

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5252452号  
(P5252452)

(45) 発行日 平成25年7月31日(2013.7.31)

(24) 登録日 平成25年4月26日(2013.4.26)

(51) Int.Cl. F I  
**G 1 O L 25/18 (2013.01)** G 1 O L 11/00 1 O 1 E  
**G 1 O L 13/06 (2013.01)** G 1 O L 13/06 2 3 O Z

請求項の数 11 (全 16 頁)

|           |                              |           |                                  |
|-----------|------------------------------|-----------|----------------------------------|
| (21) 出願番号 | 特願2009-146502 (P2009-146502) | (73) 特許権者 | 301022471                        |
| (22) 出願日  | 平成21年6月19日(2009.6.19)        |           | 独立行政法人情報通信研究機構                   |
| (65) 公開番号 | 特開2011-2703 (P2011-2703A)    |           | 東京都小金井市貫井北町4-2-1                 |
| (43) 公開日  | 平成23年1月6日(2011.1.6)          | (74) 代理人  | 100099933                        |
| 審査請求日     | 平成24年5月9日(2012.5.9)          |           | 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 |
|           |                              | 審査官       | 安田 勇太                            |
|           |                              | (56) 参考文献 | 特開平5-250677 (JP, A)              |
|           |                              |           | 特開平2-107231 (JP, A)              |

最終頁に続く

(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) 【特許請求の範囲】

【請求項1】

スペクトル信号の周波数軸上における強度に基づく分布を、周波数軸上で可変の間隔の周波数位置で発生されるパルスの密度に変換して表現するパルス密度表現を用いるスペクトル分析装置であって、

\_\_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前記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】

10

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】

20

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。

30

【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、

40

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のパラメータの中で、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】

50

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。

10

【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。

20

30

【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。

40

【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。

50

## 【 0 0 0 7 】

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】

## 【 0 0 0 8 】

【非特許文献1】大室 伸他、「積分スペクトル逆関数 ( I F I S ) 」とその応用に関する検討」、信学技報、S P 8 9 - 7 2、p . 2 3 3 0、1 9 8 9 年

## 【発明の概要】

## 【発明が解決しようとする課題】

## 【 0 0 0 9 】

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 ( I F I S ) 」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。

## 【 0 0 1 0 】

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。

## 【 0 0 1 1 】

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。

## 【 0 0 1 2 】

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のピークの間位置となり、そのピークの高さも両者のピークの間となっている。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。

## 【 0 0 1 3 】

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、I F I S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I F I Sのパラメータは、スペクトルと同程度の次元数をもつ ( 1 2 8 ~ 1 0 2 4 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、I F I S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。

## 【 0 0 1 4 】

10

20

30

40

50

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

10

20

30

40

50

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。

10

【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。

20

【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。

30

【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。

40

【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】

50

## 【 0 0 3 5 】

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 】

## 【 0 0 3 6 】

【 図 1 】 従来の音声合成システム 5 0 のブロック図である。

10

【 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 ( P D M ) 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 2 0 0 のブロック図である。

【 図 7 】 図 6 に示す P D M エンコーダ 2 2 0 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 2 4 6 のブロック図である。

20

【 図 9 】 図 6 の圧縮処理部 2 5 0 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 1 0 】 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 1 1 】 / 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 1 2 】 図 1 1 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 P D M 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 】

## 【 0 0 3 7 】

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。

30

## 【 0 0 3 8 】

## [ 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 1 6 0 及び 1 6 2 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。

## 【 0 0 3 9 】

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 ( P D M ) を行なって、スペクトル 1 6 0 及び 1 6 2 の振幅 ( 又はパワー ) をパルス密度を表すパルス列 1 7 0 及び 1 7 2 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。

40

## 【 0 0 4 0 】

パルス列 1 7 0 及び 1 7 2 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 1 7 4 が得られる。このパルス列 1 7 4 を P D M デコードして、スペクトル 1 6 0 及び 1 6 2 を平均した新たなスペクトル 1 8 0 が得られる。

## 【 0 0 4 1 】

50

〔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】

10

20

30

40

50



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、リーズ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作成する処理である。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

10

20

30

40

50

(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、リーズ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た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

10

20

30

40

50

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 9 6 に与える。

【 0 0 5 9 】

音源生成部 9 4 は音響パラメータ生成部 9 2 から与えられる F 0 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 9 6 は、この音源信号に対して P D M デコーダ 2 3 0 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。

【 0 0 6 0 】

以上のように本実施の形態によれば、H M M 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。

10

【 0 0 6 1 】

例えば、図 1 1 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 1 2 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 3 2 2 を、厳密に周波数の対応関係を付けて平均した波形 3 2 0 と比較すると、波形 3 2 0 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 3 2 2 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 3 2 4 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。

20

【 0 0 6 2 】

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全ての変更を含む。

【符号の説明】

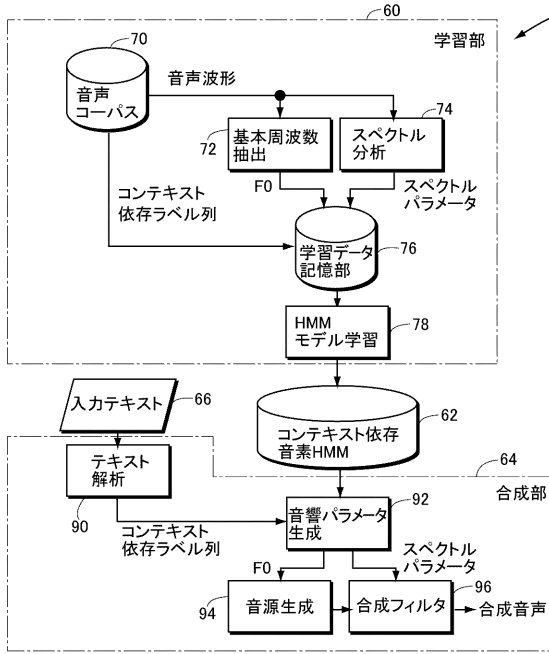
【 0 0 6 3 】

- 5 0 , 2 0 0 音声合成システム
- 6 0 , 2 1 0 学習部
- 6 2 , 2 1 2 コンテキスト依存音素 H M M
- 6 4 , 2 1 4 合成部
- 6 6 テキスト
- 7 0 音声コーパス
- 7 2 基本周波数抽出部
- 7 4 スペクトル分析部
- 7 6 , 2 2 2 学習データ記憶部
- 7 8 H M M モデル学習部
- 9 0 テキスト解析部
- 9 2 音響パラメータ生成部
- 9 4 音源生成部
- 9 6 合成フィルタ
- 1 7 0 , 1 7 2 , 1 7 4 パルス列
- 2 2 0 P D M エンコーダ
- 2 3 0 P D M デコーダ
- 2 4 6 デルタ・シグマ変調部

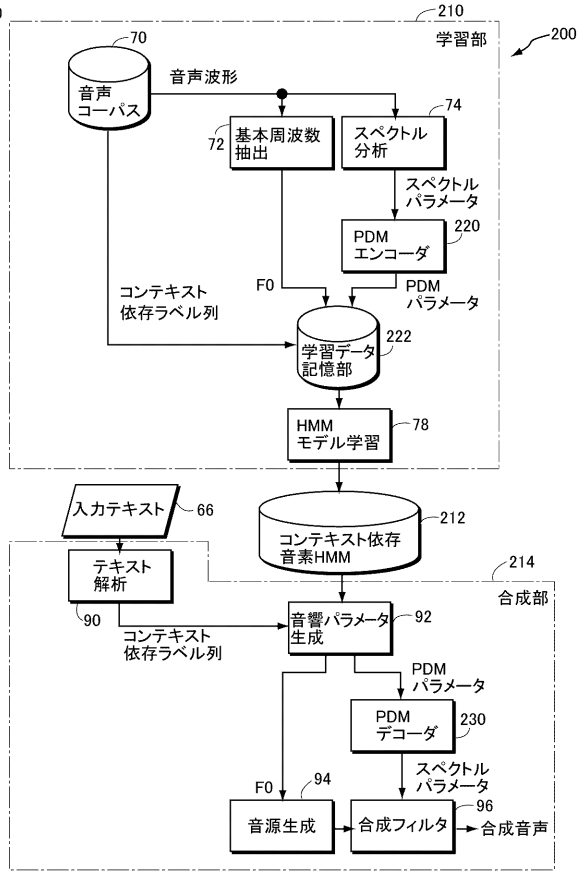
30

40

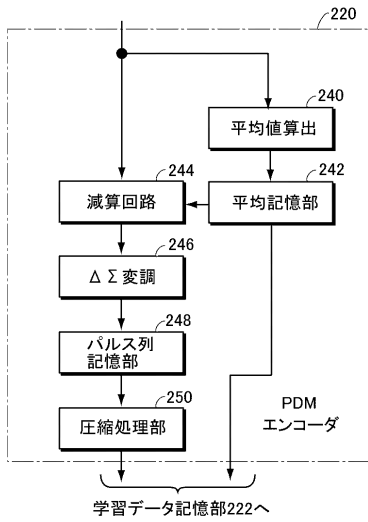
【図1】



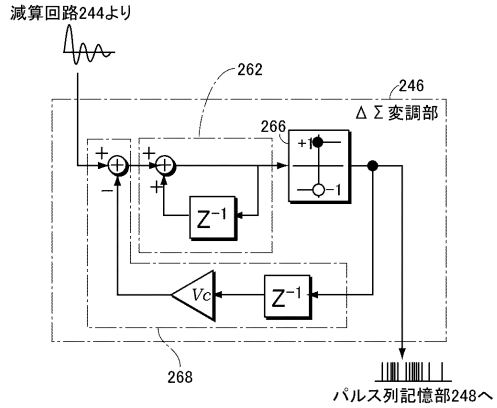
【図6】



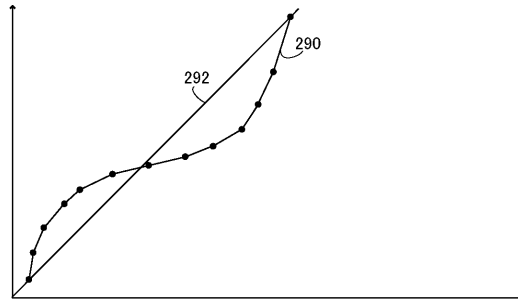
【図7】



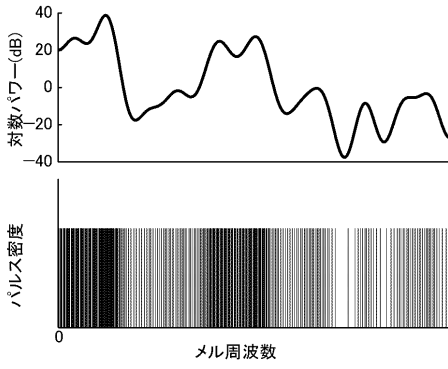
【図8】



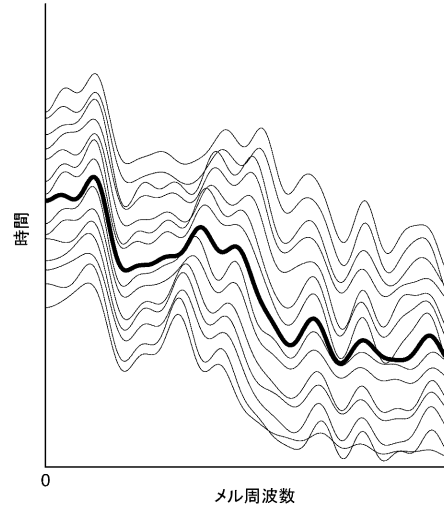
【図9】



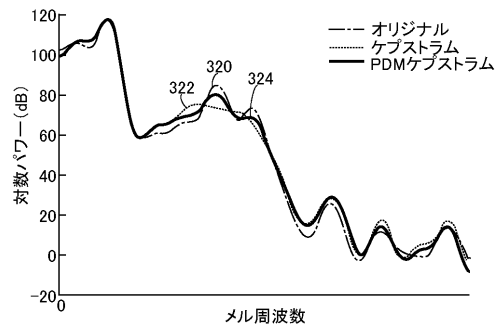
【図 1 0】



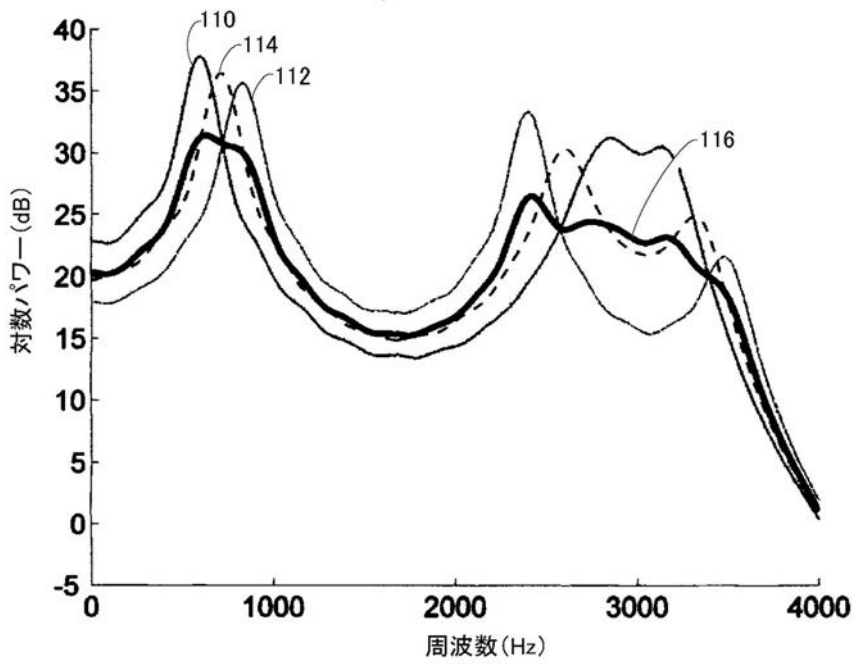
【図 1 1】



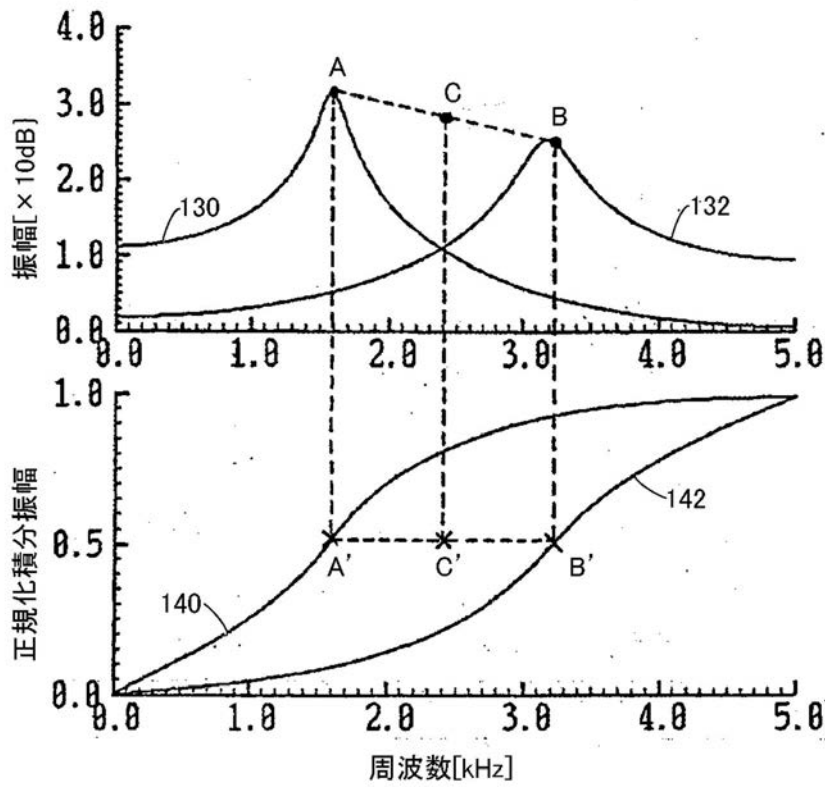
【図 1 2】



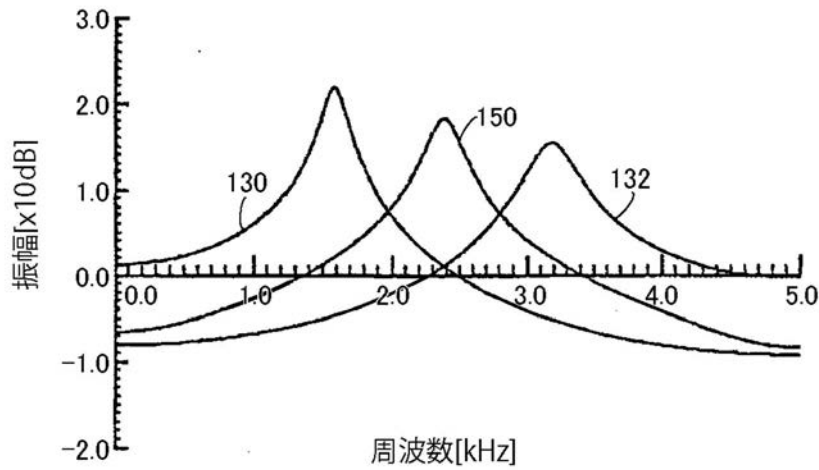
【図 2】



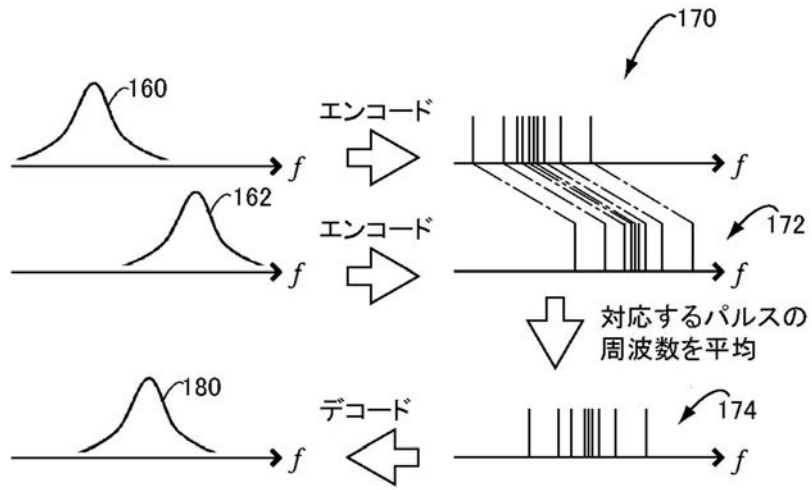
【 図 3 】



【 図 4 】



【図5】



---

フロントページの続き

(58)調査した分野(Int.Cl., DB名)

G 1 0 L 2 5 / 1 8

G 1 0 L 1 3 / 0 6