

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2003-255998

(P2003-255998A)

(43) 公開日 平成15年9月10日 (2003.9.10)

(51) Int.Cl. ⁷	識別記号	F I	テーマコード*(参考)
G 1 0 L 21/04		G 1 0 L 3/02	A 5 D 0 4 5
13/00		5/04	F
13/06		3/00	J

審査請求 未請求 請求項の数17 O L (全 21 頁)

(21) 出願番号 特願2002-52006(P2002-52006)
 (22) 出願日 平成14年2月27日 (2002.2.27)

(71) 出願人 000004075
 ヤマハ株式会社
 静岡県浜松市中沢町10番1号
 (72) 発明者 劍持 秀紀
 静岡県浜松市中沢町10番1号ヤマハ株式会社
 社内
 (72) 発明者 ジョルディ ボナダ
 パセイ デ シルコンバルーラシオ、8.
 08003 パルセロナ
 (74) 代理人 100075074
 弁理士 伊沢 敏昭

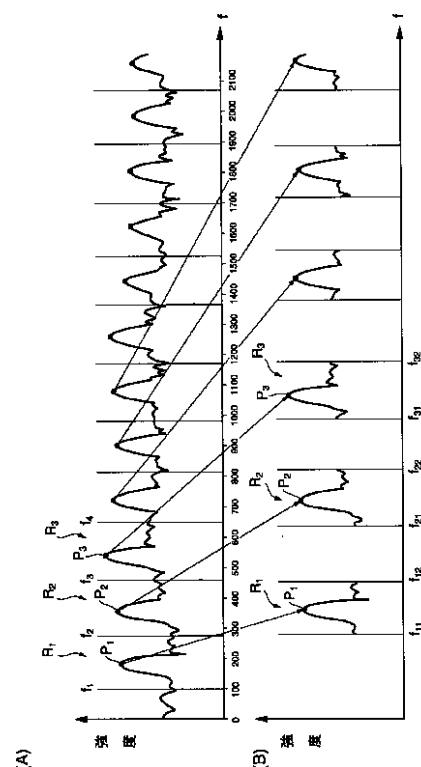
最終頁に続く

(54) 【発明の名称】 歌唱合成方法及び装置及び記録媒体

(57) 【要約】

【課題】 自然な歌唱音声又は高品質の歌唱音声を合成する。

【解決手段】 音素又は音素連鎖からなる音声素片に対応する音声波形を周波数分析して周波数スペクトル (A) を検出する。周波数スペクトル (A) 上で P_1 等の局所的ピークを検知し、これらのピークを含む R_1 等のスペクトル分布領域を指定する。各スペクトル分布領域毎に、振幅スペクトル分布を周波数軸に関して表わす振幅スペクトルデータと、位相スペクトル分布を周波数軸に関して表わす位相スペクトルデータとを生成する。各スペクトル分布領域の振幅スペクトル分布を入力音符ピッチに応じて周波数軸上で (B) の様に移動すべく振幅スペクトルデータを修正し、この修正に対応して位相スペクトルデータを修正する。所望の音色に対応するスペクトル包絡に沿うようにスペクトル強度を調整する。修正された振幅及び位相スペクトルデータを時間領域の合成音声信号に変換する。



【特許請求の範囲】

【請求項1】合成すべき音声の音声素片に対応する音声波形を周波数分析して周波数スペクトルを検出するステップと、

前記周波数スペクトル上でスペクトル強度の局所的ピークを複数検知するステップと、

各局所的ピーク毎に該局所的ピークとその前後のスペクトルとを含むスペクトル分布領域を前記周波数スペクトル上で指定し、各スペクトル分布領域毎に振幅スペクトル分布を周波数軸に関して表わす振幅スペクトルデータを生成するステップと、

各スペクトル分布領域毎に位相スペクトル分布を周波数軸に関して表わす位相スペクトルデータを生成するステップと、

前記合成すべき音声についてピッチを指定するステップと、

各スペクトル分布領域毎に前記振幅スペクトルデータが表わす振幅スペクトル分布を前記ピッチに応じて周波数軸上で移動するように前記振幅スペクトルデータを修正するステップと、

各スペクトル分布領域毎に前記位相スペクトルデータが表わす位相スペクトル分布を前記振幅スペクトルデータの修正に対応して修正するステップと、

前記修正に係る振幅スペクトルデータ及び前記修正に係る位相スペクトルデータを時間領域の合成音声信号に変換するステップとを含む歌唱合成方法。

【請求項2】合成すべき音声の音声素片に対応する振幅スペクトルデータ及び位相スペクトルデータを取得するステップであって、前記振幅スペクトルデータとして

は、前記音声素片の音声波形を周波数分析して得た周波数スペクトルにおいてスペクトル強度の複数の局所的ピークのうちの各局所的ピーク毎に該局所的ピークとその前後のスペクトルとを含むスペクトル分布領域における振幅スペクトル分布を周波数軸に関して表わすデータを取得し、前記位相スペクトルデータとしては、各スペクトル分布領域毎に位相スペクトル分布を周波数軸に関して表わすデータを取得するものと、

前記合成すべき音声についてピッチを指定するステップと、

各スペクトル分布領域毎に前記振幅スペクトルデータが表わす振幅スペクトル分布を前記ピッチに応じて周波数軸上で移動するように前記振幅スペクトルデータを修正するステップと、

各スペクトル分布領域毎に前記位相スペクトルデータが表わす位相スペクトル分布を前記振幅スペクトルデータの修正に対応して修正するステップと、

前記修正に係る振幅スペクトルデータ及び前記修正に係る位相スペクトルデータを時間領域の合成音声信号に変換するステップとを含む歌唱合成方法。

【請求項3】前記ピッチを指定するステップでは、経

時的なピッチ変化を示すピッチゆらぎデータに従って前記ピッチを指定する請求項1又は2記載の歌唱合成方法。

【請求項4】前記ピッチゆらぎデータとしては、前記合成すべき音声について音楽的表情を制御するための制御パラメータに対応したピッチゆらぎデータを用いる請求項3記載の歌唱合成方法。

【請求項5】前記振幅スペクトルデータを修正するステップでは、修正前の複数の局所的ピークを結ぶ線に対応するスペクトル包絡に沿わない局所的ピークについてスペクトル強度を該スペクトル包絡に沿うように修正する請求項1又は2記載の歌唱合成方法。

【請求項6】前記振幅スペクトルデータを修正するステップでは、予め定めたスペクトル包絡に沿わない局所的ピークについてスペクトル強度を該スペクトル包絡に沿うように修正する請求項1又は2記載の歌唱合成方法。

【請求項7】前記振幅スペクトルデータを修正するステップでは、一連の時間フレームについて経時的なスペクトル包絡の変化を示すスペクトル包絡ゆらぎデータに従ってスペクトル強度を調整することにより経時的に変化するスペクトル包絡を設定する請求項5又は6記載の歌唱合成方法。

【請求項8】前記スペクトル包絡ゆらぎデータとしては、前記合成すべき音声について音楽的表情を制御するための制御パラメータに対応したスペクトル包絡ゆらぎデータを用いる請求項7記載の歌唱合成方法。

【請求項9】合成すべき音声について音声素片及びピッチを指定する指定手段と、

30 音声素片データベースから音声素片データとして前記音声素片に対応する音声波形を表わす音声波形データを読出す読出手段と、

前記音声波形データが表わす音声波形を周波数分析して周波数スペクトルを検出する検出手段と、

前記音声波形に対応する周波数スペクトル上でスペクトル強度の局所的ピークを複数検知する検知手段と、

各局所的ピーク毎に該局所的ピークとその前後のスペクトルとを含むスペクトル分布領域を前記周波数スペクトル上で指定し、各スペクトル分布領域毎に振幅スペクトル分布を周波数軸に関して表わす振幅スペクトルデータを生成する第1の生成手段と、

各スペクトル分布領域毎に位相スペクトル分布を周波数軸に関して表わす位相スペクトルデータを生成する第2の生成手段と、

各スペクトル分布領域毎に前記振幅スペクトルデータが表わす振幅スペクトル分布を前記ピッチに応じて周波数軸上で移動するように前記振幅スペクトルデータを修正する第1の修正手段と、

50 各スペクトル分布領域毎に前記位相スペクトルデータが表わす位相スペクトル分布を前記振幅スペクトルデータ

の修正に対応して修正する第2の修正手段と、前記修正に係る振幅スペクトルデータ及び前記修正に係る位相スペクトルデータを時間領域の合成音声信号に変換する変換手段とを備えた歌唱合成装置。

【請求項10】合成すべき音声について音声素片及びピッチを指定する指定手段と、

音声素片データベースから音声素片データとして前記音声素片に対応する振幅スペクトルデータ及び位相スペクトルデータを読み出す読み出し手段であって、前記振幅スペクトルデータとしては、前記音声素片の音声波形を周波数分析して得た周波数スペクトルにおいてスペクトル強度の複数の局所的ピークのうちの各局所的ピーク毎に該局所的ピークとその前後のスペクトルとを含むスペクトル分布領域における振幅スペクトル分布を周波数軸に関して表わすデータを読み出し、前記位相スペクトルデータとしては、各スペクトル分布領域毎に位相スペクトル分布を周波数軸に関して表わすデータを読み出すものと、各スペクトル分布領域毎に前記振幅スペクトルデータが表わす振幅スペクトル分布を前記ピッチに応じて周波数軸上で移動するように前記振幅スペクトルデータを修正する第1の修正手段と、

各スペクトル分布領域毎に前記位相スペクトルデータが表わす位相スペクトル分布を前記振幅スペクトルデータの修正に対応して修正する第2の修正手段と、前記修正に係る振幅スペクトルデータ及び前記修正に係る位相スペクトルデータを時間領域の合成音声信号に変換する変換手段とを備えた歌唱合成装置。

【請求項11】前記指定手段は、前記合成すべき音声について音楽的表情を制御するための制御パラメータを指定し、前記読み出し手段は、前記音声素片及び前記制御パラメータに対応する音声素片データを読み出す請求項9又は10記載の歌唱合成装置。

【請求項12】前記指定手段は、前記合成すべき音声について音符長及び/又はテンポを指定し、前記読み出し手段は、前記音声素片データを読み出す際に前記音声素片データの一部を省略するか又は前記音声素片データの一部もしくは全部を繰返すかして前記音符長及び/又はテンポに対応する時間のあいだ前記音声素片データの読み出しを継続する請求項9又は10記載の歌唱合成装置。

【請求項13】順次に合成すべき音声のうちの各音声毎に音声素片及びピッチを指定する指定手段と、音声素片データベースから前記指定手段での指定に係る各音声素片に対応する音声波形を読み出す読み出し手段と、各音声素片に対応する音声波形を周波数分析して周波数スペクトルを検出する検出手段と、各音声素片に対応する周波数スペクトル上でスペクトル強度の局所的ピークを複数検知する検知手段と、各音声素片について各局所的ピーク毎に該局所的ピークとその前後のスペクトルとを含むスペクトル分布領域を該音声素片に対応する周波数スペクトル上で指定し、各

音声素片について各スペクトル分布領域毎に振幅スペクトル分布を周波数軸に関して表わす振幅スペクトルデータを生成する第1の生成手段と、

各音声素片について各スペクトル分布領域毎に位相スペクトル分布を周波数軸に関して表わす位相スペクトルデータを生成する第2の生成手段と、

各音声素片について各スペクトル分布領域毎に前記振幅スペクトルデータが表わす振幅スペクトル分布を該音声素片に対応するピッチに応じて周波数軸上で移動するように前記振幅スペクトルデータを修正する第1の修正手段と、

各音声素片について各スペクトル分布領域毎に前記位相スペクトルデータが表わす位相スペクトル分布を前記振幅スペクトルデータの修正に対応して修正する第2の修正手段と、

前記順次に合成すべき音声にそれぞれ対応する順次の音声素片が発音順につながるように前記修正に係る振幅スペクトルデータを接続する第1の接続手段であって、前記順次の音声素片のつながり部においてスペクトル強度を一致又は近似させるべく調整するものと、

前記順次に合成すべき音声にそれぞれ対応する順次の音声素片が発音順につながるように前記修正に係る位相スペクトルデータを接続する第2の接続手段であって、前記順次の音声素片のつながり部において位相を一致又は近似させるべく調整するものと、

前記接続に係る振幅スペクトルデータ及び前記接続に係る位相スペクトルデータを時間領域の合成音声信号に変換する変換手段とを備えた歌唱合成装置。

【請求項14】順次に合成すべき音声のうちの各音声毎に音声素片及びピッチを指定する指定手段と、

音声素片データベースから前記指定手段での指定に係る各音声素片に対応する振幅スペクトルデータ及び位相スペクトルデータを読み出す読み出し手段であって、前記振幅スペクトルデータとしては、対応する音声素片の音声波形を周波数分析して得た周波数スペクトルにおいてスペクトル強度の複数の局所的ピークのうちの各局所的ピーク毎に該局所的ピークとその前後のスペクトルとを含むスペクトル分布領域における振幅スペクトル分布を周波数軸に関して表わすデータを読み出し、前記位相スペクトルデータとしては、各スペクトル分布領域毎に位相スペクトル分布を周波数軸に関して表わすデータを読み出すものと、

各音声素片について各スペクトル分布領域毎に前記振幅スペクトルデータが表わす振幅スペクトル分布を該音声素片に対応するピッチに応じて周波数軸上で移動するように前記振幅スペクトルデータを修正する第1の修正手段と、

各音声素片について各スペクトル分布領域毎に前記位相スペクトルデータが表わす位相スペクトル分布を前記振幅スペクトルデータの修正に対応して修正する第2の修

正手段と、
前記順次に合成すべき音声にそれぞれ対応する順次の音声素片が発音順につながるように前記修正に係る振幅スペクトルデータを接続する第1の接続手段であって、前記順次の音声素片のつながり部においてスペクトル強度を一致又は近似させるべく調整するものと、
前記順次に合成すべき音声にそれぞれ対応する順次の音声素片が発音順につながるように前記修正に係る位相スペクトルデータを接続する第2の接続手段であって、前記順次の音声素片のつながり部において位相を一致又は近似させるべく調整するものと、
前記接続に係る振幅スペクトルデータ及び前記接続に係る位相スペクトルデータを時間領域の合成音声信号に変換する変換手段とを備えた歌唱合成装置。

【請求項15】 歌唱合成装置において使用される記録媒体であって、

合成すべき音声の音声素片に対応する音声波形を周波数分析して周波数スペクトルを検出するステップと、
前記周波数スペクトル上でスペクトル強度の局所的ピークを複数検知するステップと、

各局所的ピーク毎に該局所的ピークとその前後のスペクトルとを含むスペクトル分布領域を前記周波数スペクトル上で指定し、各スペクトル分布領域毎に振幅スペクトル分布を周波数軸に関して表わす振幅スペクトルデータを生成するステップと、

各スペクトル分布領域毎に位相スペクトル分布を周波数軸に関して表わす位相スペクトルデータを生成するステップと、

前記合成すべき音声についてピッチを指定するステップと、

各スペクトル分布領域毎に前記振幅スペクトルデータが表わす振幅スペクトル分布を前記ピッチに応じて周波数軸上で移動するように前記振幅スペクトルデータを修正するステップと、

各スペクトル分布領域毎に前記位相スペクトルデータが表わす位相スペクトル分布を前記振幅スペクトルデータの修正に対応して修正するステップと、

前記修正に係る振幅スペクトルデータ及び前記修正に係る位相スペクトルデータを時間領域の合成音声信号に変換するステップとを含むプログラムを記録した記録媒体。

【請求項16】 歌唱合成装置において使用される記録媒体であって、

合成すべき音声の音声素片に対応する振幅スペクトルデータ及び位相スペクトルデータを取得するステップであって、前記振幅スペクトルデータとしては、前記音声素片の音声波形を周波数分析して得た周波数スペクトルにおいてスペクトル強度の複数の局所的ピークのうちの各局所的ピーク毎に該局所的ピークとその前後のスペクトルとを含むスペクトル分布領域における振幅スペクトル

分布を周波数軸に関して表わすデータを取得し、前記位相スペクトルデータとしては、各スペクトル分布領域毎に位相スペクトル分布を周波数軸に関して表わすデータを取得するものと、

前記合成すべき音声についてピッチを指定するステップと、

各スペクトル分布領域毎に前記振幅スペクトルデータが表わす振幅スペクトル分布を前記ピッチに応じて周波数軸上で移動するように前記振幅スペクトルデータを修正するステップと、

各スペクトル分布領域毎に前記位相スペクトルデータが表わす位相スペクトル分布を前記振幅スペクトルデータの修正に対応して修正するステップと、

前記修正に係る振幅スペクトルデータ及び前記修正に係る位相スペクトルデータを時間領域の合成音声信号に変換するステップとを含むプログラムを記録した記録媒体。

【請求項17】 歌唱合成装置において使用される記録媒体であって、音声素片毎に音声合成用の振幅スペクトルデータ及び位相スペクトルデータが記録されており、

各振幅スペクトルデータとしては、対応する音声素片の音声波形を周波数分析して得た周波数スペクトルにおいてスペクトル強度の複数の局所的ピークのうちの各局所的ピーク毎に該局所的ピークとその前後のスペクトルとを含むスペクトル分布領域における振幅スペクトル分布を周波数軸に関して表わすデータが記録されており、各位相スペクトルデータとしては、各スペクトル分布領域毎に位相スペクトル分布を周波数軸に関して表わすデータ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】この発明は、フェーズボコーダ技術を用いて歌唱音声を合成する方法と装置及び記録媒体に関するものである。

【0002】

【従来の技術】従来、歌唱合成技術としては、米国特許第5029509号明細書等により公知のSMS(Spectral Modeling Synthesis)技術を用いて歌唱合成を行なうものが知られている(例えば、特許第2906970号参照)。

【0003】図21は、特許2906970号に示される技術を採用した歌唱合成装置を示すものである。ステップS1では、歌唱音声信号を入力し、ステップS2では、入力された歌唱音声信号にSMS分析処理及び区間切出し処理を施す。

【0004】SMS分析処理では、入力音声信号を一連の時間フレームに区分し、各フレーム毎にFFT(Fast Fourier Transform)等により1組の強度(マグニチュード)スペクトルデータを生成し、各フレーム毎に1組の強度スペクトルデータから複数のピークに対応する

線スペクトルを抽出する。これらの線スペクトルの振幅値及び周波数を表わすデータを調和成分 (Deterministic Component) のデータと称する。次に、入力音声波形のスペクトルから調和成分のスペクトルを差引いて残差スペクトルを得る。この残差スペクトルを非調和成分 (Stochastic Component) と称する。

【0005】区間切出し処理では、SMS分析処理で得られた調和成分のデータ及び非調和成分のデータを音声素片に対応して区分する。音声素片とは、歌詞の構成要素であり、例えば [a] , [i] のような単一の音素 (又は音韻 : Phoneme) 又は例えば 「 a_i 」 , [a_p] のような音素連鎖 (複数音素の連鎖) からなるものである。

【0006】音声素片データベースDBには、音声素片毎に調和成分のデータ及び非調和成分のデータが記憶される。

【0007】歌唱合成に際しては、ステップS3で歌詞データ及びメロディデータを入力する。そして、ステップS4では、歌詞データが表わす音素列に音素列/音声素片変換処理を施して音素列を音声素片に区分し、音声素片毎にそれに対応する調和成分のデータ及び非調和成分のデータを音声素片データとしてデータベースDBから読出す。

【0008】ステップS5では、データベースDBから読出された音声素片データ (調和成分のデータ及び非調和成分のデータ) に音声素片接続処理を施して音声素片データ同士を発音順に接続する。ステップS6では、音声素片毎に調和成分のデータと入力メロディデータの示す音符ピッチとに基づいて該音符ピッチに適合した新たな調和成分のデータを生成する。このとき、新たな調和成分のデータでは、ステップS5の処理を受けた調和成分のデータが表わすスペクトル包絡の形状をそのまま引継ぐようにスペクトル強度を調整すると、ステップS1で入力した音声信号の音色を再現することができる。

【0009】ステップS7では、ステップS6で生成した調和成分のデータとステップS5の処理を受けた非調和成分のデータとを音声素片毎に加算する。そして、ステップS8では、ステップS7で加算処理を受けたデータを音声素片毎に逆FFT等により時間領域の合成音声信号に変換する。

【0010】一例として、「サイタ」(saita) という歌唱音声合成するには、データベースDBから音声素片「#s」、「sa」、「a」、「ai」、「i」、「it」、「ta」、「a」、「a#」(#は無音を表わす) にそれぞれ対応する音声素片データを読出してステップS5で接続する。そして、ステップS6で音声素片毎に入力音符ピッチに対応するピッチを有する調和成分のデータを生成し、ステップS7の加算処理及びステップS8の変換処理を経ると、「サイタ」の歌唱合成音声信号が得られる。

【0011】

【発明が解決しようとする課題】上記した従来技術によると、調和成分と非調和成分の一体感が十分でないという問題点がある。すなわち、ステップS1で入力した音声信号のピッチをステップS6で入力音符ピッチに対応して変更し、変更したピッチを有する調和成分のデータにステップS7で非調和成分のデータを加算するため、例えば、「サイタ」の歌唱における「i」のような伸ばし音の区間で非調和成分が分離して響き、人工的な音声として聴こえるという問題点がある。

【0012】このような問題点に対処するため、非調和成分のデータが表わす低域の振幅スペクトル分布を入力音符ピッチに応じて修正することを本願出願人は先に提案した (特願2000-401041参照)。しかし、このように非調和成分のデータを修正しても、非調和成分が分離して響くのを完全に抑えるのは容易でない。

【0013】また、SMS技術にあつては、有音の摩擦音や破裂音等の分析が難しく、合成音が非常に人工的な音になってしまうという問題点もある。SMS技術は、音声信号が調和成分と非調和成分とから成り立っていることを前提にしているものであり、音声信号を調和成分と非調和成分とに完全に分離できないことは、SMS技術にとって根本的な問題といえる。

【0014】一方、フェーズボコーダ技術は、米国特許第3360610号明細書に示されている。フェーズボコーダ技術では、古くはフィルタバンクとして、新しくは入力信号のFFT結果として周波数領域で信号を表現する。最近では、フェーズボコーダ技術が楽音の時間軸圧伸 (ピッチをそのままにして時間だけ圧縮又は伸張する) やピッチ変換 (時間長はそのままにしてピッチだけ変化させる) などに広く利用されている。この種のピッチ変換技術としては、入力信号のFFT結果をそのまま用いるのではなく、FFTスペクトルを局所的ピークを中心とした複数のスペクトル分布領域に分割し、各スペクトル分布領域毎にスペクトル分布を周波数軸上で移動することによりピッチ変換を行なうものが知られている (例えば、J. Laroche and M. Dolson, "New Phase-Vocoder Techniques for Real-Time Pitch Shifting, Chorusing, Harmonizing, and Other Exotic Audio Modifications" J. Audio Eng. Soc., Vol. 47, No. 11, 1999 November 参照)。しかし、このようなピッチ変換技術と歌唱合成技術との関連性については明らかにされていない。

【0015】この発明の目的は、フェーズボコーダ技術を用いて自然で高品質の音声合成を可能にした新規な歌唱合成方法と装置及び記録媒体を提供することにある。

【0016】

【課題を解決するための手段】この発明に係る第1の歌唱合成方法は、合成すべき音声の音声素片に対応する音声波形を周波数分析して周波数スペクトルを検出するス

テップと、前記周波数スペクトル上でスペクトル強度の局所的ピークを複数検知するステップと、各局所的ピーク毎に該局所的ピークとその前後のスペクトルを含むスペクトル分布領域を前記周波数スペクトル上で指定し、各スペクトル分布領域毎に振幅スペクトル分布を周波数軸に関して表わす振幅スペクトルデータを生成するステップと、各スペクトル分布領域毎に位相スペクトル分布を周波数軸に関して表わす位相スペクトルデータを生成するステップと、前記合成すべき音声についてピッチを指定するステップと、各スペクトル分布領域毎に前記振幅スペクトルデータが表わす振幅スペクトル分布を前記ピッチに応じて周波数軸上で移動するように前記振幅スペクトルデータを修正するステップと、各スペクトル分布領域毎に前記位相スペクトルデータが表わす位相スペクトル分布を前記振幅スペクトルデータの修正に対応して修正するステップと、前記修正に係る振幅スペクトルデータ及び前記修正に係る位相スペクトルデータを時間領域の合成音声信号に変換するステップとを含むものである。

【0017】第1の歌唱合成方法によれば、音声素片（音素又は音素連鎖）に対応する音声波形が周波数分析されて周波数スペクトルが検出される。そして、周波数スペクトルに基づいて振幅スペクトルデータと、位相スペクトルデータとが生成される。所望のピッチが指定されると、指定のピッチに応じて振幅スペクトルデータ及び位相スペクトルデータが修正され、修正に係る振幅スペクトルデータ及び位相スペクトルデータに基づいて時間領域の合成音声信号が発生される。このように音声波形の周波数分析結果を調和成分と非調和成分とに分離しないで音声合成を行なうため、非調和成分が分離して響くことがなく、自然な合成音を得ることができる。また、有声の摩擦音や破裂音であっても自然な合成音が得られる。

【0018】この発明に係る第2の歌唱合成方法は、合成すべき音声の音声素片に対応する振幅スペクトルデータ及び位相スペクトルデータを取得するステップであって、前記振幅スペクトルデータとしては、前記音声素片の音声波形を周波数分析して得た周波数スペクトルにおいてスペクトル強度の複数の局所的ピークのうちの各局所的ピーク毎に該局所的ピークとその前後のスペクトルを含むスペクトル分布領域における振幅スペクトル分布を周波数軸に関して表わすデータを取得し、前記位相スペクトルデータとしては、各スペクトル分布領域毎に位相スペクトル分布を周波数軸に関して表わすデータを取得するものと、前記合成すべき音声についてピッチを指定するステップと、各スペクトル分布領域毎に前記振幅スペクトルデータが表わす振幅スペクトル分布を前記ピッチに応じて周波数軸上で移動するように前記振幅スペクトルデータを修正するステップと、各スペクトル分布領域毎に前記位相スペクトルデータが表わす位相スペ

クトル分布を前記振幅スペクトルデータの修正に対応して修正するステップと、前記修正に係る振幅スペクトルデータ及び前記修正に係る位相スペクトルデータを時間領域の合成音声信号に変換するステップとを含むものである。

【0019】第2の歌唱合成方法は、第1の歌唱合成方法において、位相スペクトルデータを生成するステップまでの処理を予め実行して振幅スペクトルデータ及び位相スペクトルデータを音声素片毎にデータベースに記憶しておいた場合、又は位相スペクトルデータを生成するステップまでの処理を他の機器で実行する場合に相当する。すなわち、第2の歌唱合成方法において、取得するステップでは、他の機器又はデータベースから合成すべき音声の音声素片に対応する振幅スペクトルデータ及び位相スペクトルデータを取得し、ピッチを指定するステップ以降の処理を第1の歌唱合成方法と同様に実行する。従って、第2の歌唱合成方法によれば、第1の歌唱合成方法と同様に自然な合成音を得られる。

【0020】第1又は第2の歌唱合成方法において、前記ピッチを指定するステップでは、経時的なピッチ変化を示すピッチゆらぎデータに従って前記ピッチを指定するようにしてもよい。このようにすると、合成音のピッチを経時的に変化させることができ、例えばピッチバンド、ビブラート等を付加することができる。また、前記ピッチゆらぎデータとしては、前記合成すべき音声について音楽的表情を制御するための制御パラメータに対応したピッチゆらぎデータを用いるようにしてもよい。このようにすると、例えば音色、ダイナミクス等の制御パラメータに応じて経時的なピッチ変化態様を異ならせることができる。

【0021】第1又は第2の歌唱合成方法において、前記振幅スペクトルデータを修正するステップでは、修正前の複数の局所的ピークを結ぶ線に対応するスペクトル包絡に沿わない局所的ピークについてスペクトル強度を該スペクトル包絡に沿うように修正するようにしてもよい。このようにすると、元の音声波形の音色を再現することができる。また、前記振幅スペクトルデータを修正するステップでは、予め定めたスペクトル包絡に沿わない局所的ピークについてスペクトル強度を該スペクトル包絡に沿うように修正するようにしてもよい。このようにすると、元の音声波形とは音色を異ならせることができる。

【0022】上記のようにスペクトル強度をスペクトル包絡に沿うように修正する場合において、前記振幅スペクトルデータを修正するステップでは、一連の時間フレームについて経時的なスペクトル包絡の変化を示すスペクトル包絡ゆらぎデータに従ってスペクトル強度を調整することにより経時的に変化するスペクトル包絡を設定するようにしてもよい。このようにすると、合成音の音色を経時的に変化させることができ、例えばトーンベン

ド等を付加することができる。また、前記スペクトル包絡ゆらぎデータとしては、前記合成すべき音声について音楽的表情を制御するための制御パラメータに対応したスペクトル包絡ゆらぎデータを用いるようにしてもよい。このようにすると、例えば音色、ダイナミクス等の制御パラメータに応じて経時的な音色変化態様を異ならせることができる。

【0023】この発明に係る第1の歌唱合成装置は、合成すべき音声について音声素片及びピッチを指定する指定手段と、音声素片データベースから音声素片データとして前記音声素片に対応する音声波形を表わす音声波形データを読み出す読み出し手段と、前記音声波形データが表わす音声波形を周波数分析して周波数スペクトルを検出する検出手段と、前記音声波形に対応する周波数スペクトル上でスペクトル強度の局所的ピークを複数検知する検知手段と、各局所的ピーク毎に該局所的ピークとその前後のスペクトルとを含むスペクトル分布領域を前記周波数スペクトル上で指定し、各スペクトル分布領域毎に振幅スペクトル分布を周波数軸に関して表わす振幅スペクトルデータを生成する第1の生成手段と、各スペクトル分布領域毎に位相スペクトル分布を周波数軸に関して表わす位相スペクトルデータを生成する第2の生成手段と、各スペクトル分布領域毎に前記振幅スペクトルデータが表わす振幅スペクトル分布を前記ピッチに応じて周波数軸上で移動するように前記振幅スペクトルデータを修正する第1の修正手段と、各スペクトル分布領域毎に前記位相スペクトルデータが表わす位相スペクトル分布を前記振幅スペクトルデータの修正に対応して修正する第2の修正手段と、前記修正に係る振幅スペクトルデータ及び前記修正に係る位相スペクトルデータを時間領域の合成音声信号に変換する変換手段とを備えたものである。

【0024】また、この発明に係る第2の歌唱合成装置は、合成すべき音声について音声素片及びピッチを指定する指定手段と、音声素片データベースから音声素片データとして前記音声素片に対応する振幅スペクトルデータ及び位相スペクトルデータを読み出す読み出し手段であって、前記振幅スペクトルデータとしては、前記音声素片の音声波形を周波数分析して得た周波数スペクトルにおいてスペクトル強度の複数の局所的ピークのうちの各局所的ピーク毎に該局所的ピークとその前後のスペクトルとを含むスペクトル分布領域における振幅スペクトル分布を周波数軸に関して表わすデータを読み出し、前記位相スペクトルデータとしては、各スペクトル分布領域毎に位相スペクトル分布を周波数軸に関して表わすデータを読み出すものと、各スペクトル分布領域毎に前記振幅スペクトルデータが表わす振幅スペクトル分布を前記ピッチに応じて周波数軸上で移動するように前記振幅スペクトルデータを修正する第1の修正手段と、各スペクトル分布領域毎に前記位相スペクトルデータが表わす位相スペ

クトル分布を前記振幅スペクトルデータの修正に対応して修正する第2の修正手段と、前記修正に係る振幅スペクトルデータ及び前記修正に係る位相スペクトルデータを時間領域の合成音声信号に変換する変換手段とを備えたものである。

【0025】第1又は第2の歌唱合成装置は、音声素片データベースを用いて前述の第1又は第2の歌唱合成方法を実施するものであり、自然な歌唱合成音を得ることができる。

10 【0026】第1又は第2の歌唱合成装置において、前記指定手段は、前記合成すべき音声について音楽的表情を制御するための制御パラメータを指定し、前記読み出し手段は、前記音声素片及び前記制御パラメータに対応する音声素片データを読み出すようにしてもよい。このようにすると、例えば音色、ダイナミクス等の制御パラメータに最適の音声素片データを用いて歌唱合成を行なうことができる。

20 【0027】第1又は第2の歌唱合成装置において、前記指定手段は、前記合成すべき音声について音符長及び/又はテンポを指定し、前記読み出し手段は、前記音声素片データを読み出す際に前記音声素片データの一部を省略するか又は前記音声素片データの一部もしくは全部を繰返すかして前記音符長及び/又はテンポに対応する時間のあいだ前記音声素片データの読み出しを継続するようにしてもよい。このようにすると、音符長及び/又はテンポに最適の発音継続時間を得ることができる。

30 【0028】この発明に係る第3の歌唱合成装置は、順次に合成すべき音声のうちの各音声毎に音声素片及びピッチを指定する指定手段と、音声素片データベースから前記指定手段での指定に係る各音声素片に対応する音声波形を読み出す読み出し手段と、各音声素片に対応する音声波形を周波数分析して周波数スペクトルを検出する検出手段と、各音声素片に対応する周波数スペクトル上でスペクトル強度の局所的ピークを複数検知する検知手段と、各音声素片について各局所的ピーク毎に該局所的ピークとその前後のスペクトルとを含むスペクトル分布領域を該音声素片に対応する周波数スペクトル上で指定し、各音声素片について各スペクトル分布領域毎に振幅スペクトルデータを生成する第1の生成手段と、各音声素片について各スペクトル分布領域毎に位相スペクトル分布を周波数軸に関して表わす位相スペクトルデータを生成する第2の生成手段と、各音声素片について各スペクトル分布領域毎に前記振幅スペクトルデータが表わす振幅スペクトル分布を該音声素片に対応するピッチに応じて周波数軸上で移動するように前記振幅スペクトルデータを修正する第1の修正手段と、各音声素片について各スペクトル分布領域毎に前記位相スペクトルデータが表わす位相スペクトル分布を前記振幅スペクトルデータの修正に対応して修正する第2の修正手段と、前記順次に合成すべき音

声にそれぞれ対応する順次の音声素片が発音順につながるように前記修正に係る振幅スペクトルデータを接続する第1の接続手段であって、前記順次の音声素片のつながり部においてスペクトル強度を一致又は近似させるべく調整するものと、前記順次に合成すべき音声にそれぞれ対応する順次の音声素片が発音順につながるように前記修正に係る位相スペクトルデータを接続する第2の接続手段であって、前記順次の音声素片のつながり部において位相を一致又は近似させるべく調整するものと、前記接続に係る振幅スペクトルデータ及び前記接続に係る位相スペクトルデータを時間領域の合成音声信号に変換する変換手段とを備えたものである。

【0029】また、この発明に係る第4の歌唱合成装置は、順次に合成すべき音声のうちの各音声毎に音声素片及びピッチを指定する指定手段と、音声素片データベースから前記指定手段での指定に係る各音声素片に対応する振幅スペクトルデータ及び位相スペクトルデータを読み出す読み出し手段であって、前記振幅スペクトルデータとしては、対応する音声素片の音声波形を周波数分析して得た周波数スペクトルにおいてスペクトル強度の複数の局所的ピークのうちの各局所的ピーク毎に該局所的ピークとその前後のスペクトルとを含むスペクトル分布領域における振幅スペクトル分布を周波数軸に関して表わすデータを読み出し、前記位相スペクトルデータとしては、各スペクトル分布領域毎に位相スペクトル分布を周波数軸に関して表わすデータを読み出すものと、各音声素片について各スペクトル分布領域毎に前記振幅スペクトルデータが表わす振幅スペクトル分布を該音声素片に対応するピッチに応じて周波数軸上で移動するように前記振幅スペクトルデータを修正する第1の修正手段と、各音声素片について各スペクトル分布領域毎に前記位相スペクトルデータが表わす位相スペクトル分布を前記振幅スペクトルデータの修正に対応して修正する第2の修正手段と、前記順次に合成すべき音声にそれぞれ対応する順次の音声素片が発音順につながるように前記修正に係る振幅スペクトルデータを接続する第1の接続手段であって、前記順次の音声素片のつながり部においてスペクトル強度を一致又は近似させるべく調整するものと、前記順次に合成すべき音声にそれぞれ対応する順次の音声素片が発音順につながるように前記修正に係る位相スペクトルデータを接続する第2の接続手段であって、前記順次の音声素片のつながり部において位相を一致又は近似させるべく調整するものと、前記接続に係る振幅スペクトルデータ及び前記接続に係る位相スペクトルデータを時間領域の合成音声信号に変換する変換手段とを備えたものである。

【0030】第3又は第4の歌唱合成装置は、音声素片データベースを用いて前述の第1又は第2の歌唱合成方法を実施するものであり、自然な歌唱合成音を得ることができる。その上、順次の音声素片が発音順につながる

ように修正に係る振幅スペクトルデータ同士、修正に係る位相スペクトルデータ同士をそれぞれ接続する際に順次の音声素片のつながり部においてスペクトル強度、位相をそれぞれ一致又は近似させるべく調整するようにしたので、合成音の発生時にノイズが発生するのを防止することができる。

【0031】

【発明の実施の形態】図1は、この発明の一実施形態に係る歌唱合成装置の回路構成を示すものである。この歌唱合成装置は、小型コンピュータ10によって動作が制御される構成になっている。

【0032】バス11には、CPU(中央処理装置)12、ROM(リード・オンリ・メモリ)14、RAM(ランダム・アクセス・メモリ)16、歌唱入力部17、歌詞・メロディ入力部18、制御パラメータ入力部20、外部記憶装置22、表示部24、タイマ26、D/A(デジタル/アナログ)変換部28、MIDI(Musical Instrument Digital Interface)インターフェース30、通信インターフェース32等が接続されている。

【0033】CPU12は、ROM14にストアされたプログラムに従って歌唱合成等に関する各種処理を実行するもので、歌唱合成に関する処理については図2~7等を参照して後述する。

【0034】RAM16は、CPU12の各種処理に際してワーキングエリアとして使用される種々の記憶部を含むものである。この発明の実施に係る記憶部としては、例えば入力部17,18,20にそれぞれ対応する入力データ記憶領域等が存在するが、詳細については後述する。

【0035】歌唱入力部17は、歌唱音声信号を入力するためのマイクロホン、音声入力端子等を有するもので、入力した歌唱音声信号をデジタル波形データに変換するA/D(アナログ/デジタル)変換器を備えている。入力に係るデジタル波形データは、RAM16内の所定領域に記憶される。

【0036】歌詞・メロディ入力部18は、文字、数字等を入力可能なキーボード、楽譜読取り可能な読取器等を備えたもので、所望の歌唱曲について歌詞を構成する音素列を表わす歌詞データとメロディを構成する音符列(休符も含む)を表わすメロディデータを入力可能である。入力に係る歌詞データ及びメロディデータは、RAM16内の所定の領域に記憶される。

【0037】制御パラメータ入力部20は、スイッチ、ボリューム等のパラメータ設定器を備えたもので、歌唱合成音について音楽的表情を制御するための制御パラメータを設定可能である。制御パラメータとしては、音色、ピッチ区分(高、中、低等)、ピッチのゆらぎ(ピッチバンド、ビブラート等)、ダイナミクス区分(音量レベルの大、中、小等)、テンポ区分(テンポの速い、

中位、遅い等)などを設定可能である。設定に係る制御パラメータを表わす制御パラメータデータは、RAM 16内の所定領域に記憶される。

【0038】外部記憶装置22は、HD(ハードディスク)、FD(フレキシブルディスク)、CD(コンパクトディスク)、DVD(デジタル多目的ディスク)、MO(光磁気ディスク)等のうち1又は複数種類の記録媒体を着脱可能なものである。外部記憶装置22に所望の記録媒体を装着した状態では、記録媒体からRAM 16へデータを転送可能である。また、装着した記録媒体がHDやFDのように書き込み可能なものであれば、RAM 16のデータを記録媒体に転送可能である。

【0039】プログラム記録手段としては、ROM 14の代わりに外部記憶装置22の記録媒体を用いることができる。この場合、記録媒体に記録したプログラムは、外部記憶装置22からRAM 16へ転送する。そして、RAM 16に記憶したプログラムにしたがってCPU 12を動作させる。このようにすると、プログラムの追加やバージョンアップ等を容易に行なうことができる。

【0040】表示部24は、液晶表示器等の表示器を含むもので、前述した歌詞データ及びメロディデータ、後述する周波数分析結果等の種々の情報を表示可能である。

【0041】タイマ26は、テンポデータTMの指示するテンポに対応した周期でテンポクロック信号TCLを発生するもので、テンポクロック信号TCLは、CPU 12に供給される。CPU 12は、テンポクロック信号TCLに基づいてD/A変換部28への信号出力処理を行なう。テンポデータTMの指示するテンポは、入力部20内のテンポ設定器により可変設定することができる。

【0042】D/A変換部28は、合成されたデジタル音声信号をアナログ音声信号に変換するものである。D/A変換部28から送出されるアナログ音声信号は、アンプ、スピーカ等を含むサウンドシステム34により音響に変換される。

【0043】MIDIインターフェース30は、この歌唱合成装置とは別体のMIDI機器36との間でMIDI通信を行なうために設けられたもので、この発明では、MIDI機器36から歌唱合成用のデータを受信するために用いられる。歌唱合成用のデータとしては、所望の歌唱曲に関する歌詞データ及びメロディデータ、音楽的表情を制御するための制御パラメータデータ等を受信可能である。これらの歌唱合成用データは、いわゆるMIDIフォーマットに従って作成されるものであり、入力部18から入力される歌詞データ及びメロディデータや入力部20から入力される制御パラメータデータについてもMIDIフォーマットを採用するのが好ましい。

【0044】MIDIインターフェース30を介して受

信される歌詞データ、メロディデータ及び制御パラメータデータについては、他のデータより時間的に先送り可能とするため、MIDIのシステムエクスクルーシブデータ(メーカーで独自に定義可能なデータ)とするのが好ましい。また、入力部20から入力される制御パラメータデータ又はMIDIインターフェース30を介して受信される制御パラメータデータのうちの1種類のデータとしては、後述するデータベースに歌手(音色)毎に音声素片データを記憶した場合に歌手(音色)指定データを用いるようにしてもよい。この場合、歌手(音色)指定データとしては、MIDIのプログラムチェンジデータを使用することができる。

【0045】通信インターフェース32は、通信ネットワーク(例えばLAN(ローカル・エリア・ネットワーク)、インターネット、電話回線等)37を介して他のコンピュータ38と情報通信を行なうために設けられたものである。この発明の実施に必要なプログラムや各種データ(例えば歌詞データ、メロディデータ、音声素片データ等)は、コンピュータ38から通信ネットワーク37及び通信インターフェース32を介してRAM 16または外部記憶装置22へダウンロード要求に応じて取込むようにしてもよい。

【0046】次に、図2を参照して歌唱分析処理の一例を説明する。ステップ40では、入力部17からマイクロホン又は音声入力端子を介して歌唱音声信号を入力してA/D変換し、入力信号の音声波形を表わすデジタル波形データをRAM 16に記憶させる。図8(A)には、入力音声波形の一例を示す。なお、図8(A)及びその他の図において、「t」は時間を表わす。

【0047】ステップ42では、記憶に係るデジタル波形データについて音声素片(音素又は音素連鎖)に対応する区間毎に区間波形を切出す(デジタル波形データを分割する)。音声素片としては、母音の音素、母音と子音又は子音と母音の音素連鎖、子音と子音の音素連鎖、母音と母音の音素連鎖、無音と子音又は母音の音素連鎖、母音又は子音と無音の音素連鎖等があり、母音の音素としては、母音を伸ばして歌唱した伸ばし音の音素もある。一例として、「サイタ」の歌唱については、音声素片「#s」、「s_a」、「a」、「a_i」、「i」、「i_t」、「t_a」、「a」、「a#」にそれぞれ対応する区間波形を切出す。

【0048】ステップ44では、区間波形毎に1又は複数の時間フレームを定め、各フレーム毎にFFT等により周波数分析を実行して周波数スペクトル(振幅スペクトルと位相スペクトル)を検出する。そして、周波数スペクトルを表わすデータをRAM 16の所定領域に記憶させる。フレーム長は、一定長であってもよく、あるいは可変長であってもよい。フレーム長を可変長とするには、あるフレームを固定長として周波数分析した後、周波数分析の結果からピッチを検出し、検出ピッチに応じ

たフレーム長を設定して再び該フレームの周波数分析を行なう方法、あるいはあるフレームを固定長として周波数分析した後、周波数分析の結果からピッチを検出し、検出ピッチに応じて次のフレームの長さを設定し、該次のフレームの周波数分析を行なう方法等を採用することができる。フレーム数は、母音のみからなる単一の音素については、1又は複数フレームとするが、音素連鎖については、複数フレームとする。図8(B)には、図8(A)の音声波形をFFTにより周波数分析して得た周波数スペクトルを示す。なお、図8(B)及びその他の図において、「f」は周波数を表わす。

【0049】次に、ステップ46では、音声素片毎に振幅スペクトルに基づいてピッチを検出し、検出ピッチを表わすピッチデータを生成し、RAM16の所定領域に記憶させる。ピッチ検出は、フレーム毎に求めたピッチを全フレームについて平均する方法等により行なうことができる。

【0050】ステップ48では、フレーム毎に振幅スペクトル上でスペクトル強度(振幅)の局所的ピークを複数検知する。局所的ピークを検知するには、近隣の複数(例えば4つ)のピークについて振幅値が最大のピークを検知する方法等を用いることができる。図8(B)には、検知した複数の局所的ピーク P_1 、 P_2 、 P_3 ...が示されている。

【0051】ステップ50では、フレーム毎に振幅スペクトル上で各局所的ピークに対応するスペクトル分布領域を指定し、該領域内の振幅スペクトル分布を周波数軸に関して表わす振幅スペクトルデータを生成し、RAM16の所定領域に記憶させる。スペクトル分布領域を指定する方法としては、隣り合う2つの局所的ピーク間で周波数軸を半分に切り、各半分を近い方の局所的ピークを含むスペクトル分布領域に割当て方法、あるいは隣り合う2つの局所的ピーク間で振幅値が最低の谷を見出し、最低の振幅値に対応する周波数を隣り合うスペクトル分布領域間の境界とする方法等を採用することができる。図8(B)には、前者の方法により局所的ピーク P_1 、 P_2 、 P_3 ...をそれぞれ含むスペクトル分布領域 R_1 、 R_2 、 R_3 ...を指定した例を示す。

【0052】ステップ52では、フレーム毎に位相スペクトルに基づいて各スペクトル分布領域内の位相スペクトル分布を周波数軸に関して表わす位相スペクトルデータを生成し、RAM16内の所定領域に記憶させる。図10(A)には、あるフレームのあるスペクトル分布領域における振幅スペクトル分布及び位相スペクトル分布がそれぞれ曲線 AM_1 及び PH_1 により示されている。

【0053】ステップ54では、音声素片毎にピッチデータ、振幅スペクトルデータ及び位相スペクトルデータを音声素片データベースに記憶させる。音声素片データベースとしては、RAM16または外部記憶装置22を使用することができる。

【0054】図3は、音声素片データベースDBSにおける記憶状況の一例を示すものである。データベースDBSには、「a」、「i」...等の単一音素に対応する音声素片データと、「a_i」、「s_a」...等の音素連鎖に対応する音声素片データとが記憶される。ステップ54では、音声素片データとして、ピッチデータ、振幅スペクトルデータ及び位相スペクトルデータが記憶される。

【0055】音声素片データの記憶に際しては、各音声素片毎に歌手(音色)、ピッチ区分、ダイナミクス区分、テンポ区分等を異にする音声素片データを記憶すると、自然な(又は高品質)の歌唱音を合成可能になる。例えば、[a]の音声素片について、歌手Aにピッチ区分を低、中、高として、ダイナミクス区分を小、中、大として、テンポ区分を遅い、中位、速いとして歌ってもらい、ピッチ区分「低」で且つダイナミクス区分「小」であっても、テンポ区分「遅い」、「中位」、「速い」にそれぞれ対応する音声素片データ M_1 、 M_2 、 M_3 を記憶し、同様にしてピッチ区分「中」、「高」やダイナミクス区分「中」、「大」についても音声素片データを記憶する。ステップ46で生成したピッチデータは、音声素片データが「低」、「中」、「高」のいずれのピッチ区分に属するか判定する際に利用される。

【0056】また、歌手Aとは音色を異にする歌手Bについても、歌手Aについて上記したと同様にピッチ区分、ダイナミクス区分、テンポ区分等を異にする多数の[a]の音声素片データをデータベースDBSに記憶させる。[a]以外の他の音声素片についても、歌手A、Bについて上記したと同様に多数の音声素片データをデータベースDBSに記憶させる。

【0057】上記した例では、入力部17から入力した歌唱音声信号に基づいて音声素片データを作成したが、歌唱音声信号は、インターフェース30又は32を介して入力し、この入力音声信号に基づいて音声素片データを作成するようにしてもよい。また、データベースDBSとしては、RAM16や外部記憶装置22に限らず、ROM14、MIDI機器36内の記憶装置、コンピュータ38内の記憶装置等を用いてもよい。

【0058】図4は、歌唱合成処理の一例を示すものである。ステップ60では、所望の歌唱曲に関して歌詞データ及びメロディデータを入力部18から入力し、RAM16に記憶させる。歌詞データ及びメロディデータは、インターフェース30又は32を介して入力することもできる。

【0059】ステップ62では、入力に係る歌詞データが表わす音素列を個々の音声素片に変換する。そして、ステップ64では、音声素片毎に対応する音声素片データ(ピッチデータ、振幅スペクトルデータ及び位相スペクトルデータ)をデータベースDBSから読出す。ステップ64では、入力部20から制御パラメータとして音

色、ピッチ区分、ダイナミクス区分、テンポ区分等のデータを入力し、該データの指示する制御パラメータに対応した音声素片データを読み出してもよい。

【0060】ところで、音声素片の発音継続時間は、音声素片データのフレーム数に対応する。すなわち、記憶に係る音声素片データをそのまま用いて音声合成を行なうと、該音声素片データのフレーム数に対応した発音継続時間が得られる。しかし、入力される音符の音価（入力音符長）や設定テンポ等によっては記憶に係る音声素片データをそのまま用いたのでは発音継続時間が不適切になることがあり、発音継続時間を変更することが必要となる。このような必要に応えるためには、入力音符長や設定テンポ等に応じて音声素片データの読み出しフレーム数を制御すればよい。

【0061】例えば、音声素片の発音継続時間を短縮するには、音声素片データを読み出す際に、一部のフレームを飛ばして読み出す。また、音声素片の発音継続時間を伸張するには、音声素片データを反復的に読み出す。なお、「a」のような単一音素の伸ばし音を合成する際には、発音継続時間を変更することが多い。伸ばし音の合成については、図14～16を参照して後述する。

【0062】ステップ66では、音声素片毎に対応する入力音符のピッチに応じて各フレームの振幅スペクトルデータを修正する。すなわち、各スペクトル分布領域毎に振幅スペクトルデータが表わす振幅スペクトル分布を入力音符ピッチに相当するピッチになる様に周波数軸上で移動する。

【0063】図10(A)、(B)は、局所的ピークの周波数が f_i であり且つ下限周波数及び上限周波数がそれぞれ f_L 及び f_U であるスペクトル分布領域についてピッチを上昇させるためスペクトル分布 AM_1 を AM_2 のように周波数軸上で高音側に移動した例を示す。この場合、スペクトル分布 AM_2 については、局所的ピークの周波数が $F_i = T \cdot f_i$ であり、 $T = F_i / f_i$ をピッチ変換比と称する。また、下限周波数 F_L 及び上限周波数 F_U は、それぞれ周波数差 $(f_i - f_L)$ 及び $(f_U - f_i)$ に対応して定める。

【0064】図9には、(A)に示すようなスペクトル分布領域(図8(B)と同じもの) $R_1, R_2, R_3 \dots$ について局所的ピーク $P_1, P_2, P_3 \dots$ をそれぞれ有するスペクトル分布を(B)に示す様に周波数軸上で高音側に移動した例を示す。図9(B)に示されるスペクトル分布領域 R_1 において、局所的ピーク P_1 の周波数、下限周波数 f_{11} 及び上限周波数 f_{12} は、図10に関して上記したと同様に定められる。このことは、他のスペクトル分布領域についても同様である。

【0065】上記した例では、ピッチを上昇させるためスペクトル分布を周波数軸上で高音側に移動したが、ピッチを低下させるためスペクトル分布を周波数軸上で低音側に移動することもできる。この場合、図11に示す

様に2つのスペクトル分布領域 R_a, R_b に部分的な重なりが生ずる。

【0066】図11の例では、局所的ピーク P_a と下限周波数 f_{a1} と上限周波数 f_{a2} とを有するスペクトル分布領域 R_a に対して、局所的ピーク P_b と下限周波数 f_{b1} ($f_{b1} < f_{a2}$)と上限周波数 f_{b2} ($f_{b2} > f_{a2}$)とを有するスペクトル分布領域 R_b が周波数 $f_{b1} \sim f_{a2}$ の領域で重なっている。このような事態を回避するため、一例として、 $f_{b1} \sim f_{a2}$ の周波数領域を中心周波数 f_c で2分割し、領域 R_a の上限周波数 f_{a2} を f_c より低い所定の周波数に変更すると共に、領域 R_b の下限周波数 f_{b1} を f_c より高い所定の周波数に変更する。この結果、領域 R_a では、 f_c より低い周波数領域でスペクトル分布 AM_a を利用可能となり、領域 R_b では、 f_c より高い周波数領域でスペクトル分布 AM_b を利用可能となる。

【0067】上記のように局所的ピークを含むスペクトル分布を周波数軸上で移動する際、周波数の設定を変更するだけではスペクトル包絡が伸び縮みすることになり、音色が入力音声波形のものとは異なる事態が生ずる。そこで、入力音声波形の音色を再現するためには、各フレーム毎に一連のスペクトル分布領域の局所的ピークを結ぶ線に相当するスペクトル包絡に沿うように1又は複数のスペクトル分布領域の局所的ピークについてスペクトル強度を調整する必要がある。

【0068】図12は、スペクトル強度調整の一例を示すもので、(A)は、ピッチ変換前の局所的ピーク $P_{11} \sim P_{18}$ に対応するスペクトル包絡 EV を示す。入力音符ピッチに応じてピッチを上昇させるため局所的ピーク $P_{11} \sim P_{18}$ をそれぞれ(B)の $P_{21} \sim P_{28}$ に示すように周波数軸上で移動する際にスペクトル包絡 EV に沿わない局所的ピークについてはスペクトル包絡 EV に沿うようにスペクトル強度を増大又は減少させる。この結果、入力音声波形と同様の音色が得られる。

【0069】図12(A)において、 R_f は、スペクトル包絡が欠如した周波数領域であり、ピッチを上昇させる場合には、図12(B)に示す様に周波数領域 R_f 内に P_{27}, P_{28} 等の局所的ピークを移動する必要がある。このような事態に対処するには、図12(B)に示す様に周波数領域 R_f について補間法によりスペクトル包絡 EV を求め、求めたスペクトル包絡 EV に従って局所的ピークのスペクトル強度の調整を行えばよい。

【0070】上記した例では、入力音声波形の音色を再現するようにしたが、入力音声波形とは異なる音色を合成音声に付与するようにしてもよい。このためには、図12に示したようなスペクトル包絡 EV を変形したスペクトル包絡を用いるか又は全く新しいスペクトル包絡を用いるかして上記したと同様に局所的ピークのスペクトル強度を調整すればよい。

【0071】スペクトル包絡を用いた処理を簡素化するには、スペクトル包絡を曲線又は直線等で表現するのが好ましい。図13には、2種類のスペクトル包絡線曲線 $E V_1$ 、 $E V_2$ を示す。曲線 $E V_1$ は、局所的ピーク間を直線で結ぶことによりスペクトル包絡を折れ線で簡単に表現したものである。また、曲線 $E V_2$ は、スペクトル包絡を3次のスプライン関数で表わしたものである。曲線 $E V_2$ を用いると、補間をより正確に行なうことができる。

【0072】次に、図4のステップ68では、音声素片毎に各フレームの振幅スペクトルデータの修正に対応して位相スペクトルデータを修正する。すなわち、図10(A)に示すようにあるフレームにおける i 番目の局所的ピークを含むスペクトル分布領域では、位相スペクトル分布 $P H_1$ が振幅スペクトル分布 $A M_1$ に対応したものであり、ステップ66で振幅スペクトル分布 $A M_1$ を $A M_2$ のように移動したときは、振幅スペクトル分布 $A M_2$ に対応して位相スペクトル分布 $P H_1$ を調整する必要がある。これは、移動先の局所的ピークの周波数で正弦波になるようにするためである。

【0073】 i 番目の局所的ピークを含むスペクトル分布領域に関する位相の補正量 θ_i は、フレーム間の時間間隔を t 、局所的ピークの周波数を f_i 、ピッチ変換比を T とすると、次の数1の式で与えられる。

【0074】

$$\text{【数1】 } \theta_i = 2 \pi f_i (T - 1) t$$

数1の式で求められた補正量 θ_i は、図10(B)に示す様に周波数 $F_L \sim F_U$ の領域内の各位相スペクトルの位相に加算され、局所的ピークの周波数 F_i では位相が $\theta_i + \theta_i$ となる。

【0075】上記のような位相の補正は、各スペクトル分布領域毎に行なわれる。例えば、あるフレームにおいて、局所的ピークの周波数が完全に調和関係にある(倍音の周波数が基音の周波数の完全な整数倍になっている)場合には、入力音声の基音周波数(すなわち音声素片データ内のピッチデータが示すピッチ)を f_0 とし、スペクトル分布領域の番号を $k = 1, 2, 3 \dots$ とすると、位相補正量 θ_i は、次の数2の式で与えられる。

【0076】

$$\text{【数2】 } \theta_i = 2 \pi f_0 k (T - 1) t$$

ステップ70では、音声素片毎に設定テンポ等に応じて発音開始時刻を決定する。発音開始時刻は、設定テンポや入力音符長等に依存し、テンポクロック信号 $T C L$ のクロック数で表わすことができる。一例として、「サイタ」の歌唱の場合、「s a」の音声素片の発音開始時刻は、入力音符長及び設定テンポで決まるノートオン時刻に「s」ではなく「a」の発音を開始されるように設定する。ステップ60でリアルタイムで歌詞データ及びメロディを入力してリアルタイムで歌唱合成を行なうときは、子音及び母音の音素連鎖について上記のような発

音開始時刻の設定が可能になるようにノートオン時刻より前に歌詞データ及びメロディデータを入力する。

【0077】ステップ72では、音声素片間でスペクトル強度のレベルを調整する。このレベル調整処理は、振幅スペクトルデータ及び位相スペクトルデータのいずれについても行なわれるもので、次のステップ74でのデータ接続に伴って合成音発生時にノイズが発生するのを回避するために行なわれる。レベル調整処理としては、スムージング処理、レベル整合処理等があるが、これらの処理については図17~20を参照して後述する。

【0078】ステップ74では、音声素片の発音順に振幅スペクトルデータ同士、位相スペクトルデータ同士をそれぞれ接続する。そして、ステップ76では、音声素片毎に振幅スペクトルデータ及び位相スペクトルデータを時間領域の合成音声信号(デジタル波形データ)に変換する。

【0079】図5は、ステップ76の変換処理の一例を示すもので、ステップ76aでは、周波数領域のフレームデータ(振幅スペクトルデータ及び位相スペクトルデータ)に逆FFT処理を施して時間領域の合成音声信号を得る。そして、ステップ76bでは、時間領域の合成音声信号に窓掛け処理を施す。この処理は、時間領域の合成音声信号に時間窓関数を乗算するものである。ステップ76cでは、時間領域の合成音声信号にオーバーラップ処理を施す。この処理は、順次の音声素片について波形をオーバーラップさせながら時間領域の合成音声信号を接続するものである。

【0080】ステップ78では、ステップ70で決定した発音開始時刻を参照して音声素片毎に合成音声信号をD/A変換部28に出力する。この結果、サウンドシステム34から合成に係る歌唱音声が発生される。

【0081】図6は、歌唱分析処理の他の例を示すものである。ステップ80では、ステップ40に関して前述したと同様にして歌唱音声信号を入力し、入力信号の音声波形を表すデジタル波形データをRAM16に記憶させる。歌唱音声信号は、インターフェース30又は32を介して入力してもよい。

【0082】ステップ82では、ステップ42に関して前述したと同様にして記憶に係るデジタル波形データについて音声素片に対応する区間ごとに区間波形を切出す。

【0083】ステップ84では、音声素片毎に区間波形を表わす区間波形データ(音声素片データ)を音声素片データベースに記憶させる。音声素片データベースとしては、RAM16や外部記憶装置22を用いることができ、所望によりROM14、MIDI機器36内の記憶装置、コンピュータ38内の記憶装置等を用いてもよい。音声素片データの記憶に際しては、図3に関して前述したと同様に各音声素片毎に歌手(音色)、ピッチ区分、ダイナミクス区分、テンポ区分等を異にする区間波

形データ $m_1, m_2, m_3 \dots$ を音声素片データベースD B Sに記憶させることができる。

【0084】次に、図7を参照して歌唱合成処理の他の例を説明する。ステップ90では、ステップ60に関して前述したと同様にして所望の歌唱曲に関して歌詞データ及びメロディデータを入力する。

【0085】ステップ92では、ステップ62に関して前述したと同様にして歌詞データが表わす音素列を個々の音声素片に変換する。そして、ステップ94では、ステップ84で記憶処理したデータベースから音声素片毎に対応する区間波形データ(音声素片データ)を讀出す。この場合、入力部20から制御パラメータとして音色、ピッチ区分、ダイナミクス区分、テンポ区分等のデータを入力し、該データの指示する制御パラメータに対応した区間波形データを讀出すようにしてもよい。また、ステップ64に関して前述したと同様に入力音符長や設定テンポ等に応じて音声素片の発音継続時間を変更するようにしてもよい。このためには、音声波形を讀出す際に音声波形の一部を省略するか又は音声波形の一部あるいは全部を繰返すかして所望の発音継続時間だけ音声波形の讀出しを継続すればよい。

【0086】ステップ96では、讀出しに係る区間波形データ毎に区間波形について1又は複数の時間フレームを定め、各フレーム毎にFFT等により周波数分析を実行して周波数スペクトル(振幅スペクトルと位相スペクトル)を抽出する。そして、周波数スペクトルを表わすデータをRAM16の所定領域に記憶させる。

【0087】ステップ98では、図2のステップ46~52と同様の処理を実行して音声素片毎にピッチデータ、振幅スペクトルデータ及び位相スペクトルデータを生成する。そして、ステップ100では、図4のステップ66~78と同様の処理を実行して歌唱音声を作成し、発音させる。

【0088】図7の歌唱合成処理を図4の歌唱合成処理と対比すると、図4の歌唱合成処理では、データベースから音声素片毎にピッチデータ、振幅スペクトルデータ及び位相スペクトルデータを取得して歌唱合成を行なうのに対し、図7の歌唱合成処理では、データベースから音声素片毎に区間波形データを取得して歌唱合成を行なっている点で両者が異なるものの、歌唱合成の手順は、両者で実質的に同一である。図4又は図7の歌唱合成処理によれば、入力音声波形の周波数分析結果を調和成分と非調和成分とに分離しないので、非調和成分が分離して響くことがなく、自然な(又は高品質の)合成音を得られる。また、有声の摩擦音や破裂音についても自然な合成音を得られる。

【0089】図14は、例えば「a」のような単一音素の伸ばし音に関するピッチ変換処理及び音色調整処理(図4のステップ66に対応)を示すものである。この場合、伸ばし音の音声素片データSDとして、図3に示

したようなピッチデータ、振幅スペクトルデータ及び位相スペクトルデータのデータ組(又は区間波形データ)をデータベース内に用意する。また、伸ばし音毎に歌手(音色)、ピッチ区分、ダイナミクス区分、テンポ区分等を異にする音声素片データをデータベースに記憶しておき、入力部20で所望の歌手(音色)、ピッチ区分、ダイナミクス区分、テンポ区分等の制御パラメータを指定すると、指定に係る制御パラメータに対応する音声素片データを讀出すようにする。

10 【0090】ステップ110では、伸ばし音の音声素片データSDに由来する振幅スペクトルデータFSPにステップ66で述べたと同様のピッチ変換処理を施す。すなわち、振幅スペクトルデータFSPに関して各フレームの各スペクトル分布領域毎にスペクトル分布を入力音符ピッチデータPTの示す入力音符ピッチに相当するピッチになるように周波数軸上で移動する。

20 【0091】音声素片データSDの時間長より長い発音継続時間の伸ばし音が要求される場合には、音声素片データSDを最後まで讀出した後最初に戻って再び讀出し、必要に応じてこのような時間的に順方向の讀出しを繰返す方法を採用することができる。別の方法としては、音声素片データSDを最後まで讀出した後最初に向かって讀出し、必要に応じてこのような時間的に順方向の讀出しと時間的に逆方向の讀出しとを繰返す方法を採用してもよい。この方法では、時間的に逆方向に讀出す際の讀出開始点をランダムに設定するようにしてもよい。

30 【0092】ステップ110のピッチ変換処理では、図3に示したデータベースDBSにおいて、例えば「a」のような伸ばし音声素片データM1(又は m_1)、M2(又は m_2)、M3(又は m_3)...にそれぞれ対応して経時的なピッチ変化を表わすピッチゆらぎデータを記憶しておき、入力部20で音色、ピッチ区分、ダイナミクス区分、テンポ区分等の制御パラメータを指定するのに応答して指定に係る制御パラメータに対応するピッチゆらぎデータを讀出すようにしてもよい。この場合、ステップ112では、讀出しに係るピッチゆらぎデータVPを入力音符ピッチデータPTに加算し、加算結果としてのピッチ制御データに応じてステップ110でのピッチ変換を制御する。このようにすると、合成音にピッチのゆらぎ(例えばピッチベンド、ビブラート等)を付加することができ、自然な合成音を得られる。また、音色、ピッチ区分、ダイナミクス区分、テンポ区分等の制御パラメータに応じてピッチのゆらぎ態様を異ならせることができるので、自然感が一層向上する。なお、ピッチゆらぎデータは、音声素片に対応する1又は複数のピッチゆらぎデータを音色等の制御パラメータに応じて補間等により改変して使うようにしてもよい。

50 【0093】ステップ114では、ステップ110でピッチ変換処理を受けた振幅スペクトルデータFSP'に

音色調整処理を施す。この処理は、図12に関して前述したように各フレーム毎にスペクトル包絡に従ってスペクトル強度を調整して合成音の音色を設定するものである。

【0094】図15は、ステップ114の音色調整処理の一例を示すものである。この例では、図3に示したデータベースDBSにおいて、例えば「a」の伸ばし音の音声素片に対応して代表的な1つのスペクトル包絡を表わすスペクトル包絡データを記憶する。

【0095】ステップ116では、伸ばし音の音声素片に対応するスペクトル包絡データをデータベースDBSから読出す。そして、ステップ118では、読出しに係るスペクトル包絡データに基づいてスペクトル包絡設定処理を行なう。すなわち、伸ばし音のフレーム群FRにおける複数n個のフレームの振幅スペクトルデータFR₁ ~ FR_nのうちの各フレームの振幅スペクトルデータ毎に、読出しに係るスペクトル包絡データの示すスペクトル包絡に沿うようにスペクトル強度を調整することによりスペクトル包絡を設定する。この結果、伸ばし音に適切な音色を付与することができる。

【0096】ステップ118のスペクトル包絡設定処理では、図3に示したデータベースDBSにおいて、例えば「a」のような伸ばし音声素片データM1（又はm1）、M2（又はm2）、M3（又はm3）...にそれぞれ対応して経時的なスペクトル包絡変化を表わすスペクトル包絡ゆらぎデータを記憶しておき、入力部20で音色、ピッチ区分、ダイナミクス区分、テンポ区分等の制御パラメータを指定するのに対応して指定に係る制御パラメータに対応するスペクトル包絡ゆらぎデータを読出すようにしてもよい。この場合、ステップ118では、各フレーム毎にステップ116での読出しに係るスペクトル包絡データに読出しに係るスペクトル包絡ゆらぎデータVEを加算し、加算結果としてのスペクトル包絡制御データに応じてステップ118でのスペクトル包絡設定を制御する。このようにすると、合成音に音色のゆらぎ（例えばトーンバンド等）を付加することができ、自然な合成音が得られる。また、音色、ピッチ区分、ダイナミクス区分、テンポ区分等の制御パラメータに応じてピッチのゆらぎ態様を異ならせることができるので、自然感が一層向上する。なお、ピッチゆらぎデータは、音声素片に対応する1又は複数のピッチゆらぎデータを音色等の制御パラメータに応じて補間等により改変して使うようにしてもよい。

【0097】図16は、ステップ114の音色調整処理の他の例を示すものである。歌唱合成では、前述した「サイタ」の歌唱例の様に音素連鎖（例えば「s_a」）- 単一音素（例えば「a」）- 音素連鎖（例えば「a_i」）の歌唱合成が典型的な例であり、このような歌唱合成例に適したのが図16の例である。図16において、前音の最終フレームの振幅スペクトルデータP

FRにおける前音とは、例えば「s_a」の音素連鎖に対応し、伸ばし音のn個のフレームの振幅スペクトルデータFR₁ ~ FR_nにおける伸ばし音とは、例えば「a」の単一音素に対応し、後音の先頭フレームの振幅スペクトルデータNFRにおける後音とは、例えば「a_i」の音素連鎖に対応する。

【0098】ステップ120では、前音の最終フレームの振幅スペクトルデータPFRからスペクトル包絡を抽出すると共に、後音の先頭フレームの振幅スペクトルデータNFRからスペクトル包絡を抽出する。そして、抽出に係る2つのスペクトル包絡を時間的に補間して伸ばし音用のスペクトル包絡を表わすスペクトル包絡データを作成する。

【0099】ステップ122では、n個のフレームの振幅スペクトルデータFR₁ ~ FR_nのうちの各フレームの振幅スペクトルデータ毎に、ステップ120での作成に係るスペクトル包絡データの示すスペクトル包絡に沿うようにスペクトル強度を調整することによりスペクトル包絡を設定する。この結果、音素連鎖間の伸ばし音に適切な音色を付与することができる。

【0100】ステップ122においても、ステップ118に関して前述したと同様にしてデータベースDBSから音色等の制御パラメータに応じてスペクトル包絡ゆらぎデータVEを読出すなどしてスペクトル包絡の設定を制御することができる。このようにすると、自然な合成音が得られる。

【0101】次に、図17~19を参照してスムージング処理（ステップ72に対応）の一例を説明する。この例では、データを扱いやすくして計算を簡単にするために、音声素片の各フレームのスペクトル包絡を図17に示すように直線（あるいは指数関数）で表現した傾き成分と指数関数などで表現した1又は複数の共鳴部分とに分解する。すなわち、共鳴部分の強度は、傾き成分を基準に計算し、傾き成分と共鳴成分を足し合わせてスペクトル包絡を表わす。また、傾き成分を0Hzまで延長した値を傾き成分のゲインと称する。

【0102】一例として、図18に示すような2つの音声素片「a_i」と「i_a」とを接続するものとする。これらの音声素片は、もともと別の録音から採取したものであるため、接続部のiの音色とレベルにミスマッチがあり、図18に示すように接続部分で波形の段差が発生し、ノイズとして聴こえる。2つの音声素片について接続部を中心として前後に何フレームにかけて、傾き成分のパラメータ同士、共鳴成分のパラメータ同士をそれぞれクロスフェードしてやれば、接続部分での段差が消え去り、ノイズの発生を防止することができる。

【0103】例えば、共鳴成分のパラメータをクロスフェードするためには、図19に示すように、接続部分で0.5となるような関数（クロスフェードパラメータ）を両音声素片の共鳴成分のパラメータに掛けて足し合わ

せてやればよい。図19に示す例では、「a_i」, 「i_a」の音声素片における第1の共鳴成分の(傾き成分を基準とした)強度の時間的変化を示す波形に対してそれぞれクロスフェードパラメータを掛けて加算することによりクロスフェードを行なった例を示している。

【0104】他の共鳴成分、傾き成分等のパラメータについても、上記したと同様にクロスフェードを行なうことができる。

【0105】図20は、レベル整合処理(ステップ72に対応)の一例を示すものである。この例では、上記と同様に「a_i」と「i_a」を接続して合成する場合について、レベル整合処理を説明する。

【0106】この場合、上記のようにクロスフェードする代わりに、音声素片の接続部分で前後の振幅がほぼ同じになる様にレベル整合を行なう。レベル整合は、音声素片の振幅に対し、一定あるいは時変の係数を掛けることにより行なうことができる。

【0107】この例では、2つの音声素片について傾き成分のゲインを合わせる処理について説明する。まず、

図20(a), (b)に示すように、「a_i」と「i_a」の各音声素片について、その最初のフレームと最終フレームの間の傾き成分のゲインを直線補間したパラメータ(図中の破線)を求め、各パラメータを基準に、実際の傾き成分のゲインとの差分を求める。

【0108】次に、[a], [i]の各音韻の代表的なサンプル(傾き成分及び共鳴成分の各パラメータ)を求める。これは、例えば、「a_i」の最初のフレームと最終フレームの振幅スペクトルデータを用いて求めてもよい。

【0109】[a], [i]の代表的なサンプルをもとに、まず、図20(c)に破線で示すように[a], [i]の間の傾き成分のゲインを直線補間したパラメータを求めると共に、[i]と[a]の間の傾き成分のゲインを直線補間したパラメータを求める。次いで、図20(a), (b)で求めた差分を直線補間に係るパラメータにそれぞれ足し込んでいけば、図20(c)に示すように、境界では必ず直線補間に係るパラメータが一致するため、傾き成分のゲインの不連続は発生しない。共鳴成分のパラメータなど他のパラメータについても、同様に不連続を防止することができる。

【0110】前述したステップ72では、振幅スペクトルデータのみならず位相スペクトルデータについても、上記のようなスムージング処理又はレベル整合処理を準用して位相の調整を行なう。この結果、ノイズ発生を回避することができ、高品質の歌唱合成が可能となる。なお、スムージング処理又はレベル整合処理において、接続部では、スペクトル強度を一致させたが近似させるだけでよいこともある。

【0111】

【発明の効果】以上のように、この発明によれば、音声

素片に対応する音声波形を周波数分析した結果に基づいて振幅スペクトルデータ及び位相スペクトルデータを生成し、指定のピッチに応じて振幅スペクトルデータ及び位相スペクトルデータを修正し、修正に係る振幅スペクトルデータ及び位相スペクトルデータに基づいて時間領域の合成音声信号を発生させるようにしたので、周波数分析結果を調和成分と非調和成分とに分離した従来例のように非調和成分が分離して響くといった事態は原理的に発生しなくなり、自然な歌唱音声又は高品質の歌唱音声合成可能となる効果が得られる。

【図面の簡単な説明】

【図1】 この発明の一実施形態に係る歌唱合成装置の回路構成を示すブロック図である。

【図2】 歌唱分析処理の一例を示すフローチャートである。

【図3】 音声素片データベース内の記憶状況を示す図である。

【図4】 歌唱合成処理の一例を示すフローチャートである。

20 【図5】 図4のステップ76の変換処理の一例を示すフローチャートである。

【図6】 歌唱分析処理の他の例を示すフローチャートである。

【図7】 歌唱合成処理の他の例を示すフローチャートである。

【図8】 (A)は、分析対象としての入力音声信号を示す波形図、(B)は、(A)の波形の周波数分析結果を示すスペクトル図である。

30 【図9】 (A)は、ピッチ変換前のスペクトル分布領域配置を示すスペクトル図、(B)は、ピッチ変換後のスペクトル分布領域配置を示すスペクトル図である。

【図10】 (A)は、ピッチ変換前の振幅スペクトル分布及び位相スペクトル分布を示すグラフ、(B)は、ピッチ変換後の振幅スペクトル分布及び位相スペクトル分布を示すグラフである。

【図11】 ピッチを低下させた場合のスペクトル分布領域の指定処理を説明するためのグラフである。

40 【図12】 (A)は、ピッチ変換前の局所的ピーク配置及びスペクトル包絡を示すグラフ、(B)は、ピッチ変換後の局所的ピーク配置及びスペクトル包絡を示すグラフである。

【図13】 スペクトル包絡曲線を例示するグラフである。

【図14】 伸ばし音に関するピッチ変換処理及び音色調整処理を示すブロック図である。

【図15】 伸ばし音に関する音色調整処理の一例を示すブロック図である。

【図16】 伸ばし音に関する音色調整処理の他の例を示すブロック図である。

50 【図17】 スペクトル包絡のモデル化を説明するため

のグラフである。

【図18】 音声素片の接続時に生ずるレベル及び音色のミスマッチを説明するためのグラフである。

【図19】 スムージング処理を説明するためのグラフである。

【図20】 レベル整合処理を説明するためのグラフである。

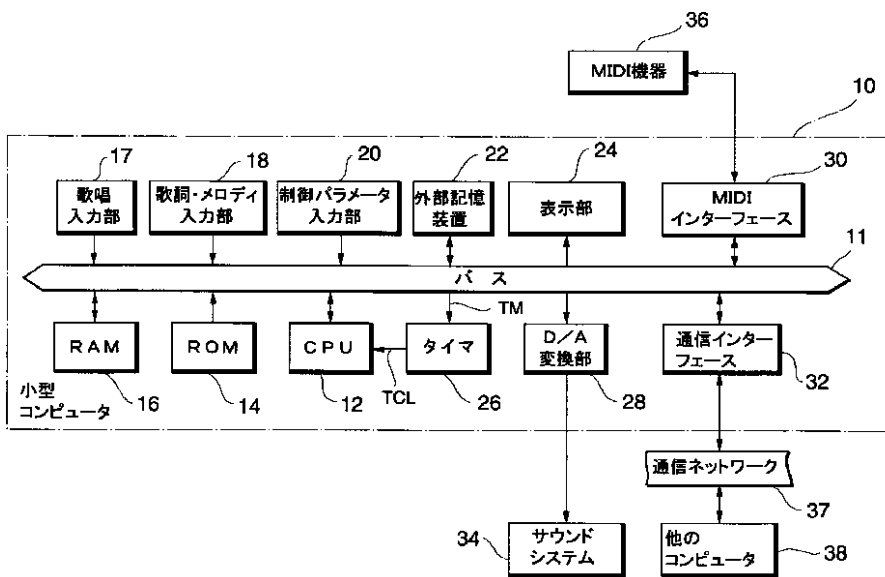
【図21】 歌唱合成処理の従来例を示すブロック図である。

【符号の説明】

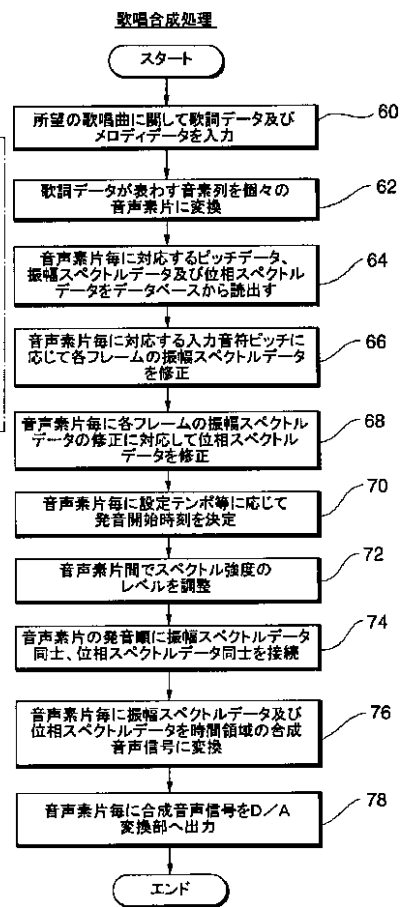
* 10：小型コンピュータ、11：バス、12：CPU、14：ROM、16：RAM、17：歌唱入力部、18：歌詞・メロディ入力部、20：制御パラメータ入力部、22：外部記憶装置、24：表示部、26：タイマ、28：D/A変換部、30：MIDIインターフェース、32：通信インターフェース、34：サウンドシステム、36：MIDI機器、37：通信ネットワーク、38：他のコンピュータ、DBS：音声素片データベース。

* 10

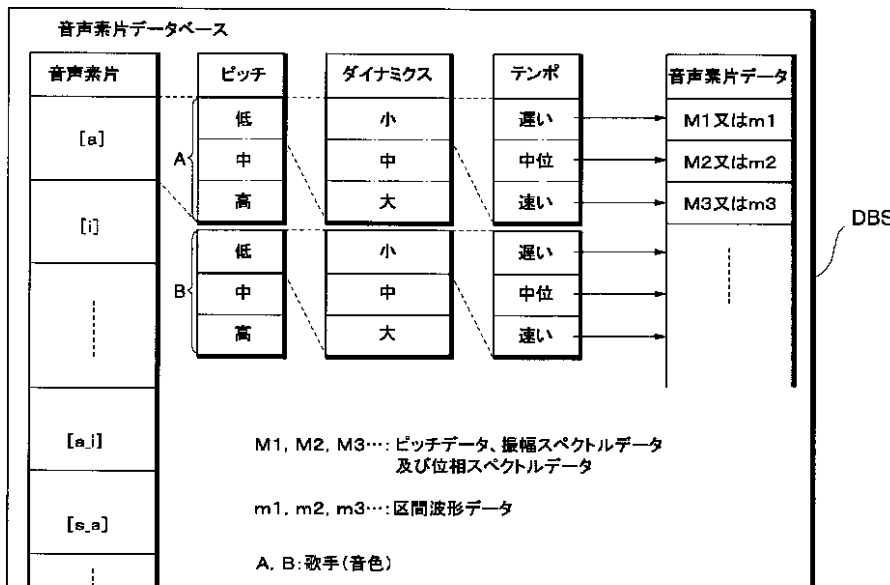
【図1】



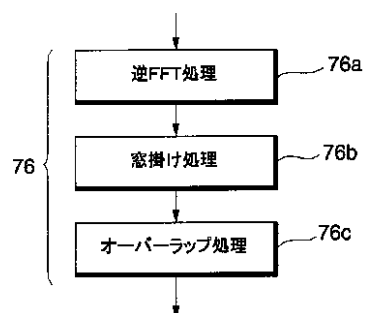
【図4】



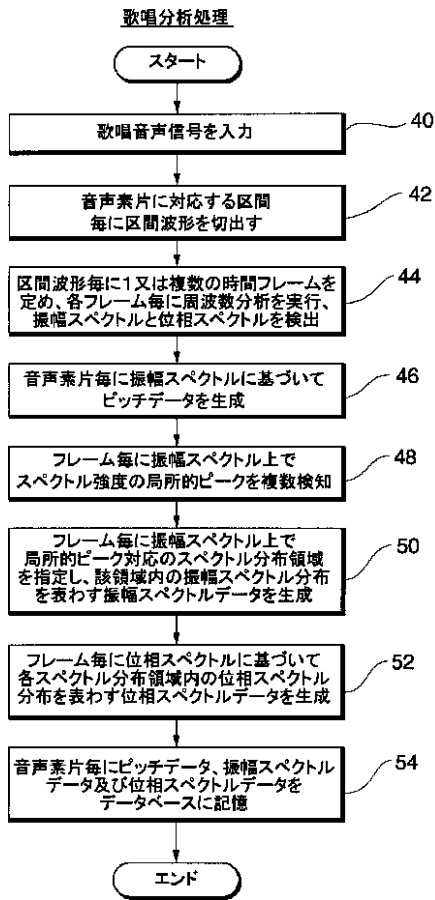
【図3】



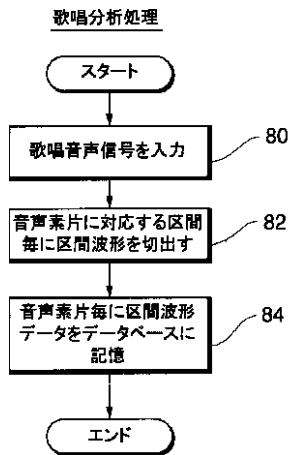
【図5】



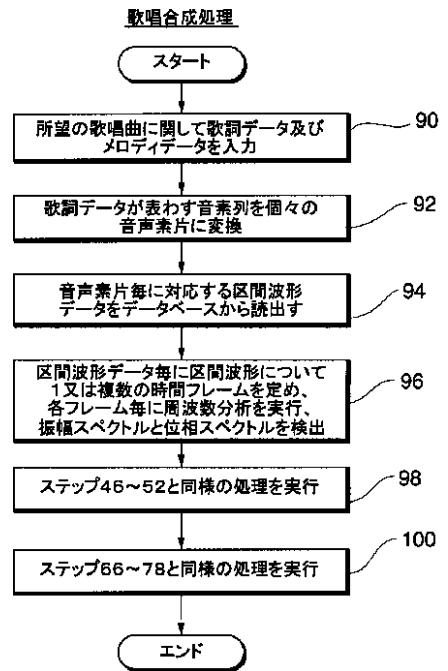
【図2】



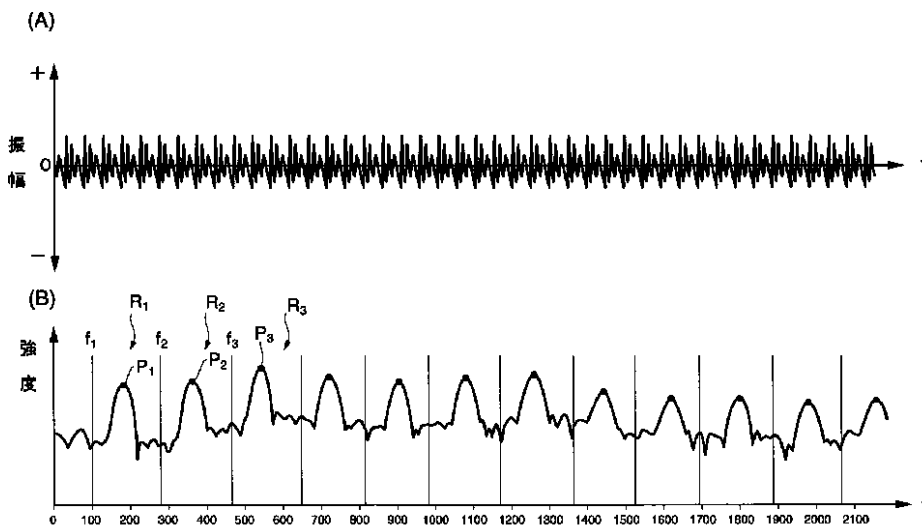
【図6】



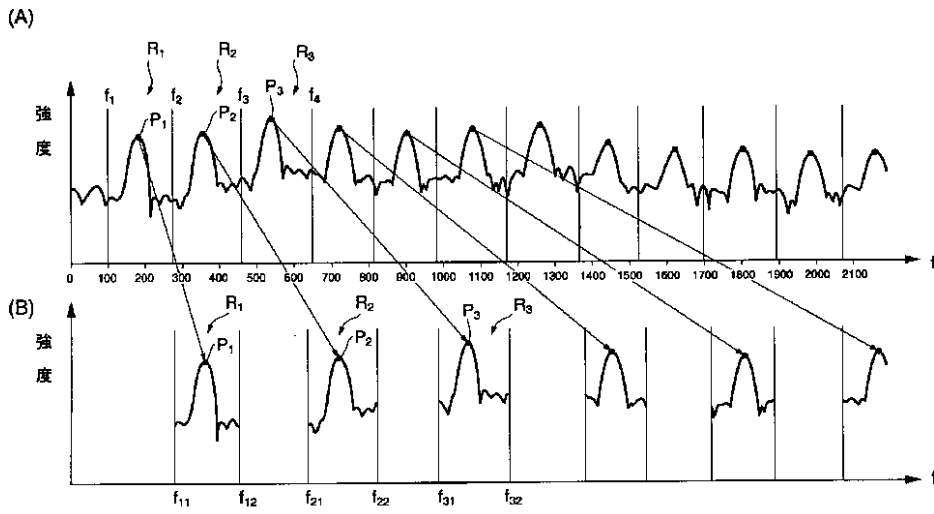
【図7】



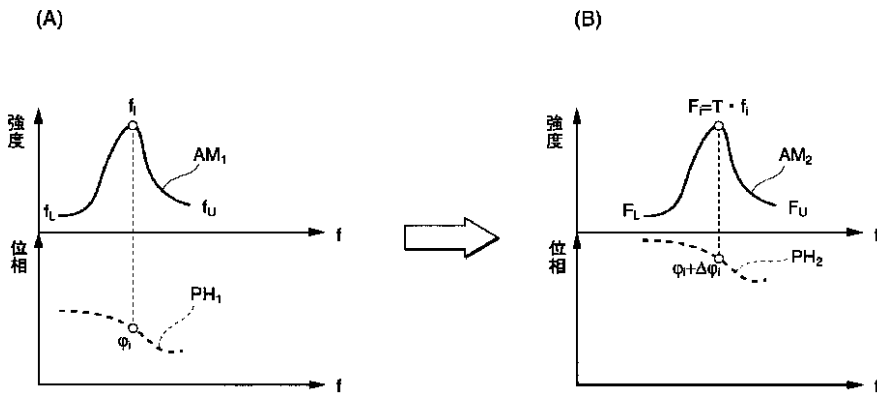
【図8】



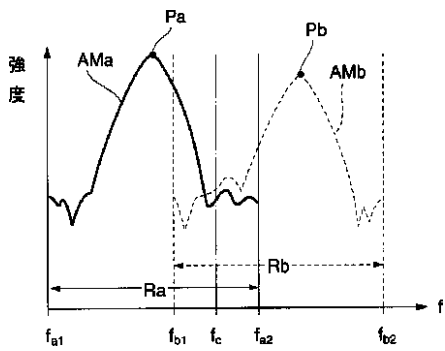
【図9】



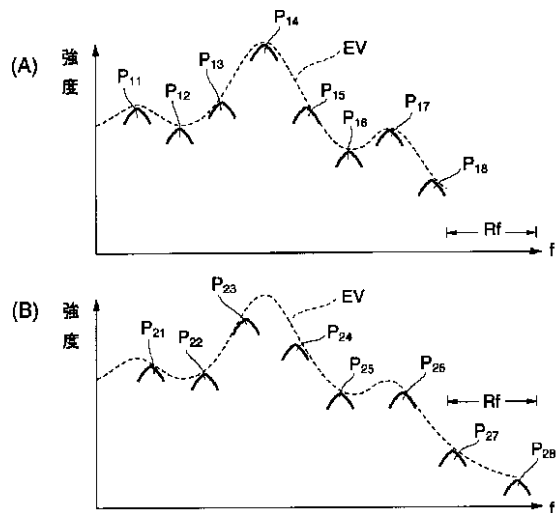
【図10】



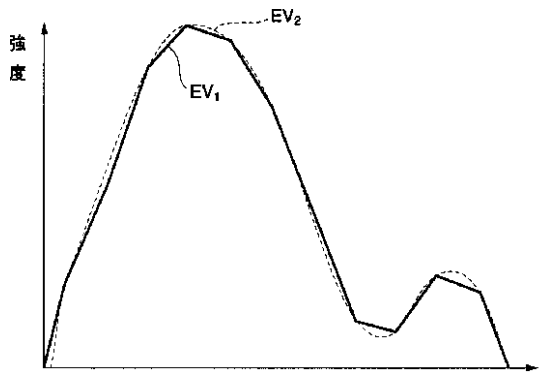
【図11】



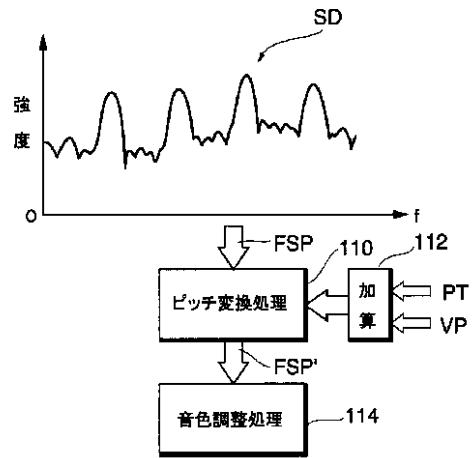
【図12】



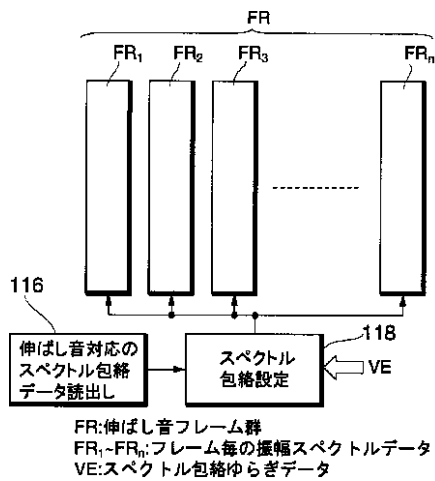
【図13】



【図14】



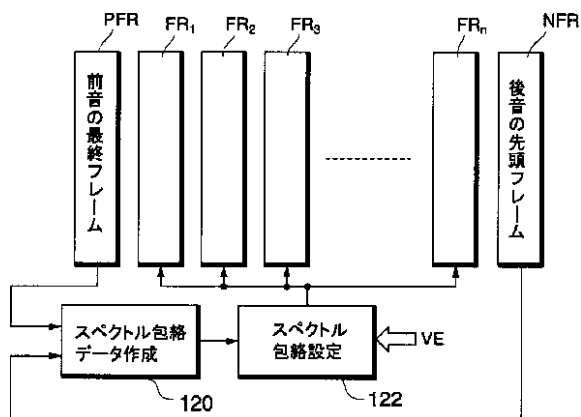
【図15】



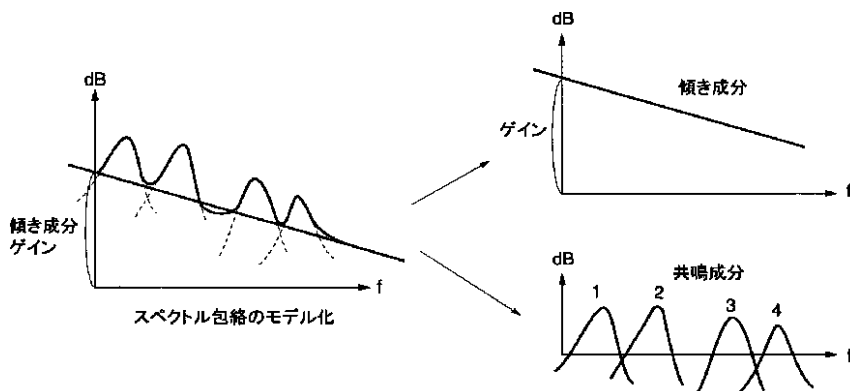
FR:伸ばし音フレーム群
 FR₁~FR_n:フレーム毎の振幅スペクトルデータ
 VE:スペクトル包絡ゆらぎデータ

SD:伸ばし音の音声素片データ
 FSP } 振幅スペクトルデータ
 FSP'
 PT:入力音符ピッチデータ
 VP:ピッチゆらぎデータ

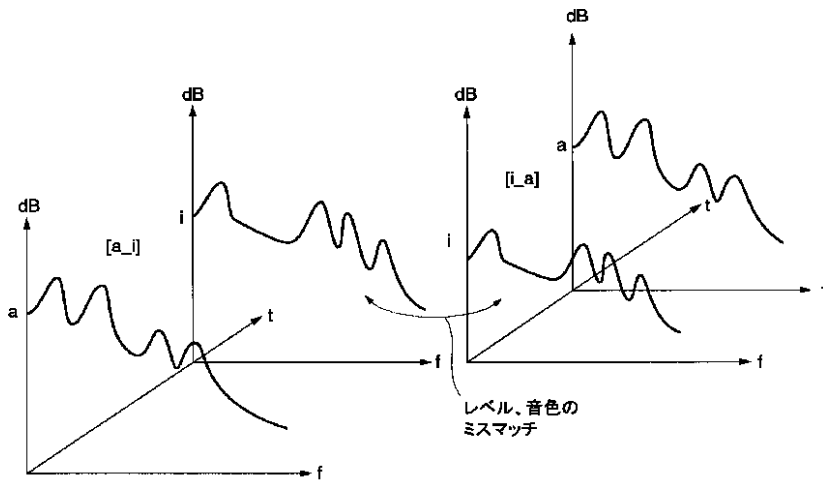
【図16】



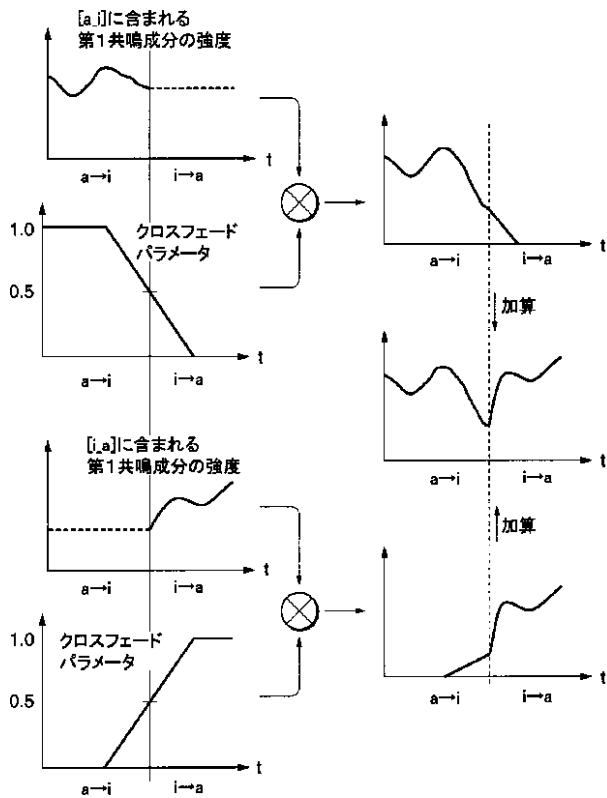
【図17】



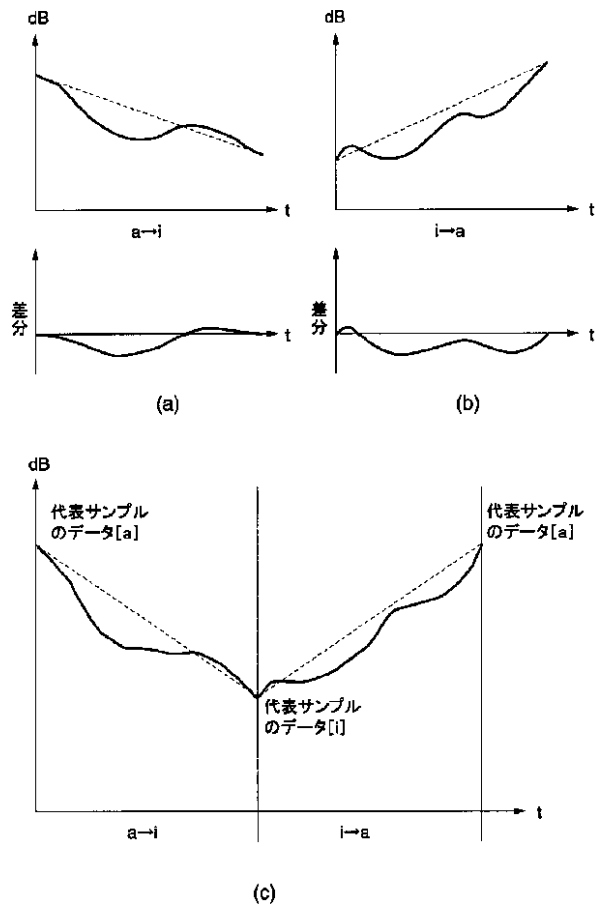
【図18】



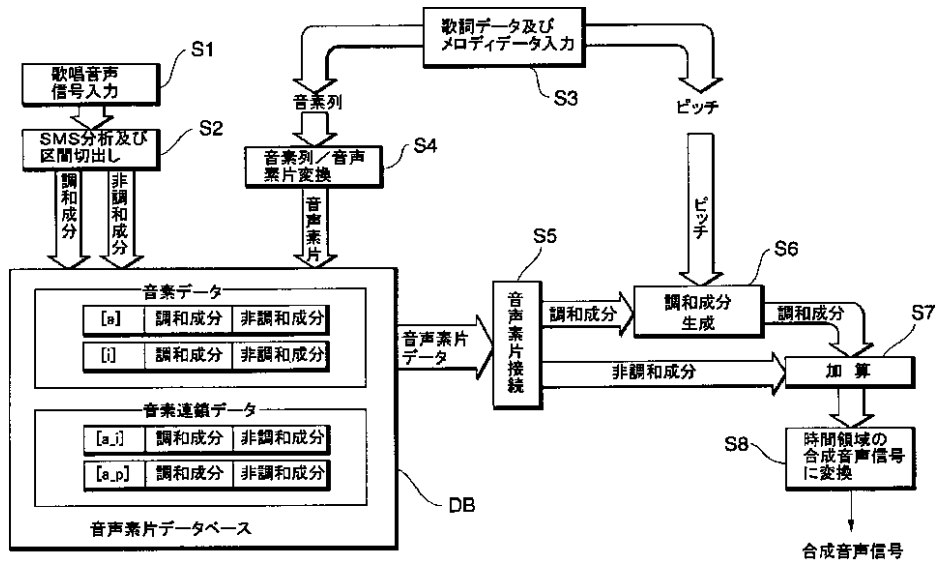
【図19】



【図20】



【図21】



フロントページの続き

(72)発明者 アレックス ロスコス
 バセイ デ シルコンバル - ラシオ、 8 .
 08003 バルセロナ

Fターム(参考) 5D045 AA07 BA10