(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第3854713号
(P3854713)
(45)発行日　平成18年12月6日(2006.12.6)　(24)登録日　平成18年9月15日(2006.9.15)

(51)Int.Cl.　F I
G10L 13/08 (2006.01)　G10L 13/08 127F
G10L 13/08 131C

請求項の数 9　(全 16 頁)

(21)出願番号　特願平10-57900
(22)出願日　平成10年3月10日(1998.3.10)
(65)公開番号　特開平11-259095
(43)公開日　平成11年9月24日(1999.9.24)
審査請求日　平成16年5月27日(2004.5.27)

(73)特許権者　000001007
キヤノン株式会社
東京都大田区下丸子３丁目３０番２号
(74)代理人　100076428
弁理士　大塚　康徳
(74)代理人　100112508
弁理士　高柳　司郎
(74)代理人　100115071
弁理士　大塚　康弘
(74)代理人　100116894
弁理士　木村　秀二
(74)代理人　100093908
弁理士　松本　研一
(74)代理人　100101306
弁理士　丸山　幸雄

最終頁に続く

(54)【発明の名称】音声合成方法および装置および記憶媒体

(57)【特許請求の範囲】

【請求項１】

音韻系列に従って音声を合成する音声合成装置であって、

音韻の種類ごとに音韻時間長の標準偏差を格納する格納手段と、

前記音韻系列の発声時間を示す発声時間情報を取得する発声時間取得手段と、

前記音韻系列の各音韻に対応する第１の音韻時間長を取得する取得手段と、

前記取得手段で取得した第１の音韻時間長の和を、前記発声時間から減じた値を、各音韻に対応する標準偏差の二乗和で割った値を係数とし、各音韻について、該係数と当該音韻の標準偏差の二乗との積を当該音韻の第１の音韻時間長に加えた値を第２の音韻時間長として設定する設定手段とを備えることを特徴とする音声合成装置。

【請求項２】

前記格納手段は更に音韻の種類ごとに音韻時間長の平均値を格納し、

前記取得手段は、前記音韻系列の各音韻の前記平均値又は重回帰分析による音韻時間長推定値のいずれかを用いて前記第１の音韻時間長を取得することを特徴とする請求項１記載の音声合成装置。

【請求項３】

前記取得手段は、各音韻の第１の音韻時間長として、当該音韻の平均値を中心に標準偏差の定数倍の範囲内に収まる音韻時間長を設定することを特徴とする請求項１記載の音声合成装置。

【請求項４】

前記格納手段は更に音韻の種類ごとに音韻時間長の最小値を格納し、
前記取得手段は、各音韻の第１の音韻時間長が当該音韻の前記最小値より小さい場合は該第１の音韻時間長を該最小値に設定することを特徴とする請求項１記載の音声合成装置。
【請求項５】
前記格納手段は各音韻の平均値、標準偏差、最小値を発声速度に基づいた分類毎に格納し、
前記取得手段は、前記音韻系列の発声時間から算出した発声速度に対応する各音韻の平均値、標準偏差、最小値を利用して各音韻の第１の音韻時間長を算出することを特徴とする請求項１乃至請求項４のいずれかに記載の音声合成装置。
【請求項６】
音声合成対象の文字系列を取得する文字系列取得手段と、
前記文字系列取得手段で取得した文字系列を音韻系列に変換する変換手段とを更に備え、
前記発声時間取得手段は、前記文字系列に含まれる発声速度を示す制御シーケンスに基づいて、前記発声時間情報を取得することを特徴とする請求項１記載の音声合成装置。
【請求項７】
音声合成対象の文字系列を取得する文字系列取得手段と、
前記文字系列取得手段で取得した文字系列を音韻系列に変換する変換手段とを更に備え、
前記発声時間取得手段は、ユーザによって設定された発声速度に基づいて、前記発声時間情報を取得することを特徴とする請求項１記載の音声合成装置。
【請求項８】
音韻系列に従って音声を合成する音声合成方法であって、
前記音韻系列の発声時間を示す発声時間情報を取得する発声時間取得工程と、
前記音韻系列の各音韻に対応する第１の音韻時間長を取得する取得工程と、
音韻の種類ごとに音韻時間長の標準偏差を格納する格納手段から前記音韻系列の各音韻に対応する標準偏差を取得し、前記取得工程で取得した第１の音韻時間長の和を、前記発声時間から減じた値を、各音韻に対応する標準偏差の二乗和で割った値を係数とし、各音韻について、該係数と当該音韻の標準偏差の二乗との積を当該音韻の第１の音韻時間長に加えた値を第２の音韻時間長として設定する設定工程とを備えることを特徴とする音声合成方法。
【請求項９】
音韻系列に従って音声を合成する音声合成方法をコンピュータに実行させるための制御プログラムを格納した記憶媒体であって、
前記音声合成方法が、
前記音韻系列の発声時間を示す発声時間情報を取得する発声時間取得工程と、
前記音韻系列の各音韻に対応する第１の音韻時間長を取得する取得工程と、
音韻の種類ごとに音韻時間長の標準偏差を格納する格納手段から前記音韻系列の各音韻に対応する標準偏差を取得し、前記取得工程で取得した第１の音韻時間長の和を、前記発声時間から減じた値を、各音韻に対応する標準偏差の二乗和で割った値を係数とし、各音韻について、該係数と当該音韻の標準偏差の二乗との積を当該音韻の第１の音韻時間長に加えた値を第２の音韻時間長として設定する設定工程とを備えることを特徴とする記憶媒体。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】
本発明は、規則合成方式による音声合成方法および音声合成装置、および、音声合成方法を実装した、コンピュータが読むことができるプログラムを格納した記憶媒体に関する。
【０００２】

【従来の技術】
従来の音声規則合成装置では、音韻時間長を制御する方法として、音韻時間長に関する統計量から導出した制御規則による方法（匂坂芳典、東倉洋一：“規則による音声合成のための音韻時間長制御”、電子通信学会論文誌、Vol.J67-A,No.7(1984)pp.629-636）、重回帰分析の一手法である数量化Ｉ類を用いる方法（洒寄哲也、佐々木昭一、北川博雄：“規則合成のための数量化Ｉ類を用いた韻律制御”、音響学会講演論文集、3-4-17(1986-10)）がある。
【０００３】
【発明が解決しようとする課題】
しかしながら、上述した従来技術においては、音韻系列の発声時間を指定することが難しいという問題がある。たとえば、制御規則による方法では、指定された発声時間に対応した制御規則の導出が難しい。また、制御規則による方法で例外的な入力がある場合や数量化Ｉ類を用いる方法で良い推定値が得られない場合に自然性を感じる音韻時間長に対する誤差が大きくなる、という問題がある。
【０００４】
制御規則を用いて音韻時間長を制御する場合、統計量（平均値や標準偏差など）に対して前後の音韻の組み合わせを考慮した重み付けや、伸縮係数の設定などが必要になってくる。音韻の組み合わせの場合分けや、重み付けや伸縮係数などのパラメータなど操作する項目が多く、しかも、操作方法（制御規則）を経験則で決めていかなければならない。音韻系列の発声時間が指定されたときに、たとえ音韻の個数が同じでも、音韻の組み合わせは膨大になる。どのような音韻の組み合わせでも、音韻時間長の和が指定された発声時間に近くなるような、制御規則の導出は困難である。
【０００５】
本発明は上記の問題点に鑑みてなされたものであり、指定した発声時間になるように音韻系列の音韻時間長を設定することを可能とし、発声時間の長短によらず自然な音韻時間長を与える音声合成方法および装置および記憶媒体を提供することを目的とする。
【０００６】
【課題を解決するための手段】
上記の目的を達成するための本発明の一態様による音声合成装置は例えば以下の構成を備える。すなわち、
音韻系列に従って音声を合成する音声合成装置であって、
音韻の種類ごとに音韻時間長の標準偏差を格納する格納手段と、
前記音韻系列の発声時間を示す発声時間情報を取得する発声時間取得手段と、
前記音韻系列の各音韻に対応する第１の音韻時間長を取得する取得手段と、
前記取得手段で取得した第１の音韻時間長の和を、前記発声時間から減じた値を、各音韻に対応する標準偏差の二乗和で割った値を係数とし、各音韻について、該係数と当該音韻の標準偏差の二乗との積を当該音韻の第１の音韻時間長に加えた値を第２の音韻時間長として設定する設定手段とを備える。
【０００７】
また、本発明によれば、上記音声合成装置で実行される音声合成方法が提供される。更に、本発明によれば、上記音声合成方法をコンピュータに実現させるための制御プログラムを格納する記憶媒体が提供される。
【０００８】
【発明の実施の形態】
以下、添付の図面を参照して本発明の好適な実施形態を説明する。
【０００９】
［第１の実施形態］
図１は、第１の実施形態の音声合成装置の構成を示すブロック図である。１０１はＣＰＵであり、本音声規則合成装置における各種制御を行なう。１０２はＲＯＭであり、各種パラメータやＣＰＵ１０１が実行する制御プログラムを格納する。１０３はＲＡＭであり、

ＣＰＵ１０１が実行する制御プログラムを格納するとともに、ＣＰＵ１０１の作業領域を提供する。１０４はハードディスク、フロッピーディスク、ＣＤ－ＲＯＭ等の外部記憶装置である。１０５は入力部であり、キーボード、マウス等から構成される。１０６はディスプレイであり、ＣＰＵ１０１の制御により各種表示を行なう。６は音声合成部であり、合成音声を生成する。１０７はスピーカであり、音声合成部６より出力される音声信号（電気信号）を音声に変換して出力する。

【００１０】

図２は、第１の実施形態による音声合成装置の機能構成を示すブロック図である。以下に示される各機能は、ＲＯＭ１０２に格納された制御プログラムあるいは外部記憶装置１０４からＲＡＭ１０３にロードされた制御プログラムをＣＰＵ１０１が実行することによって実現される。

【００１１】

１は文字系列入力部であり、入力部１０５より入力された合成すべき音声の文字系列、すなわち表音テキストの入力処理を行なう。例えば合成すべき音声が「音声」であるときには、「おんせい」というような文字系列を入力する。また、この文字系列中には、発声速度や声の高さなどを設定するための制御シーケンスなどが含まれることもある。２は制御データ格納部であり、文字系列入力部１で制御シーケンスと判断された情報や、ユーザインタフェースより入力される発声速度や声の高さなどの制御データを内部レジスタに格納する。３は音韻系列生成部であり、文字系列入力部１より入力された文字系列を音韻系列へ変換する。例えば、「おんせい」という文字系列は、「ｏ，Ｘ，ｓ，ｅ，ｉ」という音韻系列へ変換される。４は音韻系列格納部であり、音韻系列生成部３で生成された音韻系列を内部レジスタに格納する。なお、上述の各内部レジスタとしてはＲＡＭ１０３を用いることが可能である。

【００１２】

５は音韻時間長設定部であり、制御データ格納部２に格納された制御データの発声速度と音韻系列格納部４に格納された音韻の種類より、音韻時間長を設定する。６は音声合成部であり、音韻時間長設定部５で音韻時間長の設定された音韻系列と制御データ格納部２に格納された制御データの声の高さから、合成音声を生成する。

【００１３】

次に、音韻時間長設定部５で行なわれる音韻時間長の設定について説明する。以下の説明において、音韻集合をΩとする。Ωの例としては、

Ω＝｛a,e,i,o,u,X（撥音）,b,d,g,m,n,r,w,y,z,ch,f,h,k,p,s,sh,t,ts,Q（促音）｝

などを使用することができる。

【００１４】

また、音韻時間長設定区間を呼気段落（ポーズとポーズの間の区間）とする。さて、音韻時間長設定区間の音韻系列αｉ（１≦ｉ≦Ｎ）を、制御データ格納部２に格納された制御データの発声速度によって決定される発声時間Ｔで発声するように、当該音韻系列の各音韻αiの音韻時間長ｄiを決定する。すなわち、音韻系列の各αｉ（式（１ａ））の音韻時間長ｄｉ（式（１ｂ））を、式（１ｃ）を満足するように決定する。

【００１５】

【数１】

$$\alpha_i \in \Omega \quad (1 \leq i \leq N) \qquad (1\,a)$$

$$d_i \quad (1 \leq i \leq N) \qquad (1\,b)$$

$$T = \sum_{i=1}^{N} d_i \qquad (1\,c)$$

【００１６】
ここで、音韻αiの音韻時間長初期値をｄαi0とする。また、音韻αiに関して、音韻時間長の平均、標準偏差、最小値をそれぞれμαi，σαi，ｄαiminとする。そして、これらの値を用いて、以下に示す式（２）に従ってｄαiを決定し、これを新たな音韻時間長初期値とする。すなわち、音韻時間長の平均値、標準偏差値、最小値を音韻の種類毎（αi毎）に求め、これをメモリに格納しておき、これらの値を用いて音韻時間長の初期値を決定しなおす。

【００１７】
【数２】

$$d_{\alpha_i} = \begin{cases} \max(\mu_{\alpha_i} - 3\sigma_{\alpha_i}, d_{\alpha_i min}) & (d_{\alpha_i 0} < \max(\mu_{\alpha_i} - 3\sigma_{\alpha_i}, d_{\alpha_i min})) \\ d_{\alpha_i 0} & (\max(\mu_{\alpha_i} - 3\sigma_{\alpha_i}, d_{\alpha_i min}) \leq d_{\alpha_i 0} \leq \mu_{\alpha_i} + 3\sigma_{\alpha_i}) \\ \mu_{\alpha_i} + 3\sigma_{\alpha_i} & (\mu_{\alpha_i} + 3\sigma_{\alpha_i} < d_{\alpha_i 0}) \end{cases} \qquad (2)$$

【００１８】
こうして得られた音韻時間長初期値ｄαiを用いて、音韻時間長ｄiを式（３ａ）に従って設定する。なお、得られたｄiが閾値θi（＞０）に対してｄi＜θiとなるときは、式（３ｂ）に従って設定される。

【００１９】
【数３】

$$d_i = d_{\alpha_i} + \rho(\sigma_{\alpha_i})^2$$
$$\rho = \frac{\left(T - \sum_{i=1}^{N} d_{\alpha_i}\right)}{\sum_{i=1}^{N} (\sigma_{\alpha_i})^2} \qquad (3\,a)$$

$$d_i = \theta_i \qquad (3\,b)$$

【００２０】
すなわち、更新された音韻時間長の初期値の和を設定された発声時間Ｔから差引き、これを音韻時間長の標準偏差σαiの二乗和で割った値を係数ρとし、この係数ρと標準偏差σαiの二乗との積を当該音韻時間長の初期値ｄαiに加えた値を、音韻時間長ｄiとする。

【００２１】

以上の動作を、図３のフローチャートを参照して説明する。

【００２２】

まず、ステップＳ１で、文字系列入力部１より表音テキストが入力される。ステップＳ２では、外部入力された制御データ（発声速度、声の高さ）と入力された表音テキスト中の制御データが制御データ格納部２に格納される。ステップＳ３で、文字系列入力部１より入力された表音テキストから音韻系列生成部３において音韻系列が生成される。

【００２３】

次に、ステップＳ４で、次の時間長設定区間の音韻系列が音韻系列格納部４に取り込まれる。ステップＳ５で、音韻時間長設定部５において、音韻αiの種類に応じて音韻時間長初期値dαiが設定される（式（２））。ステップＳ６では、まず、制御データ格納部２に格納された制御データの発声速度から音韻時間長設定区間の発声時間Ｔを設定する。そして、音韻時間長設定区間の音韻系列の音韻時間長の和が音韻時間長設定区間の発声時間Ｔに等しくなるように、上記式（３ａ）、（３ｂ）を用いて、音韻時間長設定区間の音韻系列の各音韻時間長を設定する。

【００２４】

ステップＳ７で、音韻時間長設定部５で音韻時間長の設定された音韻系列と制御データ格納部２に格納された制御データの声の高さから、合成音声が生成される。そして、ステップＳ８で、入力された文字列に対する最後の音韻時間長設定区間であるか否かが判別され、最後の音韻時間長設定区間でない場合はステップＳ１０で外部入力された制御データが制御データ格納部２に格納されてステップＳ４に戻り、処理が続けられる。

【００２５】

一方、ステップＳ８で最後の音韻時間長設定区間であると判定された場合はステップＳ９に進み、入力が終了したか否かが判別される。入力が終了していない場合はステップＳ１に戻り、上記処理が繰り返される。

【００２６】

なお、式（２）は、音韻時間長初期値が現実にはあり得ないような値や出現確率の低い値に設定されるのを防ぐためのものである。音韻時間長の確率密度が正規分布であると仮定したときに、平均値から標準偏差の±３倍以内に入る確率は０．９９６となる。更に、音韻時間長が短くなりすぎるのを防ぐために、標本集団の最小値未満にはならないようにしている。

【００２７】

式（３ａ）は、式（２）で設定された音韻時間長初期値を平均値とする正規分布が各音韻時間長の確率密度関数であると仮定して、式（１ｃ）の制約条件のもとで最尤推定（maximum likelihood estimation）を行った結果である。本例の最尤推定について説明すると次のとおりである。

【００２８】

音韻αiの音韻時間長の標準偏差をσαiとする。音韻時間長の確率密度分布が正規分布であると仮定する（式（４ａ））。このとき、音韻時間長の対数尤度は式（４ｂ）のようになる。ここで、対数尤度を最大にするのは、式（４ｃ）のＫを最小にするのと同値である。そこで、音韻時間長の対数尤度が最大になるように上述の式（１ｃ）を満たすｄiを決定する。

【００２９】

【数４】

$$P_{\alpha_i}(d_i) = \left(\sqrt{2\pi}\sigma_{\alpha_i}\right)^{-1} \exp\left(-\frac{(d_i - d_{\alpha_i})^2}{2(\sigma_{\alpha_i})^2}\right) \qquad (4\,a)$$

$$\begin{aligned} \log(L(d_i)) &= \log\left(\prod_{i=1}^{N} P_{\alpha_i}(d_i)\right) \\ &= -\sum_{i=1}^{N} \log\left(\sqrt{2\pi}\sigma_{\alpha_i}\right) - \frac{1}{2}\sum_{i=1}^{N} \frac{(d_i - d_{\alpha_i})^2}{(\sigma_{\alpha_i})^2} \qquad (4\,b) \end{aligned}$$

$$K = \sum_{i=1}^{N} \frac{(d_i - d_{\alpha_i})^2}{(\sigma_{\alpha_i})^2} \qquad (4\,c)$$

【００３０】
今、式（５ａ）のように変数変換を行うと、式（４ｃ）及び式（１ｃ）は式（５ｂ）及び（５ｃ）のようになる。Ｋが最小となるのは、球（式５ｂ））が平面（式（５ｃ））に接するときであり、式（５ｄ）の場合である。この結果、式（３ａ）が導かれる。

【００３１】
【数５】

$$\rho_i = \frac{d_i - d_{\alpha_i}}{\sigma_{\alpha_i}} \qquad (5\,a)$$

$$K = \sum_{i=1}^{N} \rho_i^2 \qquad (5\,b)$$

$$\sum_{i=1}^{N} \rho_i \sigma_{\alpha_i} = T - \sum_{i=1}^{N} d_{\alpha_i} \qquad (5\,c)$$

$$\begin{aligned} \rho_i &= \rho\sigma_{\alpha_i} \\ \rho &= \frac{\left(T - \sum_{i=1}^{N} d_{\alpha_i}\right)}{\sum_{i=1}^{N} (\sigma_{\alpha_i})^2} \qquad (5\,d) \end{aligned}$$

【００３２】
式（２）と式（３ａ）、（３ｂ）を総合して、自然発声の標本集団から求めた統計量（平均値、標準偏差、最小値）を用いて、所望の発声時間（（１ｃ）式）を満たす最も確からしい（尤度が最大になる）値に音韻時間長が設定される。したがって、所望の発声時間（（１ｃ）式）を満たすように自然発声したときに得られる音韻時間長に対する誤差が小さい、という意味で自然な音韻時間長が得られる。

【００３３】
［第２の実施形態］
第１の実施形態では、発声速度（発声時間）や音韻のカテゴリにかかわらず、各音韻αｉの音韻時間長ｄｉを同一の規則で決定した。第２の実施形態では、発声速度や音韻のカテ

ゴリに応じて音韻時間長ｄｉの決定規則を変化させ、より自然な音声合成を可能とする。なお、第２の実施形態によるハードウエア構成、機能構成は第１の実施形態（図１、図２）と同様である。

【００３４】

音韻αｉに関して、発声速度でカテゴリーを分けて音韻時間長の平均値、標準偏差、最小値を求める。例えば、発声速度のカテゴリーを呼気段落の平均モーラ時間長で表すとして、

１：１２０ミリ秒未満、
２：１２０ミリ秒以上１４０ミリ秒未満、
３：１４０ミリ秒以上１６０ミリ秒未満、
４：１６０ミリ秒以上１８０ミリ秒未満、
５：１８０ミリ秒以上

とする。なお、上述した項目の先頭の数字を発声速度に対応するカテゴリーのインデックスとする。発声速度に対応するカテゴリーのインデックスをｎとして音韻時間長の平均値、標準偏差、最小値を求め、それぞれμαｉ(ｎ)，σαｉ(ｎ)，ｄαimin(ｎ)とする。

【００３５】

音韻αｉの音韻時間長初期値をｄαｉ0とする。音韻時間長初期値ｄαｉ0を平均値によって決定する音韻の集合をΩａ、重回帰分析の一手法である数量化Ｉ類（質的なデータから量的に測定される外的基準を予測したり、説明したりするための手法）によって決定する音韻の集合をΩｒとする。ここで、Ωの要素で、Ωａ，Ωｒのどちらにも含まれない要素や、両者に含まれる要素は存在しないようにする。すなわち、以下の式（６）を満たすようにする。

【００３６】

【数６】

$$\Omega_a \cup \Omega_r = \Omega \qquad (6a)$$

$$\Omega_a \cap \Omega_r = \emptyset \qquad (6b)$$

【００３７】

αｉ∈Ωａのとき、すなわちαｉがΩａに属するときは、平均値によって音韻時間長初期値を決定する。すなわち、音声速度に対応するカテゴリーのインデックスｎを求めて、以下の式（７）によって音韻時間長初期値を決定する。

【００３８】

【数７】

$$d_{\alpha_i 0} = \mu_{\alpha_i}(n) \qquad (7)$$

【００３９】

一方、αｉ∈Ωｒのとき、すなわちαｉがΩｒに属するときは、数量化Ｉ類によって音韻時間長初期値を決定する。ここで、要因のインデックスをｊ(１≦ｊ≦ｊ)，各要因に対応するカテゴリーのインデックスをｋ(１≦ｋ≦Ｋ(ｊ))として、(ｊ，ｋ)に対応する数量化Ｉ類の係数を、

ａjk

とする。

【００４０】

要因の一例として、
１：当該音韻の２つ前の先行音韻、
２：当該音韻の１つ前の先行音韻、
３：当該音韻、
４：当該音韻の１つ後の後続音韻、
５：当該音韻の２つ後の後続音韻、
６：呼気段落の平均モーラ時間長、
７：呼気段落内モーラ位置、
８：当該音韻を含む単語の品詞
などを使用することができる。上述した項目の先頭の数字が要因のインデックスｊに対応する。

【００４１】
さらに、各要因に対応するカテゴリーの例を述べる。音韻のカテゴリーは、
１：ａ、２：ｅ、３：ｉ、４：ｏ、５：ｕ、６：Ｘ、７：ｂ、８：ｄ、９：ｇ、10：ｍ、11：ｎ、12：ｒ、13：ｗ、14：ｙ、15：ｚ、16：＋、17：ｃ、18：ｆ、19：ｈ、20：ｋ、21：ｐ、22：ｓ、23：ｓｈ、24：ｔ、25：ｔｓ、26：Ｑ、27：ポーズ、とし、当該音韻のみ“ポーズ”をはずす。実施形態において、呼気段落を音韻時間長設定区間としているが、呼気段落はポーズを含まないので、当該音韻からポーズをはずす。なお、呼気段落という用語は、ポーズ（または文頭）とポーズ（または文末）の間の区間で、途中にポーズを含まないものという意味で使用している。

【００４２】
また、呼気段落内の平均モーラ時間長のカテゴリは、
１：１２０ミリ秒未満
２：１２０ミリ秒以上１４０ミリ秒未満
３：１４０ミリ秒以上１６０ミリ秒未満
４：１６０ミリ秒以上１８０ミリ秒未満
５：１８０ミリ秒以上
とする。

【００４３】
また、呼気段落内モーラ位置に関しては、
１：第１モーラ
２：第２モーラ
３：第３モーラ以降最後から第３番目のモーラまで
４：最後から２番目のモーラ
５：最後のモーラ
とする。

【００４４】
更に、品詞のカテゴリーを
、１：名詞、２：副詞的名詞、３：代名詞、４：固有名詞、５：数、６：動詞、７：形容詞、８：形容動詞、９：副詞、10：連体詞、11：接続詞、12：感動詞、13：助動詞、14：格助詞、15：副助詞、16：並立助詞、17：準体助詞、18：接続助詞、19：終助詞、20：接頭辞、21：接尾辞、22：形動接尾、23：サ変接尾、24：形容詞接尾、25：動詞接尾、26：助数詞
とする。

【００４５】
なお、要因（アイテムともいう）とは、数量化Ｉ類での予測に使用する質的なデータの種類を意味する。カテゴリーは、各要因毎に取りうる選択肢を意味する。したがって、上記の例に即して説明すると、次のようになる。

【００４６】
要因のインデックスｊ＝１：当該音韻の２つ前の先行音韻

インデックスｋ＝１に対応するカテゴリー：ａ
インデックスｋ＝２に対応するカテゴリー：ｅ
インデックスｋ＝３に対応するカテゴリー：ｉ
インデックスｋ＝４に対応するカテゴリー：ｏ
（中略）
インデックスｋ＝２６に対応するカテゴリー：Ｑ
インデックスｋ＝２７に対応するカテゴリー：ポーズ。
【００４７】
要因のインデックスｊ＝２：当該音韻の１つ前の先行音韻
インデックスｋ＝１に対応するカテゴリー：ａ
インデックスｋ＝２に対応するカテゴリー：ｅ
インデックスｋ＝３に対応するカテゴリー：ｉ
インデックスｋ＝４に対応するカテゴリー：ｏ
（中略）
インデックスｋ＝２６に対応するカテゴリー：Ｑ
インデックスｋ＝２７に対応するカテゴリー：ポーズ。
【００４８】
要因のインデックスｊ＝３：当該音韻
インデックスｋ＝１に対応するカテゴリー：ａ
インデックスｋ＝２に対応するカテゴリー：ｅ
インデックスｋ＝３に対応するカテゴリー：ｉ
インデックスｋ＝４に対応するカテゴリー：ｏ
（中略）
インデックスｋ＝２６に対応するカテゴリー：Ｑ
インデックスｋ＝２７に対応するカテゴリー：ポーズ。
【００４９】
要因のインデックスｊ＝４：当該音韻の１つ後の後続音韻
インデックスｋ＝１に対応するカテゴリー：ａ
インデックスｋ＝２に対応するカテゴリー：ｅ
インデックスｋ＝３に対応するカテゴリー：ｉ
インデックスｋ＝４に対応するカテゴリー：ｏ
（中略）
インデックスｋ＝２６に対応するカテゴリー：Ｑ
インデックスｋ＝２７に対応するカテゴリー：ポーズ。
【００５０】
要因のインデックスｊ＝５：当該音韻の２つ後の後続音韻
インデックスｋ＝１に対応するカテゴリー：ａ
インデックスｋ＝２に対応するカテゴリー：ｅ
インデックスｋ＝３に対応するカテゴリー：ｉ
インデックスｋ＝４に対応するカテゴリー：ｏ
（中略）
インデックスｋ＝２６に対応するカテゴリー：Ｑ
インデックスｋ＝２７に対応するカテゴリー：ポーズ。
【００５１】
要因のインデックスｊ＝６：呼気段落内の平均モーラ時間長
インデックスｋ＝１に対応するカテゴリー：120ミリ秒未満
インデックスｋ＝２に対応するカテゴリー：120ミリ秒以上140ミリ秒未満
インデックスｋ＝３に対応するカテゴリー：140ミリ秒以上160ミリ秒未満
インデックスｋ＝４に対応するカテゴリー：160ミリ秒以上180ミリ秒未満
インデックスｋ＝５に対応するカテゴリー：180ミリ秒以上。

【００５２】
要因のインデックスｊ＝７：呼気段落内モーラ位置
インデックスｋ＝１に対応するカテゴリー：第１モーラ
インデックスｋ＝２に対応するカテゴリー：第２モーラ
（中略）
インデックスｋ＝５に対応するカテゴリー：最後のモーラ。
【００５３】
要因のインデックスｊ＝８：当該音韻を含む単語の品詞
インデックスｋ＝１に対応するカテゴリー：名詞
インデックスｋ＝２に対応するカテゴリー：副詞的名詞
（中略）
インデックスｋ＝２６に対応するカテゴリー：助数詞
となる。
【００５４】
上述した項目の先頭の数字がカテゴリーのインデックスｋに対応する。
【００５５】
そして、各要因ごとに係数ａjkの平均が0になるようにする。すなわち、式（８）を満足するようにする。
【００５６】
【数８】

$$\sum_{k=1}^{K(j)} a_{jk} = 0 \quad (1 \leq j \leq J) \qquad (8)$$

【００５７】
また、音韻αiのダミー変数を、以下のように設定する。
【００５８】
【数９】

$$\delta_i(j,k) = \begin{cases} 1 & (\text{音韻}\alpha i \text{が要因}\ j \text{のカテゴリー}\ k \text{に反応するとき}) \\ 0 & (\text{上記以外のとき}) \end{cases} \qquad (9)$$

【００５９】
係数とダミー変数の積和に加える定数をｃ0とする。このとき、音韻αiの音韻時間長の数量化Ｉ類による推定値は、式（１０）となる。
【００６０】
【数１０】

$$\hat{d}_{\alpha_i} = \sum_{j=1}^{J} \sum_{k=1}^{K(j)} a_{jk}\delta_i(j,k) + c_0 \qquad (10)$$

【００６１】
そして、この推定値を用いて音韻αiの音韻時間長初期値を以下のように決定する。
【００６２】
【数１１】

$$d_{\alpha_i 0} = \hat{d}_{\alpha_i} \qquad (11)$$

【００６３】
さらに、発声速度と対応するカテゴリーのインデックスｎを求めて、当該カテゴリーの音韻時間長の平均値、標準偏差、最小値を得て、これらを用いて音韻時間長初期値ｄαi0を以下の式で更新する。こうして得られたｄαiを改めて音韻時間長初期値として設定する。

【００６４】
【数１２】

$$d_{\alpha_i} = \begin{cases} \max(\mu_{\alpha_i}(n) - r_\sigma \sigma_{\alpha_i}(n), d_{\alpha_i min}(n)) & (d_{\alpha_i 0} < \max(\mu_{\alpha_i}(n) - r_\sigma \sigma_{\alpha_i}(n), d_{\alpha_i min}(n))) \\ d_{\alpha_i 0} & (\max(\mu_{\alpha_i}(n) - r_\sigma \sigma_{\alpha_i}(n), d_{\alpha_i min}(n)) \le d_{\alpha_i 0} \le \mu_{\alpha_i}(n) + r_\sigma \sigma_{\alpha_i}(n)) \\ \mu_{\alpha_i}(n) + r_\sigma \sigma_{\alpha_i}(n) & (\mu_{\alpha_i}(n) + r_\sigma \sigma_{\alpha_i}(n) < d_{\alpha_i 0}) \end{cases} \qquad (12)$$

【００６５】
ここで、式中の標準偏差に掛ける係数のｒσは、例えば、ｒσ＝３とする。以上のようにして得られた音韻時間長初期値を用いて、第１の実施形態と類似の方法で音韻時間長を決定する。すなわち、以下の式（１３ａ）を用いて音韻時間長ｄiを決定し、閾値θｉ（＞０）に対してｄｉ＜θｉとなるときは、式（１３ｂ）により音韻時間長ｄiを決定する。

【００６６】
【数１３】

$$d_i = d_{\alpha_i} + \rho(\sigma_{\alpha_i}(n))^2$$
$$\rho = \frac{\left(T - \sum_{i=1}^{N} d_{\alpha_i}\right)}{\sum_{i=1}^{N} (\sigma_{\alpha_i}(n))^2} \qquad (13a)$$

$$d_i = \theta_i \qquad (13b)$$

【００６７】
以上の動作を、図３のフローチャートを流用して説明する。ステップＳ１で、文字系列入力部１より表音テキストが入力される。ステップＳ２で、外部入力された制御データ(発声速度、音の高さ)と入力された表音テキスト中の制御データが制御データ格納部２に格納される。ステップＳ３で、文字系列入力部１より入力された表音テキストから音韻系列生成部３において音韻系列が生成される。ステップＳ４で、次の音韻時間長設定区間の音韻系列が音韻系列格納部４に取り込まれる。

【００６８】
ステップＳ５では、音韻時間長設定部５において、制御データ格納部２に格納された制御データの発声速度、音韻時間長の平均値と標準偏差と最小値、および、数量化Ｉ類による音韻時間長推定値を用いて、上述した方法により、音韻の種類（カテゴリ）に応じて音韻時間長初期値が設定される。

【００６９】
ステップＳ６では、音韻時間長設定部５において、制御データ格納部２に格納された制御データの発声速度から音韻時間長設定区間の発声時間を設定し、音韻時間長設定区間の音韻系列の音韻時間長の和が音韻時間長設定区間の発声時間に等しくなるように、音韻時間長設定区間の音韻系列の音韻時間長を上述した方法により設定する。

【００７０】
ステップＳ７で、音韻時間長設定部５で音韻時間長の設定された音韻系列と制御データ格納部２に格納された制御データの声の高さから、合成音声が生成される。ステップＳ８で、入力された文字列に対する最後の音韻時間長設定区間であるか否かが判別される。最後の音韻時間長設定区間でない場合はステップＳ１０へ進む。ステップＳ１０では、外部入力された制御データが制御データ格納部２に格納されてステップＳ４に戻り、処理が続けられる。一方、最後の音韻時間長設定区間である場合はステップＳ９に進み、入力が終了したか否かが判別され、終了していない場合はステップＳ１に戻り、処理が続けられる。
【００７１】
なお、上記各実施形態における構成は本発明の一実施形態を示したものであり、各種変形が可能である。変形例を示せば以下の通りである。
【００７２】
（１）上述した各実施形態において音韻集合Ωは一例であり、それ以外の集合も使用でき、言語や音韻の種類に応じて音韻集合の要素を決めることができる。また、本発明は日本語以外の言語にも適用可能である。
【００７３】
（２）上述した実施形態において、呼気段落は音韻時間長設定区間の一例であり、他にも、単語、形態素、文節、文などを音韻時間長設定区間とすることができる。なお、文を音韻時間長設定区間とするときは、当該音韻のポーズを考慮する必要がある。
【００７４】
（３）上述した実施形態において、音韻時間長の初期値として設定する値として、自然発声した音声の音韻時間長を使用することができる。また、他の音韻時間長制御規則によって決定した値や数量化Ｉ類を用いて推定した値を使用することもできる。
【００７５】
（４）上述した第２の実施形態において、音韻時間長の平均値を求めるのに使用する発声速度のカテゴリーは一例を示すものであり、他のカテゴリーを用いても良い。
【００７６】
（５）上述した第２の実施形態において、数量化Ｉ類の要因とカテゴリーは一例を示すものであり、他の要因やカテゴリーを用いても良い。
【００７７】
（６）上述した実施形態において、音韻時間長初期値の設定に使用する標準偏差に掛ける係数ｒσ＝３は、一例を示すものであり、他の値を用いてもよい。
【００７８】
また、本発明の目的は、前述した実施形態の機能を実現するソフトウェアのプログラムコードを記録した記憶媒体を、システムあるいは装置に供給し、そのシステムあるいは装置のコンピュータ（またはＣＰＵやＭＰＵ）が記憶媒体に格納されたプログラムコードを読出し実行することによっても、達成されることは言うまでもない。
【００７９】
この場合、記憶媒体から読出されたプログラムコード自体が前述した実施形態の機能を実現することになり、そのプログラムコードを記憶した記憶媒体は本発明を構成することになる。
【００８０】
プログラムコードを供給するための記憶媒体としては、例えば、フロッピディスク，ハードディスク，光ディスク，光磁気ディスク，ＣＤ－ＲＯＭ，ＣＤ－Ｒ，磁気テープ，不揮発性のメモリカード，ＲＯＭなどを用いることができる。
【００８１】
また、コンピュータが読出したプログラムコードを実行することにより、前述した実施形態の機能が実現されるだけでなく、そのプログラムコードの指示に基づき、コンピュータ上で稼働しているＯＳ（オペレーティングシステム）などが実際の処理の一部または全部を行い、その処理によって前述した実施形態の機能が実現される場合も含まれることは言

うまでもない。

【００８２】

さらに、記憶媒体から読出されたプログラムコードが、コンピュータに挿入された機能拡張ボードやコンピュータに接続された機能拡張ユニットに備わるメモリに書込まれた後、そのプログラムコードの指示に基づき、その機能拡張ボードや機能拡張ユニットに備わるＣＰＵなどが実際の処理の一部または全部を行い、その処理によって前述した実施形態の機能が実現される場合も含まれることは言うまでもない。

【００８３】

【発明の効果】

以上説明したように、本発明によれば、指定した発声時間になるように音韻系列の音韻時間長を設定することが可能となり、発声時間の長短によらず自然な音韻時間長を与えることが可能である。

【００８４】

【図面の簡単な説明】

【図１】本発明の実施形態に係る音声合成装置の構成を示すブロック図である。

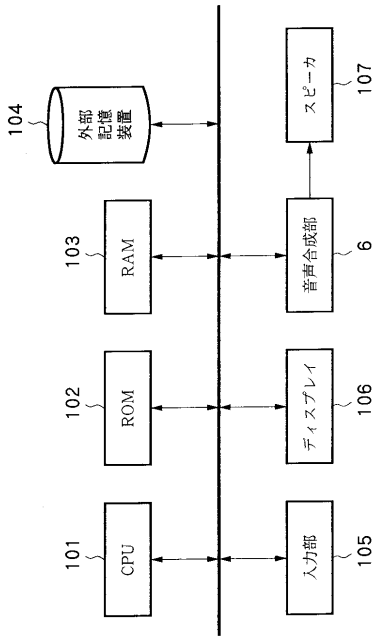
【図２】本発明の実施形態に係る音声合成装置の機能構成を示すブロック図である。

【図３】本発明の実施形態に係る音声合成手段を示すフローチャートである。

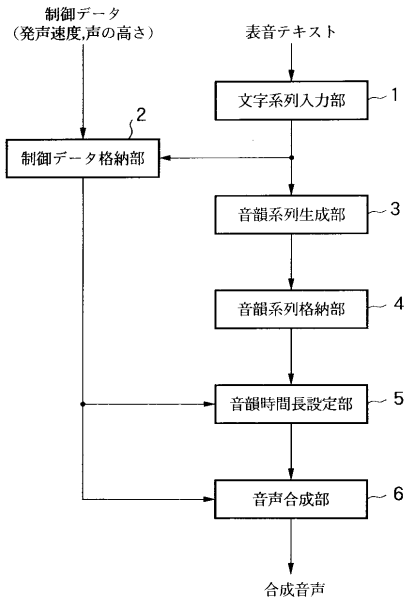
【符号の説明】

1　文字系列入力部
2　制御データ格納部
3　音韻系列生成部
4　音韻系列格納部
5　音韻時間長設定部
6　音声合成部

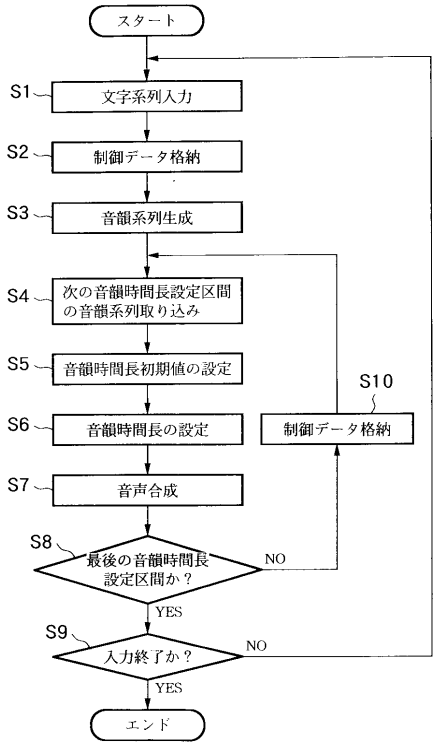
【図１】



【図２】



【図３】


スタート
S1
文字系列入力
S2
制御データ格納
S3
音韻系列生成
S4
次の音韻時間長設定区間の音韻系列取り込み
S5
音韻時間長初期値の設定
S6
音韻時間長の設定
S7
音声合成
S8
最後の音韻時間長設定区間か？
NO
YES
S9
入力終了か？
NO
YES
エンド
S10
制御データ格納

フロントページの続き

(72)発明者　大塚　充
東京都大田区下丸子３丁目３０番２号　キヤノン株式会社内

審査官　荏原　雄一

(56)参考文献　特開平１０－０３９８９６（ＪＰ，Ａ）
特開平０１－２５５８９９（ＪＰ，Ａ）
特開平０４－１７０６００（ＪＰ，Ａ）
規則合成のための数量化Ｉ類を用いた韻律制御，音講論，昭和６１年秋，３－４－１７

(58)調査した分野(Int.Cl.，ＤＢ名)
G10L 13/08
JSTPlus(JDream2)