

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4032273号
(P4032273)

(45) 発行日 平成20年1月16日(2008.1.16)

(24) 登録日 平成19年11月2日(2007.11.2)

(51) Int. Cl. F I
B 2 5 J 13/00 (2006.01) B 2 5 J 13/00 Z
A 6 3 H 3/33 (2006.01) A 6 3 H 3/33 C

請求項の数 10 (全 15 頁)

(21) 出願番号	特願平11-373779	(73) 特許権者	000002185
(22) 出願日	平成11年12月28日(1999.12.28)		ソニー株式会社
(65) 公開番号	特開2001-179667(P2001-179667A)	(74) 代理人	100082131
(43) 公開日	平成13年7月3日(2001.7.3)		弁理士 稲本 義雄
審査請求日	平成18年3月1日(2006.3.1)	(72) 発明者	山田 敬一
			東京都品川区北品川6丁目7番35号 ソニー株式会社内
		(72) 発明者	小林 賢一郎
			東京都品川区北品川6丁目7番35号 ソニー株式会社内
		(72) 発明者	新田 朋晃
			東京都品川区北品川6丁目7番35号 ソニー株式会社内

最終頁に続く

(54) 【発明の名称】 同期制御装置および方法、並びに記録媒体

(57) 【特許請求の範囲】

【請求項1】

音声信号の出力と可動部位の動作を同期させる同期制御装置において、
 言語情報を用いて複数の音韻からなる音韻情報を生成する音韻情報生成手段と、
 前記音韻情報生成手段が生成した前記音韻情報に対応して、音韻継続時間を算出する算出手段と、

前記音韻情報生成手段が生成した前記音韻情報に対応して、前記可動部位の動作時間を演算する演算手段と、

前記算出手段が算出した前記音韻継続時間と、前記演算手段が演算した前記動作時間を調整する調整手段と、

前記調整手段が調整した音韻継続時間に基づいて、合成音声情報を生成する合成音声情報生成手段と、

前記合成音声情報生成手段が生成した前記合成音声情報に基づいて、前記音声信号を合成する合成手段と、

前記調整手段が調整した動作時間に基づいて、前記可動部位の動作を制御する動作制御手段とを含み、

前記調整手段は、

前記音韻毎に対応する前記音韻継続時間と前記動作時間を比較して、長い方を用いて短い方を置換して調整する処理、

任意の前記音韻に対応する前記音韻継続時間と前記動作時間の開始タイミングまたは

終了タイミングのうち、少なくとも一方を揃えて調整する処理、

全ての前記音韻に対応する前記音韻継続時間と前記動作時間のうちの一方を、他方で置換して調整する処理、

前記音韻毎に対応する前記音韻継続時間と前記動作時間の開始タイミングまたは終了タイミングのうち、少なくとも一方を揃え、不足する時間を無処理時間で補充して調整する処理、

または、全ての前記音韻に対応する前記音韻継続時間と前記動作時間を比較して、短い方を比例的に伸張して調整する処理

のうちの少なくとも1つの処理を行う

ことを特徴とする同期制御装置。

10

【請求項2】

前記同期制御装置は、ロボットである

ことを特徴とする請求項1に記載の同期制御装置。

【請求項3】

前記ロボットの口の内部に物体が挿入されたことを検知する検知手段をさらに含み、

前記合成手段と前記動作制御手段のうち、少なくとも一方は、前記検知手段の検知結果に対応して実行中の処理を変更する

ことを特徴とする請求項2に記載の同期制御装置。

【請求項4】

音声信号の出力と可動部位の動作を同期させる同期制御装置の同期制御方法において、
言語情報を用いて複数の音韻からなる音韻情報を生成する音韻情報生成ステップと、
前記音韻情報生成ステップの処理で生成された前記音韻情報に対応して、音韻継続時間を算出する算出ステップと、

20

前記音韻情報生成ステップの処理で生成された前記音韻情報に対応して、前記可動部位の動作時間を演算する演算ステップと、

前記算出ステップの処理で算出された前記音韻継続時間と、前記演算ステップの処理で演算された前記動作時間を調整する調整ステップと、

前記調整ステップの処理で調整された音韻継続時間に基づいて、合成音声情報を生成する合成音声情報生成ステップと、

前記合成音声情報生成ステップの処理で生成された前記合成音声情報に基づいて、前記音声信号を合成する合成ステップと、

30

前記調整ステップの処理で調整された動作時間に基づいて、前記可動部位の動作を制御する動作制御ステップとを含み、

前記調整ステップは、

前記音韻毎に対応する前記音韻継続時間と前記動作時間を比較して、長い方を用いて短い方を置換して調整する処理、

任意の前記音韻に対応する前記音韻継続時間と前記動作時間の開始タイミングまたは終了タイミングのうち、少なくとも一方を揃えて調整する処理、

全ての前記音韻に対応する前記音韻継続時間と前記動作時間のうちの一方を、他方で置換して調整する処理、

40

前記音韻毎に対応する前記音韻継続時間と前記動作時間の開始タイミングまたは終了タイミングのうち、少なくとも一方を揃え、不足する時間を無処理時間で補充して調整する処理、

または、全ての前記音韻に対応する前記音韻継続時間と前記動作時間を比較して、短い方を比例的に伸張して調整する処理

のうちの少なくとも1つの処理を行う

ことを特徴とする同期制御方法。

【請求項5】

音声信号の出力と可動部位の動作を同期させる同期制御装置の制御用のプログラムであって、

50

言語情報を用いて複数の音韻からなる音韻情報を生成する音韻情報生成ステップと、
前記音韻情報生成ステップの処理で生成された前記音韻情報に対応して、音韻継続時間を算出する算出ステップと、

前記音韻情報生成ステップの処理で生成された前記音韻情報に対応して、前記可動部位の動作時間を演算する演算ステップと、

前記算出ステップの処理で算出された前記音韻継続時間と、前記演算ステップの処理で演算された前記動作時間を調整する調整ステップと、

前記調整ステップの処理で調整された音韻継続時間に基づいて、合成音声情報を生成する合成音声情報生成ステップと、

前記合成音声情報生成ステップの処理で生成された前記合成音声情報に基づいて、前記音声信号を合成する合成ステップと、

前記調整ステップの処理で調整された動作時間に基づいて、前記可動部位の動作を制御する動作制御ステップとを含み、

前記調整ステップは、

前記音韻毎に対応する前記音韻継続時間と前記動作時間を比較して、長い方を用いて短い方を置換して調整する処理、

任意の前記音韻に対応する前記音韻継続時間と前記動作時間の開始タイミングまたは終了タイミングのうち、少なくとも一方を揃えて調整する処理、

全ての前記音韻に対応する前記音韻継続時間と前記動作時間のうちの一方を、他方で置換して調整する処理、

前記音韻毎に対応する前記音韻継続時間と前記動作時間の開始タイミングまたは終了タイミングのうち、少なくとも一方を揃え、不足する時間を無処理時間で補充して調整する処理、

または、全ての前記音韻に対応する前記音韻継続時間と前記動作時間を比較して、短い方を比例的に伸張して調整する処理

のうちの少なくとも1つの処理を行う

ことを特徴とする処理を同期制御装置のコンピュータに実行させるプログラムが記録されている記録媒体。

【請求項6】

音声信号の出力と可動部位の動作を同期させる同期制御装置において、

言語情報を用いて複数の音韻からなる音韻情報を生成する音韻情報生成手段と、

前記音韻情報生成手段が生成した前記音韻情報に対応して、音韻継続時間を算出する算出手段と、

前記音韻情報生成手段が生成した前記音韻情報に対応して、前記可動部位の動作時間を演算する演算手段と、

前記算出手段が算出した前記音韻継続時間と、前記演算手段が演算した前記動作時間を調整する調整手段と、

前記調整手段が調整した音韻継続時間に基づいて、合成音声情報を生成する合成音声情報生成手段と、

前記合成音声情報生成手段が生成した前記合成音声情報に基づいて、前記音声信号を合成する合成手段と、

前記調整手段が調整した動作時間に基づいて、前記可動部位の動作を制御する動作制御手段とを含み、

前記動作制御手段は、動物の調音器官の動きを模擬した前記可動部位の動作を制御することを特徴とする同期制御装置。

【請求項7】

前記同期制御装置は、ロボットである

ことを特徴とする請求項6に記載の同期制御装置。

【請求項8】

前記ロボットの口の内部に物体が挿入されたことを検知する検知手段をさらに含み、

10

20

30

40

50

前記合成手段と前記動作制御手段のうち、少なくとも一方は、前記検知手段の検知結果に対応して実行中の処理を変更する

ことを特徴とする請求項 7 に記載の同期制御装置。

【請求項 9】

音声信号の出力と可動部位の動作を同期させる同期制御装置の同期制御方法において、言語情報を用いて複数の音韻からなる音韻情報を生成する音韻情報生成ステップと、

前記音韻情報生成ステップの処理で生成された前記音韻情報に対応して、音韻継続時間を算出する算出ステップと、

前記音韻情報生成ステップの処理で生成された前記音韻情報に対応して、前記可動部位の動作時間を演算する演算ステップと、

前記算出ステップの処理で算出された前記音韻継続時間と、前記演算ステップの処理で演算された前記動作時間を調整する調整ステップと、

前記調整ステップの処理で調整された音韻継続時間に基づいて、合成音声情報を生成する合成音声情報生成ステップと、

前記合成音声情報生成ステップの処理で生成された前記合成音声情報に基づいて、前記音声信号を合成する合成ステップと、

前記調整ステップの処理で調整された動作時間に基づいて、前記可動部位の動作を制御する動作制御ステップとを含み、

前記動作制御ステップは、動物の調音器官の動きを模擬した前記可動部位の動作を制御する

ことを特徴とする同期制御方法。

【請求項 10】

音声信号の出力と可動部位の動作を同期させる同期制御装置の制御用のプログラムであって、

言語情報を用いて複数の音韻からなる音韻情報を生成する音韻情報生成ステップと、

前記音韻情報生成ステップの処理で生成された前記音韻情報に対応して、音韻継続時間を算出する算出ステップと、

前記音韻情報生成ステップの処理で生成された前記音韻情報に対応して、前記可動部位の動作時間を演算する演算ステップと、

前記算出ステップの処理で算出された前記音韻継続時間と、前記演算ステップの処理で演算された前記動作時間を調整する調整ステップと、

前記調整ステップの処理で調整された音韻継続時間に基づいて、合成音声情報を生成する合成音声情報生成ステップと、

前記合成音声情報生成ステップの処理で生成された前記合成音声情報に基づいて、前記音声信号を合成する合成ステップと、

前記調整ステップの処理で調整された動作時間に基づいて、前記可動部位の動作を制御する動作制御ステップとを含み、

前記動作制御ステップは、動物の調音器官の動きを模擬した前記可動部位の動作を制御する

ことを特徴とする処理を同期制御装置のコンピュータに実行させるプログラムが記録されている記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、同期制御装置および方法、並びに記録媒体に関し、例えば、ロボットの頭部に設けられる調音器官の動きを模擬した部位の動作と、合成した音声の出力を同期させる場合に用いて好適な同期制御装置および方法、並びに記録媒体に関する。

【0002】

【従来の技術】

人や動物を模擬したロボットには、口や顎等を模擬した部位が可動するようなもの（例え

10

20

30

40

50

ば、顎を上下することにより口を開閉するようなもの)が存在し、口や顎等の動きとともに音声出力するものも存在する。

【0003】

【発明が解決しようとする課題】

そのようなロボットにおいて、例えば、「あ」の音の出力タイミングのときには、口等が「あ」の音を発声する形をし、「い」の音の出力タイミングのときには、口等が「い」の音を発声する形をするように、発声する言葉に対応して口等を動作するようにすれば、よりリアルに人間等を模擬することになる。しかしながら、そのようなロボットは存在していない。

【0004】

本発明はこのような状況に鑑みてなされたものであり、音声合成による発声する言葉と発声タイミングに、調音器官を模擬した部位の動作に対応させることにより、よりリアルに人間等を模擬したロボットを実現することを目的とする。

【0005】

【課題を解決するための手段】

本発明の第1の同期制御装置は、言語情報を用いて複数の音韻からなる音韻情報を生成する音韻情報生成手段と、音韻情報生成手段が生成した音韻情報に対応して、音韻継続時間を算出する算出手段と、音韻情報生成手段が生成した音韻情報に対応して、可動部位の動作時間を演算する演算手段と、算出手段が算出した音韻継続時間と、演算手段が演算した動作時間を調整する調整手段と、調整手段が調整した音韻継続時間に基づいて、合成音声情報を生成する合成音声情報生成手段と、合成音声情報生成手段が生成した合成音声情報に基づいて、音声信号を合成する合成手段と、調整手段が調整した動作時間に基づいて、可動部位の動作を制御する動作制御手段とを含み、調整手段は、音韻毎に対応する音韻継続時間と動作時間を比較して、長い方を用いて短い方を置換して調整する処理、任意の音韻に対応する音韻継続時間と動作時間の開始タイミングまたは終了タイミングのうち、少なくとも一方を揃えて調整する処理、全ての音韻に対応する音韻継続時間と動作時間のうちの一方を、他方で置換して調整する処理、音韻毎に対応する音韻継続時間と動作時間の開始タイミングまたは終了タイミングのうち、少なくとも一方を揃え、不足する時間を無処理時間で補充して調整する処理、または、全ての音韻に対応する音韻継続時間と動作時間を比較して、短い方を比例的に伸張して調整する処理のうちの少なくとも1つの処理を行うことを特徴とする。

前記同期制御装置は、ロボットであるようにすることができる。

本発明の第1の同期制御装置は、ロボットの口の内部に物体が挿入されたことを検知する検知手段をさらに含むことができ、前記合成手段と前記動作制御手段のうち、少なくとも一方は、検知手段の検知結果に対応して実行中の処理を変更するようにすることができる。

【0006】

本発明の第1の同期制御方法は、言語情報を用いて複数の音韻からなる音韻情報を生成する音韻情報生成ステップと、音韻情報生成ステップの処理で生成された音韻情報に対応して、音韻継続時間を算出する算出ステップと、音韻情報生成ステップの処理で生成された音韻情報に対応して、可動部位の動作時間を演算する演算ステップと、算出ステップの処理で算出された音韻継続時間と、演算ステップの処理で演算された動作時間を調整する調整ステップと、調整ステップの処理で調整された音韻継続時間に基づいて、合成音声情報を生成する合成音声情報生成ステップと、合成音声情報生成ステップの処理で生成された合成音声情報に基づいて、音声信号を合成する合成ステップと、調整ステップの処理で調整された動作時間に基づいて、可動部位の動作を制御する動作制御ステップとを含み、調整ステップは、音韻毎に対応する音韻継続時間と動作時間を比較して、長い方を用いて短い方を置換して調整する処理、任意の音韻に対応する音韻継続時間と動作時間の開始タイミングまたは終了タイミングのうち、少なくとも一方を揃えて調整する処理、全ての音韻に対応する音韻継続時間と動作時間のうちの一方を、他方で置換して調整する処理、音

10

20

30

40

50

韻毎に対応する音韻継続時間と動作時間の開始タイミングまたは終了タイミングのうち、少なくとも一方を揃え、不足する時間を無処理時間で補充して調整する処理、または、全ての音韻に対応する音韻継続時間と動作時間を比較して、短い方を比例的に伸張して調整する処理のうちの少なくとも1つの処理を行うことを特徴とする。

【0007】

本発明の第1の記録媒体は、言語情報を用いて複数の音韻からなる音韻情報を生成する音韻情報生成ステップと、音韻情報生成ステップの処理で生成された音韻情報に対応して、音韻継続時間を算出する算出ステップと、音韻情報生成ステップの処理で生成された音韻情報に対応して、可動部位の動作時間を演算する演算ステップと、算出ステップの処理で算出された音韻継続時間と、演算ステップの処理で演算された動作時間を調整する調整ステップと、調整ステップの処理で調整された音韻継続時間に基づいて、合成音声情報を生成する合成音声情報生成ステップと、合成音声情報生成ステップの処理で生成された合成音声情報に基づいて、音声信号を合成する合成ステップと、調整ステップの処理で調整された動作時間に基づいて、可動部位の動作を制御する動作制御ステップとを含み、調整ステップは、音韻毎に対応する音韻継続時間と動作時間を比較して、長い方を用いて短い方を置換して調整する処理、任意の音韻に対応する音韻継続時間と動作時間の開始タイミングまたは終了タイミングのうち、少なくとも一方を揃えて調整する処理、全ての音韻に対応する音韻継続時間と動作時間のうちの一方を、他方で置換して調整する処理、音韻毎に対応する音韻継続時間と動作時間の開始タイミングまたは終了タイミングのうち、少なくとも一方を揃え、不足する時間を無処理時間で補充して調整する処理、または、全ての音韻に対応する音韻継続時間と動作時間を比較して、短い方を比例的に伸張して調整する処理のうちの少なくとも1つの処理を行うことを特徴とする処理を同期制御装置のコンピュータに実行させるプログラムが記録されている。

【0008】

本発明の第1の同期制御装置、同期制御方法、および記録媒体のプログラムにおいては、言語情報を用いて複数の音韻からなる音韻情報が生成され、生成された音韻情報に対応して音韻継続時間が算出される。また、生成された音韻情報に対応して可動部位の動作時間が演算される。また、算出された音韻継続時間と演算された動作時間が調整され、調整された音韻継続時間に基づいて合成音声情報が生成され、生成された合成音声情報に基づいて音声信号が合成される。さらに、調整された動作時間に基づいて可動部位の動作が制御される。そして、前記調整では、音韻毎に対応する音韻継続時間と動作時間を比較して、長い方を用いて短い方を置換して調整する処理、任意の音韻に対応する音韻継続時間と動作時間の開始タイミングまたは終了タイミングのうち、少なくとも一方を揃えて調整する処理、全ての音韻に対応する音韻継続時間と動作時間のうちの一方を、他方で置換して調整する処理、音韻毎に対応する音韻継続時間と動作時間の開始タイミングまたは終了タイミングのうち、少なくとも一方を揃え、不足する時間を無処理時間で補充して調整する処理、または、全ての音韻に対応する音韻継続時間と動作時間を比較して、短い方を比例的に伸張して調整する処理のうちの少なくとも1つの処理が行われる。

【0009】

本発明の第2の同期制御装置は、言語情報を用いて複数の音韻からなる音韻情報を生成する音韻情報生成手段と、音韻情報生成手段が生成した音韻情報に対応して、音韻継続時間を算出する算出手段と、音韻情報生成手段が生成した音韻情報に対応して、可動部位の動作時間を演算する演算手段と、算出手段が算出した音韻継続時間と、演算手段が演算した動作時間を調整する調整手段と、調整手段が調整した音韻継続時間に基づいて、合成音声情報を生成する合成音声情報生成手段と、合成音声情報生成手段が生成した合成音声情報に基づいて、音声信号を合成する合成手段と、調整手段が調整した動作時間に基づいて、可動部位の動作を制御する動作制御手段とを含み、動作制御手段は、動物の調音器官の動きを模擬した可動部位の動作を制御することを特徴とする。

同期制御装置は、ロボットであるようにすることができる。

本発明の第2の同期制御装置は、ロボットの口の内部に物体が挿入されたことを検知す

10

20

30

40

50

る検知手段をさらに含むことができ、前記合成手段と前記動作制御手段のうち、少なくとも一方は、検知手段の検知結果に対応して実行中の処理を変更するようにすることができる。

【0010】

本発明の第2の同期制御方法は、言語情報を用いて複数の音韻からなる音韻情報を生成する音韻情報生成ステップと、音韻情報生成ステップの処理で生成された音韻情報に対応して、音韻継続時間を算出する算出ステップと、音韻情報生成ステップの処理で生成された音韻情報に対応して、可動部位の動作時間を演算する演算ステップと、算出ステップの処理で算出された音韻継続時間と、演算ステップの処理で演算された動作時間を調整する調整ステップと、調整ステップの処理で調整された音韻継続時間に基づいて、合成音声情報を生成する合成音声情報生成ステップと、合成音声情報生成ステップの処理で生成された合成音声情報に基づいて、音声信号を合成する合成ステップと、調整ステップの処理で調整された動作時間に基づいて、可動部位の動作を制御する動作制御ステップとを含み、動作制御ステップは、動物の調音器官の動きを模擬した可動部位の動作を制御することを特徴とする。

10

【0011】

本発明の第2の記録媒体は、言語情報を用いて複数の音韻からなる音韻情報を生成する音韻情報生成ステップと、音韻情報生成ステップの処理で生成された音韻情報に対応して、音韻継続時間を算出する算出ステップと、音韻情報生成ステップの処理で生成された音韻情報に対応して、可動部位の動作時間を演算する演算ステップと、算出ステップの処理で算出された音韻継続時間と、演算ステップの処理で演算された動作時間を調整する調整ステップと、調整ステップの処理で調整された音韻継続時間に基づいて、合成音声情報を生成する合成音声情報生成ステップと、合成音声情報生成ステップの処理で生成された合成音声情報に基づいて、音声信号を合成する合成ステップと、調整ステップの処理で調整された動作時間に基づいて、可動部位の動作を制御する動作制御ステップとを含み、動作制御ステップは、動物の調音器官の動きを模擬した可動部位の動作を制御することを特徴とする処理を同期制御装置のコンピュータに実行させるプログラムが記録されている。

20

【0012】

本発明の第2の同期制御装置、同期制御方法、および記録媒体のプログラムにおいては、言語情報を用いて複数の音韻からなる音韻情報が生成され、生成された音韻情報に対応して音韻継続時間が算出される。また、生成された音韻情報に対応して可動部位の動作時間が演算される。また、算出された音韻継続時間と演算された動作時間が調整され、調整された音韻継続時間に基づいて合成音声情報が生成され、生成された合成音声情報に基づいて音声信号が合成される。さらに、調整された動作時間に基づいて可動部位の動作が制御される。そして、前記動作制御では、動物の調音器官の動きを模擬した可動部位の動作が制御される。

30

【0017】

【発明の実施の形態】

図1は、本発明を適用したロボットの音声出力、および、顎、唇、喉、舌、鼻孔などの調音器官を模擬した部位の動作を制御する部分の構成例を示している。この構成例は、例えばロボットの頭部に設けられる。

40

【0018】

入力部1は、マイクロフォンと音声認識機能（いずれも不図示）を有し、マイクロフォンに入力される音声信号（例えば、「こんにちは」など、ロボットに復唱させる言葉、または、ロボットに対して問い掛ける言葉）を音声認識機能によりテキストデータに変換して、音声言語情報生成部2に供給する。なお、外部から音声言語情報生成部2に、テキストデータを入力するようにしてもよい。

【0019】

音声言語情報生成部2は、ロボットに対話させるときにおいて、入力部1から入力されるテキストデータに対して返事となる言葉の音声言語情報（発声する単語を表すもの）を

50

生成して制御部 3 に出力する。音声言語情報生成部 2 はまた、ロボットに復唱させるときにおいて、入力部 1 から入力されるテキストデータをそのまま制御部 3 に出力する。音声言語情報は仮名文字、英文字、発音記号などのテキストデータを用いて表される。

【 0 0 2 0 】

制御部 3 は、ドライブ 1 1 を制御して、磁気ディスク 1 2、光ディスク 1 3、光磁気ディスク 1 4、または半導体メモリ 1 5 に記憶されている制御用プログラムを読み出し、読み出した制御用プログラムに基づいて各部を制御する。

【 0 0 2 1 】

具体的には、制御部 3 は、音声言語情報生成部 2 から入力される音声言語情報としてのテキストデータを音声合成部 4 に供給し、音声合成部 4 からの音韻情報を調音動作生成部 5 に供給し、調音動作生成部 5 からの調音動作時間と、音声合成部 4 からの音韻情報および音韻継続時間を音声動作調整部 6 に供給する。制御部 3 はまた、音声動作調整部 6 からの調整済みの音韻継続時間を音声合成部 4 に供給し、音声動作調整部 6 からの調整済みの調音動作時間を調音動作実行部 7 に供給する。制御部 3 はさらに、音声合成部 4 からの合成音声データを音声出力部 9 に供給する。制御部 3 はさらに、外部センサ 8 からの検出情報に対応して、調音動作実行部 7 および音声出力部 9 の処理を停止、再開、または中断する。

10

【 0 0 2 2 】

音声合成部 4 は、制御部 3 から入力される、音声言語情報生成部 2 からの音声言語情報としてのテキストデータ（例えば、「こんにちは」）から、図 2 に示すように、音韻情報（いまの場合、「K O X N I C H I W A」）を生成し、各音韻の音韻継続時間を算出して制御部 3 に出力する。音声合成部 4 はまた、制御部 3 から入力される、音声動作調整部 6 からの調整済みの音韻継続時間に基づき、合成音声データを生成する。生成される合成音声データは、一般的に知られる規則合成音声によるものや録音再生音によるものなどが挙げられる。

20

【 0 0 2 3 】

調音動作生成部 5 は、制御部 3 から入力される、音声合成部 4 からの音韻情報に基づき、図 3 に示すように、各音韻に対応する調音動作命令（各調音器官を模擬した部位の動作を指示する命令）と、当該動作の動作時間である調音動作時間を算出して制御部 3 に出力する。図 3 の例においては、調音器官 1 6 として、顎、唇、喉、舌、鼻孔が記述され、調音動作命令として、顎の上下移動、唇の形状変化および開閉、舌の前後上下左右の移動、喉の振幅および上下移動、鼻の形状変化が記述されている。調音動作命令は、調音器官 1 6 のうちの 1 つに対して単独に指示するようにしてもよいし、調音器官 1 6 の複数を組み合わせて指示するようにしてもよい。

30

【 0 0 2 4 】

音声動作調整部 6 は、制御部 3 から入力される、音声合成部 4 からの音韻継続時間と調音動作生成部 5 からの調音動作時間を、所定の方法（詳細は後述する）に従って調整し、制御部 3 に出力する。例えば、音韻継続時間と調音動作時間を、音韻毎に時間が長い方を用いて短い方を置換するという方法に従って、図 2 に示した音韻継続時間と、図 3 に示した調音動作時間を調整する場合、音韻継続時間については、図 4 に示すように音韻「X」、「I」、および「W」の各音韻継続時間が調音動作時間と等しくなるように延長される。

40

【 0 0 2 5 】

調音動作実行部 7 は、制御部 3 から入力される、調音動作生成部 5 からの調音動作命令と、音声動作調整部 6 からの調整済みの調音動作時間に基づき、調音器官 1 6 を動作させる。

【 0 0 2 6 】

外部センサ 8 は、例えば、調音器官 1 6 に含まれる口の内部に設けられ、口内に挿入された物体を検知して検知情報を制御部 3 に出力する。

【 0 0 2 7 】

音声出力部 9 は、制御部 3 から入力される、音声合成部 4 からの合成音声データに対応す

50

る音声をスピーカ 10 に放音させる。

【0028】

調音器官 16 は、ロボットの頭部に設けられる顎、唇、喉、舌、鼻孔などを模擬した可動部位である。

【0029】

次に、ロボットの動作について、図 5 のフローチャートを参照して説明する。ステップ S 1 において、入力部 1 のマイクロフォンに入力された音声信号は、テキストデータに変換されて音声言語情報生成部 2 に供給される。ステップ S 2 において、音声言語情報生成部 2 は、入力部 1 から入力されたテキストデータに対応する音声言語情報を制御部 3 に出力する。制御部 3 は、音声言語情報生成部 2 から入力された音声言語情報としてのテキストデータ（例えば、「こんにちは」）を音声合成部 4 に供給する。

10

【0030】

ステップ S 3 において、音声合成部 4 は、制御部 3 から供給された音声言語情報生成部 2 からの音声言語情報としてのテキストデータから、音韻情報（いまの場合、「K O X N I C H I W A」）を生成し、各音韻の音韻継続時間を算出して制御部 3 に出力する。制御部 3 は、音声合成部 4 からの音韻情報を調音動作生成部 5 に供給する。

【0031】

ステップ S 4 において、調音動作生成部 5 は、制御部 3 から供給された音声合成部 4 からの音韻情報に基づき、各音韻に対応する調音動作命令と、調音動作時間を算出して制御部 3 に出力する。制御部 3 は、調音動作生成部 5 からの調音動作時間と、音声合成部 4 からの音韻情報および音韻継続時間を音声動作調整部 6 に供給する。

20

【0032】

ステップ S 5 において、音声動作調整部 6 は、制御部 3 から供給された音声合成部 4 からの音韻継続時間と調音動作生成部 5 からの調音動作時間を、所定の規則に従って調整し、制御部 3 に出力する。

【0033】

ここで、音韻継続時間と調音動作時間を調整する第 1 乃至第 5 の方法について、図 6 乃至図 11 を参照して説明する。なお、以下の説明においては、ステップ S 3 で生成された音韻継続時間を図 6 (A) に示すものとし、ステップ S 4 で生成された調音動作時間を図 6 (B) に示すものとする。

30

【0034】

第 1 の方法は、各音韻の音韻継続時間と調音動作時間を比較して、長い方の時間を用いて短い方の時間を置き換えて調整する方法である。図 7 は、第 1 の方法による調整結果を示している。例えば、図 6 の例において、音韻「K」、「CH」、「W」については、音韻継続時間の方が調音動作時間よりも長いので、図 7 (B) に示すように、音韻継続時間が調音動作時間で置き換えられる。反対に、図 6 の例において、音韻「O」、「X」、「N」、「I」、「I」、および「A」については、調音動作時間の方が音韻継続時間よりも長いので、図 7 (A) に示すように、調音動作時間が音韻継続時間で置き換えられる。

【0035】

第 2 の方法は、任意の音韻の開始タイミングまたは終了タイミングを同期させて調整する方法である。図 8 は、第 2 の方法による調整結果を示している。例えば、図 8 に示すように、音韻「X」の開始タイミングを同期させた場合、音韻「K」の音韻継続時間の開始タイミング以前と、音韻「A」の音韻継続時間の終了タイミング以降に不足部分が生じるが、それらの間は音声を発声せず、調音動作のみ実行するように調整する。なお、開始タイミングを同期させる音韻の指定は、ユーザが指定するようにしてもよいし、例えば、制御部 3 が所定の規則に基づいて決定するようにしてもよい。

40

【0036】

第 3 の方法は、全ての音韻に対して音韻継続時間または調音動作時間のうち、一方を他方で置き換えて調整する方法である。図 9 は、第 3 の方法により、調音動作時間を優先し、全ての音韻に対して音韻継続時間を調音動作時間で置き換えることによって調整した結果

50

を示している。音韻継続時間または調音動作時間のうち、いずれを優先させるかについては、ユーザが選択するようにしてもよいし、例えば、制御部 3 が所定の規則に基づいて選択するようにしてもよい。

【 0 0 3 7 】

第 4 の方法は、各音韻の開始タイミングまたは終了タイミングを音韻継続時間と調音動作時間で同期させ、不足する時間を空白（発声や調音動作を実行しない時間を指す）で埋めることにより調整する方法である。図 10 は、第 4 の方法による調整結果を示している。例えば、図 10 (B) に示すように、音韻「K」については、調音動作時間の開始タイミング以前に生じる不足する時間を空白で埋め、図 10 (A) に示すように、音韻「O」、「X」、「N」、および「I」については、音韻継続時間の開始タイミング以前に生じる不足する時間を空白で埋めるようにする。

10

【 0 0 3 8 】

第 5 の方法は、音韻情報の中心に位置する音韻の開始タイミングまたは終了タイミングを揃えて、かつ、全音韻継続時間と全調音動作時間を比較して短い方の時間を長い方の時間に等しくなるように比例的に伸張する方法である。具体的には、例えば、音韻情報「K O X N I C H I W A」の中心に位置する音韻「I」の開始タイミングを揃え、いまの場合、全音韻継続時間（300 ミリ秒）の方が調音動作時間（550 ミリ秒）よりも短いので、音韻継続時間を 550 ミリ秒まで伸張する。より具体的には、音韻「I」以前の音韻「K」、「O」、「X」、および「N」の音韻継続時間を 2（= 300 / 150）倍に伸張し、音韻「I」以降の音韻「I」、「CH」、「I」、「W」、および「A」の音韻継続時間を 1.25（= 250 / 200）倍に伸張する。

20

【 0 0 3 9 】

以上のように、第 1 乃至第 4 の方法のうちのいずれかの方法、あるいは第 1 乃至第 4 の方法を適宜組み合わせた方法によって、音韻継続時間と調音動作時間は調整されて制御部 3 に供給される。

【 0 0 4 0 】

図 5 に戻り、ステップ S 6 において、制御部 3 は、音声動作調整部 6 からの調整済みの音韻継続時間を音声合成部 4 に供給し、音声動作調整部 6 からの調整済みの調音動作時間と調音動作生成部 5 からの調音動作命令を調音動作実行部 7 に供給する。音声合成部 4 は、制御部 3 から入力された音声動作調整部 6 からの調整済みの音韻継続時間に基づき、合成音声データを生成して制御部 3 に出力する。制御部 3 はまた、音声合成部 4 からの合成音声データを音声出力部 9 に供給する。音声出力部 9 は、制御部 3 から入力された、音声合成部 4 からの合成音声データに対応する音声をスピーカ 10 に放音させる。それと同期して、調音動作実行部 7 は、制御部 3 から入力された、調音動作生成部 5 からの調音動作命令と、音声動作調整部 6 からの調整済みの調音動作時間に基づき、調音器官 16 を動作する。

30

【 0 0 4 1 】

以上のように、ロボットが動作することにより、より自然に人や動物の発声動作を模擬することが可能となる。

【 0 0 4 2 】

なお、ステップ S 6 の処理を実行中、外部センサ 8 が調音器官 16 である口の内部に物体が挿入されたことを検知した場合、検知情報が制御部 3 に供給され、それに対応して、制御部 3 は、調音動作実行部 7 および音声出力部 9 の処理を停止、再開、または中断する。このようにすることで、物体が挿入されて口が動かさないときには、声が出せない状態となるので、よりリアリティを増すことができる。また、外部センサ 8 からの検出情報だけでなく、何らかの外力により、調音器官 16 の動作が妨げられた場合にも、音声出力部 9 の処理を停止、再開、または中断するようにしてもよい。

40

【 0 0 4 3 】

このような制御は、いわば、調音動作の変更に対応して、発声処理を変更する制御であるが、反対に、例えば、発声する言葉が急に変更された場合、調音動作も直ちに変更される

50

ような、いわば、発声処理の変更に対応して調音動作を変更するような制御を行うようにしてもよい。

【0044】

なお、本実施の形態においては、音声言語情報生成部2の出力を、例えば「こんにちは」のようなテキストデータとしたが、「K O X N I C H I W A」のような音韻情報を出力するようにしてもかまわない。

【0045】

また、本発明は、発声する音声の音韻と、調音器官以外の部位の動作を同期させる場合にも適用することが可能である。すなわち、例えば、図12に示すように、音声の音韻と首の動作、あるいは手の動作を同期させる場合にも適用することが可能である。

10

【0046】

さらに、本発明は、ロボットに限らず、例えば、コンピュータグラフィクスにより表現されるキャラクタの発声する音声の音韻と、当該キャラクタの動作を同期させる場合にも適用することが可能である。

【0047】

ところで、上述した一連の処理は、ハードウェアにより実行させることもできるが、ソフトウェアにより実行させることもできる。一連の処理をソフトウェアにより実行させる場合には、そのソフトウェアを構成するプログラムが、専用のハードウェアに組み込まれているコンピュータ、または、各種のプログラムをインストールすることで、各種の機能を実行することが可能な、例えば汎用のパーソナルコンピュータなどに、記録媒体からインストールされる。

20

【0048】

この記録媒体は、図1に示すように、コンピュータとは別に、ユーザにプログラムを提供するために配布される、プログラムが記録されている磁気ディスク12(フロッピディスクを含む)、光ディスク13(CD-ROM(Compact Disc-Read Only Memory)、DVD(Digital Versatile Disc)を含む)、光磁気ディスク14(MD(Mini Disc)を含む)、もしくは半導体メモリ15などよりなるパッケージメディアにより構成されるだけでなく、コンピュータに予め組み込まれた状態でユーザに提供される、プログラムが記録されているROMやハードディスクなどで構成される。

【0049】

なお、本明細書において、記録媒体に記録されるプログラムを記述するステップは、記載された順序に従って時系列的に行われる処理はもちろん、必ずしも時系列的に処理されなくとも、並列的あるいは個別に実行される処理をも含むものである。

30

【0050】

【発明の効果】

以上のように、本発明によれば、音声合成による発声する言葉と発声タイミングに、調音器官を模擬した部位の動作を同期させることが可能となり、よりリアルなロボットを実現することが可能となる。

【図面の簡単な説明】

【図1】本発明を適用したロボットの音声出力、および調音器官を模擬した部位の動作を制御する部分の構成例を示すブロック図である。

40

【図2】音韻情報および音韻継続時間の例を示す図である。

【図3】調音動作命令と調音動作時間の例を示す図である。

【図4】調整された音韻継続時間の例を示す図である。

【図5】本発明を適用してロボットの動作を説明するフローチャートである。

【図6】対応する音韻継続時間と調音動作時間の例を示す図である。

【図7】第1の方法によって調整された音韻継続時間と調音動作時間を示す図である。

【図8】第2の方法によって調整された音韻継続時間と調音動作時間を示す図である。

【図9】第3の方法によって調整された音韻継続時間と調音動作時間を示す図である。

【図10】第4の方法によって調整された音韻継続時間と調音動作時間を示す図である。

50

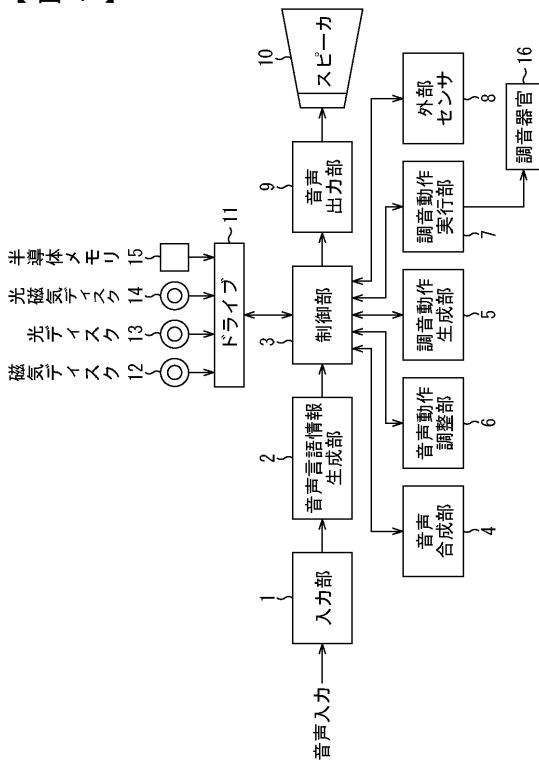
【図11】第5の方法によって調整された音韻継続時間と調音動作時間を示す図である。

【図12】音韻情報と調音器官以外の動作を同期させる例を示す図である。

【符号の説明】

1 入力部, 2 音声言語情報生成部, 3 制御部, 4 音声合成部, 5 調音動作生成部, 6 音声動作調整部, 7 調音動作実行部, 8 外部センサ, 9 音声出力部, 10 スピーカ, 16 調音器官

【図1】



【図2】

音韻情報	K	O	X	N	I	CH	I	W	A
音韻継続時間 (単位 msec)	87.0	126.0	79.0	40.0	91.0	114.0	42.0	32.0	190.0

音韻情報
音韻継続時間
(単位 msec)

【図 3】

音韻情報	K		O		X		N		I		CH		I		W		A	
	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T
顎 (上下)	下	20	上	80	上	80	下	20	上	40	下	30	上	40	下	50	上	50
唇 (上下)	下	30	下	30	上	60	下	0	下	30	上	0	上	50	下	40	上	0
(張暖)	暖	40	0	張	50	張	20	張	40	張	40	張	20	暖	40	暖	40	暖
(閉開)	閉	30	0	閉	70	閉	10	閉	30	閉	40	閉	50	閉	50	閉	0	0
喉 (振幅)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
舌 (上下)	上	40	上	30	上	30	上	0	上	0	上	0	上	0	上	0	下	40
(左右)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
(前後)	後	30	前	40	0	0	0	0	0	0	0	0	0	後	50	0	前	60
鼻 (閉開)	閉	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	閉	50

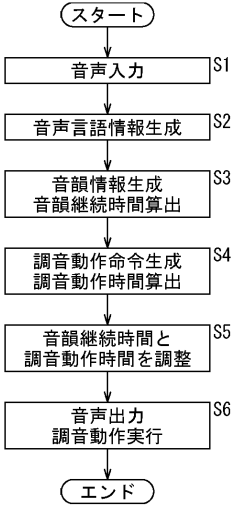
A: 動作命令 T: 動作時間 (msec)

【図 4】

音韻情報	K	O	X	N	I	CH	I	W	A
音韻継続時間 (単位 msec)	87.0	126.0	80.0	40.0	91.0	114.0	50.0	40.0	190.0

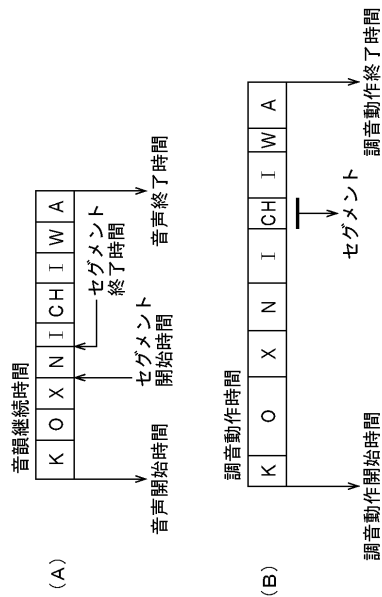
音韻継続時間 (単位 msec)

【図 5】



【図 6】

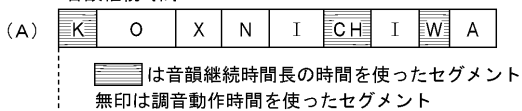
(例) 音韻列とそれに対する音韻継続時間、調音動作時間が以下のような場合



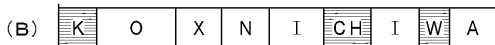
【 図 7 】

開始点と長い方で揃える場合

音韻継続時間



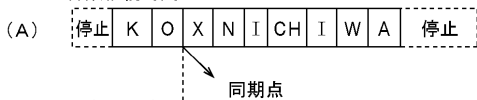
調音動作時間



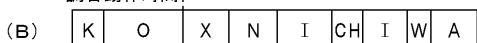
【 図 8 】

任意点をあわせ、音声、動作停止で揃える場合

音韻継続時間



調音動作時間



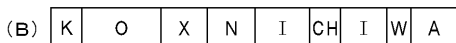
【 図 9 】

調音動作が優先の場合

音韻継続時間

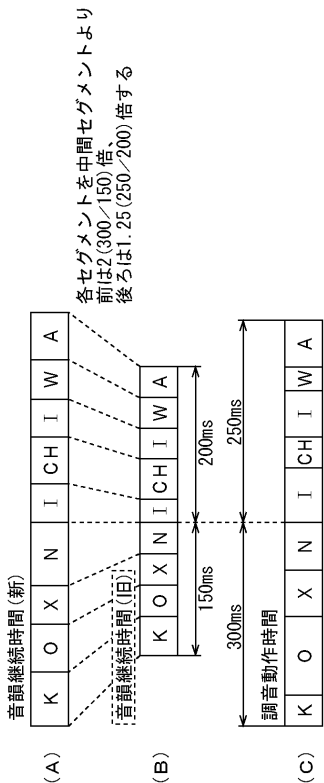


調音動作時間



【 図 1 1 】

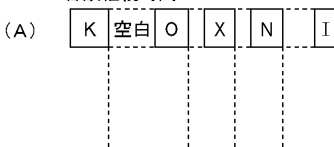
中間セグメントで合わせ、比例して伸ばす



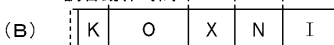
【 図 1 0 】

セグメント単位で終了をあわせ、空白を埋める場合

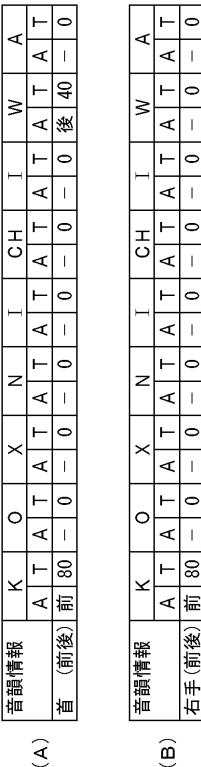
音韻継続時間



調音動作時間



【 図 1 2 】



フロントページの続き

- (72)発明者 赤羽 誠
東京都品川区北品川6丁目7番35号 ソニー株式会社内
- (72)発明者 島川 真人
東京都品川区北品川6丁目7番35号 ソニー株式会社内
- (72)発明者 山崎 信英
東京都品川区北品川6丁目7番35号 ソニー株式会社内
- (72)発明者 小林 恵理香
東京都品川区北品川6丁目7番35号 ソニー株式会社内

審査官 八木 誠

(56)参考文献 特開平10-171486(JP,A)

(58)調査した分野(Int.Cl., DB名)

B25J1/00-21/02

A63H1/00-37/00

G10L11/00-13/08,19/00-21/06