

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号
特許第4644879号
(P4644879)

(45) 発行日 平成23年3月9日 (2011.3.9)

(24) 登録日 平成22年12月17日 (2010.12.17)

(51) Int.Cl.

F I

G 1 O L 13/08 (2006.01)

G 1 O L 11/00 (2006.01)

G 1 O L 13/08 1 5 O Z

G 1 O L 13/08 1 3 1 Z

G 1 O L 11/00 5 O 1

請求項の数 5 (全 23 頁)

(21) 出願番号	特願2005-329011 (P2005-329011)	(73) 特許権者	393031586
(22) 出願日	平成17年11月14日 (2005.11.14)		株式会社国際電気通信基礎技術研究所
(65) 公開番号	特開2007-133328 (P2007-133328A)		京都府相楽郡精華町光台二丁目2番地2
(43) 公開日	平成19年5月31日 (2007.5.31)	(74) 代理人	100099933
審査請求日	平成19年11月16日 (2007.11.16)		弁理士 清水 敏
(出願人による申告) 平成17年度独立行政法人情報通信研究機構、研究テーマ「人間情報コミュニケーションの研究開発」に関する委託研究、産業活力再生特別措置法第30条の適用を受ける特許出願		(72) 発明者	竹本 浩典
			京都府相楽郡精華町光台二丁目2番地2
			株式会社国際電気通信基礎技術研究所内
		(72) 発明者	北村 達也
			京都府相楽郡精華町光台二丁目2番地2
			株式会社国際電気通信基礎技術研究所内
		(72) 発明者	本多 清志
			京都府相楽郡精華町光台二丁目2番地2
			株式会社国際電気通信基礎技術研究所内
最終頁に続く			

(54) 【発明の名称】 調音パラメータ補間用データ生成装置及びコンピュータプログラム

(57) 【特許請求の範囲】

【請求項1】

第1の音素から第2の音素まで連続的に変化する音声を合成するための調音パラメータを、前記第1及び第2の音素の音声合成のための既知の調音パラメータの間の補間によって生成する際の補間用データを生成するための調音パラメータ補間用データ生成装置であって、

連続する前記第1の音素及び前記第2の音素を含む入力音声信号から、前記第1の音素から前記第2の音素までの所定のスペクトル変化率の推移を算出するための変化率算出手段と、

前記変化率を用いて前記補間用データを作成するための補間用データ作成手段とを含み

、前記変化率算出手段は、

前記入力音声信号を所定時間ごとにフレーム化するためのフレーム化手段と、

前記フレーム化手段によりフレーム化された音声信号からフレームごとに前記所定のスペクトル変化率を算出するためのスペクトル変化率算出手段とを含み、

前記スペクトル変化率算出手段は、

前記第1の音素から前記第2の音素までの前記音声信号から前記フレーム化手段によりフレーム化された音声信号のフレームの内、前記所定のスペクトル変化率の隣り合う極小値を与える第1及び第2の二つのフレームと、当該二つのフレームの間に存在し、前記所定のスペクトル変化率の極大値を与える第3のフレームとを含む複数のフレームを組合せ

10

20

るためのフレーム組合せ手段と、

前記フレーム組合せ手段により組合されたフレームの前記極大値と極小値とにより定められるガウス関数で前記所定のスペクトル変化率の推移を近似するための関数近似手段とを含み、

前記補間用データ作成手段は、

前記関数近似手段により得られた前記ガウス関数の、前記第1のフレームから前記第2のフレームまでを積分して積分関数を求め、さらに当該積分関数をその最大値で正規化した正規化積分関数を求めるための積分・正規化手段を含む、調音パラメータ補間用データ生成装置。

【請求項2】

10

前記関数近似手段は、

前記フレーム組合せ手段により組合されたフレームのうち、前記第1のフレームと前記第3のフレームとの間で、前記所定のスペクトル変化率を近似する第1のガウス関数を求めるための手段と、

前記フレーム組合せ手段により組合されたフレームのうち、前記第3のフレームと前記第2のフレームとの間で、前記所定のスペクトル変化率を近似する第2のガウス関数を求めるための手段と、

前記第3のフレームにおいて前記第1のガウス関数と前記第2のガウス関数とを接続することにより、前記極大値と2つの極小値との間で前記所定のスペクトル変化率の推移を近似する前記ガウス関数を求めるための手段とを含む、請求項1に記載の長音パラメータ補間用データ生成装置。

20

【請求項3】

前記補間用データ作成手段はさらに、前記第1の音素のための調音パラメータと、前記第2の音素のための調音パラメータとの間を前記正規化積分関数によって補間する事により、前記第1の音素と前記第2の音素との間の所定の時刻における混合比を算出するための手段を含む、請求項1又は請求項2に記載の調音パラメータ補間用データ生成装置。

【請求項4】

前記補間用データ作成手段はさらに、前記第1の音素のための調音パラメータと、前記第2の音素のための調音パラメータとの間を前記混合比によって補間する事により、前記所定の時刻における調音パラメータである声道断面積関数を算出するための手段を含む、請求項3に記載の調音パラメータ補間用データ生成装置。

30

【請求項5】

コンピュータにより実行されると、当該コンピュータを、請求項1～請求項4のいずれかに記載の調音パラメータ補間用データ生成装置として動作させる、コンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

この発明は、音声合成に際し、音源に対する変調を行ない、合成される音声をより滑らかにするためのパラメータを生成する技術に関する。

40

【背景技術】

【0002】

人間とコンピュータシステムに代表される機械系との間のインターフェイスとして、近年種々のものが提案されている。それらの中で最近特に使用される頻度が高くなっているものに、音声がある。音声を用いる事によって、人間と機械系との間のコミュニケーションを、人間同士の間でのそれに近い形で実現できる。

【0003】

音声によるコミュニケーションを実現するための音声合成技術には、以下の2つの方式がある。1つは、あらかじめ録音された音声の音声波形から音素片を取り出し、それらをつなぎ合わせる事によって音声を合成する方式である。もう1つは、人間の発話器官の形

50

状の変化をシミュレートする事によって音声を合成する方式である。

【0004】

音声波形の変化に比べて、発話器官の形状の変化は緩やかである事から、後者の方式の方が、聴覚上、より滑らかな音声を合成する事ができる。そのため、この方式は近年特に注目されている。

【0005】

この方式の音声合成方法では、音源からの信号を、発話器官の形状を表わすパラメータ（例えば声道断面積関数、声道長、開口面積等。以下「調音パラメータ」という）により特性が変化する電気回路からなるフィルタに通す事によって、音声信号を合成する。例えば、「あ」という音を発声しているときと、「い」という音を発声しているときとでは、調音パラメータが違う。従って、調音パラメータを変化させる事によって、種々の音声を合成する事ができる。もちろん、この方式は、現在ではコンピュータとソフトウェアとによるデジタル技術で実現される。

10

【0006】

より自然な連続音声を合成するためには、ある音声を発話しているときの調音パラメータと、次の音声を発話しているときの調音パラメータとを、連続音声発話時の発話器官の形状変化の時間的な推移と一致した形で補間する必要がある。そのためには、実際に連続音声を発話している際の発話器官の形状変化の速度を反映した物理量とその時間的推移とを、何らかの方法で算出する必要がある。

【0007】

20

発話器官の形状変化の速度を利用する先行技術として非特許文献1に開示のものがある。これは、連続音声を合成するために、声道断面積関数を時間補間したものであり、時間補間のためのパラメータをMRI（Magnetic Resonance Imaging：核磁気共鳴画像）動画の平均輝度値の時間的な変化から求めたものである。この方法によると、実際の調音動作に基づいた合成であるので、自然な連続音を得られるという利点がある。

【非特許文献1】声道断面積関数の時間補間による音声合成、竹本浩典、本多清志、北村達也、パーハム・モクタリ、平井啓之、日本音響学会講演論文集、pp157-158、2005年3月。

【発明の開示】

30

【発明が解決しようとする課題】

【0008】

非特許文献1に開示の技術によれば、実際の調音動作に基づいた自然な連続音を得る事ができるという大きな利点がある。しかしこの方法では、人間に発話をさせながらMRI動画を撮る必要がある。周知の様にMRI動画の撮影には時間がかかる。また、MRI動画を撮るために必要な設備も大型で、その利用には面倒な手順がいる。さらに、MRI動画の画像処理は大変な上、コストもかかる。

【0009】

非特許文献1に開示された技術と同様の効果を、より簡単に、簡略な処理で得る事ができる様にする事が望ましい。

40

【0010】

そこで、この発明の目的は、MRI動画等の透過撮影手法よりも簡略な処理により、実際の発話動作による発話器官の形状変化を反映する調音パラメータ補間用のデータを生成できる補間用データ生成装置を提供する事である。

【課題を解決するための手段】

【0011】

本発明の第1の局面に係る調音パラメータ補間用データ生成装置は、第1の音素から第2の音素まで連続的に変化する音声を合成するための調音パラメータを、第1及び第2の音素の音声合成のための既知の調音パラメータの間の補間によって生成する際の補間用データを生成するための装置であって、連続する第1の音素及び第2の音素を含む入力音声

50

信号から、第1の音素から第2の音素までの所定の音響特徴量の変化率の推移を算出するための変化率算出手段と、変化率を用いて補間用データを作成するための補間用データ作成手段とを含む。

【0012】

この補間用データ生成装置によると、各々独立した第1の音素と第2の音素との間を連続的につなぐ様な補間用データを音声から生成できる。音声は、MRI動画と異なり、容易に取得できる。また、音響特徴量の変化率は、発話器官の形状変化をよく反映すると考えられる。従って、MRI動画等の透過撮影手法を用いるよりも簡略な方法で実際の発話動作による発話器官の形状変化を反映する補間用データを生成する事ができる。

【0013】

好ましくは、変化率算出手段は、入力音声信号を所定時間ごとにフレーム化するためのフレーム化手段と、フレーム化手段によりフレーム化された音声信号からフレームごとに所定のスペクトル変化率を算出するためのスペクトル変化率算出手段とを含む。

【0014】

この補間用データ生成装置によると、上述した変化率として、フレームごとのスペクトル変化率を算出する事ができる。そして、このスペクトル変化率は、発話器官の形状変化の速度をよく反映すると考えられる上、その算出には確立された技術を用いる事ができる。従って、MRI動画等の透過撮影手法を用いるよりも簡略な、音の特徴を使用して実際の発話動作による発話器官の形状変化を反映する補間用データを生成する事ができる。

【0015】

さらに好ましくは、スペクトル変化率算出手段は、第1の音素から第2の音素までの音声信号からフレーム化手段によりフレーム化された音声信号のフレームの内、所定のスペクトル変化率の隣り合う極小値を与える第1及び第2の二つのフレームと、当該二つのフレームの間に存在し、所定のスペクトル変化率の極大値を与える第3のフレームとを含む複数のフレームを組合せるためのフレーム組合せ手段と、フレーム組合せ手段により組合せられたフレームの極大値と極小値とにより定められる所定の連続関数で所定のスペクトル変化率の推移を近似するための関数近似手段とを含む。

【0016】

この補間用データ生成装置によると、音中心にあたる二つの極小値を両端とする連続関数で所定のスペクトル変化率の推移を近似する。そこで、補間用データ生成の際に、スペクトル変化率そのものよりもデータ量の小さい近似関数を使用する事ができる。それゆえ、補間用データ生成の際に必要なリソースを簡略化する事ができる。従って、より簡略な方法を使用して、実際の発話動作による発話器官の形状変化を反映する補間用データを生成する事ができる。

【0017】

さらに好ましくは、フレーム組合せ手段は、変化率に基づき、第1及び第2のフレームを検出するための極小値検出手段と、変化率に基づき、第3のフレームを検出するための極大値検出手段と、第1のフレームに第1の音素を、第2のフレームに第2の音素を、それぞれ割当てするための音割当手段と、第1のフレームから第2のフレームまでのフレームを組合せてフレームシーケンスを作成するためのフレームシーケンス作成手段とを含む。

【0018】

この補間用データ生成装置によると、第1のフレーム及び第2のフレームに具体的な音素を割当て、上記二つのフレームを含む様にフレームシーケンスを作成する。ゆえに、補間用データ生成の際に、補間の対象を自動的に定めて補間する事ができる。従って、より簡略な方法で実際の発話動作による発話器官の形状変化を反映する補間用データを生成する事ができる。

【0019】

さらに好ましくは、関数近似手段は、第1のフレームにおける変化率と、第3のフレームにおける変化率とに基づいて、第1のフレームにおいて所定の定数を取り、第3のフレームにおいて第3のフレームにおける変化率の値と所定の関係を有する値をとる第1の近

10

20

30

40

50

似関数を求める第1の近似手段と、第3のフレームにおける変化率と第2のフレームにおける変化率とに基づいて、第2のフレームにおいて所定の定数を取り、第3のフレームにおいて第1の近似関数と同じ値をとる第2の近似関数を求める第2の近似手段とを含む。

【0020】

この補間用データ生成装置によると、該当するフレームにおける変化率に基づいて、近似関数の種類を適切に変化させる事ができる。変化率が最大となる箇所は、補間対象の二つの音素の中間位置とは限らず、その前後にずれる事が多い。そうした場合でも、上記した様に変化率が最大となる箇所の前後で異なる近似関数を適用する事により、簡単な近似関数を用いて精度高く変化率を近似できる。ゆえに、近似の精度が上がり、この近似関数を用いた補間用データ生成処理にかかる時間を短縮する事ができる。従って、より簡略な方法で実際の発話動作による発話器官の形状変化を反映する補間用データを生成する事ができる。

10

【0021】

さらに好ましくは、補間用データ作成手段は、関数近似手段により得られた所定の関数の、第1のフレームから第2のフレームまでの積分関数を正規化した正規化積分関数を求めるための正規化積分手段を含む。

【0022】

この補間用データ生成装置によると、補間用データとして、正規化積分関数を算出するのみであるので、調音パラメータそのものの補間までする場合と比較して、補間用データ生成のためにかかる時間が短縮できる。また、補間用データをデータベース（以下DBと記載）に収納する場合には、DBの容量が削減できる。従って、より簡略な処理により実際の発話動作による発話器官の形状変化を反映する補間用データを生成する事ができる。

20

【0023】

さらに好ましくは、補間用データ作成手段はさらに、第1の音素のための調音パラメータと、第2の音素のための調音パラメータとの間を正規化積分関数によって補間する事により、第1の音素と第2の音素との間の所定の時刻における混合比を算出するための手段を含む。

【0024】

この補間用データ生成装置によると、補間用データとして、混合比を算出するのみである。そこで、補間用データ生成のためにかかる時間が短縮できる。また、補間用データをDBに収納する場合には、DBの容量が削減できる。さらに、正規化積分関数の算出までで処理をとどめておく場合と比較して、音声の合成時には混合比の算出をする必要がない。従って、より簡略な処理により実際の発話動作による発話器官の形状変化を反映するとともに、音声合成時の処理量を削減可能な補間用データを生成する事ができる。

30

【0025】

さらに好ましくは、補間用データ作成手段はさらに、第1の音素のための調音パラメータと、第2の音素のための調音パラメータとの間を混合比によって補間する事により、所定の時刻における調音パラメータである声道断面積関数を算出するための手段を含む。

【0026】

この補間用データ生成装置によると、音声の特徴量から補間用データとして、声道断面積関数を算出する。ゆえに、この声道断面積関数を使用して後の音声合成を行なう場合には、音声合成にかかる時間を削減する事ができる。従って、効率的な音声合成を行なう事ができる。その結果、MRI動画等による処理よりも簡略な処理によって、実際の発話動作による発話器官の形状変化を反映する補間用データを生成する事ができる。

40

【0027】

この発明の第2の局面に係るコンピュータプログラムは、コンピュータにより実行されると、当該コンピュータを、上記したいずれかの調音パラメータ補間用データ生成装置として動作させるものである。従って、上記した調音パラメータ補間用データ生成装置と同様の効果を奏する事ができる。

【発明を実施するための最良の形態】

50

【0028】

<スペクトル変化率に基づく補間方法>

本発明の一実施の形態に係る音声合成システムでは、音声信号の時間変化から、補間用パラメータである声道断面積関数を算出し、その結果を用いて音声合成を行なう。音声信号はMRI動画像と比較してはるかに容易に得る事ができるし、信号処理も容易である。本実施の形態の詳細を説明する前に、まず、本実施の形態で採用した補間方法の原理について詳述する。なお以下の説明及び図面において、同じ部品には同じ参照番号を付す。

【0029】

(1) 混合比の算出

混合比とは、発話音声第1の音素から第2の音素へ移行する途中の音素を合成するために、移行途中の調音パラメータを、第1の音素と第2の音素との調音パラメータの間の補間により生成するための、第1の音素と第2の音素との調音パラメータを混合する割合の事をいう。例えば実際の発話において、「あ」から「い」に移行する際には、徐々に「あ」の音の要素が減り、代わりに「い」の音の要素が増えて、連続的に音が移行する。混合比を使用すると、母音中心の間での連続音を生成するための調音パラメータをスムーズに補間できる。本実施の形態では、混合比を発話音声のスペクトル変化率から得た値で近似する。スペクトル変化率とは、発話音声のスペクトルがどの程度の速度で変化しているかを表す値であり（式については後述する。）、発話時の発話器官の動きを反映している。以下で、スペクトル変化率を用いた混合比の算出方法について述べる。

【0030】

図1に、発話音声波形の一例と、この音声波形から得られたスペクトル変化率との関係を示す。図1を参照して、図中で/a/、/i/、/u/、/e/、及び/o/で表わされるフレームは各々の母音の母音中心が存在する位置を示している。母音の発声の間は、発話器官形状は安定しており、従ってスペクトル変化率は小さい。一方、ある母音から別の母音への移行時には、発話器官の形状が変化する。この変化の速度は一定ではなく、変化の最初と最後とでは変化速度が小さく、途中のどこかで極大となる事が知られている。図1に示す例では、スペクトル変化率の極大値30、32、34及び36で発話器官の形状変化速度が最大に達する。

【0031】

前述した様に混合比は、隣接する母音中心の調音パラメータの間を補間して連続的で自然な音声合成をするためのパラメータである。従って、混合比算出にあたっては、隣接する二つの母音の発声中心に対応するフレーム（スペクトル変化率が極小となるフレーム）が両端に配される様にとられたフレームシーケンスのスペクトル変化率を使用してその間の音声を補間するための混合比を算出すればよい。ただし、この場合、フレームシーケンスの最後のフレームは、次のフレームシーケンスの先頭フレームに相当するので、混合比算出の対象とはしない。

【0032】

また、処理対象となる音声信号にはノイズが含まれている事が多く、そのために、音声信号から得られるスペクトル変化率のグラフは変形している。そうしたノイズの影響を除去し、かつ演算を簡略にするために、スペクトル変化率の描く曲線を何らかの関数で近似する。本実施の形態では、スペクトル変化率の描く曲線をガウス関数で近似する。

【0033】

図2に、母音/a/から母音/i/まで音声に変化する際の音声信号から得られたスペクトル変化率の一例（スペクトル変化率42）と、このスペクトル変化率の描く曲線を近似するガウス関数44とを示す。図2を参照して、母音/a/の母音中心がグラフの左端に、母音/i/の母音中心がグラフの右端に配されている。なお、前述の理由により、グラフの右端は母音/i/の母音中心を含むフレームよりも一つ前のフレームとなっている。

【0034】

ガウス分布は左右対称の分布を示す。しかし、スペクトル変化率は極大値フレーム40

10

20

30

40

50

の左右で異なった分布を持つ事が一般的である。そこで、極大値フレームの左右で異なるガウス関数を割当てて。この近似に用いられるガウス関数は、スペクトル変化率の極大値と極小値とから定められる。なお、ガウス関数は、x 軸に漸近するが決して 0 にはならない。一方、ガウス関数による近似では、フレームシーケンスの両端の母音位置で調音パラメータがそれぞれ両者の調音パラメータと一致する事が望ましく、そのためにはフレームシーケンスの両端の母音位置では近似関数の値が 0 となる必要がある。そこで、例えば上記したガウス関数を一旦求めた後、さらにフレームシーケンスの両端での関数の値が 0 となる様に、ガウス関数から定数を引いたり、ガウス関数に所定倍数を乗じた後に定数を引いたりして、関数の値がフレームシーケンスの両端（母音位置）で 0 となる様にする。

【0035】

10

この様にしてスペクトル変化率を近似するガウス関数を求めた後、このガウス関数から以下の様にして混合比を算出する。

【0036】

図 3 の上段に図 2 に示すスペクトル変化率 50 の曲線と、対応するガウス関数による近似（近似関数 52）とを示す。この近似関数 52 を積分し、最終的な積分の値が 1 となる様に正規化した曲線 54（図 3 下段に示す）により、母音 / a / から母音 / i / までの移行途中の混合比が得られる。

【0037】

以上のガウス関数の割当から混合比までの算出をまとめると、次の様になる。隣接する母音中心間に m 個のフレームがあり、その n 番目にスペクトル変化率の極大値が存在するものとする。このとき、母音間のフレーム番号を k（ $0 \leq k < n$ ）、曲線 60 に対応する関数（すなわち混合比）を $br(k)$ とすると、混合比 $br(k)$ は次の式（1）及び（2）で表わされる。

20

【0038】

【数 1】

$$br(k) = 0 \quad (k = 1) \quad (1)$$

【0039】

【数 2】

$$br(k) = \frac{1}{2 \sum_{l=2}^m (y(l-1) + y(l))} \sum_{l=2}^k (y(l-1) + y(l)) \quad (k > 1) \quad (2) \quad 30$$

また、 $y(k)$ は次の式（3）及び（4）で表わされる。

【0040】

【数 3】

$$y(k) = gw_{2(n-1)+1}(k) \quad (k \leq n) \quad (3)$$

【0041】

【数 4】

$$y(k) = gw_{2(m-n)+1}(k+m-2n+1) \quad (k > n) \quad (4) \quad 40$$

ここで、 $gw_n(k)$ は、n 点の Gaussian Window の k 番目の点である。

【0042】

（2）声道断面積関数の算出

次に、この混合比 br を用いて、声道断面積関数を算出する方法について説明する。

【0043】

図 4 及び図 5 に、声道断面積関数の例を示す。声道断面積関数は時刻の関数であり、ある時刻における声門からの距離と声道断面積との間の関係を示す。図 4 及び図 5 では、縦軸に声道断面積をとり、横軸に声門からの距離をとっている。図 4 に示された声道断面積

50

関数は時刻 t_1 の時のものである。図 5 に示された声道断面積関数は時刻 t_2 の時のものである。声道断面積関数を求めるためには、各時刻における声門からの距離と、その位置における声道の断面積とを求める必要がある。本実施の形態では、声道を所定個数のセクションに分け、各時刻における各セクションでの声道断面積によって声道断面積関数を表す。また本実施の形態では、セクション数は一定であるものとし、声道長の変化にはセクション長の変化で対応する。

【0044】

以下では、混合比 br を時刻 t の関数 $br(t)$ として考える。このとき、声門から数えて n 番目のセクションにおける声門からの距離 ($xpos$) は次の式 (5) で表わされる。

【0045】

【数 5】

$$xpos(t, n) = (1 - br(t)) \cdot xpos(t1, n) + br(t) \cdot xpos(t2, n) \quad (5)$$

また声道の断面積 ($area$) は次の式 (6) で表わされる。

【0046】

【数 6】

$$area(t, n) = (1 - br(t)) \cdot area(t1, n) + br(t) \cdot area(t2, n) \quad (6)$$

以上の式 (1) ~ (6) を用いる事により、ある母音から他の母音に移行する間の任意の時点における声道の各セクションにおける断面積を、両母音に対する調音パラメータの間の補間により算出する事ができる。

【0047】

< 第 1 の実施の形態 >

以下、上記した補間方法を利用した、本発明の一実施の形態に係る音声合成装置について詳述する。

【0048】

[構成]

(1) 音声合成システム 70

図 6 に、本発明の第 1 の実施の形態に係る音声合成システム 70 のブロック図を示す。なお、本実施の形態では、ある話者が連続音声を発声したものを録音し、その録音から各母音の調音パラメータと母音間の混合比とを算出する。その連続音声の発話テキスト 84 は予め与えられるものとする。

【0049】

図 6 を参照して、音声合成システム 70 は、音声信号と発話テキスト 84 とを用いて、音声合成のための調音パラメータを補間するために用いられるデータを生成するための補間用データ生成装置 86 と、補間用データ生成装置 86 によって生成されたデータを保持するための補間用 DB 88 と、入力されたテキスト 90 に対し、補間用 DB 88 内のデータを用いて合成音声信号 94 を出力するための音声合成装置 92 とを含む。

【0050】

(2) 補間用データ生成装置 86

図 7 に、補間用データ生成装置 86 のブロック図を示す。補間用データ生成装置 86 は、音声信号からそのスペクトル変化率を生成するためのスペクトル変化率生成部 60 と、生成されたスペクトル変化率を保持するためのスペクトル変化率 DB 62 と、スペクトル変化率と発話テキスト 84 とから、各母音中心の調音パラメータと補間用データとを生成するための補間用データ生成部 64 とを含む。母音中心の調音パラメータも補間比が 0 または 1 の補間用データと考える事ができるので、本実施の形態では母音中心の調音パラメータも含めて補間用データと呼ぶ。

【0051】

(3) スペクトル変化率生成部 60

図 8 に、スペクトル変化率生成部 60 のブロック図を示す。スペクトル変化率生成部 60 は、与えられた音声信号をフレーム化するためのフレーム化部 100 と、フレーム化さ

10

20

30

40

50

れた音声信号のスペクトルを抽出するためのスペクトル抽出部102と、抽出されたスペクトルからケプストラムを算出するためのケプストラム算出部104と、算出されたケプストラムから音声特徴量の差であるデルタケプストラムを算出するためのデルタケプストラム算出部106と、算出されたデルタケプストラムからスペクトル変化率を算出し、その算出結果をスペクトル変化率DB62に与えるためのスペクトル変化率算出部108を含む。

【0052】

ここで、デルタケプストラムはフレーム間の音声特徴量であるケプストラムの差を表すパラメータであるので、デルタケプストラムの値が大きいほどフレーム間のスペクトル変化が大きいと言える。デルタケプストラム $dcep(n, m)$ は次の式(7)で表わされる。ここで m はフレーム番号を表わし、 n はケプストラムの次数を表わす。

【0053】

【数7】

$$dcep(n, m) = \sum_{k=-K}^K (k \cdot cep(n, m+k)) \quad (7)$$

このデルタケプストラムを使用して、スペクトル変化率 $s(m)$ は次の式(8)で表わされる。

【0054】

【数8】

$$s(m) = \frac{1}{\sum_{k=-K}^K k^2} \sum_{n=1}^N dcep(n, m)^2 \quad (8)$$

(4) 補間用データ生成部64

図9に、補間用データ生成部64のブロック図を示す。補間用データ生成部64は、スペクトル変化率DB62に保持されたスペクトル変化率を参照して、母音中心を示す調音運動極小値を含むフレームを検出するための極小値検出部110と、スペクトル変化率DB62に保持されたスペクトル変化率を参照して、発話母音間の境界を示す調音運動極大値を含むフレームを検出するための極大値検出部112と、発話テキストを比較参照して、調音運動が極小値になるフレームに順に母音を割当てするための母音割当部114と、全フレームのうち、第1の音素の母音中心であるスペクトル変化率最小値を含むフレームから始まって、スペクトル変化率最大値を含むフレームを経て、第2の音素の母音中心であるスペクトル最小値を含むフレームの一つ前のフレームで終わる様に複数のフレームを一組のフレームシーケンスとして組合せるためのフレームシーケンス作成部116を含む。

【0055】

補間用データ生成部64はさらに、フレームシーケンス作成部116により組合わされたフレームシーケンス中で先頭のフレームに極小値を、スペクトル変化率の極大値のフレームにガウス関数の極大値を、それぞれ持ち、かつ極小値が0となるガウス関数を割当てするためのガウス関数割当部118と、スペクトル変化率の極大値のフレームに極大値を、次のフレームシーケンスの先頭のフレームに極小値を、それぞれ持ち、かつ極小値が0となるガウス関数を割当てするためのガウス関数割当部120とを含む。ガウス関数割当部118により割当てられたガウス関数のうちの前半部(極小値から極大値まで)と、ガウス関数割当部120により割当てられたガウス関数のうち後半部(極大値から極小値の1フレーム前まで)とを極大値部分で接続する事により、発話器官の動きを表す関数が得られる。以下、この関数を「発話器官変化率関数」と呼ぶ。

【0056】

補間用データ生成部64はさらに、この発話器官変化率関数を先頭フレームから最終フレーム(次のフレームシーケンスの1つ前のフレーム)まで積分して最大値が1になる様に正規化して得られる関数(以下これを「発話器官形状関数」と呼ぶ。)を出力するため

10

20

30

40

50

の積分・正規化部 1 2 2 と、積分・正規化部 1 2 2 により出力された発話器官形状関数から各フレームにおける混合比を算出するための混合比算出部 1 2 4 と、種々の音素に対して予め算出された調音パラメータを保持するための調音パラメータ D B 1 2 8 と、母音割当部 1 1 4 で割当てられた音素に対応する調音パラメータを調音パラメータ D B 1 2 8 から取り出し、対応する混合比を用いた補間により各時刻における各音素間の声道断面積関数を算出し、その結果を補間用 D B 8 8 に与えるための声道断面積関数算出部 1 2 6 とを含む。この声道断面積関数が、本実施の形態における補間用データである。

【0057】

(5) 音声合成装置 9 2

図 1 0 に、音声合成装置 9 2 のブロック図を示す。図 1 0 を参照して、テキスト 9 0 には、各音素を発声すべき時間情報が付されている。音声合成装置 9 2 は、入力されたテキスト 9 0 を音素単位に分割し、隣接する 2 音素ごとに、当該 2 音素間の補間用データを補間用 D B 8 8 から抽出して、補間調音パラメータ 1 3 2 として出力するための補間パラメータ抽出部 1 3 0 と、所定周期のクロック信号を発生するためのクロック部 1 4 0 と、合成する連続音声の発音長等に応じて、クロック部 1 4 0 からのクロックにより定まるタイミングで、補間調音パラメータ 1 3 2 を順番に出力してフィルタ 1 3 8 に与えるための出力部 1 3 4 と、発話のための音源となる音源信号を出力するための音源 1 3 6 と、出力部 1 3 4 によって与えられる調音パラメータに従って変化する特性で音源 1 3 6 からの音源信号を変調し、合成音声信号 9 4 を出力するためのフィルタ 1 3 8 とを含む。

【0058】

なお、音声にイントネーションを付けたり音の高さを調整する必要がある場合、音源 1 3 6 からの音源信号の周波数を変化させる必要がある。

【0059】

[動作]

本実施の形態に係る音声合成システム 7 0 の動作には、2 つの局面がある。すなわち、第 1 の局面は、音声信号及び対応する発話テキストから、調音パラメータ補間用のデータ（補間後の調音パラメータ）を生成し、補間用 D B を作成する局面（補間用データ生成装置 8 6 の動作に相当する。）である。第 2 の局面は、補間用 D B 8 8 のデータを用いて、入力テキスト 9 0 の連続音声を合成する局面（音声合成装置 9 2 の動作に相当する。）である。以下、順に説明する。

【0060】

(1) 補間用データ生成装置 8 6 の動作

本実施の形態に係る補間用データ生成装置 8 6 は以下の様に動作する。

【0061】

図 8 を参照して、音声信号が与えられると、フレーム化部 1 0 0 は、音声信号をフレーム化する。スペクトル抽出部 1 0 2 は、フレーム化された音声信号の各フレームから発話スペクトルを抽出する。ケプストラム算出部 1 0 4 が、各フレームから抽出された発話スペクトルに基づいてフレームごとのケプストラムを算出する。デルタケプストラム算出部 1 0 6 は、算出されたケプストラムから前述した式 (7) によってデルタケプストラムを算出する。スペクトル変化率算出部 1 0 8 は、算出されたデルタケプストラムから前述した式 (8) によってスペクトル変化率を算出する。算出されたスペクトル変化率は、スペクトル変化率 D B 6 2 に与えられ、保持される。

【0062】

次に、図 9 を参照して、極小値検出部 1 1 0 は、スペクトル変化率 D B 6 2 に保持されたスペクトル変化率から、変化率が極小であるフレームを検出する。同様に、極大値検出部 1 1 2 は、スペクトル変化率 D B 6 2 に保持されたスペクトル変化率から、変化率が極大であるフレームを検出する。

【0063】

母音割当部 1 1 4 は、検出された調音運動極小値に発話テキスト 8 4 に含まれた音を順番に対応させて、極小値である母音中心に該当する音を割当てる。

【0064】

フレームシーケンス作成部116は、スペクトル変化率が極小となるフレームのうち、隣接する極小値を与えるフレームの対と、その間のすべてのフレームとからなるフレームシーケンスを作成する。その中にはスペクトル変化率が極大になるフレームも含まれる。この処理により、一つのフレームシーケンスの組が作成される。この処理を繰返す事により、すべてのフレームシーケンスの組を作る。

【0065】

ガウス関数割当部118は、組合された各組のフレームシーケンスのうちで、先頭のフレームと極大値を含むフレームとの間のスペクトル変化率にガウス関数を割当てる。同様に、ガウス関数割当部120は、極大値を含むフレームと、当該フレームシーケンスの最後のフレームより一つ前のフレームまでに含まれるスペクトル変化率にガウス関数を割当てる。積分・正規化部122は、この二つのガウス関数を接続して得られる発話器官変化率関数についてフレームシーケンスの先頭から最後（最後の極小値を与えるフレームの一つ前のフレーム）まで積分し、さらに最大値が1になる様に正規化する事により、発話器官形状関数を生成する。

【0066】

混合比算出部124は、この様にして得られた発話器官形状関数から、前述した式（1）、（2）、（3）、及び（4）によって混合比を算出する。声道断面積関数算出部126は、調音パラメータDB128に保持された調音パラメータから、母音割当部114で割当てられた音素に該当する調音パラメータを読み出す。そして読み出された調音パラメータを使用して、算出された混合比から、前述した式（5）及び（6）によってフレームシーケンスに含まれる各フレームでの各音素間の声道断面積関数を算出する。算出された声道断面積関数は補間用DB88に与えられ、保持される。

【0067】

この処理がすべてのフレームシーケンスに対して実行される。

【0068】

（2）音声合成装置92の動作

本実施の形態に係る音声合成装置92は以下の様に動作する。

【0069】

図10を参照して、テキスト90が入力されると、補間パラメータ抽出部130は、入力テキスト90を音素単位に分割する。さらに補間パラメータ抽出部130は、入力テキスト90内において隣接する2音素の組の各々について、その2音素を補間するための補間用データ（補間後の調音パラメータ）である声道断面積関数を補間用DB88から抽出する。この抽出作業を入力テキスト90内において隣接する2音素の組み合わせのすべてについて行ない、補間調音パラメータ132として出力する。

【0070】

出力部134は、出力された補間調音パラメータ132を順に読み込み、補間調音パラメータ132に付された、そのパラメータの2音素間における位置情報及び合成すべき音声の長さ等から、クロック部140からのクロックに従って適切な時期に各補間調音パラメータをフィルタ138に与える。フィルタ138は、与えられた補間調音パラメータに従ってその特性を変化させて音源136からの音源信号を変調し、合成音声信号94を出力する。この合成音声信号94を図示しない増幅器を介してスピーカに与える事により、連続音声が発生される。

【0071】

なお、フレームシーケンス作成部116（図9参照）で行なわれる、フレームシーケンスの組を作る処理は、手動で行なわれてもよい。

【0072】

〔第1の実施の形態の効果〕

この様にして、本発明の第1の実施の形態に係る音声合成システム70によれば、実際の人間の発話における発話器官の動きと一致する様な方法で調音パラメータである声道断

10

20

30

40

50

面積関数を補間する。そのため、聴覚上、滑らかで自然な連続音声を合成することができる。さらに、この第1の実施の形態では、声道断面積関数算出まで予め行ない、実際の音声合成時にはこの声道断面積関数を読出すだけでよい。その結果、実際の音声合成時の計算量が削減されるという効果がある。また、従来技術と異なり、音声信号から補間パラメータの抽出を行なう。その結果、MRI動画像の撮影に伴う手間及びコストを削減することができる。

【0073】

〔コンピュータによる実現〕

本発明の第1の実施の形態に係る音声合成システム70は、コンピュータと、当該コンピュータ上で実行されるコンピュータプログラムとにより実現できる。以下、図11～図15を参照して音声合成システム70を実現するコンピュータプログラムの制御構造を説明する。

10

【0074】

(1) 補間用データ生成部64を実現するプログラム

図11に、補間パラメータを算出する処理の詳細なフローチャートを示す。図11を参照して、補間パラメータ算出処理が開始されると、まずステップ150にて初期処理を行なう。すなわち、ワークエリアのクリア、使用する変数のクリア等を行なう。ここで、極小値の配列の添字となる変数*i*には1を代入しておく。続いて、ステップ152では、変数*i*に1を加算した値を変数*j*に代入して、ステップ154へ進む。

【0075】

20

ステップ154では、極小値の配列において、*j*番目のデータがセットされているかどうかを判断する。セットされていれば、処理はステップ156に進み、さもなければ、補間パラメータ算出処理を終了する。

【0076】

ステップ156では、極小値の配列の*i*番目の調音パラメータ（これを調音パラメータ（*i*）と呼ぶ。）と調音パラメータ（*j*）との間を、前述した式（1）～式（6）を用いて補間する。具体的には、まず、極小値の配列の*i*番目のフレーム番号（フレーム番号（*i*）と呼ぶ。）とフレーム番号（*j*）とを参照して、フレーム番号（*i*）とフレーム番号（*j*）との間のフレームのスペクトル変化率に2種類のガウス関数を割当てて。すなわち、フレーム番号（*i*）とフレーム番号（*j*）との間のスペクトル変化率と、その間のスペクトル変化率の極大値と、その極大値を与えるフレーム番号とを特定し、フレーム番号（*i*）から極大値を与えるフレームまでについてのガウス関数と、極大値を与えるフレームからフレーム番号（*j*）－1までについてのガウス関数とを割当てて。それらガウス関数を極大値部分で接続して発話器官変化率関数を得る。この発話器官変化率関数に基づき、前述の式（1）～式（4）を用いて各フレームにおける混合比を算出する。さらに、調音パラメータ（*i*）及び調音パラメータ（*j*）と、算出された混合比とを用いて、前述の式（5）及び（6）によってそのフレームにおける声道断面積関数を算出する。この様にして、フレーム番号（*i*）とフレーム番号（*j*）との間の全てのフレームにおいて、補間された声道断面積関数を算出する。

30

【0077】

40

続いて、ステップ158において、変数*i*の値に1を加算し、再びステップ152の処理に戻る。

【0078】

この様にして、すべての極小値の間のフレームについて、補間用データを算出する処理を繰り返す。

【0079】

(2) 音声合成装置92を実現するプログラム

図12に、音声合成装置92を実現するコンピュータプログラムのフローチャートを示す。図12を参照して、音声合成処理が開始されると、ステップ160において、入力発話テキスト90に応じて、補間用DBから補間パラメータである声道断面積関数を抽出す

50

る処理が行なわれる。続いて、ステップ162において、ステップ160で抽出された声道断面積関数をクロックに従ってフィルタに出力し、合成音声信号を発生させる出力処理を行なう。

【0080】

図13に、ステップ160の補間パラメータを抽出する処理の詳細なフローチャートを示す。図13を参照して、補間パラメータ抽出処理が開始されると、まずステップ170で初期処理を行なう。すなわち、ワークエリアのクリア、使用する変数のクリア等を行なう。ここで、後述する音素の配列の添字となる変数*i*には1を代入しておく。ステップ172で、入力テキスト90を読出す。ステップ174では、テキスト90を音素単位に分割し、それらの音素を順に配列にセットする。処理はステップ176へ進む。

10

【0081】

ステップ176では、変数*i*に1を加算した値を変数*j*に代入する。ステップ178で、ステップ174でセットした音素の配列から、*i*番目の音素（これを音素（*i*）と呼ぶ。）及び音素（*j*）を参照する。このとき、音素（*j*）に音素がセットされているかどうかを判定する（ステップ180）。音素（*j*）に値がなければ（すなわち終了であれば）、補間パラメータ抽出処理を終了し、さもなければ、処理はステップ182へ進む。

【0082】

ステップ182では、音素（*i*）・音素（*j*）間の補間調音パラメータを、補間用DBよりすべて抽出し、ワークエリアに順に蓄積していく。続いてステップ184で、変数*i*に1を加算し、処理はステップ176へ戻る。

20

【0083】

この様にして、入力テキストに係る補間調音パラメータを全て順に抽出して、ワークエリアに順に出力し、蓄積していく。

【0084】

図12に示すステップ162の出力処理では、ステップ160の補間パラメータ抽出処理で抽出されワークエリアに蓄積された補間調音パラメータを使用して、合成音声信号を発生させる。なお、この処理の詳細については、前述の音声合成装置92の構成・動作の説明から処理内容が明らかであるため、ここでは詳細な説明は繰返さない。

【0085】

〔コンピュータハードウェア構成〕

30

上記したコンピュータプログラムを実行するコンピュータシステムの外観の一例を図14に、そのブロック図の例を図15に、それぞれ示す。

【0086】

図14を参照して、このコンピュータシステム330は、FD（フレキシブルディスク）ドライブ352及びCD-ROM（コンパクトディスク読出専用メモリ）ドライブ350を有するコンピュータ340と、キーボード346と、マウス348と、モニタ342と、スピーカ372とを含む。

【0087】

図15を参照して、コンピュータ340は、FDドライブ352及びCD-ROMドライブ350に加えて、CPU（中央処理装置）356と、CPU356、FDドライブ352及びCD-ROMドライブ350に接続されたバス366と、ブートアッププログラム等を記憶する読出専用メモリ（ROM）358と、バス366に接続され、プログラム命令、システムプログラム、及び作業データ等を記憶するランダムアクセスメモリ（RAM）360と、バス366に接続され、スピーカ372に接続されるサウンドボード368を含む。コンピュータシステム330はさらに、図示しないプリンタを含んでいる。

40

【0088】

ここでは示さないが、コンピュータ340はさらにローカルエリアネットワーク（LAN）への接続を提供するネットワークアダプタボードを含んでもよい。

【0089】

コンピュータシステム330に、図7に示す補間用データ生成装置86又は図10に示

50

す音声合成装置 9 2 としての動作を行なわせるためのコンピュータプログラムは、CDROM ドライブ 3 5 0 又は FD ドライブ 3 5 2 に挿入される CD-ROM 3 6 2 又は FD 3 6 4 に記憶され、さらにハードディスク 3 5 4 に転送される。又は、プログラムは図示しないネットワークを通じてコンピュータ 3 4 0 に送信されハードディスク 3 5 4 に記憶されてもよい。プログラムは実行の際に RAM 3 6 0 にロードされる。CD-ROM 3 6 2 から、FD 3 6 4 から、又はネットワークを介して、直接に RAM 3 6 0 にプログラムをロードしてもよい。

【0090】

このプログラムは、コンピュータ 3 4 0 にこの実施の形態の補間用データ生成装置 8 6 又は音声合成装置 9 2 としての動作を行なわせる複数の命令を含む。この方法を行なわせるのに必要な基本的機能のいくつかはコンピュータ 3 4 0 上で動作する OS 又はサードパーティのプログラム、もしくはコンピュータ 3 4 0 にインストールされる各種ツールキットのモジュールにより提供される。従って、このプログラムはこの実施の形態の補間用データ生成装置 8 6 又は音声合成装置 9 2 を実現するのに必要な機能全てを必ずしも含まなくてよい。このプログラムは、命令のうち、所望の結果が得られる様に制御されたやり方で適切な機能又は「ツール」を呼出す事により、上記した補間用データ生成装置 8 6 又は音声合成装置 9 2 を実現する命令のみを含んでいればよい。コンピュータシステム 3 3 0 の動作は周知であるので、ここでは繰返さない。

【0091】

＜第 2 の実施の形態＞

上記した第 1 の実施の形態では、補間用 DB 8 8 には、実際に補間式を適用して補間パラメータである声道断面積関数を算出して補間用データとして蓄積し、それを使用して連続音声合成した。しかし、本発明はこのような実施の形態には限定されない。例えば、補間用 DB 8 8 には、声道断面積関数ではなく、声道断面積関数を算出するための混合比（前述した式（1）～式（4）によって算出されるもの）を蓄積しておき、それを連続音声合成の際に使用する方法も考えられる。以下、この方法を適用した、第 2 の実施の形態について説明する。

【0092】

[構成]

本発明の第 2 の実施の形態に係る音声合成システムは、図 6 に示す第 1 の実施の形態に係る音声合成システム 7 0 と類似した構成である。ただし、この第 2 の実施の形態に係るシステムの補間用データ生成装置は、第 1 の実施の形態における補間用データ生成装置 8 6 と類似した構成を持つが、補間用データ生成部 6 4（図 9 参照）に代えて図 1 6 に示す構成を有する補間用データ生成部 1 9 0 を含む点で異なる。またこの第 2 の実施の形態に係るシステムは、第 1 の実施の形態における音声合成装置 9 2（図 1 0 参照）に代えて、図 1 7 に示す構成を有する音声合成装置 2 1 0 を含む。

【0093】

（1）補間用データ生成部 1 9 0

図 1 6 に、本実施の形態に係る補間用データ生成部 1 9 0 のブロック図を示す。図 1 6 を参照して、補間用データ生成部 1 9 0 は、第 1 の実施の形態に係る補間用データ生成部 6 4 と同様の構成を持つが、図 9 に示す声道断面積関数算出部 1 2 6 を含まず、混合比算出部 1 2 4 の出力をそのまま補間用 DB 2 0 8 に保持させる点で補間用データ生成部 6 4 と異なる。すなわち、第 2 の実施の形態においては、第 1 の実施の形態とは異なり、調音パラメータである声道断面積関数の補間までは行わず、フレームごとの混合比を算出するにとどめる様にしたものである。

【0094】

（2）音声合成装置 2 1 0

図 1 7 に、本実施の形態に係る音声合成装置 2 1 0 のブロック図を示す。図 1 7 を参照して、音声合成装置 2 1 0 は、図 1 0 に示す第 1 の実施の形態に係る音声合成装置 9 2 と類似した構成を持つ。異なるのは、音声合成装置 2 1 0 が、図 1 0 に示す補間パラメータ

抽出部 1 3 0 に代えて、隣接する音素の組の各々に対して補間用 D B 2 0 8 から 2 音素間の声道断面積関数を補間するための混合比を抽出するための混合比抽出部 2 1 2 と、種々の音素に対して予め算出された調音パラメータを保持するための調音パラメータ D B 2 1 6 と、テキスト 9 0 により与えられた音素に対応する調音パラメータを調音パラメータ D B 2 1 6 から取り出し、対応する混合比を用いた補間により音素間の声道断面積関数を算出し、補間調音パラメータ 1 3 2 として出力するための声道断面積関数算出部 2 1 4 とを含む点である。

【0095】

[動作]

第 2 の実施の形態のうち、スペクトル変化率生成部 6 0 は第 1 の実施の形態に係る音声合成システムに含まれるスペクトル変化率生成部 6 0 と同様に動作する。従ってこれについての詳細な説明は繰返さない。

【0096】

(1) 補間用データ生成部 1 9 0 の動作

図 1 6 に示す補間用データ生成部 1 9 0 は、第 1 の実施の形態に含まれる補間用データ生成部 6 4 とは声道断面積関数算出部 1 2 6 を含むか否かの点に違いがあるのみである。従って、動作の詳細な説明は繰返さない。ただし、ここでは補間用データとして混合比算出部 1 2 4 で混合比を算出するのみであるので、補間用 D B 2 0 8 には声道断面積関数の代わりに混合比が保持される。

【0097】

(2) 音声合成装置 2 1 0 の動作

音声合成装置 2 1 0 のうち、出力部 1 3 4、音源 1 3 6、フィルタ 1 3 8、及びクロック部 1 4 0 は、第 1 の実施の形態に係る音声合成装置 9 2 における場合と同様に動作する。従ってそれらについての詳細な説明は繰返さない。

【0098】

図 1 7 を参照して、テキスト 9 0 は、発声すべき音素列からなるテキストと、各音素を発声するための時間情報とを含む。テキスト 9 0 が入力されると、このテキストと時間情報とは混合比抽出部 2 1 2 及び声道断面積関数算出部 2 1 4 に与えられる。

【0099】

混合比抽出部 2 1 2 は、テキスト 9 0 により与えられた音素列のうち隣接する 2 音素の組合せの各々について、その 2 音素を補間するための補間用データである混合比を補間用 D B 2 0 8 からすべて抽出する。声道断面積関数算出部 2 1 4 は、調音パラメータ D B 2 1 6 に保持された調音パラメータから、テキスト 9 0 により与えられた音素に該当する調音パラメータを読み出す。そして読み出された調音パラメータを使用して、式 (5) 及び式 (6) によって抽出された混合比から各音素間の声道断面積関数を算出する。そして、テキスト 9 0 に含まれる 2 音素間を声道断面積関数で補間して補間調音パラメータ 1 3 2 を出力する。この作業を、テキスト 9 0 内の音素列で隣接する 2 音素の組合せの全てについて行ない、補間調音パラメータ 1 3 2 として出力部 1 3 4 に与える。その後の動作は、音声合成装置 9 2 と同様である。

【0100】

なお、フレームシーケンス作成部 1 1 6 (図 1 6 参照) で行なわれる、フレームシーケンスの組を作る処理は、手動で行なわれてもよい。

【0101】

[第 2 の実施の形態の効果]

本実施の形態に係る音声合成システムにおいても、実際の人間の発話における発話器官の動きと一致する様な方法で調音パラメータである声道断面積関数を補間する。そのため、聴覚上、滑らかで自然な連続音声を合成する事ができる。また補間用データ生成装置では 2 音素間の各フレームに対応する混合比のみを算出し、声道断面積関数までは算出しない。そのため、補間用データの生成に要する時間が短くて済み、補間用 D B 2 0 8 として必要な容量も削減できる。また、本実施の形態においても、従来技術と異なり、音声信号

10

20

30

40

50

から補間パラメータの抽出を行なう。その結果、MRI 動画像の撮影に伴う手間及びコストを削減することができる。

【0102】

[コンピュータによる実現]

本発明の第2の実施の形態に係る音声合成システムも、第1の実施の形態と同様に、コンピュータと、当該コンピュータ上で実行されるコンピュータプログラムとにより実現できる。なお、本実施の形態に係る音声合成システムを実現するコンピュータプログラムの制御構造については、第1の実施の形態の説明に基づいて、当業者には容易に実現できると思われる。コンピュータのハードウェア構成についても、第1の実施の形態で説明したものと同様である。従って、ここではそれらについての詳細な説明は繰返さない。

10

【0103】

<第3の実施の形態>

上記した第2の実施の形態では、補間用DB208には、2音素間で声道断面積関数の補間を行なうための混合比を蓄積し、実際の補間時にそれを使用して声道断面積関数を算出し連続音声合成した。しかし、本発明はこのような実施の形態には限定されない。例えば、補間用DBには、声道断面積関数又はその算出に用いる混合比ではなく、スペクトル変化率にガウス関数を割当て、さらに積分及び正規化して得られた発話器官形状関数を蓄積しておき、連続音声合成の際にこれらを用いて調音パラメータである声道断面積関数を算出する方法も考えられる。以下、この方法を適用した第3の実施の形態について説明する。

20

【0104】

[構成]

本発明の第3の実施の形態に係る音声合成システムも、図6に示す第1の実施の形態に係る音声合成システム70と類似した構成である。ただし、この第3の実施の形態に係るシステムの補間用データ生成装置は、第1の実施の形態における補間用データ生成装置86と類似した構成を持つが、補間用データ生成部64（図9参照）に代えて図18に示す補間用データ生成部230を含む点で異なる。またこの第3の実施の形態に係るシステムは、第1の実施の形態における音声合成装置92（図10参照）に代えて図19に示す構成を有する音声合成装置250を含む。

【0105】

(1) 補間用データ生成部230

図18に、本実施の形態に係る補間用データ生成部230のブロック図を示す。図18を参照して、補間用データ生成部230は、第1の実施の形態における補間用データ生成部64と同様の極小値検出部110と、極大値検出部112と、母音割当部114と、フレームシーケンス作成部116と、ガウス関数割当部118と、ガウス関数割当部120と、積分・正規化部124とを含む。第3の実施の形態においては、第1の実施の形態とは異なり、混合比の算出及び声道断面積関数の算出までは行わず、発話器官変化率関数の積分・正規化処理をして発話器官形状関数の算出をするにとどめ、補間用データとしてこの発話器官形状関数を補間用DB246に出力する様にしたものである。

30

【0106】

(2) 音声合成装置250

図19に、本実施の形態に係る音声合成装置250のブロック図を示す。図19を参照して、この音声合成装置250は、図10に示す第1の実施の形態に係る音声合成装置92と類似した構成を持つ。異なるのは、図10に示す補間パラメータ抽出部130に代えて、入力テキスト90の中の隣接する音素の組の各々に対して補間用DB246から積分・正規化された発話器官形状関数を抽出するための補間用データ抽出部252と、抽出された補間用データから混合比を算出するための混合比算出部254と、種々の音素に対して予め算出された調音パラメータを保持するための調音パラメータDB216と、テキスト90により与えられた音素に対応する調音パラメータを調音パラメータDB216から取り出し、混合比を用いた補間により各音素間の声道断面積関数を算出し補間調音パラメ

40

50

ータ 1 3 2 として出力するための、第 2 の実施の形態で使用したものと同一声道断面積関数算出部 2 1 4 (図 1 7 参照) とを含む点である。

【0 1 0 7】

[動作]

第 3 の実施の形態に係る音声合成システムのうち、スペクトル変化率生成部 6 0 は第 1 の実施の形態及び第 2 の実施の形態に係る音声合成システムに含まれるスペクトル変化率生成部 6 0 と同様に動作する。従ってこれについての詳細な説明は繰返さない。

【0 1 0 8】

(1) 補間用データ生成部 2 3 0 の動作

図 1 8 に示す補間用データ生成部 2 3 0 は、第 1 の実施の形態に含まれる補間用データ生成部 6 4 とは混合比算出部 1 2 4 及び声道断面積関数算出部 1 2 6 (図 9 参照) を含むか否かの点に違いがあるのみである。従って、その動作の詳細な説明は繰返さない。ただし、ここでは補間用データとして積分・正規化部 1 2 4 で発話器官変化率関数を積分・正規化して得られた発話器官形状関数を算出するのみであるので、補間用 D B 2 4 6 には混合比の代わりに発話器官形状関数が保持される。

【0 1 0 9】

(2) 音声合成装置 2 5 0 の動作

図 1 9 を参照して、テキスト 9 0 は発声すべき音素列からなるテキストと、各音素を発声するための時間情報とを含む。テキスト 9 0 が入力されると、そのテキストと時間情報とは補間用データ抽出部 2 5 2 及び声道断面積関数算出部 2 1 4 に与えられる。

【0 1 1 0】

補間用データ抽出部 2 5 2 は、テキスト 9 0 を参照し、テキスト中で隣接する各音素の組に対し、補間用 D B 2 4 6 に保持された発話器官形状関数を抽出する。混合比算出部 2 5 4 は前述した式 (1) ~ (4) によって算出されたデータから、二つの音素に対応する調音パラメータの混合比を算出して、声道断面積関数算出部 2 1 4 に与える。その後の動作は、第 2 の実施の形態における音声合成装置 2 1 0 と同様である。

【0 1 1 1】

なお、フレームシーケンス作成部 1 1 6 (図 1 8 参照) で行なわれる、フレームシーケンスの組を作る処理は、手動で行なわれてもよい。

【0 1 1 2】

[第 3 の実施の形態の効果]

本実施の形態に係る音声合成システムにおいても、実際の人間の発話における発話器官の動きと一致する様な方法で調音パラメータである声道断面積関数を補間する。そのため、聴覚上、より滑らかで自然な連続音声を作成できる。補間用データ生成部 2 3 0 ではガウス関数でスペクトル変化率を近似する事により得られた発話器官変化率関数を積分・正規化して発話器官形状関数を算出するのみであり、混合比又は声道断面積関数は算出しない。そのため、補間用データの生成に要する時間が短くて済み、補間 D B 2 4 6 として必要な容量も削減できる。ただし、音声合成装置 2 5 0 において混合比の算出と声道断面積関数の算出との双方を行なうので、第 1 の実施の形態又は第 2 の実施の形態と比較して計算量は大きくなる。また、本実施の形態においても、従来技術と異なり、音声信号から補間パラメータの抽出を行なう。その結果、M R I 動画像の撮影に伴う手間及びコストを削減できる。

【0 1 1 3】

[コンピュータによる実現]

本発明の第 3 の実施の形態に係る音声合成システムにおいても、第 1 の実施の形態と同様に、コンピュータと、当該コンピュータ上で実行されるコンピュータプログラムとにより実現できる。なお、本実施の形態に係る音声合成システムを実現するコンピュータプログラムの制御構造については、第 1 の実施の形態の説明に基づいて、当業者には容易に実現できると思われる。コンピュータのハードウェア構成についても、第 1 の実施の形態で説明したものと同様である。従って、ここではそれらについての詳細な説明は繰返さない

10

20

30

40

50

。

【0114】

今回開示された実施の形態は単に例示であって、本発明が上記した実施の形態のみに制限されるわけではない。本発明の範囲は、発明の詳細な説明の記載を参酌した上で、特許請求の範囲の各請求項によって示され、そこに記載された文言と均等の意味及び範囲内のすべての変更を含む。

【図面の簡単な説明】

【0115】

【図1】発話音声波形の一例とスペクトル変化率の一例とを示すグラフである。

【図2】混合比算出のためのスペクトル変化率の一例を示すグラフである。

10

【図3】スペクトル変化率50、そのガウス関数割当てによる正規分布52、及び混合比算出のために積分処理されたガウス関数の一例を示すグラフである。

【図4】声道断面積関数の例を示す図である。

【図5】声道断面積関数の例を示す図である。

【図6】本発明の第1の実施の形態に係る音声合成システム70のブロック図である。

【図7】補間用データ生成装置86のブロック図である。

【図8】スペクトル変化率生成部60のブロック図である。

【図9】本発明の第1の実施の形態に係る補間用データ生成部64のブロック図である。

【図10】本発明の第1の実施の形態に係る音声合成装置92のブロック図である。

【図11】補間パラメータを算出する処理の詳細なフローチャートである。

20

【図12】音声合成装置92を実現するコンピュータプログラムのフローチャートである。

。

【図13】ステップ160の補間パラメータを抽出する処理の詳細なフローチャートである。

【図14】コンピュータシステムの外観の一例を示す図である。

【図15】図14に示すコンピュータシステムのブロック図である。

【図16】本発明の第2の実施の形態に係る補間用データ生成部190のブロック図である。

【図17】本発明の第2の実施の形態に係る音声合成装置210のブロック図である。

【図18】本発明の第3の実施の形態に係る補間用データ生成部230のブロック図である。

30

【図19】本発明の第3の実施の形態に係る音声合成装置250のブロック図である。

【符号の説明】

【0116】

60 スペクトル変化率生成部

64 補間用データ生成部

100 フレーム化部

108 スペクトル変化率算出部

110 極小値検出部

112 極大値検出部

40

116 フレームシーケンス作成部

118 ガウス関数割当部

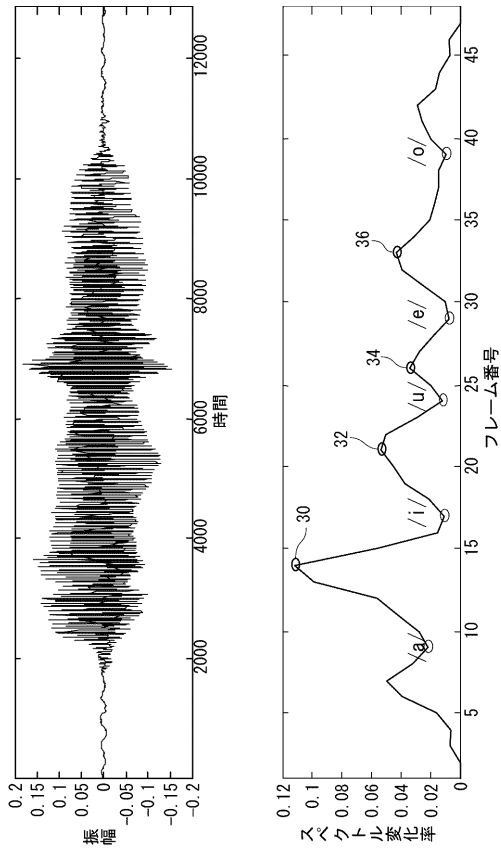
120 ガウス関数割当部

122 積分・正規化部

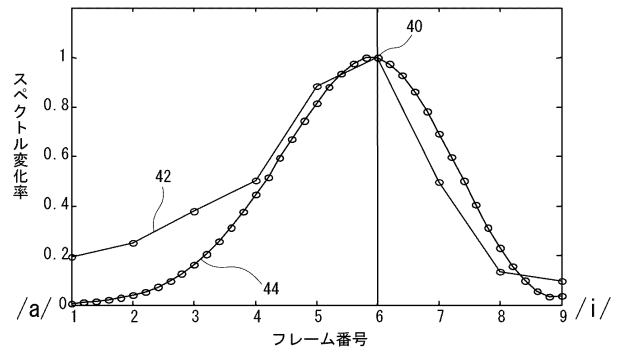
124 混合比算出部

126 声道断面積関数算出部

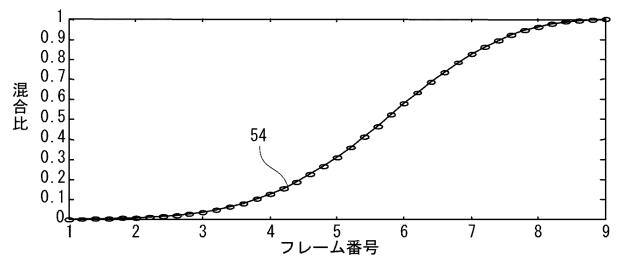
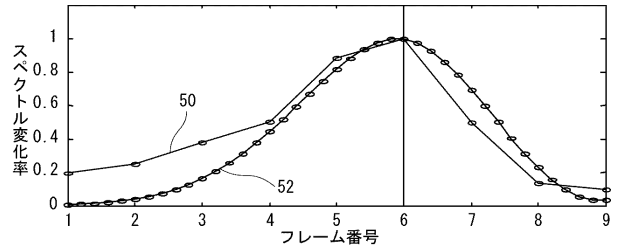
【図 1】



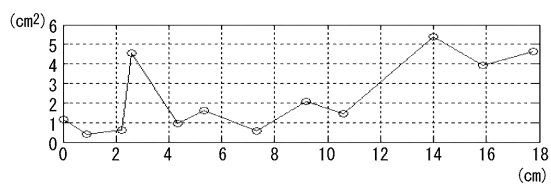
【図 2】



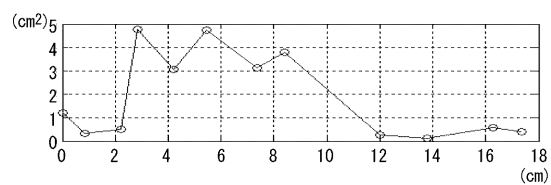
【図 3】



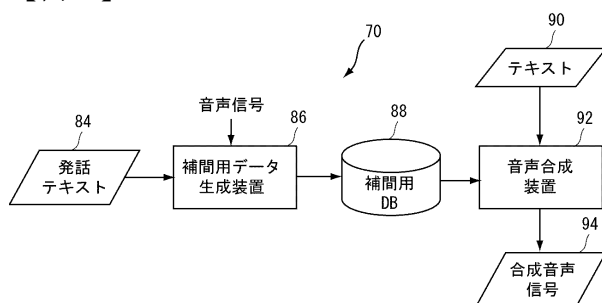
【図 4】



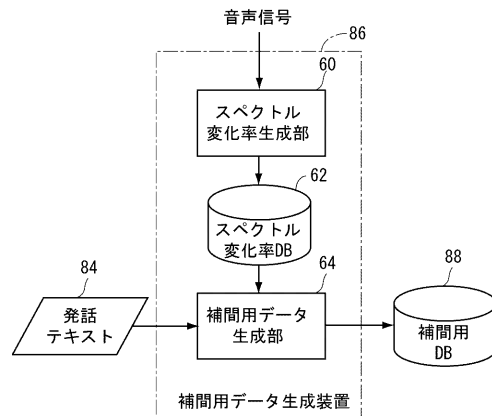
【図 5】



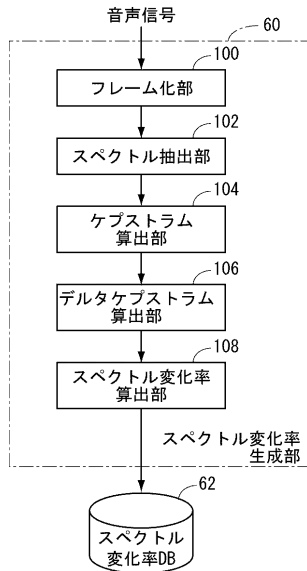
【図 6】



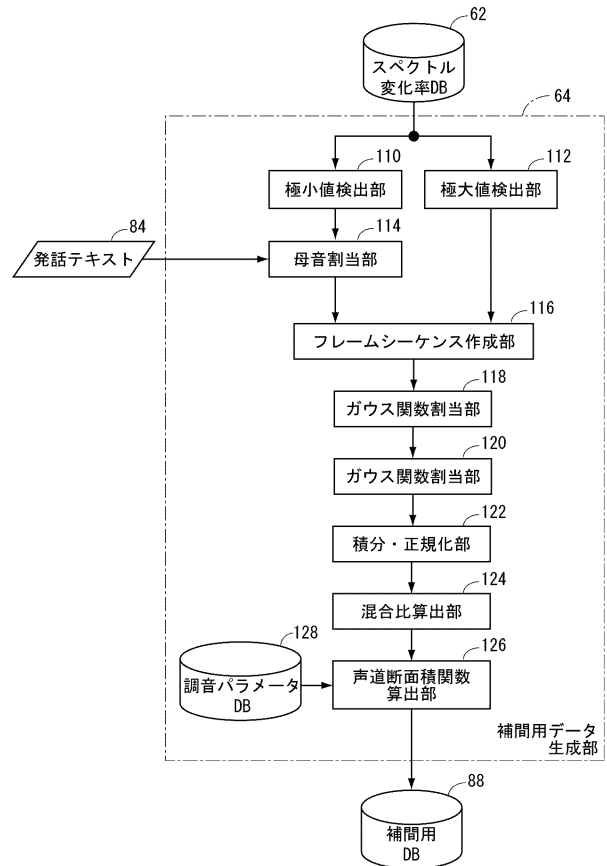
【図 7】



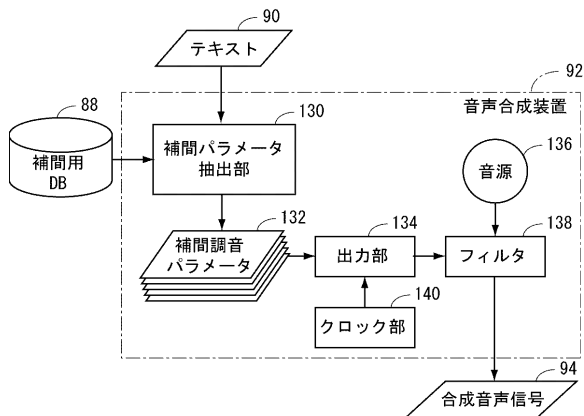
【図 8】



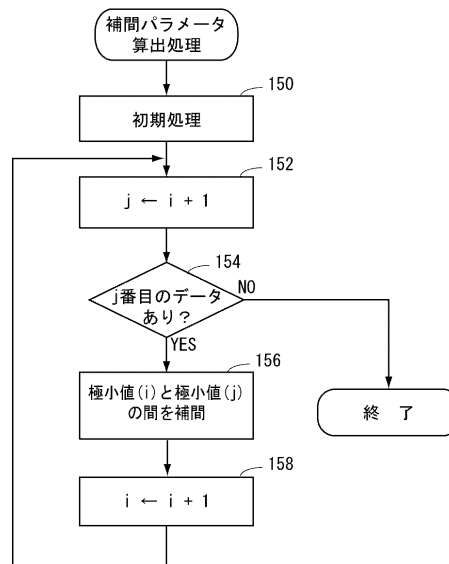
【図 9】



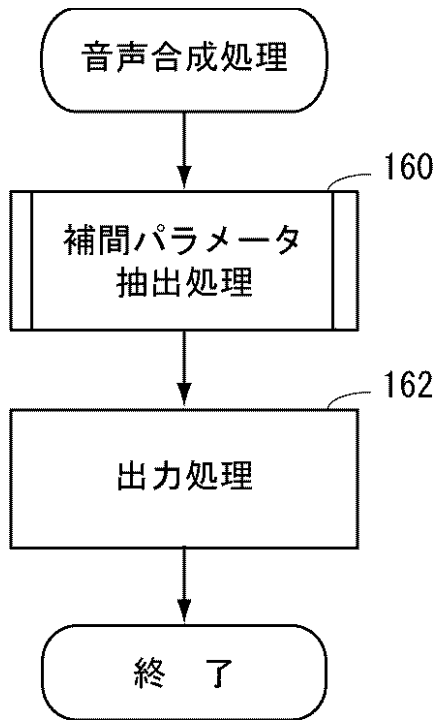
【図 10】



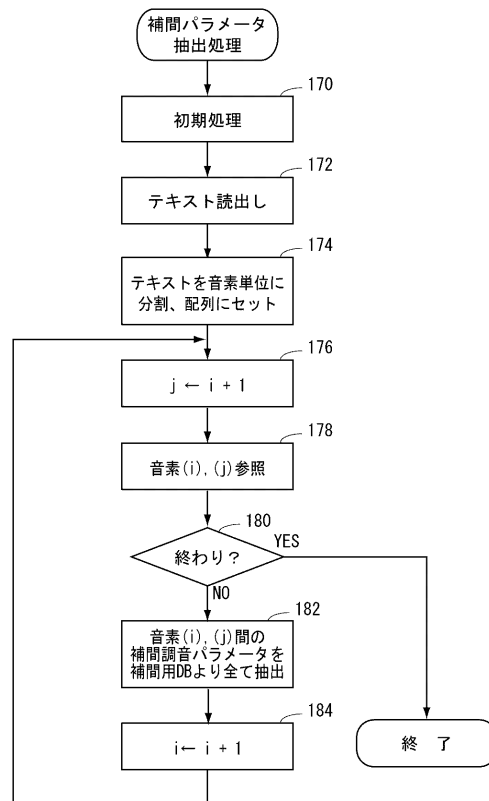
【図 11】



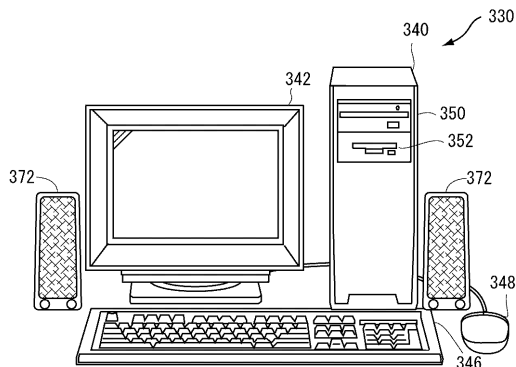
【図 1 2】



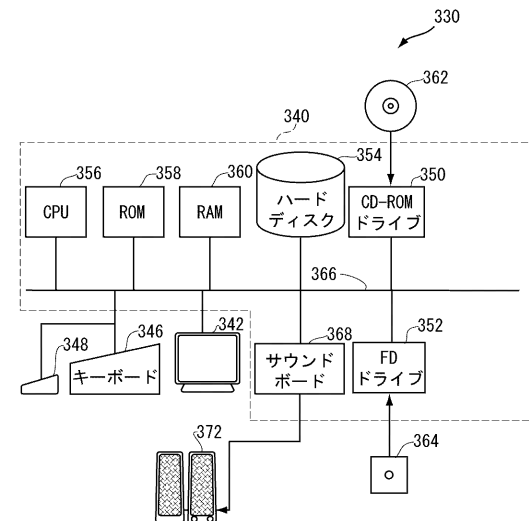
【図 1 3】



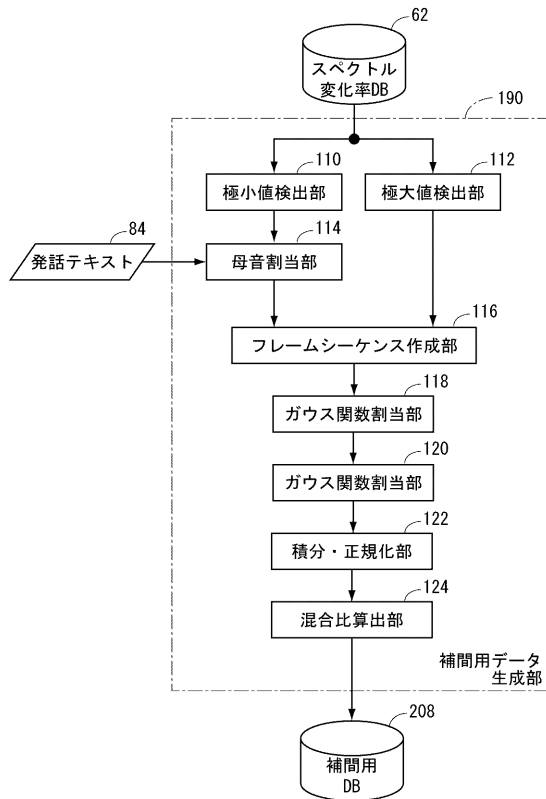
【図 1 4】



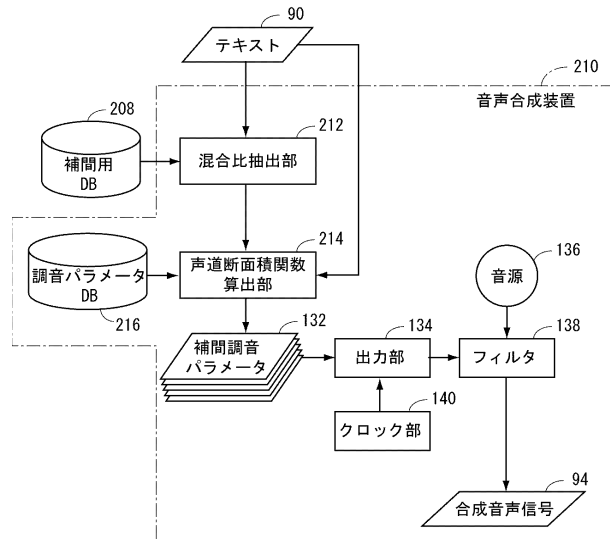
【図 1 5】



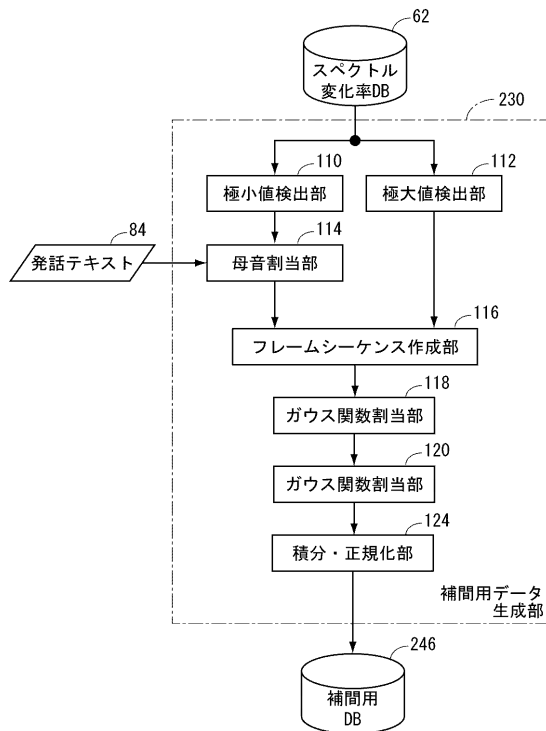
【図 16】



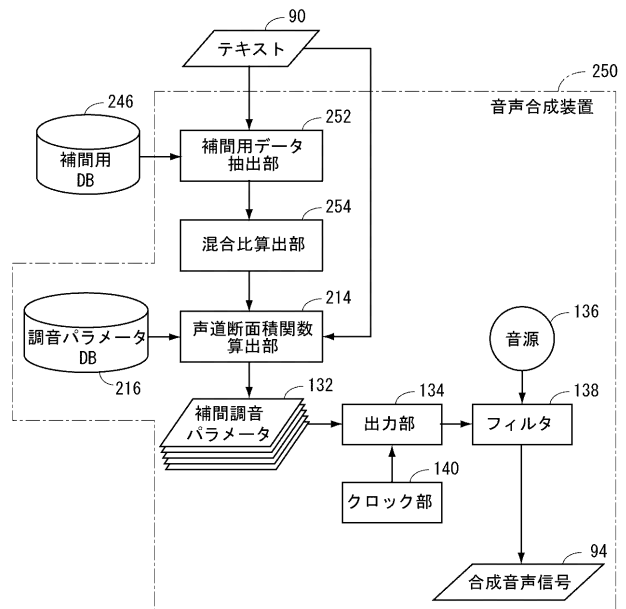
【図 17】



【図 18】



【図 19】



フロントページの続き

(72)発明者 パーハム・モクタリ

京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内

審査官 菊池 智紀

(56)参考文献 特開昭63-210900 (JP, A)

特開2005-134685 (JP, A)

竹本浩典 他, "声道断面積関数の時間補間による音声合成", 日本音響学会2005年春季研究発表
会講演論文集-I, 2005年 3月 8日, p.157-158

(58)調査した分野(Int.Cl., DB名)

G10L 11/00-21/06

JSTPlus (JDreamII)