

(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

特開平 7 - 8 6 9 5 2

(43)公開日 平成7年(1995)3月31日

(51)Int. Cl.<sup>6</sup>  
H 0 3 M 3/04

識別記号 庁内整理番号  
8842 - 5 J

F I

技術表示箇所

審査請求 未請求 請求項の数 2

O L

(全 6 頁)

(21)出願番号 特願平5-227577

(22)出願日 平成5年(1993)9月13日

(71)出願人 000004226

日本電信電話株式会社  
東京都千代田区内幸町一丁目1番6号

(72)発明者 片岡 章俊

東京都千代田区内幸町一丁目1番6号 日本  
電信電話株式会社内

(72)発明者 守谷 健弘

東京都千代田区内幸町一丁目1番6号 日本  
電信電話株式会社内

(72)発明者 林 伸二

東京都千代田区内幸町一丁目1番6号 日本  
電信電話株式会社内

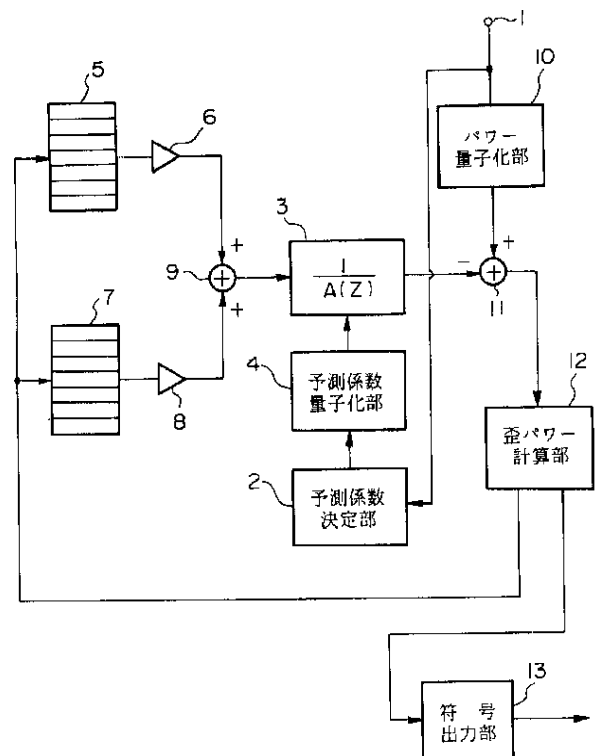
(74)代理人 弁理士 志賀 正武

(54)【発明の名称】音声の予測符号化方法

(57)【要約】

【目的】 異なる周波数特性を有する音声を符号化して伝送した場合でも、伝送先で良い品質の復号音声を得る。

【構成】 予測係数決定部 2，予測係数量子化部 4 は、合成フィルタ 3 に予測係数を設定する。適応符号帳 5，雑音符号帳 7 からピッチ周期ベクトル，雑音波形ベクトルが出力され、利得部 6，8 においてそれぞれ利得が乗算される。利得部 6，8 の出力は加算された後、合成フィルタ 3 に供給され、合成音声ベクトルが合成される。パワーが量子化された入力音声ベクトルから合成音声ベクトルが減算されて得られた歪は、その重み付けの度合いが入力音声ベクトルの周波数特性に基づいて適応的に制御された聴覚重み付けがなされた後、パワーが計算され、このパワーが最小になるように、適応符号帳 5，雑音符号帳 7 からピッチ周期ベクトル，雑音波形ベクトルが選択され、利得部 6，8 の利得が設定される。



## 【特許請求の範囲】

【請求項 1】 入力音声を線形予測分析して予測係数を算出し、該予測係数を合成フィルタに設定し、該合成フィルタを、複数のピッチ周期ベクトルが記憶された適応符号帳と、複数の雑音波形ベクトルが記憶された雑音符号帳とからそれぞれ選択されたピッチ周期ベクトルおよび雑音波形ベクトルによって、前記入力音声の複数サンプルからなるフレーム単位に駆動して合成音声を合成することを利用して音声を符号化する音声の予測符号化方法において、

前記合成音声と前記入力音声との歪が最小となるように前記適応符号帳および前記雑音符号帳から前記ピッチ周期ベクトルおよび前記雑音波形ベクトルを選択するために前記歪に聴覚的な重み付けを行う際に、その重み付けの度合いを前記入力音声の周波数特性に基づいて適応的に制御することを特徴とする音声の予測符号化方法。

【請求項 2】 前記重み付けの度合いを、パーコール係数を用いて適応的に制御することを特徴とする請求項 1 記載の音声の予測符号化方法。

## 【発明の詳細な説明】

## 【0001】

【産業上の利用分野】本発明は、自動車電話等のデジタル移動通信などに用いられ、音声を高効率に符号化する音声の予測符号化方法に関する。

## 【0002】

【従来の技術】近年、デジタル移動通信などの技術分野においては、電波を有効利用するなどの目的で、種々の高効率符号化方法が用いられている。これらの高効率符号化方法のうち、8 k b i t / s 程度の符号化速度で音声を符号化する高効率符号化方法としては、符号駆動型線形予測 ( C E L P ) 符号化方法やベクトル加算駆動型線形予測 ( V S E L P ) 符号化方法、あるいはマルチパス符号化方法等がある。

【0003】なお、C E L P 符号化方法の詳細については、たとえば、M. R. Schroeder と B. S. Atal とが著した "Code-Excited Linear Prediction (CELP) : High-quality Speech at Very Low Rates" (Proc. ICASSP '85, 25.1. 1, pp. 937-940, 1985) を、V S E L P 符号化方法の詳細については、たとえば、I. A. Gerson と M. A. Jasiuk とが著した "Vector Sum Excited Linear Prediction (VSELP) Speech Coding at 8kps" (Proc. ICASSP '90, S9.3, p. 461-464, 1990) を、マルチパス符号化方法の詳細については、たとえば、小澤一範と荒関卓とが著した "ピッチ情報を用いた 9 . 6 - 4 . 8 k b i t / s マルチパス音声符号化方式" (信学誌 ( D - I I ) , J 7 2 - D - I I , 8 , p p . 1 1 2 5 - 1 1 3 2 , 1 9 8 9 ) をそれぞれ参照されたい。

【0004】図 1 は、従来の C E L P 符号化方法を用いた音声の符号化装置の構成例を表すブロック図である。アナログの音声信号がサンプリング周波数 8 k H z でサ

ンプリングされて生成された入力音声データが入力端子 1 から入力される。予測係数決定部 2 において、入力端子 1 から入力された入力音声データの複数のサンプルが 1 フレームとして 1 つのベクトルにまとめられ (以下、入力音声ベクトルという)、この入力音声ベクトルについて線形予測分析がなされ、伝達関数  $\{ 1 / A ( z ) \}$  を有する合成フィルタ 3 の予測係数 (線形予測符号化 ( L P C ) 係数、または線スペクトル対 ( L S P ) 係数) が計算され、決定される。これにより、予測係数量子化部 4 において、予測係数が量子化され、合成フィルタ 3 に設定される。

【0005】適応符号帳 5 は、音声の有声区間のピッチ周期に対応した複数のピッチ周期ベクトルが記憶されるように構成されている。この適応符号帳 5 から、後述する歪パワー計算部 1 2 によって選択され、取り出されたピッチ周期ベクトルに、利得部 6 において、同じく歪パワー計算部 1 2 によって設定された利得が乗算され、利得部 6 から出力される。

【0006】いっぽう、雑音符号帳 7 には、音声の無声区間に対応した複数の雑音波形ベクトル (たとえば、乱数ベクトル) があらかじめ記憶されている。この雑音符号帳 7 から、後述する歪パワー計算部 1 2 によって選択され、取り出された雑音波形ベクトルに、利得部 8 において、歪パワー計算部 1 2 によって設定された利得が乗算され、利得部 8 から出力される。そして、利得部 6 の出力ベクトルと、利得部 8 の出力ベクトルとが加算器 9 において加算され、加算器 9 の出力ベクトルが合成フィルタ 3 に駆動ベクトルとして供給され、合成フィルタ 3 において、設定された予測係数に基づいて音声ベクトル (以下、合成音声ベクトルという) が合成される。

【0007】また、パワー量子化部 1 0 において、入力音声ベクトルのパワーが計算された後、そのパワーが量子化され、これにより、量子化された入力音声ベクトルのパワーが用いられ入力音声ベクトルとピッチ周期ベクトルとが正規化される。そして、減算器 1 1 において、正規化され、パワー量子化部 1 0 から出力された入力音声ベクトルから合成音声ベクトルが減算されて、歪データが求められる。

【0008】次に、歪パワー計算部 1 2 は、歪データのパワーを計算し、この歪データのパワーが最も小さくなるように、適応符号帳 5 および雑音符号帳 7 それぞれからピッチ周期ベクトルおよび雑音波形ベクトルをそれぞれ選択するとともに、利得部 6 および 8 のそれぞれの利得を設定する。これにより、符号出力部 1 3 において、予測係数、入力音声ベクトルのパワー、ピッチ周期ベクトルおよび雑音波形ベクトルそれぞれに対して選択された情報 (コード) と利得などがビット系列の符号に変換されて出力され、これらの符号が伝送される。

【0009】ところで、歪パワー計算部 1 2 において、合成音声ベクトルと入力音声ベクトルとの差である歪デ

ータを評価する際、歪データが最小になること、すなわち、SNが最大になることだけで評価すると、量子化雑音が周波数軸上で一様に分布することになる。また、音声信号は、低域に多くのパワーを有するが、周波数の増加に従ってパワーは減少する。そのため、量子化雑音が周波数軸上で一様に分布していると、高域において量子化雑音レベルが音声レベルより相対的に高く、このことが符号化音声を劣化させる原因となる。

【0010】そこで、従来では、図2に示すように、歪パワー計算部12において、聴覚重み付けフィルタ14を用いて歪データを入力音声ベクトルのスペクトルに基づいて重み付けを行った後、パワー計算部15において評価する。つまり、音声パワーの大きい低域では、一様分布の時より多少量子化雑音レベルが大きくなって音声によってマスキングされるため、量子化雑音は聞こえない。逆に、高域では、一様分布より低くなるように重み付けを行う。図2において、eは減算器11から出力された歪データ、e'は重み付きの歪データである。

【0011】聴覚重み付けフィルタ14の伝達関数W(z)は、(1)式によって表される。

【数1】

$$W(z) = \frac{A\left(\frac{z}{\gamma_1}\right)}{A\left(\frac{z}{\gamma_2}\right)} \cdots (1)$$

ここで、

【数2】

$$A\left(\frac{z}{\gamma_1}\right) = 1 - \sum_1^{10} (\alpha_i \gamma_1^i) z^{-i} \cdots (2)$$

【数3】

$$A\left(\frac{z}{\gamma_2}\right) = 1 - \sum_1^{10} (\alpha_i \gamma_2^i) z^{-i} \cdots (3)$$

(1)~(3)式において、係数 $\alpha_i$ は、予測係数決定部2において得られた量子化されていないLPC係数である。また、係数 $\gamma_1$ および $\gamma_2$ には、 $0 < \gamma_2 < \gamma_1 < 1$ の値が用いられる。また、係数 $\gamma_1$ および $\gamma_2$ によって聴覚重み付けフィルタ14の特性が左右されるため、これらの係数 $\gamma_1$ および $\gamma_2$ の値は、試験によって経験的に決定される。

【0012】

【発明が解決しようとする課題】ところで、電話機の音声には、従来から、国際電信電話諮問委員会(CCITT)において標準化されているIRS特性を有する音声を用いられてきた。しかしながら、最近では、小型のエレクトレット・マイクロホンの普及に伴って、IRS特性とは異なる周波数特性(以下、NON-IRS特性と

いう)を有する音声も用いられるようになってきている。そして、このような異なる周波数特性を有する音声を符号化する場合、上述した聴覚重み付けフィルタ14の係数 $\gamma_1$ および $\gamma_2$ の最適値も当然異なる。

【0013】しかしながら、上述した従来の音声の予測符号化方法においては、既に説明したように、聴覚重み付けフィルタ14の係数 $\gamma_1$ および $\gamma_2$ の値は、入力音声の周波数特性に関係なく、入力音声に応じて試験を行って経験的に一定の値に固定されていた。したがって、上記聴覚重み付けフィルタ14の係数 $\gamma_1$ および $\gamma_2$ の値に適合しない周波数特性を有する音声を符号化して伝送した場合には、伝送先で良い品質の復号音声を得ることができないという問題があった。本発明は、このような背景の下になされたもので、異なる周波数特性を有する音声を符号化して伝送した場合でも、伝送先で良い品質の復号音声を得ることができる音声の予測符号化方法を提供することを目的とする。

【0014】

【課題を解決するための手段】請求項1記載の発明は、入力音声を線形予測分析して予測係数を算出し、該予測係数を合成フィルタに設定し、該合成フィルタを、複数のピッチ周期ベクトルが記憶された適応符号帳と、複数の雑音波形ベクトルが記憶された雑音符号帳とからそれぞれ選択されたピッチ周期ベクトルおよび雑音波形ベクトルによって、前記入力音声の複数サンプルからなるフレーム単位に駆動して合成音声を合成することを利用して音声を符号化する音声の予測符号化方法において、前記合成音声と前記入力音声との歪が最小となるように前記適応符号帳および前記雑音符号帳から前記ピッチ周期ベクトルおよび前記雑音波形ベクトルを選択するために前記歪に聴覚的な重み付けを行う際に、その重み付けの度合いを前記入力音声の周波数特性に基づいて適応的に制御することを特徴としている。請求項2記載の発明は、請求項1記載の発明において、前記重み付けの度合いを、パーコール係数を用いて適応的に制御することを特徴としている。

【0015】

【作用】本発明によれば、合成音声と入力音声との歪に施される聴覚的な重み付けの度合いが入力音声の周波数特性に基づいて適応的に制御されるので、異なる周波数特性を有する音声を符号化して伝送した場合でも、伝送先で良い品質の復号音声が得られる。

【0016】

【実施例】以下、図面を参照して、本発明の実施例について説明する。本発明においては、音声の符号化装置の構成は、図1および図2とほぼ同様であるが、歪パワー計算部12を構成する聴覚重み付けフィルタ14の係数 $\gamma_2$ の値が、以下に示す第1~第3の実施例によってそれぞれ適応的に制御される。

【0017】(1)第1の実施例(フレーム内処理)

聴覚重み付けフィルタ14の係数 $\gamma_1$ および $\gamma_2$ の値を決定する際の試聴の結果、IRS特性を有する音声においては、係数 $\gamma_1$ および $\gamma_2$ の値は、 $\gamma_1 = 0.9$ 、 $\gamma_2 = 0.6$ が好ましい。いっぽう、NON-IRS特性を有する音声においては、 $\gamma_1 = 0.9$ 、 $\gamma_2 = 0.4$ が好ましい。そこで、IRS特性およびNON-IRS特性の両特性に最適となるように、入力音声に応じて係数 $\gamma_2$ の値を制御すればよい。つまり、入力音声はIRS特性を有する場合には、 $\gamma_2 = 0.6$ と設定し、入力音声はNON-IRS特性を有する場合には、 $\gamma_2 = 0.4$ と設定すればよい。

【0018】また、IRS特性を有する音声およびNON-IRS特性を有する音声を分析した結果、両者には、PARCOR係数(パーコール係数)の1次の係数 $k_1$ の出現確率の分布に大きな違いがあることがわかった。すなわち、このPARCOR係数 $k_1$ は、必ず $-1 < k_1 < 1$ の範囲に存在するが、NON-IRS特性を有する音声を分析して得られたPARCOR係数 $k_1$ は、値+1付近に集中する傾向があり、いっぽう、IRS特性を有する音声を分析して得られたPARCOR係数 $k_1$ には、そのような傾向がない。

【0019】ここで、図3に実際の音声データを処理して得られるPARCOR係数 $k_1$ の値をその出現確率の分布別に統計を取った時の値+1付近の、IRS特性を有する音声を分析して得られたPARCOR係数 $k_1$ の出現確率の分布(曲線a)、およびNON-IRS特性を有する音声を分析して得られたPARCOR係数 $k_1$ の出現確率の分布(曲線b)を示す。図3からわかるように、NON-IRS特性を有する音声は、PARCOR係数 $k_1 > 0.9$ において多く存在するが、IRS特性を有する音声は、PARCOR係数 $k_1 > 0.9$ の出現が減少している。

10

\*【0020】したがって、このPARCOR係数 $k_1$ の相違の特徴を用いて、聴覚重み付けフィルタ14の係数 $\gamma_2$ の値を、入力音声によって適応的に制御する。つまり、入力音声を分析して得られたPARCOR係数 $k_1$ の値がしきい値Th(たとえば、 $tH = 0.9$ )以上である場合には、聴覚重み付けフィルタ14の係数 $\gamma_2$ の値を0.4に設定し、係数 $k_1$ の値がしきい値Thより小さい場合には、係数 $\gamma_2$ の値を0.6に設定する。なお、PARCOR係数 $k_1$ は、予測係数決定部2において線形予測分析を行う時に求めることができる。また、従来の技術において既に説明したように、音声の符号化はフレーム単位で行われるため、この実施例においては、聴覚重み付けフィルタ14の係数 $\gamma_2$ の適応的制御もフレーム単位で行われる。

【0021】(2)第2の実施例(フレーム間処理) 図1に示す予測係数決定部2においては、入力音声データについてフレーム単位で線形予測分析がなされ、合成フィルタ3の予測係数が計算されるが、入力音声データの子音部や無声区間などでは、このフレーム単位での線形予測分析が必ずしも有効ではない。そのため、NON-IRS特性を有する音声においても、フレーム単位では、PARCOR係数 $k_1$ の値が値+1付近に集中するとは限らない。また、聴覚重み付けフィルタ14の係数 $\gamma_2$ の値がフレーム単位で大きく変化した場合、復号音声の連続性が失われて望ましくない。

【0022】そこで、この実施例においては、係数 $\gamma_2$ の値を(4)式で表されるように、現在処理しようとしているフレーム(現フレーム)の入力音声データの係数 $\gamma_2$ と、過去に処理されたM個のフレームのそれぞれの入力音声データの係数 $\gamma_2$ の和によって表現する。

【数4】

$$\gamma_2 = W_0 \cdot \gamma_2(n) + W_1 \cdot \gamma_2(n-1) + W_2 \cdot \gamma_2(n-2) + \dots + W_M \cdot \gamma_2(n-M) \dots (4)$$

(4)式において、nはフレーム番号(現フレームのフレーム番号はn)、 $\gamma_2(n)$ はフレーム番号nのフレームの入力音声データを分析して得られたPARCOR係数 $k_1$ によって決まる係数 $\gamma_2$ 、Mは次数(たとえば、 $M = 3$ )、 $W_i$ は重み係数である。このように、係数 $\gamma_2$ を和の形で表現することにより、フレーム単位での係数 $\gamma_2$ の急激な変化が避けられるので、復号音声の連続性が失われることはない。

30

あるいはNON-IRS特性を有するか不明である。しかしながら、たとえば、使用する電話機によって音声の周波数特性が決まってくるので、電話機の機種が決定されれば、入力音声データがIRS特性を有するか、あるいはNON-IRS特性を有するかが決まり、以後、両特性が逆転することはあまり考えられない。

【0024】そこで、この実施例においては、図3に示すPARCOR係数 $k_1$ の出現確率の分布に基づいて、聴覚重み付けフィルタ14の係数 $\gamma_2$ の値を(5)式に表すように適応的に制御する。

【数5】

【0023】(3)第3の実施例

ところで、実際の使用状況において、音声の符号化装置に入力される入力音声データがIRS特性を有するか、

40

(5)

7

8

$$\gamma_2(n) = \begin{cases} \gamma_{2min} & \gamma_2(n-1) = \gamma_{2min} \\ \gamma_2(n-1) - \alpha_1 & a_1 \leq k_1 \\ \gamma_2(n-1) - \alpha_2 & a_2 \leq k_1 < a_1 \\ \gamma_2(n-1) & a_3 \leq k_1 < a_2 \\ \gamma_2(n-1) + \beta & k_1 < a_3 \\ \gamma_{2max} & \gamma_2(n-1) = \gamma_{2max} \end{cases} \dots (5)$$

(5)式において、 $\gamma_{2min} = 0.4$ 、 $\gamma_{2max} = 0.6$ 、 $\alpha_1 = 0.05$ 、 $\alpha_2 = 0.01$ 、 $\beta = 0.001$ 、 $a_1 = 0.97$ 、 $a_2 = 0.95$ 、 $a_3 = 0.90$ 、初期値 $\gamma_2(0) = 0.6$ とする。また、(5)式において、 $n$ はフレーム番号である。

【0025】入力音声データがIRS特性を有する場合には、係数 $\gamma_2$ は、 $a_1 = 0.97$ 以上の値が出現する確率はゼロであり、 $a_2 = 0.95$ 以上の値が出現する確率も非常に少ない。そのため、係数 $\gamma_2(n)$ はほとんど変化しない。つまり、入力音声データがIRS特性を有する場合には、係数 $\gamma_2(n)$ が0.6に保たれる。いっぽう、入力音声データがNON-IRS特性を有する場合には、係数 $\gamma_2$ は、 $a_2 = 0.95$ 以上の値が出現する確率が非常に高く、 $\alpha_1 = 0.05$ ならば、4フレーム後には、 $\gamma_2(n) = 0.4$ となる。そして、音声区間によって係数 $\gamma_2$ の値が0.9以下である場合でも、増加が $\beta = 0.001$ と小さいので、急激に変化しない。また、係数 $\gamma_2$ の値が0.05を越えるフレームがあれば、係数 $\gamma_2(n)$ は0.4に収束する。以上、

20

30

数特性を有する音声を符号化して伝送した場合でも、伝送先で良い品質の復号音声を得ることができるという効果がある。

【図面の簡単な説明】

【図1】本発明の第1～第3の実施例および従来例による音声の予測符号化方法を適用した音声の符号化装置の構成を表すブロック図である。

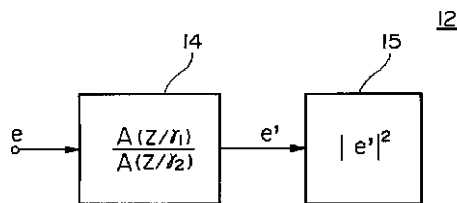
【図2】歪パワー計算部12の構成の一例を示すブロック図である。

【図3】入力音声の周波数特性の違いによるPARCOR係数 $k_1$ の出現確率の分布の一例を示す図である。

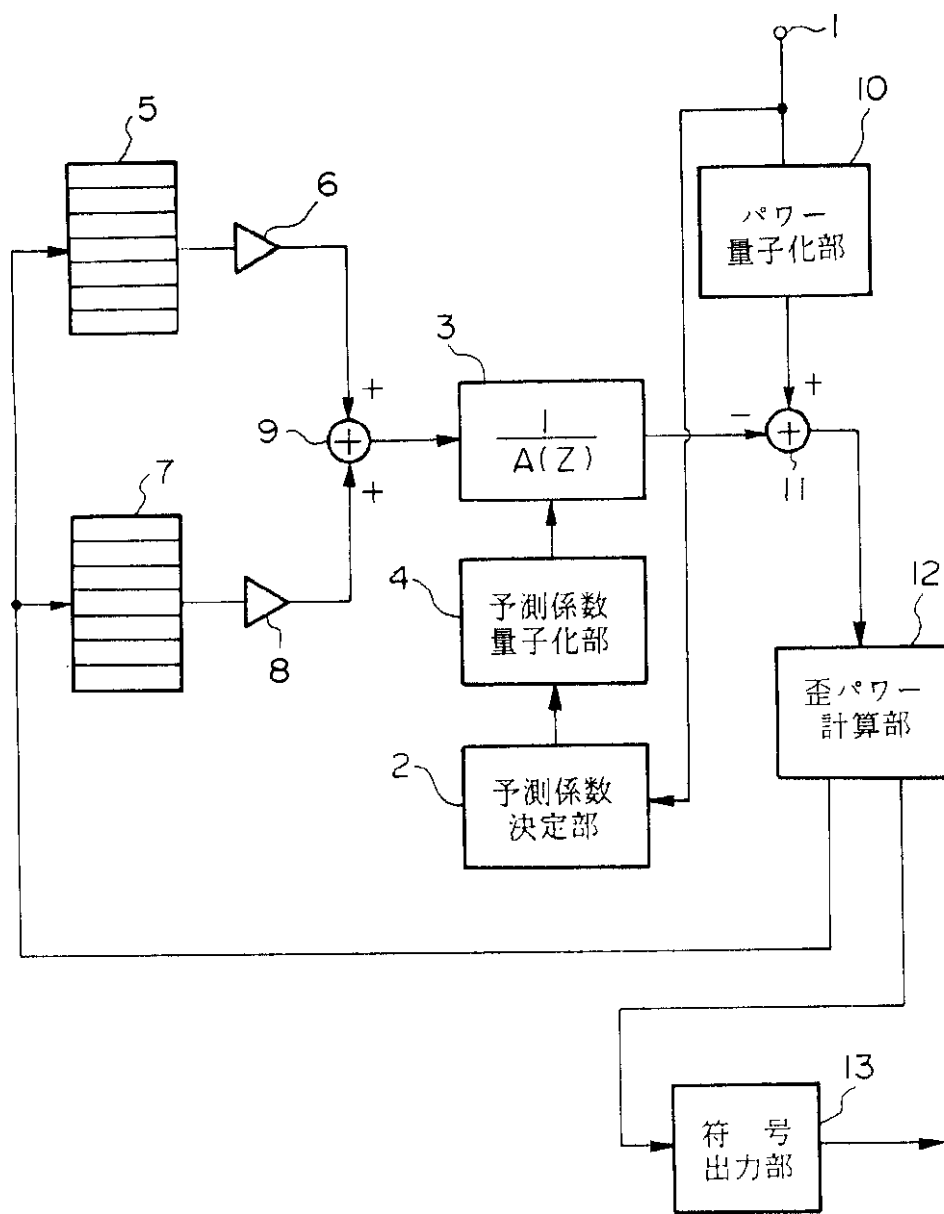
【符号の説明】

- 1 入力端子
- 2 予測係数決定部
- 3 合成フィルタ
- 4 予測係数量子化部
- 5 適応符号帳
- 6, 8 利得部
- 7 雑音符号帳
- 9 加算器
- 10 パワー量子化部
- 11 減算器
- 12 歪パワー計算部
- 13 符号出力部
- 14 聴覚重み付けフィルタ
- 15 パワー計算部

【図2】



【図1】



【図3】

