

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第3622365号
(P3622365)

(45) 発行日 平成17年2月23日(2005.2.23)

(24) 登録日 平成16年12月3日(2004.12.3)

(51) Int.CI.⁷

F 1

G 1 O L 19/00

G 1 O L 9/00

N

G 1 O L 19/12

G 1 O L 9/14

S

G 1 O L 9/18

A

請求項の数 7 (全 10 頁)

(21) 出願番号

特願平8-254967

(22) 出願日

平成8年9月26日(1996.9.26)

(65) 公開番号

特開平10-105193

(43) 公開日

平成10年4月24日(1998.4.24)

審査請求日

平成15年2月6日(2003.2.6)

(73) 特許権者 000004075

ヤマハ株式会社

静岡県浜松市中沢町10番1号

(74) 代理人 100092820

弁理士 伊丹 勝

(72) 発明者 山内 健一

静岡県浜松市中沢町10番1号 ヤマハ株式会社内

審査官 山下 剛史

最終頁に続く

(54) 【発明の名称】音声符号化伝送方式

(57) 【特許請求の範囲】

【請求項1】

音声信号を送信部で符号化して伝送路を介して伝送し受信部で受信して復号することより前記音声信号を再生する音声符号化伝送方式において、

前記送信部は、前記音声信号を符号化して得た符号化出力を、前記音声信号を逐次再生可能な低ビットレートの概略データと、この概略データと組み合わせて前記音声信号を高品質で再生するための詳細データとに分割し、前記概略データを前記詳細データに先立ってまとめて伝送し、その後に前記詳細データをまとめて伝送するものであり、

前記受信部は、受信された前記概略データを前記詳細データの受信を待つことなく逐次復号して前記音声信号を逐次再生し、前記概略データ及び詳細データが共に伝送されたら、両者を合成して前記音声信号を高品質で再生するものである

ことを特徴とする音声符号化伝送方式。

【請求項2】

前記送信部は、前記伝送路の状況に応じて前記概略データのビットレートを制御するものであることを特徴とする請求項1記載の音声符号化伝送方式。

【請求項3】

前記送信部は、前記音声信号をアナログ・ディジタル変換して得られたnビット(nは2以上の整数)のPCM(Pulse Code Modulation)データの上位mビット(m < n)を前記概略データ、下位n-mビットを前記詳細データとするものであり、

前記受信部は、前記PCMデータの上位mビットから前記音声信号を逐次再生し、前記PCMデータの上位mビットと下位n-mビットの合成データから前記音声信号を高品質で再生するものである

ことを特徴とする請求項1又は2記載の音声符号化伝送方式。

【請求項4】

前記送信部は、ステレオ音声を得るための複数チャネルの音声信号を加算したモノラル音声信号を前記概略データ、前記複数チャネルの音声信号の差分信号を前記詳細データとするものであり、

前記受信部は、前記モノラル信号から前記音声信号を逐次再生し、前記モノラル信号及び差分信号からステレオ音声を再生するものである

ことを特徴とする請求項1又は2記載の音声符号化伝送方式。

【請求項5】

前記送信部は、音声信号の低域成分を前記概略データ、前記音声信号の高域成分を前記詳細データとするものであり、

前記受信部は、前記音声信号の低域成分から前記音声信号を逐次再生し、前記音声信号の低域成分及び高域成分から前記音声信号を高品質に再生するものである

ことを特徴とする請求項1又は2記載の音声符号化伝送方式。

【請求項6】

前記送信部は、第1及び第2の符号帳を有し、前記音声信号に基づくターゲットベクトルに対して第1の符号帳から最適なベクトル選択を行い、次にそのベクトルと組み合わせて最もターゲットベクトルに近づくサブベクトルを第2の符号帳から選択する2段ベクトル量子化処理を実行し、前記第1の符号帳から選択されたベクトルのインデックスを前記概略データ、前記第2の符号帳から選択されたベクトルのインデックスを前記詳細データとするものであり、

前記受信部は、前記第1の符号帳から選択されたベクトルのインデックスから前記音声信号を逐次再生し、前記第1及び第2の符号帳から選択されたベクトルのインデックスから前記音声信号を高品質に再生するものである

ことを特徴とする請求項1又は2記載の音声符号化伝送方式。

【請求項7】

前記送信部は、前記受信部から送られる詳細データの伝送要求を待って前記詳細データを前記受信部に伝送するものであることを特徴とする請求項1乃至6のいずれか1項記載の音声符号化伝送方式。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

この発明は、音声信号を符号化して伝送する音声符号化伝送方式に関し、特に音声情報のリアルタイム伝送を行うのに好適な音声符号化伝送方式に関する。

【0002】

【従来の技術】

音声情報を通信回線を介して伝送する場合、従来は、全ての音声情報が伝送されないと音声情報の再生が行われない。このため、音声情報をリアルタイムに伝送して再生する場合には、通信帯域を考慮して最適な伝送ビットレート、即ち符号化圧縮率を選択することが重要となる。例えば、ある程度の再生品質を望む場合、符号化圧縮率をあまり高めることができず、音声情報が全て伝送されるまでに待ち時間が発生する。再生品質を多少落としても、軽く試聴してみたいという要求に対しては、符号化圧縮率を高めて再生品質を低下させればよい。

【0003】

従来、このような目的で、復号時の再生品質や符号化圧縮率を任意に選択できるスケーラブルな音声符号化方式が提案されている（例えば「スケーラブルな階層構造を持つ楽音・音声の符号化」神明夫、三樹聰：日本音響学会講演論文集，H7.9音声B3-1-5，

p p 2 7 7 - 2 7 8)。この方式は、3階層の符号化方式を用い、第1階層では入力信号をダウンサンプリングした狭帯域信号を符号化し、その他の上位階層では、下位層よりも帯域を広げた入力信号から下位層までの再生音の差分をとり、これを符号化する。このような階層構造によって、低ビットレートの場合（通信帯域に余裕がない場合）は、下位層のみの情報を用いて低品質の符号化を実行し、高ビットレートの場合（通信帯域に余裕がある場合）は、上位層の情報も用いて、再生音を広帯域且つ高品質で再生する。

【0004】

【発明が解決しようとする課題】

しかしながら、上述した従来の音声符号化伝送方式では、受信側で音声情報をリアルタイムで軽く試聴した結果、より品質の良い音声情報を伝送を望む場合には、再度、高ビットレートのビットストリームを最初から伝送し直す必要があり、その際の待ち時間が長くなるという問題がある。10

【0005】

この発明は、このような問題点に鑑みなされたもので、低速な伝送路であっても受信側で待ち時間なしに音声情報を試聴することができ、試聴の結果、同じ音声情報を高品質で再度聴きたい場合にも、少ない待ち時間で高品質の音声情報を聴くことが出来る音声符号化伝送方式を提供することを目的とする。

【0006】

【課題を解決するための手段】

この発明に係る音声符号化伝送方式は、音声信号を送信部で符号化して伝送路を介して伝送し受信部で受信して復号することより前記音声信号を再生する音声符号化伝送方式において、前記送信部が、前記音声信号を符号化して得た符号化出力を、前記音声信号を逐次再生可能な低ビットレートの概略データと、この概略データと組み合わせて前記音声信号を高品質で再生するための詳細データとに分割し、前記概略データを前記詳細データに先立ってまとめて伝送し、その後に前記詳細データをまとめて伝送するものであり、前記受信部が、受信された前記概略データを前記詳細データの受信を待つことなく逐次復号して前記音声信号を逐次再生し、前記概略データ及び詳細データが共に伝送されたら、両者を合成して前記音声信号を高品質で再生することを特徴とする。20

【0007】

この発明によれば、送信部が逐次再生可能な低ビットレートの概略データを送信すると、受信部がこの概略データを、詳細データの受信を待つことなく逐次再生するので、音楽情報の概要を待ち時間なしに軽く試聴することができる。この結果、受信側のユーザが、更に高品質で同じ音声情報を聴きたい場合には、詳細データの伝送を待って音声情報の再生を行えばよい。詳細データは、概略データと合成されて再生されるので、従来のように全ての音声情報を再度伝送する場合に比べ、待ち時間は格段に軽減される。また、受信された概略データの試聴の結果、受信側のユーザの期待に添わなかった場合には、以後のデータの受信を中止することができ、これによるデータの無駄な伝送が防止できる。30

【0008】

なお、インターネットのように通信帯域が刻々と変動するようなネットワークを使用して音声情報を伝送する場合には、伝送路の状況に応じて概略データのビットレートを調整すればよい。この場合、伝送するビットストリームにデータレートの情報又はその変化の情報を適宜挿入しておけばよい。40

【0009】

概略データとしては、例えばnビットのPCMデータの上位mビットのデータ、ステレオ音声における左右チャネルの加算データ、音声信号の低域成分、2段ベクトル量子化における第1段のベクトルインデックス等を用いることができる。また、詳細データとしては、これらに対応して例えば上記PCMデータの下位n-mビットのデータ、上記ステレオ音声における左右チャネルの差分データ、音声信号の高域成分、2段ベクトル量子化における第2段のベクトルインデックス等をそれぞれ用いることができる。このように、詳細データは、音声情報の一部を構成するものであるから、それ自体の情報量は、同一の品質50

を得るために再送する場合の音声情報の全情報量に比べて格段に少なく、その結果、待ち時間も少なくすることができる。

【0010】

また、送信部が受信部から送られる詳細データの転送要求を待って詳細データを受信部に伝送するものであれば、無駄なデータの伝送を更に効果的に防ぐことができる。

【0011】

【発明の実施の形態】

以下、図面を参照して、この発明の好ましい実施の形態について説明する。

図1は、この発明の実施例に係る音声符号化伝送方式における送信部の構成を示すプロック図である。なお、この送信部としては、ネットワークの状況の如何に拘わらず、固定的な低速ビットレートの概略データを最初に送信する方式と、ネットワークの状況に応じて概略データのビットレートを調整する方式とが考えられる。そこで、以下、上記2通りの方式について送信部の構成をそれぞれ説明する。

【0012】

(1) 概略データのビットレートが固定の場合

音声入力信号は、スケーラブル・エンコーダ1に入力され、ここで所定の符号化方式に基づいて符号化されると共に、予め指示されたビットレート情報に基づいて固定的なビットレートの概略データと、この概略データを補完して高品質の再生を可能にするための詳細データとに分割される。概略データは、ヘッダや境界識別子等の付加情報と共にマルチプレクサ2に供給される。詳細データは、記憶装置3に一旦記憶されたのち、並べ替え部4の制御のもとで読み出され、マルチプレクサ2に供給される。マルチプレクサ2は、供給されたデータをヘッダ、概略データ、境界識別子、詳細データの順番に、データ送信部5に順次供給する。データ送信部5は、マルチプレクサ2から供給されたビットストリームを伝送路を介して受信部に送信する。ビットレート制御部6は、指示入力に従って概略データのビットレートを決定し、ビットレート情報をスケーラブル・エンコーダ1に与える。なお、ビットレートは、伝送を行うネットワークが低速であっても、リアルタイム伝送が可能な情報量となるように送信側で設定する。

【0013】

この送信部のマルチプレクサ2で組み立てられるビットストリームのフォーマット例を図2(a)に示す。ビットストリームは、ヘッダ、概略データ、境界識別子、詳細データの順に組み立てられる。

【0014】

(2) 概略データのビットレートが可変の場合

この場合、ネットワーク状況監視部7がデータ送信部5を介してネットワークの状況を常に監視し、通信帯域が変動した場合には、ビットレート制御部6に指示情報を与えて概略データのビットレートを調整する。ビットレートの調整方法としては、符号化ビット数を落としたり、部分的にデータを間引く等の方法が考えられる。この場合、概略データと対をなす詳細データのビットレートは、概略データのビットレートの変化を吸収するような形で作成されることになる。

【0015】

ビットレートが可変の場合のマルチプレクサ2で組み立てられるビットストリームのフォーマットは、図2(b)のようになる。即ち、ヘッダに続く概略データには、ビットレートの変化する直前にビットレートが変化することを示す情報が挿入される。これにより、受信部で概略データのビットレートを認識することが可能になる。

【0016】

なお、詳細データは、更にいくつかに分割され、重要度の高い順に並び替えて読み出されるようにしても良いし、曲の中の重要度の高い部分が最初に送信されるようにフレームを並べ替えるようにしても良い。音楽データの場合、例えば図3(a)に示すように、データ作成者がリスナーに高品質で聴いてほしい、いわゆる“さび”の部分(この例ではフレーム番号16～20)を、同図(b)に示すように詳細データの最初に配置する。これによ

10

20

30

40

50

り、詳細データのダウンロードを途中で中止した場合でも、作成者が聴いてほしい部分については、高品質な再生音で聞くことが出来る。このような詳細データの並べ替えは、並べ替え部4にて行われる。

【0017】

図4は、受信部の構成を示すブロック図である。

ネットワークを介して伝送されたビットストリームは、データ受信部11で受信され、記憶装置12に一旦格納されたのち、逐次デコーダ13に供給される。逐次デコーダ13は、ヘッダが検出されてから境界識別子が検出されるまでのデータを概略データであると認識して、逐次デコードを開始する。これにより、リアルタイムでのデコードが実現され、再生品質は若干落ちるもの、軽く試聴するには十分の再生音がリアルタイムで得られる。

【0018】

また、ネットワークの混雑状況によっては、ビットレートをかなり下げてもリアルタイムの伝送が不可能になる場合がある。この場合には、ネットワーク状況監視部14がネットワークの状況を監視して、その監視結果をデコード制御部15に伝送する。デコード制御部15は、ネットワークの状況に応じて、例えば1分待ち、2分待ち、…のように、待ち時間を設定し、この待ち時間の間は、データを受信するだけで逐次デコードを行わないようにしてと共に、待ち時間が経過した後は連続的な逐次デコードを開始するように逐次デコーダ15を制御する。これにより、音切れのないほぼリアルタイムに近いデコードが可能になる。

【0019】

詳細データは、この概略データに統いて受信されることになるが、受信部での試聴結果によってリストナが更に高品質の音声情報の伝送を希望するか否かを選択できるようにしておき、この選択要求が受信部から送信部に伝送されるのを待って、詳細データを送信するようにも良い。

【0020】

詳細データの送信も完了し、記憶装置12に詳細データが全て格納された場合には、データ合成部16が既に伝送されている概略データと伝送が終了した詳細データとを合成する。詳細データを更に複数に分割した場合には、データ合成部16で概略データと複数の詳細データとが合成されることになる。合成されたデータは、デコーダ17によってデコードされる。これにより、音声信号を高品質で再生することができる。

【0021】

次に、スケーラブル・エンコーダ1の更に具体的な実施例について説明する。

図5は、PCMデータを上位ビットと下位ビットとに分割し、上位ビットを概略データ、下位ビットを詳細データとしたスケーラブル・エンコーダ1の構成例を示す図である。

音声入力信号は、A/D変換器21で所定サンプリング周波数でアナログ・ディジタル変換され、例えば16ビットのPCMデータとなる。このPCMデータは、ビット分割部22において、例えば上位8ビットの概略データと、下位8ビットの詳細データとに分割される。付加情報生成部23は、ヘッダや境界識別子等の付加情報を生成するもので、音声入力信号の先頭にヘッダを生成して出力し、音声入力信号の終わりに境界識別子を生成して出力する。

【0022】

ビットレートが固定的である場合には、図6(a)に示すように、ビットストリームは、ヘッダ、PCMデータの上位8ビットのデータ、境界識別子、PCMデータの下位8ビットのデータの順に並べられて伝送され、受信部では、上位8ビットのPCMデータをディジタル・アナログ変換してリアルタイムで音声を再生し、下位8ビットのPCMデータも伝送されたら、対応するフレームの上位8ビットと下位8ビットとで16ビットのPCMデータを合成してディジタル・アナログ変換することにより、高品質の音声を再生する。概略データとしての上位8ビットのPCMデータは、フルビットで送信する場合に比べてビットレートを1/2まで低下させることができるので、低速のネットワークでもリアル

タイムでの音声伝送が可能になる。また、詳細データとしての下位 8 ビットの P C M データの传送待ち時間も、フルビットのデータを再送する場合に比べ、1 / 2 に短縮することができる。

【 0 0 2 3 】

ネットワーク状況に応じてビットレートを変更する場合には、ビットレート情報に応じてビット分割部 2 2 が上位のビット数と下位のビット数とを変更し、付加情報生成部 2 3 は、ビットレートの変更を示す情報を出力する。P C M データが 1 6 ビットであるとともに、例えば図 6 (b) に示すように、ヘッダに続く最初の概略データは、上位 8 ビットに固定しておき、8 ビットのデータの境界に “ 0 ” (ビットレート変更なし) 又は “ 1 ” (ビットレートの変更あり) のデータを挿入しておく。そして、“ 1 ” (ビットレートの変更あり) が挿入された場合には、それに続く 2 ビットをビットレートの変更を示す情報として、10

0 0	1 ビット増加
0 1	2 ビット増加
1 0	1 ビット減少
1 1	2 ビット減少

のように、増減するビット数を指示するようにすればよい。

【 0 0 2 4 】

図 7 は、多チャネルの音声信号のうち 1 チャネルを概略データ、残りのチャネルを詳細データとした場合のスケーラブル・エンコーダ 1 の例である。20

第 1 及び第 2 チャネルの音声信号は、例えば左右チャネルのステレオ信号であり、第 3 チャネルは、例えばスーパーウォーハー信号である。これら各チャネルの信号は、A / D 変換器 3 1 , 3 2 , 3 3 でそれぞれ A / D 変換される。第 1 及び第 2 チャネルのデジタルデータは、加算器 3 4 で加算されてモノラル信号となり、このモノラル信号が概略データとして出力される。一方、第 1 及び第 2 チャネルのデジタルデータは、減算器 3 5 で差分信号 (1 c h - 2 c h) となる。そして、この差分信号と第 3 チャネルの信号とが詳細データとして出力される。付加情報生成部 3 6 は、モノラル信号の先頭でヘッダを生成出し、末尾で境界識別子を生成出力する。これにより、生成されるビットストリームは、図 8 のように、ヘッダ、(1 c h + 2 c h) のデータ、境界識別子、(1 c h - 2 c h) 及び 3 c h のデータの順に続くデータとなる。30

【 0 0 2 5 】

図 9 は、M P E G (M o v i n g P i c t u r e E x p e r t s G r o u p) - A u d i o にこの発明を適用した場合のスケーラブル・エンコーダ 1 のブロック図である。入力音声は、サブバンド分割部 4 1 で例えば 3 2 の周波数帯域に分割され、各バンド毎にサブサンプリングされて 1 バンド 1 2 サンプル程度のデータとなる。このデータがスケーリング部 4 2 で、正規化された波形と倍率とに分離され、量子化部 4 3 に供給される。一方、入力信号は、F F T (F a s t F o u r i e r T r a n s f o r m) 部 4 4 において、高速フーリエ変換される。求められた音声スペクトラムと、スケーリング部 4 2 で求められたバンド毎の波形及び倍率に基づき、マスキング効果も加味した聴覚心理モデル 4 5 に応じたビット割り当てがビット割り当て部 4 6 にて行われる。このビット割り当てに応じて量子化部 4 3 が各バンド毎の波形と倍率とを量子化する。ビットストリーム構成部 4 7 は、このうち、低域部分のデータを概略データとして出力し、高域部分のデータを詳細データとして出力する。低域部分にどの程度のビットレートを割り当てるかは、ビットレート情報に基づき、ビット割り当て部 4 6 及びビットストリーム構成部 4 7 が決定する。生成されたビットストリームのフォーマットは、図 1 0 (a) のように、ヘッダ、低域部分のデータ、境界識別子、高域部分のデータが順に並んだものとなる。また、ビットレートを可変にした場合には、同図 (b) のように、ヘッダに続く低域部分のデータのビットレートの変化の直前部分にビットレートの変化を示す情報が挿入されることになる。40

【 0 0 2 6 】

10

20

30

40

50

図11は、多段ベクトル量子化にこの発明を適用した場合のスケーラブル・エンコーダ1の構成を示すブロック図である。

音声入力信号又はこれをMDCT (Modified Discrete Cosine Transform) 等で直交変換したデータをターゲットベクトルとし、第1段出力ベクトル決定部51では、このターゲットベクトルに対して第1の符号帳52から最適なベクトル選択を行い、第2段出力ベクトル決定部53では、前段で選択されたベクトルとの組み合わせで最もターゲットベクトルに近づくサブベクトルを歪計算部55での歪計算結果に基づいて第2の符号帳54から選択する。第1の符号帳52から選択されたメインインデックスと第2の符号帳54から選択されたサブインデックスとがビットストリーム構成部56でビットレート情報に基づいて合成されてビットストリームが生成される。この場合、メインインデックスだけでも、ある程度の品質の音声の再生は可能であるため、メインインデックスを概略データ、サブインデックスを詳細データとして用いることができる。

【0027】

ビットレートを固定的に設定した場合には、ビットストリームは、図12に示すように、ヘッダ情報、概略データとしてのメインインデックス、境界識別子、詳細データとしてのサブインデックスの順に伝送される。ビットレートをネットワーク状況に応じて変化させる場合には、メインインデックスを適当に間引きしたり、メインインデックスにサブインデックスも加えて概略データとすればよい。

【0028】

【発明の効果】

以上述べたように、この発明によれば、送信部が逐次再生可能な低ビットレートの概略データを送信し、受信部がこの概略データを、詳細データの受信を待つことなく逐次再生するので、音楽情報の概要を待ち時間なしに軽く試聴することができ、しかも受信側のユーザが更に高品質で同じ音声情報を聴きたい場合には、詳細データの伝送を待って音声情報の再生を行えばよく、この場合の待ち時間も従来に比べて格段に軽減され、且つ受信された概略データの試聴の結果、受信側のユーザの期待に添わなかった場合には、以後のデータの受信を中止することができ、これによるデータの無駄な伝送が防止できるという効果を奏する。

【図面の簡単な説明】

【図1】この発明の一実施例に係る音声符号化伝送方式の送信部のブロック図である。

【図2】同送信部から送信されるビットストリームのフォーマットを示す図である。

【図3】同ビットストリームにおける詳細データの並べ替えの例を示す図である。

【図4】同音声符号化伝送方式の受信部のブロック図である。

【図5】同送信部におけるスケーラブル・エンコーダの第1の実施例を示すブロック図である。

【図6】同エンコーダにより生成されるビットストリームのフォーマットを示す図である。

【図7】同送信部におけるスケーラブル・エンコーダの第2の実施例を示すブロック図である。

【図8】同エンコーダにより生成されるビットストリームのフォーマットを示す図である。

【図9】同送信部におけるスケーラブル・エンコーダの第3の実施例を示すブロック図である。

【図10】同エンコーダにより生成されるビットストリームのフォーマットを示す図である。

【図11】同送信部におけるスケーラブル・エンコーダの第4の実施例を示すブロック図である。

【図12】同エンコーダにより生成されるビットストリームのフォーマットを示す図である。

10

20

30

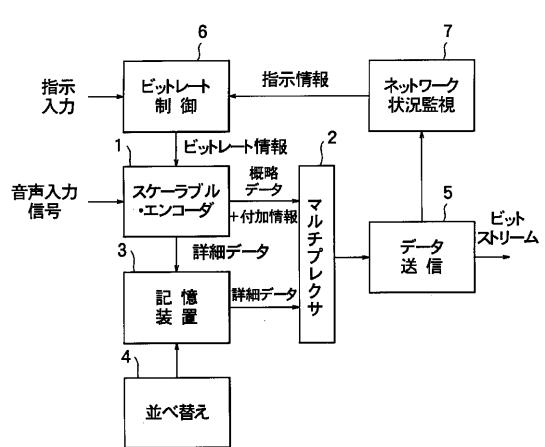
40

50

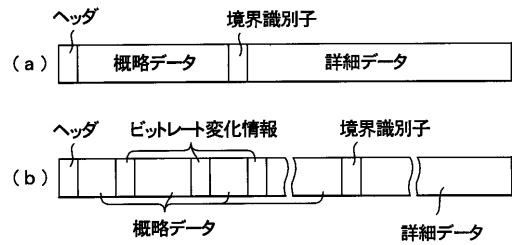
【符号の説明】

1 …スケーラブル・エンコーダ、2 …マルチプレクサ、3 , 1 2 …記憶装置、4 …並べ替え部、5 …データ送信部、6 …ビットレート制御部、7 , 1 4 …ネットワーク状況監視部、1 1 …データ受信部、1 3 …逐次デコーダ、1 5 …デコード制御部、1 6 …データ合成功部、1 7 …デコーダ。

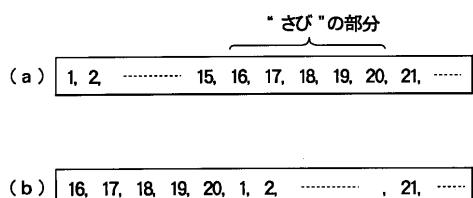
【 図 1 】



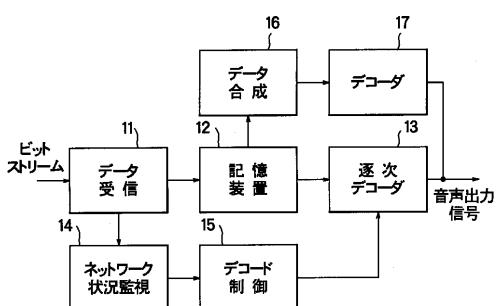
【 図 2 】



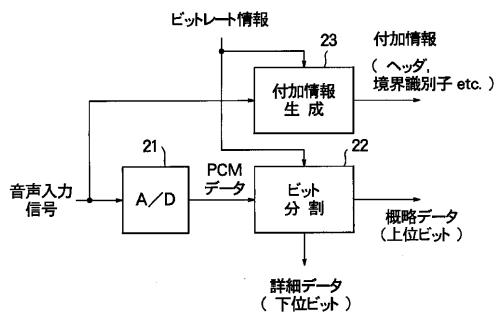
【 図 3 】



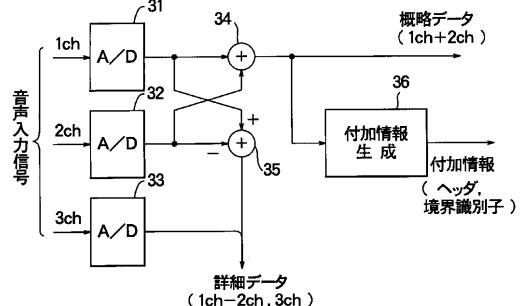
【 図 4 】



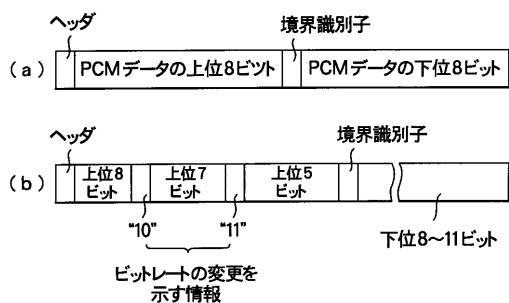
【図5】



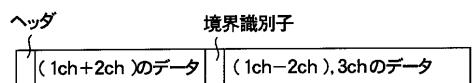
【図7】



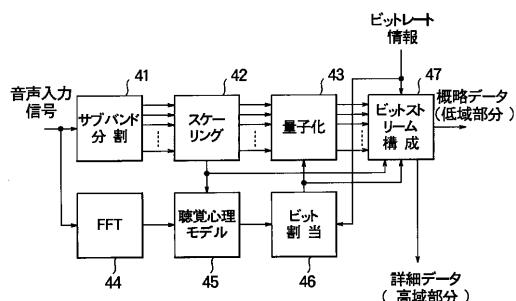
【図6】



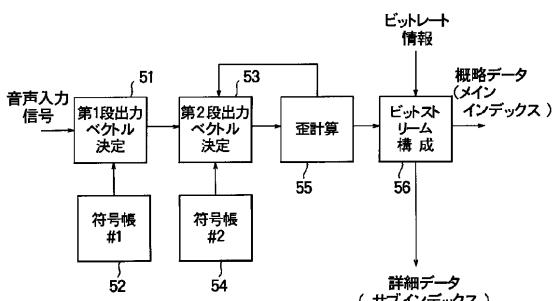
【図8】



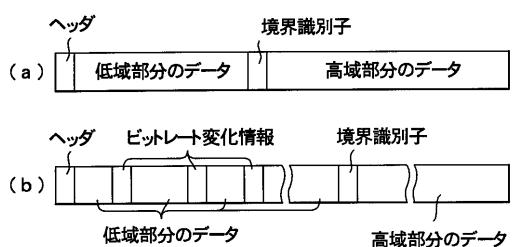
【図9】



【図11】



【図10】



【図12】



フロントページの続き

(56)参考文献 特開平03-044187(JP,A)
特開平02-143735(JP,A)
特開平06-069811(JP,A)
特開平01-215141(JP,A)
国際公開第95/17745(WO,A1)

(58)調査した分野(Int.Cl.⁷, DB名)

G10L 19/00-19/14

H03M 7/30

H04B 14/04