

(19)日本国特許庁(JP)

(12)特許公報 ( B 2 )

(11)特許番号

特許第 3 2 0 9 2 4 7 号

( P 3 2 0 9 2 4 7 )

(45)発行日 平成13年9月17日(2001.9.17)

(24)登録日 平成13年7月13日(2001.7.13)

(51)Int. Cl. <sup>7</sup>	識別記号	F I
G 1 0 L	19/12	G 1 0 L 9/14 S
	19/00	9/18 E

請求項の数 1

(全 5 頁)

(21)出願番号	特願平5-165581	(73)特許権者	000004226 日本電信電話株式会社 東京都千代田区大手町二丁目3番1号
(22)出願日	平成5年7月5日(1993.7.5)	(72)発明者	守谷 健弘 東京都千代田区内幸町1丁目1番6号 日本 電信電話株式会社内
(65)公開番号	特開平7-20895	(72)発明者	片岡 章俊 東京都千代田区内幸町1丁目1番6号 日本 電信電話株式会社内
(43)公開日	平成7年1月24日(1995.1.24)	(74)代理人	100066153 弁理士 草野 卓
審査請求日	平成11年11月10日(1999.11.10)	審査官	渡邊 聡
		(56)参考文献	特開 平5 - 265496 ( J P , A ) 特開 平5 - 100697 ( J P , A ) 特開 平1 - 319799 ( J P , A )

最終頁に続く

(54)【発明の名称】音声の励振信号符号化法

1

(57)【特許請求の範囲】

【請求項 1】 ピッチ周期成分ベクトルと雑音励振ベクトルを線形予測合成フィルタの励振源とする音声合成モデルを持ち、合成音声と入力音声の誤差を最小化するように雑音励振ベクトルを符号帳のなかから選択する音声の励振信号符号化法において、雑音励振ベクトル毎にその合成後の波形と入力音声波形の内積を求め、前記合成フィルタの類型を決定し、予め、雑音励振ベクトル毎、および合成フィルタの類型毎に作られた励振ベクトルの合成後のエネルギー推定値の表または推定値の逆数の表から推定値を、前記決定した類型で各雑音励振ベクトル毎に読み出し、前記内積値の 2 乗を前記エネルギーの推定値でわった値を基準に雑音励振ベクトルの選択または予備選択を行な

2

うことを特徴とする音声の励振信号符号化法。

【発明の詳細な説明】

【 0 0 0 1 】

【産業上の利用分野】この発明は音声信号をできるだけ少ない情報量でデジタル符号化する高能率音声符号化法で、特に線形予測合成フィルタに供給する励振信号を決定する励振信号符号化方法に関するものである。

【 0 0 0 2 】

【従来の技術】 8 k b i t / s 程度以下の音声符号化法として、CELP (Code Excited Linear Prediction: 符号励振線形予測) 符号化が良く知られている。この符号化法は図 5 に示すように、フレーム単位で雑音符号帳 1 1 と適応符号帳 1 2 から励振ベクトルを選択して、それぞれ利得調整手段 1 3 , 1 4 で利得を調整して線形予測合成フィルタ 1 5 に

入力して音声合成する。この合成音声と入力音声との差を差回路 16 で求め、その差出力を聴感補正フィルタ 17 を通し、そのフィルタ出力から歪計算部 18 で歪を計算し、その歪が小さくなるように励振信号を選択し利得調整する。励振ベクトルの選択基準は合成された信号と入力信号の聴感上の誤差を最小化することである。このように最終的に合成される波形をフィードバックして励振ベクトルを決定するため、高い品質が得られる。

【0003】この半面、大量の雑音励振ベクトルを記憶しておく必要があることや、励振ベクトルの探索の演算量が膨大となるという難点がある。このため、記憶容量や演算量を削減するためにいくつかの手法が考えられているが、以下で探索の基本原則を示し、この発明と関連のある代表的な演算量削減法について整理して紹介する。

$$d = \| X - gHC^j \|^2 \tag{1}$$

$$H = \begin{bmatrix} h_0 & 0 & \dots & \dots & 0 \\ h_1 & h_0 & 0 & \dots & 0 \\ h_2 & h_1 & h_0 & \dots & 0 \\ \dots & \dots & \dots & \dots & 0 \\ h_{n-1} & h_{n-2} & h_{n-3} & \dots & h_0 \end{bmatrix} \tag{2}$$

ここで、最適な利得が与えられ、雑音励振ベクトル決定が最適な利得となり ( $X^T$  は  $X$  の転置行列)、このときとなる。結局第 2 項の

$$g = X^T HC^j / ((HC^j)^T (HC^j)) \tag{3}$$

$$d = X^T X - (X^T HC^j)^2 / ((HC^j)^T (HC^j)) \tag{4}$$

$$f = (X^T HC^j)^2 / ((HC^j)^T (HC^j)) \tag{5}$$

を最大とする  $j$  を捜せばよいことになる。ここで  $HC^j$  の合成演算には  $n^2 / 2$  回程度の積和演算が必要ですべての  $j$  について  $C^j$  を合成するには莫大な演算量を必要とする。そこで種々の提案がなされている。

【0006】まず式 (5) の分子は  $X^T H$  を 1 回だけ先に計算し、そのあとで  $C^j$  と内積をとれば近似誤差を発生させずに演算量を削減する手法が知られている。内積は  $n$  回の演算ですむためである。ところが、分母の計算で演算量を削減しようとすると、近似誤差が生ずる。例えば分母の項を  $(C^j)^T H^T HC^j$  として、 $H^T H$  をテブリッツ型の相関行列で近似し、 $H$  の自己相関関数と  $C^j$  の自己相関関数の内積で計算する方法が知られている。これにより、分母の項を  $n$  回の積和で計算できるが、近似誤差が大きく、歪の増加と品質の劣化を招く。また雑音励振ベクトル  $C^j$  毎に相関値を記憶しておくために、符号帳のための記憶容量と同程度の記憶容量の追加が必要となる。

【0007】これとは別に  $C$  にローパスフィルタを掛けてダウンサンプルした励振ベクトルをもち、分母の項をこのベクトルで近似する手法も知られている。ダウンサンプルした比率で演算量を削減できるが、この手法でも近似誤差やダウンサンプルしたベクトルのための記憶容

\*【0004】CELP 符号化における雑音励振ベクトルの探索では、前述したように線形予測合成フィルタ 15 の出力合成波形と入力音声との差をとり、その差出力を聴感補正フィルタ 17 を通したあとの歪が最小となるベクトルを選択する。具体的には、あるフレーム (フレーム長は  $n$  サンプル) で量子化の目標とする  $n$  次元の縦ベクトル (入力音声から、前のフレームからの応答成分や適応符号ベクトルからの合成成分を差し引いた現フレームの入力音声を、聴感補正フィルタに通した信号ベクトル) を  $X$ 、符号帳中の  $m$  種類の雑音励振ベクトルの  $n$  次元縦ベクトルを  $C^j$  ( $j = 0, \dots, m - 1$ )、聴感補正フィルタ 17 も一括した合成フィルタ 15 のインパルス応答行列 ( $n \times n$ ) を  $H$ 、雑音励振ベクトルの利得を  $g$  とすると、下記の歪み  $d$  を最小とする  $j$  を探索する。

【0005】

後にその利得を量子化することとすると、

の歪は

量が増加するという問題がある。また探索を 2 段階にわけ、第 1 段では近似計算で複数の  $C^j$  を候補として予備選択し、その候補についてのみ第 2 段で式 (5) を最大とするものを選択する方法も知られている。第 1 段の近似法としては分子の項だけを用いる方法や、上記で紹介した近似法を用いる方法などが考えられる。第 1 段の近似誤差を小さくすることや候補数を増やすことで歪を小さくできるが、演算量削減効果が小さくなってしまふ。

【0008】

【発明が解決しようとする課題】この発明の目的は音声を少ない情報量で符号化するとき、できるだけ演算量を抑えたまま符号化音声の品質を向上させ、特に上記従来法の中で紹介した 2 段階の探索と組合せ、歪の劣化が小さい割に大幅な演算量の削減ができる音声の励振信号符号化方法を提供することにある。

【0009】

【課題を解決するための手段】この発明では雑音励振ベクトルを選択する (または予備選択する) 際に式 (5) の分母の項すなわちベクトルのエネルギーの項も考慮するが、そのエネルギーの項は予め蓄えておいた推定値を読み出すことで求めることが特徴である。このエネルギーの推定値は雑音励振ベクトル毎に、また合成フィルタ

の分類毎に予め蓄えておく。図 1 に 2 段階の探索を行なう場合について、この発明方法と従来法との比較を示す。従来においては図 1 B に示すようにエネルギー項を全く無視して、式 ( 5 ) の分子の計算結果のみで予備選択するか、図 1 C に示すように、近似演算を行なうことでエネルギー項を求め、これを用いて式 ( 5 ) を近似計算して予備選択することが知られていたが、この発明では図 1 A に示すようにインパルス応答の分類に基づいた表の参照でエネルギー項を推定し、これを用いて式 ( 5 ) を近似計算して予備選択する。

【 0 0 1 0 】

【実施例】図 2 にこの発明の最も基本的な実施例の要部の構成を示す。通常の m 個の n 次元雑音励振ベクトルからなる符号帳 1 1 のほかに、この実施例では合成フィルタの種類を示すものとして複数種類のインパルス応答のパターンが蓄えられた記憶部 2 1 およびその記憶部 2 1 からインパルス応答系列と対応したコードを選択する選択器 2 2、インパルス応答パターンを縦軸、雑音励振ベクトルを横軸とし、各要素には合成後のエネルギー推定値またはその逆数が予め計算されて記憶されているエネルギー項表 2 3 とが備えられる。2 段階の探索を行なうとしたときの探索手順は以下ようになる。

【 0 0 1 1 】 1 . あるフレームで合成フィルタや聴感補正フィルタの特性が与えられると、合成フィルタのイン\*

$$e = \| h - h^i \|^2$$

またこの変形として次式のように高次の係数が減少する三角窓 W の重みをつけることや、低次のみでの距離計算

$$e = W \| h - h^i \|^2$$

これはインパルス応答行列が下三角であるため、低次の係数ほどエネルギー計算に大きく貢献していることを考慮したものである。

$$e = \sum_{i=0}^p |h_i - h_i^i|^2$$

ここで、 $h_i$  は  $i = 0$  から  $p$  まで、 $p$  はモデルの次数、 $h_i^i$  は代表モデルの  $i$  次の線形予測係数、 $h_i^i$  はインパルス応答の  $i$  次の自己相関関数であり、この尺度は音声認識で頻繁に使われる尺度である。また予め用意しておく少数の代表パターンの作成には通常のベクトル量子化で用いられている代表パターンの設計アルゴリズムをそのまま利用すればよい。

【 0 0 1 6 】 以上の実施例は図 5 に示した基本的な C E L P 符号化を前提に説明したが、雑音励振ベクトルの探

$$f = (X^T H C^j - H P)^2 / ((H C^j)^T (H C^j) - (H C^j)^T (H P)) \quad (9)$$

ここで

$$= (H C^j)^T (H P) / ((H P)^T (H P)) \quad (10)$$

である。式 ( 9 ) でも分子の項は式 ( 5 ) の場合と同様に正確に比較的少ない演算で計算できる。また分母の以下の第 2 項も積和の計算も同様である。従って分母の第 1 項を第 1 の実施例と同じ手順で表から読み出せば演算量を削減できる。また分母の第 2 項の計算を全く省略

\* パルス応答が決まる。このインパルス応答と記憶部 2 1 に予め蓄えられた応答パターンとの照合を選択器 2 2 でき、現在のフレームのインパルス応答のパターンを決定する。

2 . 次に式 ( 5 ) の分子の項を全ての雑音励振ベクトルに対して演算する。

【 0 0 1 2 】 3 . 選択器 2 2 からの決定されたインパルス応答に対応するエネルギー推定値の表 2 3 を参照して求め、各雑音励振ベクトルに対して、(分子の項 / エネルギー推定値) を算出し、その算出値の大きいものから設定された数の候補を残す。

4 . 予備選択した候補についてのみ式 ( 5 ) の計算を行ない、最適な雑音励振ベクトルの符号を決定する。

【 0 0 1 3 】 上記処理中で、インパルス応答を分類する、つまりインパルス応答のパターンを決定する方法としてはインパルス応答すなわち F I R フィルタ係数の距離尺度を用いる方法や、インパルス応答を近似する全極型のスペクトルモデルの尺度を用いる方法がある。フレーム毎に決まるインパルス応答ベクトルを  $h$  (ただし  $h^T = \{ h_0, \dots, h_{n-1} \}$ )、M 種類の代表パターンとして蓄えておくインパルス応答ベクトルを  $h^i$  ( $i = 0, \dots, M - 1$ ) とすると、前者の場合  $e$  を最小とする代表パターンを選ぶことになる。

【 0 0 1 4 】

$$(6)$$

も考えられる。

$$(7)$$

【 0 0 1 5 】 一方、全極型のスペクトルモデルを用いるときは以下の  $e$  を最小とする最尤距離尺度で代表パターンを選ぶ。

$$(8)$$

索の前に、すべての雑音励振ベクトルをピッチ成分と直交化させる手法や、複数チャンネルの雑音励振ベクトルをもつ場合にもこの発明を適用できる。すなわち第 2 の実施例として雑音励振ベクトルをピッチ成分 (適応符号帳からの出力) n 次元ベクトル P と直交化させる場合について説明する。この場合には最終的に最大化する式 ( 5 ) が次のように変形される。

【 0 0 1 7 】

$$f = (X^T H C^j - H P)^2 / ((H C^j)^T (H C^j) - (H C^j)^T (H P)) \quad (9)$$

ここで

$$= (H C^j)^T (H P) / ((H P)^T (H P)) \quad (10)$$

して、第 1 項のみで評価しても歪の増加はほとんどない。

【 0 0 1 8 】 第 3 の実施例として 2 チャンネルの雑音励振ベクトルをもつ場合について示す。ここでも 2 段階の探索を行なうと図 4 に示すような処理を行なうことにな

る。第1チャンネルの励振ベクトルをC<sup>j</sup>、第2チャンネルの励振ベクトルをC<sup>k</sup>とすると、最終的な評価尺度は以下ようになる。

【0019】  
【数1】

$$f = \frac{(X^T H C^j + X^T H C^k)^2}{(H C^j)^T (H C^j) + 2(H C^j)^T (H C^k) + (H C^k)^T (H C^k)} \quad (11)$$

しかしながら各チャンネル毎の予備選択ではそれぞれ、第1の実施例と同じ処理を行なう。すなわち分母のエネルギー項の概算値を表から読みだし、式(5)を最大とする候補を複数個選択する。そして、各チャンネルからの候補の組合せで式(11)が最大となるjとkの組合せを探索する。

【0020】上述において、エネルギー項の表23には雑音励振ベクトルの合成後のエネルギー推定値の逆数を各励振ベクトル及び各合成フィルタの類型ごとに記憶しておいてもよい。また場合によっては式(5)を正しく計算することなく、前記予備選択を最終選択としてもよい。

【0021】

【発明の効果】この発明を用いた場合の入力音声と符号化歪のSNRと音声符号化処理全体の演算量(単位MOPS: Mega Operation Per Second: 実時間処理に必要な1秒あたり100万回単位の演算処理回数)との関係を従来法と比較したものを図4に示す。この場合、雑音励振ベクトルは2チャンネルからなり、各チャンネル毎に7ビット(極性を除く)を配分した場合である。雑音励振ベクトルの次元数nは40でピッチ成分との直交化処理を併用している。またインパルス応答の分類は4種類とした。ここでの従来法は式(5)の分子の項のみで予備選択を行なう方法であ

る。図中の数字は各チャンネル毎に128個の中から残す候補数である。この図からこの発明により従来の予備選択と比較してほとんど演算量を増加させることなく、SNRを改善していることがわかる。この発明で残す候補数が同一であれば、増加する演算量の内分けは、インパルス応答の分類のための距離計算がフレームあたり1回、分母の項の逆数の積が予備選択での各雑音励振ベクトル毎に1回ずつで、探索の距離計算の演算量と比較してごく少ない量である。またこの発明で増加する記憶容量はインパルス応答の分類のための表21とエネルギーの表23の記憶容量である。例えば図3の場合のようにインパルス応答を4種類に分類し、40次元の雑音励振ベクトルを用いる時には、記憶容量の増加する割合は約1割である。

【図面の簡単な説明】

【図1】2段階の探索におけるこの発明の処理を従来法と対比させて示す図。

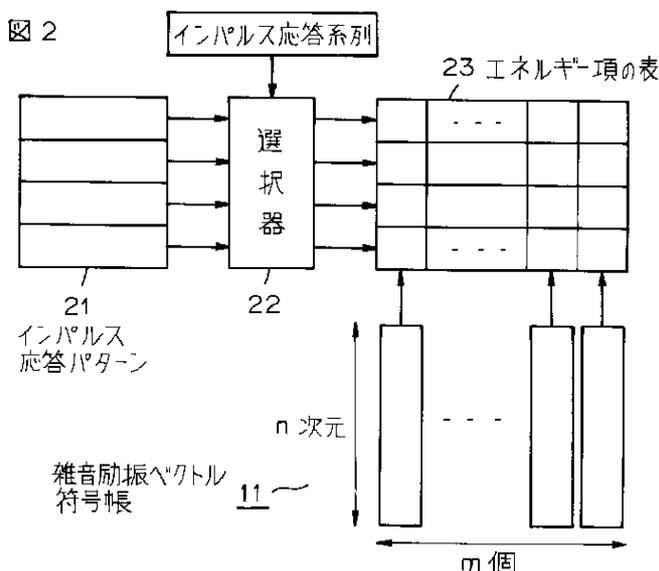
【図2】この発明の実施例の要部を示すブロック図。

【図3】この発明を2チャンネルの雑音励振ベクトルの場合への適用した実施例の要部を示すブロック図。

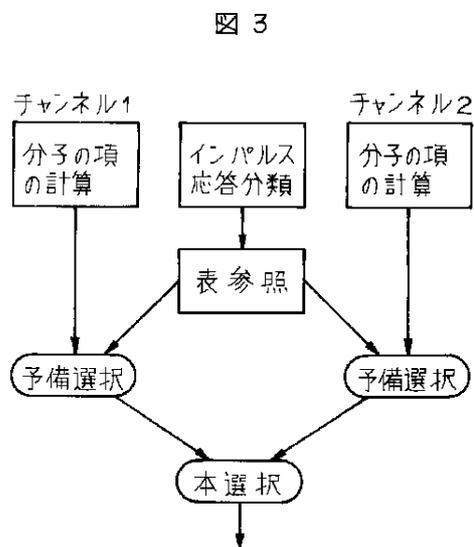
【図4】この発明によるSNRと演算量の関係を従来法と比較した図。

【図5】CELP符号化法の基本原理を示すブロック図。

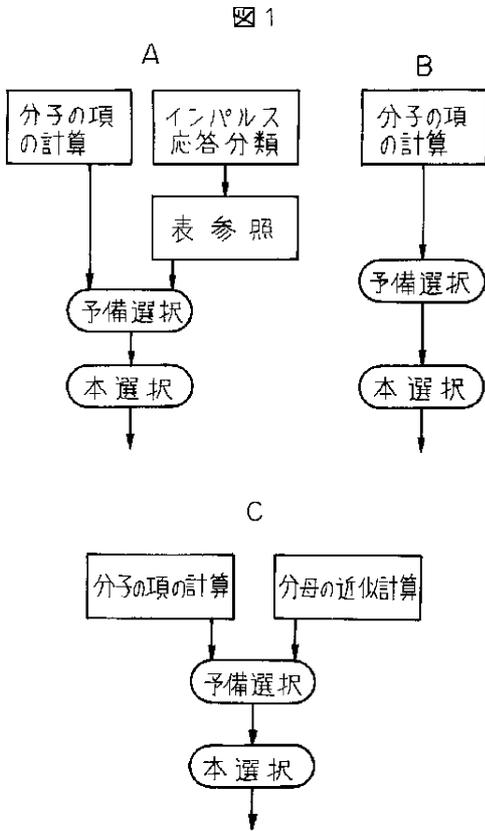
【図2】



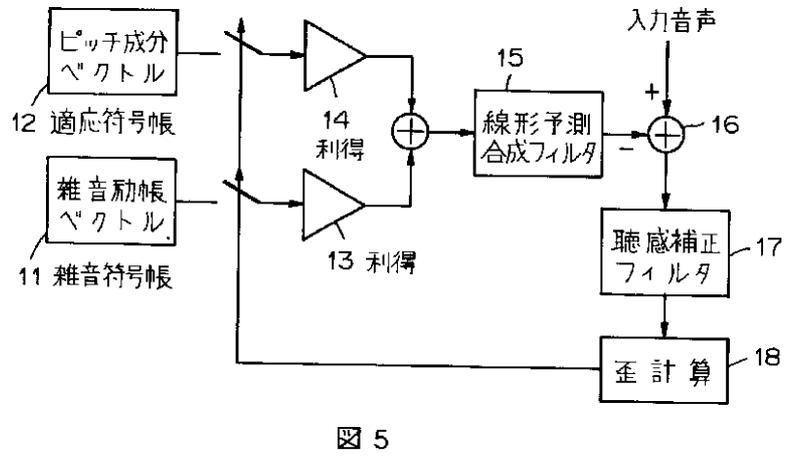
【図3】



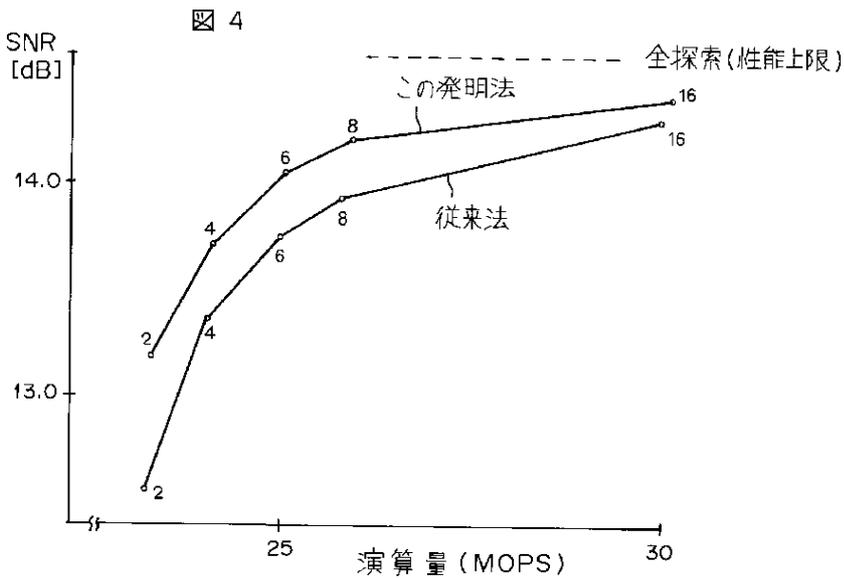
【図1】



【図5】



【図4】



フロントページの続き

(58)調査した分野(Int.Cl.<sup>7</sup>, DB名)