

(19) 世界知的所有権機関
国際事務局



(43) 国際公開日
2006年5月26日 (26.05.2006)

PCT

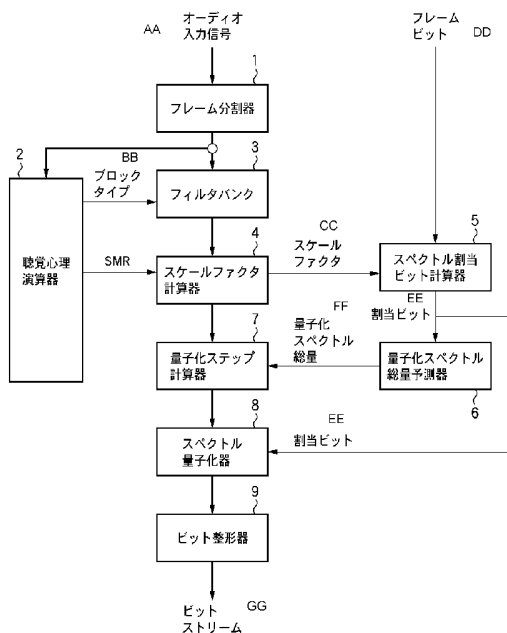
(10) 国際公開番号
WO 2006/054583 A1

- (51) 国際特許分類:
G10L 19/02 (2006.01)
- (21) 国際出願番号: PCT/JP2005/021014
- (22) 国際出願日: 2005年11月16日 (16.11.2005)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2004-335005
2004年11月18日 (18.11.2004) JP
特願 2005-328945
2005年11月14日 (14.11.2005) JP
- (71) 出願人 (米国を除く全ての指定国について): キヤノン株式会社 (CANON KABUSHIKI KAISHA) [JP/JP]; 〒1468501 東京都大田区下丸子3-30-2 Tokyo (JP).
- (72) 発明者; および
- (75) 発明者/出願人 (米国についてのみ): 船越正伸 (FU-NAKOSHI, Masanobu) [JP/JP]; 〒1468501 東京都大田区下丸子3-30-2 キヤノン株式会社内 Tokyo (JP).
- (74) 代理人: 大塚康徳 (OHTSUKA, Yasunori); 〒1020094 東京都千代田区紀尾井町3番6号 秀和紀尾井町パークビル7F Tokyo (JP).
- (81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, KE, KG, KM, KN, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

[続葉有]

(54) Title: AUDIO SIGNAL ENCODING APPARATUS AND METHOD

(54) 発明の名称: オーディオ信号符号化装置および方法



- AA AUDIO INPUT SIGNAL
- 1 FRAME DIVIDING DEVICE
- 2 AUDITORY MENTALITY ARITHMETIC DEVICE
- BB BLOCK TYPE
- 3 FILTER BANK
- 4 SCALE FACTOR CALCULATING DEVICE
- CC SCALE FACTOR
- DD FRAME BIT
- 5 SPECTRUM ASSIGNMENT BIT CALCULATING DEVICE
- EE ASSIGNMENT BIT
- FF TOTAL QUANTIZED SPECTRUM AMOUNT
- 6 TOTAL QUANTIZED SPECTRUM AMOUNT PREDICTING DEVICE
- 7 QUANTIZATION STEP CALCULATING DEVICE
- 8 SPECTRUM QUANTIZING DEVICE
- 9 BIT SHAPING DEVICE
- GG BITSTREAM

(57) Abstract: There are included a frame dividing part (1); an auditory mentality arithmetic part (2); a filter bank (3); a scale factor calculating part (4) for weighting the spectrum of each frequency band by use of an arithmetic result of the auditory mentality arithmetic part (2); a quantization step deciding part (7) for deciding a quantization step of the whole frame prior to a spectrum quantization by subtracting, from the auditory information amount of the whole weighted spectrum before the quantization, the information amount of the whole spectrum after the quantization and then multiplying the result of the subtraction by a factor obtained from a quantization roughness increment width; a spectrum quantizing part (8); and a bit shaping part (9) for providing a bitstream obtained by shaping the quantized spectrum. The quantization step deciding part (7) predicts, based on the bit amount assigned to a frame to be encoded, the information amount of the whole quantized spectrum.

(57) 要約: フレーム分割部(1)、聴覚心理演算部(2)、フィルタバンク(3)、聴覚心理演算部(2)の演算結果によって各周波数帯域のスペクトルを重み付けするスケールファクタ算出部(4)、重み付けされた量子化前のスペクトル全体が持つ聴覚情報量から量子化後のスペクトル全体の情報量を減じ、量子化粗さの刻み幅から得られる係数を積算することによって、フレーム全体の量子化ステップをスペクトル量子化前に決定する量子化ステップ決定部(7)、スペクトル量子化部(8)、量子化スペクトルを整形したビットストリームを出力するビット整形部(9)、を備え、量子化ステップ決定部(7)は、符号化対象であるフレームに割り当てられたビット量に基づいて量子化スペクトル全体の情報量を予測する。



WO 2006/054583 A1



(84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

2文字コード及び他の略語については、定期発行される各PCTガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。

添付公開書類:

— 国際調査報告書

明 細 書

オーディオ信号符号化装置および方法

技術分野

[0001] 本発明は、オーディオ信号の符号化装置および方法に関する。

背景技術

[0002] 近年、高音質かつ高効率なオーディオ信号符号化技術が、DVD-Videoの音声トラック、携帯オーディオプレーヤー、音楽配信、家庭内LANにおけるホームサーバへの楽曲蓄積などに広く利用され、幅広く普及するとともにその重要性も増している。

[0003] このようなオーディオ信号符号化技術の多くは、変換符号化技術を利用して時間周波数変換を行っている。例えば、MPEG-2 AACやDolby Digital(AC-3)などでは、MDCT (Modified Discrete Cosine Transform) などの直交変換単体でフィルタバンクを構成している。また、MPEG-1 Audio Layer III (MP3) やATRAC (MD(ミニディスク)に利用されている符号化方式) では、QMF (Quadrature Mirror Filter)などのサブバンド分割フィルタと直交変換を多段接続してフィルタバンクを構成している。

[0004] これらの変換符号化技術では、人間の聴覚特性を利用したマスキング分析が行われる。そして、マスクされると判断されるスペクトル成分を取り除く、あるいはマスクされる量子化誤差を許容することにより、スペクトル表現のための情報量を削減し、圧縮効率を高めている。

[0005] また、これらの変換符号化技術では、その多くが、スペクトル成分を非線形量子化することにより、スペクトルが持つ情報量を圧縮している。例えば、MP3やAACでは、各スペクトル成分を0.75乗することにより情報量を圧縮している。

[0006] また、これらの変換符号化技術では、フィルタバンクによって周波数成分に変換された入力信号を、人間の聴覚の周波数分解能に基づいて設定される分割周波数帯域ごとにまとめる。そして、量子化時に各分割周波数帯域毎の正規化係数を聴覚分析結果から決定し、正規化係数と量子化スペクトルの組み合わせで周波数成分を表現することで情報量を削減している。この正規化係数は、実際には分割帯域毎の量子化粗さの調整を行う変数であり、正規化係数が1変化することによって、量子化粗

さは1ステップ分変化することになる。MPEG-2 AACでは、この分割周波数帯域をスケールファクタバンド(SFB)と呼び、正規化係数をスケールファクタと呼称する。

[0007] また、これらの変換符号化方式では、符号化単位である1フレーム全体の量子化粗さを制御することによって符号量を制御している。多くの変換符号化方式では、量子化粗さは、ある基数の整数乗幅でステップ状に制御されており、この整数を量子化ステップと呼ぶ。MPEGオーディオ規格では、この、フレーム全体の量子化粗さを設定する量子化ステップを「グローバルゲイン」もしくは「コモンスケールファクタ」と呼称している。また、前述のスケールファクタは量子化ステップとの相対値で表現することによって、これらの変数の符号に必要な情報量を削減している。

[0008] 例えば、MP3やAACではこれらの変数が1変化することによって、実際の量子化粗さは2の3/16乗分変化する。

[0009] 変換符号化方式の量子化処理では、スケールファクタを制御して聴覚演算の結果を反映して量子化誤差がマスクされるように量子化歪みを制御する。またこれと同時に、量子化ステップを制御してフレーム全体の量子化粗さを適宜調整することによってフレーム全体の符号量制御を行わなければならない。これらの量子化粗さを決める二種類の数値は、符号化品質に重大な影響を及ぼすため、慎重かつ正確に、この二つの制御を同時に効率よく行うことが求められる。

[0010] MPEG-1 Audio Layer III(MP3)の規格書(ISO/IEC 11172-3)やMPEG-2 AACの規格書(ISO/IEC 13818-7)を参照されたい。そこには、量子化時にスケールファクタとグローバルゲインを適宜制御する方法として、歪み制御ループ(アウトーループ)と符号量制御ループ(インナーループ)の二重ループによって繰り返し処理を行う方法が紹介されている。以下、この方法を図面を用いて説明する。なお、便宜上、MPEG-2 AACの場合を例にとって説明を行う。

[0011] 図19は、ISO/IEC規格書に記載されている量子化処理を簡単なフローチャートにしたものである。

[0012] まず、ステップS501では全てのSFBのスケールファクタと、グローバルゲインが0に初期化され、歪み制御ループ(アウトーループ)に入る。

[0013] 歪み制御ループでは、まず、符号量制御ループ(インナーループ)が実行される。

[0014] 符号量制御ループでは、まず、ステップS502において、1フレーム分、すなわち、1024個のスペクトル成分が、下記の量子化式に従って量子化される。

$$[0015] \quad X_q = \text{Int} \left[\left[|x_i| \cdot 2^{-\frac{1}{4} \cdot (\text{global_gain} - \text{scalefac})} \right]^{\frac{3}{4}} + 0.4054 \right] \quad (1)$$

[0016] ただし、(1)式において X_q は量子化スペクトル、 x_i は量子化前のスペクトル(MDCT係数)、 global_gain はグローバルゲイン、 scalefac はこのスペクトル成分が含まれるSFBのスケールファクタである。

[0017] 次に、ステップS503において、これらの量子化スペクトルをハフマン符号化した時の1フレーム分の使用ビット数が計算され、ステップS504でフレームに割り当てられたビット数と比較する。割り当てられたビット数より使用ビット数が大きい場合は、ステップS505においてグローバルゲインを1増加して、量子化粗さを粗くして、再びステップS502のスペクトル量子化に戻る。この繰り返しは量子化後に必要なビット数が割り当てられたビット数より少なくなるまで行われ、この時点でのグローバルゲインが決定されて、符号量制御ループが終了する。

[0018] ステップS506では、符号量制御ループによって量子化されたスペクトルを逆量子化して、量子化前のスペクトルとの差分を取ることで量子化誤差を算出する。この量子化誤差は、SFB毎にまとめられる。

[0019] ステップS507では、全てのSFBでスケールファクタが0より大きくなったか、もしくは、量子化誤差が許容誤差範囲内に納まっているかどうかを調べる。このいずれの条件も満たさないSFBがある場合は、ステップS508に進み、許容誤差範囲内に量子化誤差が納まっていないSFBのスケールファクタを1増やし、再び歪み制御ループ処理を繰り返す。なお、SFB毎の許容誤差は聴覚演算によって量子化処理の前に求められている。

[0020] 以上説明したように、ISO規格書に記載されている量子化処理方法は二重ループで構成されており、しかも、グローバルゲインとスケールファクタは1刻みの制御しか行われぬ。そのため、この処理が収束するまでに、スペクトル量子化とビット計算が幾度となく延々と繰り返されることになる。

- [0021] ここで、例えばMPEG-2 AACの場合では、スペクトル量子化は1回処理するたびに式(1)の計算を1024回行うことになるため、計算量が多い処理である。また、ビット計算時に検索されるハフマン符号表が11種も存在するため、ハフマン符号表を全探索するとビット計算もやはり計算量が必然的に多くなる。
- [0022] さらに、歪み制御ループでは逆量子化後に量子化誤差の計算を行っているが、この処理も計算量が多い。そのため、この二重ループが収束するまでには膨大な処理量がかかってしまう。
- [0023] この問題を解決するために、二重ループの繰り返し回数を削減することによって、処理量を削減しようとする様々な試みがなされている。
- [0024] 例えば、特開2003-271199号公報は、ハフマン符号表の特性に応じて決定したステップ数によって、コモンスケールファクタやスケールファクタを1刻みではなく飛び飛びに制御する技術を開示する。これにより、二重ループそれぞれのループ回数を減らし、処理量を削減している。
- [0025] また、特開2001-184091号公報は、最初に量子化ステップの推定値を算出した後、スケールファクタをMNRに応じて計算後、通常のインナーループを実行する方法を開示している。
- [0026] また、A.D.Duenes、R.Perez、B.Rivas 等、"A robust and efficient implementation of MPEG-2/4 AAC Natural Audio Coders", AES 112th Convention Paper(2002)は、式(1)を変形した式と、聴覚分析によって求められるSFB毎の許容誤差エネルギーを用いることによって、スケールファクタをスペクトル量子化に先行して適宜計算する技術を開示する。これにより、二重ループの外側の歪み制御ループを取り除き、処理量を削減している。
- [0027] これらの従来技術を用いることによって、量子化処理の二重ループの収束を早め、量子化処理の処理量をある程度までは削減することができる。

発明の開示

発明が解決しようとする課題

- [0028] しかしながら、従来の技術では、ISO規格書に記載されている二重ループを完全に繰り返さないようにすることはできない。そのため、スペクトル量子化を数回から数十

回繰り返さないと量子化処理を終えることができず、符号化処理全体に占める量子化処理の処理量は依然として大きかった。

- [0029] 特に、二重ループのうち、聴覚演算結果を利用して事前にスケールファクタを計算することによって、外側の歪み制御ループを無くすことは可能であるが、量子化ステップを量子化前に計算することは従来技術では不可能であった。
- [0030] そのため、従来技術では符号量制御ループにおけるスペクトル量子化とビット計算をやはり繰り返し行っており、処理量を浪費しているという課題がある。
- [0031] ところで、量子化処理とともに処理量のかかる処理として、聴覚心理分析処理がある。そこで、符号化効率よりも処理量削減が優先されるような場合、具体的には例えば、比較的安価な携帯ビデオ撮影機器などにおいて音質よりも消費電力の低減が優先されるような場合では、聴覚心理分析を全く行わずに符号化することも可能である。このとき、量子化処理においては、全ての分割周波数帯域においてスケールファクタを一律に同一値に設定することによって、外側の歪み制御ループを取り除き、さらに処理量を削減することができる。
- [0032] このように聴覚心理分析を行わない構成においても上記の問題は同様に生じる。全ての分割周波数帯域においてスケールファクタを一律に同じ値とした場合であっても、外側の歪み制御ループのみが省略できるだけであり、量子化ステップを量子化前に計算することは従来技術では不可能である。そのため、従来技術では符号量制御ループにおけるスペクトル量子化とビット計算をやはり繰り返し行っており、処理量を浪費しているという課題がある。
- [0033] さらに、聴覚心理分析を行わない構成では、符号量制御の根拠となるPE(聴覚エントロピー)が算出されないため、ビットリザーバに蓄積されている余剰ビットをフレームに割り当てることができず、さらに音質が劣化してしまうという問題も生じる。
- [0034] 以上より、本発明の目的は、オーディオ信号符号化における量子化処理にかかる演算量を削減することにある。
- [0035] また、本発明の別の目的は、聴覚心理分析を行わないように構成されたオーディオ信号符号化において、聴覚心理分析を行わないことによる音質劣化を最小限に留めつつ、量子化にかかる演算量を削減することにある。

課題を解決するための手段

- [0036] 本発明の一側面に係るオーディオ信号符号化装置は、オーディオ入力信号をチャンネルごとに処理単位フレームに分割するフレーム分割部と、オーディオ入力信号を分析し、変換ブロック長の決定および聴覚マスキングの計算を行う聴覚心理演算部と、前記聴覚心理演算部で決定された前記変換ブロック長に従って、処理対象フレームをブロック化し、フレーム中の時間領域信号を1または2以上の周波数スペクトルの組に変換するフィルタバンク部と、前記フィルタバンク部より出力された周波数スペクトルを、複数の周波数帯域に分割して、前記聴覚心理演算部の演算結果によって各周波数帯域のスペクトルを重み付けするスケールファクタ算出部と、前記スケールファクタ算出部によって重み付けされた量子化前のスペクトル全体が持つ聴覚情報量から量子化後のスペクトル全体の情報量を減じ、量子化粗さの刻み幅から得られる係数を積算することによって、フレーム全体の量子化ステップをスペクトル量子化前に決定する量子化ステップ決定部と、前記スケールファクタと前記量子化ステップとを利用して前記周波数スペクトル列を量子化するスペクトル量子化部と、前記スペクトル量子化部から出力される量子化スペクトルを規定のフォーマットに従って整形したビットストリームを作成出力するビット整形部とを備え、前記量子化ステップ決定部は、符号化対象であるフレームに割り当てられたビット量に基づいて前記量子化スペクトル全体の情報量を予測する量子化スペクトル情報量予測部とを含む。
- [0037] 本発明の別の側面に係るオーディオ信号符号化装置は、オーディオ入力信号をチャンネルごとに処理単位フレームに分割するフレーム分割部と、前記フレーム分割部より得られた連続する2フレームの時間領域信号を周波数スペクトルに変換する処理を、1フレームずつずらしながら行うフィルタバンク部と、前記フィルタバンク部より出力された周波数スペクトルの情報量を、量子化前のスペクトル情報量として算出するスペクトル情報量算出部と、ビットレートとサンプリングレートとから算出されるフレーム平均ビット量に基づいて、量子化後のスペクトル情報量を予測する量子化スペクトル情報量予測部と、前記スペクトル情報量算出部で算出された前記量子化前のスペクトル情報量から前記量子化スペクトル情報量予測部で予測された前記量子化後のスペクトル情報量を減じ、その減算結果に、量子化粗さの刻み幅から得られる係数を乗

じること、フレーム全体の量子化ステップをスペクトル量子化前に決定する量子化ステップ決定部と、前記量子化ステップ決定部で決定された前記量子化ステップを利用して前記周波数スペクトルを量子化するスペクトル量子化部と、符号化規格に準じた余剰ビット量を前記規格に合致するように管理するビットリザーバと、前記スペクトル量子化部で量子化された周波数スペクトルを所定のフォーマットに従って整形したビットストリームを生成するビット整形部と、前記フレーム平均ビットに、前記ビットリザーバに蓄積されている余剰ビット量の一部を随時加算してスペクトル割当ビットを計算するスペクトル割当ビット計算部とを備え、前記スペクトル量子化部は、前記スペクトル割当ビット計算部で計算された前記スペクトル割当ビット量に基づいて符号量制御を行う。

[0038] 本発明の他の特徴及び利点は、添付の図面を用いてなされる以下の説明から明らかとなる。

図面の簡単な説明

[0039] [図1]図1は、本発明の第1の実施形態におけるオーディオ信号符号化装置の一構成例を示す図である。

[図2]図2は、本発明の第2の実施形態におけるオーディオ信号符号化処理のフローチャートである。

[図3]図3は、本発明の第2の実施形態における量子化ステップ予測処理のフローチャートである。

[図4]図4は、本発明の第2の実施形態におけるスペクトル量子化処理のフローチャート、

[図5]図5は、本発明の第2の実施形態におけるオーディオ信号符号化装置の一構成例を示す図である。

[図6]図6は、本発明の第2の実施形態におけるオーディオ信号符号化処理プログラムを格納した記憶媒体の内容構成例を示す図である。

[図7]図7は、本発明の第2の実施形態におけるオーディオ信号符号化処理プログラムのPCへの導入を示す模式図である。

[図8]図8は、本発明の第2の実施形態におけるメモリマップの例を示す図である。

[図9]図9は、本発明の第2の実施形態における入力信号バッファの構成例を示す図である。

[図10]図10は、本発明の第3の実施形態におけるオーディオ信号符号化装置の一構成例を示す図である。

[図11]図11は、本発明の第4の実施形態における量子化ステップ予測処理のフローチャートである。

[図12]図12は、本発明の第4の実施形態におけるスペクトル割当ビット算出処理のフローチャートである。

[図13]図13は、本発明の第5の実施形態におけるオーディオ信号符号化装置の一構成例を示す図である。

[図14]図14は、本発明の第6の実施形態におけるオーディオ信号符号化処理のフローチャートである。

[図15]図15は、本発明の第6の実施形態における量子化ステップ予測処理のフローチャートである。

[図16]図16は、本発明の第6の実施形態におけるスペクトル量子化処理のフローチャートである。

[図17]図17は、本発明の第6の実施形態におけるメモリマップの例を示す図である。

[図18]図18は、本発明の第6の実施形態における入力信号バッファの構成例を示す図である。

[図19]図19は、従来のISO規格書に従う量子化処理のフローチャートである。

発明を実施するための最良の形態

- [0040] 本発明は、基本的には量子化前の情報量を量子化後の情報量で割ることによって、全体の量子化粗さを求めることができるという考えに基づき、量子化ステップを実際の量子化前に求めようとするものである。ここで、量子化粗さは一般的に基数を量子化ステップ乗したものであるため、量子化ステップを求めるために底をこの基数にした対数をとると、情報量の除算は情報量の差分に変化する。この差分に、量子化の刻み幅によって決定される係数を積算すると正確な量子化ステップを求めることができる。さらに、実際の量子化後の情報量は量子化後でないと求めることができないが、

フレームに割り当てられた符号量から予測することができるため、本発明はこの予測を利用して量子化前に正確な量子化ステップを求めるものである。

[0041] また、本発明は、量子化前の予測時にはフレーム平均符号量を利用し、実際の量子化時にはビットリザーバに蓄積されている余剰ビット量の一部を足し込み、この値を基準にして符号量を制御する。これにより、量子化ステップの予測値に多少の誤差が生じた場合でも一回のスペクトル量子化で量子化処理を終了し、かつ、情報量が多いフレームには聴覚分析せずとも自動的に余剰ビットの一部が割り当てられるようにする。

[0042] 本発明においては、最初にスケールファクタを算出、確定した後に、その値を使用した計算で量子化ステップをほぼ正確に算出することができるので、ほぼ一回のスペクトル量子化とビット計算で量子化を終了することが可能になる。

[0043] 以下、図面を参照して本発明の好適な実施形態について詳細に説明する。なお、本発明は以下の実施形態に限定されるものではなく、本発明の実施に有利な具体例を示すにすぎない。また、以下の実施形態の中で説明されている特徴の組み合わせの全てが本発明の課題解決手段として必須のものであるとは限らない。

[0044] (第1の実施形態)

図1は、本実施形態におけるオーディオ信号符号化装置の一構成例を示す図である。同図において、太線はデータ信号、細線は制御信号を示す。

[0045] 図示の構成において、1はオーディオ入力信号を処理単位であるフレームに分割するフレーム分割器である。ここでフレーム単位に分割されたオーディオ入力信号は後述する聴覚心理演算器2とフィルタバンク3へ送出される。

[0046] 2は聴覚心理演算器であり、オーディオ入力信号をフレーム単位に分析し、SFBよりも詳細な分割周波数帯域でマスキング計算を行う。この演算の結果、ブロックタイプがフィルタバンク3へ、また、SFB毎の信号対マスク比(SMR)がスケールファクタ計算器4へ出力される。

[0047] 3はフィルタバンクであり、フレーム分割器1から入力された時間信号に対して、聴覚心理演算器2によって指定されたブロックタイプのウィンドウ掛けを行った後、指定されたブロック長で時間周波数変換を行い、周波数スペクトルに変換する。

- [0048] 4はスケールファクタ計算器であり、SFB毎のSMR(信号対マスク比)と、周波数スペクトルより、SFB毎の許容誤差エネルギーを算出し、それに基づいて全てのSFBのスケールファクタを決定する。
- [0049] 5はスペクトル割当ビット計算器であり、量子化スペクトル符号に割り当てられるビット数を計算する。
- [0050] 6は量子化スペクトル総量予測器であり、スペクトル割当ビット数に基づいて、量子化後のスペクトル総量を予測する。
- [0051] 7は量子化ステップ計算器であり、量子化前のスペクトルが持つ聴覚情報量を計算し、量子化後のスペクトル総量から求めた量子化後のスペクトル情報量を減じることにより量子化ステップを求める。
- [0052] 8はスペクトル量子化器であり、各周波数スペクトルを量子化する。
- [0053] 9はビット整形器であり、スケールファクタと量子化スペクトルを適宜規定のフォーマットに整形してビットストリームを作成し、出力する。
- [0054] 上記構成によるオーディオ信号符号化装置におけるオーディオ信号の処理動作を以下に説明する。
- [0055] なお、本実施形態では、説明の便宜のために符号化方式としてMPEG-2 AACを例にとって説明するが、同様な量子化手法を適用可能なその他の符号化方式についても全く同様な方法で実現可能である。
- [0056] まず、処理に先立ち、各部の初期化を行う。初期化によって、量子化ステップと全てのスケールファクタの値は0にセットされる。
- [0057] オーディオPCM信号などのオーディオ入力信号はフレーム分割器1によってフレーム単位に分割され、聴覚心理演算器2とフィルタバンク3に送出される。MPEG-2 AAC LC(Low-Complexity)プロファイルの場合、1フレームは1024サンプルのPCM信号で構成され、この信号が送出される。
- [0058] 聴覚心理演算器2では、フレーム分割器1から送出された入力信号を適宜分析し、聴覚マスキング分析を行い、ブロックタイプをフィルタバンク3に、また、SFB毎の信号対マスク比(SMR)をスケールファクタ計算器4に各々出力する。なお、聴覚心理演算器2で行われる分析やマスキング計算は当分野において公知であるため、これらの

詳細な説明は行わない。

[0059] フィルタバンク3は、聴覚心理演算器2が出力するブロックタイプに従い、フレーム分割器1からの現フレームの入力信号と先行フレームの入力信号の2フレーム分、2048サンプルの時間領域信号を周波数領域信号へ変換する。なお、本実施形態において、先行フレームの入力信号はフィルタバンク3内のバッファに保持されている。ここで、ブロックタイプが長いブロック長を用いる場合は、入力信号の2048サンプルを一つのブロックとして、ブロックタイプに応じた形の窓掛けを実行後、MDCTを行い、1024個の周波数スペクトルを出力する。短いブロック長を用いる場合には、入力信号の2048サンプルのうち、448番目のサンプルを先頭として256サンプルを一つのブロックとして窓掛けを行う。その後、MDCTを行い、128個の周波数成分を出力する変換を、入力信号を128サンプルずつずらしながら都合8回行う。これにより8組の周波数スペクトルが得られる。

[0060] スケールファクタ計算器4は、フィルタバンク3から出力されるスペクトル成分と聴覚心理演算器2から出力されるSFB毎のSMR値からSFB毎の許容誤差エネルギーを算出し、これに基づき、SFB毎のスケールファクタを計算する。許容誤差エネルギーに基づくスケールファクタの算出方法は、当分野では公知であるため、ここでは詳細は述べない。例えば、前述した非特許文献1に記載された手法をとるならば、MPEG-2 AACにおいて、SFB b におけるスケールファクタ $scalefac[b]$ は次式で求めることができる。

$$[0061] \quad scalefac[b] = \text{Int} \left[-\frac{16}{3} \cdot \left[\frac{1}{2} \cdot \log_2 x_{min}[b] + \log_2 \left(\frac{3}{4} \right) - \frac{1}{4} \cdot \log_2 x_{avg} \right] \right] \quad (2)$$

[0062] ただし、式(2)において、 x_{avg} はSFB b に含まれるスペクトル成分の平均レベルである。また、 $x_{min}[b]$ はSFB b の許容誤差エネルギーであり、SFB b のスペクトルエネルギーを $energy[b]$ 、信号対マスク比を $SMR[b]$ 、含まれるスペクトル本数を $sfb_width[b]$ とすると、この $x_{min}[b]$ は次式によって求められる。

$$[0063] \quad x_{min}[b] = \frac{energy[b]}{SMR[b]} / sfb_width[b] \quad (3)$$

[0064] スペクトル割当ビット計算器5は、スケールファクタ計算器4から出力されたスケールファクタをハフマン符号化したときのビット数を計算し、指定されたフレームビット数から減じる。これにより、量子化スペクトルに割り当てられるビット数を計算し、量子化スペクトル総量予測器6へ出力する。

[0065] 量子化スペクトル総量予測器6は、スペクトル割当ビット計算器5から出力されたビット数に基づいて量子化スペクトル総量の予測計算を行う。本実施形態において、この計算は、従来の量子化器によって量子化した際の、スペクトル割当ビット数と量子化スペクトル総量との関係を実際に測定し、その結果に基づいて作成した近似式によって計算する。例えば、この近似式をF(x)として、スペクトル割当ビットをspectrum_bitsとすると、量子化後スペクトル予測総量は以下の式によって求めることができる。

$$[0066] \quad \sum_i X_q \approx F(\text{spectrum_bits}) \quad (4)$$

[0067] 量子化ステップ計算器7では、まず、フィルタバンク3から出力された各周波数スペクトルに、スケールファクタによって聴覚上の重み付けをした値の総計をとり、これに基づいて量子化前の周波数スペクトルが持つ聴覚情報量を計算する。次に、量子化後スペクトル総量予測器6から出力された量子化スペクトル総量に基づいて量子化スペクトルが持つ情報量を計算する。

[0068] 最後に、量子化前スペクトルの聴覚情報量から量子化スペクトルの情報量を減じ、量子化粗さの刻み幅から得られる係数を掛けることによって、フレーム全体の量子化粗さである量子化ステップを計算する。

[0069] 具体的には、MPEG-2 AACの場合は、量子化ステップの予測値は次式を計算することによって得られる。

$$[0070] \quad \text{global_gain} = \text{Int} \left[\frac{16}{3} \cdot \left[\log_2 \sum_i \left[|x_i|^3 \cdot 2^{\frac{3}{16} \text{scalefac}} \right] - \log_2 \sum_i X_q \right] \right] \quad (5)$$

[0071] ただし、式(5)において、Xqは量子化スペクトル、xiは量子化前のスペクトル、global_gainはグローバルゲイン(量子化ステップ)、scalefacはこのスペクトル成分が含まれるSFBのスケールファクタである。また、総計を取るiの範囲は1フレーム分、すなわち0 ≤

$i \leq 1023$ である。

ここで、式(5)において、以下に示す右辺の第1項

$$\log_2 \sum_i \left[|x_i|^{3/4} \cdot 2^{\frac{3}{16} \cdot \text{scalefac}} \right]$$

が、量子化前のスペクトル全体が持つ聴覚情報量であり、各スペクトルに、スケールファクタによって聴覚上の重み付けがなされた値の総計である。また、右辺の第2項である $\log_2 \sum_i X_q$ が、量子化後のスペクトルが持つ情報量であり、このうち、 $\sum_i X_q$ は量子化スペクトルの総計であり、量子化スペクトル総量予測器6によって予測された値である。この値は前述したように例えば近似式(4)を計算することによって得られる。

[0072] なお、式(5)はスペクトル量子化式(1)を適宜変形することによって得ることができる。

[0073] スペクトル量子化器8は、スケールファクタ計算器4が出力したスケールファクタと量子化ステップ計算器7が出力した量子化ステップに従って、1024本の周波数スペクトルを量子化する。具体的には、例えば、MPEG-2 AACの場合では式(1)によって量子化スペクトルを算出し、フレーム全体で消費されるビット数をカウントする。

[0074] ここで、使用ビット数がスペクトル割当ビット数を超過してしまった場合には、使用ビット数がスペクトル割当ビット数に収まるまで量子化ステップを増加して再度スペクトル量子化を行う。しかしながら、量子化ステップ計算器7の計算が正確であるため、多くの場合1回のみ量子化スペクトル計算とビット計算が行われる。

[0075] 各SFBのスケールファクタと量子化スペクトルはビット整形器8によって定められた書式に従ってビットストリームに整形されて、出力される。

[0076] 以上説明したように、本実施形態におけるオーディオ信号符号化装置は、フレームに割り当てられたビット量から、量子化後のスペクトル総量を予測し、これを用いて量子化前後のスペクトル全体が持つ情報量の差分を計算する。これにより、スペクトル量子化の前に量子化ステップがほぼ正確に予測される。このため、量子化ステップの調整のための繰り返しを行うことが減るために、迅速に量子化処理を終了することができる。

[0077] (第2の実施形態)

本発明は、パーソナルコンピュータ(PC)等の汎用的な計算機上で動作するソフトウェアプログラムとして実施することも可能である。以下、この場合について図面を用いて説明する。

[0078] 図5は、本実施形態におけるオーディオ信号符号化装置の構成例を示す図である。

[0079] 図示の構成において、100はCPUであり、オーディオ信号符号化処理のための演算、論理判断等を行い、102のバスを介して各構成要素を制御する。

[0080] 101はメモリであり、本実施形態の構成例における基本I/Oプログラムや、実行しているプログラムコード、プログラム処理時に必要なデータなどを格納する。

[0081] 102はバスであり、CPU100の制御の対象とする構成要素を指示するアドレス信号を転送し、CPU100の制御の対象とする各構成要素のコントロール信号を転送し、各構成機器相互間のデータ転送を行う。

[0082] 103は端末であり、装置の起動、各種条件や入力信号の設定、符号化開始の指示を行う。

[0083] 104はデータやプログラム等を記憶するための外部記憶領域を提供する外部記憶装置であり、例えばハードディスク装置などによって実現される。ここに、OSをはじめとするプログラムやデータ等が保管され、また、保管されたデータやプログラムは必要な時にCPU100によって呼び出される。また、後述するように、オーディオ信号符号化処理プログラムもこの外部記憶装置104にインストールされることになる。

[0084] 105はメディアドライブである。記録媒体(例えば、CD-ROM)に記録されているプログラムやデータ、デジタルオーディオ信号などはこのメディアドライブ105が読み取ることにより本オーディオ信号符号化装置にロードされる。また、外部記憶部104に蓄えられた各種データや実行プログラムを、記録媒体に書き込むこともできる。

[0085] 106はマイクロフォンであり、実際の音を收音してオーディオ信号に変換する。107はスピーカーであり、任意のオーディオ信号データを実際の音にして出力することができる。

[0086] 108は通信網であり、LAN、公衆回線、無線回線、放送電波などで構成されている

。109は通信インタフェースであり、通信網108に接続されている。本実施形態におけるオーディオ信号符号化装置はこの通信インタフェース109を介して通信網108を経由し、外部機器と通信を行い、データやプログラムを送受信することができる。

[0087] かかる構成を備えるオーディオ信号符号化装置は、端末103からの各種の入力に応じて作動する。端末103からの入力が増加されると、インタラプト信号がCPU100に送られることによって、CPU100がメモリ101内に記憶してある各種の制御信号を読み出し、それらの制御信号に従って、各種の制御が行われる。

[0088] 本実施形態のオーディオ信号符号化装置は、CPU100が、メモリ101に格納されている基本I/Oプログラムを実行し、これより外部記憶装置104に記憶されているOSをメモリ101にロードしてこれを実行することによって、動作する。具体的には、本装置の電源がONにされると、基本I/Oプログラム中のIPL(イニシャルプログラムローディング)機能により外部記憶部104からOSがメモリ101に読み込まれ、OSの動作が開始される。

[0089] オーディオ信号符号化処理プログラムは、図2に示されるオーディオ信号符号化処理手順のフローチャートに基づいてプログラムコード化されたものである。

[0090] 図6は、オーディオ信号符号化処理プログラムおよび関連データを記録媒体に記録したときの内容構成例を示す図である。本実施形態において、オーディオ信号符号化処理プログラムおよびその関連データは記録媒体に記録されている。図示したように記録媒体の先頭領域には、この記録媒体のディレクトリ情報が記録されており、その後この記録媒体のコンテンツであるオーディオ信号符号化処理プログラムと、オーディオ信号符号化処理関連データがファイルとして記録されている。

[0091] 図7は、オーディオ信号符号化処理プログラムのオーディオ信号符号化装置(PC)への導入を示す模式図である。記録媒体に記録されたオーディオ信号符号化処理プログラムおよびその関連データは、図7に示したようにメディアドライブ105を通じて本装置にロードすることができる。この記録媒体110をメディアドライブ105にセットすると、OS及び基本I/Oプログラムの制御のもとにオーディオ信号符号化処理プログラムおよびその関連データが記録媒体110から読み出され、外部記憶部104に格納される。その後、再起動時にこれらの情報がメモリ101にロードされて動作可能となる

。

- [0092] 図8は、本実施形態におけるオーディオ信号符号化処理プログラムがメモリ101にロードされ実行可能となった状態のメモリマップを示す図である。図示のように、メモリ101のワークエリアには例えば、基準ビットレート、基準サンプリングレート、ビットレート、サンプリングレート、が格納されている。このワークエリアにはさらに、割当ビット上限値、平均割当ビット、PEビット、使用ビット、スケールファクタビット、スペクトル割当ビット、量子化前スペクトル聴覚情報量、量子化後スペクトル予測情報量が格納されている。さらに、許容誤差エネルギー、スペクトルバッファ、量子化スペクトル、入力信号バッファ、スケールファクタ、量子化ステップ、ブロックタイプ、SMR、PE、リザーブビット量も格納されている。
- [0093] 図9は、本実施形態におけるオーディオ信号符号化装置における入力信号バッファの一構成例を示す図である。図示の構成において、バッファサイズは1024×3サンプルであり、説明の便宜上1024サンプル毎に縦線で区切っている。入力信号は右から1フレーム分の1024サンプルずつ入力されて、左から逐次処理される。なお、図示の構成は1チャンネル分の入力信号バッファを模式的に示したものであり、本実施形態では入力信号のチャンネル分だけ同様なバッファが用意される。
- [0094] 以下、本実施形態においてCPU100で実行されるオーディオ信号符号化処理をフローチャートを用いて説明する。
- [0095] 図2は、本実施形態におけるオーディオ信号符号化処理のフローチャートである。このフローチャートに対応するプログラムはオーディオ信号符号化処理プログラムに含まれ、上記のとおりメモリ101にロードされCPU100によって実行される。
- [0096] まず、ステップS1は、符号化する入力オーディオ信号をユーザが端末103を用いて指定する処理である。本実施形態において、符号化するオーディオ信号は、外部記憶104に格納されているオーディオPCMファイルでも良いし、マイク106で捉えたリアルタイムの音声信号をアナログ・デジタル変換した信号でも良い。この処理を終えると、ステップS2へ進む。
- [0097] ステップS2は、符号化する入力オーディオ信号が終了したかどうかを判定する処理である。入力信号が終了している場合は、ステップS11へ処理が進む。未終了の

場合は、ステップS3へ処理が進む。

- [0098] ステップS3は、図9に示した入力信号バッファにおいて、右から2フレーム分、すなわち2048サンプルの時間信号を1フレーム分左にシフトするとともに、新たに1フレーム分、すなわち1024サンプルを右側に読み込む入力信号シフト処理である。この処理は入力信号に含まれる全てのチャンネルに対して行われる。処理を終えると、ステップS4へ処理が進む。
- [0099] ステップS4は、入力信号バッファに格納されている時間信号を分析し、現行フレームの聴覚心理演算を行う処理である。この演算の結果、現行フレームのブロックタイプと、聴覚エントロピー (PE) と、SFBごとのSMR値が算出され、メモリ101上のワークエリアに格納される。ここで、SMR値は、現行フレームのブロック長が短い場合はショートブロック時の8組が、若しくは、ブロックタイプがそれ以外の場合はロングブロック時の1組が算出される。このような聴覚演算は当分野において公知であるため、詳細な説明は行わない。処理を終えると、ステップS5へ処理が進む。
- [0100] ステップS5では、ステップS4で得られたブロックタイプに従って、現行フレームの時間信号、すなわち、図10の現行フレーム先頭ポインタから右に2048サンプル(2フレーム分)の信号に対して窓掛け後、時間周波数変換を行う。この結果、MPEG-2 AACの場合、変換ブロック長が短いときは、128の周波数成分に分割されたスペクトルの組が8組得られる。それ以外のブロック長が長いブロックタイプの場合は、1024の周波数成分に分割されたスペクトルの組が1組得られる。双方の場合とも、算出された計1024本のスペクトルは、メモリ101上のワークエリアにあるスペクトルバッファに格納される。処理を終えると、ステップS6に処理が進む。
- [0101] ステップS6は、ステップS5で得られた周波数スペクトルとステップS4で得られたSFB毎のSMRから許容誤差エネルギーを計算したのち、これを利用して各SFB毎のスケールファクタを算出する処理である。例えば、MPEG-2 AACの場合は前述の実施形態1の式(2)によってスケールファクタを算出する。この処理において算出したSFB毎の許容誤差エネルギーとスケールファクタはメモリ101上のワークエリア上に格納される。処理を終えると、ステップS7へ進む。
- [0102] ステップS7は、量子化前のスペクトルが持つ聴覚情報量と量子化後のスペクトルが

持つ情報量との差分から量子化ステップを計算する処理である。この処理の詳細は図3を用いて後述する。処理を終えると、ステップS8へ進む。

- [0103] ステップS8は、ステップS6で求めたスケールファクタとステップS7で求めた量子化ステップに従って、1024本の周波数スペクトルを量子化して使用ビットを計算する。この使用ビットがメモリ101上のワークエリアに格納されている割当ビットを超えた場合のみ、量子化ステップの増加と再量子化を行う。この処理の詳細は図4を用いて後述する。処理を終えると、ステップS9へ処理が進む。
- [0104] ステップS9は、ステップS8で算出された量子化スペクトルと、スケールファクタを、符号化方式によって定められたフォーマットに従って整形し、ビットストリームとして出力する処理である。本実施形態において、この処理によって出力されるビットストリームは、外部記憶装置104に格納されても良いし、あるいは、通信インタフェース109を介して通信網108に繋がっている外部機器に出力されても良い。処理を終えると、ステップS10へ処理が進む。
- [0105] ステップS10は、ステップS9で出力されたビットストリームに使用されたビット量と符号化ビットレートから、メモリ101上に格納されている蓄積ビット数の補正を行う処理である。処理を終えると、処理はステップS2へと戻る。
- [0106] ステップS11は、聴覚心理演算や直交変換などで生じる遅延によってまだ出力されていない量子化スペクトルがメモリ上に残っているため、それらをビットストリームに整形して出力する処理である。処理を終えると、オーディオ信号符号化処理を終了する。
- [0107] 図3は、上記したステップS7の量子化ステップ予測処理の詳細を示すフローチャートである。
- [0108] まず、ステップS101は、メモリ101上のワークエリアに保存されているスケールファクタを符号化形式によって定められているフォーマットに従って符号化したときに使用するビット数を算出する処理である。算出されたビット数は、メモリ101上のワークエリアに保存される。処理を終えると、ステップS102へ進む。
- [0109] ステップS102は、フレームに割り当てられたビット数からメモリ101上に格納されたスケールファクタビット数を引いて、スペクトル符号に割り当てられるビット数を算出す

る処理である。算出されたスペクトル割当ビット数はメモリ101上のワークエリアに保存される。処理を終えると、ステップS103へ進む。

[0110] ステップS103は、メモリ101上のスペクトル割当ビット数を用いて、量子化スペクトル総量の予測計算を行う処理である。この予測計算は、予め実験を実施することによって求めた近似式によって行う。例えば、この近似式を $F(x)$ として、スペクトル割当ビットを $spectrum_bits$ とすると、量子化後スペクトル予測総量は以下の式によって求めることができる。

$$[0111] \quad \sum_i X_q \approx F(spectrum_bits) \quad (4)$$

[0112] 算出された量子化スペクトル予測総量はメモリ101上のワークエリアに格納される。処理を終えると、ステップS104へ進む。

[0113] ステップS104は、量子化前のスペクトルが持つ聴覚情報量を算出する処理である。量子化前のスペクトルの聴覚情報量は、各スペクトル成分に、そのスペクトル成分が含まれるSFBのスケールファクタによる量子化粗さの減少分を積算し、1フレーム分の総量を求め、その対数を算出することによって求められる。例えば、MPEG-2 AACの場合、量子化前のスペクトルが持つ聴覚情報量は以下の式を計算することによって求めることができる。

$$[0114] \quad \log_2 \sum_i \left[|x_i|^{\frac{3}{4}} \cdot 2^{\frac{3}{16} \cdot scalefac} \right] \quad (6)$$

[0115] 算出された量子化前スペクトルの聴覚情報量はメモリ101上のワークエリアに保存される。処理を終えると、ステップS105へ進む。

[0116] ステップS105は、ステップS103で求めた量子化スペクトルの予測総量の対数を計算し、量子化スペクトルの予測情報量を算出する処理である。例えば、MPEG-2 AACの場合は以下の式を計算することによって算出することができる。

$$[0117] \quad \log_2 \sum_i X_q \quad (7)$$

[0118] すなわち、ステップS103で得られた量子化スペクトル総量の対数を計算することに

よって、量子化スペクトル予測情報量が得られる。この処理によって算出された量子化後のスペクトル情報量はメモリ101上のワークエリアに保存される。処理を終えると、ステップS106へ進む。

[0119] ステップS106では、ステップS104で求めた量子化前スペクトルの聴覚情報量から、ステップS105で求めた量子化スペクトル予測情報量を減じる。その結果にステップS107で、量子化粗さの刻み幅によって決定される係数をかけ、グローバルゲイン、すなわち量子化ステップの予測値が算出される。MPEG-2 AACの場合は、この予測値は結局実施形態1と同じく式(5)を計算したことになる。

$$[0120] \quad global_gain = Int \left[\frac{16}{3} \cdot \left[\log_2 \sum_i \left[|x_i|^{\frac{3}{4}} \cdot 2^{\frac{3}{16} \cdot scalefac} \right] - \log_2 \sum_i X_q \right] \right] \quad (5)$$

[0121] 算出された量子化ステップ予測値は、メモリ101上のワークエリアに量子化ステップとして格納される。処理を終えると、量子化ステップ予測処理を終了し、リターンする。

[0122] 図4は、上記したステップS8のスペクトル量子化処理を詳細化したフローチャートである。

[0123] ステップS201は、メモリ101上に格納されている量子化ステップとスケールファクタに従って、スペクトルバッファに格納されている1024本のスペクトル成分を量子化する処理である。MPEG-2 AACの場合は、前出の式(1)に従って量子化スペクトルが計算される。処理を終えると、ステップS202へ進む。

[0124] ステップS202は、ステップS201で計算された量子化スペクトル全てを符号化した時に使用されるビット数を計算する処理である。例えば、MPEG-2 AACの場合は、量子化スペクトルは複数個をまとめた上でハフマン符号化されるため、この処理においてハフマンコード表の探索が行われ、符号化ビット数の総計が計算される。計算された使用ビット数はメモリ101上のワークエリアに格納される。処理を終えると、ステップS203へ進む。

[0125] ステップS203は、メモリ101上のスペクトル割当ビットと使用ビットの大きさを比較する処理である。この比較の結果、使用ビットが割り当てられたビットよりも大きい場合は、ステップS204へ進み、符号量を削減するためにメモリ101に格納されている量子

化ステップを増加させる。その後、ステップS201に戻り再度スペクトルの量子化を行うが、前述の量子化ステップ予測処理によってほぼ正確な量子化ステップが予測されるため、ステップS204が実際に実行されることは少ない。

- [0126] ステップS203の比較において使用ビットが割り当てられたビットよりも小さい場合は、スペクトル量子化処理を終了してリターンする。
- [0127] 以上説明したように、本実施形態におけるオーディオ信号符号化処理では、スペクトル符号に割り当てられたビット数から量子化後のスペクトルが持つ情報量を予測し、さらに、量子化前の聴覚情報量との差分がとられる。これによって量子化ステップを実際の量子化を行う前にほぼ正確に予測することによって、量子化ステップの調整を極力避けることが可能になるため、量子化処理にかかる処理量を大幅に削減することができる。
- [0128] (第3の実施形態)
- 固定ビットレートで符号化するとき、ビットリザーバに蓄積されている蓄積ビットを入力信号の特性によって各フレームに適宜配分する場合でも、本発明の技術を適用することが可能である。本実施形態ではこの場合について図面を用いて説明する。
- [0129] 図10は、本実施形態におけるオーディオ信号符号化装置の一構成例を示す図である。第1の実施形態に係る図1と同じく、図中の太線はデータの流れを、細線は制御信号の流れを示す。また、図10においては、図1と同様の機能を持つ構成要素には同じ番号を振っている。
- [0130] 図示の構成において、1はフレーム分割器、2は聴覚心理演算器、3はフィルタバンク、4はスケールファクタ計算器、7は量子化ステップ計算器、8はスペクトル量子化器、9はビット整形器である。
- [0131] 11はPEビット計算器であり、フレームの聴覚エントロピー(PE)に基づいてフレームの予測発生符号量であるPEビットを計算する。
- [0132] 12はスペクトル割当ビット計算器であり、ビットレートやPEビット、蓄積ビット量、スケールファクタなどに基づいてスペクトル符号に割り当てられるビット数を計算する。
- [0133] 13はビットリザーバであり、符号化方式に従って規定されている蓄積ビット量を逐次管理する。

- [0134] 14は量子化スペクトル総量予測器であり、条件により、フレーム割当ビット、もしくはPEビットに基づいて量子化スペクトル総量を予測する。
- [0135] 以上のような構成のオーディオ信号符号化装置における処理動作を以下に説明する。なお、本実施形態においても説明の便宜のために符号化方式としてMPEG-2 AACを例にとって説明するが、非線形量子化を行うその他の符号化方式においても全く同様な方法で実現可能である。
- [0136] まず、処理に先立ち、各部の初期化が行われる。初期化によって、量子化ステップと全てのスケールファクタは0にセットされる。
- [0137] オーディオ入力信号はフレーム分割器1によってフレーム単位に分割され、聴覚心理演算器2とフィルタバンク3に出力される。
- [0138] 聴覚心理演算器2では、フレーム分割器1から出力された入力信号に対して適宜聴覚マスキング分析を行い、ブロックタイプとSFB毎のSMR、PEを出力する。
- [0139] フィルタバンク3はフレーム分割器1から出力された1フレームとフィルタバンク3に保持されている先行1フレームとの2フレーム分の入力信号を聴覚心理分析器2から出力されたブロックタイプに従って時間周波数変換を行い、周波数スペクトルに変換する。
- [0140] スケールファクタ計算器4は、フィルタバンク3から出力される周波数スペクトルと聴覚心理演算器2から出力されるSFB毎のSMR値に基づいて、スケールファクタを実施形態1と同様に適宜算出する。
- [0141] PEビット計算器11は聴覚心理演算器3が出力されるPEから、PEビットを計算する。すなわち、処理中のフレームの入力信号が持つ聴覚的な情報量を聴覚上完全に符号化した場合の予想符号量に変換する。MPEG-2 AACの場合、ISO規格書に記載されているPEビットの計算式は次のようになる。
- [0142] ブロック長がロングのとき：

$$pe_bits = 0.3 \cdot PE + 6.0 \cdot \sqrt{PE} \quad (8)$$

ブロック長がショートのとき：

$$pe_bits = 0.6 \cdot PE + 24.0 \cdot \sqrt{PE} \quad (9)$$

- [0143] 本実施形態では、この計算式をこのまま用いて、ブロックタイプのブロック長に応じてPEビットを算出する。
- [0144] スペクトル割当ビット計算器12では、まず、スケールファクタ計算器4から出力されるスケールファクタを符号化するために必要なビット数を計算する。次に、ビットレートに基づく1フレーム・チャンネル当りの平均ビット量との差分を求めて平均スペクトル割当ビットを計算する。
- [0145] 次に、この値とPEビット計算器11が出力するPEビットを比較し、PEビットが大きい場合はPEビットをビットリザーバ13に蓄積されている蓄積ビット量によって決定される最大値まで割り当てる。PEビットが小さい場合は平均スペクトル割当ビットをそのまま割り当てる。
- [0146] すなわち、本実施形態において、スペクトル割当ビットは具体的には以下の手順で計算される。
- [0147] 1. 蓄積ビット使用許容量を蓄積ビット量から計算する。
ブロック長がロングのとき: 蓄積ビット量の10%、
ブロック長がショートのとき: 蓄積ビット量の25%、
を蓄積ビット使用許容量とする。これをusable_bitsとする。
- [0148] 2. 平均スペクトル割当ビット量をaverage_bitsとすると、スペクトル割当ビット量、spectrum_bitsは以下の要領で決定される。
pe_bits > (average_bits + usable_bits) のとき、
spectrum_bits = average_bits + usable_bits;
pe_bits < average_bits のとき、
spectrum_bits = average_bits;
それ以外、average_bits ≤ pe_bits ≤ (average_bits + usable_bits) のとき、
spectrum_bits = pe_bits;
- [0149] 次に、スペクトル割当ビット計算器12は、PEビットが平均スペクトル割当ビット量よりも少ない場合は、PEビットを量子化スペクトル総量予測器14に出力する。一方、PEビットが平均スペクトル割当ビット以上の場合は上記手順で計算したスペクトル割当ビットを量子化スペクトル総量予測器14に出力する。このとき、どちらのビット数を量子化

スペクトル総量予測器14に出力したかを示すフラグであるビット選択情報(以下、単に「選択情報」という。)を同時に出力する。

[0150] 量子化スペクトル総量予測器14は、入力された選択情報とビット数に基づいて量子化スペクトル総量を予測する。この予測計算は、第1の実施形態で示した方法と同様に、実験によって求めた近似式によって行うが、本実施形態における量子化スペクトル総量予測器14では、この近似式を選択情報によって切り換えて予測計算を行う。例えば、スペクトル割当ビットによる量子化スペクトル総量の近似式をF(x)、PEビットによる量子化スペクトル総量の近似式をG(x)とすると、スペクトル予測総量は以下の式で求められる。

[0151] 選択情報がスペクトル割当ビットの選択を示している場合:

$$\sum_i X_q \approx F(\text{spectrum_bits}) \quad (4)$$

選択情報がPEビットの選択を示している場合:

$$\sum_i X_q \approx \frac{\text{bit_rate}}{\text{base_bit_rate}} \cdot \frac{\text{base_sampling_rate}}{\text{sampling_rate}} \cdot G(\text{pe_bits}) \quad (10)$$

[0152] ただし、式(10)において、bit_rateは処理中の入力信号のビットレート、sampling_rateは処理中の入力信号のサンプリングレートである。また、base_bit_rateは基準ビットレート、base_sampling_rateは基準サンプリングレートである。また、基準ビットレートと基準サンプリングレートは、PEビットによる量子化スペクトル総量予測式G(x)を実験によって求めたときの入力信号のビットレートとサンプリングレートである。これは、本実施形態におけるオーディオ信号符号化装置において予め定められている値である。

[0153] ここで、本実施形態において上記のような量子化スペクトルの予測方法を取る理由を以下説明する。

[0154] 本実施形態では、スペクトル割当ビット計算器12においてPEビットを基準にしたビットの割当が行われる。従って、スペクトル割当ビットには通常PEビットの大きさ、すなわち、処理中のフレームにおける入力信号が持つ聴覚上の発生符号量が反映される。ところが、固定ビットレート制御において、PEビットの大きさが平均スペクトル割当ビットを下回る場合には、スペクトル割当ビットには平均スペクトル割当ビットがそのま

ま割り当てられる。したがって、この場合はスペクトル割当ビットには入力信号の聴覚上の発生符号量が反映されないため、スペクトル割当ビットを用いて量子化スペクトル総量を予測すると予測誤差が大きくなってしまふ。そこで、この場合にはPEビットを用いて量子化スペクトル総量を予測することで、より正確な量子化スペクトル総量を予測することができる。

[0155] また、スペクトル割当ビットはビットレートやサンプリングレートの制約を考慮して計算されるため、ビットレートやサンプリングレートの変化に追随する特性を持っている。一方、PEビットは、サンプリングレートの変化によって元となるPE自体の値は変化するものの、ビットレートやサンプリングレートが変化しても式(8)、(9)自体は変化しない。そこで、PEビットによる予測を行う場合は式(10)に示すように、基準となるビットレートやサンプリングレートからの変化率の影響を考慮した予測を行っている。

[0156] このようにすることで、一つの近似式 $G(x)$ をあらゆるビットレートやサンプリングレートに適用することが可能になる。

[0157] 図10の説明に戻る。量子化ステップ計算器7は第1の実施形態と同様に、フィルタバンク3が出力する周波数スペクトルにスケールファクタ計算器4から出力されたスケールファクタによって重み付けをした値の総量を計算する。量子化ステップ計算器7はさらにその対数を計算して量子化前のスペクトルが持つ聴覚情報量を算出する。次に、量子化スペクトル総量予測器14によって予測された量子化スペクトル総量の対数を計算して量子化後のスペクトル情報量を算出する。さらに、この差分を取り、量子化粗さの刻み幅によって決定される係数を掛けることによって量子化ステップを計算する。具体的には、前出の式(5)の計算を行う。

[0158] 第1の実施形態と同様に、スペクトル量子化器8はスケールファクタ計算器4が出力するスケールファクタと量子化ステップ計算器7が出力する量子化ステップを用いてフィルタバンク3が出力する周波数スペクトルを量子化し、必要ビット数をカウントする。この必要ビット数は、スペクトル割当ビット計算器12から出力されるスペクトル割当ビットと比較される。ここで、必要ビット数がスペクトル割当ビットを超えてしまった場合は、量子化ステップを適宜増加して再度量子化を行う。ただし、前述したように量子化ステップ計算器7による量子化ステップの予測値はほぼ正確であるため、この再量子化

が行われることは少ない。

[0159] スペクトル量子化器8によって最終的に出力される量子化スペクトルとスケールファクタ、量子化ステップは、ビット整形器9によってエントロピー符号化後、符号化方式が定めるビットストリーム形式に適宜整形され、出力される。

[0160] このとき、ビットリザーバ13に実際に符号に使用したビット数が通知され、ビットリザーバ13はフレームビットとの差分を計算し、増減分を蓄積ビット量に加減することで適宜蓄積ビット量を調整する。

[0161] 以上説明したように、本実施形態のように固定ビットレートにおいて、入力信号に応じてビットリザーバに蓄積されている蓄積ビットを適宜フレームに割り当てる場合においても、量子化前に正確に量子化スペクトル総量を予測する。これによって、量子化前に正確に量子化ステップを決定することが可能になり、スペクトル量子化とビット計算の繰り返しを避けて、効率的に量子化を行うことができる。

[0162] (第4の実施形態)

第3の実施形態で説明したオーディオ信号符号化装置もまた、PC等の汎用的なコンピュータ上で動作するソフトウェアプログラムとして実施することが可能である。以下、この場合について図面を用いて説明する。

[0163] 本実施形態におけるオーディオ信号符号化装置の構成およびオーディオ信号符号化処理プログラムの処理内容等については、第2の実施形態と概ね共通している。したがって本実施形態では、第2の実施形態で説明した図5、図2、図6～9を援用することとし、それらの詳細な説明は省略する。第2の実施形態と相違する点は、ステップS7の量子化ステップ予測処理の内容にある。そこで以下では、このステップS7の量子化ステップ予測処理の説明のみを行う。

[0164] 図11は、本実施形態におけるステップS7の量子化ステップ予測処理の詳細を示すフローチャートである。

[0165] まず、ステップS301は、ステップS4の聴覚心理演算処理によって得られたメモリ101上のPEとブロックタイプより、PEビットを算出する処理である。具体的には、第3の実施形態と同様にブロックタイプによって、前出の式(9)もしくは式(10)を選択してPEビットを算出する。算出したPEビットはメモリ101上のワークエリアに格納される。処

理を終えるとステップS302へ進む。

[0166] ステップS302は、メモリ101上のワークエリアに保存されているスケールファクタを符号化方式規定のフォーマットに符号化したときに使用するビット数を算出する処理である。この処理によって算出されたスケールファクタビット数は、メモリ101上のワークエリアに保存される。処理を終えると、ステップS303へ進む。

[0167] ステップS303は、フレームに割り当てられる平均ビット数からメモリ101上に格納されたスケールファクタビット数を引いて、スペクトル符号に割り当てられるビット数、すなわち平均スペクトル割当ビット数(平均割当ビット)を算出する処理である。算出された平均割当ビット数はメモリ101上のワークエリアに保存される。処理を終えると、ステップS304へ進む。

[0168] ステップS304は、メモリ101上の平均割当ビット数とPEビット数の大きさを比較する処理である。この比較の結果、PEビット数の方が大きい場合はステップS305へ進む。それ以外の場合は、ステップS307へ進む。

[0169] ステップS305は、メモリ101上のPEビットと平均割当ビット、蓄積ビット量よりスペクトル割当ビットを算出する処理である。この処理の詳細は図12を用いて後述する。処理を終えると、ステップS306へ進む。

[0170] ステップS306は、メモリ101上のスペクトル割当ビット数を用いて、量子化スペクトル総量の予測計算を行う処理である。この予測計算は、予め実験を実施することによって求めた近似式によって行う。例えば、この近似式を $F(x)$ として、スペクトル割当ビットを $spectrum_bits$ とすると、量子化後スペクトル予測総量は第2の実施形態と同様に式(4)によって求めることができる。

[0171]
$$\sum_i X_q \approx F(spectrum_bits) \quad (4)$$

[0172] 算出された量子化スペクトル予測総量はメモリ101上のワークエリアに格納される。処理を終えると、ステップS309へ進む。

[0173] 一方のステップS307は、メモリ101上の平均割当ビットをスペクトル割当ビットとしてメモリ101に格納する処理である。すなわち、平均割当ビットの値をスペクトル割当ビットにコピーする。処理を終えると、ステップS308へ進む。

[0174] ステップS308は、メモリ101上のPEビット数を用いて、量子化スペクトル総量の予測計算を行う処理である。この予測計算も、予め実験を実施することによって求めた近似式によって行う。この近似式をG(x)として、PEビットをpe_bitsとすると、量子化後スペクトル予測総量は第3の実施形態と同様に式(10)によって求めることができる。

$$[0175] \quad \sum_i X_q \approx \frac{\text{bit_rate}}{\text{base_bit_rate}} \cdot \frac{\text{base_sampling_rate}}{\text{sampling_rate}} \cdot G(\text{pe_bits}) \quad (10)$$

[0176] 算出された量子化スペクトル予測総量はメモリ101上のワークエリアに格納される。処理を終えると、ステップS309へ進む。

[0177] ステップS309は、量子化前のスペクトルが持つ聴覚情報量を算出する処理である。第2の実施形態と同様に、量子化前スペクトルの聴覚情報量は、各スペクトル成分に、そのスペクトル成分が含まれるSFBのスケールファクタによる量子化粗さの減少分を積算し、1フレーム分の総量を求め、その対数を算出することによって求められる。例えば、MPEG-2 AACの場合、量子化前のスペクトルが持つ聴覚情報量は以下の式を計算することによって求めることができる。

$$[0178] \quad \log_2 \sum_i \left[|x_i|^{\frac{3}{4}} \cdot 2^{\frac{3}{16} \cdot \text{scalefac}} \right] \quad (6)$$

[0179] 算出された量子化前スペクトルの聴覚情報量はメモリ101上のワークエリアに保存される。処理を終えると、ステップS310へ進む。

[0180] ステップS310は、ステップS306もしくはステップS308で求めた量子化スペクトルの予測総量の対数を計算し、量子化スペクトルの予測情報量を算出する処理である。例えば、MPEG-2 AACの場合は以下の式を計算することによって算出することができる。

$$[0181] \quad \log_2 \sum_i X_q \quad (7)$$

[0182] この処理によって算出された量子化後のスペクトル予測情報量はメモリ101上のワークエリアに保存される。処理を終えると、ステップS311へ進む。

[0183] ステップS311では、ステップS309で求めた量子化前スペクトルの聴覚情報量から

、ステップS310で求めた量子化スペクトル予測情報量を引く。さらに、その結果に量子化粗さの刻み幅によって決定される係数を掛け、グローバルゲイン、すなわち量子化ステップの予測値が算出される。MPEG-2 AACの場合は、この予測値は結局第1の実施形態と同じく式(5)を計算したことになる。

$$[0184] \quad global_gain = Int \left[\frac{16}{3} \cdot \left[\log_2 \sum_i \left[|x_i|^{\frac{3}{4}} \cdot 2^{\frac{3}{16} \cdot scalefac} \right] - \log_2 \sum_i X_q \right] \right] \quad (5)$$

[0185] 算出された量子化ステップ予測値は、メモリ101上のワークエリアに量子化ステップとして格納される。処理を終えると、量子化ステップ予測処理を終了し、リターンする。

[0186] 図12は、本実施形態におけるステップS305のスペクトル割当ビット算出処理の詳細を示すフローチャートである。

[0187] ステップS401は、このフレームに割り当てることができる蓄積ビット数を、メモリ101上の蓄積ビット量とブロックタイプに従って計算し、この値を平均割当ビットに足すことで、スペクトル割当ビットの上限値を計算する処理である。本実施形態において、蓄積ビット数は第3の実施形態と同じく以下の要領で決定される。

[0188] ブロック長がロングのとき：蓄積ビット量の10%、
ブロック長がショートのとき：蓄積ビット量の25%

[0189] 上記手順で求められた値をメモリ101上の平均割当ビットに加えることで、スペクトル割当ビット上限値が得られる。

[0190] この計算によって得られたスペクトル割当ビット上限値はメモリ101に格納される。処理を終えると、ステップS402へ進む。

[0191] ステップS402は、メモリ101上のPEビットとスペクトル割当ビット上限値を比較する処理である。この比較の結果、PEビット数がスペクトル割当ビット上限値よりも少ない場合はステップS403へ進む。そうでない場合は、ステップS404へ進む。

[0192] ステップS403は、メモリ101上のPEビットをスペクトル割当ビットとして格納する処理である。すなわち、PEビットの値をスペクトル割当ビットにコピーする。処理を終えると、スペクトル割当ビット算出処理を終えて、リターンする。

[0193] ステップS404は、メモリ101上のスペクトル割当ビット上限値をスペクトル割当ビット

として格納する処理である。すなわち、スペクトル割当ビット上限値をスペクトル割当ビットにコピーする。処理を終えると、スペクトル割当ビット算出処理を終えて、リターンする。

[0194] 本処理では、上述したようにPEビットによって割り当てられるビット数に上限値を設定することによって、蓄積ビットが枯渇してビットリザーバが破綻することを防止する効果がある。

[0195] 以上説明したように、本実施形態によれば、固定ビットレートにおいて、入力信号の特性に応じてビットリザーバに蓄積されている蓄積ビットを適宜フレームに割り当てる場合においても、量子化前に正確に量子化スペクトル総量を予測する。これによって、量子化前に正確に量子化ステップを決定することが可能になり、スペクトル量子化とビット計算の繰り返しを避けて、効率的に量子化を行うことができる。

[0196] 以上説明したように、本発明のオーディオ信号符号化処理では、フレームに割り当てられたビット量から、量子化後のスペクトル総量を予測する。これにより、量子化前後のスペクトル全体が持つ情報量の差分が計算され、スペクトル量子化の前にフレーム全体の量子化ステップがほぼ正確に予測される。したがって、ほぼ一回のスペクトル量子化処理を行うだけで量子化処理を終了することが可能になる。これにより、従来の技術と同等の符号化品質を保ちながら、従来の技術に比べて量子化処理にかかる処理量を大幅に削減することができる。

[0197] (第5の実施形態)

以下では、聴覚心理演算器2を省いた構成のオーディオ信号符号化装置の実施形態を説明する。図13は、本実施形態におけるオーディオ信号符号化装置の構成を示す図である。なお、上述の実施形態における構成要素と同様の構成要素には共通の参照符号が付されている。

[0198] 図示の構成において、フレーム分割器1はオーディオ入力信号を処理単位であるフレームに分割する。フレーム単位に分割された入力信号はフィルタバンク3へ送出される。フィルタバンク3は、フレーム分割器1から入力された時間信号に対して、ウィンドウ掛けを行った後、所定のブロック長で時間一周波数変換を行い、周波数スペクトルに変換する。

- [0199] 量子化スペクトル情報量計算器15は、フィルタバンク3から出力された各周波数スペクトルの総計をとり、これに基づいて量子化前の周波数スペクトルが持つ情報量を計算する。量子化ステップ計算器7は、量子化スペクトル情報量計算器15で求めた量子化前のスペクトルが持つ情報量から、後述の量子化スペクトル情報量予測器16で予測した量子化後のスペクトル情報量を減じることによって量子化ステップを求める。スペクトル量子化器8は、各周波数スペクトルを量子化する。ビット整形器9は、スケールファクタと量子化スペクトルを適宜規定のフォーマットに整形してビットストリームを作成し、出力する。ビットリザーバ13は、各符号化規格により規定される余剰ビット(リザーブビット)数を管理する。
- [0200] スペクトル割当ビット計算器12は、ビットリザーバ13から通知される余剰ビット量とフレーム平均ビットから量子化スペクトル符号に割り当てられるビット数を計算する。量子化スペクトル情報量予測器16は、各フレームに割り当てられる平均ビット数に基づいて量子化スペクトル情報量の予測計算を行う。
- [0201] 次に、上記構成によるオーディオ信号符号化装置における、オーディオ信号の符号化動作を説明する。なお、ここでは符号化方式としてMPEG-2 AACを例にとって説明するが、同様な量子化手法を適用可能なその他の符号化方式についても全く同様な方法で実現可能である。
- [0202] まず、処理に先立ち、各部の初期化を行う。初期化によって、量子化ステップと全てのスケールファクタの値は0にセットされる。
- [0203] オーディオPCM信号などのオーディオ入力信号はフレーム分割器1によってフレーム単位に分割され、フィルタバンク2に送出される。MPEG-2 AAC LC(Low-Complexity)プロファイルの場合、1フレームは1024サンプルのPCM信号で構成され、この信号が送出される。
- [0204] フィルタバンク3では、フレーム分割器1から送出される1フレーム分の現入力信号と、前回の変換時に受け取った先行フレームの入力信号とを合わせて2フレーム分、すなわち2048サンプルの時間信号が1024サンプルの周波数成分に変換される。なお、本実施形態において、先行フレームの入力信号はフィルタバンク3内の図示しないバッファに保持されている。フィルタバンク3は、入力信号の2048サンプルを1つのブ

ロックとして、窓掛けを実行後、MDCTを行い、1024個の周波数スペクトルを出力する。

[0205] スペクトル情報量計算器15は、フィルタバンク3から出力された各周波数スペクトルの総計をとり、これに基づいて量子化前の周波数スペクトルが持つ情報量を計算する。MPEG-2 AACの場合、量子化前のスペクトル全体が持つ情報量は次式で計算できる。

$$[0206] \quad \log_2 \sum_i |x_i|^{\frac{3}{4}} \quad (11)$$

[0207] ただし、 x_i は量子化前のスペクトルを示し、総計をとる i の範囲は1フレーム分、すなわち $0 \leq i \leq 1023$ である。これは、各スペクトルの総計に対して底が2の対数をとったものである。

[0208] 量子化スペクトル情報量予測器15は、各フレームに割り当てられる平均ビット数に基づいて量子化スペクトル情報量の予測計算を行う。この計算では、まず、フレーム平均ビットに基づいて量子化スペクトル総量の予測計算が行われる。本実施形態において、この計算は、従来の量子化器によって量子化した際の、フレームビットと量子化スペクトル総量との関係を実際に測定し、その結果に基づいて作成した近似式によって計算する。例えば、この近似式を $F(x)$ として、フレーム平均ビットを $average_bits$ とすると、量子化スペクトル予測総量は次式によって求めることができる。

$$[0209] \quad \sum_i X_q \approx F(average_bits) \quad (12)$$

[0210] ただし、 X_q は量子化スペクトルであり、総計をとる i の範囲は1フレーム分、すなわち $0 \leq i \leq 1023$ である。なお、本実施形態において、フレーム平均ビットはシステム初期化時にビットレートとサンプリングレートと入力チャネル数とから予め計算されている。この計算は当分野において公知であるため、ここでは詳述しない。システム上に保持されているフレーム平均ビットは、初期化時に計算された値が符号化処理中は不変のまま利用される。

[0211] 次に、量子化スペクトル総量を量子化スペクトル情報量に変換する。本実施形態に

において、この計算は(12)式で求めた量子化スペクトル総量に対し、底が2の対数をとることによって行われる。すなわち、量子化スペクトル情報量は次のように表される。

$$[0212] \quad \log_2 \sum_i X_q \quad (13)$$

[0213] 量子化ステップ計算器7は、スペクトル情報量計算器15から出力された量子化前スペクトルの情報量から、量子化スペクトル情報量予測器16から出力された量子化スペクトル情報量を減じる。その後、その減算結果に、量子化粗さの刻み幅から得られる係数を掛けることによって、フレーム全体の量子化粗さである量子化ステップを計算する。

[0214] 具体的には、MPEG-2 AACの場合は、量子化ステップの予測値は次式によって得られる。

$$[0215] \quad \text{global_gain} = \text{Int} \left[\frac{16}{3} \cdot \left[\log_2 \sum_i |x_i|^{\frac{3}{4}} - \log_2 \sum_i X_q \right] \right] \quad (14)$$

[0216] ただし、 X_q は量子化スペクトル、 x_i は量子化前のスペクトル、 global_gain はグローバルゲイン(量子化ステップ)である。また、総計をとる i の範囲は1フレーム分、すなわち $0 \leq i \leq 1023$ である。

[0217] ここで、(14)式における右辺の第1項は次のとおりである。

$$[0218] \quad \log_2 \sum_i |x_i|^{\frac{3}{4}} \quad (15)$$

[0219] これは、量子化前のスペクトル全体が持つ情報量であり、スペクトル情報量計算器15によって(11)式により計算された値である。また、右辺の第2項は次のとおりである。

$$[0220] \quad \log_2 \sum_i X_q \quad (16)$$

[0221] これは、量子化後のスペクトルが持つ情報量であり、量子化スペクトル情報量予測器16によって(13)式により予測された値である。

- [0222] なお、(14)式は先述のスペクトル量子化式(1)を適宜変形し、スケールファクタ $scale_{fac}$ に一律に0を代入することによって得ることができる。
- [0223] スペクトル割当ビット計算器12は、ビットリザーバ13によって管理されている現在の余剰ビット量をビットリザーバ13から通知され、例えばそのうちの2割をフレーム平均ビットに加えてこれを割当ビットとし、スペクトル量子化器8に通知する。
- [0224] スペクトル量子化器8は量子化ステップ計算器7が出力した量子化ステップに従って、1024本の周波数スペクトルを量子化する。例えば、MPEG-2 AACの場合では(1)式によって量子化スペクトルを算出し、フレーム全体で消費されるビット数をカウントする。
- [0225] ここで、使用ビット数がスペクトル割当ビット計算器12から通知された割当ビット数を超過してしまった場合には、使用ビット数がスペクトル割当ビット数に収まるまで量子化ステップを増加して再度スペクトル量子化を行う。しかしながら、量子化ステップ計算器7の計算が正確であり、かつ、量子化ステップの予測計算が行われた時のビット量に加えて、余剰ビット量の一部が割当ビットに加算されている。このため、多くの場合、1回の量子化スペクトル計算とビット計算が行われるだけで量子化が完了する。
- [0226] また、量子化ステップ計算器7で計算された量子化ステップでスペクトル量子化した場合に使用ビット量が足りなくなるようなフレームは、必然的に情報量が元々平均的なフレームよりも多いフレームである。そのため、余剰ビットの一部を割り当てビットに加算し、この値を基準にしてスペクトル量子化処理を行うことによって、このようなフレームには自動的により多くのビットが割り当てられることになる。
- [0227] 各SFBのスケールファクタと量子化スペクトルはビット整形器9によって定められた書式に従ってビットストリームに整形されて、出力される。
- [0228] 最後に、ビット整形器9は実際に使用したビット量をビットリザーバ13に通知する。ビットリザーバ13はビット整形器9から通知された使用ビット量とフレーム平均ビット量から実際に使用された余剰ビット量を計算し、リザーブビットを適宜加減する。
- [0229] 以上説明した本実施形態におけるオーディオ信号符号化装置は、処理負荷の重い聴覚心理分析を一切行わない。しかも、フレームに割り当てられたビット量から量子化後のスペクトル情報量を予測し、これを用いて量子化前後のスペクトル全体が持

つ情報量の差分を計算することによって、スペクトル量子化の前に量子化ステップをほぼ正確に予測する。このため、量子化ステップの調整のための繰り返しを行うことが減るため、迅速に量子化処理を終了することができる。よって、符号化処理にかかる演算量を大幅に削減することができる。

[0230] また、本実施形態におけるオーディオ信号符号化装置は、フレーム平均ビット量に基づいて量子化ステップを予測しておき、余剰ビット量の一部を一律に足してから実際のスペクトル量子化を行う。これにより、多少の予測誤差が生じても量子化処理が1回の処理で済むとともに、元々の情報量が多いフレームに自動的にリザーブビットが割当てられることになるため、聴覚心理分析を行わないことによる音質劣化を最小限に留めることができる。

[0231] (第6の実施形態)

なお、上述の第5の実施形態は、第2の実施形態と同様、パーソナルコンピュータ(PC)等の汎用的な計算機上で動作するソフトウェアプログラムにより実現することも可能である。

[0232] 本実施形態におけるオーディオ信号符号化装置の構成は、第2の実施形態と同様であるため、ここでは図5、図6を援用する。

[0233] 図17は、本実施形態におけるオーディオ信号符号化処理プログラムがメモリ101にロードされ実行可能となった状態のメモリマップを示す図である。図示のように、メモリ101のワークエリアには例えば、量子化前スペクトル聴覚情報量、量子化後スペクトル予測情報量、スペクトル割当ビット、スペクトルバッファ、量子化スペクトル、入力信号バッファが格納される。この他に、使用ビット、量子化ステップ、ビットレート、サンプリングレート、平均割当ビット、リザーブビット量も格納されている。

[0234] 図18は、本実施形態におけるオーディオ信号符号化装置における入力信号バッファの一構成例を示す図である。図示の構成において、バッファサイズは1024×2サンプルであり、説明の便宜上1024サンプル毎に縦線で区切っている。入力信号は1フレーム分の1024サンプルずつ右側から入力されて、左から逐次処理される。太線の矢印は、入力信号の流れを示している。なお、図示の構成は1チャンネル分の入力信号バッファを模式的に示したものであり、本実施形態では入力信号のチャンネル分だけ

同様なバッファが用意される。

[0235] 以下、本実施形態においてCPU100で実行されるオーディオ信号符号化処理をフローチャートを用いて説明する。

[0236] 図14は、本実施形態におけるオーディオ信号符号化処理のフローチャートである。このフローチャートに対応するプログラムはオーディオ信号符号化処理プログラムに含まれ、上記のとおりメモリ101にロードされCPU100によって実行される。

[0237] まず、ステップS1は、符号化する入力オーディオ信号をユーザが端末103を用いて指定する処理である。本実施形態において、符号化するオーディオ信号は、外部記憶104に格納されているオーディオPCMファイルでも良いし、マイク106で捉えたリアルタイムの音声信号をアナログ・デジタル変換した信号でも良い。この処理を終えると、ステップS2へ進む。

[0238] ステップS2は、符号化する入力オーディオ信号が終了したかどうかを判定する処理である。入力信号が終了している場合は、ステップS11へ処理が進む。未終了の場合は、ステップS3へ処理が進む。

[0239] ステップS3は、図18に示した入力信号バッファにおいて、右から2フレーム分、すなわち2048サンプルの時間信号を1フレーム分左にシフトするとともに、新たに1フレーム分、すなわち1024サンプルを右側に読み込む入力信号シフト処理である。この処理は入力信号に含まれる全てのチャンネルに対して行われる。処理を終えると、ステップS5へ処理が進む。

[0240] ステップS5では、現行フレームの時間信号、すなわち、図18の入力信号バッファに格納されている2048サンプル(2フレーム分)の信号に対して窓掛けを行った後、時間一周波数変換を行う。この結果、MPEG-2 AACの場合、1024の周波数成分に分割されたスペクトルの組が1組得られる。なお本実施形態では、ブロックタイプは全て長いブロック長に設定されている。算出された計1024本のスペクトルは、メモリ101上のワークエリアにあるスペクトルバッファに格納される。このステップS5を終えると、処理はステップS7へと進む。

[0241] ステップS7は、量子化前のスペクトルが持つ情報量と量子化後のスペクトルが持つ情報量との差分から量子化ステップを計算する処理である。この処理の詳細は図15

を用いて後述する。このステップS7を終えると、処理はステップS8へと進む。

[0242] ステップS8では、ステップS7で求めた量子化ステップに従って、1024本の周波数スペクトルを量子化して、使用ビットを計算する。さらに、その使用ビットがメモリ101上のワークエリアに格納されている割当ビットを超えた場合のみ、量子化ステップの増加と再量子化を行う。この処理の詳細は図16を用いて後述する。このステップS8を終えると、処理はステップS9へと進む。

[0243] ステップS9は、ステップS8で算出された量子化スペクトルと、スケールファクタとを、符号化方式によって定められたフォーマットに従って整形し、ビットストリームとして出力する処理である。本実施形態において、この処理によって出力されるビットストリームは、外部記憶装置104に格納されても良いし、あるいは、通信インタフェース109を介して通信網108に繋がっている外部機器に出力されても良い。このステップS9を終えると、処理はステップS10へと進む。

[0244] ステップS10は、ステップS9で出力されたビットストリームに使用されたビット量とフレーム平均ビットから、メモリ101上に格納されている余剰ビットの補正を行う処理である。このステップS10を終えると、処理はステップS2へと戻る。

[0245] ステップS11は、直交変換などで生じる遅延によってまだ出力されていない量子化スペクトルがメモリ101上に残っているため、それらをビットストリームに整形して出力する処理である。このステップS11を終えると、オーディオ信号符号化処理は終了する。

[0246] 図15は、上記したステップS7の量子化ステップ予測処理の詳細を示すフローチャートである。

[0247] ステップS100は、量子化前のスペクトルが持つ情報量を算出する処理である。量子化前のスペクトル情報量は、各スペクトル成分の総量を求め、その対数を算出することによって求められる。例えば、MPEG-2 AACの場合、量子化前のスペクトル情報量は次式によって求めることができる。

[0248]
$$\log_2 \sum_i |x_i|^{\frac{3}{4}} \quad (17)$$

[0249] 算出された量子化前スペクトル情報量はメモリ101上のワークエリアに保存される。

このステップS100を終えると、処理はステップS103へ進む。

[0250] ステップS103は、メモリ101上のフレーム平均ビット数を用いて、量子化スペクトル総量の予測計算を行う処理である。この予測計算は、予め実験を実施することによって求めた近似式によって行う。例えば、この近似式をF(x)として、フレーム平均ビットを *average_bits* とすると、量子化後スペクトル予測総量は次式によって求めることができる。

$$[0251] \quad \sum_i X_q \approx F(\textit{average_bits}) \quad (18)$$

[0252] 算出された量子化スペクトル予測総量はメモリ101上のワークエリアに格納される。このステップS103を終えると、処理はステップS105へと進む。

[0253] ステップS105は、ステップS103で求めた量子化スペクトル予測総量の対数を計算し、量子化スペクトル予測情報量を算出する処理である。例えば、MPEG-2 AACの場合は次式によって算出することができる。

$$[0254] \quad \log_2 \sum_i X_q \quad (19)$$

[0255] この処理によって算出された量子化後のスペクトル情報量はメモリ101上のワークエリアに保存される。このステップS105を終えると、処理はステップS108へと進む。

[0256] ステップS108では、ステップS100で求めた量子化前スペクトル情報量から、ステップS105で求めた量子化スペクトル予測情報量を減じる処理を行う。次に、ステップS109で、ステップS108の減算結果に量子化粗さの刻み幅によって決定される係数を乗じ、グローバルゲイン、すなわち量子化ステップの予測値を算出する。MPEG-2 AACの場合は、この予測値は結局第1の実施形態と同じく式(5)を計算したことになる。

$$[0257] \quad \textit{global_gain} = \textit{Int} \left[\frac{16}{3} \cdot \left[\log_2 \sum_i \left[|x_i|^{\frac{3}{4}} \cdot 2^{\frac{3}{16} \cdot \textit{scalefac}} \right] - \log_2 \sum_i X_q \right] \right] \quad (5)$$

[0258] 算出された量子化ステップ予測値は、メモリ101上のワークエリアに量子化ステップ

として格納される。以上でこの量子化ステップ予測処理を終了し、リターンする。

[0259] 図4は、上記したステップS8のスペクトル量子化処理の詳細を示すフローチャートである。

[0260] ステップS200は、メモリ101上に格納されているフレーム平均ビットに、余剰ビット量の一部を加算して、スペクトル割当ビットを計算する処理である。例えば、本実施形態では、余剰ビット量の2割を一律にフレーム平均ビットに加算してスペクトル割当ビットとする。計算されたスペクトル割当ビットはメモリ101上のワークエリアに格納される。このステップS200を終えると、処理はステップS201へ進む。

[0261] ステップS201は、メモリ101上に格納されている量子化ステップに従って、スペクトルバッファに格納されている1024本のスペクトル成分を量子化する処理である。MPEG-2 AACの場合は、前出の(1)式に従って量子化スペクトルが計算される。このステップS201を終えると、処理はステップS202へ進む。

[0262] ステップS202は、ステップS202で計算された量子化スペクトル全てを符号化した時に使用されるビット数を計算する処理である。例えば、MPEG-2 AACの場合は、量子化スペクトルは複数個をまとめた上でハフマン符号化されるため、この処理においてハフマンコード表の探索が行われ、符号化ビット数の総計が計算される。計算された使用ビット数はメモリ101上のワークエリアに格納される。このステップS202を終えると、処理はステップS203へ進む。

[0263] ステップS203は、メモリ101上のスペクトル割当ビットと使用ビットとの大きさを比較する処理である。この比較の結果、使用ビットが割り当てられたビットよりも大きい場合は、ステップS204へ進み、符号量を削減するためにメモリ101に格納されている量子化ステップを増加した後、ステップS201に戻り再度スペクトルの量子化を行う。ただし、図15に示した前述の量子化ステップ予測処理(ステップS7)によってほぼ正確な量子化ステップが予測されており、かつ、フレーム平均ビットに基づいて量子化ステップの予測が行われている。これに対し、ステップS203では、それに余剰ビットの一部を加えたスペクトル割当ビットを基準にして符号量の制御を行っているため、ステップS204が実際に実行されることは極めて少ないであろう。

[0264] また、予測した量子化ステップで量子化した結果、使用したビットがフレーム平均ビ

ットを超えてしまう場合も、余剰ビットの追加分を超えなければ1回のスペクトル量子化で量子化が終了することになる。かつ、このようなフレームは元々情報量が多いフレームであり、結果的に情報量が多いフレームに自動的により多くのビットが割当てられることになる。

- [0265] ステップS203の比較において使用ビットが割り当てられたビットよりも小さい場合は、このスペクトル量子化処理を終了してリターンする。
- [0266] 以上説明した本実施形態におけるオーディオ信号符号化処理は、聴覚心理分析処理を一切省いたものである。そして、フレーム平均ビットから量子化後のスペクトルが持つ情報量を予測し、さらに、量子化前のスペクトル情報量との差分をとることによって量子化ステップを実際の量子化を行う前にほぼ正確に予測する。これによって、聴覚心理演算を行わなくても、量子化ステップの調整を極力避けることが可能になり、符号化処理全体にかかる処理量を大幅に削減することができる。
- [0267] また、本実施形態におけるオーディオ信号符号化装置は、フレーム平均ビット量に基づいて量子化ステップを予測しておき、リザーブビット量の一部を一律に足してから実際のスペクトル量子化を行う。これにより、多少の予測誤差が生じても量子化処理が1回の処理で済むとともに、元々の情報量が多いフレームに自動的にリザーブビットが割り当てられることになるため、聴覚心理分析を行わないことによる音質劣化を最小限に留めることができる。
- [0268] (他の実施形態)
本発明はその要旨を逸脱しない範囲で種々変形して実施することができる。
- [0269] たとえば、上述の実施形態ではブロックスイッチングを全く行っていないが、聴覚分析を行わず、比較的簡易に入力信号の過渡状態を検知して、ブロックスイッチングを行うように構成された装置にも、本発明を同様に適用することが可能である。
- [0270] また、本発明は、複数の機器から構成されるシステムに適用してもよいし、また、一つの機器からなる装置に適用してもよい。
- [0271] なお、本発明は、前述した実施形態の各機能を実現するプログラムを、システムまたは装置に直接または遠隔から供給し、そのシステムまたは装置に含まれるコンピュータがその供給されたプログラムコードを読み出して実行することによっても達成され

る。

- [0272] 従って、本発明の機能・処理をコンピュータで実現するために、そのコンピュータにインストールされるプログラムコード自体も本発明を実現するものである。つまり、上記機能・処理を実現するためのコンピュータプログラム自体も本発明の一つである。
- [0273] その場合、プログラムの機能を有していれば、オブジェクトコード、インタプリタにより実行されるプログラム、OSに供給するスクリプトデータ等、プログラムの形態を問わない。
- [0274] プログラムを供給するための記録媒体としては、例えば、フレキシブルディスク、ハードディスク、光ディスク、光磁気ディスク、MO、CD-ROM、CD-R、CD-RWなどがある。また、記録媒体としては、磁気テープ、不揮発性のメモ리카ード、ROM、DVD(DVD-ROM, DVD-R)などもある。
- [0275] また、プログラムは、クライアントコンピュータのブラウザを用いてインターネットのホームページからダウンロードしてもよい。すなわち、ホームページから本発明のコンピュータプログラムそのもの、もしくは圧縮され自動インストール機能を含むファイルをハードディスク等の記録媒体にダウンロードしてもよい。また、本発明のプログラムを構成するプログラムコードを複数のファイルに分割し、それぞれのファイルを異なるホームページからダウンロードする形態も考えられる。つまり、本発明の機能・処理をコンピュータで実現するためのプログラムファイルを複数のユーザに対してダウンロードさせるWWWサーバも、本発明の構成要件となる場合がある。
- [0276] また、本発明のプログラムを暗号化してCD-ROM等の記憶媒体に格納してユーザに配布してもよい。この場合、所定条件をクリアしたユーザにのみ、インターネットを介してホームページから暗号化を解く鍵情報をダウンロードさせ、その鍵情報で暗号化されたプログラムを復号して実行し、プログラムをコンピュータにインストールしてもよい。
- [0277] また、コンピュータが、読み出したプログラムを実行することによって、前述した実施形態の機能が実現されてもよい。なお、そのプログラムの指示に基づき、コンピュータ上で稼動しているOSなどが、実際の処理の一部または全部を行ってもよい。もちろん、この場合も、前述した実施形態の機能が実現され得る。

[0278] さらに、記録媒体から読み出されたプログラムが、コンピュータに挿入された機能拡張ボードやコンピュータに接続された機能拡張ユニットに備わるメモリに書き込まれてもよい。そのプログラムの指示に基づき、その機能拡張ボードや機能拡張ユニットに備わるCPUなどが実際の処理の一部または全部を行ってもよい。このようにして、前述した実施形態の機能が実現されることもある。

優先権の主張

[0279] 本願は、2004年11月18日提出の日本国特許出願特願2004-335005、および、2005年11月14日提出の日本国特許出願特願2005-328945を基礎として優先権を主張するものであり、その記載内容の全てを、ここに援用する。

請求の範囲

- [1] オーディオ入力信号をチャンネルごとに処理単位フレームに分割するフレーム分割部と、
- オーディオ入力信号を分析し、変換ブロック長の決定および聴覚マスキングの計算を行う聴覚心理演算部と、
- 前記聴覚心理演算部で決定された前記変換ブロック長に従って、処理対象フレームをブロック化し、フレーム中の時間領域信号を1または2以上の周波数スペクトルの組に変換するフィルタバンク部と、
- 前記フィルタバンク部より出力された周波数スペクトルを、複数の周波数帯域に分割して、前記聴覚心理演算部の演算結果によって各周波数帯域のスペクトルを重み付けするスケールファクタ算出部と、
- 前記スケールファクタ算出部によって重み付けされた量子化前のスペクトル全体が持つ聴覚情報量から量子化後のスペクトル全体の情報量を減じ、量子化粗さの刻み幅から得られる係数を積算することによって、フレーム全体の量子化ステップをスペクトル量子化前に決定する量子化ステップ決定部と、
- 前記スケールファクタと前記量子化ステップとを利用して前記周波数スペクトル列を量子化するスペクトル量子化部と、
- 前記スペクトル量子化部から出力される量子化スペクトルを規定のフォーマットに従って整形したビットストリームを作成出力するビット整形部と、
- を備え、
- 前記量子化ステップ決定部は、符号化対象であるフレームに割り当てられたビット量に基づいて前記量子化スペクトル全体の情報量を予測する量子化スペクトル情報量予測部を含むことを特徴とするオーディオ信号符号化装置。
- [2] オーディオ入力信号をチャンネルごとに処理単位フレームに分割するフレーム分割部と、
- オーディオ入力信号を分析し、変換ブロック長の決定および聴覚マスキングの計算を行う聴覚心理演算部と、
- 前記聴覚心理演算部で決定された変換ブロック長に従って、処理対象フレームを

ブロック化し、フレーム中の時間領域信号を1または2以上の周波数スペクトルの組に変換するフィルタバンク部と、

前記フィルタバンク部より出力された周波数スペクトルを、複数の周波数帯域に分割して、前記聴覚心理演算部の演算結果によって各周波数帯域のスペクトルを重み付けするスケールファクタ算出部と、

符号化対象であるフレームに割り当てられたビット量に基づいて前記量子化スペクトル全体の情報量を量子化する前に予測する量子化スペクトル情報量予測部と、

前記スケールファクタ算出部によって重み付けされた量子化前のスペクトル全体が持つ聴覚情報量から前記量子化後のスペクトル全体の情報量を減じ、量子化粗さの刻み幅から得られる係数を積算することによって、フレーム全体の量子化ステップをスペクトル量子化前に決定する量子化ステップ決定部と、

前記スケールファクタと前記量子化ステップとを利用して前記周波数スペクトル列を量子化するスペクトル量子化部と、

前記スペクトル量子化部から出力される量子化スペクトルを規定のフォーマットに従って整形したビットストリームを作成出力するビット整形部と、

を備え、

前記量子化スペクトル情報量予測部は、固定ビットレート符号化時に、入力信号の予測符号量が平均フレーム割当ビットに満たない場合は、聴覚エントロピーに基づいて量子化スペクトル情報量を予測することを特徴とするオーディオ信号符号化装置。

[3] オーディオ入力信号をチャンネルごとに処理単位フレームに分割するフレーム分割部と、

オーディオ入力信号を分析し、変換ブロック長の決定および聴覚マスキングの計算を行う聴覚心理演算部と、

前記聴覚心理演算部で決定された変換ブロック長に従って、処理対象フレームをブロック化し、フレーム中の時間領域信号を1または2以上の周波数スペクトルの組に変換するフィルタバンク部と、

前記フィルタバンク部より出力された周波数スペクトルを、複数の周波数帯域に分割して、前記聴覚心理演算部の演算結果によって各周波数帯域のスペクトルを重み

付けするスケールファクタ算出部と、

符号化対象であるフレームに割り当てられたビット量に基づいて前記量子化スペクトル全体の情報量を量子化する前に予測する量子化スペクトル情報量予測部と、

前記スケールファクタ算出部によって重み付けされた量子化前のスペクトル全体が持つ聴覚情報量から前記量子化後のスペクトル全体の情報量を減じ、量子化粗さの刻み幅から得られる係数を積算することによって、フレーム全体の量子化ステップをスペクトル量子化前に決定する量子化ステップ決定部と、

前記スケールファクタと前記量子化ステップを利用して前記周波数スペクトル列を量子化するスペクトル量子化部と、

前記スペクトル量子化部から出力される量子化スペクトルを規定のフォーマットに従って整形したビットストリームを作成出力するビット整形部と、

を備え、

前記スペクトル量子化部は、量子化スペクトルに使用される符号量が割り当てられた符号量を超えた場合に、前記量子化ステップを調整してスペクトルの再量子化を行うことを特徴とするオーディオ信号符号化装置。

- [4] 符号化形式がMPEG-1 Audio Layer IIIであることを特徴とする請求項1から3までのいずれかに記載のオーディオ信号符号化装置。
- [5] 符号化形式がMPEG-2/4 AACであることを特徴とする請求項1から3までのいずれかに記載のオーディオ信号符号化装置。
- [6] オーディオ入力信号をチャンネルごとに処理単位フレームに分割するフレーム分割ステップと、
オーディオ入力信号を分析し、変換ブロック長の決定および聴覚マスキングの計算を行う聴覚心理演算ステップと、
前記聴覚心理演算ステップで決定された前記変換ブロック長に従って、処理対象フレームをブロック化し、フレーム中の時間領域信号を1または2以上の周波数スペクトルの組に変換するフィルタバンク処理ステップと、
前記フィルタバンク処理ステップで得られた周波数スペクトルを、複数の周波数帯域に分割して、前記聴覚心理演算ステップでの演算結果によって各周波数帯域のス

ペクトルを重み付けするスケールファクタ算出ステップと、

前記スケールファクタ算出ステップによって重み付けされた量子化前のスペクトル全体が持つ情報量から量子化後のスペクトル全体の情報量を減じ、量子化粗さの刻み幅から得られる係数を積算することによって、フレーム全体の量子化ステップをスペクトル量子化前に決定する量子化ステップ決定ステップと、

前記スケールファクタと前記量子化ステップとを利用して前記周波数スペクトル列を量子化するスペクトル量子化ステップと、

前記スペクトル量子化ステップで得られた量子化スペクトルを規定のフォーマットに従って整形したビットストリームを作成出力するビット整形ステップと、

を有し、

前記量子化ステップ決定ステップは、符号化対象であるフレームに割り当てられた情報量に基づいて前記量子化スペクトル全体の情報量を予測する量子化スペクトル総量予測ステップを含むことを特徴とするオーディオ信号符号化方法。

[7] オーディオ入力信号をチャンネルごとに処理単位フレームに分割するフレーム分割ステップと、

オーディオ入力信号を分析し、変換ブロック長の決定および聴覚マスキングの計算を行う聴覚心理演算ステップと、

前記聴覚心理演算ステップで決定された前記変換ブロック長に従って、処理対象フレームをブロック化し、フレーム中の時間領域信号を1または2以上の周波数スペクトルの組に変換するフィルタバンク処理ステップと、

前記フィルタバンク処理ステップで得られた周波数スペクトルを、複数の周波数帯域に分割して、前記聴覚心理演算ステップでの演算結果によって各周波数帯域のスペクトルを重み付けするスケールファクタ算出ステップと、

符号化対象であるフレームに割り当てられたビット量に基づいて前記量子化スペクトル全体の情報量を量子化する前に予測する量子化スペクトル情報量予測ステップと、

前記スケールファクタ算出ステップによって重み付けされた量子化前のスペクトル全体が持つ聴覚情報量から前記量子化後のスペクトル全体の情報量を減じ、量子

化粗さの刻み幅から得られる係数を積算することによって、フレーム全体の量子化ステップをスペクトル量子化前に決定する量子化ステップ決定ステップと、

前記スケールファクタと前記量子化ステップとを利用して前記周波数スペクトル列を量子化するスペクトル量子化ステップと、

前記スペクトル量子化で得られた量子化スペクトルを規定のフォーマットに従って整形したビットストリームを作成出力するビット整形ステップと、

を有し、

前記量子化スペクトル情報量予測ステップでは、固定ビットレート符号化時に、入力信号の予測符号量が平均フレーム割当ビットに満たない場合は、聴覚エントロピーに基づいて量子化スペクトル情報量を予測することを特徴とするオーディオ信号符号化方法。

[8] オーディオ入力信号をチャンネルごとに処理単位フレームに分割するフレーム分割ステップと、

オーディオ入力信号を分析し、変換ブロック長の決定および聴覚マスキング計算を行う聴覚心理演算ステップと、

前記聴覚心理演算ステップで決定された前記変換ブロック長に従って、処理対象フレームをブロック化し、フレーム中の時間領域信号を1または2以上の周波数スペクトルの組に変換するフィルタバンク処理ステップと、

前記フィルタバンク処理ステップで得られた周波数スペクトルを、複数の周波数帯域に分割して、前記聴覚心理演算ステップでの演算結果によって各周波数帯域のスペクトルを重み付けするスケールファクタ算出ステップと、

符号化対象であるフレームに割り当てられたビット量に基づいて前記量子化スペクトル全体の情報量を量子化する前に予測する量子化スペクトル情報量予測ステップと、

前記スケールファクタ算出ステップによって重み付けされた量子化前のスペクトル全体が持つ聴覚情報量から前記量子化後のスペクトル全体の情報量を減じ、量子化粗さの刻み幅から得られる係数を積算することによって、フレーム全体の量子化ステップをスペクトル量子化前に決定する量子化ステップ決定ステップと、

前記スケールファクタと前記量子化ステップとを利用して前記周波数スペクトル列を量子化するスペクトル量子化ステップと、

前記スペクトル量子化ステップで得られた量子化スペクトルを規定のフォーマットに従って整形したビットストリームを作成出力するビット整形ステップと、

を有し、

前記スペクトル量子化ステップでは、量子化スペクトルに使用される符号量が割り当てられた符号量を超えた場合に、前記量子化ステップを調整してスペクトルの再量子化を行うことを特徴とするオーディオ信号符号化方法。

[9] 請求項6から8までのいずれかに記載のオーディオ信号符号化方法をコンピュータに実行させるためのプログラム。

[10] 請求項9に記載のプログラムを格納したコンピュータ読み取り可能な記憶媒体。

[11] オーディオ入力信号をチャンネルごとに処理単位フレームに分割するフレーム分割部と、

前記フレーム分割部より得られた連続する2フレームの時間領域信号を周波数スペクトルに変換する処理を、1フレームずつずらしながら行うフィルタバンク部と、

前記フィルタバンク部より出力された周波数スペクトルの情報量を、量子化前のスペクトル情報量として算出するスペクトル情報量算出部と、

ビットレートとサンプリングレートとから算出されるフレーム平均ビット量に基づいて、量子化後のスペクトル情報量を予測する量子化スペクトル情報量予測部と、

前記スペクトル情報量算出部で算出された前記量子化前のスペクトル情報量から前記量子化スペクトル情報量予測部で予測された前記量子化後のスペクトル情報量を減じ、その減算結果に、量子化粗さの刻み幅から得られる係数を乗じることで、フレーム全体の量子化ステップをスペクトル量子化前に決定する量子化ステップ決定部と、

前記量子化ステップ決定部で決定された前記量子化ステップを利用して前記周波数スペクトルを量子化するスペクトル量子化部と、

符号化規格に準じた余剰ビット量を前記規格に合致するように管理するビットリザーバと、

前記スペクトル量子化部で量子化された周波数スペクトルを所定のフォーマットに従って整形したビットストリームを生成するビット整形部と、

前記フレーム平均ビットに、前記ビットリザーバに蓄積されている余剰ビット量の一部を随時加算してスペクトル割当ビットを計算するスペクトル割当ビット計算部と、
を備え、

前記スペクトル量子化部は、前記スペクトル割当ビット計算部で計算された前記スペクトル割当ビット量に基づいて符号量制御を行うことを特徴とするオーディオ信号符号化装置。

[12] 符号化形式がMPEG-1 Audio Layer IIIであることを特徴とする請求項11に記載のオーディオ信号符号化装置。

[13] 符号化形式がMPEG-2 AACであることを特徴とする請求項11に記載のオーディオ信号符号化装置。

[14] オーディオ入力信号をチャンネルごとに処理単位フレームに分割するフレーム分割ステップと、

前記フレーム分割ステップで得られた連続する2フレームの時間領域信号を周波数スペクトルに変換する処理を、1フレームずつずらしながら行う時間一周波数変換ステップと、

前記時間一周波数変換ステップで得られた周波数スペクトルの情報量を、量子化前のスペクトル情報量として算出するスペクトル情報量算出ステップと、

ビットレートとサンプリングレートとから算出されるフレーム平均ビット量に基づいて、量子化後のスペクトル情報量を予測する量子化スペクトル情報量予測ステップと、

前記スペクトル情報量算出ステップで算出された前記量子化前のスペクトル情報量から前記量子化スペクトル情報量予測ステップで予測された前記量子化後のスペクトル情報量を減じ、その減算結果に、量子化粗さの刻み幅から得られる係数を乗じることで、フレーム全体の量子化ステップをスペクトル量子化前に決定する量子化ステップ決定ステップと、

前記量子化ステップ決定ステップで決定された前記量子化ステップを利用して前記周波数スペクトルを量子化するスペクトル量子化ステップと、

前記スペクトル量子化ステップで量子化された周波数スペクトルを所定のフォーマットに従って整形したビットストリームを生成するビット整形ステップと、

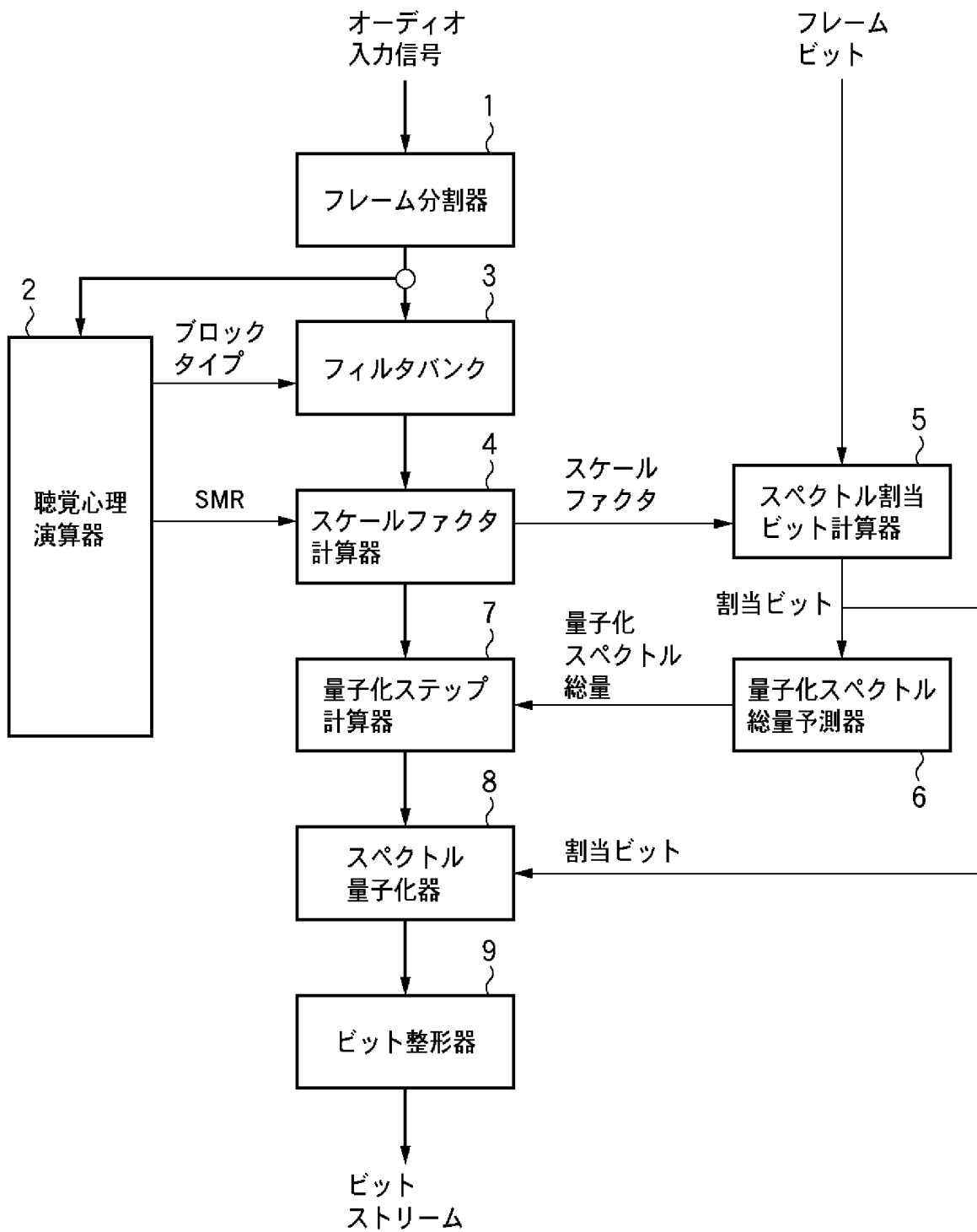
前記フレーム平均ビットに、符号化規格に準じた余剰ビット量を前記規格に合致するように管理するビットリザーバに蓄積されている余剰ビット量の一部を随時加算してスペクトル割当ビットを計算するスペクトル割当ビット計算ステップと、

を備え、

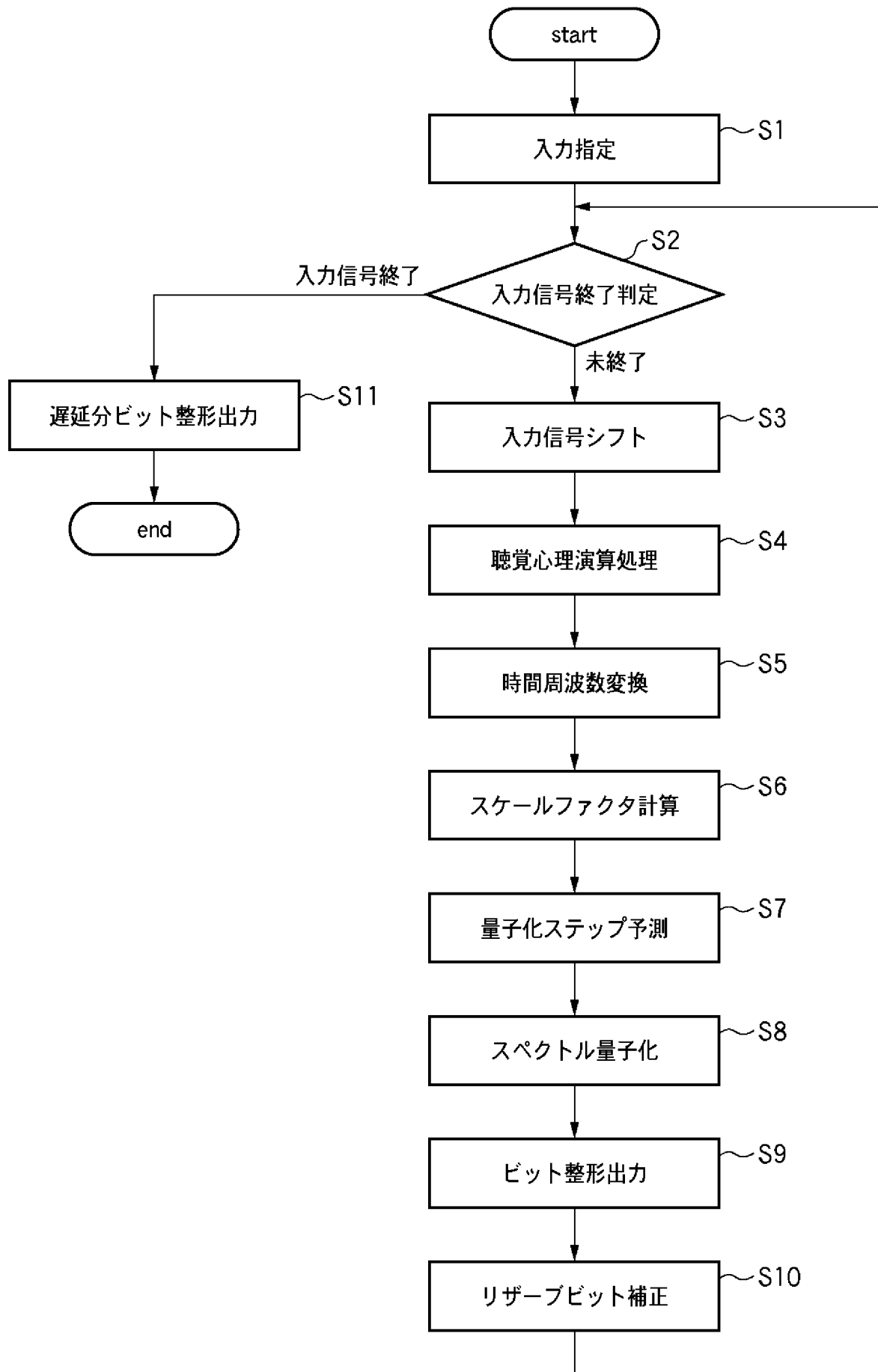
前記スペクトル量子化ステップは、前記スペクトル割当ビット計算ステップで計算された前記スペクトル割当ビット量に基づいて符号量制御を行うことを特徴とするオーディオ信号符号化方法。

- [15] 請求項14に記載のオーディオ信号符号化方法をコンピュータに実行させるためのプログラム。
- [16] 請求項15に記載のプログラムを格納したコンピュータ読み取り可能な記憶媒体。

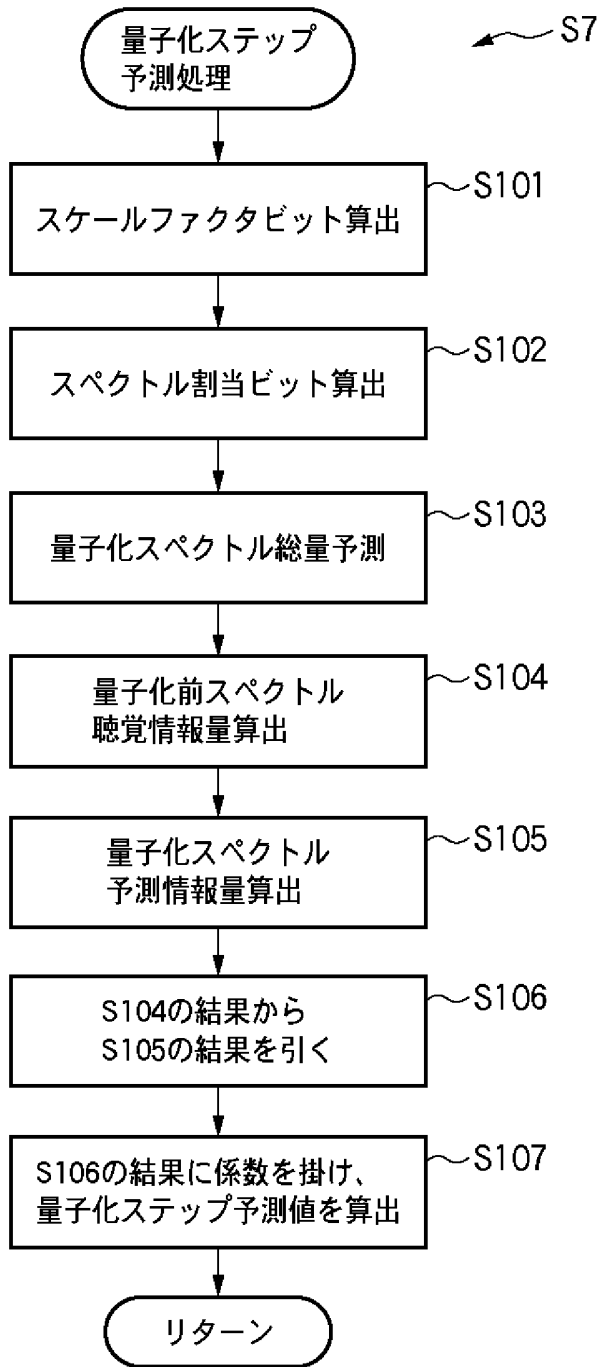
[図1]



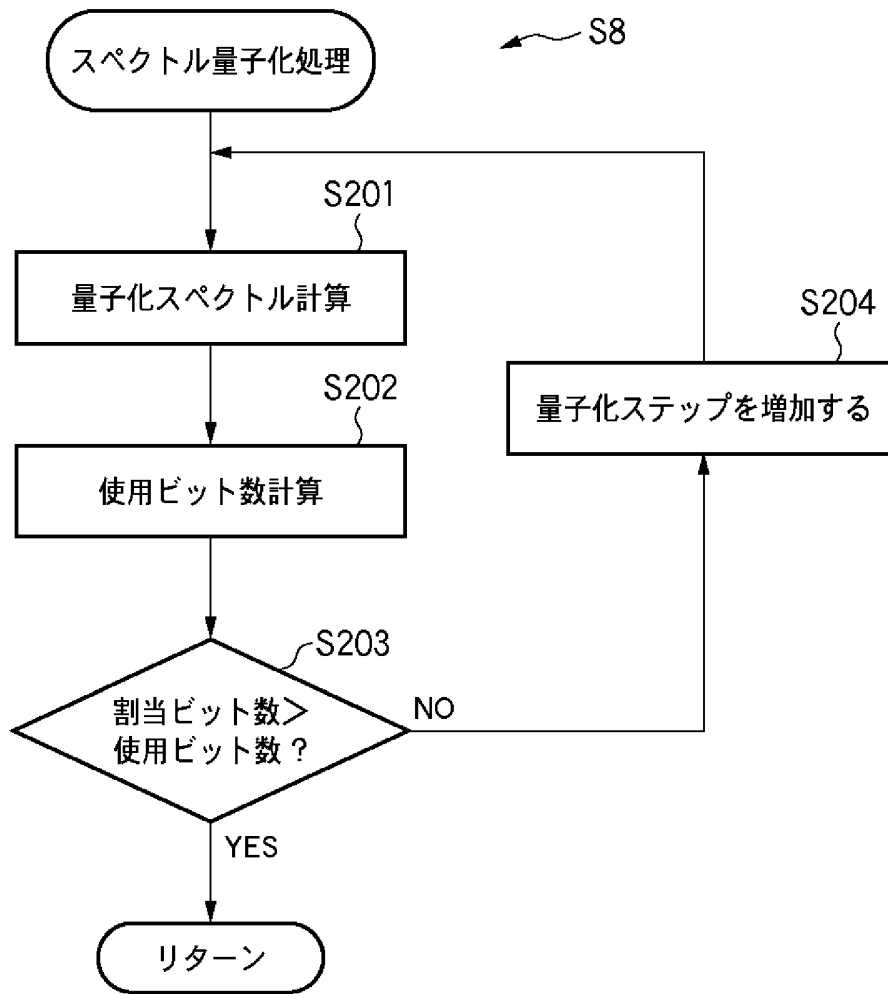
[図2]



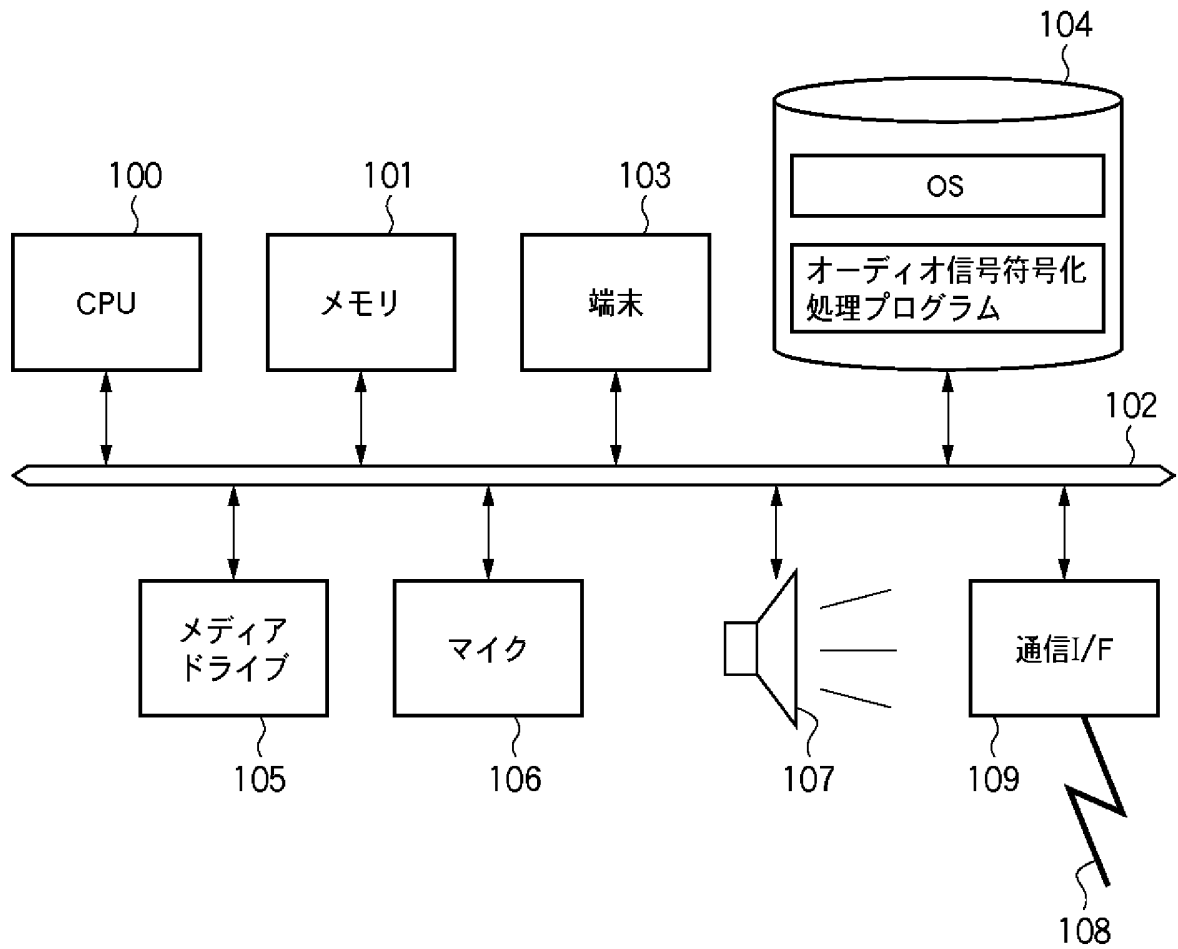
[図3]



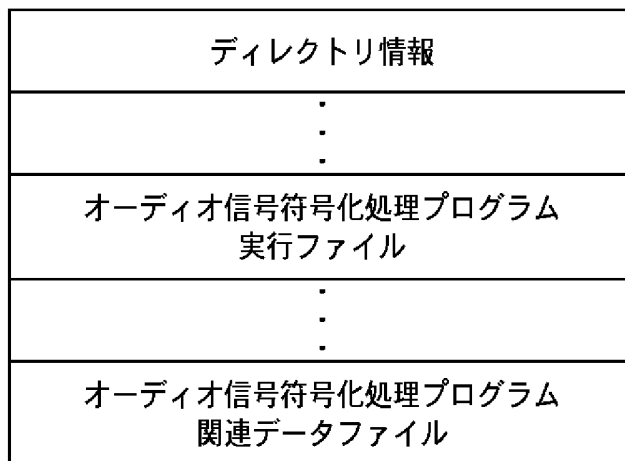
[図4]



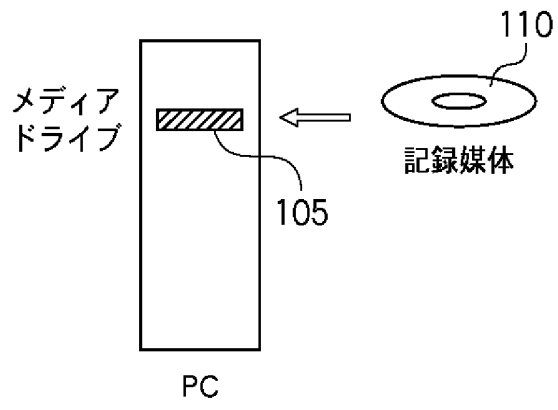
[図5]



[図6]



[図7]



[図8]

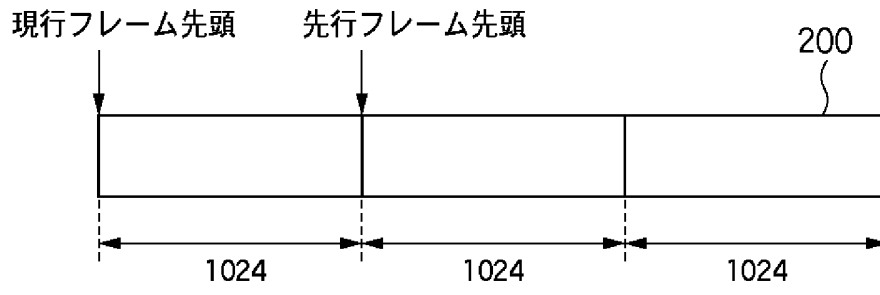
上位アドレス



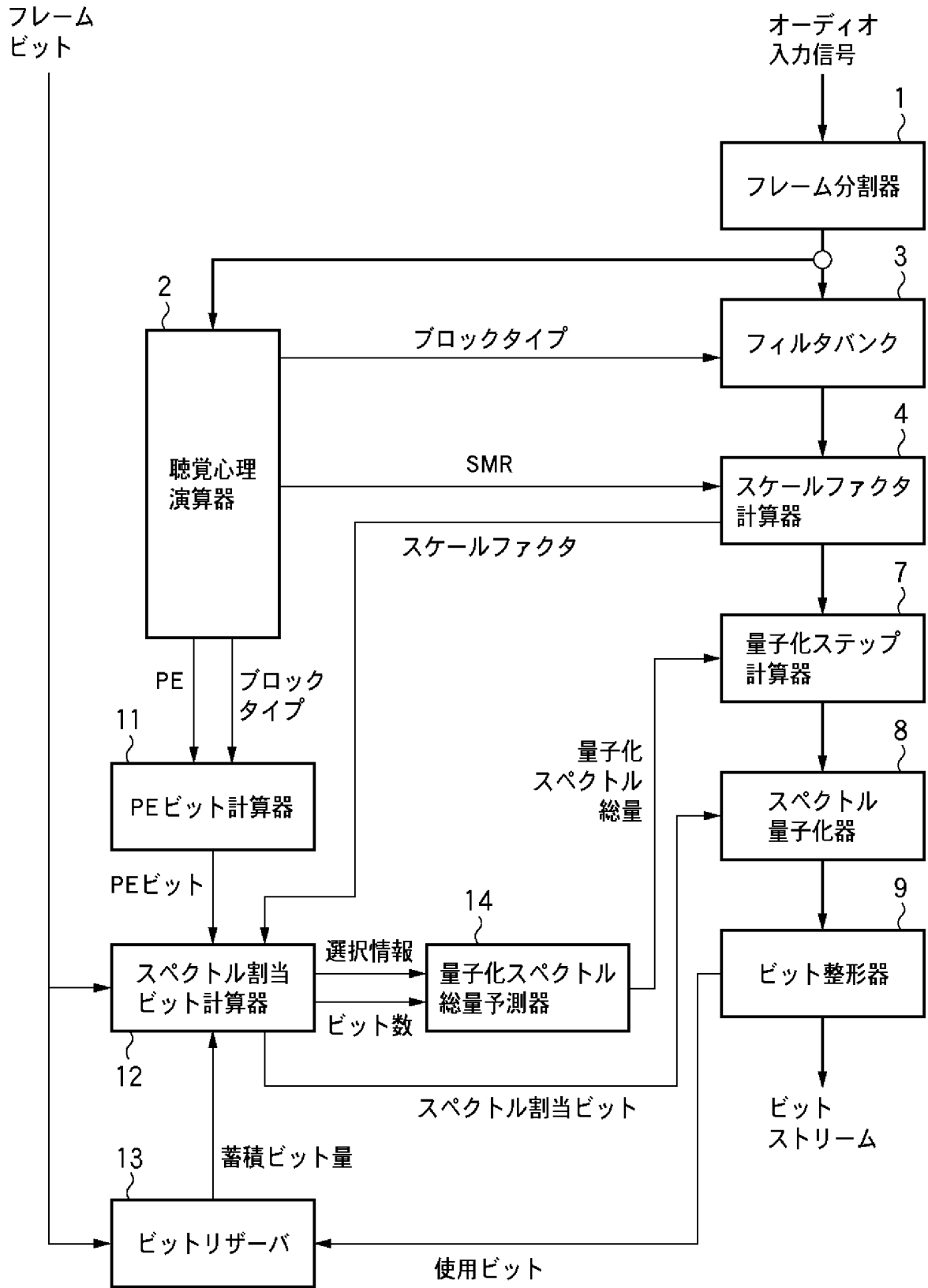
基本I/Oプログラム	
OS	
オーディオ信号符号化処理プログラム	
関連データ	
ワークエリア	
基準 ビットレート	基準 サンプリングレート
ビットレート	サンプリングレート
割当ビット上限値	平均割当ビット
PEビット	使用ビット
スケールファクタ ビット	スペクトル 割当ビット
量子化前スペクトル 聴覚情報量	量子化後スペクトル 予測情報量
許容誤差エネルギー	スペクトルバッファ
量子化スペクトル	入力信号バッファ
スケールファクタ	量子化ステップ
ブロックタイプ	SMR
PE	リザーブビット量

下位アドレス

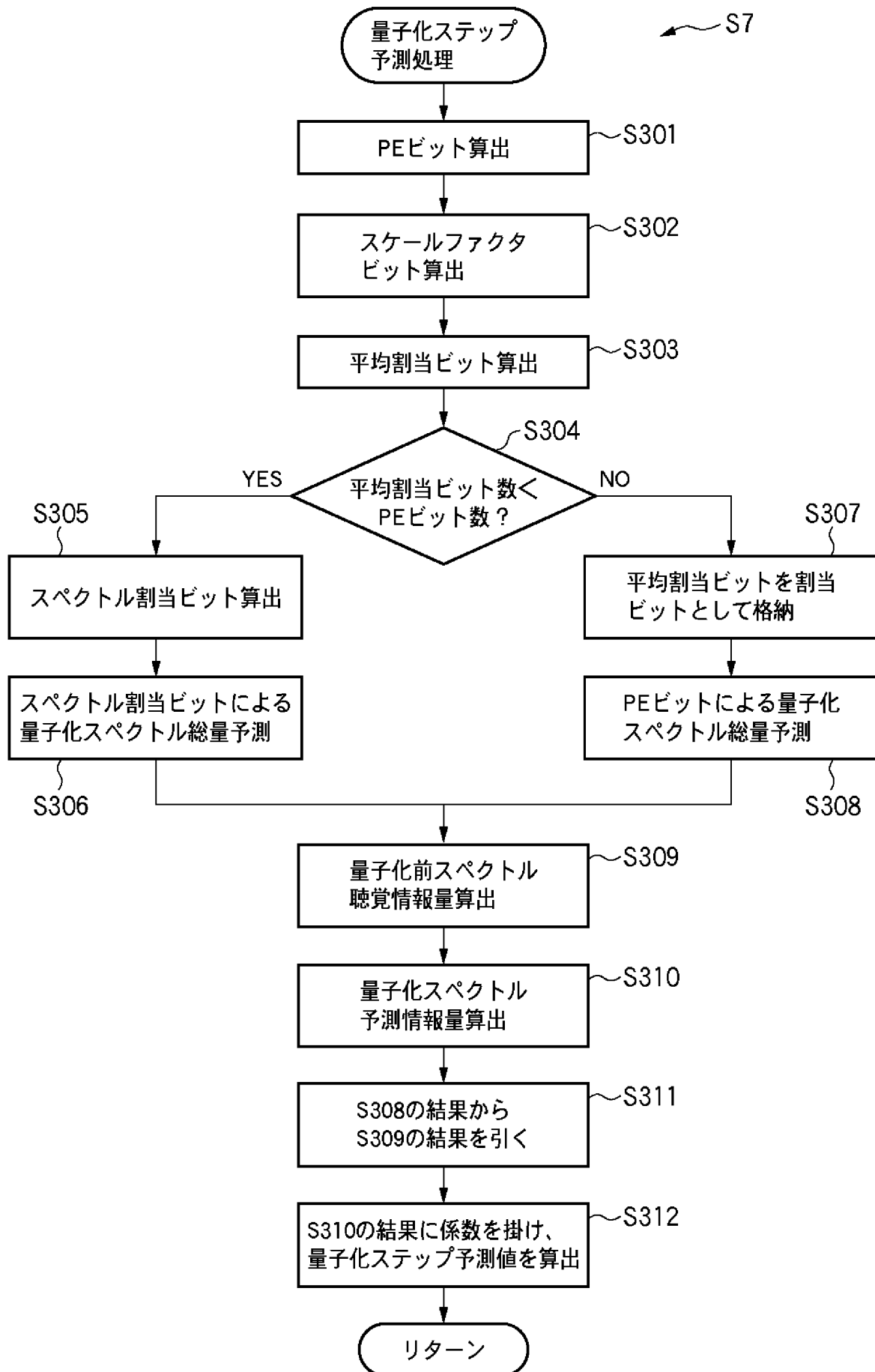
[図9]



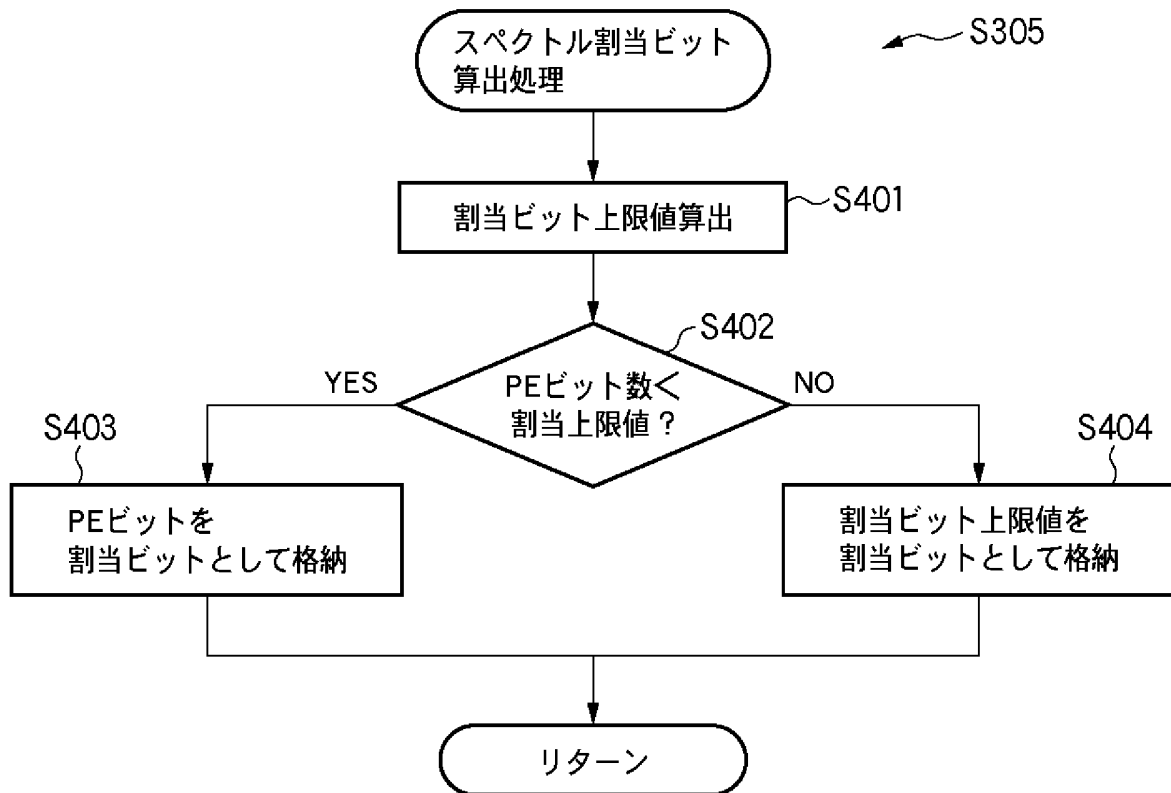
[図10]



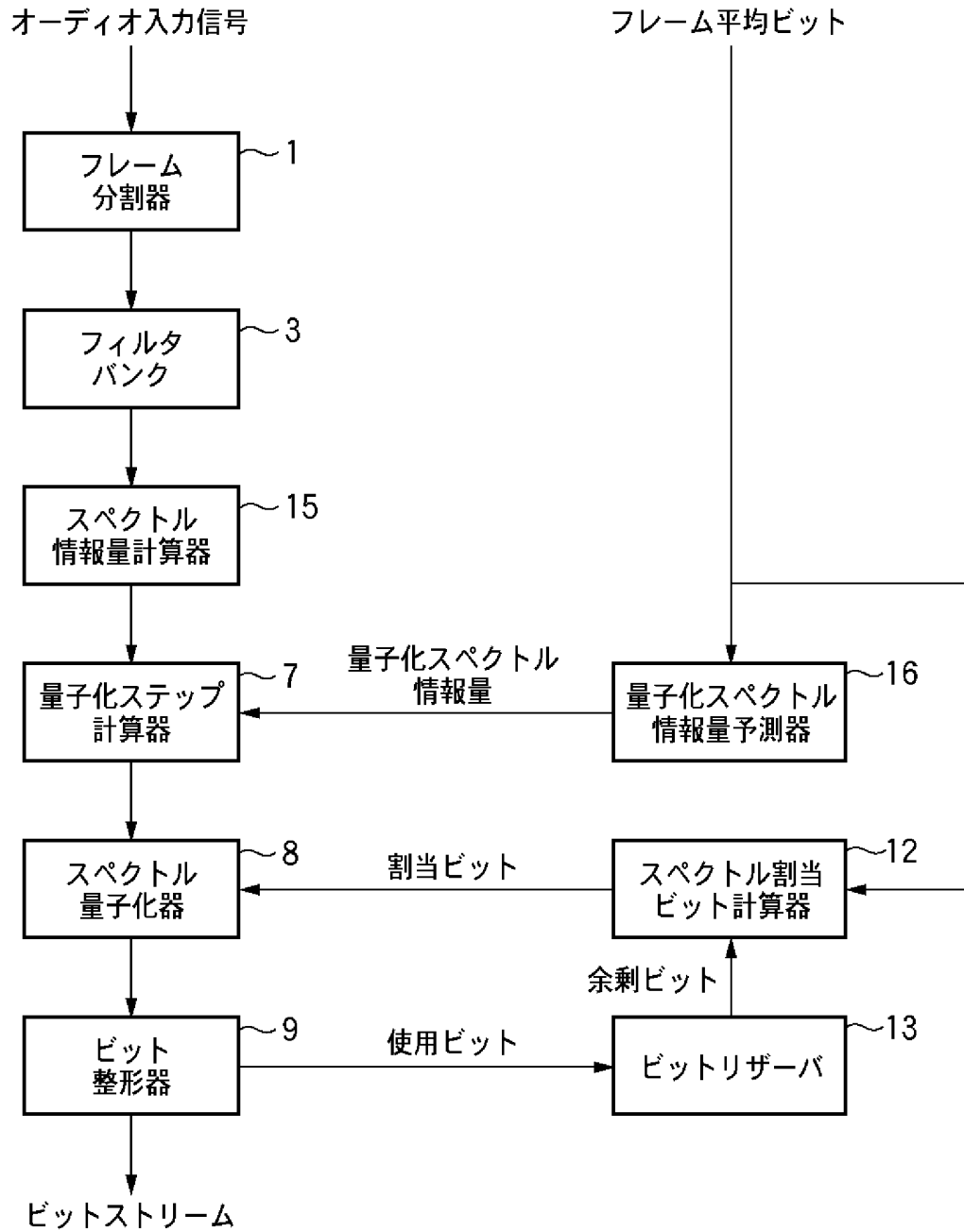
[図11]



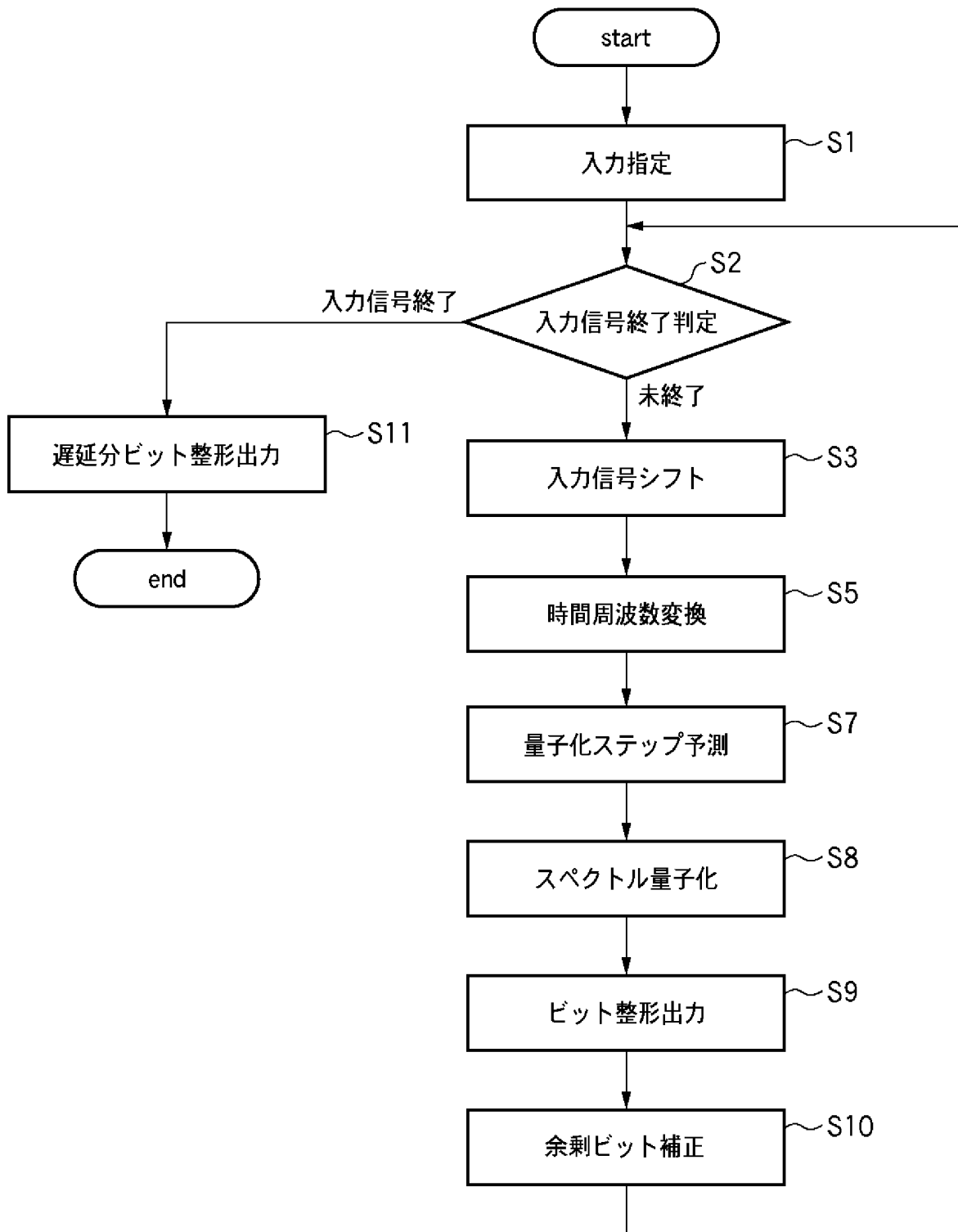
[図12]



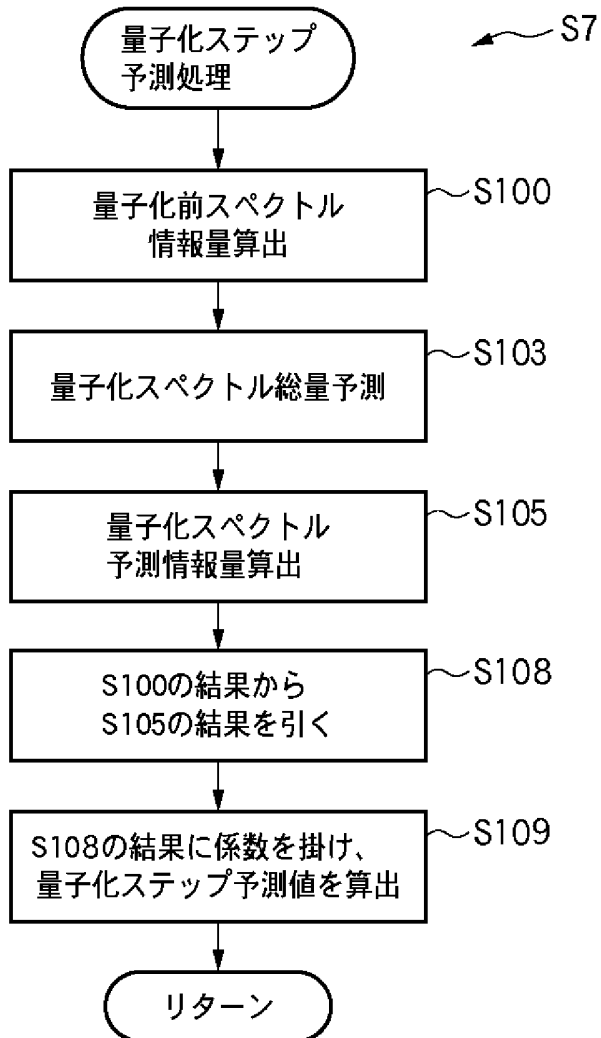
[図13]



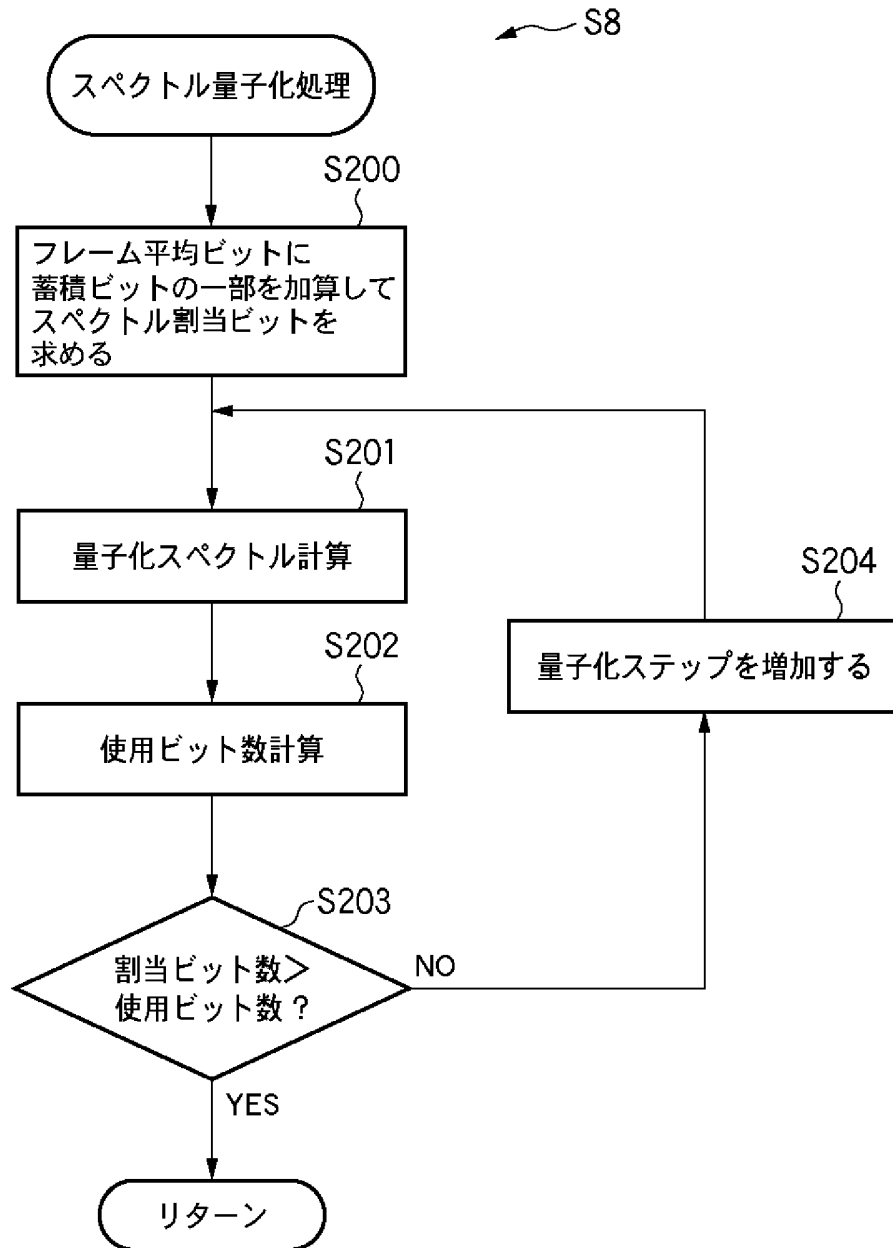
[図14]



[図15]



[図16]



[図17]

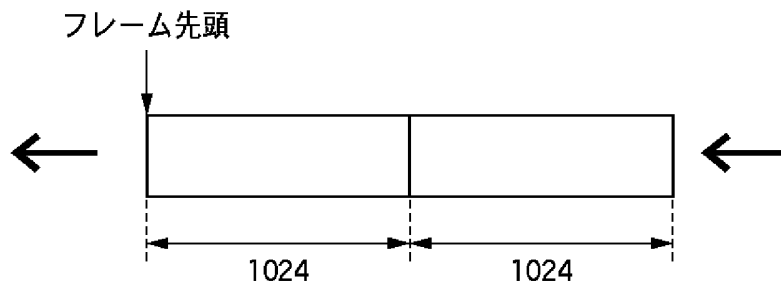
上位アドレス



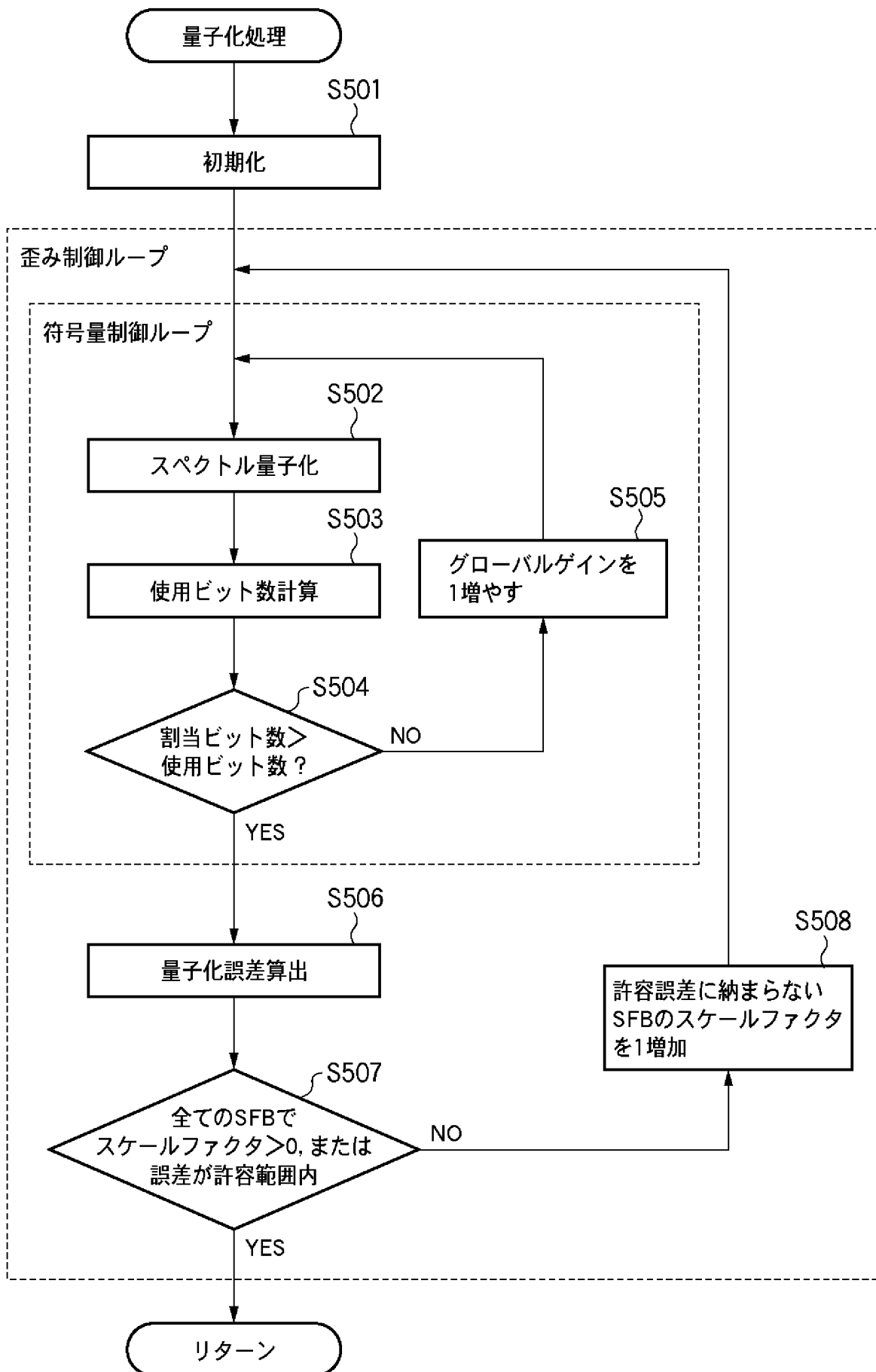
基本I/Oプログラム	
OS	
オーディオ信号符号化処理プログラム	
関連データ	
ワークエリア	
量子化前スペクトル 聴覚情報量	量子化後スペクトル 予測情報量
スペクトル割当ビット	スペクトルバッファ
量子化スペクトル	入力信号バッファ
使用ビット	量子化ステップ
ビットレート	サンプリングレート
フレーム平均ビット	リザーブビット

下位アドレス

[図18]



[図19]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2005/021014

A. CLASSIFICATION OF SUBJECT MATTER G10L19/02 (2006.01)				
According to International Patent Classification (IPC) or to both national classification and IPC				
B. FIELDS SEARCHED				
Minimum documentation searched (classification system followed by classification symbols) G10L19/00-19/14 (2006.01), H03M7/30 (2006.01)				
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Jitsuyo Shinan Koho 1922-1996 Jitsuyo Shinan Toroku Koho 1996-2006 Kokai Jitsuyo Shinan Koho 1971-2006 Toroku Jitsuyo Shinan Koho 1994-2006				
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)				
C. DOCUMENTS CONSIDERED TO BE RELEVANT				
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.		
X	JP 2002-182695 A (Matsushita Electric Industrial Co., Ltd.), 26 June, 2002 (26.06.02), Full text; Figs. 1 to 25 (Family: none)	1-16		
A	JP 6-252773 A (Sony Corp.), 09 September, 1994 (09.09.94), Full text; Figs. 1 to 7 (Family: none)	1-16		
A	JP 6-259099 A (Mitsubishi Electric Corp.), 16 September, 1994 (16.09.94), Full text; Figs. 1 to 34 (Family: none)	1-16		
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.				
<table style="width:100%; border:none;"> <tr> <td style="width:50%; border:none; vertical-align:top;"> * Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed </td> <td style="width:50%; border:none; vertical-align:top;"> "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family </td> </tr> </table>			* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family			
Date of the actual completion of the international search 31 January, 2006 (31.01.06)		Date of mailing of the international search report 07 February, 2006 (07.02.06)		
Name and mailing address of the ISA/ Japanese Patent Office		Authorized officer		
Facsimile No.		Telephone No.		

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2005/021014

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 8-237131 A (Sharp Corp.), 13 September, 1996 (13.09.96), Full text; Figs. 1 to 21 & JP 3146121 B2	1-16
A	JP 10-233692 A (Sony Corp.), 02 September, 1998 (02.09.98), Full text; Figs. 1 to 23 & US 6301558 B1	1-16
A	JP 2001-242895 A (Toshiba Corp.), 07 September, 2001 (07.09.01), Full text; Figs. 1 to 7 (Family: none)	1-16

A. 発明の属する分野の分類 (国際特許分類 (IPC))
 Int.Cl. G10L19/02(2006.01)

B. 調査を行った分野
 調査を行った最小限資料 (国際特許分類 (IPC))
 Int.Cl. G10L19/00-19/14(2006.01), H03M7/30(2006.01)

最小限資料以外の資料で調査を行った分野に含まれるもの
 日本国実用新案公報 1922-1996年
 日本国公開実用新案公報 1971-2006年
 日本国実用新案登録公報 1996-2006年
 日本国登録実用新案公報 1994-2006年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
X	JP 2002-182695 A (松下電器産業株式会社) 2002.06.26, 全文, 図1-25 (ファミリーなし)	1-16
A	JP 6-252773 A (ソニー株式会社) 1994.09.09, 全文, 図1-7 (ファミリーなし)	1-16
A	JP 6-259099 A (三菱電機株式会社) 1994.09.16, 全文, 図1-34 (ファミリーなし)	1-16

C欄の続きにも文献が列挙されている。

パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー
 「A」 特に関連のある文献ではなく、一般的技術水準を示すもの
 「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
 「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)
 「O」 口頭による開示、使用、展示等に言及する文献
 「P」 国際出願日前で、かつ優先権の主張の基礎となる出願日の後に公表された文献
 「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
 「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
 「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
 「&」 同一パテントファミリー文献

国際調査を完了した日
 31.01.2006

国際調査報告の発送日
 07.02.2006

国際調査機関の名称及びあて先
 日本国特許庁 (ISA/JP)
 郵便番号100-8915
 東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)
 山下 剛史
 電話番号 03-3581-1101 内線 3541

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	JP 8-237131 A (シャープ株式会社) 1996.09.13, 全文, 図1-21 & JP 3146121 B2	1-16
A	JP 10-233692 A (ソニー株式会社) 1998.09.02, 全文, 図1-23 & US 6301558 B1	1-16
A	JP 2001-242895 A (株式会社東芝) 2001.09.07, 全文, 図1-7 (ファミリーなし)	1-16