(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2006-251375
(P2006-251375A)
(43)公開日　平成18年9月21日(2006.9.21)

(51)Int.Cl.		F I		テーマコード（参考）
G10L 21/04	(2006.01)	G10L 21/04	120C	
G10L 21/02	(2006.01)	G10L 21/02	400	

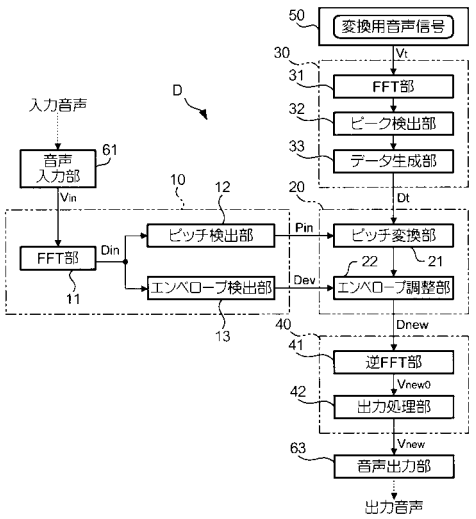
審査請求 未請求　請求項の数 8　OL　（全 15 頁）

(21)出願番号　特願2005-67907(P2005-67907)
(22)出願日　平成17年3月10日(2005.3.10)

(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
(74)代理人　100098084
弁理士　川▲崎▼　研二
(72)発明者　劔持　秀紀
静岡県浜松市中沢町１０番１号　ヤマハ株式会社内
(72)発明者　吉岡　靖雄
静岡県浜松市中沢町１０番１号　ヤマハ株式会社内
(72)発明者　ジョルディ　ボナダ
スペイン　バルセロナ　パセイ　デ　シルコンバルーラシオ　８

(54)【発明の名称】音声処理装置およびプログラム

(57)【要約】
【課題】　入力音声を簡易な構成によって多人数での合唱音や合奏音に変換する。
【解決手段】　ピッチ検出部１２は、音声入力部６１から供給される入力音声信号Ｖinのピッチ Ｐinを検出する。エンベロープ検出部１３は、入力音声信号Ｖinのスペクトルエンベロープを検出する。スペクトル取得手段３０は、並列に発生された複数の音声を含む変換用音声の周波数スペクトルを取得する。ピッチ変換部２１は、スペクトル取得手段３０が取得した周波数スペクトルの各ピークの周波数をピッチ Ｐinに応じて変化させる。エンベロープ調整部２２は、ピッチ変換部２１による処理後の周波数スペクトルのスペクトルエンベロープをエンベロープ検出部１３が検出したスペクトルエンベロープと略一致するように調整する。音声生成手段４０は、エンベロープ調整部２２による調整後の周波数スペクトルから出力音声信号Ｖnewを生成する。
【選択図】　図１



【特許請求の範囲】
【請求項１】
入力された入力音声のスペクトルエンベロープを検出するエンベロープ検出手段と、
並列に発生した複数の音声を含む変換用音声の周波数スペクトルである変換用スペクトルを取得するスペクトル取得手段と、
前記スペクトル取得手段が取得した変換用スペクトルのスペクトルエンベロープを前記エンベロープ検出手段が検出したスペクトルエンベロープと略一致するように調整するエンベロープ調整手段と、
前記エンベロープ調整手段による調整後の変換用スペクトルから出力音声信号を生成する音声生成手段と
を具備する音声処理装置。
【請求項２】
前記入力音声のピッチを検出するピッチ検出手段と、
前記スペクトル取得手段が取得した変換用スペクトルに含まれる各ピークの周波数を前記ピッチ検出手段が検出したピッチに応じて変化させるピッチ変換手段と
を具備し、
前記エンベロープ調整手段は、前記ピッチ変換手段による処理後の変換用スペクトルのスペクトルエンベロープを調整する
請求項１に記載の音声処理装置。
【請求項３】
前記ピッチ変換手段は、前記ピッチ検出手段が検出したピッチに応じて変換用スペクトルを周波数軸の方向に伸長または縮小する
請求項２に記載の音声処理装置。
【請求項４】
前記ピッチ変換手段は、前記変換用スペクトルにおける各ピークの周波数を含むスペクトル分布領域の各々を前記ピッチ検出手段が検出したピッチに応じて周波数軸の方向に移動させる
請求項２に記載の音声処理装置。
【請求項５】
前記入力音声のピッチを検出するピッチ検出手段を具備し、
前記スペクトル取得手段は、各々のピッチが相違する複数の変換用音声のうち前記ピッチ検出手段が検出したピッチに近似するピッチの変換用音声の変換用スペクトルを取得する
請求項１に記載の音声処理装置。
【請求項６】
入力された入力音声のスペクトルエンベロープを検出するエンベロープ検出手段と、
変換用音声の周波数スペクトルである第１変換用スペクトル、および、前記第１変換用スペクトルが示す変換用音声と略同一ピッチの音声の周波数スペクトルであり前記第１変換用スペクトルよりも各ピークの帯域幅が広い第２変換用スペクトルの何れかを取得するスペクトル取得手段と、
前記スペクトル取得手段が取得した変換用スペクトルのスペクトルエンベロープを前記エンベロープ検出手段が検出したスペクトルエンベロープと略一致するように調整するエンベロープ調整手段と、
前記エンベロープ調整手段による調整後の変換用スペクトルから出力音声信号を生成する音声生成手段と
を具備する音声処理装置。
【請求項７】
コンピュータに、
入力された入力音声のスペクトルエンベロープを検出するエンベロープ検出処理と、
並列に発生した複数の音声を含む変換用音声の周波数スペクトルである変換用スペクト

ルを取得するスペクトル取得処理と、
前記スペクトル取得処理にて取得した変換用スペクトルのスペクトルエンベロープを前記エンベロープ検出処理にて検出したスペクトルエンベロープと略一致するように調整するエンベロープ調整処理と、
前記エンベロープ調整処理後の変換用スペクトルから出力音声信号を生成する音声生成処理と
を実行させるプログラム。
【請求項８】
コンピュータに、
入力された入力音声のスペクトルエンベロープを検出するエンベロープ検出処理と、
変換用音声の周波数スペクトルである第１変換用スペクトル、および、前記第１変換用スペクトルが示す変換用音声と略同一ピッチの音声の周波数スペクトルであり前記第１変換用スペクトルよりも各ピークの帯域幅が広い第２変換用スペクトルの何れかを取得するスペクトル取得処理と、
前記スペクトル取得処理にて取得した変換用スペクトルのスペクトルエンベロープを前記エンベロープ検出処理にて検出したスペクトルエンベロープと略一致するように調整するエンベロープ調整処理と、
前記エンベロープ調整処理後の変換用スペクトルから出力音声信号を生成する音声生成処理と
を実行させるプログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、音声の特性を変化させる技術に関する。
【背景技術】
【０００２】
利用者が発声した音声（以下「入力音声」という）に音楽的な効果を付与するための種々の技術が従来から提案されている。例えば特許文献１には、入力音声のピッチを変換することによって生成された協和音（入力音声と和音を構成する音声）を入力音声と加算して出力する技術が開示されている。この構成によれば、実際の発声者がひとりであっても、恰も複数人にて別個の旋律を合唱しているかのような音声を出力することができる。また、例えば入力音声を楽器の演奏音とすれば、複数の楽器によって別個の旋律を合奏しているかのような音声が生成される。
【特許文献１】特開平１０－７８７７６号公報（段落００１３および図１参照）
【発明の開示】
【発明が解決しようとする課題】
【０００３】
ところで、合唱や合奏の形態としては、各歌唱者や演奏者が別個の旋律を歌唱または演奏する形態（いわゆるコーラス）のほか、複数の歌唱者や演奏者が同一の旋律を歌唱または演奏するユニゾンと呼ばれる形態がある。特許文献１に記載された構成においては、入力音声のピッチを変換することによって協和音が生成されるため、複数人が別個の旋律を歌唱ないし演奏したときの音声を生成することはできるものの、複数人が共通の旋律を歌唱または演奏するユニゾンの効果を入力音声に付与することはできない。なお、特許文献１に記載された構成においても、例えば入力音声のピッチを変更せずに音響的な特性（声質）のみを変換した音声を入力音声とともに出力すれば、複数人が共通の旋律を歌唱または演奏しているかのような効果を付与することも一応は可能である。しかしながら、この場合には、ユニゾンを構成する音声ごとに入力音声の特性を変換するための仕組みを用意することが不可欠となる。したがって、多人数によるユニゾンを実現しようとすれば、ＤＳＰ（Digital Signal Processor）などのハードウェアによって入力音声の特性が変換される構成においてはその回路規模が肥大化し、この変換がソフトウェアによって実現され

る構成においては演算装置の処理負荷が過大になるといった問題がある。本発明は、このような事情に鑑みてなされたものであり、入力音声を簡易な構成によって多人数での合唱音や合奏音に変換することを目的としている。

【課題を解決するための手段】

【０００４】

この課題を解決するために、本発明に係る音声処理装置は、入力された入力音声のスペクトルエンベロープを検出するエンベロープ検出手段と、並列に発生した複数の音声を含む変換用音声の周波数スペクトルである変換用スペクトルを取得するスペクトル取得手段と、スペクトル取得手段が取得した変換用スペクトルのスペクトルエンベロープをエンベロープ検出手段が検出したスペクトルエンベロープと略一致するように調整するエンベロープ調整手段と、エンベロープ調整手段による調整後の変換用スペクトルから出力音声信号を生成する音声生成手段とを具備することを特徴としている。なお、本発明にいう「音声」には、人間が発声した音声や楽器の演奏音といった種々の音響が含まれる。

この構成によれば、並列に発生した複数の音声を含む変換用音声の変換用スペクトルのエンベロープが入力音声のスペクトルエンベロープと略一致するように調整されるから、入力音声と同様の音韻をもった複数の音声（すなわち合奏音や合唱音）を示す出力音声信号を生成することができる。しかも、複数の音声の各々について入力音声の特性を変換するための仕組みは原理的に不要であるから、音声処理装置の構成は特許文献１の構成と比較して大幅に簡素化される。

なお、エンベロープ検出手段が検出したスペクトルエンベロープと変換用スペクトルのスペクトルエンベロープとが「略一致する」とは、エンベロープ調整手段による調整後の周波数スペクトルから生成された出力音声信号に基づいて実際に音声が放音されたときに、その音声の音韻が聴感上において入力音声の音韻と同一であると知覚される程度に近似（理想的には一致）していることを意味する。したがって、入力音声のスペクトルエンベロープとエンベロープ調整手段による調整後のスペクトルエンベロープとは厳密な意味で完全に一致している必要は必ずしもない。

【０００５】

本発明に係る音声処理装置において、音声生成手段が生成した出力音声信号は、例えばスピーカやイヤホンなどの放音機器に供給されて音声（以下「出力音声」という）として出力される。ただし、この出力音声信号が利用の態様は任意である。例えば、出力音声信号が記録媒体に記憶されたうえで、当該記憶手段を再生する他の装置にて出力音声が出力される態様としてもよいし、出力音声信号が通信回線を介して他の装置に送信されて当該装置にて音声として再生される態様としてもよい。

【０００６】

音声生成手段が生成する出力音声信号のピッチ（換言すると出力音声のピッチ）は、入力音声のピッチとは無関係なピッチであってもよいが、より好適には入力音声に応じたピッチ（例えば入力音声と略一致するピッチや入力音声と協和音を構成するピッチ）とされる。この好適な態様においては、例えば、入力音声のピッチを検出するピッチ検出手段と、スペクトル取得手段が取得した変換用スペクトルに含まれる各ピークの周波数をピッチ検出手段が検出したピッチに応じて変化させるピッチ変換手段とが更に設けられ、エンベロープ調整手段は、ピッチ変換手段による処理後の変換用スペクトルのスペクトルエンベロープを調整する。この態様によれば、出力音声信号が入力音声に対応したピッチとされるから、この出力音声信号に基づいて放音される音声を聴感上において心地よい音声とすることができる。

より具体的な態様において、ピッチ変換手段は、ピッチ検出手段が検出したピッチに応じて変換用スペクトルを周波数軸の方向に伸長または縮小する。この態様によれば、変換用スペクトルの各周波数に対して入力音声のピッチに応じた数値を乗算するという簡易な処理によって変換用スペクトルのピッチを調整することができる。また、他の態様において、ピッチ変換手段は、変換用スペクトルにおける各ピークの周波数を含むスペクトル分布領域（例えばピークの周波数を中心とする所定幅の周波数帯域）の各々をピッチ検出手

段が検出したピッチに応じた周波数軸の方向に移動させる（図８参照）。この態様によれば、変換用スペクトルの各ピークの周波数を所期の周波数に合致させることができるから、変換用スペクトルのピッチを所望のピッチに精度よく調整することができる。

【０００７】

もっとも、出力音声を入力音声に応じたピッチとするための構成は任意である。例えば、入力音声のピッチを検出するピッチ検出手段を設けたうえで、スペクトル取得手段が、各々のピッチが相違する複数の変換用音声のうちピッチ検出手段が検出したピッチに近似（理想的には一致）するピッチの変換用音声の変換用スペクトルを取得する態様としてもよい（図６参照）。この態様によれば、変換用音声のピッチを変換するための仕組みを不要とすることができる。ただし、変換用スペクトルのピッチを変換する構成と、各々のピッチが異なる複数の変換用音声の何れかを選択する構成とを組み合わせてもよい。例えば、各々が異なるピッチに対応する複数の変換用スペクトルのうち入力音声のピッチに近似するピッチに対応した変換用スペクトルをスペクトル取得手段が取得し、この選択した変換用スペクトルのピッチをピッチ変換手段がピッチデータに応じて変換する構成も採用される。

【０００８】

ところで、複数の歌唱者や演奏者から略同一のピッチにて同時に（並列に）発せられた音声の周波数スペクトルは、その各ピークの帯域幅（例えば図３に示される帯域幅Ｗ2）が、単一の歌唱者や演奏者から発せられた音声の周波数スペクトルにおける各ピークの帯域幅（例えば図２に示される帯域幅Ｗ1）よりも広い場合が多い。いわゆるユニゾンにおいては、各歌唱者や各演奏者の音声のピッチが厳密には一致していないからである。このような観点から、本発明に係る音声処理装置は、入力された入力音声のスペクトルエンベロープを検出するエンベロープ検出手段と、変換用音声の周波数スペクトルである第１変換用スペクトル、および、第１変換用スペクトルが示す変換用音声と略同一ピッチの音声の周波数スペクトルであり第１変換用スペクトルよりも各ピークの帯域幅が広い第２変換用スペクトルの何れかを取得するスペクトル取得手段と、スペクトル取得手段が取得した変換用スペクトルのスペクトルエンベロープをエンベロープ検出手段が検出したスペクトルエンベロープと略一致するように調整するエンベロープ調整手段と、エンベロープ調整手段による調整後の変換用スペクトルから出力音声信号を生成する音声生成手段とを具備する構成としても特定される。なお、この構成の具体例は第２実施形態（図５）として後述される。

この構成によれば、出力音声信号を生成するための周波数スペクトルとして第１変換用スペクトルおよび第２変換用スペクトルの何れかが選択されるから、第１変換用スペクトルに応じた特性の出力音声信号と第２変換用スペクトルに応じた特性の出力音声信号とを選択的に生成することができる。例えば、第１変換用スペクトルが選択された場合には単一の歌唱者または演奏者から発せられた出力音声を生成することができ、第２変換用スペクトルが選択された場合には複数の歌唱者や演奏者から発せられた出力音声を生成することができる。なお、ここでは第１変換用スペクトルと第２変換用スペクトルとが特定されているが、更に他の変換用スペクトルが選択手段による選択の対象とされる構成としてもよい。例えば、それぞれ各ピークの帯域幅が相違する複数の変換用スペクトルを記憶手段に記憶させておき、このうちの何れかを選択手段が選択して出力音声信号の生成に利用するといった構成も採用される。

【０００９】

本発明に係る音声処理装置は、音声処理に専用されるＤＳＰなどのハードウェアによって実現されるほか、パーソナルコンピュータなどのコンピュータとプログラムとの協働によっても実現される。このプログラムは、入力された入力音声のスペクトルエンベロープを検出するエンベロープ検出処理と、並列に発生した複数の音声を含む変換用音声の周波数スペクトルである変換用スペクトルを取得するスペクトル取得処理と、スペクトル取得処理にて取得した変換用スペクトルのスペクトルエンベロープをエンベロープ検出処理にて検出したスペクトルエンベロープと略一致するように調整するエンベロープ調整処理と

、エンベロープ調整処理後の変換用スペクトルから出力音声信号を生成する音声生成処理とをコンピュータに実行させる内容となる。また、他の態様に係るプログラムは、入力された入力音声のスペクトルエンベロープを検出するエンベロープ検出処理と、変換用音声の周波数スペクトルである第１変換用スペクトル、および、第１変換用スペクトルが示す変換用音声と略同一ピッチの音声の周波数スペクトルであり第１変換用スペクトルよりも各ピークの帯域幅が広い第２変換用スペクトルの何れかを取得するスペクトル取得処理と、スペクトル取得処理にて取得した変換用スペクトルのスペクトルエンベロープをエンベロープ検出処理にて検出したスペクトルエンベロープと略一致するように調整するエンベロープ調整処理と、エンベロープ調整処理後の変換用スペクトルから出力音声信号を生成する音声生成処理とをコンピュータに実行させる内容となる。これらのプログラムは、コンピュータ読取り可能な記録媒体（例えばＣＤ－ＲＯＭ）に格納された態様にて利用者に提供されてコンピュータにインストールされるほか、ネットワークを介した配信の形態にてサーバ装置から提供されてコンピュータにインストールされる。

【００１０】

また、本発明は、入力音声を処理するための方法としても特定される。この方法は、入力された入力音声のスペクトルエンベロープを検出するエンベロープ検出過程と、並列に発生した複数の音声を含む変換用音声の周波数スペクトルである変換用スペクトルを取得するスペクトル取得過程と、スペクトル取得過程にて取得した変換用スペクトルのスペクトルエンベロープをエンベロープ検出過程にて検出したスペクトルエンベロープと略一致するように調整するエンベロープ調整過程と、エンベロープ調整過程における調整後の変換用スペクトルから出力音声信号を生成する音声生成過程とを有する。他の観点に基づく音声処理方法は、入力された入力音声のスペクトルエンベロープを検出するエンベロープ検出過程と、変換用音声の周波数スペクトルである第１変換用スペクトル、および、第１変換用スペクトルが示す変換用音声と略同一ピッチの音声の周波数スペクトルであり第１変換用スペクトルよりも各ピークの帯域幅が広い第２変換用スペクトルの何れかを取得するスペクトル取得過程と、スペクトル取得過程にて取得した変換用スペクトルのスペクトルエンベロープをエンベロープ検出過程にて検出したスペクトルエンベロープと略一致するように調整するエンベロープ調整過程と、エンベロープ調整過程における調整後の変換用スペクトルから出力音声信号を生成する音声生成過程とを有する。

【発明の効果】

【００１１】

以上のように、本発明によれば、簡易な構成によって多人数での合唱や合奏を実現することができる。

【発明を実施するための最良の形態】

【００１２】

＜Ａ：第１実施形態＞

まず、図１を参照して、本発明の第１実施形態に係る音声処理装置の構成および動作を説明する。同図に示される音声処理装置の各部は、例えばＣＰＵ（Central Processing Unit）などの演算回路がプログラムを実行することによって実現されてもよいし、ＤＳＰなど音声処理に専用されるハードウェアによって実現されてもよい。後述する各実施形態においても同様である。

【００１３】

図１に示されるように、音声処理装置Ｄは、周波数分析手段１０と、スペクトル変換手段２０と、スペクトル取得手段３０と、音声生成手段４０と、記憶手段５０とを有する。このうち周波数分析手段１０には音声入力部６１が接続される。この音声入力部６１は、利用者が発する入力音声に応じた信号（以下「入力音声信号」という）Ｖinを出力する手段であり、例えば、入力音声の時間軸上における波形を表わすアナログの電気信号を出力する収音機器（マイクロホン）と、この電気信号をデジタルの入力音声信号Ｖinに変換するＡ／Ｄ変換器とを有する。

【００１４】

周波数分析手段１０は、音声入力部６１から供給される入力音声信号Ｖinのピッチ Ｐin およびスペクトルエンベロープＥＶinを特定する手段であり、ＦＦＴ（Fast Fourier Transform）部１１とピッチ検出部１２とエンベロープ検出部１３とを有する。このうちＦＦＴ部１１は、音声入力部６１から供給される入力音声信号Ｖinを所定の時間長（例えば５ｍｓないし１０ｍｓ）のフレームに切り出し、各フレームの入力音声信号Ｖinに対してＦＦＴ処理を含む周波数分析を実行して周波数スペクトル（以下「入力スペクトル」という）ＳＰinを検出する。入力音声信号Ｖinの各フレームは時間軸上において相互に重なり合うように選定される。これらのフレームは簡易的には同一の時間長とされるが、入力音声信号Ｖinのピッチ Ｐin（後述するようにピッチ検出部１２によって検出される）に応じて時間長が変化する構成としてもよい。図２には、ひとりの利用者が発声した入力音声のうちひとつのフレームについて特定された入力スペクトルＳＰinが例示されている。この場合の入力スペクトルＳＰinは、基音および倍音に相当する各周波数においてスペクトル強度Ｍの局所的なピークｐが極めて狭い帯域幅Ｗ1にて現れる。ＦＦＴ部１１は、入力音声信号Ｖinの入力スペクトルＳＰinを表わすデータ（以下「入力スペクトルデータ」という）Ｄinをフレームごとにピッチ検出部１２とエンベロープ検出部１３とに出力する。入力スペクトルデータＤinは複数の単位データを含む。各単位データは、周波数軸上に所定の間隔ごとに選定された複数の周波数Ｆinの各々と当該周波数における入力スペクトルＳＰinのスペクトル強度Ｍinとが組み合わされたデータである。

【００１５】

図１に示されるピッチ検出部１２は、ＦＦＴ部１１から供給される入力スペクトルデータＤinに基づいて入力音声のピッチ Ｐinを検出する手段である。更に詳述すると、ピッチ検出部１２は、図２に示されるように、入力スペクトルデータＤinが示す入力スペクトルＳＰinのうち基音に相当するピークｐ（すなわち周波数が最小であるピークｐ）の周波数をピッチ Ｐinとして検出する。一方、エンベロープ検出部１３は、入力音声のスペクトルエンベロープ（スペクトル包絡）ＥＶinを検出する手段である。スペクトルエンベロープＥＶinは、図２に示されるように、入力スペクトルＳＰinのピークｐを連結した包絡線である。このスペクトルエンベロープＥＶinを検出する方法としては、例えば、入力スペクトルＳＰinのうち周波数軸上において相互に隣接するピークｐの間隙を直線的に補間することによってスペクトルエンベロープＥＶinを折線として検出する方法や、各ピークｐを通過する曲線を３次のスプライン補間など各種の補間処理によって算定してスペクトルエンベロープＥＶinを検出する方法などが採用される。エンベロープ検出部１３は、図２に示されるように、こうして検出したスペクトルエンベロープＥＶinを示すデータ（以下「エンベロープデータ」という）Ｄevを出力する。エンベロープデータＤevは、入力スペクトルデータＤinと同様に複数の単位データＵevを含む。各単位データＵevは、周波数軸上に所定の間隔ごとに選定された複数の周波数Ｆin（Ｆin1, Ｆin2, ……）の各々と当該周波数Ｆinにおけるスペクトルエンベロープ ＥＶinのスペクトル強度Ｍev（Ｍev1, Ｍev2, ……）とが組み合わされたデータである。

【００１６】

次に、図１に示されるスペクトル変換手段２０は、入力音声の特性を変化させた出力音声の周波数スペクトル（以下「出力スペクトル」という）ＳＰnewを示すデータ（以下「新規スペクトルデータ」という）Ｄnewを生成する手段である。本実施形態におけるスペクトル変換手段２０は、予め用意された特定の音声（以下「変換用音声」という）の周波数スペクトル（以下「変換用スペクトル」という）ＳＰtと入力音声のスペクトルエンベロープＥＶinとに基づいて出力音声の周波数スペクトルＳＰnewを特定する。なお、周波数スペクトルＳＰnewを生成する手順については後述する。

【００１７】

一方、スペクトル取得手段３０は、変換用スペクトルＳＰtを取得するための手段であり、ＦＦＴ部３１とピーク検出部３２とデータ生成部３３とを有する。このうちＦＦＴ部３１には、記憶手段５０（例えばハードディスク装置）から読み出された変換用音声信号Ｖtが供給される。この変換用音声信号Ｖtは、変換用音声の波形を特定の区間にわたって

表わす時間領域の信号であり、予め記憶手段５０に格納されている。ＦＦＴ部３１は、入力音声に係る手順と同様に、記憶手段５０から順次に供給される変換用音声信号Ｖｔを所定の時間長のフレームに切り出し、各フレームの変換用音声信号Ｖｔに対してＦＦＴ処理を含む周波数分析を実行することによって変換用スペクトルＳＰｔを検出する。一方、ピーク検出部３２は、ＦＦＴ部３１によって特定された変換用スペクトルＳＰｔのピークｐｔを検出してその周波数を特定する。ピークｐｔを検出する方法としては、例えば、周波数軸上において近接する所定数のピークのうちスペクトル強度が最大となるものをピークｐｔとして検出する方法が採用される。

【００１８】

本実施形態においては、多数の発声者が略同一のピッチＰｔにて発声した音声（すなわち合唱や合奏といったユニゾンの音声）をマイクロホンなどの収音機器によって収音した信号が変換用音声信号Ｖinとして記憶手段５０に記憶されている場合を想定する。このような変換用音声信号ＶｔにＦＦＴ処理を施して得られる変換用スペクトルＳＰｔは、図３に示されるように、変換用音声のピッチＰｔに応じた基音および倍音に相当する各周波数においてスペクトル強度Ｍの局所的なピークｐｔが現れる点で図１の入力スペクトルＳＰinと共通するが、各ピークｐｔの帯域幅Ｗ2が入力スペクトルＳＰinの各ピークｐの帯域幅Ｗ1よりも広いという特性を有する。このようにピークｐｔの帯域幅Ｗ2が広いのは、多数の発声者によって発声された各音声のピッチが完全には一致しないからである。

【００１９】

図１に示されるデータ生成部３３は、変換用スペクトルＳＰｔを示すデータ（以下「変換用スペクトルデータ」という）Ｄｔを生成するための手段である。変換用スペクトルデータＤｔは、図３に示されるように、複数の単位データＵｔと指示子Ａとを含む。各単位データＵｔは、エンベロープデータＤevと同様に、周波数軸上に所定の間隔ごとに選定された複数の周波数Ｆｔ（Ｆｔ1，Ｆｔ2，……）の各々と当該周波数Ｆｔにおける変換用スペクトルＳＰｔのスペクトル強度Ｍｔ（Ｍｔ1，Ｍｔ2，……）とが組み合わされたデータ構造となっている。一方、指示子Ａは、変換用スペクトルＳＰｔのピークｐｔを指示するためのデータ（例えばフラグ）であり、変換用スペクトルデータＤｔに含まれる総ての単位データＵｔのうちピーク検出部３２によって検出されたピークｐｔに対応する単位データＵｔに対して選択的に付加される。例えば、ピーク検出部３２が周波数Ｆｔ3にピークｐｔを検出した場合、図３に示されるように、周波数Ｆｔ3を含む単位データＵｔに指示子Ａが付加され、これ以外の単位データＵｔ（つまりピークｐｔ以外の周波数に対応する単位データＵｔ）に指示子Ａは付加されない。

【００２０】

図１に示されるように、スペクトル変換手段２０は、ピッチ変換部２１とエンベロープ調整部２２とを有する。スペクトル取得手段３０から出力された変換用スペクトルデータＤｔはピッチ変換部２１に入力される。このピッチ変換部２１は、変換用スペクトルデータＤｔが示す変換用スペクトルＳＰｔの各ピークｐｔの周波数を、ピッチ検出部１２が検出したピッチＰinに応じて変化させる手段である。本実施形態におけるピッチ変換部２１は、変換用スペクトルデータＤｔが示す変換用音声のピッチＰｔがピッチ検出部１２によって検出されたピッチＰinと略一致するように変換用スペクトルＳＰｔを変形する。この変換の具体的な手順について図４を参照して説明する。

【００２１】

図４の部分（ｂ）には、図３に示した変換用スペクトルＳＰｔが図示されている。また、図４の部分（ａ）には、入力スペクトルＳＰin（図２に示したもの）が変換用スペクトルＳＰｔとの対比のために併記されている。入力音声のピッチＰinは利用者の発声に応じて変動するから、図４の部分（ａ）および部分（ｂ）に示されるように、入力スペクトルＳＰinの各ピークｐの周波数と変換用スペクトルＳＰｔの各ピークｐｔの周波数とは必ずしも一致しない。そこで、ピッチ変換部２１は、変換用スペクトルＳＰｔを周波数軸の方向に伸長または縮小することによって当該変換用スペクトルＳＰｔの各ピークｐｔの周波数を入力スペクトルＳＰinの各ピークｐの周波数に合致させる。更に詳述すると、ピッチ変換

部２１は、ピッチ検出部１２が検出したピッチＰinと変換用音声のピッチＰtとの比「Ｐin／Ｐt」を算定し、変換用スペクトルデータＤtを構成する各単位データＵtの周波数Ｆtに対して当該比を乗算する。なお、変換用音声のピッチＰtは、例えば、変換用スペクトルＳＰtの多数のピークｐtのうち基音に相当するピークｐt（すなわち周波数が最小であるピークｐt）の周波数として特定される。この処理により、図４の部分（ｃ）に示されるように、変換用スペクトルＳＰtの各ピークｐtは入力スペクトルＳＰinの各ピークｐの周波数まで移動し、この結果として変換用音声のピッチＰtは入力音声のピッチＰinに略一致することになる。ピッチ変換部２１は、こうしてピッチを変換した変換用スペクトルＳＰtを示す変換用スペクトルデータＤtをエンベロープ調整部２２に出力する。

【００２２】

エンベロープ調整部２２は、この変換用スペクトルデータＤtが示す変換用スペクトルＳＰtのスペクトル強度Ｍ（換言すればスペクトルエンベロープＥＶt）を調整することによって新規スペクトルＳＰnewを生成する手段である。更に詳述すると、エンベロープ調整部２２は、図４の部分（ｄ）に示されるように、新規スペクトルＳＰnewのスペクトルエンベロープが、エンベロープ検出部１３によって検出されたスペクトルエンベロープＥＶinと略一致するように、変換用スペクトルＳＰtのスペクトル強度Ｍを調整する。スペクトル強度Ｍを調整する方法の具体例は以下の通りである。

【００２３】

エンベロープ調整部２２は、まず、変換用スペクトルデータＤtのうち指示子Ａが付加されたひとつの単位データＵtを選定する。この単位データＵtは、変換用スペクトルＳＰtの何れかのピークｐt（以下では特に「注目ピークｐt」という）の周波数Ｆtおよびスペクトル強度Ｍtを含む（図３参照）。次いで、エンベロープ調整部２２は、エンベロープ検出部１３から供給されるエンベロープデータＤevのうち注目ピークｐtの周波数Ｆtに近似または一致する周波数Ｆinを含む単位データＵevを選定する。そして、エンベロープ調整部２２は、この選定した単位データＵevに含まれるスペクトル強度Ｍevと注目ピークｐtのスペクトル強度Ｍtとの比「Ｍev／Ｍt」を算定し、注目ピークｐtを中心とした所定の帯域に属する変換用スペクトルＳＰtの各単位データＵtのスペクトル強度Ｍtに対して当該比を乗算する。この一連の処理を変換用スペクトルＳＰtの総てのピークｐtについて繰り返すことにより、新規スペクトルＳＰnewは、図４の部分（ｄ）に示されるように、各ピークの頂点がスペクトルエンベロープＥＶin上に位置する形状となる。エンベロープ調整部２２は、この新規スペクトルＳＰnewを示す新規スペクトルデータＤnewを出力する。

【００２４】

ピッチ変換部２１やエンベロープ調整部２２による処理は入力音声信号Ｖinを区分したフレームごとに実施される。ところで、変換用音声のフレーム数は記憶手段５０に記憶された変換用音声信号Ｖtの時間長に応じて制約されるのに対して入力音声のフレーム数は利用者による発声の期間に応じて変化するため、入力音声のフレーム数と変換用音声のフレーム数とは一致しない場合が多い。変換用音声のフレーム数が入力音声のフレーム数よりも多い場合には、余ったフレームに対応する変換用スペクトルデータＤtを破棄すれば足りる。一方、変換用音声のフレーム数が入力音声のフレーム数よりも少ない場合には、総てのフレームに対応する変換用スペクトルデータＤtの使用後に最初のフレームの変換用スペクトルデータＤtを使用するといった具合に、変換用スペクトルデータＤtをループさせて使用すればよい。

【００２５】

次に、図１に示される音声生成手段４０は、新規スペクトルＳＰnewに基づいて時間領域の出力音声信号Ｖnewを生成する手段であり、逆ＦＦＴ部４１と出力処理部４２とを有する。このうち逆ＦＦＴ部４１は、エンベロープ調整部２２からフレームごとに出力される新規スペクトルデータＤnewに対して逆ＦＦＴ処理を施して時間領域の出力音声信号Ｖnew0を生成する。出力処理部４２は、こうして生成されたフレームごとの出力音声信号Ｖnew0に時間窓関数を乗算し、これらを時間軸上において相互に重なり合うように連結して出力音声信号Ｖnewを生成する。この出力音声信号Ｖnewは音声出力部６３に供給される。

音声出力部６３は、出力音声信号Ｖnewをアナログの電気信号に変換するＤ／Ａ変換器と、このＤ／Ａ変換器からの出力信号に基づいて放音する放音機器（例えばスピーカやヘッドフォン）とを有する。

【００２６】

以上に説明したように、本実施形態においては、多数の発声者によって並列に発せられた複数の音声を含む変換用音声のスペクトルエンベロープＥＶtが入力音声のスペクトルエンベロープＥＶinと略一致するように調整されるから、入力音声と同様の音韻をもった複数の音声（すなわち合唱音や合奏音）を示す出力音声信号Ｖnewを生成することができる。したがって、ひとりの利用者による音声や演奏音が入力音声とされた場合であっても、恰も多数の発声者や演奏者によって合唱や合奏が行なわれているかのような出力音声を音声出力部６３から出力することができる。しかも、複数の音声の各々について入力音声の特性を変化させるための仕組みは原理的に不要である。したがって、音声処理装置Ｄの構成は特許文献１の構成と比較して大幅に簡素化される。さらに、本実施形態においては、入力音声のピッチＰinに応じて変換用音声のピッチＰtが変換されるから、任意のピッチの合唱音や合奏音を生成することができる。また、このピッチの変換が、変換用スペクトルＳＰtを周波数軸の方向に伸長するという簡素な処理（乗算処理）によって実現されるという利点もある。

【００２７】

＜Ｂ：第２実施形態＞

次に、本発明の第２実施形態に係る音声処理装置について説明する。なお、本実施形態のうち第１実施形態と同様の要素については共通の符号を付してその説明を適宜に省略する。

【００２８】

図５は、本実施形態に係る音声処理装置Ｄの構成を示すブロック図である。同図に示されるように、この音声処理装置Ｄは、記憶手段５０の記憶内容およびスペクトル取得手段３０の構成が第１実施形態の音声処理装置Ｄとは相違するが、他の要素は同様の構成である。本実施形態においては、第１変換用音声信号Ｖt1と第２変換用音声信号Ｖt2とが記憶手段５０に記憶される。第１変換用音声信号Ｖt1と第２変換用音声信号Ｖt2とは、互いに略同一のピッチＰtにて発せられた変換用音声を収音した信号である。ただし、第１変換用音声信号Ｖt1は、図２に示した入力音声信号Ｖinと同様に、単一の音声（ひとりの発声者からの音声やひとつの楽器からの演奏音）の波形を示す信号であるのに対し、第２変換用音声信号Ｖt2は、第１実施形態の変換用音声信号Ｖtと同様に、各々が並列に発せられた複数の音声（多数の発声者からの音声や多数の楽器からの演奏音）からなる変換用音声を収音した信号である。したがって、第２変換用音声信号Ｖt2から特定される変換用スペクトルＳＰtの各ピークの帯域幅（図３に示す帯域幅W2）は、第１変換用音声信号Ｖt1から特定される変換用スペクトルＳＰtの各ピークの帯域幅（図１に示す帯域幅W1）よりも広い。

【００２９】

また、本実施形態におけるスペクトル取得手段３０はＦＦＴ部３１の前段に選択部３４を有する。この選択部３４は、外部から供給される選択信号に基づいて、第１変換用音声信号Ｖt1および第２変換用音声信号Ｖt2の何れかを選択して記憶手段５０から読み出す手段である。選択信号は、例えば、入力機器６７に対する操作に応じて供給される。この選択部３４によって読み出された変換用音声信号ＶtがＦＦＴ部３１に供給される。これ以後の構成および動作は第１実施形態と同様である。

【００３０】

このように、本実施形態においては、第１変換用音声信号Ｖt1および第２変換用音声信号Ｖt2の何れかが選択的に新規スペクトルＳＰnewの生成に利用される。そして、第１変換用音声信号Ｖt1が選択された場合には、入力音声の音韻と変換用音声の周波数特性とを兼ね備えた単一の出力音声が出力される一方、第２変換用音声信号Ｖt2が選択された場合には、第１実施形態と同様に、入力音声の音韻を維持した多数の音声からなる出力音声が

出力される。すなわち、本実施形態においては、出力音声を単一の音声とするか複数の音声とするかを利用者が任意に選択することができる。

【００３１】

なお、本実施形態においては入力機器６７への操作に応じて変換用音声信号Ｖｔが選択される構成を例示したが、この選択の基準となる要素は任意に変更される。例えば、所定の時間間隔にて発生するタイマ割込を契機として第１変換用音声信号Ｖｔ1および第２変換用音声信号Ｖｔ2の一方から他方に切り替える構成としてもよい。さらに、本実施形態に係る音声処理装置Ｄをカラオケ装置に適用した場合には、カラオケ演奏される楽曲の進行に同期して第１変換用音声信号Ｖｔ1および第２変換用音声信号Ｖｔ2の一方から他方に切り替える構成も採用される。また、本実施形態においては、単一の音声を示す第１変換用音声信号Ｖｔ1と複数の音声を示す第２変換用音声信号Ｖｔ2とが記憶手段５０に記憶された構成を例示したが、各変換用音声信号Ｖｔが示す音声数はこれに限られない。例えば、各々が並列に発せられた所定数の音声からなる変換用音声を示す第１変換用音声信号Ｖｔ1と、これよりも多数の音声からなる変換用音声を示す第２変換用音声信号Ｖｔ2とを利用してもよい。

【００３２】

＜Ｃ：変形例＞

各実施形態に対しては種々の変形が加えられる。具体的な変形の態様は以下の通りである。なお、以下の各態様を適宜に組み合わせてもよい。

【００３３】

（１）各実施形態においてはひとつのピッチＰｔの変換用音声信号Ｖｔ（またはＶｔ1，Ｖｔ2）が記憶手段５０に記憶された構成を例示したが、図６に示されるように、各々のピッチＰｔ（Ｐｔ1，Ｐｔ2，……）が相違する複数の変換用音声信号Ｖｔを記憶手段５０に記憶させた構成も採用される。各変換用音声信号Ｖｔは、並列に発生した多数の音声を含む変換用音声を収音したものである。図６の構成においては、ピッチ検出部１２によって検出されたピッチＰｉｎがスペクトル取得手段３０の選択部３４にも供給されるようになっている。この選択部３４は、入力音声のピッチＰｉｎに近似または一致するピッチＰｔの変換用音声信号Ｖｔを選択的に記憶手段５０から読み出してＦＦＴ部３１に出力する手段である。この構成によれば、新規スペクトルＳＰｎｅｗの生成に利用される変換用音声信号Ｖｔのピッチ Ｐｔを入力音声信号ＶｉｎのピッチＰｉｎに近づけることができるから、ピッチ変換部２１による処理にて変換用スペクトルＳＰｔの各ピークｐｔの周波数を変化させる量が低減される。したがって、自然な形状の新規スペクトルＳＰｎｅｗを生成することができるという利点がある。なお、ここでは変換用音声信号Ｖｔの選択に加えてピッチ変換部２１による処理も実行する構成としたが、多数のピッチＰｔの変換用音声信号Ｖｔが記憶手段５０に記憶されていれば変換用音声信号Ｖｔの選択のみによって所望のピッチの出力音声を生成することができるから、ピッチ変換部２１は必ずしも必要ではない。

【００３４】

（２）各実施形態においては、変換用スペクトルデータＤｔの各単位データＵｔに含まれる周波数Ｆｔに特定の数値（Ｐｉｎ／Ｐｔ）を乗算することによって変換用スペクトルＳＰｔを周波数軸の方向に伸長または縮小する構成を例示したが、変換用スペクトルＳＰｔのピッチＰｔを変換する方法は任意に変更される。例えば、各実施形態に示した方法においては、変換用スペクトルＳＰｔが全帯域にわたって同率に伸長または縮小されるため、各ピークｐｔの帯域幅が元のピークｐｔの帯域幅よりも著しく広がってしまう場合が生じ得る。例えば、図７の部分（ａ）に示される変換用スペクトルＳＰｔのピッチＰｔを第１実施形態の方法によって２倍のピッチに変換した場合、図７の部分（ｂ）に示されるように各ピークｐｔの帯域幅は２倍となる。このように各ピークｐｔのスペクトル形状が大幅に変化すると変換用音声の特性とは著しく相違する出力音声が生成されることになる。このような問題を解消するために、ピッチ変換部２１が、特定の数値（Ｐｉｎ／Ｐｔ）を乗算して得られた変換用スペクトルＳＰｔ（図７の部分（ｂ）に示される周波数スペクトル）の各ピークｐｔについて、図７の部分（ｃ）に矢印Ｂにて示されるように、当該ピークｐｔの帯域幅をピ

ッチ変換前のピークｐｔの帯域幅まで狭めるための演算処理を各単位データＵｔの周波数Ｆｔに施してもよい。この構成によれば、変換用音声の特性を忠実に再現した出力音声を生成することができる。

【００３５】

また、ここでは各単位データＵｔの周波数Ｆｔに対する乗算処理によってピッチＰｔを変換する場合を例示したが、図８の部分（ａ）に示されるように、変換用スペクトルＳＰｔを周波数軸上にて複数の帯域（以下「スペクトル分布領域」という）Ｒに区分し、各スペクトル分布領域Ｒを周波数軸の方向に移動させることによってピッチＰｔを変化させてもよい。各スペクトル分布領域Ｒは、ひとつのピークｐｔとその前後の帯域とを含むように選定される。ピッチ変換部２１は、図８の部分（ｂ）に示されるように、各スペクトル分布領域Ｒに属するピークｐｔの周波数が、入力スペクトルＳＰｉｎ（図８の部分（ｃ））に現れる各ピークｐの周波数と略一致するように、各スペクトル分布領域Ｒを周波数軸の方向に移動させる。なお、図８の部分（ｂ）に示されるように、相互に隣接するスペクトル分布領域Ｒの間隙には周波数スペクトルが存在しない帯域が生じ得るが、この帯域についてはスペクトル強度Ｍを所定値（例えばゼロ）に選定すればよい。この処理によれば、変換用スペクトルＳＰｔの各ピークｐｔの周波数を確実に入力音声のピークｐｔの周波数に一致させることができるから、所望のピッチの出力音声を精度よく生成することができるという利点がある。

【００３６】

（３）各実施形態においては、記憶手段５０に記憶された変換用音声信号Ｖｔから変換用スペクトルＳＰｔが特定される構成を例示したが、変換用スペクトルＳＰｔを示す変換用スペクトルデータＤｔが予めフレームごとに記憶手段５０に記憶された構成も採用される。この構成におけるスペクトル取得手段３０は、記憶手段５０から変換用スペクトルデータＤｔを読み出してスペクトル変換手段２０に出力する構成であれば足り、ＦＦＴ部３１やピーク検出部３２やデータ生成部３３を備えている必要はない。また、ここでは記憶手段５０に変換用スペクトルデータＤｔが記憶された構成を例示したが、スペクトル取得手段３０は、例えば通信回線を介して接続された通信装置から変換用スペクトルデータＤｔを取得する手段であってもよい。このように、本発明におけるスペクトル取得手段３０は、変換用スペクトルＳＰｔを取得する手段であれば足り、その取得の方法や取得先の如何は不問である。

【００３７】

（４）各実施形態においては入力音声の周波数スペクトルＳＰｉｎからピッチＰｉｎを検出する構成を例示したが、このピッチＰｉｎを検出する方法は任意に変更される。例えば、音声入力部６１から入力された時間領域の入力音声信号Ｖｉｎからピッチ Ｐｉｎを検出する構成としてもよい。ピッチＰｉｎを検出する方法としては、公知である各種の方法が採用される。

【００３８】

（５）各実施形態においては変換用音声のピッチＰｔを入力音声のピッチＰｉｎに一致させる構成を例示したが、変換用音声のピッチＰｔをこれ以外のピッチに変換してもよい。例えば、ピッチ変換部２１が、入力音声のピッチＰｉｎと協和音を構成するピッチとなるように変換用音声のピッチＰｔを変換する構成も採用される。この構成に加え、出力処理部４２から出力された出力音声信号Ｖｎｅｗと音声入力部６１から入力された入力音声信号Ｖｉｎとを加算したうえで音声出力部６３から放音する構成を採用すれば、利用者が発声した入力音声とともにコーラス音を出力することができる。このように、本発明のうちピッチ変換部２１を備えた態様においては、このピッチ変換部２１が変換用音声のピッチＰｔを入力音声のピッチＰｉｎに応じて（すなわちピッチＰｉｎの変化に伴なって変換用音声のピッチＰｔが変化するように）変化させる構成であれば足りる。

【図面の簡単な説明】

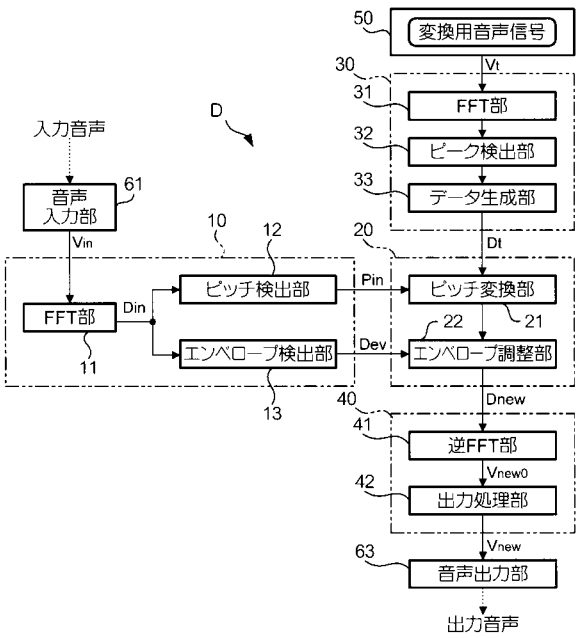
【００３９】

【図１】第１実施形態に係る音声処理装置の構成を示すブロック図である。

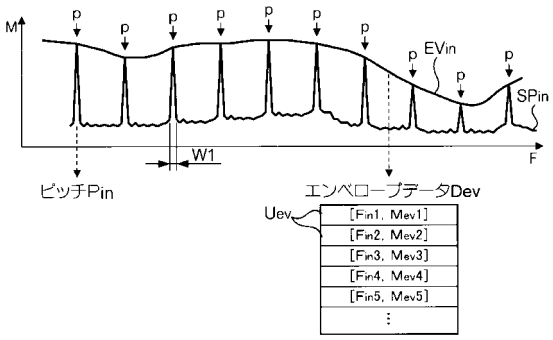
【図２】入力音声に関する処理を説明するための図である。

【図３】変換用音声信号に関する処理を説明するための図である。
【図４】スペクトル変換手段による処理の内容を説明するための図である。
【図５】第２実施形態に係る音声処理装置の構成を示すブロック図である。
【図６】変形例に係る音声処理装置の構成を示すブロック図である。
【図７】変形例に係る音声処理装置におけるピッチ変換について説明するための図である。
【図８】変形例に係る音声処理装置におけるピッチ変換について説明するための図である。
【符号の説明】
【００４０】
Ｄ……音声処理装置、１０……周波数分析手段、１１……ＦＦＴ部、１２……ピッチ検出部、１３……エンベロープ検出部、２０……スペクトル変換手段、２１……ピッチ変換部、２２……エンベロープ調整部、３０……スペクトル取得手段、３１……ＦＦＴ部、３２……ピーク検出部、３３……データ生成部、３４……選択部、４０……音声生成手段、４１……逆ＦＦＴ部、４２……出力処理部、５０……記憶手段、６１……音声入力部、６３……音声出力部、６７……入力機器、Ｖin……入力音声信号、Ｖt，Ｖt1，Ｖt2……変換用音声信号、Ｖnew……出力音声信号、ＳＰin……入力スペクトル、ＳＰt……変換用スペクトル、ＳＰnew……新規スペクトル、ＥＶin……スペクトルエンベロープ、Ｄin……入力スペクトルデータ、Ｄt……変換用スペクトルデータ、Ｄnew……新規スペクトルデータ、Ｄev……エンベロープデータ、Ｒ……スペクトル分布領域。

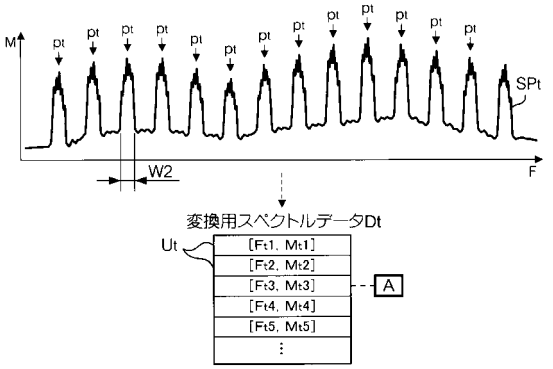
【図１】



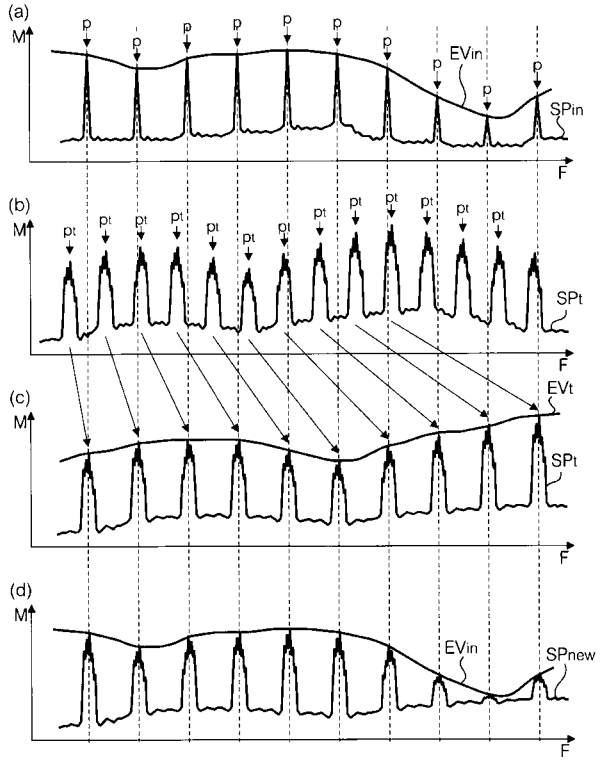
【図２】



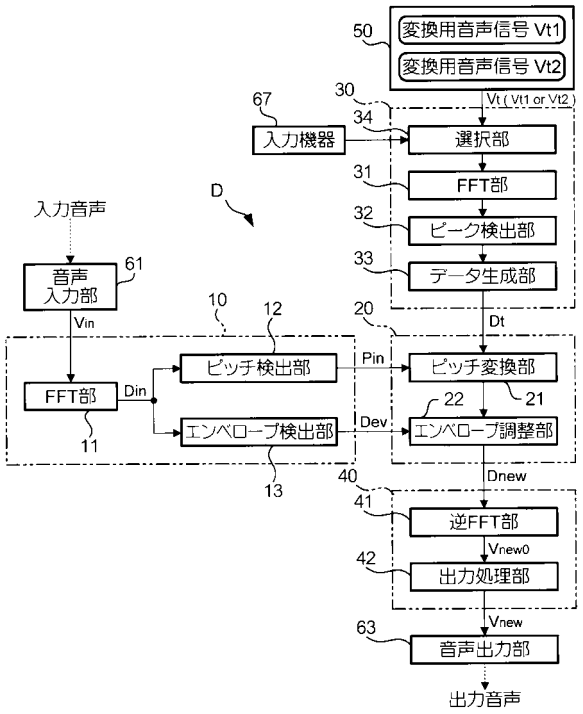
【図３】



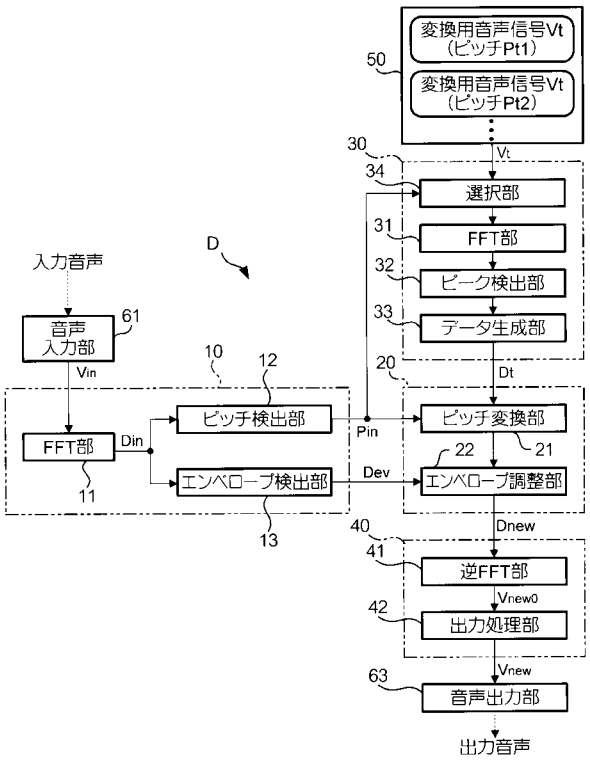
【図４】


(a)
M
p
EVin
SPin
F
(b)
pt
SPt
(c)
EVt
(d)
SPnew

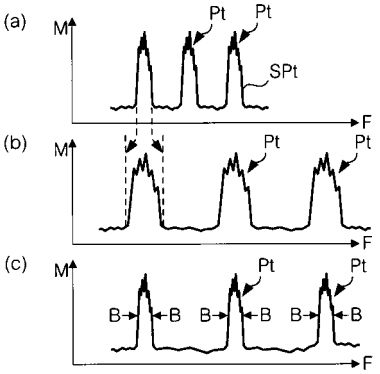
【図５】


50
変換用音声信号 Vt1
変換用音声信号 Vt2
Vt (Vt1 or Vt2)
30
34
選択部
67
入力機器
31
FFT部
32
ピーク検出部
33
データ生成部
D
入力音声
音声入力部
61
Vin
10
12
20
Dt
ピッチ検出部
Pin
ピッチ変換部
FFT部
Din
22
21
Dev
エンベロープ検出部
エンベロープ調整部
11
13
40
Dnew
41
逆FFT部
Vnew0
42
出力処理部
Vnew
63
音声出力部
出力音声

【図６】


変換用音声信号Vt (ピッチPt1)
変換用音声信号Vt (ピッチPt2)
50
Vt
30
34
選択部
31
FFT部
32
ピーク検出部
33
データ生成部
D
入力音声
音声入力部
61
Vin
10
12
20
Dt
ピッチ検出部
Pin
ピッチ変換部
FFT部
Din
22
21
Dev
エンベロープ検出部
エンベロープ調整部
11
13
40
Dnew
41
逆FFT部
Vnew0
42
出力処理部
Vnew
63
音声出力部
出力音声

【図７】


(a)
M
Pt
SPt
F
(b)
(c)
B

【図８】