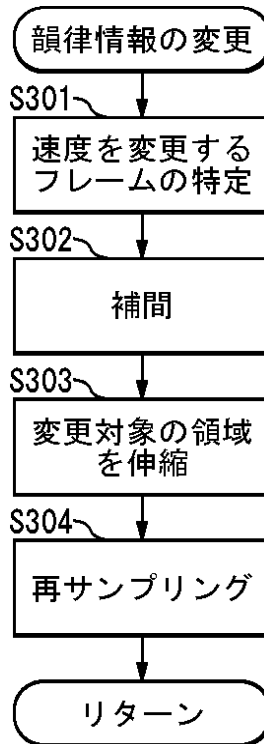
【図2】



【図3】



【図4】



【図5】

<fast>次の東京行きのみ100号は<keyword>10時30分に</keyword>8番線から発車の予定です</fast>

【図6】

開始時刻	音素	保護フラグ
0	sil	
0.105	ts	
0.18	u	
0.225	g	
0.27	i	
0.33	n	
0.375	o	
0.45	t	
0.535	o	
0.595	o	
0.655	ky	
0.75	o	
0.835	o	
0.905	i	
0.95	k	
1.045	i	
1.11	n	
1.165	o	
1.25	z	
1.315	o	
1.385	m	
1.445	i	
1.52	hy	
1.61	a	
1.69	k	
1.77	u	
1.825	g	
1.875	o	
1.955	o	
2.03	w	
2.1	a	
2.21	pau	
2.315	j	
2.395	u	
2.475	u	
2.535	j	
2.6	i	
2.67	s	
2.76	a	
2.855	N	
2.93	j	

開始時刻	音素	保護フラグ
2.975	u	
3.045	Q	
3.14	p	
3.175	u	
3.255	N	
3.3	n	
3.345	i	
3.475	pau	
3.575	h	
3.645	a	
3.705	ch	
3.81	i	
3.845	b	
3.89	a	
3.985	N	
4.035	s	
4.125	e	
4.205	N	
4.265	k	
4.345	a	
4.405	r	
4.43	a	
4.515	h	
4.585	a	
4.645	Q	
4.675	sh	
4.8	a	
4.865	n	
4.9	o	
4.975	y	
5.055	o	
5.115	t	
5.195	e	
5.27	e	
5.33	d	
5.375	e	
5.465	s	
5.55	U	
5.625	sil	
5.735	END	

【図7】

開始時刻	音素	保護フラグ
0	sil	
0.105	ts	
0.18	u	
0.225	g	
0.27	i	
0.33	n	
0.375	o	
0.45	t	
0.535	o	
0.595	o	
0.655	ky	
0.75	o	
0.835	o	
0.905	i	
0.95	k	
1.045	i	
1.11	n	
1.165	o	
1.25	z	
1.315	o	
1.385	m	
1.445	i	
1.52	hy	
1.61	a	
1.69	k	
1.77	u	
1.825	g	
1.875	o	
1.955	o	
2.03	w	
2.1	a	
2.21	pau	
2.315	j	*
2.395	u	*
2.475	u	*
2.535	j	*
2.6	i	*
2.67	s	*
2.76	a	*
2.855	N	*
2.93	j	*

開始時刻	音素	保護フラグ
2.975	u	*
3.045	Q	*
3.14	p	*
3.175	u	*
3.255	N	*
3.3	n	*
3.345	i	*
3.475	pau	
3.575	h	
3.645	a	
3.705	ch	
3.81	i	
3.845	b	
3.89	a	
3.985	N	
4.035	s	
4.125	e	
4.205	N	
4.265	k	
4.345	a	
4.405	r	
4.43	a	
4.515	h	
4.585	a	
4.645	Q	
4.675	sh	
4.8	a	
4.865	n	
4.9	o	
4.975	y	
5.055	o	
5.115	t	
5.195	e	
5.27	e	
5.33	d	
5.375	e	
5.465	s	
5.55	U	
5.625	sil	
5.735	_END_	

【図8】

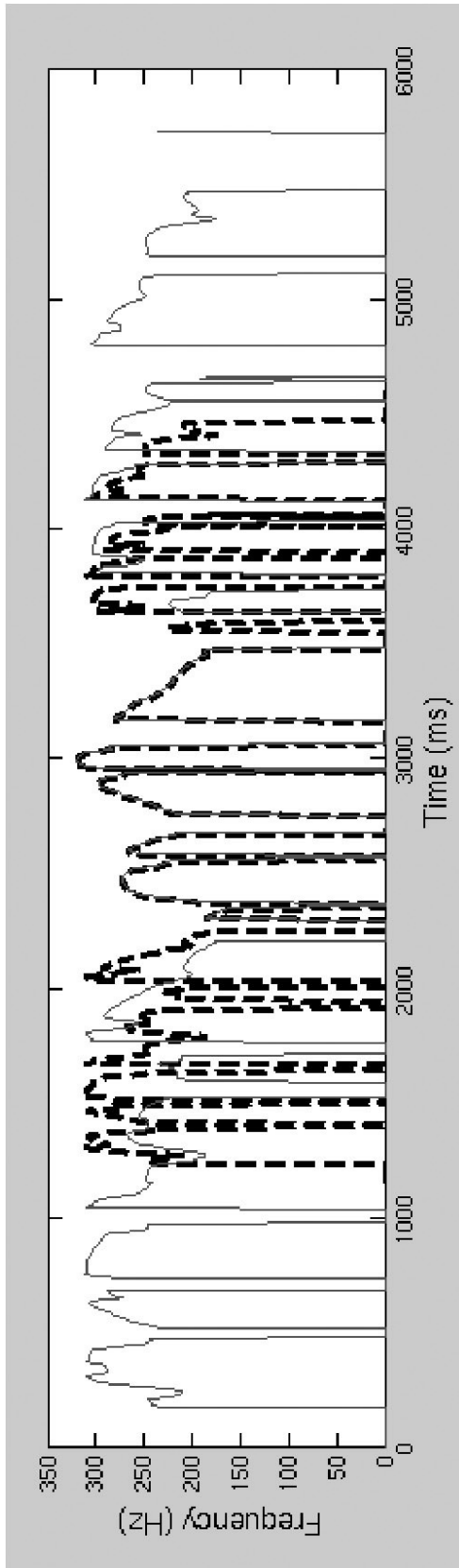
開始時刻	音素	保護フラグ
0	sil	
0.052	ts	
0.09	u	
0.113	g	
0.135	i	
0.165	n	
0.188	o	
0.225	t	
0.268	o	
0.297	o	
0.328	ky	
0.375	o	
0.417	o	
0.453	i	
0.475	k	
0.522	i	
0.555	n	
0.583	o	
0.625	z	
0.657	o	
0.693	m	
0.723	i	
0.76	hy	
0.805	a	
0.845	k	
0.885	u	
0.912	g	
0.938	o	
0.978	o	
1.015	w	
1.05	a	
1.105	pau	
1.157	j	*
1.238	u	*
1.318	u	*
1.378	j	*
1.443	i	*
1.512	s	*
1.602	a	*
1.698	N	*
1.773	j	*

開始時刻	音素	保護フラグ
1.818	u	*
1.887	Q	*
1.983	p	*
2.018	u	*
2.098	N	*
2.143	n	*
2.188	i	*
2.318	pau	
2.368	h	
2.403	a	
2.433	ch	
2.485	i	
2.503	b	
2.525	a	
2.573	N	
2.598	s	
2.643	e	
2.683	N	
2.713	k	
2.753	a	
2.783	r	
2.795	a	
2.838	h	
2.873	a	
2.903	Q	
2.918	sh	
2.98	a	
3.013	n	
3.03	o	
3.068	y	
3.108	o	
3.138	t	
3.178	e	
3.215	e	
3.245	d	
3.268	e	
3.313	s	
3.355	U	
3.393	sil	
3.448	_END_	





【図 11】



## フロントページの続き

- (56)参考文献 特開平06-337876(JP,A)  
特開平08-179789(JP,A)  
特開平10-274999(JP,A)  
特開平10-210154(JP,A)  
特開2003-208192(JP,A)  
河井 恒 Hisashi KAWAI,大規模コーパスを用いた音声合成システムXIMERA XIMERA:  
A Concatenative Speech Synthesis System with Large Scale Corpora,電子情報通信学会論文  
誌 (J89-D) 第12号 THE IEICE TRANSACTIONS ON INFORMATION AND SYSTEMS (JAPAN  
ESE EDITION),日本,社団法人電子情報通信学会 THE INSTITUTE OF ELECTRONICS,INFORMATION  
AND COMMUNICATION ENGINEERS,2006年12月 1日,p.2688-2698  
平井 俊男 Toshio Hirai,コーパス・ベース多言語音声合成システムXIMERA,日本音  
響学会 2006年 春季研究発表会講演論文集CD-ROM [CD-ROM],日本,社団  
法人日本音響学会,2006年 3月14日,p.333-334  
小林 隆夫 Takao KOBAYASHI,コーパスベース音声合成技術の動向[IV] Technology Tre  
nds in Corpus-based Speech Synthesis [IV],電子情報通信学会誌 第87巻 第4号 THE J  
OURNAL OF THE INSTITUTE OF ELECTRONICS,INFORMATION AND COMMUNICATION ENGINEERS,日本,  
社団法人電子情報通信学会 DENSHI-JOHO-TSUSHIN-GAKKAI,2004年,p.322-327
- (58)調査した分野(Int.Cl.,DB名)  
G10L 13/00-13/08