

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4009781号
(P4009781)

(45) 発行日 平成19年11月21日(2007.11.21)

(24) 登録日 平成19年9月14日(2007.9.14)

(51) Int. Cl.

G 1 0 L 19/02 (2006.01)

F I

G 1 0 L 19/02 1 4 3

G 1 0 L 19/02 1 4 4

請求項の数 8 (全 17 頁)

(21) 出願番号 特願2003-365983 (P2003-365983)
 (22) 出願日 平成15年10月27日(2003.10.27)
 (65) 公開番号 特開2005-128404 (P2005-128404A)
 (43) 公開日 平成17年5月19日(2005.5.19)
 審査請求日 平成17年5月26日(2005.5.26)

(73) 特許権者 000001443
 カシオ計算機株式会社
 東京都渋谷区本町1丁目6番2号
 (74) 代理人 100090033
 弁理士 荒船 博司
 (74) 代理人 100093045
 弁理士 荒船 良男
 (72) 発明者 井手 博康
 東京都羽村市栄町3丁目2番1号 カシオ
 計算機株式会社 羽村技術センター内

審査官 新川 圭二

最終頁に続く

(54) 【発明の名称】 音声処理装置及び音声符号化方法

(57) 【特許請求の範囲】

【請求項1】

入力された音声信号をフレームに分割するフレーム化部と、
 前記フレーム化部により得られたフレーム毎に、フレームに含まれる音声信号の振幅の最大値に基づいて音声信号のレベルを調整するレベル調整部と、
 前記レベル調整部によりレベルが調整された音声信号に対し、1フレームの信号について複数のブロックに分けて周波数変換を施す周波数変換部と、
 前記周波数変換により各ブロックから得られた1フレーム分の周波数変換係数について、同一周波数帯域の係数をまとめてベクトル化して、ベクトル量子化を施すベクトル量子化部と、
 前記ベクトル量子化により得られた音声信号に対し、エントロピー符号化を施すエントロピー符号化部と、
 前記エントロピー符号化により得られた音声信号のデータ量が、予め設定された目標データ量より大きいかが否かを判定する判定部と、
 前記判定部により、前記エントロピー符号化により得られた音声信号のデータ量が前記目標データ量より大きいと判定された場合、前記ベクトル量子化により得られた音声信号のうち、エネルギーが所定値以下の周波数帯域の音声信号を削除するデータ削除部と、
 を備え、
 前記エントロピー符号化部は前記データ削除部で削除されなかった音声信号を再度符号化することを特徴とする音声処理装置。

10

20

【請求項 2】

前記周波数変換部は、周波数変換として変形離散コサイン変換を用いることを特徴とする請求項 1 に記載の音声処理装置。

【請求項 3】

高域周波数用の音声パターンを示すベクトルを格納した第 1 のテーブルと、
低域周波数用の音声パターンを示すベクトルを格納した第 2 のテーブルと、を備え、
前記ベクトル量子化部は、前記第 1 のテーブルと前記第 2 のテーブルを用いてベクトル量子化を行うことを特徴とする請求項 1 又は 2 に記載の音声処理装置。

【請求項 4】

周波数帯域別に、音声パターンを示すベクトルを格納したテーブルを備え、
前記ベクトル量子化部は、前記テーブルを用いてベクトル量子化を行うことを特徴とする請求項 1 又は 2 に記載の音声処理装置。

10

【請求項 5】

前記エントロピー符号化部は、エントロピー符号化としてハフマン符号化を用いることを特徴とする請求項 1 ~ 4 の何れか一項に記載の音声処理装置。

【請求項 6】

前記エントロピー符号化部は、エントロピー符号化としてレンジコーダによる符号化を用いることを特徴とする請求項 1 ~ 4 の何れか一項に記載の音声処理装置。

【請求項 7】

前記エントロピー符号化部は、入力信号を示す各記号の出現確率を格納した生起確率テーブルを有し、記号の入力に伴って前記生起確率テーブルを更新することを特徴とする請求項 6 に記載の音声処理装置。

20

【請求項 8】

入力された音声信号をフレームに分割し、
フレーム毎に、フレームに含まれる音声信号の振幅の最大値に基づいて音声信号のレベルを調整し、
レベルが調整された音声信号に対し、1 フレームの信号について複数のブロックに分けて周波数変換を施し、
前記周波数変換により各ブロックから得られた 1 フレーム分の周波数変換係数について、同一周波数帯域の係数をまとめてベクトル化して、ベクトル量子化を施し、
前記ベクトル量子化により得られた音声信号に対し、エントロピー符号化を施し、
前記エントロピー符号化により得られた音声信号のデータ量が、予め設定された目標データ量より大きいと判定し、前記エントロピー符号化により得られた音声信号のデータ量が前記目標データ量より大きいと判定された場合、前記ベクトル量子化により得られた音声信号のうち、エネルギーが所定値以下の周波数帯域の音声信号を削除して、削除されなかった音声信号を再度符号化することを特徴とする音声符号化方法。

30

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、音声処理装置及び音声符号化方法に関する。

40

【背景技術】

【0002】

従来より、音声信号の圧縮方式として、 μ -law、ADPCM (Adaptive Differential Pulse Code Modulation)、音楽で利用される MP3 (MPEG Audio Layer-3)、携帯電話等で利用される VSELP (Vector Sum Excited Linear Prediction)、G.729 等の CELP (Code-Excited Linear Prediction) 系の圧縮方式が実用化されている。特許文献 1 には、音声圧縮技術として、ベクトル量子化を用いた技術が開示されている。

【特許文献 1】特開平 10 - 63299 号公報

【発明の開示】

【発明が解決しようとする課題】

50

【0003】

語学学習において会話等の録音を行う場合、16kHz程度のサンプリング周波数が、各言語の特徴を保ちつつ、多くのデータ量を必要としない適度な周波数であると考えられている。しかしながら、CELP系の圧縮方式に現れる圧縮ノイズは、語学学習用には適切ではないという問題があった。また、 μ -law、ADPCMは、十分な音質であるが、符号化レートが高いため、携帯機器でこれらの圧縮方式を利用する場合、録音時間が短くなってしまうという問題があった。また、MP3は、主に高品質の音声の圧縮を対象としており、16kHz程度のサンプリング周波数では、効果的に圧縮を行うことができないという問題があった。

【0004】

本発明の課題は、語学学習用会話等の録音に適切なサンプリング周波数で、符号化レートを低減させる音声圧縮を可能にすることである。

【課題を解決するための手段】

【0005】

本発明に係る音声処理装置は、入力された音声信号をフレームに分割するフレーム化部と、前記フレーム化部により得られたフレーム毎に、フレームに含まれる音声信号の振幅の最大値に基づいて音声信号のレベルを調整するレベル調整部と、前記レベル調整部によりレベルが調整された音声信号に対し、1フレームの信号について複数のブロックに分けて周波数変換を施す周波数変換部と、前記周波数変換により各ブロックから得られた1フレーム分の周波数変換係数について、同一周波数帯域の係数をまとめてベクトル化して、ベクトル量子化を施すベクトル量子化部と、前記ベクトル量子化により得られた音声信号に対し、エントロピー符号化を施すエントロピー符号化部と、前記エントロピー符号化により得られた音声信号のデータ量が、予め設定された目標データ量より大きいか否かを判定する判定部と、前記判定部により、前記エントロピー符号化により得られた音声信号のデータ量が前記目標データ量より大きいと判定された場合、前記ベクトル量子化により得られた音声信号のうち、エネルギーが所定値以下の周波数帯域の音声信号を削除するデータ削除部と、を備え、前記エントロピー符号化部は前記データ削除部で削除されなかった音声信号を再度符号化することを特徴とする。

【0006】

また、前記周波数変換部は、周波数変換として変形離散コサイン変換を用いるのが好ましい。

【0007】

また、高域周波数用の音声パターンを示すベクトルを格納した第1のテーブルと、低域周波数用の音声パターンを示すベクトルを格納した第2のテーブルと、を備え、前記ベクトル量子化部は、前記第1のテーブルと前記第2のテーブルを用いてベクトル量子化を行うのが好ましい。

【0008】

また、周波数帯域別に、音声パターンを示すベクトルを格納したテーブルを備え、前記ベクトル量子化部は、前記テーブルを用いてベクトル量子化を行うようにしてもよい。

【0009】

また、前記エントロピー符号化部は、エントロピー符号化としてハフマン符号化を用いるのが好ましい。

【0010】

また、前記エントロピー符号化部は、エントロピー符号化としてレンジコードによる符号化を用いるのが好ましい。

【0011】

また、レンジコードによる符号化を行うエントロピー符号化部は、入力信号を示す各記号の出現確率を格納した生起確率テーブルを有し、記号の入力に伴って前記生起確率テーブルを更新するのが好ましい。

【0012】

10

20

30

40

50

また、前記エントロピー符号化により得られた音声信号のデータ量が、予め設定された目標データ量より大きいか否かを判定する判定部と、前記判定部により、前記エントロピー符号化により得られた音声信号のデータ量が前記目標データ量より大きいと判定された場合、前記周波数変換により得られた各周波数帯域の音声信号のうち、エネルギーが所定値以下の帯域の音声信号を削除するデータ削除部と、を備えるようにしてもよい。

【0014】

本発明に係る音声符号化方法は、入力された音声信号をフレームに分割し、フレーム毎に、フレームに含まれる音声信号の振幅の最大値に基づいて音声信号のレベルを調整し、レベルが調整された音声信号に対し、1フレームの信号について複数のブロックに分けて周波数変換を施し、前記周波数変換により各ブロックから得られた1フレーム分の周波数変換係数について、同一周波数帯域の係数をまとめてベクトル化して、ベクトル量子化を施し、前記ベクトル量子化により得られた音声信号に対し、エントロピー符号化を施し、前記エントロピー符号化により得られた音声信号のデータ量が、予め設定された目標データ量より大きいか否かを判定し、前記エントロピー符号化により得られた音声信号のデータ量が前記目標データ量より大きいと判定された場合、前記ベクトル量子化により得られた音声信号のうち、エネルギーが所定値以下の周波数帯域の音声信号を削除して、削除されなかった音声信号を再度符号化することを特徴とする。

10

【発明の効果】

【0015】

本発明によれば、語学学習用会話等の録音に適切なサンプリング周波数で、符号化レートを低減させることができる。

20

【発明を実施するための最良の形態】

【0016】

以下、図面を参照して、本発明の実施形態について詳細に説明する。

まず、本実施形態における構成について説明する。

【0017】

図1は、本発明の実施形態の音声処理装置100の構成を示すブロック図である。音声処理装置100は、図1に示すように、A/D変換部1、DC(Direct Current)除去部2、フレーム化部3、レベル調整部4、周波数変換部5、周波数並べ替え部6、ベクトル量子化部7、エントロピー符号化部8、レートコントローラ9、データ削除部10により構成される。

30

【0018】

A/D変換部1は、入力された音声アナログ信号をデジタル信号に変換し、DC除去部2に出力する。サンプリング周波数は、16kHz程度が望ましいが、11.025kHz、22.05kHz等でもかまわない。

【0019】

DC除去部2は、A/D変換部1から入力された音声信号の直流成分を除去し、フレーム化部3に出力する。音声信号の直流成分を除去するのは、直流成分が音質にほとんど無関係であることによる。直流成分の除去は、例えば、高域通過フィルタによって実現することができる。高域通過フィルタには、例えば、式(1)で表されるものがある。

40

【数1】

$$H(z) = \frac{0.464 - 0.927z^{-1} + 0.464z^{-2}}{1 - 1.906z^{-1} + 0.911z^{-2}} \quad (1)$$

【0020】

フレーム化部3は、DC除去部2から入力された信号を、圧縮の処理単位であるフレームに分割し、レベル調整部4に出力する。ここで、1つのフレームには、1つ以上、好ましくは4つ以上のブロックが含まれる長さにする。1ブロックは、1回のMDCT(Modified Discrete Cosine Transform: 変形離散コサイン変換)を行う単位であり、MDCT

50

の次数分の長さを有する。以下、1フレームを構成する各ブロックをMDC Tブロックと呼ぶことにする。図2に、入力信号と各フレームとの関係を示し、図3に、1フレームと各MDC Tブロックとの関係を示す。図3に示すように、各MDC Tブロックは直前のMDC Tブロックと、MDC Tブロックの半分の長さの重複部分を有する。また、図2に示すように、各フレームは、直前のフレームと、MDC Tブロックの半分の長さの重複部分を有する。

【0021】

レベル調整部4は、フレーム毎に、入力された音声信号のレベル調整を行い、レベル調整された信号を周波数変換部5に出力する。レベル調整とは、1フレーム中に含まれる信号の振幅の最大値を、指定されたビット（以下、制圧目標ビット）数に収まるようにすることである。レベル調整は、例えば、1フレーム中の信号の最大振幅をnbit、制圧目標ビット数をNとすると、フレーム中の信号を全て、式(2)を満たすshift_bit数分LSB (Least Significant Bit: 最下位ビット) 側にシフトすることによって実現できる。

【数2】

$$\text{shift_bit} = \begin{cases} 0 & (n \leq N) \\ N - n & (n > N) \end{cases} \quad (2)$$

音声再生時には、振幅が制圧目標ビット以下に制圧された信号を元に戻す必要があるため、shift_bitを表す信号を、音声圧縮信号の一部として出力する必要がある。

【0022】

周波数変換部5は、レベル調整部4から入力された信号に対し周波数変換を施し、周波数並べ替え部6に出力する。本実施形態では、周波数変換としてMDC T (Modified Discrete Cosine Transform: 変形離散コサイン変換) を用いる場合を示す。MDC Tブロックの長さをM、入力信号を $\{x_n \mid n=0, \dots, M-1\}$ とすると、MDC T係数 $\{X_k \mid n=0, \dots, M/2-1\}$ は式(3)で表される。

【数3】

$$X_k = \sum_{n=0}^{M-1} x_n \cdot h_n \cdot \cos \left\{ \frac{2\pi}{M} \left(\frac{k+1}{2} \right) \left(n + \frac{M}{4} + \frac{1}{2} \right) \right\} \quad (3)$$

ここで、 h_n は窓関数であり、式(4)で表される。

【数4】

$$h_n = \sin \left\{ \frac{\pi}{M} \left(n + \frac{1}{2} \right) \right\} \quad (4)$$

なお、ブロック長Mは、16kHz程度のサンプリング周波数の音声では、256程度の値が考えられる。

【0023】

周波数並べ替え部6は、周波数変換部5から入力されたMDC T係数を周波数毎に並べ替え、同一周波数帯域の係数をまとめてベクトル化し、ベクトル量子化部7に出力する。このように、同一周波数帯域の信号をまとめてベクトル化すると、例えば、定常信号を多く含む場合、後のベクトル量子化の精度が向上する。1フレームにMDC Tブロックがm個あり、各MDC TでMDC T係数がM/2個算出された場合、i番目のMDC Tブロックのj番目のMDC T係数を X_{ij} とすると、j番目の周波数帯域をまとめたベクトル F_j は、 $F_j = \{X_{ij} \mid i=0, \dots, m-1\}, j=0, \dots, M/2-1$ となる。

【0024】

ベクトル量子化部7は、複数の音声パターンを示す代表ベクトルを格納したVQ (Vector Quantization) テーブルを有し、周波数並べ替え部6で作成されたベクトル F_j と、VQテーブルに格納された各代表ベクトルを比較し、最も類似した代表ベクトルが示すインデックスを符号としてエントロピー符号化部8に出力する。

10

20

30

40

50

【 0 0 2 5 】

例えば、ベクトル長 N の符号化対象のベクトルを $\{s_j \mid j=1, \dots, N\}$ 、 VQ テーブルに格納された k 個の代表ベクトルを $\{V_i \mid i=1, \dots, k\}$ 、 $V_i = \{v_{ij} \mid j=1, \dots, N\}$ とすると、符号化対象のベクトルと、 VQ テーブルに格納された i 番目の代表ベクトルの各要素 v_{ij} の誤差 e_i が最小となるような i (インデックス) を、出力する符号とする。誤差 e_i の算出式を式 (5) に示す。

【数 5】

$$e_i = \sum_{j=1}^N (s_j - v_{ij})^2 \quad (5)$$

10

代表ベクトルの数 k とベクトル長 N は、ベクトル量子化に要する処理時間や VQ テーブルの容量等を勘案して決定される。例えば、ベクトル長を 2 にして代表ベクトル数を 256 にしたり、ベクトル長を 4 にして代表ベクトル数を 8192 ($= 2^{13}$) にしたりするなど、自由な組み合わせが考えられる。

【 0 0 2 6 】

音声は、高域周波数部分と低域周波数部分で異なる特性がある場合が多いため、本実施形態では、高域と低域で異なる VQ テーブルを用いることにする。高域用の代表ベクトルが格納された VQ テーブルを高域用 VQ テーブル 7 a、低域用の代表ベクトルが格納された VQ テーブルを低域用 VQ テーブル 7 b とする。周波数並べ替え部 6 で作成されたベクトル $F_j = \{X_{ij} \mid i=0, \dots, m-1\}$, $j=0, \dots, M/2-1$ において、高域と低域の境界は、周波数帯域を示す j を単純に半分に分ければよい。即ち、 $F_0, F_1, \dots, F_{M/4-1}$ を低域、 $F_{M/4}, F_{M/4+1}, \dots, F_{M/2-1}$ を高域とすればよい。従って、低域のベクトル $F_0, F_1, \dots, F_{M/4-1}$ は、低域用 VQ テーブル 7 b に格納された各代表ベクトルと比較され、最も類似した代表ベクトルが示すインデックスが符号として出力される。同様に、高域のベクトル $F_{M/4}, F_{M/4+1}, \dots, F_{M/2-1}$ は、高域用 VQ テーブル 7 a に格納された各代表ベクトルと比較され、最も類似した代表ベクトルが示すインデックスが符号として出力される。

20

【 0 0 2 7 】

エントロピー符号化部 8 は、ベクトル量子化部 7 から入力された信号に対してエントロピー符号化を施し、レートコントローラ 9 に出力する。エントロピー符号化とは、信号の統計的性質を利用して、符号をより短い符号へと変換する符号化方式であり、ハフマン (Huffman) 符号化、算術符号化、レンジコーダ (Range Coder) による符号化等がある。エントロピー符号化の詳細については、後に図 4 ~ 図 8 を参照して説明する。

30

【 0 0 2 8 】

レートコントローラ 9 は、エントロピー符号化で得られた符号のデータ量が、予め設定された目標データ量より大きいかなかを判定し、エントロピー符号化で得られた符号のデータ量が目標データ量より大きいと判定した場合、データ削除部 10 に対し、 j 番目の周波数帯域 F_j のエネルギーを $|F_j|^2$ としたとき、 $|F_j|^2$ が所定値以下の帯域の信号の削除を要求する。エントロピー符号化で得られた符号のデータ量が目標データ量以下であると判定した場合は、レートコントローラ 9 は、エントロピー符号化で得られた符号を音声圧縮信号として出力する。レートコントローラ 9 から出力された音声圧縮信号は、記録媒体に記録されたり、通信ネットワークを介して外部装置に伝送されたりする。

40

【 0 0 2 9 】

データ削除部 10 は、エネルギー $|F_j|^2$ が所定値以下の帯域の信号を削除する処理を行い、処理後の信号をエントロピー符号化 8 に出力し、再度、エントロピー符号化を要求する。エネルギー $|F_j|^2$ が所定値以下の帯域の信号を削除する処理は、例えば、式 (6) に示すように、エネルギー $|F_j|^2$ が閾値 t 未満の係数を 0 にするようにする。

$$F_j = \{0, 0, \dots, 0\}; \quad |F_j|^2 < t \quad (6)$$

【 0 0 3 0 】

エントロピー符号化

50

以下では、本実施形態で適用されるエントロピー符号化の例として、ハフマン符号化、レンジコーダによる符号化について説明する。

【0031】

(ハフマン符号化)

ハフマン符号化とは、出現頻度の高い記号には短い符号を割り当て、出現頻度の低い記号には長い符号を割り当てることで、全体のデータ量を圧縮する方式である。例えば、4つの記号{a、b、c、d}からなる100文字のデータがあったとする。全ての記号に同じ長さの2進数の符号(固定長符号)を割り当てる場合、4つの記号を表すには2ビットが必要であるため、100文字のデータ量は、 $2[\text{bit}] \times 100 = 200[\text{bit}]$ となる。

【0032】

ハフマン符号化では、各記号の出現頻度に応じて2進数の符号が割り当てられる。図4に、100文字のデータ中の各記号a、b、c、dの出現頻度が、それぞれ、10、70、1、19である場合に各記号に割り当てられた2進数の符号の例を示す。図4に示すように、記号a、b、c、dに、それぞれ、符号100、0、101、11が割り当てられた場合、100文字のデータ量は、 $3[\text{bit}] \times 10 + 1[\text{bit}] \times 70 + 3[\text{bit}] \times 1 + 2[\text{bit}] \times 19 = 141[\text{bit}]$ となり、データ量は、固定長符号のデータ量の70%に圧縮される。

【0033】

(レンジコーダによる符号化)

符号化前の元信号に含まれる記号の集合を $S = \{s_i | i=1, \dots, n\}$ とし、各記号 s_i の出現確率を p_i とする。また、元信号に含まれる各記号 s_i を予め決められた順番に並べ替えた記号列 $\{s_1, s_2, \dots, s_n\}$ において、記号 s_k ($k \geq 2$)より前に並んでいる各記号の出現確率の合計を F_k とする。即ち、 F_k は、式(7)のように表される。

【数6】

$$F_k = \sum_{i=1}^{k-1} p_i \quad (k \geq 2), \quad F_1 = 0 \quad (7)$$

【0034】

レンジコーダによる符号化では、記号毎に出現確率 p_i と F_i を対応付けて格納したテーブル(以下、生起確率テーブルという。)に基づいて、入力済みの信号が示す記号列に、数値で示す範囲(下限、幅)を設定する処理を行う。入力済みの信号に設定される範囲(下限、幅)は、直前に入力された信号に設定された範囲と生起確率テーブルに基づいて決定される。

【0035】

符号化対象の信号 s_k が入力されたときに設定される幅を range' 、下限を low' とし、その信号 s_k の1つ前の信号が入力されたときに設定された幅を range 、下限を low とすると、幅 range' 、下限 low' は、それぞれ、式(8)、式(9)のように表される。

$$\text{range}' = \text{range} \times p_k \quad (8)$$

$$\text{low}' = \text{low} + \text{range} \times F_k \quad (9)$$

式(8)及び式(9)で算出された range' 、 low' が、次の信号が入力されたときの range 、 low となる。

【0036】

式(8)及び式(9)で示す算出処理は、入力信号がなくなるまで行われ、最後の信号が入力されたときに算出された range 、 low に基づいて決定される範囲 $\text{low} \sim \text{low} + \text{range}$ の間の値が符号値として出力される。

【0037】

図5に、レンジコーダ符号化の例を示す。図5(a)に、元信号に含まれる記号の集合が $S = \{s_1=a, s_2=b, s_3=c, s_4=d\}$ であるときの生起確率テーブルの一例を示す。また、図5(b)に、記号列{b a c a}に対する符号化の一例を示す。図5(b)では、

10

20

30

40

50

記号列を示す符号を10進数とし、lowの初期値を0、rangeの初期値を 10^6 とした場合を示している。図5(b)において、「入力信号」項目は、入力された記号を示し、「記号列」項目は、これまでに入力された記号列を示し、「low」項目は、式(9)により算出されるlow'を示し、「range」項目は、式(8)により算出されるrange'を示す。また、「範囲」項目は、low及びrangeから決定される符号値の範囲を示す。図5(b)において、 $[x, y)$ という表記は、符号値Zが $x \leq Z < y$ を満たすことを意味する。図5(b)によると、 $593750 \leq Z < 603125$ を満たす符号値Zのうちの1つ(例えば、600000)が、記号列{b a c a}を符号化した結果として出力されることになる。
【0038】

このように、レンジコードによる符号化では、予め決められた出現確率を利用して入力される各記号を符号化しているため、元信号に含まれる各記号の出現確率が固定された情報源からの発生であれば非常に有効である。しかしながら、符号化対象となる信号が、出現確率が一定の情報源から発生されていることは極めてまれである。よって、上述のレンジコードによる符号化では、各記号の出現確率が符号化対象となる信号に適應していない。そこで、本実施形態では、レンジコード符号化において、信号が入力される度に出現確率を更新させるようにすることによって、実際の信号に適應可能にした。以下、本実施形態のレンジコードによる符号化について説明する。

【0039】

上述と同様に、符号化前の元信号に含まれる記号の集合を $S = \{s_i | i=1, \dots, n\}$ とする。元信号に含まれる記号 s_i の出現頻度を f_i 、出現頻度 f_i の合計をcum、各記号 s_i の出現確率を p_i とすると、cum、 p_i は、それぞれ、式(10)、式(11)のように表される。

【数7】

$$cum = \sum_{i=1}^n f_i \quad (10)$$

【数8】

$$p_i = \frac{f_i}{cum} \quad (11)$$

【0040】

エントロピー符号化部8は、入力された信号に幅range及び下限lowを設定するためのテーブルとして、図6に示すような生起確率テーブル81を有する。生起確率テーブル81は、図6に示すように、各記号毎に、出現頻度 f_i 、出現確率 p_i 、 F_i の各項目を対応付けて格納している。 F_i の定義は、式(7)で示したとおりである。

【0041】

エントロピー符号化部8に符号化対象の信号 s_k が入力されたときに設定される幅をrange'、下限をlow'とし、その信号 s_k の1つ前の信号が入力されたときに設定された幅をrange、下限をlowとすると、幅range'、下限low'は、それぞれ、式(12)、式(13)のように表される。

【数9】

$$range' = range \times p_k = range \times \frac{f_k}{cum} \quad (12)$$

【数10】

$$low' = low + range \times F_k = low + \left(\frac{range}{cum} \right) \sum_{i=1}^{k-1} f_i \quad (13)$$

10

20

30

40

50

式(12)及び式(13)で算出されたrange'、low'が、次の信号が入力されたときのrange、lowとなる。

【0042】

信号 s_k の入力によりrange、lowが算出されると、エントロピー符号化部8は、式(14)に示すように、出現頻度 f_k に1を加算し、算出された出現頻度 f_k' を新たな f_k とする。

$$f_k' = f_k + 1 \quad (14)$$

エントロピー符号化部8は、出現頻度 f_k の加算に伴い、cum、出現確率 p_i 、 F_i を再計算し、生起確率テーブル81を更新する。エントロピー符号化部8は、これらの処理を、入力信号がなくなるまで行い、最後の信号が入力されたときに算出されたrange、lowに基づいて決定される範囲low~low+rangeの間の値を符号値として出力する。

【0043】

図7及び図8に、本実施形態のレンジコード符号化の例を示す。図7(a)に、元信号に含まれる記号の集合が $S = \{s_1=a, s_2=b, s_3=c, s_4=d\}$ であるときのデフォルトの生起確率テーブル81の一例を示す。図7(a)に示すデフォルトの生起確率テーブル81の p_i 及び F_i は、図5(a)に示す生起確率テーブルと同一であるものとする。また、図7(b)には、図5(b)に示した記号列と同一の記号列{b a c a}に対する符号化の一例を示す。図7(b)においても、記号列を示す符号を10進数とし、lowの初期値を0、rangeの初期値を 10^6 とする。図7(b)において、「入力信号」項目は、入力された記号を示し、「記号列」項目は、これまでに入力された記号列を示し、「low」項目は、式(13)により算出されるlow'を示し、「range」項目は、式(12)により算出されるrange'を示す。また、「範囲」項目は、low及びrangeから決定される符号値の範囲を示す。また、「生起確率テーブル」項目は、記号の入力毎に更新された生起確率テーブルを示す。図8に、記号の入力毎に更新された生起確率テーブルを示す。図7(b)によると、記号の入力毎に生起確率テーブルを更新することで、記号列{b a c a}が示す「範囲」は、図5(b)に示した生起確率テーブルが固定された場合と異なり、 $591992 \leq Z < 599757$ を満たす符号値Zのうちの1つが、記号列{b a c a}を符号化した結果として出力されることになる。

【0044】

図9に、音声処理装置100により圧縮された音声信号を復号する復号装置200の構成を示す。復号装置200は、図9に示すように、エントロピー復号部11、逆ベクトル量子化部12、時間順並べ替え部13、周波数逆変換部14、レベル再現部15、フレーム合成部16、D/A変換部17により構成される。なお、音声処理装置100と復号装置200を、一つの筐体に一体的に備えるような構造としてもよいし、各々を別体として設けるようにしてもよい。

【0045】

エントロピー復号部11は、エントロピー符号化により符号化された信号を復号し、逆ベクトル量子化部12に出力する。逆ベクトル量子化部12は、複数の音声パターンを示す代表ベクトルを格納したテーブルとして、高域用VQテーブル12a、低域用VQテーブル12bを有し、エントロピー復号部11から入力された信号(インデックス)に対応する代表ベクトルを抽出し、時間順並べ替え部13に出力する。

【0046】

時間順並べ替え部13は、逆ベクトル量子化部12から入力されたベクトルを時間順に並べ替え、周波数逆変換部14に出力する。周波数逆変換部14は、時間順並べ替え部13から入力された信号(ベクトル)に対し、逆MDCTを施し、レベル再現部15に出力する。レベル再現部15は、周波数逆変換部14から入力された信号のレベル調節を行って、元のレベルに戻し、フレーム合成部16に出力する。フレーム合成部16は、符号化及び復号化の処理単位であったフレームを合成し、合成後の信号をD/A変換部17に出力する。D/A変換部17は、フレーム合成部16から入力されたデジタル信号をアナログ信号に変換し、音声再生信号として出力する。

10

20

30

40

50

【 0 0 4 7 】

次に、本実施形態における動作について説明する。

まず、図 10 のフローチャートを参照して、音声処理装置 100 において実行される音声圧縮処理について説明する。以下のフローチャートでは、エントロピー符号化として、レンジコードによる符号化が用いられる場合を示す。

【 0 0 4 8 】

まず、音声アナログ信号が入力されると、A/D変換部 1 において、入力された音声アナログ信号が音声デジタル信号に変換される（ステップ S 1）。以下、符号化対象の音声デジタル信号を単に音声信号と呼ぶことにする。次いで、DC除去部 2 において、音声信号の直流成分が削除され（ステップ S 2）、フレーム化部 3 において、直流成分削除後の音声信号がフレームに分割される（ステップ S 3）。

10

【 0 0 4 9 】

次いで、レベル調整部 4 において、フレーム毎に、入力された音声信号のレベルが調整され（ステップ S 4）、周波数変換部 5 において、レベル調整後の音声信号に対し、MDCTが施される（ステップ S 5）。次いで、周波数並べ替え部 6 において、MDCT係数が周波数毎に並べ替えられ（ステップ S 6）、同一周波数帯域の係数がまとめてベクトル化される。

【 0 0 5 0 】

次いで、ベクトル量子化部 7 において、高域のMDCT係数のベクトルと高域用VQテーブル 7 a に格納された代表ベクトルが比較されるとともに、低域のMDCT係数のベクトルと低域用VQテーブル 7 b に格納された代表ベクトルが比較され、最も類似した代表ベクトルが示すインデックスが符号として出力される（ステップ S 7）。

20

【 0 0 5 1 】

次いで、生起確率テーブル 8 1 が初期化され（ステップ S 8）、レンジコードの現在の生起確率テーブル 8 1 の内容が一旦保存される（ステップ S 9）。そして、エントロピー符号化部 8 において、ベクトル量子化後の音声信号に対し、フレーム毎にレンジコード符号化が施され（ステップ S 10）、レンジコード符号化後の音声圧縮信号がレートコントローラ 9 に出力される。レンジコード符号化については、後に図 11 を参照して詳細に説明する。

【 0 0 5 2 】

次いで、レートコントローラ 9 において、エントロピー符号化部 8 から入力された 1 フレーム分の音声圧縮信号が予め決められた目標データ量以下であるか否かが判定される（ステップ S 11）。ステップ S 11 において、入力された音声圧縮信号が目標データ量より大きいと判定された場合（ステップ S 11 ; NO）、データ削除部 10 において、エネルギーが小さい帯域、即ち、 $|F_j|^2$ が所定値以下の帯域の信号が削除され（ステップ S 12）、生起確率テーブル 8 1 が、ステップ S 9 で保存していた値に戻され（ステップ S 13）、レンジコードの現在の生起確率テーブル 8 1 の値が保存され（ステップ S 9）、再度、該当するフレームのレンジコード符号化処理が行われる（ステップ S 10）。

30

【 0 0 5 3 】

ステップ S 11 において、入力された音声圧縮信号が目標データ量以下であると判定された場合（ステップ S 11 ; YES）、エントロピー符号化部 8 に次のフレームの音声信号が入力されたか否かが判定される（ステップ S 14）。ステップ S 14 において、エントロピー符号化部 8 に次のフレームの音声信号が入力されたと判定された場合（ステップ S 14 ; YES）、レンジコードの現在の生起確率テーブル 8 1 の内容が一旦保存され（ステップ S 9）、再度、そのフレームに対するレンジコード符号化処理が行われる（ステップ S 10）。ステップ S 14 において、エントロピー符号化部 8 に入力された全てのフレームに対するレンジコード符号化が終了したと判定された場合（ステップ S 14 ; YES）、本音声圧縮処理が終了する。

40

【 0 0 5 4 】

50

次に、図 11 のフローチャートを参照して、エントロピー符号化部 8 において実行されるレンジコーダ符号化 (図 10 のステップ S 10) の詳細について説明する。

【 0055 】

まず、low 及び range が初期化され (ステップ S 20) 、式 (10) により、各記号の出現確率の合計 cum が算出される (ステップ S 21) 。音声信号を示す記号 s_k が入力されると (ステップ S 22) 、式 (12) により range' が算出され、式 (13) により low' が算出され、算出された range' 及び low' が、それぞれ、次の記号が入力されたとき range、low のとなる (ステップ S 23) 。

【 0056 】

次いで、記号 s_k の出現頻度 f_k がインクリメントされる (ステップ S 24) 。出現頻度 f_k がインクリメントされたことにより、各記号の出現確率 p_i と F_i が再計算され、生起確率テーブル 81 が更新される。次いで、1 フレーム分の記号列に対して、ステップ S 21 ~ S 24 までの処理が終了したか否かが判定される (ステップ S 25) 。

【 0057 】

ステップ S 25 において、1 フレーム分の記号列に対する処理が終了していないと判定された場合 (ステップ S 25 ; NO) 、ステップ S 21 に戻り、次に入力される記号に関して、ステップ S 21 ~ S 24 までの処理が繰り返される。ステップ S 25 において、1 フレーム分の記号列に対する処理が終了したと判定された場合 (ステップ S 25 ; YES) 、low ~ low+range の範囲の一つの数値が出力され (ステップ S 26) 、本レンジコーダ符号化処理が終了する。

【 0058 】

図 11 に示すレンジコーダ符号化では、生起確率テーブルの初期化後に入力された全ての記号に基づいて生起確率テーブルが更新される例を示したが、出現確率の変動が局所的に大きい場合や、出現確率が過去の入力にほとんど依存しない場合 (例えば、2 つの異なる内容の音声ファイルが連結されて一つのファイルになっている場合) は、過去の入力に影響されない。従って、このような場合、一定期間以上過去に処理した記号の出現頻度の影響を生起確率テーブルから取り除くようにすれば、符号化効率のよい生起確率テーブルを生成することが可能になる。

【 0059 】

以下、図 12 のフローチャートを参照して、図 11 のレンジコーダ符号化の変形例として、一定期間以上過去に処理した記号の出現頻度の影響を生起確率テーブルから取り除く場合のレンジコーダ符号化処理 (レンジコーダ符号化処理 2 と呼ぶ。) について説明する。以下では、エントロピー符号化部 8 に、入力信号を記録するための F I F O (First-In First-Out) が備えられているものとする。

【 0060 】

まず、low 及び range が初期化され (ステップ S 30) 、式 (10) により、各記号の出現確率の合計 cum が算出される (ステップ S 31) 。音声信号を示す記号 s_k が入力されると (ステップ S 32) 、式 (12) により range' が算出され、式 (13) により low' が算出され、算出された range' 及び low' が、それぞれ、次の記号が入力されたとき range、low のとなる (ステップ S 33) 。

【 0061 】

次いで、記号 s_k の出現頻度 f_k がインクリメントされ (ステップ S 34) 、F I F O に記号 s_k が入力され、 x 個前に F I F O に入力された記号 s_y が F I F O から取り出される (ステップ S 35) 。ここで、 x は、2 以上の整数である。記号 s_y が F I F O から取り出されると、記号 s_y の出現頻度 f_y がデクリメントされる (ステップ S 36) 。

【 0062 】

出現頻度 f_k がインクリメントされ、出現頻度 f_y がデクリメントされたことにより、各記号の出現確率 p_i と F_i が再計算され、生起確率テーブル 81 が更新される。次いで、1 フレーム分の記号列に対して、ステップ S 31 ~ S 36 までの処理が終了したか否かが判定される (ステップ S 37) 。

10

20

30

40

50

【 0 0 6 3 】

ステップ S 3 7 において、1 フレーム分の記号列に対する処理が終了していないと判定された場合（ステップ S 3 7 ; N O）、ステップ S 3 1 に戻り、次に入力される記号に関して、ステップ S 3 1 ~ S 3 6 までの処理が繰り返される。ステップ S 3 7 において、1 フレーム分の記号列に対する処理が終了したと判定された場合（ステップ S 3 7 ; Y E S）、 $low \sim low + range$ の範囲の一つの数値が出力され（ステップ S 3 8）、本レンジコーダ符号化処理 2 が終了する。

【 0 0 6 4 】

以上のように、本実施形態の音声処理装置 1 0 0 によれば、入力された音声信号をフレームに分割し、フレーム毎にレベル調整し、レベル調整された音声信号に M D C T を施し、M D C T 係数を周波数毎に並べ替えたベクトルにベクトル量子化を施し、ベクトル量子化により得られた音声信号にエントロピー符号化を施すことにより、語学学習用会話等の録音に適切なサンプリング周波数で、符号化レートを低減させることができる。例えば、本実施形態の音声処理装置 1 0 0 により、1 6 k H z 程度のサンプリング周波数の音声信号を 1 6 k b p s 程度のレートに圧縮することが可能になる。

10

【 0 0 6 5 】

また、エントロピー符号化としてレンジコーダ符号化を用いることにより、ビット操作が不要になり、処理負荷を軽減させることができる。特に、1 記号が入力される度に生起確率テーブルを更新するようにしたことにより、話者の変更等による音声信号の変化に対応することができる。

20

【 0 0 6 6 】

更に、エントロピー符号化で得られた符号のデータ量と目標データ量を比較し、エントロピー符号化で得られた符号のデータ量が目標データ量より大きい場合、エネルギーが小さい帯域の信号を削除するようにしたため、再度、ベクトル量子化を行う必要がなく、符号化処理を高速に行うことが可能になる。

【 0 0 6 7 】

なお、本実施形態における記述内容は、本発明の趣旨を逸脱しない範囲で適宜変更可能である。

【 0 0 6 8 】

例えば、上述の実施形態では、周波数毎にまとめられた M D C T 係数のうち、エネルギーが小さい（エネルギーが所定値以下）帯域の M D C T 係数を削除することで、データ量を削減するようにしたが、周波数毎にまとめられた M D C T 係数のうち、存在するデータ数が少ない、即ち、存在するデータ数が所定値以下の帯域の M D C T 係数を削除するようにしてもよい。

30

【 0 0 6 9 】

また、上述の実施形態では、ベクトル量子化の際、高域周波数と低域周波数で異なる V Q テーブルを用いるようにしたが、音声の全ての周波数帯域で同一の V Q テーブルを用いるようにしてもよい。また、周波数帯域を更に細かく分け、周波数帯域毎に異なる V Q テーブルを用いるようにしてもよい。

【 図面の簡単な説明 】

40

【 0 0 7 0 】

【 図 1 】 本発明の実施形態の音声処理装置 1 0 0 の構成を示すブロック図。

【 図 2 】 入力信号のフレーム分割を示す図。

【 図 3 】 1 フレームと各 M D C T ブロックの関係を示す図。

【 図 4 】 ハフマン符号の一例を示す図。

【 図 5 】 従来のレンジコーダによる符号化の一例を示す図。

【 図 6 】 本実施形態のレンジコーダ符号化に必要な生起確率テーブル 8 1 のデータ構成を示す図。

【 図 7 】 デフォルトの生起確率テーブル 8 1 の一例（同図（ a ））と、符号化の一例（同図（ b ））を示す図。

50

【図8】 生起確率テーブル81の更新例を示す図。

【図9】 圧縮された音声信号を復号する復号装置200の構成を示すブロック図。

【図10】 音声処理装置100において実行される音声圧縮処理を示すフローチャート。

【図11】 図10のレンジコーダ符号化処理の詳細を示すフローチャート。

【図12】 図11のレンジコーダ符号化処理の変形例を示すフローチャート。

【符号の説明】

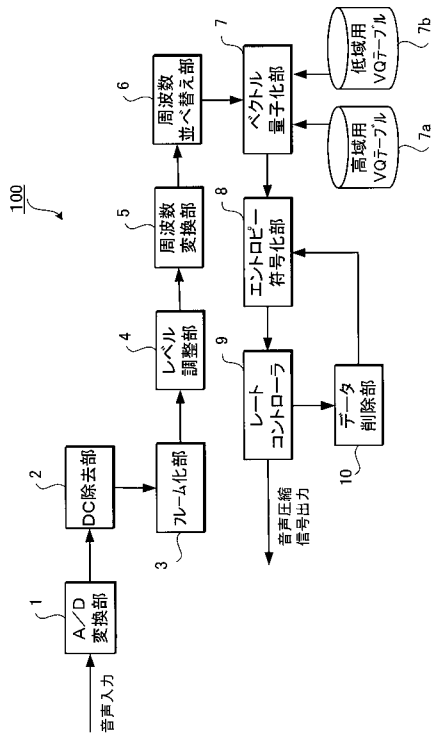
【0071】

- 1 A/D変換部
- 2 DC除去部
- 3 フレーム化部
- 4 レベル調整部
- 5 周波数変換部
- 6 周波数並べ替え部
- 7 ベクトル量子化部
- 7a 高域用VQテーブル(第1のテーブル)
- 7b 低域用VQテーブル(第2のテーブル)
- 8 エントロピー符号化部
- 81 生起確率テーブル
- 9 レートコントローラ(判定部)
- 10 データ削除部
- 100 音声処理装置

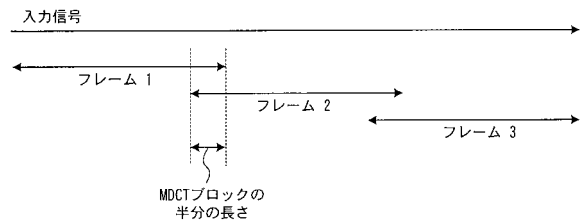
10

20

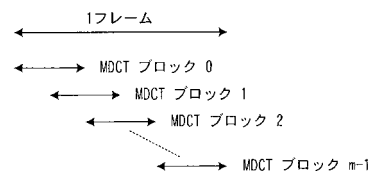
【図1】



【図2】



【図3】



【 図 4 】

ハフマン符号の例

記号	出現頻度	符号	符号長[bit]
a	10	100	3
b	70	0	1
c	1	101	3
d	19	11	2

【 図 5 】

(a) 生起確率テーブル

s_i	p_i	F_i
a	0.5	0
b	0.25	0.5
c	0.15	0.75
d	0.1	0.9

(b) 記号列 [bacal] の符号化の例
初期値 : low=0, range= 10^6

入力信号	記号例	low	range	範囲
b	b	500000	250000	[500000, 750000)
a	ba	500000	125000	[500000, 625000)
c	bac	593750	018750	[593750, 612500)
a	baca	593750	009375	[593750, 603125)

【 図 6 】

生起確率テーブル 81

s_i	f_i	p_i	F_i
s_1	f_1	p_1	0
s_2	f_2	p_2	p_1
s_3	f_3	p_3	p_1+p_2
\vdots	\vdots	\vdots	\vdots
s_n	f_n	p_n	$\sum_{i=1}^{n-1} p_i$

【 図 7 】

(a) 生起確率テーブル 81 (デフォルト)

s_i	f_i	p_i	F_i
a	10	10/20	0
b	5	5/20	10/20
c	3	3/20	15/20
d	2	2/20	18/20

(b) 記号列 [bacal] の符号化の例
初期値 : low=0, range= 10^6

入力信号	記号例	low	range	範囲	生起確率テーブル
b	b	500000	250000	[500000, 750000)	81-1
a	ba	500000	119048	[500000, 619048)	81-2
c	bac	591992	016234	[591992, 608226)	81-3
a	baca	591992	007765	[591992, 599757)	81-4

【 図 8 】

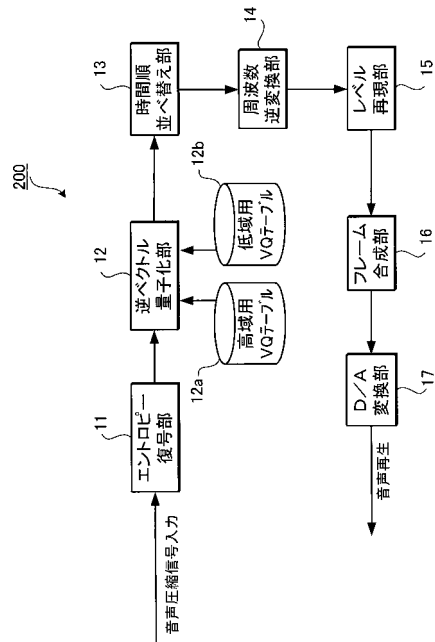
81-1			
s_i	f_i	p_i	F_i
a	10	10/21	0
b	6	6/21	10/21
c	3	3/21	16/21
d	2	2/21	19/21

81-2			
s_i	f_i	p_i	F_i
a	11	11/22	0
b	6	6/22	11/22
c	3	3/22	17/22
d	2	2/22	20/22

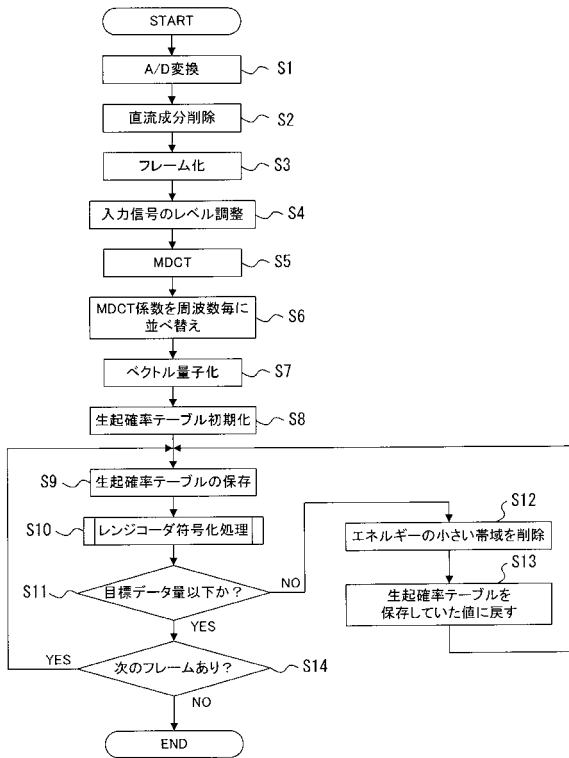
81-3			
s_i	f_i	p_i	F_i
a	11	11/23	0
b	6	6/23	11/23
c	4	4/23	17/23
d	2	2/23	21/23

81-4			
s_i	f_i	p_i	F_i
a	12	12/24	0
b	6	6/24	12/24
c	4	4/24	18/24
d	2	2/24	22/24

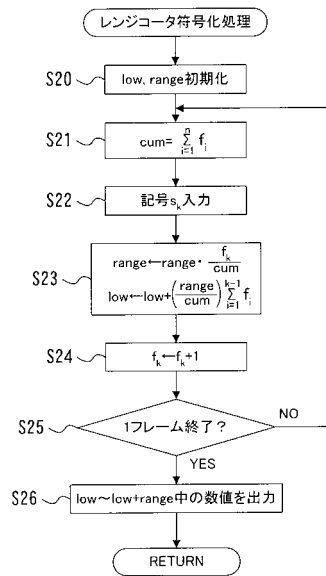
【 図 9 】



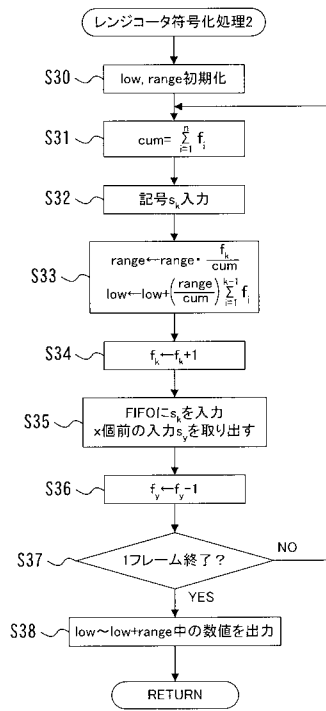
【 図 10 】



【 図 11 】



【 図 1 2 】



フロントページの続き

- (56)参考文献 特開平9 - 130260 (JP, A)
特開平7 - 183854 (JP, A)
特開平6 - 268606 (JP, A)
特開平11 - 252573 (JP, A)
特開平7 - 202823 (JP, A)
特開2002 - 91497 (JP, A)
特開2000 - 132195 (JP, A)
特開平3 - 65822 (JP, A)
特開平3 - 121624 (JP, A)
特開昭64 - 53642 (JP, A)
特開2002 - 202799 (JP, A)

(58)調査した分野(Int.Cl., DB名)

G10L 19/00 - 19/14
IEEE
JST7580 (JDream2)
JSTPlus (JDream2)