(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2011-2703
(P2011-2703A)
(43)公開日　平成23年1月6日(2011.1.6)

(51)Int.Cl.	F I	テーマコード（参考）
G10L 19/02 (2006.01)	G10L 19/02 190	

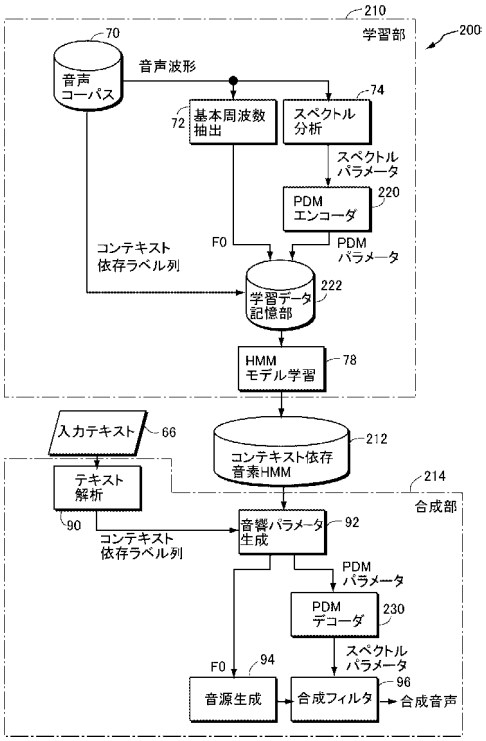
審査請求 未請求　請求項の数 11　ＯＬ　（全 15 頁）

(21)出願番号　特願2009-146502 (P2009-146502)
(22)出願日　平成21年6月19日 (2009.6.19)

(71)出願人　301022471
独立行政法人情報通信研究機構
東京都小金井市貫井北町４－２－１
(74)代理人　100099933
弁理士　清水　敏
(72)発明者　志賀　芳則
東京都小金井市貫井北町４－２－１　独立行政法人情報通信研究機構内

(54)【発明の名称】スペクトル分析装置及びスペクトル演算装置

(57)【要約】
【課題】形状の特徴的な部分を失うことなく、複数のスペクトルの間で形状の補間を行なうことが容易にできるパラメータを出力可能なスペクトル分析装置を提供する。
【解決手段】スペクトル分析装置６０は、音声信号に対するスペクトル分析を行なって、音声のスペクトル包絡を表すスペクトル信号を出力するスペクトル分析部７４と、スペクトル部７４により出力されたスペクトル信号に対し、周波数軸を時間軸とみなしてパルス密度変調（ＰＤＭ）し、各パルスが出力されたときの周波数を含む周波数列を音声信号のスペクトル包絡を表すパラメータとして出力するＰＤＭエンコーダ２２０とを含む。
【選択図】　図６



【特許請求の範囲】
【請求項１】
音声信号に対するスペクトル分析を行なって、音声のスペクトル包絡を表すスペクトル信号を出力するためのスペクトル分析手段と、
前記スペクトル分析手段により出力されたスペクトル信号のパルス密度表現における、各パルス位置に対応する周波数を、前記音声信号のスペクトル包絡を表すパラメータとして出力するためのパラメータ生成手段とを含む、スペクトル分析装置。
【請求項２】
前記スペクトル信号を入力とし、所定のしきい値により量子化を行なうデルタ・シグマ変調に基づいて得られるパルス密度表現の、各パルス位置に対応する周波数を、前記音声信号のスペクトル包絡を表すパラメータとして、前記パラメータ生成手段が出力することを特徴とする請求項１に記載のスペクトル分析装置。
【請求項３】
前記スペクトル分析手段が出力するスペクトル信号が、音声のスペクトル包絡を表すケプストラム係数列であることを特徴とする、請求項１又は請求項２に記載のスペクトル分析装置。
【請求項４】
前記パラメータ生成手段は、
前記スペクトル分析手段が出力するケプストラム係数の内、第０次のケプストラム係数を記憶する第１の記憶手段と、
前記スペクトル分析手段の出力するケプストラム係数の内、第１次以降、所定次数までのケプストラム係数により表されるスペクトル包絡のパルス密度表現における、各パルス位置に対応する周波数を周波数列として記憶する第２の記憶手段とを備え、
前記第１の記憶手段に記憶された前記第０次のケプストラム係数と、前記第２の記憶手段に記憶した周波数列とを、前記パラメータとして出力することを特徴とする、請求項３に記載のスペクトル分析装置。
【請求項５】
前記パラメータ生成手段は、
前記スペクトル分析手段が出力するスペクトル包絡の平均値を記憶する第１の記憶手段と、
前記スペクトル分析手段の出力するスペクトル包絡から、平均値を差し引いたスペクトルのパルス密度表現における、各パルス位置に対応する周波数を周波数列として記憶する第２の記憶手段とを備え、
前記第１の記憶手段に記憶された前記スペクトル包絡平均値と、前記第２の記憶手段に記憶した周波数列とを、前記パラメータとして出力することを特徴とする、請求項１又は請求項２に記載のスペクトル分析装置。
【請求項６】
前記パラメータ生成手段が出力する周波数列に対して、周波数列データを圧縮する処理を行なうパラメータ圧縮処理手段をさらに含み、前記圧縮された周波数列データを、前記スペクトル包絡を表すパラメータの全部または一部として出力する、請求項１～請求項５のいずれかに記載のスペクトル分析装置。
【請求項７】
前記パラメータ圧縮処理手段は、前記パラメータ生成手段が出力する周波数列を、三角級数展開に基づいて圧縮することを特徴とする、請求項６に記載のスペクトル分析装置。
【請求項８】
前記スペクトル分析手段が出力する音声のスペクトル包絡に対して、該スペクトル包絡の傾きを含む大局的な特徴を抑圧又は除去するスペクトル成形手段をさらに備え、該スペクトル成形手段において大局的な特徴が抑圧あるいは除去されたスペクトル包絡を、前記パラメータ生成手段へ入力することを特徴とする、請求項１～請求項７のいずれかに記載のスペクトル分析装置。

【請求項９】
前記スペクトル成形手段は、前記スペクトル分析手段が出力する音声のスペクトル包絡を表すケプストラムに対して、該ケプストラムの低次の係数を減じることによって、前記スペクトル包絡の傾きを含む大局的な特徴を抑圧又は除去することを特徴とする、請求項８に記載のスペクトル分析装置。
【請求項１０】
請求項１～請求項９のいずれかに記載のスペクトル分析装置と、
前記スペクトル分析装置が第１及び第２のスペクトルに対してそれぞれ出力する第１及び第２のパラメータを受け、当該第１及び第２のパラメータ間で所定の補間演算をするための補間手段とを含む、スペクトル演算装置。
【請求項１１】
前記補間手段は、前記第１及び第２のパラメータの内で、対応するパラメータの平均を演算するための平均手段を含む、請求項１０に記載のスペクトル演算装置。
【発明の詳細な説明】
【技術分野】
【０００１】
この発明は音声関連技術に関し、特に、音声を統計的に処理する際のパラメータ化の改善技術に関する。
【背景技術】
【０００２】
音声スペクトルの表現（パラメータ）としてケプストラムがよく用いられる。例えば音声認識に用いられる音響モデルは隠れマルコフモデル（ＨＭＭ）によることが多いが、その学習のための音響パラメータとしてケプストラムが用いられることが多い。ケプストラムを用いた音声のパラメータ化技術はよく研究されており、そのために必要なソフトウェア等も充実している。なお、音声認識等で用いられるケプストラム解析の際には、周波数を聴覚周波数スケールで変換したメル周波数表現が用いられることが多く、それに対するケプストラム解析で得られるケプストラム係数はメルケプストラムと呼ばれる。
【０００３】
ＨＭＭは、音声認識だけではなく音声合成にも用いられる。図１に、ＨＭＭを用いた従来の音声合成システムの概略構成を示す。図１を参照して、ＨＭＭを用いた従来の音声合成システム５０は、コンテキストに依存した音素ＨＭＭ６２を記憶する記憶装置と、この音素ＨＭＭ６２の学習を行なうための学習部６０と、入力されたテキスト６６にしたがって、学習が完了した音素ＨＭＭ６２を使用して音声合成を行なうための合成部６４とを含む。
【０００４】
学習部６０は、多数の発話を記憶した音声コーパス７０と、音声コーパス７０内の各音素の音声波形に対して基本周波数抽出処理を行ない、基本周波数パラメータＦ０を出力するための基本周波数抽出部７２と、音声コーパス７０内の各音素の音声波形に対してスペクトル分析を行ない、音声の対数パワースペクトルの包絡を表すスペクトルパラメータ（ケプストラム係数）を出力するためのスペクトル分析部７４とを含む。さらに学習部６０は、基本周波数抽出部７２からのＦ０パラメータ、スペクトル分析部７４からのスペクトルパラメータ、及び音声コーパス７０の各音素のコンテキストに依存した音素ラベル（以下このラベルを「コンテキスト依存ラベル」と呼ぶ。）を含む学習データを記憶するための学習データ記憶部７６と、学習データ記憶部７６に記憶された学習データに対する統計処理を行なって、音素ＨＭＭ６２の各コンテキスト依存音素モデルの確率密度関数等のパラメータの計算を行なうためのＨＭＭモデル学習部７８とを含む。コンテキストとしては、当該音素を含む文節のアクセント型、当該音素を含む単語の品詞、文の長さ、文内での当該音素の位置等が含まれる。
【０００５】
合成部６４は、入力されたテキスト６６に対してテキスト解析を行ない、テキスト６６

に対する音素列を示す音素ラベル列であって、テキスト６６内で各音素のおかれたコンテキストに応じた音素ラベル列（「コンテキスト依存ラベル列」と呼ぶ。）を出力するためのテキスト解析部９０と、テキスト解析部９０からのコンテキスト依存ラベル列に応じ、音素ＨＭＭ６２内の音素ＨＭＭを連結し、与えられたコンテキスト依存ラベル列に対して最も尤度が高くなる音響パラメータ（Ｆ０及びスペクトルパラメータ）列をこれらＨＭＭ列から推定するための音響パラメータ生成部９２と、音響パラメータ生成部９２から出力されるＦ０にしたがって音源生成を行なう音源生成部９４と、音源生成部９４からの音源波形に対し、音響パラメータ生成部９２から出力されるスペクトルパラメータにしたがって変調することにより、合成音声信号を出力するための合成フィルタ９６とを含む。

【０００６】

このような音声合成システム５０では、多数の音声により音素ＨＭＭ６２の学習を行なうことが必要である。この学習時には、結局のところ、特定音素コンテキストの音声スペクトルの、全サンプルにわたる平均が計算される。しかしそのような処理をケプストラムで行なうと、スペクトルの山（フォルマント）の位置（周波数）が異なる複数の音声スペクトルがケプストラム領域で平均されることになる。この場合、次のような問題が生じる。

【０００７】

図２を参照して、２つのスペクトル１１０及び１１２を考える。これらはそれぞれフォルマントに対応するピークを持つが、その周波数軸上の位置は互いにずれている。これらを単純に平均すると、スペクトル１１６が得られる。スペクトル１１６では、スペクトル１１０及びスペクトル１１２で明確に存在するピークがなまってしまっている。このスペクトルで仮に音声合成を行なうと、音質が低くなることは明らかである。本来は、スペクトル１１４のように、ピークが明確に生じるように両者の平均を算出すべきである。

【先行技術文献】

【非特許文献】

【０００８】

【非特許文献１】大室　仲他、「積分スペクトル逆関数（ＩＦＩＳ）」とその応用に関する検討」、信学技報、ＳＰ８９－７２、ｐ．２３－３０、１９８９年

【発明の概要】

【発明が解決しようとする課題】

【０００９】

こうした問題を解決する１つの手法が非特許文献１に開示されている。非特許文献１は、スペクトルを補間するために、「積分スペクトル逆関数（ＩＦＩＳ）」と呼ばれるパラメータを使用することを提案している。平均は補間の一部と考えることができるため、非特許文献１に提案されたパラメータを上記した処理に適用できる可能性がある。

【００１０】

図３を参照して、この手法によれば、２つのスペクトル１３０及び１３２の平均を算出する時には、まずそれらのグラフを全体にわたり積分する。積分の結果得られた曲線１４０及び１４２において、元のスペクトル１３０及び１３２のピークＡ及びＢに対応する周波数の値を求め（Ａ′及びＢ′）、これらの周波数軸上での平均Ｃ′を算出する。この周波数Ｃ′が、スペクトル１３０及び１３２を平均したスペクトルのピークＣの中心周波数位置となる。非特許文献１によれば、さらに、スペクトル１３０及び１３２をこの手法を使用して平均する場合、結果として得られるスペクトルの各周波数における高さは、これらスペクトル１３０及び１３２のその周波数における高さの調和平均となる。

【００１１】

すなわち、非特許文献１による手法は、「２つのスペクトルをそれぞれ周波数０から積分し、積分値が等しくなった２点の振幅の調和平均をとる」手法であるということができる。

【００１２】

この手法によってスペクトル１３０及び１３２を平均して得られたスペクトルの例を図

４に示す。図４において、スペクトル１５０のピークはスペクトル１３０及び１３２のピークの中間位置となり、そのピークの高さも両者のピークの中間となっている。そのため、図２に示すような例と比較すると、スペクトルのピークがなまるおそれは小さい。なお、図４に示すスペクトル１３０及び１３２は試験のためのデータであるため、通常のスペクトルの曲線とは異なっている。

【００１３】

確かに非特許文献１による手法によれば、２つのスペクトルを「平均」してもピークがなまってしまうことはなく、ＨＭＭの学習には好ましいと思われる。しかしこの非特許文献１の開示では、まず、ＩＦＩＳの数値計算方法が明らかにされていない。スペクトルを実際に積分してその逆関数をとる場合には、多くの計算量を必要とする問題がある。また、得られるＩＦＩＳのパラメータは、スペクトルと同程度の次元数をもつ（１２８～１０２４次元程度）。こうした高い次元数の音響パラメータをＨＭＭの学習に用いると莫大な処理量（処理時間）を必要とし問題となる。さらに、ＩＦＩＳはスペクトルの単純な積分に基づくため、対象となるスペクトルが例えば大きな傾斜をもっている場合に、パワーの小さな周波数領域において周波数解像度が悪くなる問題がある。

【００１４】

したがって、図４に示されるような結果を効率よく計算し、ＨＭＭ学習に適したパラメータとして得ることができ、かつ全周波数帯域にわたって十分な周波数解像度を得ることができるような音声のパラメータ化技術が必要である。

【００１５】

それゆえに本発明の目的は、複数のスペクトルについて、形状の特徴的な部分を失うことなく、複数のスペクトルの間で、形状の補間を行なうことが容易にできるパラメータを出力可能なスペクトル分析装置を提供することである。

【００１６】

本発明の他の目的は、複数のスペクトルについて、形状の特徴的な部分を失うことなく、複数のスペクトルの間で形状の補間を行なうことが容易にできるスペクトル演算装置を提供することである。

【課題を解決するための手段】

【００１７】

本発明の第１の局面に係るスペクトル分析装置は、音声信号に対するスペクトル分析を行なって、音声のスペクトル包絡を表すスペクトル信号を出力するためのスペクトル分析手段と、スペクトル分析手段により出力されたスペクトル信号のパルス密度表現における、各パルス位置に対応する周波数を、音声信号のスペクトル包絡を表すパラメータとして出力するためのパラメータ生成手段とを含む。

【００１８】

スペクトル分析手段は、入力された音声信号に対するスペクトル分析を行ない、スペクトル信号を出力する。このスペクトル信号は音声のスペクトル包絡を表す。パラメータ生成手段は、スペクトル分析手段により出力されたスペクトル信号のパルス密度表現における、各パルス位置に対応する周波数を、音声信号のスペクトル包絡を表すパラメータとして出力する。この出力が、音声信号のスペクトル包絡を表すパラメータとして使用される。

【００１９】

音声信号のスペクトル包絡を、パルス密度表現における、各パルス位置に対応する周波数の形でパラメータとして表す。スペクトル包絡の特徴を一連の周波数列で表すため、形状が類似しているが特徴となる部分の周波数位置が異なるような複数のスペクトルについて、特徴となる部分の対応関係を的確に表すことができる。その結果、複数のスペクトルについて、形状の特徴的な部分を失うことなく、形状の補間を行なうことが容易にできるパラメータを出力可能なスペクトル分析装置を提供できる。

【００２０】

好ましくは、パラメータ生成手段は、スペクトル信号を入力とし、所定のしきい値によ

り量子化を行なうデルタ・シグマ変調に基づいて得られるパルス密度表現の、各パルス位置に対応する周波数を、音声信号のスペクトル包絡を表すパラメータとして出力する。

【００２１】

時間領域の信号に対してよく利用されるデルタ・シグマ変調を利用して、スペクトル信号を周波数データに変換することができる。

【００２２】

より好ましくは、スペクトル分析手段が出力するスペクトル信号が、音声のスペクトル包絡を表すケプストラム係数列である。

【００２３】

音声信号の解析にはケプストラム解析が多用されており、ケプストラム解析によって得られたスペクトル情報を処理することで、既存の手段を有効に利用しながら、スペクトル包絡を周波数列で表す新たなパラメータにより、音声信号の特徴を表すことができる。

【００２４】

さらに好ましくは、パラメータ生成手段は、スペクトル分析手段が出力するケプストラム係数の内、第０次のケプストラム係数を記憶する第１の記憶手段と、スペクトル分析手段の出力するケプストラム係数の内、第１次以降、所定次数までのケプストラム係数により表されるスペクトル包絡のパルス密度表現における、各パルス位置に対応する周波数を周波数列として記憶する第２の記憶手段とを備え、第１の記憶手段に記憶された第０次のケプストラム係数と、第２の記憶手段に記憶した周波数列とを、パラメータとして出力する。

【００２５】

第０次のケプストラム係数は、スペクトルの平均値を表す。平均値を除いてパルス密度変調することにより、スペクトルの平均値は０となり、パラメータ化する際の情報量と処理量とを削減できる。

【００２６】

パラメータ生成手段は、スペクトル分析手段が出力するスペクトル包絡の平均値を記憶する第１の記憶手段と、スペクトル分析手段の出力するスペクトル包絡から、平均値を差し引いたスペクトルのパルス密度表現における、各パルス位置に対応する周波数を周波数列として記憶する第２の記憶手段とを備え、第１の記憶手段に記憶されたスペクトル包絡平均値と、第２の記憶手段に記憶した周波数列とを、パラメータとして出力してもよい。

【００２７】

好ましくは、スペクトル分析装置は、パラメータ生成手段が出力する周波数列に対して、周波数列データを圧縮する処理を行なうパラメータ圧縮処理手段をさらに含み、圧縮された周波数列データを、スペクトル包絡を表すパラメータの全部または一部として出力する。

【００２８】

さらに好ましくは、パラメータ圧縮処理手段は、パラメータ生成手段が出力する周波数列を、三角級数展開に基づいて圧縮する。

【００２９】

スペクトル分析装置は、スペクトル分析手段が出力する音声のスペクトル包絡に対して、該スペクトル包絡の傾きを含む大局的な特徴を抑圧又は除去するスペクトル成形手段をさらに備え、該スペクトル成形手段において大局的な特徴が抑圧あるいは除去されたスペクトル包絡を、パラメータ生成手段へ入力するようにしてもよい。

【００３０】

スペクトル成形手段は、スペクトル分析手段が出力する音声のスペクトル包絡を表すケプストラムに対して、該ケプストラムの低次の係数を減じることによって、スペクトル包絡の傾きを含む大局的な特徴を抑圧又は除去してもよい。

【００３１】

本発明の第２の局面に係るスペクトル演算装置は、上記したいずれかのスペクトル分析装置と、スペクトル分析装置が第１及び第２のスペクトルに対してそれぞれ出力する第１

及び第２のパラメータを受け、当該第１及び第２のパラメータ間で所定の補間演算をするための補間手段とを含む。

【００３２】

スペクトル分析装置は、複数のスペクトルについて、パルス密度表現における各パルス位置に対応する周波数を、音声信号のスペクトル包絡を表すパラメータとして出力する。このパラメータは、スペクトル包絡の特徴的な部分を失うことなく補間ができる性質を持つ。補間手段は、スペクトル分析装置によって第１及び第２のスペクトルから得られた第１及び第２のパラメータの間で所定の補間演算を行なう。したがって、第１及び第２のスペクトルについて、特徴部分を失うことなく補間処理を行なうことができる。

【００３３】

好ましくは、補間手段は、第１及び第２のパラメータの内で、対応するパラメータの平均を演算するための平均手段を含む。

【００３４】

補間演算として平均が計算される。複数のスペクトルの平均を演算する際に、それらスペクトルの特徴部分を失うことなく、平均のスペクトルを得ることができる。

【発明の効果】

【００３５】

以上のように本発明によれば、スペクトル包絡の特徴を一連の周波数列で表すので、形状が類似しているが特徴となる部分の周波数位置が異なるような複数のスペクトルについて、特徴となる部分の対応関係を的確に表すことができる。その結果、複数のスペクトルについて、形状の特徴的な部分を失うことなく、形状の補間を行なうことが容易にできるようなパラメータを出力可能なスペクトル分析装置を提供できる。

【図面の簡単な説明】

【００３６】

【図１】従来の音声合成システム５０のブロック図である。

【図２】従来の手法でスペクトルを平均するときの問題点を示すスペクトルのグラフである。

【図３】非特許文献１に提案されたパラメータ化手法を説明するための図である。

【図４】非特許文献１により提案されたパラメータ化手法によって平均されたスペクトルを説明するためのグラフである。

【図５】本発明の実施の形態で採用するパルス密度変調（ＰＤＭ）によるスペクトルの平均の算出方法を説明するための図である。

【図６】本発明の実施の形態に係る音声合成システム２００のブロック図である。

【図７】図６に示すＰＤＭエンコーダ２２０のブロック図である。

【図８】図７に示すデルタ・シグマ変調部２４６のブロック図である。

【図９】図６の圧縮処理部２５０で行なう正弦級数展開を説明するためのグラフである。

【図１０】スペクトルと、本発明の実施の形態によってこのスペクトルから得られたパルス列とを対比して示す図である。

【図１１】／ｒ／から／ｌ／への過渡部の連続スペクトルを示す図である。

【図１２】図１１に示す連続スペクトルの真の平均であるスペクトルと、連続スペクトルをケプストラム及び本発明の実施の形態に係るＰＤＭパラメータを用いてそれぞれ平均化して得られるスペクトルとを対比して示すグラフである。

【発明を実施するための形態】

【００３７】

本明細書及び図面では、同一の部品には同一の参照番号を付してある。それらの名称及び機能もそれぞれ同一である。したがってそれらについての詳細な説明は繰返さない。

【００３８】

［基本的な考え方］

図５を参照して、本発明の実施の形態におけるスペクトルのパラメータ化方法、及びそ

のパラメータを用いたスペクトルの平均の計算方法について説明する。２つのスペクトル１６０及び１６２の平均を求める場合を考える。両者のピークは図からも明らかなように周波数軸上で異なった位置にある。

【００３９】

本実施の形態では、音声スペクトルの包絡をまずリフタリングして対極的なスペクトルの特性を抑制した後、パルス密度変調（ＰＤＭ）を行なって、スペクトル１６０及び１６２の振幅（又はパワー）をパルス密度を表すパルス列１７０及び１７２にそれぞれ変換する。各スペクトルを予め正規化しておくことで、１つのスペクトルについて出力されるパルス数が一定となるようにし、各パルスが出力されたときの周波数を記憶しておく。このためにはパルスが出力されたときの周波数のみを記憶しておけばよい。

【００４０】

パルス列１７０及び１７２の各パルスの間には１対１の対応関係が付く。対応するパルス対の周波数を全てのパルス対について平均することで、新たなパルス列１７４が得られる。このパルス列１７４をＰＤＭデコードして、スペクトル１６０及び１６２を平均した新たなスペクトル１８０が得られる。

【００４１】

［構成］

図６は、本発明の実施の形態に係る音声合成システム２００のブロック図である。図６を参照して、ＨＭＭを用いた、本発明の実施の形態に係る音声合成システム２００は、コンテキストに依存した音素ＨＭＭであって、かつ音響パラメータとして図１に示す音素ＨＭＭ６２と異なり、上記したパルス列の周波数を使用した音素ＨＭＭ２１２を記憶する記憶装置と、この音素ＨＭＭ２１２の学習を行なうための学習部２１０と、入力されたテキスト６６にしたがって、学習が完了した音素ＨＭＭ２１２を使用して音声合成を行なうための合成部２１４とを含む。

【００４２】

学習部２１０は、図１に示す従来の学習部６０と同様の構成に加え、図１のスペクトル分析部７４の出力を受けるように接続され、スペクトル分析部７４から出力されるスペクトルパラメータに対して、ＰＤＭエンコードを行なって、スペクトルを表すパルス列の周波数情報（これらを「ＰＤＭパラメータ」又は「ＰＤＭケプストラム」と呼ぶ。）に変換して出力するためのＰＤＭエンコーダ２２０をさらに含む点と、図１の学習データ記憶部７６に代えて、基本周波数抽出部７２から出力されるある音声波形のＦ０パラメータ、音声コーパス７０から与えられる、対応する音声波形に付与されたコンテキスト依存ラベル列、及びＰＤＭエンコーダ２２０から与えられるＰＤＭパラメータを学習データとしてまとめて記憶するための学習データ記憶部２２２を含む点とで異なる。

【００４３】

学習部２１０は図１の学習部６０と同様のＨＭＭモデル学習部７８を含んでいる。ＨＭＭモデル学習部７８自体は、図１に示すものと全く同じ機能を持つが、スペクトルに関する音響パラメータとして通常のケプストラムではなくＰＤＭパラメータを含む学習データを用いて音素ＨＭＭ２１２の学習を行なう。そのため、音素ＨＭＭ２１２の内部パラメータは図１に示す音素ＨＭＭ６２の内部パラメータとは異なったものとなる。特に、音素ＨＭＭ２１２は、その出力が、ケプストラムではなくＰＤＭパラメータである点で図１の音素ＨＭＭ６２と異なる。

【００４４】

合成部２１４も、図１に示す合成部６４と同様の構成を持つが、音素ＨＭＭ２１２から出力されるスペクトルに関する音響パラメータがケプストラムでなくＰＤＭパラメータであるため、以下に述べる点で合成部６４と異なっている。すなわち、合成部２１４は、図１に示す合成部６４の構成に加え、音響パラメータ生成部９２から出力されるＰＤＭパラメータをデコードし、スペクトルパラメータに変換して出力し合成フィルタ９６に与えるためのＰＤＭデコーダ２３０をさらに含む。ＰＤＭデコーダ２３０は、ＰＤＭデコーダ２３０から与えられるＰＤＭパラメータ（パルスの周波数データ）にしたがってパルスを発

生した後、そのパルス列をローパスフィルタに通すという簡単な構成で実現できる。なお、ここでは、音声のスペクトルを音声信号とみなし、周波数と時間とを対応付けてパルス列を発生させるようにすればよい。

【００４５】

図７は、図６に示すＰＤＭエンコーダ２２０のブロック図である。図７を参照して、図６に示すＰＤＭエンコーダ２２０は、スペクトル分析部７４から出力されるスペクトルパラメータの平均値を算出して平均値を示す信号を出力するための平均値算出回路２４０と、平均値算出回路２４０から出力される平均値信号を記憶するための平均記憶回路２４２と、スペクトル分析部７４から出力されるスペクトルパラメータから平均記憶回路２４２に記憶されている平均値を減算するための減算回路２４４とを含む。実際には、スペクトル分析部７４から出力されるケプストラム係数の内、第０次の係数C_0がこの平均値に相当するため、平均値算出回路２４０が第０次の係数C_0のみを抽出する処理を行ない、減算回路２４４がケプストラム係数の内、第１次以降の係数のみを抽出する処理を行なうようにすることでこの処理を実現できる。平均記憶回路２４２は第０次のケプストラム係数を記憶し、ＰＤＭパラメータの一部として学習データ記憶部２２２に出力する。

【００４６】

ＰＤＭエンコーダ２２０はさらに、減算回路２４４から出力される、平均値を減算した後の音響パラメータを入力とし、スペクトルの周波数を時間軸とみなしてデルタ・シグマ変調を行ない、スペクトルを表すパルス列に変換するためのデルタ・シグマ変調部２４６と、処理するスペクトルごとに、デルタ・シグマ変調部２４６からパルスが出力されたときのスペクトルの周波数を記憶するためのパルス列記憶部２４８と、パルス列記憶部２４８に記憶されたパルス列の発生位置（すなわち周波数）情報に対して正弦級数展開を行なってデータを圧縮し、学習データ記憶部２２２に記憶させるための圧縮処理部２５０とを含む。

【００４７】

図８は、図７に示すデルタ・シグマ変調部２４６のブロック図である。図８を参照して、デルタ・シグマ変調部２４６は、本実施の形態では一次のデルタ・シグマ変調を行なうものであって、減算回路２４４の出力を受ける積分器２６２と、積分器２６２の出力が一定のしきい値を越えると＋１のパルスを出力する量子化器２６６と、量子化器２６６の出力を積分器２６２の入力にフィードバックするフィードバック部２６８とを含む。パルス列記憶部２４８は、スペクトル信号と量子化器２６６の出力とを受けており、量子化器２６６がパルスを出力したときのスペクトルの周波数を記憶する。積分器２６２によってスペクトル波形が積分され、あるしきい値を越えたところでパルスが出力される。したがって、図５に示すように、スペクトルのピーク部分では出力されるパルスの密度が高く、スペクトルの値が小さくなると出力されるパルスの密度は低くなる。

【００４８】

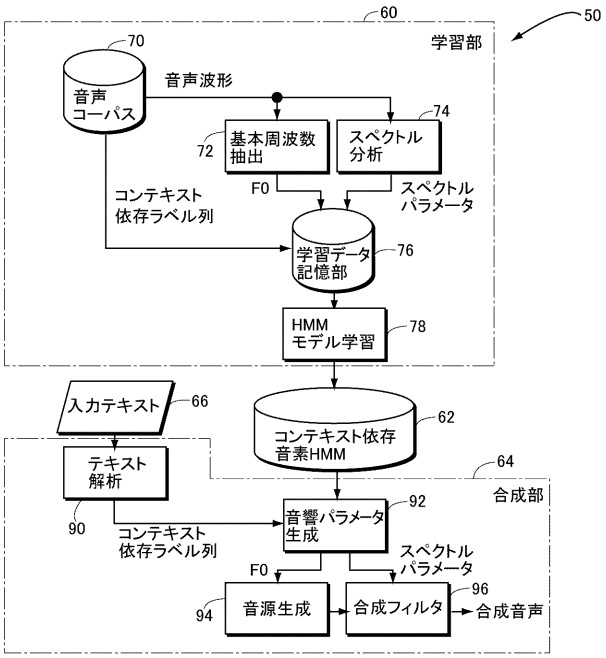
図７に示す圧縮処理部２５０は、パルス列を表すデータ（周波数列）を正弦級数展開によって圧縮する。図９を参照して、横軸に正規化されたスペクトルの積分値、縦軸に周波数軸をとって各パルスが出力された点をプロットし、それらを結んだ曲線２９０を考える。この曲線の内、最初の点と最後の点とを結ぶ線分２９２を考えると、曲線２９０は線分２９２を中心としてその上方と下方とに分けられる。図９に示す例は図５に示すスペクトル１６０に対応するものであり、曲線２９０が線分２９２の上方に存在する領域と、線分２９２の下方に存在する領域との２つの領域に分けられる。線分２９２を曲線２９０から減算すると、ちょうど正弦級数に似た曲線が得られる。従って、曲線２９０は正弦級数展開の低次の項（十数～数十次程度）で近似でき、曲線２９０を構成する各点を示すデータを圧縮することができる。曲線２９０が線分２９２によって上下の複数箇所に分けられた場合もこれと同様である。この圧縮処理部２５０によって、前記周波数列データ（百数十～千次程度）を少数次元のデータで表すことができ、その結果、ＨＭＭ学習における処理量が膨大になるのを防ぎ、リーゾナブルな計算時間でＨＭＭ学習を完了することができる。

【００４９】
［動作］
音声合成システム２００の動作には２つのフェーズがある。第１フェーズは音素ＨＭＭ２１２の学習である。第２フェーズは音素ＨＭＭ２１２を使用して、テキスト６６にしたがった音声を合成する処理である。以下、これらフェーズにおける音声合成システム２００の動作を説明する。
【００５０】
－学習－
音素ＨＭＭ２１２の学習時、音声合成システム２００は以下のように動作する。音声コーパス７０内の発話内の音声の各フレームには、予めコンテキスト依存ラベルが付されている。各フレームの音声波形データは基本周波数抽出部７２とスペクトル分析部７４とにそれぞれ与えられる。基本周波数抽出部７２は、与えられた音声からＦ０パラメータを算出し、学習データ記憶部２２２に与える。スペクトル分析部７４は、音声のスペクトルパラメータを算出し、ＰＤＭエンコーダ２２０に与える。
【００５１】
図７を参照して、平均値算出回路２４０は、与えられた対数パワースペクトルの平均値（第０次のケプストラム係数）を算出し、その値を平均記憶回路２４２が記憶する。減算回路２４４は、対数パワースペクトルからその平均値を減算し、周波数の低い方から始めてデルタ・シグマ変調部２４６に与える。
【００５２】
図８を参照して、積分器２６２は、与えられるパワースペクトルを積分する。積分器２６２の出力は量子化器２６６に与えられる。量子化器２６６は、積分器２６２の出力がしきい値より高くなるとパルスを出力する。このパルスはフィードバック部２６４により積分器２６２の入力にフィードバックされる。その結果、積分器２６２の出力からは、しきい値に相当する値が減算される。このようにしてデルタ・シグマ変調部２４６は、スペクトル波形の低周波数領域から高周波数領域に向かって波形を積分し、積分値がしきい値を越えるとパルスを出力する。パルスが出力されると、積分値からしきい値に相当する値が減算されるので、結果としてスペクトル波形を低周波数側から積分していったときに、その積分値がしきい値となった時点でパルスが出力される。パルス列記憶部２４８はこのときのスペクトルの周波数を記憶する。こうして、パルス列記憶部２４８はスペクトルについて出力されたパルス列を、それらが出力されたときの周波数列の形で記憶する。圧縮処理部２５０は、正弦級数展開により、パルス列記憶部２４８に記憶された周波数列データ（百数十～千次程度）を少数次元のデータに変換する。この圧縮処理の結果、ＨＭＭ学習における処理量が膨大になるのを防ぎ、リーゾナブルな計算時間でＨＭＭ学習を完了することができる。
【００５３】
ＰＤＭエンコーダ２２０は、パルス列記憶部２４８に記憶され、圧縮処理部２５０により圧縮された周波数列データをＰＤＭパラメータとして学習データ記憶部２２２に与える。学習データ記憶部２２２は、各フレームごとに、基本周波数抽出部７２からのＦ０パラメータと、音声コーパス７０からコンテキスト依存ラベルとをそれぞれ受け、さらにＰＤＭエンコーダ２２０からＰＤＭパラメータとを受けてこれらを一まとめの学習データとして保存する。
【００５４】
このようにして、音声コーパス７０に保存されている発話データの全フレームに対して学習データが作成されると、ＨＭＭモデル学習部７８はこの学習データを使用して、音素ＨＭＭ２１２の学習を行なう。
【００５５】
図１０に、ある音声の対数パワースペクトルのグラフと、このスペクトルに対してＰＤＭを行なったときに得られるパルス列との例を示す。図１０に示すように、スペクトルの対数パワーが大きいときにはパルス密度は高く、小さいときにはパルス密度は低くなる。

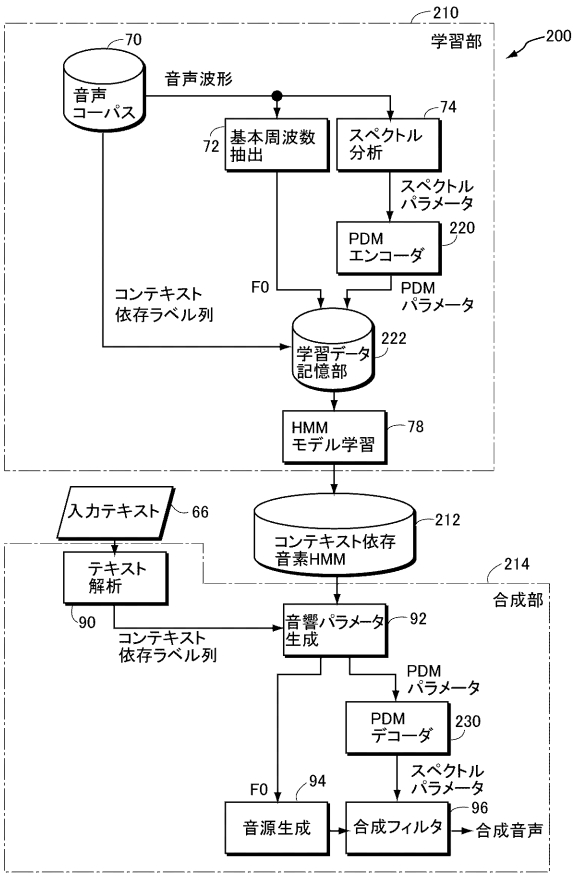
【００５６】
－音声合成－
音声合成時には、音声合成システム２００は以下の様に動作する。音声合成の対象となるテキスト６６が与えられると、合成部２１４のテキスト解析部９０は、公知のテキスト解析処理を行ない、合成すべき音素列を含むコンテキスト依存ラベル列を生成して音響パラメータ生成部９２に与える。
【００５７】
音響パラメータ生成部９２は、与えられたコンテキスト依存ラベル列にしたがって音素ＨＭＭ２１２内のコンテキスト依存ＨＭＭを連結する。音響パラメータ生成部９２はさらに、連結後のコンテキスト依存ＨＭＭに基づいて、最も尤度の高い音響パラメータ列（Ｆ０パラメータ列及びＰＤＭパラメータ列）を生成する。このとき、音素ＨＭＭ２１２が音響パラメータとしてＰＤＭパラメータを用いた学習を行なっているため、音響パラメータ生成部９２の出力する音響パラメータ列は、従来の技術で示したようなケプストラム列ではなく、ＰＤＭパラメータ列となる。
【００５８】
ＰＤＭデコーダ２３０は、音響パラメータ生成部９２の出力するＰＤＭパラメータ列に対するデコードを行なってスペクトルパラメータ列に変換し、合成フィルタ９６に与える。
【００５９】
音源生成部９４は音響パラメータ生成部９２から与えられるＦ０パラメータ列にしたがって音源信号を生成し、合成フィルタ９６は、この音源信号に対してＰＤＭデコーダ２３０から与えられるスペクトルパラメータに依存した特性のフィルタ処理を行ない、合成音声信号を出力する。この合成音声信号をアナログ変換・増幅してスピーカに与えることにより、合成音声の発声が行なわれる。
【００６０】
以上のように本実施の形態によれば、ＨＭＭのモデル学習のときに行なわれるスペクトルの平均処理において、スペクトルの平坦化を緩和できる。したがって、合成音声の音質が改善するという効果を得ることができる。
【００６１】
例えば、図１１に示したような、／ｒ／から／ｌ／への過渡部の連続スペクトルを平均化する場合を考える。図１２を参照して、従来のようにケプストラム領域でスペクトルを平均化して得られる波形３２２を、厳密に周波数の対応関係を付けて平均した波形３２０と比較すると、波形３２０では第２フォルマント及び第３フォルマントが区別できているのに対し、波形３２２ではこれらが平坦化され区別できなくなっている。一方、本実施の形態の方法にしたがって得られた波形３２４では、これらフォルマントがきちんと区別され、厳密に計算した波形の特徴をよく残すことができている。
【００６２】
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での全ての変更を含む。
【符号の説明】
【００６３】
５０，２００　音声合成システム
６０，２１０　学習部
６２，２１２　コンテキスト依存音素ＨＭＭ
６４，２１４　合成部
６６　テキスト
７０　音声コーパス
７２　基本周波数抽出部

７４　スペクトル分析部
７６，２２２　学習データ記憶部
７８　ＨＭＭモデル学習部
９０　テキスト解析部
９２　音響パラメータ生成部
９４　音源生成部
９６　合成フィルタ
１７０，１７２，１７４　パルス列
２２０　ＰＤＭエンコーダ
２３０　ＰＤＭデコーダ
２４６　デルタ・シグマ変調部

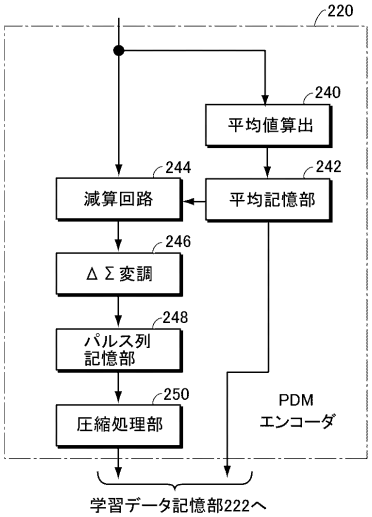
【図１】



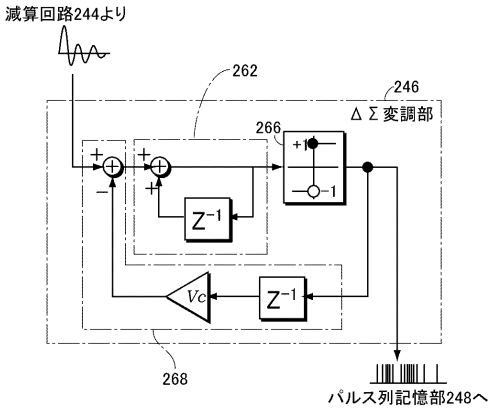
【図６】



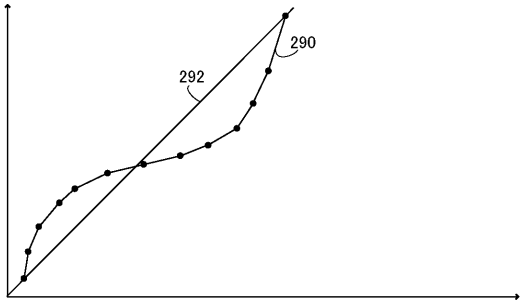
【図７】



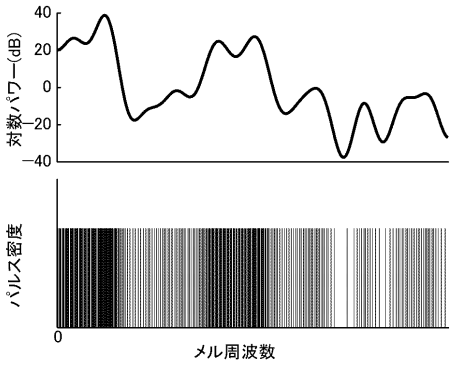
【図８】



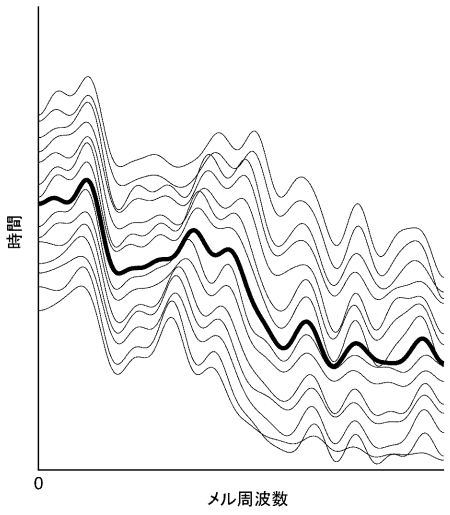
【図９】



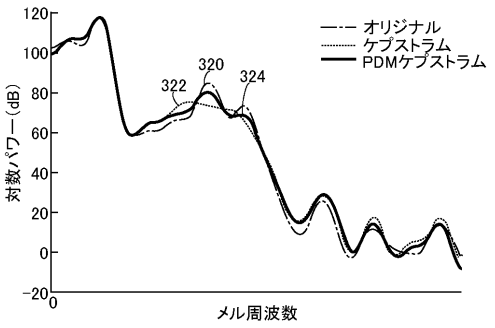
【図１０】



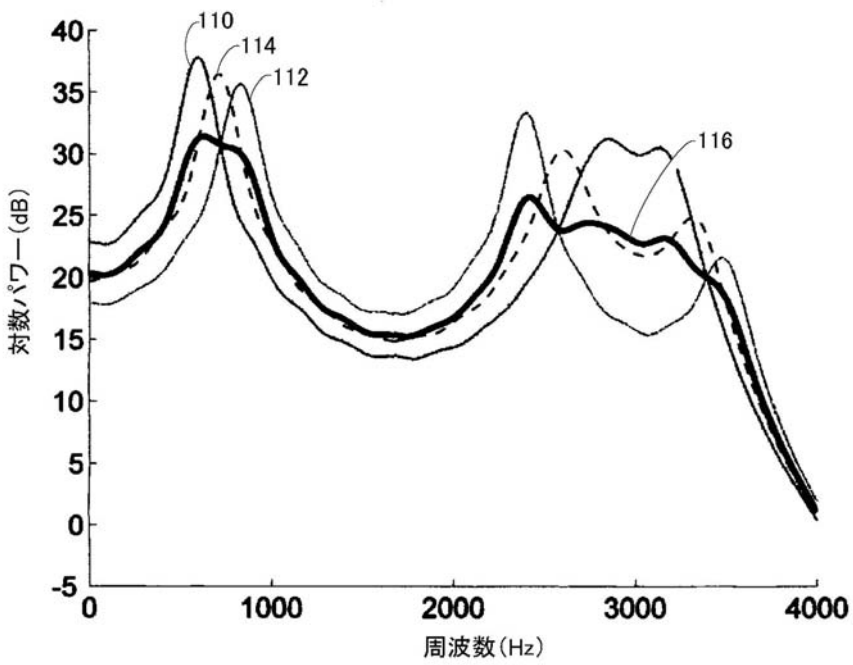
【図１１】



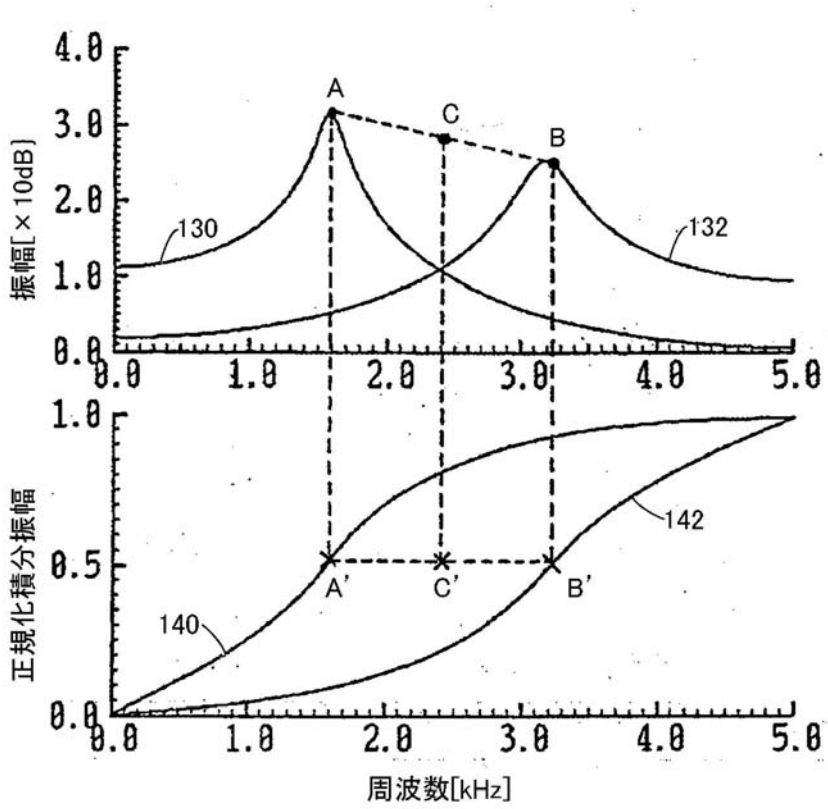
【図１２】



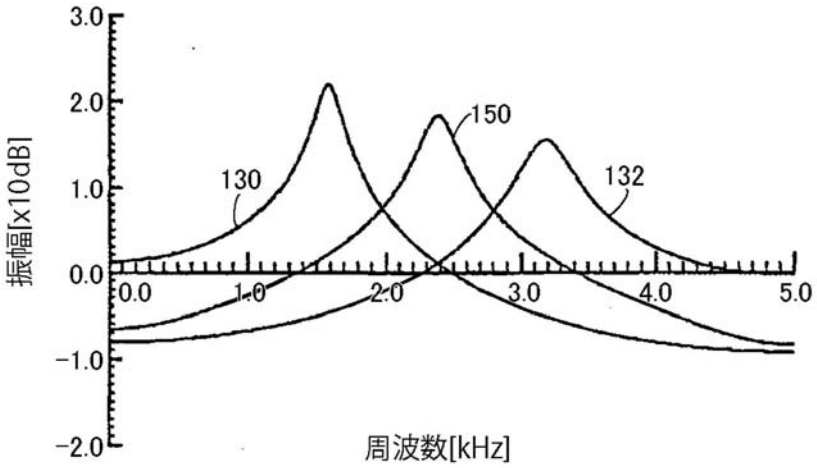
【図２】


110
114
112
116
対数パワー(dB)
周波数(Hz)

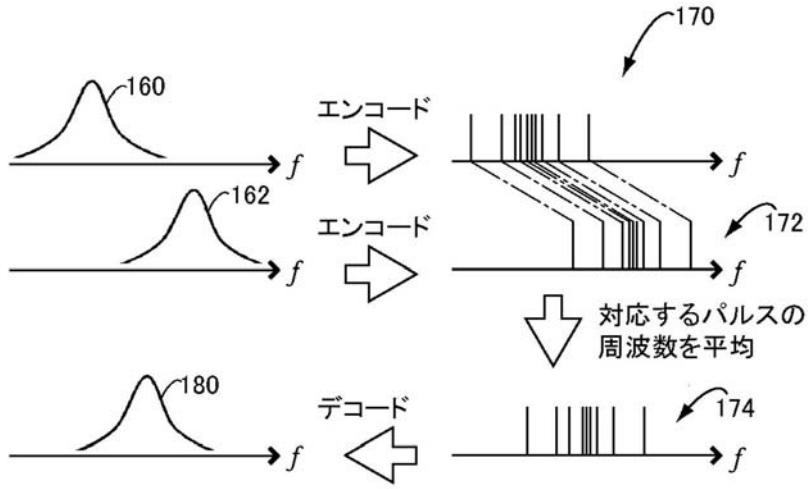
【図３】


A
C
B
130
132
振幅[×10dB]
A'
C'
B'
140
142
正規化積分振幅
周波数[kHz]

【図４】


3.0
2.0
1.0
0.0
−1.0
−2.0
振幅[x10dB]
130
150
132
0.0
1.0
2.0
3.0
4.0
5.0
周波数[kHz]

【図５】


170
160
エンコード
f
162
エンコード
172
対応するパルスの
周波数を平均
180
デコード
174