

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第3963850号
(P3963850)

(45) 発行日 平成19年8月22日(2007.8.22)

(24) 登録日 平成19年6月1日(2007.6.1)

(51) Int. Cl. F I
G 1 O L 11/00 (2006.01) G 1 O L 11/00 1 O 1 E
G 1 O L 11/02 (2006.01) G 1 O L 11/00 4 O 2 L
 G 1 O L 11/02

請求項の数 3 (全 27 頁)

(21) 出願番号	特願2003-64643 (P2003-64643)	(73) 特許権者	000005223 富士通株式会社
(22) 出願日	平成15年3月11日(2003.3.11)		神奈川県川崎市中原区上小田中4丁目1番1号
(65) 公開番号	特開2004-272052 (P2004-272052A)	(74) 代理人	100092152 弁理士 服部 毅麿
(43) 公開日	平成16年9月30日(2004.9.30)	(72) 発明者	大谷 猛 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内
審査請求日	平成16年10月26日(2004.10.26)	(72) 発明者	鈴木 政直 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内
		(72) 発明者	大田 恭士 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

最終頁に続く

(54) 【発明の名称】 音声区間検出装置

(57) 【特許請求の範囲】

【請求項 1】

音声区間の検出を行う音声区間検出装置において、
 入力信号の周波数分布を算出する周波数分布算出部と、
 周波数分布から周波数分布の平坦さを算出する平坦さ算出部と、
 周波数分布の平坦さとしきい値とを比較して、音声と雑音の判定を行い、入力信号の音声区間を検出する音声/雑音判定部と、

を有し、

前記平坦さ算出部は、前記周波数分布の最大値を求め、前記周波数分布と最大値との差分の総和を、前記周波数分布の平坦さとする、

ことを特徴とする音声区間検出装置。

【請求項 2】

音声区間の検出を行う音声区間検出装置において、
 入力信号の周波数分布を算出する周波数分布算出部と、
 周波数分布から周波数分布の平坦さを算出する平坦さ算出部と、
 周波数分布の平坦さとしきい値とを比較して、音声と雑音の判定を行い、入力信号の音声区間を検出する音声/雑音判定部と、

を有し、

前記平坦さ算出部は、前記周波数分布の隣接帯域間の差分の総和を、前記周波数分布の平坦さとする、

ことを特徴とする音声区間検出装置。

【請求項 3】

音声区間の検出を行う音声区間検出装置において、
入力信号の周波数分布を算出する周波数分布算出部と、
周波数分布から周波数分布の平坦さを算出する平坦さ算出部と、
周波数分布の平坦さとしきい値とを比較して、音声と雑音の判定を行い、入力信号の音
声区間を検出する音声 / 雑音判定部と、
を有し、
前記平坦さ算出部は、前記周波数分布から平均値を求め、前記平均値からしきい値を生
成し、前記周波数分布のうち前記しきい値を超える帯域数を前記周波数分布の平坦さとす
る、

10

ことを特徴とする音声区間検出装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、音声区間検出装置に関し、特に音声区間または雑音区間を検出する音声区間検出装置に関する。

【0002】

【従来の技術】

近年、携帯電話機をはじめとする移動体通信の加入者数は、爆発的に増加している。また、携帯電話機の高機能化が進んでおり、モバイル分野におけるマルチメディアサービスへの発展が期待されている。

20

【0003】

移動体通信などの音声処理の技術として、VOX (Voice Operated Transmitter)、ノイズキャンセラがある。VOXとは、音声の有無に応じて送信信号出力のON/OFFを行う技術のことで(例えば、音声を検出したときのみ信号を発信し、装置周辺が無音の時は信号を発信しないなど)、送信部の省電力化を図ることができる。また、ノイズキャンセラは、装置周辺の雑音を抑圧して、通話中に音声を聴こえやすくする技術のことである。

【0004】

これらVOXやノイズキャンセラでは、通話中に音声が存在する区間(音声区間)または雑音区間を検出する必要がある。音声区間の検出としては、例えば、入力信号の電力を算出し、電力の大きい区間を音声区間として扱うこともあるが、単純な電力の比較だけでは誤検出が多くなる。

30

【0005】

この対策として、従来、入力音声を一定の時間毎に、電力と周波数特性形状とを抽出し、前フレームの電力及び周波数特性形状から現フレームへの変化量を計測し、判定部でしきい値と比較することで音声の有無を検出する技術が提案されている(例えば、特許文献1)。

【0006】

また、入力信号の極性反転回数(零交差数)を計測し、このピッチ情報を判定部でしきい値と比較することで音声の有無を検出する技術が提案されている(例えば、特許文献2)。

40

【0007】

【特許文献 1】

特開昭60-200300号公報(第3頁-第6頁,第5図)

【特許文献 2】

特開平1-286643号公報(第3頁-第4頁,第1図)

【0008】

【発明が解決しようとする課題】

50

しかし、上記のような従来技術（特開昭60-200300号公報）では、環境騒音が大きい場合や音声小さい場合などには、雑音区間と音声区間との音声特徴量の差が小さくなり、音声区間と無音区間を精度よく判定することは困難であった。また、従来技術（特開平1-286643号公報）では、入力信号に低周波の雑音が含まれる場合、極性反転回数は低周波の雑音の電力に応じて変化してしまうので、音声区間と無音区間を精度よく判定することは困難であった。

【0009】

本発明はこのような点に鑑みてなされたものであり、音声区間を高精度に検出して、通話品質の向上を図った音声区間検出装置を提供することを目的とする。

【0010】

【課題を解決するための手段】

本発明では上記課題を解決するために、図1に示すような、音声区間の検出を行う音声区間検出装置10において、入力信号の周波数分布を算出する周波数分布算出部11と、周波数分布から周波数分布の平坦さを算出する平坦さ算出部12と、周波数分布の平坦さとしきい値とを比較して、音声と雑音の判定を行い、入力信号の音声区間を検出する音声/雑音判定部13と、を有し、平坦さ算出部12は、周波数分布の最大値を求め、周波数分布と最大値との差分の総和を、周波数分布の平坦さとする、ことを特徴とする音声区間検出装置10が提供される。

【0011】

ここで、周波数分布算出部11は、入力信号の周波数分布を算出する。平坦さ算出部12は、周波数分布から周波数分布の平坦さを算出する。音声/雑音判定部13は、周波数分布の平坦さとしきい値とを比較して、音声か雑音かを判定し、入力信号の音声区間を検出する。また、平坦さ算出部12は、周波数分布の最大値を求め、周波数分布と最大値との差分の総和を、周波数分布の平坦さとする。

【0012】

【発明の実施の形態】

以下、本発明の実施の形態を図面を参照して説明する。図1は音声区間検出装置の原理図である。音声区間検出装置10は、信号中の音声が存在する区間である音声区間を検出する装置である。

【0013】

周波数分布算出部11は、入力信号（音声、雑音を含む）から電力の周波数分布を算出する。平坦さ算出部12は、電力の周波数分布から周波数分布の平坦さ（平坦度合い）を算出する。なお、周波数分布とは、信号の周波数軸上における電力の分布状態のことを指す。

【0014】

音声/雑音判定部13は、周波数分布の平坦さと、しきい値とを比較して、音声か雑音かを判定し、入力信号の音声区間を検出する。ここで、周波数分布の平坦さが強い（周波数分布が平坦に近い）場合は、その部分は雑音とみなすことができ、周波数分布の平坦さが弱い（周波数分布が平坦でない）場合は、その部分は音声とみなすことができる。

【0015】

音声区間検出装置10では、入力信号の電力の周波数分布の平坦さにもとづき、測定区間が音声であるか雑音であるかを判定することで、高精度の音声区間の検出を行うものである。

【0016】

次に周波数分布算出部11について説明する。周波数分布算出部11は、入力信号の各フレームに対して、周波数帯域毎の電力（電力の周波数分布）を求める。この場合、フレーム毎に周波数分析を行う方法と、バンドパスフィルタ（帯域通過フィルタ）を利用して1フレームを帯域分割し、分割された帯域毎の信号から電力を算出する方法とがある（どちらを用いてもよい）。まず、周波数分析を行う方法について説明する。

【0017】

10

20

30

40

50

周波数分析によって、電力の周波数分布を算出する方法としては、高速フーリエ変換 (FFT: Fast Fourier Transform) やウェーブレット (Wavelet) 変換を用いることができる。以下、FFTの場合について説明する。

【0018】

時系列の信号にフーリエ変換を施すと、周波数領域に変換されて、該当周波数に対するスペクトルが求まる。ここで、時系列の入力データ (1フレーム) x をFFTして、周波数空間上のデータ X に変換したとする。 k を周波数、 N を全周波数帯域数とすると、 $X = \{X[k] \mid k = 1, 2, \dots, N\}$ と表せる。また、周波数 k に対応する電力を $P[k]$ とする。

【0019】

図2は電力 $P[k]$ を示す図である。FFT後の $X[k]$ は、複素数値を含む関数であるから、リアルパート (実数領域) とイマジナリパート (複素数領域) からなり、 $X[k]$ は実軸 Re と虚軸 Im 上の複素平面上にプロットすることができる。このとき、 $X[k]$ の原点からの距離の2乗が、 $X[k]$ の電力 $P[k]$ となる。したがって、周波数 k に対応する電力 $P[k]$ は、次式から求められる。

【0020】

【数1】

$$P[k] = (\text{Re}(X[k]))^2 + (\text{Im}(X[k]))^2$$

$$(k = 1, 2, \dots, N) \quad \text{---- (1)}$$

【0021】

次にバンドパスフィルタにより入力信号を帯域分割して電力を算出する場合について説明する。図3は帯域分割による電力算出の概念を示す図である。入力信号の1フレームに対し、複数のバンドパスフィルタで複数の周波数帯域に分割する。例えば、周波数帯域を N 分割するものとして (図中の i は帯域分割番号であり、 $1 \leq i \leq N$)、周波数帯域 $k = 1 \sim k = N$ の N 個のバンドパスフィルタでフィルタリングを施し、フィルタ出力としてそれぞれの信号 $x_{\text{bpf}}[i]$ を取り出す。そして、分割後の各周波数帯域の電力 $P[k]$ を求めることで、電力の周波数分布を取得する。

【0022】

バンドパスフィルタには、FIR (Finite Impulse Response) フィルタを用いる。ここで、入力信号を $x[n]$ 、各帯域に分割するバンドパスフィルタ係数 (フィルタの特性を決める係数) を $b_{\text{bpf}}[i][j]$ とすると、帯域分割後の信号 $x_{\text{bpf}}[i][n]$ は次式で表せる。なお、 i は帯域分割番号、 j はサンプリング番号であり、 n は時間に対応する添え字である。

【0023】

【数2】

$$x_{\text{bpf}}[i][n] = \sum_j b_{\text{bpf}}[i][j] * x[n-j]$$

$$\text{---- (2)}$$

【0024】

図4は式(2)の内容を説明するための図である。図に示す波形に対し、信号 $x[n]$ のサンプリング番号 j が0のときの信号の値は、 $x[n-0] = 0$ である。また、 $j = 1$ のときの信号の値は $x[n-1] = -1$ であり、 $j = 2$ のときの信号の値は $x[n-2] = 1, \dots$ である。

【0025】

さらに、バンドパスフィルタ係数 $b_{\text{bpf}}[i][j]$ に対し、 $j = 0$ のとき $b_{\text{bpf}}[i][0] = 1$ 、 $j = 1$ のとき $b_{\text{bpf}}[i][1] = 1$ 、 $j = 2$ のとき $b_{\text{bpf}}[i][2] = 0, \dots$ とする。

10

20

30

40

50

【 0 0 2 6 】

F I Rフィルタの出力 $x_{bpf}[i][n]$ は、サンプリングポイントの信号値にフィルタ係数を乗算した値の総和であるから、一般式は式(2)となり、この例の場合では、図中に示すような計算が行われることになる。

【 0 0 2 7 】

なお、バンドパスフィルタの周波数特性を決める場合には、以下の式(3)で求めることができる。

【 0 0 2 8 】

【 数 3 】

$$\text{amp}_{BPF}[i][k] = \sqrt{(\text{real}[i][k])^2 + (\text{imag}[i][k])^2} \quad \text{---- (3)} \quad 10$$

【 0 0 2 9 】

ただし、式(3)中の $\text{real}[i][k]$ と $\text{imag}[i][k]$ は、式(4a)、(4b)で示される。

【 0 0 3 0 】

【 数 4 】

$$\text{real}[i][k] = \sum_j \left(b_{p f}[i][j] * \cos\left(\frac{2\pi \cdot k \cdot j}{N}\right) \right) \quad \text{--- (4 a)} \quad 20$$

$$\text{imag}[i][k] = \sum_j \left(b_{p f}[i][j] * \sin\left(\frac{2\pi \cdot k \cdot j}{N}\right) \right) \quad \text{--- (4 b)} \quad 30$$

【 0 0 3 1 】

図5はバンドパスフィルタの周波数特性の例を示す図である。縦軸は利得、横軸は周波数であり、実線が1つのバンドパスフィルタの特性を示している。バンドパスフィルタは i 個用いるので、点線で示すバンドパスフィルタと合わせてフィルタリングを行うことになる。

【 0 0 3 2 】

一方、バンドパスフィルタによって取り出した帯域毎の電力 $P[k]$ は、 i を k に置き換えた $x_{bpf}[k][n]$ ($k = 1, 2, \dots, N$: N は全周波数帯域数) の自乗和であるから式(5)で求めることができる。

【 0 0 3 3 】

【 数 5 】

$$P[k] = \sum_n (x_{bpf}[k][n])^2 \quad (k = 1, 2, \dots, N) \quad \text{---- (5)} \quad 50$$

【 0 0 3 4 】

以上、周波数分析による算出方法とバンドパスフィルタを利用した算出方法とを説明した。いずれかの方法で求めた電力の周波数分布の例を図 6 に示す。

次に平坦さ算出部 1 2 について説明する。平坦さ算出部 1 2 は、周波数分布算出部 1 1 で求めた電力の周波数分布から周波数分布の平坦さを算出する。平坦さの算出には以降に示す〔 1 〕～〔 1 1 〕の求め方があり、いずれを選んでもよい。また、平坦さを算出するための帯域は、1 フレーム中のすべての帯域を対象にしても、または 1 フレーム中の特定箇所の帯域を対象にしてもかまわない。

〔 1 〕周波数分布の平均を求め、周波数分布（周波数分布の電力）と平均値との差分の総和を、周波数分布の平坦さとする。図 7 は周波数分布と平均値との差分の総和から平坦さを求める際の概要を説明するための図である。グラフの横軸は周波数 k 、縦軸は電力 $P[k]$ であり、信号 $X 1$ の電力の周波数分布 $R 1$ を示している。また、周波数分布 $R 1$ の電力の平均値を P_m とする。なお、横軸上の L は周波数帯域の下限値、 M は周波数帯域の上限値である。

10

【 0 0 3 5 】

周波数分布と平均値との差分を $d[k]$ とする。例えば、周波数 $k 1$ のときの差分 $d[k 1]$ は $|P[k 1] - P_m|$ である。同様に周波数 $k 2$ のときの差分 $d[k 2]$ は $|P[k 2] - P_m|$ であり、周波数 $k 3$ のときの差分 $d[k 3]$ は $|P[k 3] - P_m|$ である。したがって、 L から M の間の信号 $X 1$ に対する、周波数分布 $R 1$ と平均値 P_m との差分の総和は、図に示す斜線部の面積とほぼ等しい（離散値による総和なので）ことがわかる。そして

20

【 0 0 3 6 】

上記のことを式で表すと平均値 P_m は、以下の式（ 6 ）で求まる。 L は周波数帯域の下限値、 M は周波数帯域の上限値、 ave は平均算出を示す。また、周波数分布の平坦さを求める式は式（ 7 ）となる。

【 0 0 3 7 】

【 数 6 】

$$P_m = \text{ave}_{k=L}^M (P[k]) \quad \text{---- (6)}$$

30

【 0 0 3 8 】

【 数 7 】

$$FLT = \sum_{k=L}^M \left(|P[k] - P_m| \right) \quad \text{---- (7)}$$

【 0 0 3 9 】

このような周波数分布の平坦さを計算することで、音声区間と雑音区間との判別を行うことができる。以下、周波数分布の平坦さ及び音声 / 雑音区間の関係について説明する。一般に音声はスペクトル包絡やピッチ構造を有し、周波数分布が一様でないことが知られている。

40

【 0 0 4 0 】

スペクトル包絡とは、声の音色を示すもので、声道（声帯から口までの器官）の形状により生じる性質である。声道の形状に応じて音色が変わるのは、形状に対応した伝達特性が変わること、声道での共鳴の仕方が変わり、周波数的にエネルギーの強弱が生じるからである。

【 0 0 4 1 】

また、ピッチ構造とは、声の高さを示すもので、声帯の振動周期により生じる性質であ

50

る。ピッチ構造が時間的に変化することでアクセントやイントネーションなどの声の性質を付与することになる。一方、環境雑音は、白色雑音やピンク雑音などで近似されることが多いように、比較的周波数分布が一様であることが知られている。

【 0 0 4 2 】

したがって、ある区間における周波数分布を測定したとき、音声が存在する区間の周波数分布は平坦になりやすく、雑音が存在する区間の周波数分布は平坦になりやすいといえる。本発明では、音声と雑音に対するこれらの特徴を利用して、音声区間の検出を行うものである。

【 0 0 4 3 】

図 8 は信号の周波数分布を示す図である。横軸は周波数 k 、縦軸は電力 $P[k]$ である。信号 X_2 の電力の周波数分布 R_2 を示している。また、周波数分布 R_2 の電力の平均値を P_{m2} とする。信号 X_2 の周波数帯域毎の電力 $P[k]$ は、平均値 P_{m2} の近傍に集中して存在している（信号 X_2 は雑音とみなせる）。そして、信号 X_2 の周波数分布における周波数分布と平均値の差分の総和は、図中の斜線部の面積となり、この面積を信号 X_2 の平坦さ FLT_2 とする。

10

【 0 0 4 4 】

ここで図 7 で上述した信号 X_1 の平坦さ FLT_1 と、図 8 の信号 X_2 の平坦さ FLT_2 とを比較すれば、あきらかに $FLT_1 > FLT_2$ である。したがって、この場合、 FLT_1 を求めた際の信号 X_1 は音声であり、 FLT_2 を求めた際の信号 X_2 は雑音として判別することができる。

20

【 0 0 4 5 】

このように、算出した平坦さ FLT の値（この例では面積）が大きいほど平坦さが弱く（周波数分布が平坦でない）、平坦さ FLT の値が小さいほど平坦さが強い（周波数分布が平坦である）ので、周波数分布の平坦さを求めて比較することで、音声区間の検出を行うことが可能になる（なお、実際には、周波数分布の平坦さと、あらかじめ設定してあるしきい値とを音声 / 雑音判定部 13 で比較することで音声区間を判別することになる）。

〔 2 〕周波数分布の平均を求め、周波数分布と平均値との差分の自乗和を、周波数分布の平坦さとする。図 9 は周波数分布と平均値との差分の自乗和から、平坦さを求める際の概要を説明するための図である。グラフの横軸は周波数 k 、縦軸は電力 $P[k]$ であり、信号 X_1 の電力の周波数分布 R_1 を示している。周波数分布と平均値との差分の自乗和を求めるということは、平均値から周波数分布へ向かうベクトルの長さを求めることである。

30

【 0 0 4 6 】

例えば、周波数 k_1 のとき、平均値 m_1 、周波数分布上の電力 $P[m_1]$ とし、周波数 k_2 のとき、平均値 $m_2 (= m_1)$ 、周波数分布上の電力 $P[m_2]$ とする。そして、 m_1 を x 軸、 m_2 を y 軸方向にとって、 (m_1, m_2) と $(P[m_1], P[m_2])$ をプロットすると、図のようなベクトル v となり、ベクトル v の距離は $((P[m_1] - m_1)^2 + (P[m_2] - m_2)^2)^{1/2}$ となる。これらのことを全周波数帯域数の N まで繰り返してベクトルの距離の総和を求め、これを平坦さ FLT とする。上記のことを式で表すと以下の式 (8) となる。なお、式 (8) ではルートは除いてある（大小関係がわかればよいので）。また、このように算出した平坦さは、音声区間の平坦さを FLT_v 、雑音区間の平坦さを FLT_n とすれば $FLT_v > FLT_n$ である。

40

【 0 0 4 7 】

【 数 8 】

$$FLT = \sum_{k=L}^M \left(\left| P[k] - P_m \right| \right)^2 \quad \text{---- (8)}$$

【 0 0 4 8 】

50

〔 3 〕周波数分布の平均を求め、周波数分布と平均値との差分の最大値を、周波数分布の平坦さとする。図 1 0 は周波数分布と平均値との差分の最大値から平坦さを求める際の概要を説明するための図である。グラフの横軸は周波数 k 、縦軸は電力 $P[k]$ であり、信号 $X 1$ の電力の周波数分布 $R 1$ 及び信号 $X 2$ の電力の周波数分布 $R 2$ を示している。

【 0 0 4 9 】

図の場合、周波数分布 $R 1$ では、信号 $X 1$ の周波数分布 $R 1$ と平均値との差分の最大値は、周波数 $k a$ のとき $M A X a$ である。また、周波数分布 $R 2$ では、信号 $X 2$ の周波数分布 $R 2$ と平均値との差分の最大値は、周波数 $k b$ のとき $M A X b$ である。そして、これら $M A X a$ 、 $M A X b$ を周波数分布の平坦さ $F L T$ とする。上記のことを式で表すと以下の式 (9) となる。なお、このように算出した平坦さは、音声区間の平坦さを $F L T v$ 、雑音区間の平坦さを $F L T n$ とすれば $F L T v > F L T n$ である。

【 0 0 5 0 】

【 数 9 】

$$FLT = \max_{k=L}^M \left(\left| P[k] - P_m \right| \right) \quad \text{---- (9)}$$

【 0 0 5 1 】

〔 4 〕周波数分布の最大を求め、周波数分布と最大値との差分の総和を、周波数分布の平坦さとする。図 1 1 は周波数分布と最大値との差分の総和から平坦さを求める際の概要を説明するための図である。グラフの横軸は周波数 k 、縦軸は電力 $P[k]$ であり、信号 $X 1$ の電力の周波数分布 $R 1$ 及び信号 $X 2$ の電力の周波数分布 $R 2$ を示している。また、 $P_{MAX 1}$ 、 $P_{MAX 2}$ は、それぞれの最大値である。

【 0 0 5 2 】

上述の〔 1 〕～〔 3 〕までは周波数分布の平均値を基準にして平坦さを求めたが、〔 4 〕は周波数分布の最大値を基準にして平坦さを求めるものである (以下の〔 5 〕、〔 6 〕も同様) 。

【 0 0 5 3 】

周波数分布と最大値との差分の総和は、図に示す斜線部の面積であり、この面積を平坦さ $F L T$ とする。電力の周波数分布の最大値 P_{MAX} は以下の式 (1 0) で求め、周波数分布と最大値との差分の総和である平坦さ $F L T$ は以下の式 (1 1) で求まる。なお、このように算出した平坦さは、音声区間の平坦さを $F L T v$ 、雑音区間の平坦さを $F L T n$ とすれば $F L T v > F L T n$ である。

【 0 0 5 4 】

【 数 1 0 】

$$P_{MAX} = \max_{k=L}^M (P[k]) \quad \text{--- (10)}$$

【 0 0 5 5 】

【 数 1 1 】

$$FLT = \sum_{k=L}^M \left(\left| P[k] - P_{MAX} \right| \right) \quad \text{--- (11)}$$

【 0 0 5 6 】

〔 5 〕周波数分布の最大を求め、周波数分布と最大値との差分の自乗和を、周波数分布の平坦さとする。〔 2 〕では周波数分布と平均値との差分の自乗和を、周波数分布の平坦さ

10

20

30

40

50

としたが、〔 5 〕では平均値を最大値としたものであり、考え方は〔 2 〕と同様なので概要説明は省略する。〔 5 〕によって平坦さを求める際の式は以下の式 (1 2) となる。

【 0 0 5 7 】

【 数 1 2 】

$$FLT = \sum_{k=L}^M \left(\left| P[k] - P_{MAX} \right| \right)^2 \quad \text{--- (12)}$$

【 0 0 5 8 】

10

〔 6 〕周波数分布の最大を求め、周波数分布と周波数分布の最大値との差分の最大値を、周波数分布の平坦さとする。〔 3 〕では、周波数分布と平均値との差分の最大値を、周波数分布の平坦さとしたが、〔 6 〕では平均値を最大値としたものであり、考え方は〔 3 〕と同様なので概要説明は省略する。〔 6 〕によって平坦さを求める際の式は以下の式 (1 3) となる。

【 0 0 5 9 】

【 数 1 3 】

$$FLT = \max_{k=L}^M \left(\left| P[k] - P_{MAX} \right| \right) \quad \text{--- (13)}$$

20

【 0 0 6 0 】

〔 7 〕周波数分布の隣接帯域間の差分の総和を、周波数分布の平坦さとする。図 1 2 は周波数分布の隣接帯域間の差分の総和から平坦さを求める際の概要を説明するための図である。グラフの横軸は周波数 k 、縦軸は電力 $P[k]$ であり、信号 X_1 の電力の周波数分布 R_1 を示している。

【 0 0 6 1 】

例えば、周波数 k_1 と周波数 k_2 の電力差分は d_1 、周波数 k_2 と周波数 k_3 の電力差分は d_2 、周波数 k_3 と周波数 k_4 の電力差分は d_3 というように、隣接帯域間の差分を求め、この差分の総和を平坦さ FLT とする。このことを式で表すと以下の式 (1 4) となる。

30

【 0 0 6 2 】

なお、このように算出した平坦さは、音声区間の平坦さを FLT_v 、雑音区間の平坦さを FLT_n とすれば $FLT_v > FLT_n$ である (音声は周波数の電力変動が大きく、雑音は周波数の電力変動が小さいので、〔 7 〕により算出した平坦さで音声 / 雑音の判別を行うことができる)。

【 0 0 6 3 】

【 数 1 4 】

$$FLT = \sum_{k=L}^{M-1} \left(\left| P[k] - P[k+1] \right| \right) \quad \text{--- (14)}$$

40

【 0 0 6 4 】

〔 8 〕周波数分布の隣接帯域間の差分の最大値を、周波数分布の平坦さとする。図 1 3 は周波数分布の隣接帯域間の差分の最大値から平坦さを求める際の概要を説明するための図である。グラフの横軸は周波数 k 、縦軸は電力 $P[k]$ であり、信号 X_1 の電力の周波数分布 R_1 を示している。

【 0 0 6 5 】

50

例えば、周波数 k_5 と周波数 k_6 の差分 d_{max} が、全周波数帯域における最大値であり、これを平坦さ FLT とする。このことを式で表すと以下の式 (15) となる。なお、このように算出した平坦さは、音声区間の平坦さを FLT_v 、雑音区間の平坦さを FLT_n とすれば $FLT_v > FLT_n$ である。

【0066】

【数15】

$$FLT = \max_{k=L}^{M-1} \left(\left| P[k] - P[k+1] \right| \right) \quad \text{--- (15)}$$

10

【0067】

〔9〕周波数分布の平坦さを周波数分布の平均で除算する、またはフレームの平均電力で除算して、除算（正規化）した結果を平坦さとする。〔9〕では、上述の〔1〕～〔8〕で求めた平坦さを、さらに周波数分布の平均値またはフレームの平均電力で除算して、除算した値を平坦さとするものである。

【0068】

音声には、大きい音（声）、小さい音があるので、例えば、〔8〕のような隣接帯域間の差分の最大値を周波数分布の平坦さとする、大きい音声の隣接帯域間の差分の最大値の方が、小さい音声のそれよりも大きくなる。平坦さの算出と全体の音量とは関係ないので、平坦さを算出する際に音量に依存しないようにするには、〔1〕～〔8〕で求めた平坦さを、その平坦さを求めたときの音の大きさ（周波数分布の平均値またはフレームの平均電力）で除算して正規化すれば、音の大きさによらない処理ができ、さらに高精度に平坦さを算出することが可能になる。

20

〔10〕周波数分布から平均値を求め、この平均値に定数を乗算または加算した値をしきい値とし、周波数分布のうちしきい値を超える帯域数を周波数分布の平坦さとする。図14は周波数分布の平均値から求めたしきい値を用いて平坦さを求める際の概要を説明するための図である。グラフの横軸は周波数 k 、縦軸は電力 $P[k]$ であり、信号 X_1 の電力の周波数分布 R_1 と信号 X_2 の周波数分布 R_2 を示している。

【0069】

30

周波数分布 R_1 の平均値を P_{m1} とし、電力 P_{m1} に定数を乗算または加算して生成したしきい値を t_{h1} とする。また、周波数分布 R_2 の平均値を P_{m2} とし、電力 P_{m2} に定数を乗算または加算して生成したしきい値を t_{h2} とする。

【0070】

周波数分布 R_1 に対し、しきい値 t_{h1} が図の位置にあるとする。この場合、しきい値 t_{h1} と周波数帯域の電力との比較を行い、電力がしきい値 t_{h1} を上回る帯域数を数え、この個数を信号 X_1 の周波数分布 R_1 の平坦さ FLT_1 とする。

【0071】

また、周波数分布 R_2 に対し、しきい値 t_{h2} が図の位置にあるとする。この場合、しきい値 t_{h2} と周波数帯域の電力との比較を行い、電力がしきい値 t_{h2} を上回る帯域数を数え、この個数を信号 X_2 の周波数分布 R_2 の平坦さ FLT_2 とする。

40

【0072】

図からわかるように、 $FLT_1 < FLT_2$ である。すなわち、しきい値を上回る帯域数が多いほど周波数分布の平坦さは強く、その信号は雑音とみなすことができる（〔1〕～〔9〕の場合は、音声区間の平坦さを FLT_v 、雑音区間の平坦さを FLT_n とすれば $FLT_v > FLT_n$ であったが、〔10〕の場合は $FLT_v < FLT_n$ となることに注意）。

【0073】

これらのことを式で表すと以下の式 (16) で平坦さが求まる。式中の $count$ とは、括弧内の条件を満たした事象を数える手段を表す。また、しきい値を求める式は式 (1

50

7 a)、(17 b)である。なお、COEFFは乗算用定数、CONSTは加算用定数である。

【0074】

【数16】

$$FLT = \text{count}_{k=L}^{M-1} \left(P[k] > THR \right) \quad \text{--- (16)}$$

【0075】

【数17】

$$THR = P_m * COEFF \quad \text{--- (17 a)}$$

$$THR = P_m + CONST \quad \text{--- (17 b)}$$

【0076】

〔11〕周波数分布から最大値を求め、この最大値に定数を乗算または加算した値をしきい値とし、周波数分布のうちしきい値を超える帯域数を周波数分布の平坦さとする。〔10〕では周波数分布から平均値を求め、この平均値からしきい値を生成したが、〔11〕では周波数分布から最大値を求め、この最大値からしきい値を生成して、しきい値を超える帯域数を周波数分布の平坦さとするものであり、考え方は〔10〕と同様なので概要説明は省略する。〔11〕によって平坦さを求める際の式は以下の式(18)となり、しきい値の算出式は式(19 a)、(19 b)となる。

【0077】

【数18】

$$FLT = \text{count}_{k=L}^{M-1} \left(P[k] > THR \right) \quad \text{--- (18)}$$

【0078】

【数19】

$$THR = P_{MAX} * COEFF \quad \text{--- (19 a)}$$

$$THR = P_{MAX} + CONST \quad \text{--- (19 b)}$$

【0079】

次に音声/雑音判定部13について説明する。音声/雑音判定部13では、平坦さ算出部12によって上述の〔1〕～〔11〕のいずれかから求めた周波数分布の平坦さに対し、あらかじめ用意しておいたしきい値との比較を行うことで、その区間における信号が音声であるか雑音であるかを判定し、判定に応じたフラグを出力する。

【0080】

図15は音声区間、雑音区間の判定処理例を示す図である。縦軸は電力、横軸はフレーム(時間)である。音声/雑音判定部13は、しきい値THによって、図に示すように音声区間、雑音区間を判別する。

【0081】

次に音声区間検出装置を適用した具体的な装置例について説明する。図16はVOX装置の構成を示す図である。VOX装置20は、区間毎に入力信号を分析し、音声の有無を判定し、判定結果に応じて送信出力のON/OFFを行うことで送信部の省電力化を図る

10

20

30

40

50

装置である。なお、この装置では電力の周波数分布を求めるためにFFTを用い、式(7)で周波数分布の平坦さを求め、かつ正規化を行っている例を示す。

【0082】

V O X装置20は、マイク21、A/D部22、音声区間検出部23(図1の音声区間検出装置10に該当)、エンコーダ24、送信部25から構成される。音声区間検出部23は、FFT部23a、電力スペクトル算出部23b、平均値算出部23c、差分算出部23d、差分総和算出部23e、正規化部23f、音声/雑音判定部23gから構成される。なお、FFT部23a、電力スペクトル算出部23bは、図1の周波数分布算出部11に該当し、平均値算出部23c、差分算出部23d、差分総和算出部23e、正規化部23fは、図1の平坦さ算出部12に該当し、音声/雑音判定部23gは、図1の音声/雑音判定部13に該当する。

10

〔S1〕マイク21から入力された音声が入力されたA/D部22にてデジタル信号に変換され、入力が得られる。

〔S2〕FFT部23aは、FFTを用いて、一定時間(フレーム)毎に入力信号を周波数分析する。

〔S3〕電力スペクトル算出部23bは、各フレーム毎に得られた入力信号の周波数分析結果から電力を求めることで電力スペクトル(周波数分布)を得る。

〔S4〕平均値算出部23cは、電力スペクトルの平均を算出する(式(6)により)。

〔S5〕差分算出部23dは、電力スペクトルから電力スペクトルの平均の差分を算出し、差分総和算出部23eは、差分の総和を算出して平坦さを求める(式(7)により)。

20

〔S6〕正規化部23fは、平坦さを電力スペクトルの平均で除算して正規化する。

〔S7〕音声/雑音判定部23gは、各フレーム毎に得られる平坦さと、あらかじめ用意しておいたしきい値とを比較することで、該当フレームが音声であるか雑音であるかを判定し、判定結果(フラグ)を出力する。例えば、受信した平坦さがしきい値以上では音声フラグを、しきい値以下では雑音フラグを出力する。

〔S8〕エンコーダ24は、入力信号に対して音声符号化を行い、符号データを出力する。

〔S9〕送信部25は、エンコーダ24より得られる符号データと、音声/雑音判定部23gより得られる判定フラグを受け取り、音声フラグの場合、判定フラグと符号データを送信し、雑音フラグの場合、判定フラグのみを送信する。

30

【0083】

一般に、携帯電話機では、信号を送信するために大きな電力を消費するが、上記のV O X装置20を用いることで、雑音判定時には符号データを送信しないので、電力消費を抑えることができる。

【0084】

また、V O X装置20を用いることで、高精度の音声/雑音の判定を行うため、音声が含まれるフレームで雑音のフレームであると誤判定して、そのフレームの音声情報を送信しないなどといった現象を起すことがない。これにより、音切れの原因をなくすことができ、通話品質(音質)の向上を図ることが可能になる。

【0085】

40

次にノイズキャンセラ装置について説明する。図17はノイズキャンセラ装置の構成を示す図である。ノイズキャンセラとは、入力信号から雑音成分を抑圧することで、音声の明瞭度の向上を図る機能である。この機能は、雑音学習と雑音抑圧(n-1ステップ目で検出した雑音成分を用いて、nステップ目の信号に含まれる雑音を除去すること)の切り換えに利用される。なお、この装置では電力の周波数分布を求めるためにバンドパスフィルタによる帯域分割を行い、式(12)で周波数分布の平坦さを求める場合の例を示す。

【0086】

ノイズキャンセラ装置30は、信号受信部31、デコーダ32、雑音区間検出部33(図1の音声区間検出装置10に該当)、(雑音)抑圧量算出部34、雑音抑圧部35、D/A部36、スピーカ37から構成される。

50

【 0 0 8 7 】

また、雑音区間検出部 3 3 は、帯域分割部 3 3 a、狭帯域別フレームパワー算出部 3 3 b、最大値算出部 3 3 c、差分算出部 3 3 d、自乗和算出部 3 3 e、音声 / 雑音判定部 3 3 f から構成される。雑音抑圧量算出部 3 4 は、狭帯域雑音パワー推定部 3 4 a、抑圧量算出部 3 4 b から構成される。雑音抑圧部 3 5 は、抑圧部 3 5 a - 1 ~ 3 5 a - n、加算器 3 5 b から構成される。

【 0 0 8 8 】

なお、帯域分割部 3 3 a、狭帯域別フレームパワー算出部 3 3 b は、図 1 の周波数分布算出部 1 1 に該当し、最大値算出部 3 3 c、差分算出部 3 3 d、自乗和算出部 3 3 e は、図 1 の平坦さ算出部 1 2 に該当し、音声 / 雑音判定部 3 3 f は、図 1 の音声 / 雑音判定部 1 3 に該当する。

〔 S 1 1 〕デコーダ 3 2 は、信号受信部 3 1 から得られる符号化データを復号し、雑音区間検出部 3 3 へ送信する。

〔 S 1 2 〕帯域分割部 3 3 a は、フレーム毎に各帯域に分割し、狭帯域別フレームパワー算出部 3 3 b は、帯域毎のフレームパワー（周波数分布）を算出する。

〔 S 1 3 〕最大値算出部 3 3 c は、フレームパワーの最大値を算出する（式（ 1 0 ）により）。差分算出部 3 3 d は、フレームパワーからフレームパワーの最大値の差分の絶対値を求め、自乗和算出部 3 3 e は、絶対値の自乗和を求め平坦さとして出力する（式（ 1 2 ）により）。

〔 S 1 4 〕音声 / 雑音判定部 3 3 f は、フレーム毎に得られる平坦さと、あらかじめ用意しておいたしきい値とを比較することで、該当フレームが音声であるか雑音であるかを判定し、判定フラグを出力する。

〔 S 1 5 〕狭帯域雑音パワー推定部 3 4 a は、判定フラグが雑音の場合にのみ、各帯域の雑音のパワーを推定し、狭帯域雑音パワーを得る。推定の方法として、例えば、過去に雑音と判定されたフレームでの帯域毎のフレームパワーを平均する方法などがある。

〔 S 1 6 〕抑圧量算出部 3 4 b は、狭帯域雑音パワー推定部 3 4 a で得られた狭帯域雑音パワーと、狭帯域別フレームパワー算出部 3 3 b からの各帯域のフレームパワーとを比較し、帯域毎の抑圧量を算出する。例えば、各帯域において、狭帯域雑音パワーよりフレームパワーの方が小さかった場合には、抑圧量を 1 5 d B とし、それ以外の場合には 0 d B（抑圧なし）とする。

〔 S 1 7 〕抑圧部 3 5 a - 1 ~ 3 5 a - n は、帯域毎に、帯域分割部 3 3 a で得られた入力の帯域分割信号に抑圧量算出部 3 4 b で得られた抑圧量をかけることで、入力信号のうち、雑音の成分のみを抑圧する。

〔 S 1 8 〕加算器 3 5 b は、帯域毎の雑音抑圧後の信号を足し合わせる。

〔 S 1 9 〕 D / A 部 3 6 は、加算器 3 5 b より得られるデジタル信号をアナログ信号に変換し、スピーカ 3 7 は音声を出力する。

【 0 0 8 9 】

以上説明したように、ノイズキャンセラ装置 3 0 では、高精度の音声 / 雑音の判定処理を行うので、例えば、音声が含まれるフレームで雑音のフレームであると誤判定して、そのフレームの音声を抑圧してしまうなどといった現象を起すことがない。また、雑音学習の精度を落とすことがないので、雑音抑圧の性能も向上することができ、音声時に抑圧しすぎたり、音切れが発生したり、雑音が残留したりするようなことを防止できるので、通話品質の向上を図ることが可能になる。

【 0 0 9 0 】

図 1 8 はノイズキャンセラ装置の構成を示す図である。この例のノイズキャンセラ装置 4 0 は、電力の周波数分布を求めるために F F T を使用し、式（ 1 5 ）で周波数分布の平坦さを求めている。

【 0 0 9 1 】

ノイズキャンセラ装置 4 0 は、信号受信部 4 1、デコーダ 4 2、雑音区間検出部 4 3（図 1 の音声区間検出装置 1 0 に該当）、（雑音）抑圧量算出部 4 4、雑音抑圧部 4 5、D

10

20

30

40

50

/ A部46、スピーカ47から構成される。

【0092】

また、雑音区間検出部43は、FFT部43a、電力スペクトル算出部43b、隣接帯域間差分算出部43c、最大値算出部43d、音声/雑音判定部43eから構成される。雑音抑圧量算出部44は、雑音電力スペクトル推定部44a、抑圧量算出部44bから構成される。雑音抑圧部45は、抑圧部45a、IFFT (Inverse Fast Fourier Transform) 部45bから構成される。

【0093】

なお、FFT部43a、電力スペクトル算出部43bは、図1の周波数分布算出部11に該当し、隣接帯域間差分算出部43c、最大値算出部43dは、図1の平坦さ算出部12に該当し、音声/雑音判定部43eは、図1の音声/雑音判定部13に該当する。 10

〔S21〕デコーダ42は、信号受信部41から得られる符号化データを復号し、雑音区間検出部43へ送信する。

〔S22〕FFT部43aは、FFTを用いてフレーム毎に入力信号を周波数分析する。電力スペクトル算出部43bは、フレーム毎に得られた入力信号の周波数分析結果から電力を求め、電力スペクトルを求める。

〔S23〕隣接帯域間差分算出部43cは、電力スペクトルから隣接帯域間の差分を求め、最大値算出部43dは、差分の最大値を求め、これを平坦さとして出力する(式(15)により)。

〔S24〕音声/雑音判定部43eは、フレーム毎に得られる平坦さと、あらかじめ用意しておいたしきい値とを比較することで、該当フレームが音声であるか雑音であるかを判定し、判定フラグを出力する。 20

〔S25〕雑音電力スペクトル推定部44aは、音声/雑音判定部43eから得られる判定フラグが雑音の場合に、雑音の電力スペクトルの推定を更新する。

〔S26〕抑圧量算出部44bは、雑音の電力スペクトルと該当フレームの電力スペクトルとを比較することで、各帯域の抑圧量を算出する。

〔S27〕抑圧部45aは、FFT部43aで得られた周波数分析された入力信号に、抑圧量算出部44bで得られた抑圧量をかけることで、入力信号のうち、雑音の成分のみを抑圧する。IFFT部45bは、抑圧後のフーリエ変換対に逆フーリエ変換を施す。

〔S28〕D/A部46は、IFFT部45bより得られるデジタル信号をアナログ信号に変換し、スピーカ47は音声を出力する。 30

【0094】

次にトーン検出装置について説明する。図19はトーン検出装置の構成を示す図である。トーン検出機能とは、トーン信号を検出した場合には、受信信号に加工を加えず、そのまま出力し、トーン信号を検出しなかった場合にのみ、ノイズキャンセラ等の音声信号処理を行うことで、DTMF (Dual Tone-Multiple Frequency) やFAX信号を透過させるための機能である。なお、この装置では電力の周波数分布を求めるためにFFTを使用し、式(18)で周波数分布の平坦さを求める場合の例を示す。

【0095】

トーン検出装置50は、信号受信部51、デコーダ52、トーン信号検出部53、信号出力部54、D/A部55、スピーカ56から構成される。トーン信号検出部53は、FFT部53a、電力スペクトル算出部53b、最大値算出部53c、しきい値決定部53d、帯域数カウント部53e、トーン判定部53fから構成される。信号出力部54は、ノイズキャンセル部54a、IFFT部54b、スイッチ54cから構成される。 40

【0096】

なお、FFT部53a、電力スペクトル算出部53bは、図1の周波数分布算出部11に該当し、最大値算出部53c、しきい値決定部53d、帯域数カウント部53eは、図1の平坦さ算出部12に該当し、トーン判定部53fは、図1の音声/雑音判定部13に該当する。

〔S31〕デコーダ52は、信号受信部51から得られる符号化データを復号し、トーン 50

信号検出部 5 3 へ送信する。

〔 S 3 2 〕 F F T 部 5 3 a は、 F F T を用いてフレーム毎に入力信号を周波数分析する。電力スペクトル算出部 5 3 b は、フレーム毎に得られた入力信号の周波数分析結果から電力を求めることで電力スペクトルを求める。

〔 S 3 3 〕 最大値算出部 5 3 c は、電力スペクトルの最大値を求める（式（ 1 0 ）により）。しきい値決定部 5 3 d は最大値にもとづきしきい値を算出する（式（ 1 9 a ）、（ 1 9 b ）のいずれかにより）。帯域数カウント部 5 3 e は、電力スペクトルとしきい値とを比較して帯域数をカウントし、カウント結果を平坦さとして出力する（式（ 1 8 ）により）。

〔 S 3 4 〕 トーン判定部 5 3 f は、フレーム毎に得られる平坦さと、あらかじめ用意しておいたしきい値とを比較することで、該当フレームがトーン信号であるか否かを判定し、判定フラグを出力する。 10

〔 S 3 5 〕 ノイズキャンセル部 5 4 a は、 F F T 部 5 3 a によるフレーム毎に得られた入力信号の周波数分析結果に、音声処理としてノイズキャンセル処理を施し、雑音を抑圧する。 I F F T 部 5 4 b は、雑音抑圧後のフーリエ変換対に逆フーリエ変換を施す。

〔 S 3 6 〕 スイッチ部 5 4 c は、判定フラグがトーン信号の場合には、デコーダ 5 2 からの出力を選択し、判定フラグがトーン信号でない場合には、 I F F T 部 5 4 b からの出力を選択する。

〔 S 3 7 〕 D / A 部 5 5 は、スイッチ 5 4 c より得られるデジタル信号をアナログ信号に変換し、スピーカ 5 6 は音声を出力する。 20

【 0 0 9 7 】

図 2 0 はトーン信号区間の判定処理を示す図である。縦軸は電力、横軸はフレームである。図からわかるように入力信号がトーン信号の場合は明らかに周波数分布の平坦さが弱くなるので、精度よくトーン信号を検出することが可能になる。

【 0 0 9 8 】

次にエコーキャンセラ装置について説明する。図 2 1 はエコーキャンセラ装置の構成を示す図である。エコーキャンセル機能とは、受信信号に電気信号や音声の出力が入力機器に拾われて起こるエコー発生やハウリングの現象を防止する機能のことである。

【 0 0 9 9 】

エコーキャンセラ装置 6 0 は、マイク 6 1、 A / D 部 6 2、エコーキャンセル部 6 3、 30
入力音声区間検出部 6 4（図 1 の音声区間検出装置 1 0 に該当）、出力音声区間検出部 6 5（図 1 の音声区間検出装置 1 0 に該当）、符号化部 6 6、復号化部 6 7、 D / A 部 6 8、スピーカ 6 9 から構成される。また、エコーキャンセル部 6 3 は、エコーキャンセラ 6 3 a、状態制御部 6 3 b から構成され、入力音声区間検出部 6 4 は、電力スペクトル算出部 6 4 a、区間検出部 6 4 b から構成され、出力音声区間検出部 6 5 は、電力スペクトル算出部 6 5 a、区間検出部 6 5 b から構成される。

【 0 1 0 0 】

なお、入力音声区間検出部 6 4 の電力スペクトル算出部 6 4 a は、図 1 の周波数分布算出部 1 1 に該当し、区間検出部 6 4 b は図 1 の平坦さ算出部 1 2 及び音声 / 雑音判定部 1 3 に該当する。また、出力音声区間検出部 6 5 の電力スペクトル算出部 6 5 a は、図 1 の周波数分布算出部 1 1 に該当し、区間検出部 6 5 b は図 1 の平坦さ算出部 1 2 及び音声 / 雑音判定部 1 3 に該当する。 40

〔 S 4 1 〕 マイク 6 1 から入力された音声は A / D 部 6 2 にてデジタル信号に変換され、エコーキャンセラ 6 3 a 及び電力スペクトル算出部 6 4 a に入力される。

〔 S 4 2 〕 電力スペクトル算出部 6 4 a は、 F F T を行って入力音より電力スペクトルを算出し、区間検出部 6 4 b に電力スペクトルを送信する。

〔 S 4 3 〕 区間検出部 6 4 b は、電力スペクトルより、その平坦さを算出し、現フレームが音声区間であるか否かを判定し、入力音に対する判定フラグ（入力音フラグ）を状態制御部 6 3 b へ送信する。

〔 S 4 4 〕 復号化部 6 7 は、受信信号（符号データ）を復号化し、電力スペクトル算出部 50

65 a、エコーキャンセラ63 a、D/A部68へ送信する。なお、D/A部68は、出力音をアナログ音にして、スピーカ69は、アナログ音を出力する。

〔S45〕電力スペクトル算出部65 aは、出力音より電力スペクトルを算出し、区間検出部65 bに電力スペクトルを送信する。

〔S46〕区間検出部65 bは、電力スペクトルより、その平坦さを算出し、現フレームが音声区間であるか否かを判定し、出力音に対する判定フラグ（出力音フラグ）を状態制御部63 bへ送信する。

〔S47〕状態制御部63 bは、入力音及び出力音の判定フラグから入出力の状態を検知し、図22に示すテーブルT1にしたがって、制御信号をエコーキャンセラ63 aに送信する。

〔S48〕エコーキャンセラ63 aは、制御信号（減算）がONの場合、出力音にエコー経路特性をかけることで疑似エコー信号を作成し、入力音から疑似エコー信号を減算する。また、制御信号（学習）がONの場合、エコーキャンセル後の信号から、推定したエコー経路を更新する（更新されたエコー経路は、次ステップで入力音からエコーを取り除く場合の疑似エコー信号の生成に用いられる）。

〔S49〕エコーキャンセル後の信号は、符号化部66によって符号化され送信される。

【0101】

以上説明したように、エコーキャンセラ装置60は、入出力の状態を高精度に検知し、検知した状態に合わせて減算・学習の制御を行うので、検知に失敗して、異音や音切れを発生したりするようなことがなく、通話品質の向上を図ることが可能になる。

【0102】

以上説明したように、本発明によれば、フレームが音声であるか雑音であるかを判定するための物理量として、周波数分布の平坦さを利用した。これにより、簡単な計算で精度よく音声区間・雑音区間の検出が可能になる。また、電力の周波数分布にもとづき、音声/雑音区間検出を行うので、特に、入力音声の電力が小さい場合や、入力雑音の電力が大きい場合でも誤検出しにくく、効果が大きい。さらに、ノイズキャンセラなどのように、信号の周波数変換を含む音声信号処理に利用する場合には、あらたに時間-周波数変換を行う必要がないので、制御構成を簡略化することができる。

【0104】

なお、上記の説明では、音声区間検出装置10をVOX装置、ノイズキャンセラ、トーン検出装置、エコーキャンセラ装置に適用した例を示したが、これらに限らず、その他の音声処理を行う多様な装置について幅広く適用可能である。

【0105】

（付記1） 音声区間の検出を行う音声区間検出装置において、
入力信号の周波数分布を算出する周波数分布算出部と、
周波数分布から周波数分布の平坦さを算出する平坦さ算出部と、
周波数分布の平坦さとしきい値とを比較して、音声と雑音の判定を行い、入力信号の音声区間を検出する音声/雑音判定部と、
を有することを特徴とする音声区間検出装置。

【0106】

（付記2） 前記周波数分布算出部は、フレーム毎の入力信号に対する周波数分析、またはバンドパスフィルタで入力信号を帯域分割し、分割された帯域毎の信号からフレーム毎の電力算出のいずれかを行って、前記周波数分布を算出することを特徴とする付記1記載の音声区間検出装置。

【0107】

（付記3） 前記平坦さ算出部は、前記周波数分布の平均を求め、前記周波数分布と平均値との差分の総和を、前記周波数分布の平坦さとする特徴とする付記1記載の音声区間検出装置。

【0108】

（付記4） 前記平坦さ算出部は、前記周波数分布の平均を求め、前記周波数分布と平

10

20

30

40

50

均値との差分の自乗和を、前記周波数分布の平坦さとすることを特徴とする付記 1 記載の音声区間検出装置。

【 0 1 0 9 】

(付記 5) 前記平坦さ算出部は、前記周波数分布の平均を求め、前記周波数分布と平均値との差分の最大値を、前記周波数分布の平坦さとすることを特徴とする付記 1 記載の音声区間検出装置。

【 0 1 1 0 】

(付記 6) 前記平坦さ算出部は、前記周波数分布の最大を求め、前記周波数分布と最大値との差分の総和を、前記周波数分布の平坦さとすることを特徴とする付記 1 記載の音声区間検出装置。

10

【 0 1 1 1 】

(付記 7) 前記平坦さ算出部は、前記周波数分布の最大を求め、前記周波数分布と最大値との差分の自乗和を、前記周波数分布の平坦さとすることを特徴とする付記 1 記載の音声区間検出装置。

【 0 1 1 2 】

(付記 8) 前記平坦さ算出部は、前記周波数分布の最大を求め、前記周波数分布と最大値との差分の最大値を、前記周波数分布の平坦さとすることを特徴とする付記 1 記載の音声区間検出装置。

【 0 1 1 3 】

(付記 9) 前記平坦さ算出部は、前記周波数分布の隣接帯域間の差分の総和を、前記周波数分布の平坦さとすることを特徴とする付記 1 記載の音声区間検出装置。

20

【 0 1 1 4 】

(付記 10) 前記平坦さ算出部は、前記周波数分布の隣接帯域間の差分の最大値を、前記周波数分布の平坦さとすることを特徴とする付記 1 記載の音声区間検出装置。

【 0 1 1 5 】

(付記 11) 前記平坦さ算出部は、前記周波数分布の平坦さを周波数分布の平均で除算して正規化することを特徴とする付記 1 記載の音声区間検出装置。

(付記 12) 前記平坦さ算出部は、前記周波数分布の平坦さをフレームの平均電力で除算して正規化することを特徴とする付記 1 記載の音声区間検出装置。

【 0 1 1 6 】

(付記 13) 前記平坦さ算出部は、前記周波数分布から平均値を求め、前記平均値からしきい値を生成し、前記周波数分布のうち前記しきい値を超える帯域数を前記周波数分布の平坦さとすることを特徴とする付記 1 記載の音声区間検出装置。

30

【 0 1 1 7 】

(付記 14) 前記平坦さ算出部は、前記周波数分布から最大値を求め、前記最大値からしきい値を生成し、前記周波数分布のうち前記しきい値を超える帯域数を前記周波数分布の平坦さとすることを特徴とする付記 1 記載の音声区間検出装置。

【 0 1 1 8 】

(付記 15) 音声の有無に応じて送信信号出力の ON / OFF を行う VOX 装置において、

40

入力信号の周波数分布を算出する周波数分布算出部と、周波数分布から周波数分布の平坦さを算出する平坦さ算出部と、周波数分布の平坦さとしきい値とを比較して、音声か雑音かを判定し、音声区間を検出した場合は音声フラグを、雑音区間を検出した場合は雑音フラグを出力する音声 / 雑音判定部と、から構成される音声区間検出部と、

入力信号をエンコードして、符号化データを生成するエンコーダと、

前記音声フラグを受信した場合は、前記符号化データと前記音声フラグとを送信し、前記雑音フラグを受信した場合は、前記雑音フラグのみ送信する送信部と、

を有することを特徴とする VOX 装置。

【 0 1 1 9 】

(付記 16) 信号中の雑音成分を抑圧するノイズキャンセラ装置において、

50

入力信号をバンドパスフィルタを用いて帯域分割し、周波数分布を帯域毎に算出する周波数分布算出部と、周波数分布から周波数分布の平坦さを算出する平坦さ算出部と、周波数分布の平坦さとしきい値とを比較して、音声か雑音かを判定し、雑音区間を検出した場合は雑音フラグを出力する音声/雑音判定部と、から構成される雑音区間検出部と、

前記雑音フラグを受信した場合、入力信号の帯域毎の雑音パワーを推定し、前記雑音パワーと帯域毎のフレームパワーとにもとづき抑圧量を算出する抑圧量算出部と、

入力信号を帯域毎に前記抑圧量に応じて抑圧することで、入力信号のうち雑音成分のみ抑圧する雑音抑圧部と、

を有することを特徴とするノイズキャンセラ装置。

【0120】

(付記17) 信号中の雑音成分を抑圧するノイズキャンセラ装置において、

入力信号の周波数分析を行って、周波数分布を算出する周波数分布算出部と、周波数分布から周波数分布の平坦さを算出する平坦さ算出部と、周波数分布の平坦さとしきい値とを比較して、音声か雑音かを判定し、雑音区間を検出した場合は雑音フラグを出力する音声/雑音判定部と、から構成される雑音区間検出部と、

前記雑音フラグを受信した場合、入力信号の雑音の雑音電力スペクトルを推定し、前記雑音電力スペクトルとフレーム電力スペクトルとにもとづき抑圧量を算出する抑圧量算出部と、

入力信号を前記抑圧量に応じて抑圧することで、入力信号のうち雑音成分のみ抑圧する雑音抑圧部と、

を有することを特徴とするノイズキャンセラ装置。

【0121】

(付記18) トーン信号を検出するトーン検出装置において、

入力信号の周波数分布を算出する周波数分布算出部と、周波数分布から周波数分布の平坦さを算出する平坦さ算出部と、周波数分布の平坦さとしきい値とを比較して、トーン信号の有無を判定し、トーン信号を検出した場合はトーン検出フラグを出力するトーン判定部と、から構成されるトーン信号検出部と、

入力信号をデコードして、復号化データを生成するデコーダと、

前記トーン検出フラグを受信した場合は、前記復号化データを出力し、前記トーン検出フラグを受信しなかった場合は、前記復号化データに音声処理を施して出力する信号出力部と、

を有することを特徴とするトーン検出装置。

【0122】

(付記19) エコーの発生を抑止するエコーキャンセラ装置において、

入力音の周波数分布を算出する入力音周波数分布算出部と、周波数分布から周波数分布の平坦さを算出する入力音平坦さ算出部と、周波数分布の平坦さとしきい値とを比較して、音声と雑音の判定を行い、入力音の音声区間を検出した場合は入力音フラグを出力する入力音判定部と、から構成される入力音声区間検出部と、

出力音の周波数分布を算出する出力音周波数分布算出部と、周波数分布から周波数分布の平坦さを算出する出力音平坦さ算出部と、周波数分布の平坦さとしきい値とを比較して、音声と雑音の判定を行い、出力音の音声区間を検出した場合は出力音フラグを出力する出力音判定部と、から構成される出力音声区間検出部と、

前記入力音フラグと前記出力音フラグから入出力状態を認識し、入出力状態に応じて、出力音にエコー経路特性を乗算することで疑似エコー信号を生成して入力音から前記疑似エコー信号を減算する減算処理、またはエコー経路を更新する学習処理を行うエコーキャンセル部と、

を有することを特徴とするエコーキャンセラ装置。

【0123】

(付記20) 音声区間の検出を行う音声区間検出方法において、

入力信号の周波数分布を算出し、

10

20

30

40

50

周波数分布から周波数分布の平坦さを算出し、

周波数分布の平坦さとしきい値とを比較して、音声と雑音の判定を行い、入力信号の音声区間を検出することを特徴とする音声区間検出方法。

【 0 1 2 4 】

(付記 2 1) 前記周波数分布を算出する際は、フレーム毎の入力信号に対する周波数分析、またはバンドパスフィルタで入力信号を帯域分割して分割された帯域毎の信号からフレーム毎による電力算出、のいずれかを行うことを特徴とする付記 2 0 記載の音声区間検出方法。

【 0 1 2 5 】

(付記 2 2) 前記周波数分布の平坦さを算出する際は、前記周波数分布の平均を求めた後に、前記周波数分布と平均値との差分の総和、前記周波数分布と平均値との差分の自乗和、前記周波数分布と平均値との差分の最大値、のいずれかを求めることを特徴とする付記 2 0 記載の音声区間検出方法。

【 0 1 2 6 】

(付記 2 3) 前記周波数分布の平坦さを算出する際は、前記周波数分布の最大を求めた後に、前記周波数分布と最大値との差分の総和、前記周波数分布と最大値との差分の自乗和、前記周波数分布と最大値との差分の最大値、のいずれかを求めることを特徴とする付記 2 0 記載の音声区間検出方法。

【 0 1 2 7 】

(付記 2 4) 前記周波数分布の平坦さを算出する際は、前記周波数分布の隣接帯域間の差分の総和、前記周波数分布の隣接帯域間の差分の最大値、のいずれかを求めることを特徴とする付記 2 0 記載の音声区間検出方法。

【 0 1 2 8 】

(付記 2 5) 前記周波数分布の平坦さを周波数分布の平均で除算、またはフレームの平均電力で除算して正規化することを特徴とする付記 2 0 記載の音声区間検出方法。

【 0 1 2 9 】

(付記 2 6) 前記周波数分布の平坦さを算出する際は、前記周波数分布から平均値を求め、前記平均値からしきい値を生成し、前記周波数分布のうち前記しきい値を超える帯域数を前記周波数分布の平坦さとする特徴とする付記 2 0 記載の音声区間検出方法。

【 0 1 3 0 】

(付記 2 7) 前記周波数分布の平坦さを算出する際は、前記周波数分布から最大値を求め、前記最大値からしきい値を生成し、前記周波数分布のうち前記しきい値を超える帯域数を前記周波数分布の平坦さとする特徴とする付記 2 0 記載の音声区間検出方法。

【 0 1 3 1 】

【 発明の効果 】

以上説明したように、本発明の音声区間検出装置は、入力信号の周波数分布を算出し、周波数分布の平坦さを算出し、この場合、周波数分布の最大値を求め、周波数分布と最大値との差分の総和を、周波数分布の平坦さとする。そして、周波数分布の平坦さとしきい値とを比較して、音声か雑音かを判定し、入力信号の音声区間を検出する構成とした。これにより、周波数分布の最大値を求め、周波数分布と最大値との差分の総和を、周波数分布の平坦さとして、音声 / 雑音の判定を行うため、音声区間を高精度に検出することができ、通話品質の向上を図ることが可能になる。

また、本発明の音声区間検出装置は、入力信号の周波数分布を算出し、周波数分布の平坦さを算出し、この場合、周波数分布の隣接帯域間の差分の総和を、周波数分布の平坦さとする。そして、周波数分布の平坦さとしきい値とを比較して、音声か雑音かを判定し、入力信号の音声区間を検出する構成とした。これにより、周波数分布の隣接帯域間の差分の総和を、周波数分布の平坦さとして、音声 / 雑音の判定を行うため、音声区間を高精度に検出することができ、通話品質の向上を図ることが可能になる。

10

20

30

40

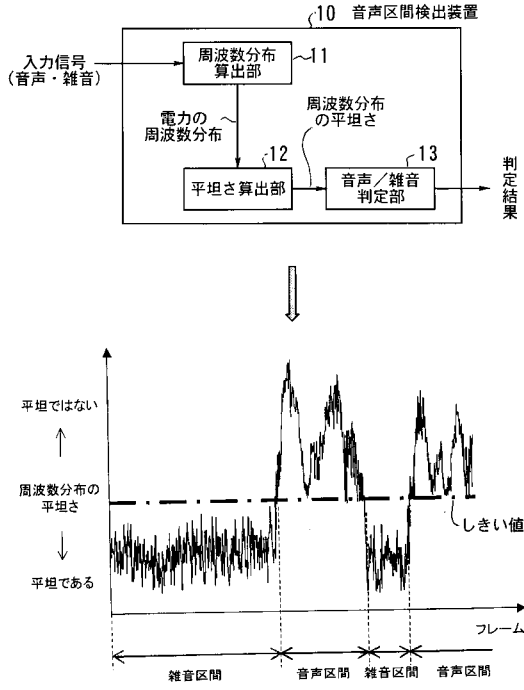
50

さらに、本発明の音声区間検出装置は、入力信号の周波数分布を算出し、周波数分布の平坦さを算出し、この場合、周波数分布から平均値を求め、平均値からしきい値を生成し、周波数分布のうちしきい値を超える帯域数を周波数分布の平坦さとする。そして、周波数分布の平坦さとしきい値とを比較して、音声か雑音かを判定し、入力信号の音声区間を検出する構成とした。これにより、周波数分布から平均値を求め、平均値からしきい値を生成し、周波数分布のうちしきい値を超える帯域数を周波数分布の平坦さとして、音声/雑音の判定を行うため、音声区間を高精度に検出することができ、通話品質の向上を図ることが可能になる。

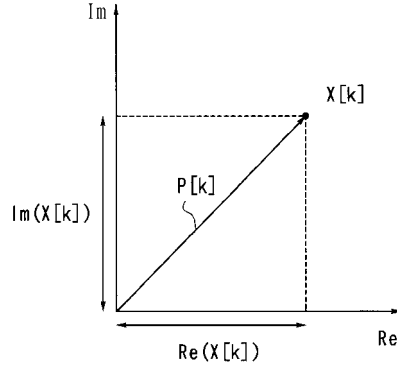
【図面の簡単な説明】

- 【図 1】 音声区間検出装置の原理図である。 10
- 【図 2】 電力 $P[k]$ を示す図である。
- 【図 3】 帯域分割による電力算出の概念を示す図である。
- 【図 4】 式 (2) の内容を説明するための図である。
- 【図 5】 バンドパスフィルタの周波数特性の例を示す図である。
- 【図 6】 電力の周波数分布の例を示す図である。
- 【図 7】 周波数分布と平均値との差分の総和から平坦さを求める際の概要を説明するための図である。
- 【図 8】 信号の周波数分布を示す図である。
- 【図 9】 周波数分布と平均値との差分の自乗和から、平坦さを求める際の概要を説明するための図である。 20
- 【図 10】 周波数分布と平均値との差分の最大値から平坦さを求める際の概要を説明するための図である。
- 【図 11】 周波数分布と最大値との差分の総和から平坦さを求める際の概要を説明するための図である。
- 【図 12】 周波数分布の隣接帯域間の差分の総和から平坦さを求める際の概要を説明するための図である。
- 【図 13】 周波数分布の隣接帯域間の差分の最大値から平坦さを求める際の概要を説明するための図である。
- 【図 14】 周波数分布の平均値から求めたしきい値を用いて平坦さを求める際の概要を説明するための図である。 30
- 【図 15】 音声区間、雑音区間の判定処理例を示す図である。
- 【図 16】 V O X 装置の構成を示す図である。
- 【図 17】 ノイズキャンセラ装置の構成を示す図である。
- 【図 18】 ノイズキャンセラ装置の構成を示す図である。
- 【図 19】 トーン検出装置の構成を示す図である。
- 【図 20】 トーン信号区間の判定処理を示す図である。
- 【図 21】 エコーキャンセラ装置の構成を示す図である。
- 【図 22】 制御テーブルを示す図である。
- 【符号の説明】
- 10 音声区間検出装置 40
- 11 周波数分布算出部
- 12 平坦さ算出部
- 13 音声/雑音判定部

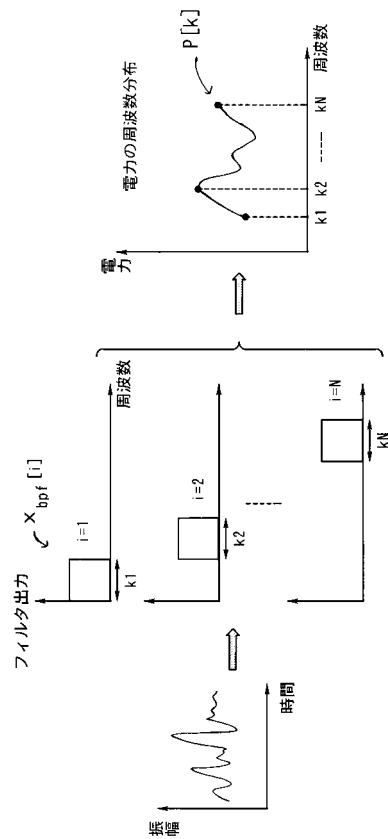
【 図 1 】



【 図 2 】



【 図 3 】



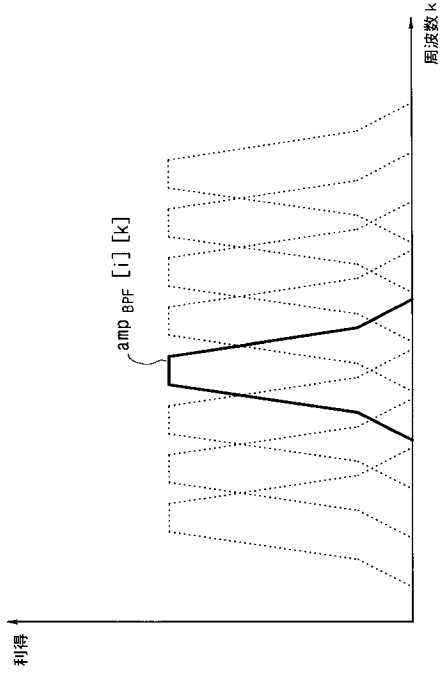
【 図 4 】

$$\begin{cases} b_{p f}[i][0] = 1 \\ b_{p f}[i][1] = 1 \\ b_{p f}[i][2] = 0 \\ \vdots \\ b_{p f}[i][j] = \sim \end{cases}$$

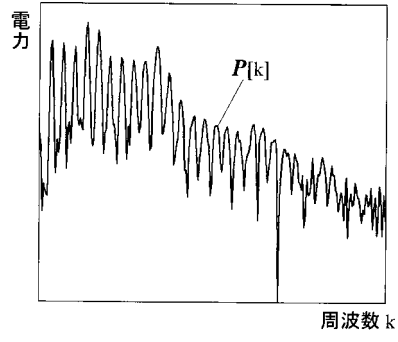
フィルタ係数

$$\begin{aligned}
 x_{b p f}[i][n] &= \sum_j b_{p f}[i][j] * x[n-j] \\
 &= b_{p f}[i][0] * x[n-0] \\
 &+ b_{p f}[i][1] * x[n-1] \\
 &+ b_{p f}[i][2] * x[n-2] \\
 &+ b_{p f}[i][j] * x[n-j] \\
 &= (1 * 0) + (1 * (-1)) + (0 * 1) + \dots
 \end{aligned}$$

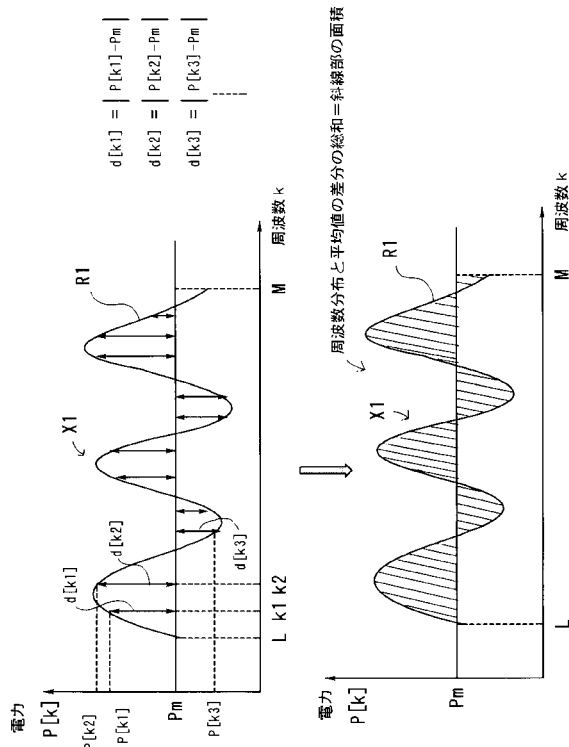
【 図 5 】



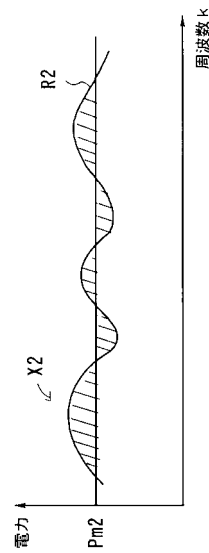
【 図 6 】



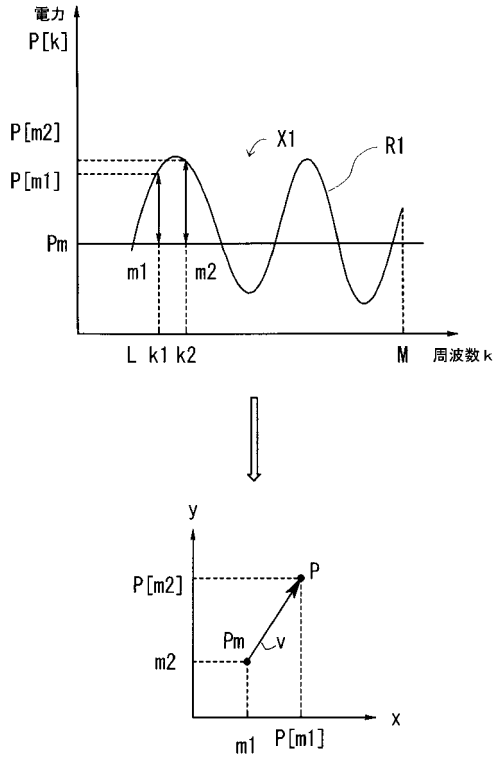
【 図 7 】



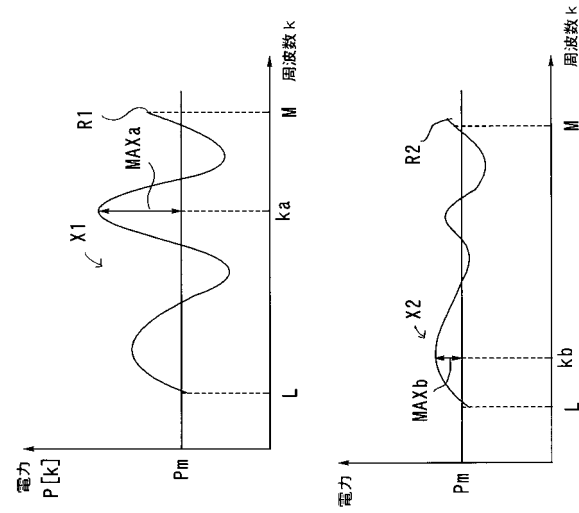
【 図 8 】



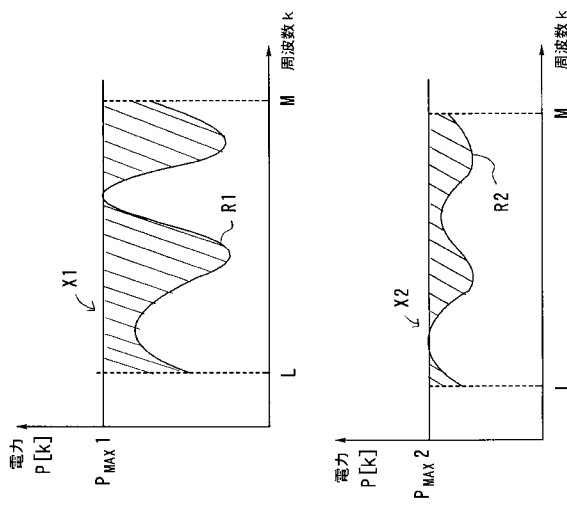
【 図 9 】



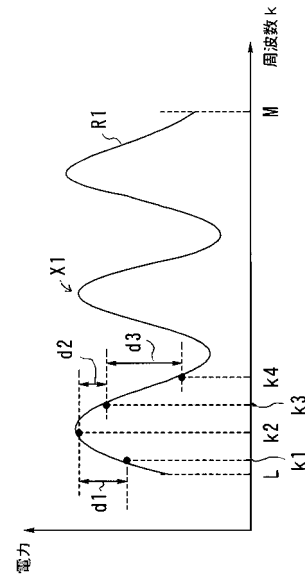
【 図 10 】



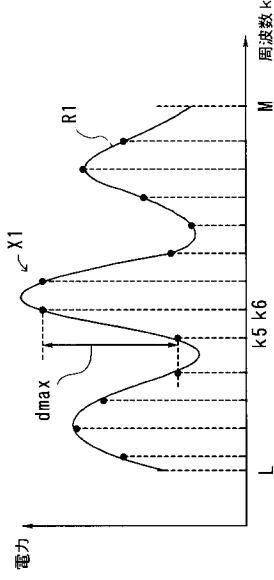
【 図 11 】



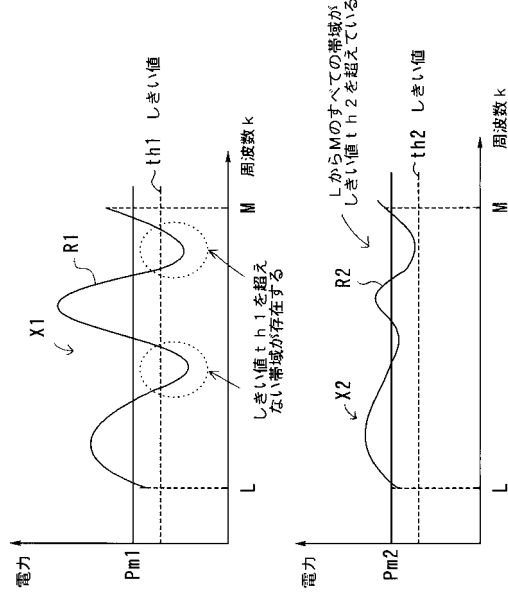
【 図 12 】



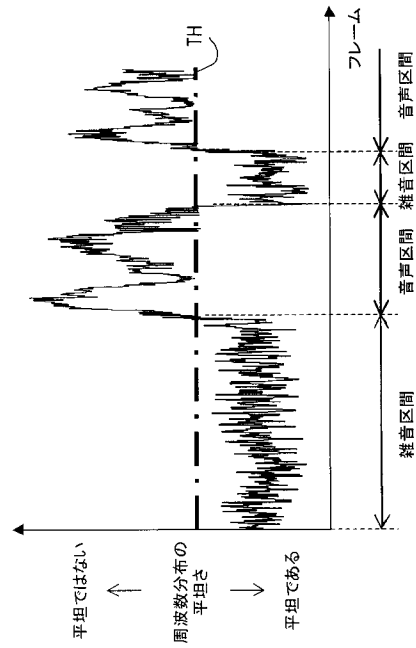
【 図 1 3 】



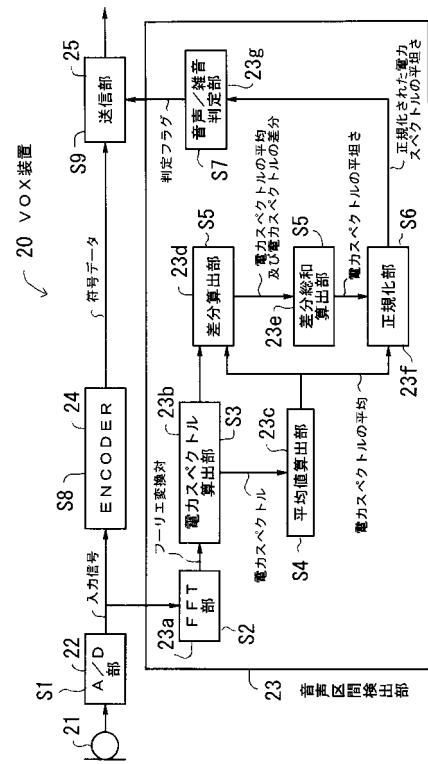
【 図 1 4 】



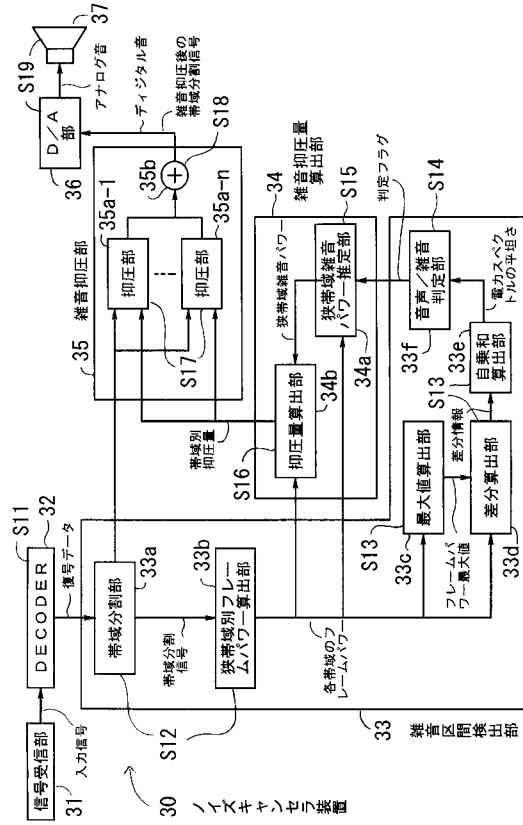
【 図 1 5 】



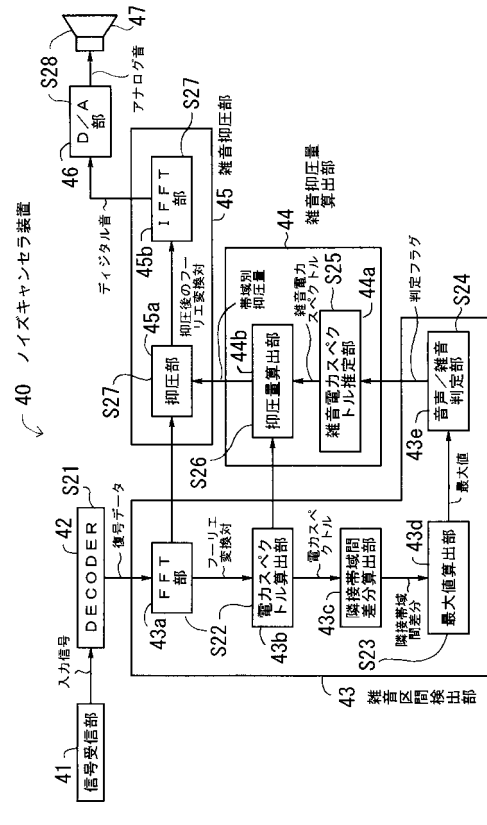
【 図 1 6 】



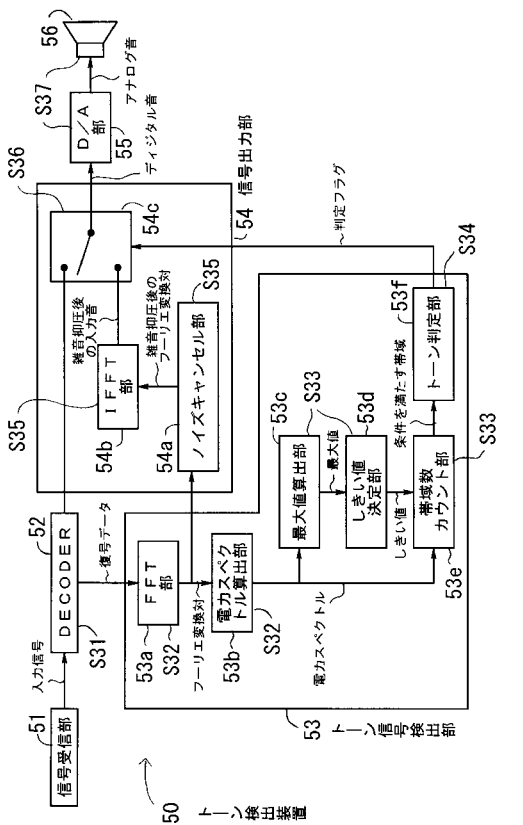
【図 17】



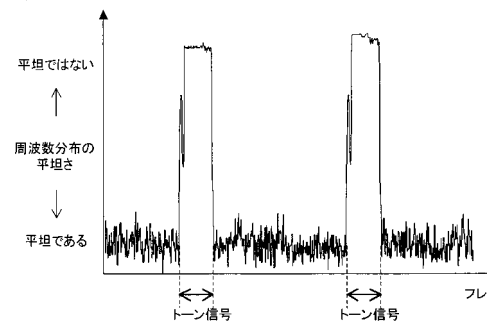
【図 18】



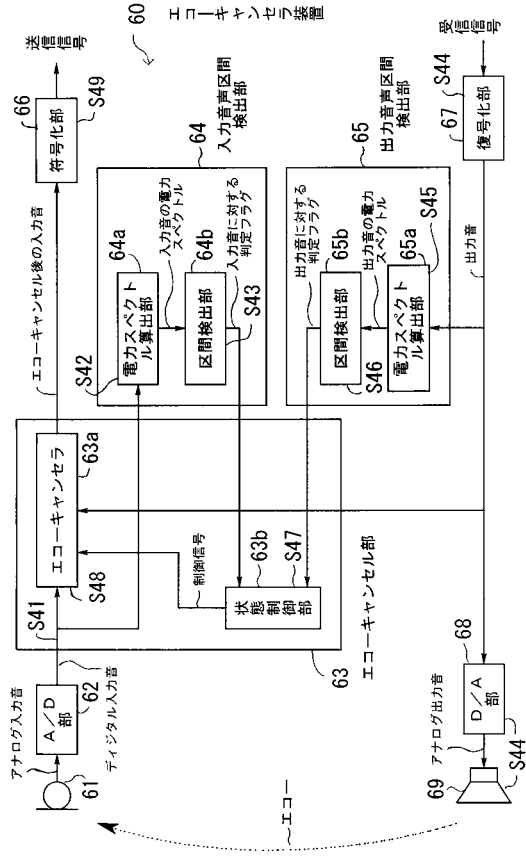
【図 19】



【図 20】



【図 2 1】



【図 2 2】

T1

入力音	出力音	制御信号	
		減算	学習
あり	あり	ON	OFF
あり	なし	OFF	OFF
なし	あり	ON	ON
なし	なし	OFF	OFF

フロントページの続き

審査官 山下 剛史

(56)参考文献 特開平03 - 114100 (JP, A)
特開平7 - 143009 (JP, A)

(58)調査した分野(Int.Cl., DB名)
G10L 11/00~21/06