

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5302980号  
(P5302980)

(45) 発行日 平成25年10月2日 (2013. 10. 2)

(24) 登録日 平成25年6月28日 (2013. 6. 28)

(51) Int. Cl. F I  
 G 1 0 L 19/00 (2013. 01) G 1 0 L 19/00 4 0 0 Z  
 H 0 4 R 3/00 (2006. 01) H 0 4 R 3/00

請求項の数 16 (全 44 頁)

(21) 出願番号 特願2010-549055 (P2010-549055)  
 (86) (22) 出願日 平成21年3月4日 (2009. 3. 4)  
 (65) 公表番号 特表2011-513780 (P2011-513780A)  
 (43) 公表日 平成23年4月28日 (2011. 4. 28)  
 (86) 国際出願番号 PCT/EP2009/001533  
 (87) 国際公開番号 W02009/109373  
 (87) 国際公開日 平成21年9月11日 (2009. 9. 11)  
 審査請求日 平成23年1月12日 (2011. 1. 12)  
 (31) 優先権主張番号 61/033, 590  
 (32) 優先日 平成20年3月4日 (2008. 3. 4)  
 (33) 優先権主張国 米国 (US)

(73) 特許権者 500341779  
 フラウンホーファー・ゲゼルシャフト・ツ  
 ール・フェルデルング・デル・アンゲヴァ  
 ンテン・フォルシュング・アインゲトラ  
 ー  
 ゲネル・フェライン  
 ドイツ連邦共和国, 80686 ミュンヘ  
 ン, ハンザシュトラッセ 27ツェー  
 (74) 代理人 100085464  
 弁理士 野口 繁雄  
 (72) 発明者 マルクス・シュネル  
 ドイツ連邦共和国 91052 エルラン  
 ゲン ハルトマンシュトラッセ 7

最終頁に続く

(54) 【発明の名称】 複数の入力データストリームのミキシングのための装置

(57) 【特許請求の範囲】

【請求項 1】

第1の入力データストリーム(510-1)の第1のフレーム(540-1)と第2の入力データストリーム(510-2)の第2のフレーム(540-2)とをミキシングして、出力データストリーム(530)の出力フレーム(550)を得るための装置(500)であって、

前記出力フレーム(550)を生成するように構成された処理ユニット(520)を備えており、

前記第1のフレーム(540-1)が、第1のオーディオ信号の第1のスペクトルについて第1のクロスオーバー周波数までの下方部分を記述する第1のスペクトルデータと、前記第1のスペクトルについて前記第1のクロスオーバー周波数から始まる上方部分を記述する第1のスペクトル帯域複製(SBR)データとを含んでおり、前記第2のフレーム(540-2)が、第2のオーディオ信号の第2のスペクトルについて第2のクロスオーバー周波数までの下方部分を記述する第2のスペクトルデータと、前記第2のスペクトルについて前記第2のクロスオーバー周波数から始まる上方部分を記述する第2のSBRデータとを含んでおり、前記第1及び第2のSBRデータが、前記第1及び第2のスペクトルのそれぞれの上方部分を、時間/周波数格子分解能におけるエネルギー関連の値によって記述しており、前記第1のクロスオーバー周波数が、前記第2のクロスオーバー周波数とは異なっており、

前記出力フレーム(550)が、出力スペクトルについて出力クロスオーバー周波数ま

での下方部分を記述する出力スペクトルデータを含むとともに、前記出力スペクトルについて前記出力クロスオーバー周波数よりも上の上方部分を、出力時間/周波数格子分解能におけるエネルギー関連の値によって記述する出力SBRデータをさらに含んでおり、

前記処理ユニット(520)は、前記第1のクロスオーバー周波数、前記第2のクロスオーバー周波数及び前記出力クロスオーバー周波数の最小値を下回る周波数に対応する前記出力スペクトルデータを、前記第1及び第2のスペクトルデータに基づいてスペクトル領域において生成するように構成されており、

前記処理ユニット(520)は、前記第1のクロスオーバー周波数、前記第2のクロスオーバー周波数及び前記出力クロスオーバー周波数の最大値を上回る周波数に対応する前記出力SBRデータを、前記第1及び第2のSBRデータに基づいてSBR領域において処理するようにさらに構成されており、

10

前記処理ユニット(520)は、前記最小値及び前記最大値の間の周波数領域について、第1及び第2のスペクトルデータの少なくとも一方から少なくとも1つのSBR値を推定し、前記出力SBRデータの該当のSBR値を少なくとも前記推定したSBR値に基づいて生成するようにさらに構成されている装置(500)。

#### 【請求項2】

前記処理ユニット(520)が、前記少なくとも1つのSBR値の推定を、該推定すべきSBR値に対応する周波数成分に対応するスペクトル値に基づいて行うように構成されている請求項1に記載の装置(500)。

#### 【請求項3】

20

第1の入力データストリーム(510-1)の第1のフレーム(540-1)と第2の入力データストリーム(510-2)の第2のフレーム(540-2)とをミキシングして、出力データストリーム(530)の出力フレーム(550)を得るための装置(500)であって、

前記出力フレーム(550)を生成するように構成された処理ユニット(520)を備えており、

前記第1のフレーム(540-1)が、第1のオーディオ信号の第1のスペクトルについて第1のクロスオーバー周波数までの下方部分を記述する第1のスペクトルデータと、前記第1のスペクトルについて前記第1のクロスオーバー周波数から始まる上方部分を記述する第1のスペクトル帯域複製(SBR)データとを含んでおり、前記第2のフレーム(540-2)が、第2のオーディオ信号の第2のスペクトルについて第2のクロスオーバー周波数までの下方部分を記述する第2のスペクトルデータと、前記第2のスペクトルについて前記第2のクロスオーバー周波数から始まる上方部分を記述する第2のSBRデータとを含んでおり、前記第1及び第2のSBRデータが、前記第1及び第2のスペクトルのそれぞれの上方部分を、時間/周波数格子分解能におけるエネルギー関連の値によって記述しており、前記第1のクロスオーバー周波数が、前記第2のクロスオーバー周波数とは異なっており、

30

前記出力フレーム(550)が、出力スペクトルについて出力クロスオーバー周波数までの下方部分を記述する出力スペクトルデータを含むとともに、前記出力スペクトルについて前記出力クロスオーバー周波数よりも上の上方部分を、出力時間/周波数格子分解能におけるエネルギー関連の値によって記述する出力SBRデータをさらに含んでおり、

40

前記処理ユニット(520)は、前記第1のクロスオーバー周波数、前記第2のクロスオーバー周波数及び前記出力クロスオーバー周波数の最小値を下回る周波数に対応する前記出力スペクトルデータを、前記第1及び第2のスペクトルデータに基づいてスペクトル領域において生成するように構成されており、

前記処理ユニット(520)は、前記第1のクロスオーバー周波数、前記第2のクロスオーバー周波数及び前記出力クロスオーバー周波数の最大値を上回る周波数に対応する前記出力SBRデータを、前記第1及び第2のSBRデータに基づいてSBR領域において処理するようにさらに構成されており、

さらに当該装置(500)は、前記最小値及び前記最大値の間の周波数領域について、

50

前記第 1 及び第 2 のフレームの少なくとも一方からの少なくとも 1 つのスペクトル値を、該当のフレームの前記 S B R データに基づいて推定し、前記出力スペクトルデータの該当のスペクトル値を、少なくとも前記推定したスペクトル値に基づいて、該推定したスペクトル値をスペクトル領域において処理することによって生成するように構成されている装置 ( 5 0 0 )。

【請求項 4】

前記処理ユニットは、前記少なくとも 1 つのスペクトル値を、前記 S B R データ及び前記それぞれのフレームの前記それぞれのスペクトルの前記下方部分の前記スペクトルデータに基づいてスペクトル成分について少なくとも 1 つのスペクトル値を再現することに基づいて、推定するように構成されている請求項 3 に記載の装置。

10

【請求項 5】

前記処理ユニット ( 5 2 0 ) は、前記出力クロスオーバー周波数を前記第 1 のクロスオーバー周波数又は前記第 2 のクロスオーバー周波数になるように決定するように構成されている請求項 1 から 4 のいずれか一項に記載の装置 ( 5 0 0 )。

【請求項 6】

前記処理ユニット ( 5 2 0 ) は、前記出力クロスオーバー周波数を第 1 及び第 2 のクロスオーバー周波数のうちの低い方のクロスオーバー周波数へと設定し、又は前記出力クロスオーバー周波数を前記第 1 及び第 2 のクロスオーバー周波数のうちの高い方へと設定するように構成されている請求項 1 から 5 のいずれか一項に記載の装置 ( 5 0 0 )。

【請求項 7】

前記処理ユニット ( 5 2 0 ) は、前記出力時間 / 周波数格子分解能を、前記第 1 又は第 2 のフレームの前記時間 / 周波数格子分解能によって示される過渡の過渡位置に適合するように決定するように構成されている請求項 1 から 6 のいずれか一項に記載の装置 ( 5 0 0 )。

20

【請求項 8】

前記処理ユニット ( 5 2 0 ) は、前記第 1 及び第 2 のフレームの前記時間 / 周波数格子分解能が 2 つ以上の過渡の存在を示している場合に、前記時間 / 周波数格子分解能を、前記第 1 及び第 2 のフレームの前記時間 / 周波数格子分解能によって示される早い方の過渡に適合するように設定するように構成されている請求項 7 に記載の装置 ( 5 0 0 )。

【請求項 9】

前記処理ユニット ( 5 2 0 ) は、S B R 周波数領域又は S B R 領域における線形結合に基づいて、スペクトルデータ又は S B R データを出力するように構成されている請求項 1 から 8 のいずれか一項に記載の装置 ( 5 0 0 )。

30

【請求項 10】

前記処理ユニット ( 5 2 0 ) は、前記第 1 及び第 2 のフレームの正弦曲線関連の S B R データの線形結合に基づいて、正弦曲線関連の S B R データを含む前記出力 S B R データを生成するように構成されている請求項 1 から 9 のいずれか一項に記載の装置 ( 5 0 0 )。

【請求項 11】

前記処理ユニット ( 5 2 0 ) は、前記第 1 及び第 2 のフレームのノイズ関連の S B R データの線形結合に基づいて、ノイズ関連の S B R データを含む前記出力 S B R データを生成するように構成されている請求項 1 から 10 のいずれか一項に記載の装置 ( 5 0 0 )。

40

【請求項 12】

前記処理ユニット ( 5 2 0 ) は、前記第 1 及び第 2 のフレームのそれぞれの S B R データの関連性の心理音響的推定に基づく正弦曲線関連又はノイズ関連の S B R データを含むように構成されている請求項 10 又は 11 に記載の装置 ( 5 0 0 )。

【請求項 13】

前記処理ユニット ( 5 2 0 ) は、平滑化フィルター処理に基づいて前記出力 S B R データを生成するように構成されている請求項 1 から 12 のいずれか一項に記載の装置 ( 5 0 0 )。

50

## 【請求項 1 4】

複数の入力データストリーム(510)を処理するように構成されており、該複数の入力データストリームが、3つ以上の入力データストリームを含んでおり、該複数の入力データストリームが、前記第1及び第2の入力データストリーム(510-1、510-2)を含んでいる請求項1から13のいずれか一項に記載の装置(500)。

## 【請求項 1 5】

第1の入力データストリーム(510-1)の第1のフレーム(540-1)と第2の入力データストリーム(510-2)の第2のフレーム(540-2)とをミキシングして出力データストリーム(530)の出力フレーム(550)を得るための方法であって、前記第1のフレームが、第1のオーディオ信号のスペクトルについて第1のクロスオーバー周波数までの下方部分を記述する第1のスペクトルデータと、前記スペクトルについて前記第1のクロスオーバー周波数から始まる上方部分を記述する第1のスペクトル帯域複製(SBR)データとを含んでおり、前記第2のフレームが、第2のオーディオ信号の第2のスペクトルについて第2のクロスオーバー周波数までの下方部分を記述する第2のスペクトルデータと、第2のスペクトルについて前記第2のクロスオーバー周波数から始まる上方部分を記述する第2のSBRデータとを含んでおり、前記第1及び第2のSBRデータが、前記それぞれのスペクトルの前記それぞれの上方部分を、時間/周波数格子分解能におけるエネルギー関連の値によって記述しており、前記第1のクロスオーバー周波数が、前記第2のクロスオーバー周波数とは異なっており、

当該方法は、

出力スペクトルの出力クロスオーバー周波数までの下方部分を記述する出力スペクトルデータを含むとともに、該出力スペクトルについて前記出力クロスオーバー周波数よりも上の上方部分を、出力時間/周波数格子分解能におけるエネルギー関連の値によって記述する出力SBRデータをさらに含む前記出力フレームを生成するステップ、

前記第1のクロスオーバー周波数、前記第2のクロスオーバー周波数及び出力クロスオーバー周波数の最小値を下回る周波数に対応するスペクトルデータを、前記第1及び第2のスペクトルデータに基づいてスペクトル領域において生成するステップ、

前記第1のクロスオーバー周波数、前記第2のクロスオーバー周波数及び前記出力クロスオーバー周波数の最大値を上回る周波数に対応する出力SBRデータを、前記第1及び第2のSBRデータに基づいてSBR領域において生成するステップ、並びに

前記最小値及び前記最大値の間の周波数領域の周波数について、第1及び第2のスペクトルデータの少なくとも一方から少なくとも1つのSBR値を推定し、前記出力SBRデータの該当のSBR値を少なくとも前記推定したSBR値に基づいて生成するステップ、又は

前記最小値及び前記最大値の間の周波数領域の周波数について、該当のフレームの前記SBRデータに基づいて前記第1及び第2のフレームの少なくとも一方からの少なくとも1つのスペクトル値を推定し、前記出力スペクトルデータのスペクトル値を、少なくとも前記推定したスペクトル値に基づいて、該推定したスペクトル値をスペクトル領域において処理することによって生成するステップを含んでいる方法。

## 【請求項 1 6】

第1の入力データストリームの第1のフレームと第2の入力データストリームの第2のフレームとのミキシングのための請求項15に記載の方法をプロセッサに実行させるためのプログラム。

## 【発明の詳細な説明】

## 【技術分野】

## 【0001】

本発明による実施の形態は、複数の入力データストリームをミキシングして、出力データストリームを得るための装置に関する。そのような装置を、例えばビデオ会議システム及びテレビ会議システムなどの会議システムの分野において使用することができる。

## 【背景技術】

## 【 0 0 0 2 】

多くの用途において、2つ以上のオーディオ信号が、複数のオーディオ信号から1つの信号又は少なくともより少ない数の信号が生成されるような方法で処理される。これは、多くの場合、「ミキシング」と称される。したがって、オーディオ信号のミキシングの処理を、いくつかの個別のオーディオ信号を結果としての信号へと束ねるものと称することができる。このプロセスは、例えば、コンパクトディスクのために楽曲を生成する場合に使用される（「合成録音」）。この場合、典型的には、種々の楽器からの種々のオーディオ信号が、声楽演奏（歌唱）を含む1つ以上のオーディオ信号と一緒に歌曲へとミックスされる。

## 【 0 0 0 3 】

ミキシングが重要な役割を果たすさらなる応用の分野は、ビデオ会議システム及びテレビ会議システムである。そのようなシステムは、典型的には、登録済みの参加者から到来するビデオ及びオーディオデータを適切にミックスし、得られた信号を各々の参加者へ返送する中央サーバを使用することによって、何人かの空間的に離れた会議の参加者を接続することができる。この得られた信号又は出力信号は他のすべての会議参加者のオーディオ信号を含んでいる。

## 【 0 0 0 4 】

現代のデジタル会議システムにおいては、いくつかの部分的に相反する目標及び態様が互いに競合する。さまざまな種類のオーディオ信号（例えば、一般的なオーディオ信号及び音楽信号と比べたスピーチ信号）について、再生されるオーディオ信号の品質、ならびにいくつかのコーディング及びデコーディングの技法の適用可能性及び有用可能性を考慮しなければならない。会議システムの設計及び実施の際に考慮が必要であると考えられるさらなる局面は、利用可能な帯域幅及び遅延の問題である。

## 【 0 0 0 5 】

例えば、品質と帯域幅とをバランスさせるとき、多くの場合に妥協は避けられない。しかしながら、AAC-ELD技法（AAC＝アドバンスド・オーディオ・コーディング；ELD＝エンハンスド・ロー・ディレイ）などの最新のコーディング及びデコーディング技法を実施することによって、品質に関する改善を達成することが可能である。しかしながら、達成できる品質は、そのような最新の技法を使用するシステムにおいて、より基本的な問題及び見地による悪影響を受ける。

## 【 0 0 0 6 】

達成すべき課題を1つだけ挙げると、すべてのデジタル信号伝送は、量子化が必要であるという問題に直面する。そのような量子化は、少なくとも原理的には、ノイズのないアナログシステムにおいて理想的な環境のもとでは回避することができる。量子化プロセスによって、或る量の量子化ノイズが処理対象の信号へ持ち込まれることは避けられない。生じうる可聴なひずみに対処するために、量子化のレベル数を増やし、すなわち量子化の分解能を高めることが考えられる。しかしながら、そのようにすることで、伝送すべき信号値の数が多くなり、伝送すべきデータの量が多くなる。換言すると、量子化ノイズによって持ち込まれる可能性があるひずみを少なくすることによって品質を高めることは、特定の環境下では、伝送されるデータの量を増加させ、伝送システムに課された帯域幅の制約に最終的に違反する可能性がある。

## 【 0 0 0 7 】

会議システムの場合には、典型的に2つ以上の入力オーディオ信号を処理しなければならないという事実によって、品質、利用可能な帯域幅及び他のパラメータの間のトレードオフを改善するという課題がさらに複雑になる可能性がある。すなわち、会議システムによって生成される出力信号又は得られる信号を生成するときに、2つ以上のオーディオ信号によって課される境界条件を考慮しなければならない可能性がある。

## 【 0 0 0 8 】

特に、会議の参加者間の直接的なやり取りを、参加者が容認できないと考えるかもしれない実質的な遅延を持ち込むことなく可能にするために、遅延が十分に少ない会議システ

10

20

30

40

50

ムを実現するというさらなる課題に照らすと、課題はさらに大きくなる。

【 0 0 0 9 】

遅延の少ない会議システムの実現においては、遅延の原因が、典型的には、それらの数に関して限定され、これが、他方では、オーディオ信号のミキシングをそれぞれの信号の重畳又は加算によって達成できる時間領域の外部でのデータの処理という課題につながりうる。

【 0 0 1 0 】

一般的なオーディオ信号の場合において、品質とビットレートとの間のトレードオフを改善するために、そのような矛盾するパラメータ（再生される信号の品質、ビットレート、遅延、計算の複雑さ、及びさらなるパラメータ、など）の間のトレードオフをさらに改善することができるかなりの数の技法が存在している。

10

【 0 0 1 1 】

上述のトレードオフを改善するためのきわめて柔軟なツールが、いわゆるスペクトル帯域表現ツール（SBR）である。SBRモジュールは、典型的には、MPEG-4 AACエンコーダなどの中央エンコーダの一部として実現されるのではなく、むしろ追加のエンコーダ及びデコーダである。SBRは、オーディオ信号内のより高い周波数とより低い周波数との間の相関を利用する。SBRは、信号のより高い周波数が、地盤の振動の単なる整数倍であり、したがってより高い周波数を、より低いスペクトルに基づいて再現することができるという仮定に基づいている。さらに、人間の耳の可聴分解能は、より高い周波数の場合において対数的であるため、より高い周波数範囲に関する小さな差は、きわめて熟練の聴取者でなければ実感することができず、したがってSBRエンコーダによって持ち込まれる不正確さは、おそらくは、圧倒的多数の聴取者には知覚されない。

20

【 0 0 1 2 】

SBRエンコーダは、MPEG-4エンコーダへ供給されるオーディオ信号を前処理し、入力信号を周波数範囲へと分ける。より低い周波数範囲又は周波数帯は、上方の周波数帯又は周波数範囲から、いわゆるクロスオーバー周波数によって分離される。クロスオーバー周波数は、利用可能なビットレート及びさらなるパラメータに応じて、さまざまに設定することが可能である。SBRエンコーダは、典型的には直交ミラー・フィルター・バンド（QMF）となるように実現される周波数を分析するためのフィルターバンクを使用する。

30

【 0 0 1 3 】

SBRエンコーダは、上方の周波数範囲の周波数表現からエネルギー値を抽出し、これが後に、この周波数範囲を下方の周波数帯に基づいて再現するために使用される。

【 0 0 1 4 】

したがって、SBRエンコーダは、SBRデータ又はSBRパラメータをフィルター処理されたオーディオ信号又はフィルター処理されたオーディオデータと一緒にコアエンコーダへともたらし、これが、元のオーディオ信号のサンプリング周波数の半分に基づいて下方の周波数帯へと適用される。これは、処理されるサンプル値を大幅に少なくする機会をもたらし、したがって個々の量子化レベルをより高精度に設定することができる。SBRエンコーダによってもたらされる追加のデータ、すなわちSBRパラメータは、副情報として、MPEG-4エンコーダ又は他の任意のエンコーダによって得られるビットストリームに保存される。これは、適切なビットマルチプレクサを使用することによって達成できる。

40

【 0 0 1 5 】

デコーダ側では、到着するビットストリームが、最初にビットデマルチプレクサによって分離され、少なくともSBRデータが分離されて、SBRデコーダへともたらされる。しかしながら、SBRデコーダがSBRパラメータを処理する前に、最初に下方の周波数帯がコアデコーダによってデコードされ、下方の周波数帯のオーディオ信号が再現される。SBRデコーダそのものが、SBRエネルギー値（SBRパラメータ）及び下方の周波数範囲のスペクトル情報に基づいて、オーディオ信号のスペクトルの上方部分を計算する

50

。換言すると、SBRデコーダが、オーディオ信号の上方のスペクトル帯域を、上述のビットストリームにて伝達される下方の帯域及びSBRパラメータに基づいて複製する。上述したSBRモジュールの可能性の他に、再現されるオーディオ信号の全体としてのオーディオの知覚を向上させるために、SBRは、さらなるノイズ源ならびに個々の正弦曲線をエンコードする可能性をさらに提供する。

**【0016】**

したがって、SBRは、品質とビットレートとの間のトレードオフを改善するためのきわめて柔軟なツールに相当し、このことが、SBRを会議システムの分野における応用のための興味深い候補にもしている。しかしながら、複雑さならびにきわめて多数の可能性及び選択肢ゆえに、SBRによってエンコードされたオーディオ信号は、これまでのところ、それぞれのオーディオ信号を時間領域の信号へと完全にデコードし、この領域において実際のミキシング処理を実行し、その後ミックスされた信号をSBRによってエンコードされた信号へと再びエンコードすることによって、時間領域においてのみミックスされている。信号を時間領域へとエンコードすることに起因して持ち込まれる追加の遅延の他に、エンコードされたオーディオ信号のスペクトル情報の再現がかなりの計算の複雑さを必要とする可能性があり、このことが、例えば可搬の用途や他のエネルギー効率又は効率的な計算が求められる用途の場合に、魅力的でないかもしれない。

10

**【発明の概要】****【発明が解決しようとする課題】****【0017】**

したがって、本発明の目的は、SBRによってエンコードされたオーディオ信号をミキシングする際に必要な計算の複雑さを軽減することにある。

20

**【課題を解決するための手段】****【0018】**

この目的は、請求項1もしくは3に記載の装置、請求項15に記載の方法、又は請求項16に記載のプログラムによって達成される。

**【0019】**

本発明による実施の形態は、関係するクロスオーバー周波数の最小値を下回る周波数についてはスペクトル領域においてスペクトル情報をミキシングすることによってミキシングを実行し、最大のクロスオーバー周波数を上回る周波数についてはSBR領域においてミキシングを実行し、最小値と最大値との間の領域の周波数については、少なくとも1つのSBR値を推定して、少なくとも推定によるSBR値に基づいて該当のSBR値を生成するか、又はそれぞれのSBRデータに基づいてスペクトル値もしくはスペクトル情報を推定して、この推定によるスペクトル値もしくはスペクトル情報に基づいてスペクトル情報のスペクトル値を生成することによってミキシングを実行することで、計算の複雑さを軽減できるという発見に基づいている。

30

**【0020】**

換言すると、本発明による実施の形態は、最大のクロスオーバー周波数を上回る周波数について、ミキシングをSBR領域において実行できる一方で、最小のクロスオーバー周波数を下回る周波数について、ミキシングを対応するスペクトル値を直接的に処理することによってスペクトル領域において実行できるという発見に基づいている。さらに、本発明の実施の形態による装置は、前記最大及び最小値の間の周波数について、該当するSBR値からスペクトル値を推定し、又はスペクトル値からSBR値を推定し、実際のミキシングをSBR領域もしくはスペクトル領域において前記推定した値に基づいて実行することによって、ミキシングをSBR領域又はスペクトル領域において実行することができる。この文脈において、出力クロスオーバー周波数が、入力データストリームのクロスオーバー周波数のいずれかであってよく、あるいは他の値であってよいことに、注意すべきである。

40

**【0021】**

結果として、すべての関連のクロスオーバー周波数よりも上及び下の実際のミキシング

50

が、それぞれの領域における直接的なミキシングに基づいて実行される一方で、推定は、関係するすべてのクロスオーバー周波数のうちの最小値とすべてのクロスオーバー周波数のうちの最大値との間の中間領域においてのみ実行すればよいため、装置によって実行すべき工程の数が少なくなり、したがって必要な計算の複雑さが軽減される。上記推定にもとづき、実際のSBR値又は実際のスペクトル値が計算又は決定される。したがって、多くの場合に、上記中間周波数領域においても、推定及び処理を関係するすべての入力データストリームについて実行する必要は典型的にはないため、計算の複雑さが軽減される。

【0022】

本発明の実施の形態による態様において、出力クロスオーバー周波数は、入力データストリームのクロスオーバー周波数のうちの1つに等しくてよく、あるいは出力クロスオーバー周波数を、例えば心理音響的な評価の結果を考慮して、独自に選択することができる。さらに、本発明による実施の形態においては、生成されたSBRデータ又は生成されたスペクトル値を、中間周波数範囲のSBRデータ又はスペクトル値を平滑化又は変更するために、さまざまに適用することが可能である。

【0023】

本発明による実施の形態を、以下の図面を参照しつつ、以下で説明する。

【図面の簡単な説明】

【0024】

【図1】会議システムのブロック図を示している。

【図2】一般的なオーディオコーデックに基づく会議システムのブロック図を示している。

【図3】ビットストリームミキシング技術を使用して周波数領域で動作する会議システムのブロック図を示している。

【図4】複数のフレームを含んでいるデータストリームの概略図を示している。

【図5】スペクトル成分ならびにスペクトルデータ又は情報の異なる形態を示している。

【図6A】第1の入力データストリームの第1のフレームと第2の入力データストリームの第2のフレームとをミキシングするための本発明の実施の形態による装置の簡単なブロック図を示している。

【図6B】データストリームのフレームの時間/周波数格子分解能のブロック図を示している。

【図7】本発明の実施の形態による装置のさらに詳細なブロック図を示している。

【図8】会議システムの文脈において複数の入力データストリームをミキシングするための本発明のさらなる実施の形態による装置のブロック図を示している。

【図9A】本発明の実施の形態による装置へもたらされるような第1の入力データストリームの第1のフレームを示している。

【図9B】本発明の実施の形態による装置へもたらされるような第2の入力データストリームの第2のフレームを示している。

【図9C】図9A及び9Bに示した入力フレームの重ね合わせの状況を示している。

【図9D】本発明の実施の形態による装置によって生成されるような出力フレームを示しており、出力クロスオーバー周波数が、入力フレームの2つのクロスオーバー周波数のうちの低い方である。

【図9E】本発明の実施の形態による装置によって生成されるような出力フレームを示しており、出力クロスオーバー周波数が、入力フレームのクロスオーバー周波数のうちの高い方である。

【図10】低い周波数格子分解能及び高い周波数格子分解能の照合を示している。

【発明を実施するための形態】

【0025】

図4から10に関して、本発明による種々の実施の形態を、さらに詳しく説明する。しかしながら、これらの実施の形態をさらに詳しく説明する前に、最初に図1から3に関して、会議システムの枠組みにおいて重要になるであろう課題及び要望に照らして、簡単な

10

20

30

40

50



序論を提示する。

【 0 0 2 6 】

図 1 は、多地点制御ユニット ( M C U ) とも称することができる会議システム 1 0 0 のブロック図を示している。その機能に関する説明から明らかになるとおり、図 1 に示されているような会議システム 1 0 0 は、時間領域において機能するシステムである。

【 0 0 2 7 】

図 1 に示されているような会議システム 1 0 0 は、適切な数の入力 1 1 0 - 1、1 1 0 - 2、1 1 0 - 3、・・・ ( 図 1 には、そのうちの 3 つだけが示されている。 ) を介して複数の入力データストリームを受け取るように構成されている。入力 1 1 0 の各々は、それぞれのデコーダ 1 2 0 へと接続されている。より正確には、第 1 の入力データストリームのための入力 1 1 0 - 1 が第 1 のデコーダ 1 2 0 - 1 へ接続され、第 2 の入力 1 1 0 - 2 が第 2 のデコーダ 1 2 0 - 2 へ接続され、第 3 の入力 1 1 0 - 3 が第 3 のデコーダ 1 2 0 - 3 へ接続されている。

10

【 0 0 2 8 】

さらに、会議システム 1 0 0 は、適切な数の加算器 1 3 0 - 1、1 3 0 - 2、1 3 0 - 3、・・・ ( 図 1 には、やはりそのうちの 3 つだけが示されている。 ) を備えている。各々の加算器が、会議システム 1 0 0 の入力 1 1 0 のうちの 1 つに組み合わせられている。例えば、第 1 の加算器 1 3 0 - 1 が、第 1 の入力 1 1 0 - 1 及び対応するデコーダ 1 2 0 - 1 に組み合わせられている。

【 0 0 2 9 】

各々の加算器 1 3 0 は入力 1 1 0 が接続されているデコーダ 1 2 0 を除くすべてのデコーダ 1 2 0 の出力へ接続されている。換言すると、第 1 の加算器 1 3 0 - 1 は第 1 のデコーダ 1 2 0 - 1 を除くすべてのデコーダ 1 2 0 へと接続されている。したがって、第 2 の加算器 1 3 0 - 2 は第 2 のデコーダ 1 2 0 - 2 を除くすべてのデコーダ 1 2 0 へ接続されている。

20

【 0 0 3 0 】

さらに、各々の加算器 1 3 0 はそれぞれ 1 つのエンコーダ 1 4 0 へ接続された出力を備えている。すなわち、第 1 の加算器 1 3 0 - 1 の出力は第 1 のエンコーダ 1 4 0 - 1 へ接続されている。したがって、第 2 の加算器 1 3 0 - 2 及び第 3 の加算器 1 3 0 - 3 もそれぞれ第 2 のエンコーダ 1 4 0 - 2 及び第 3 のエンコーダ 1 4 0 - 3 へ接続されている。

30

【 0 0 3 1 】

次いで、各々のエンコーダ 1 4 0 はそれぞれの出力 1 5 0 へ接続されている。換言すると、例えば第 1 のエンコーダは、例えば第 1 の出力 1 5 0 - 1 へ接続されている。第 2 のエンコーダ 1 4 0 - 2 及び第 3 のエンコーダ 1 4 0 - 3 もそれぞれ第 2 の出力 1 5 0 - 2 及び第 3 の出力 1 5 0 - 3 へ接続されている。

【 0 0 3 2 】

図 1 に示されているような会議システム 1 0 0 の動作をさらに詳しく説明できるよう、図 1 は第 1 の参加者の会議端末 1 6 0 をさらに示している。会議端末 1 6 0 は、例えばデジタル電話 ( 例えば、I S D N 電話 ( I S D N = 総合デジタル通信網 ) )、ボイスオーバー IP インフラストラクチャーを備えているシステム、又は同様の端末とすることができる。

40

【 0 0 3 3 】

会議端末 1 6 0 は会議システム 1 0 0 の第 1 の入力 1 1 0 - 1 へ接続されたエンコーダ 1 7 0 を備えている。さらに、会議端末 1 6 0 は、会議システム 1 0 0 の第 1 の出力 1 5 0 - 1 へ接続されたデコーダ 1 8 0 を備えている。

【 0 0 3 4 】

同様の会議端末 1 6 0 が、さらなる参加者の場所にも存在することができる。それらの会議端末は、単に簡素化のために、図 1 には示されていない。また、会議システム 1 0 0 と会議端末 1 6 0 は、決して互いに物理的に近くに存在する必要がないことに注意すべきである。会議端末 1 6 0 と会議システム 1 0 0 を、例えば W A N 技術 ( W A N = 広域ネッ

50

トワーク)によってのみ接続することができる異なる場所に配置することができる。

【0035】

人間であるユーザとのより分かり易い方法でのオーディオ信号の交換を可能にするために、さらに会議端末160は、マイクロホン、増幅器、及びスピーカー又はヘッドホンなど、追加の部品を備えることができ、又はこのような追加の部品へ接続することができる。それらは、単に簡素化のために、図1には示されていない。

【0036】

すでに示したように、図1に示した会議システム100は、時間領域において機能するシステムである。例えば、第1の参加者がマイクロホン(図1には示されていない)に話しかけると、会議端末160のエンコーダ170がそれぞれのオーディオ信号を対応するビットストリームへとエンコードし、このビットストリームを会議システム100の第1の入力110-1へ伝達する。

10

【0037】

会議システム100の内部において、ビットストリームが第1のデコーダ120-1によってデコードされ、再び時間領域へ変換される。第1のデコーダ120-1が第2のミキサー130-2及び第3のミキサー130-3へ接続されているため、第1の参加者によって生成されたとおりのオーディオ信号は、その再現されたオーディオ信号を第2及び第3の参加者のそれぞれからのさらなる再現オーディオ信号と単純に加えることによって、時間領域においてミックスすることができる。

【0038】

20

このことは、それぞれ第2の参加者及び第3の参加者によってもたらされ、第2の入力110-2及び第3の入力110-3によって受信され、第2のデコーダ120-2及び第3のデコーダ120-3によって処理されるオーディオ信号にも当てはまる。次いで、第2の参加者及び第3の参加者のこれらの再現オーディオ信号が第1のミキサー130-1へもたらされ、第1のミキサー130-1は時間領域の合計のオーディオ信号を第1のエンコーダ140-1へもたらず。エンコーダ140-1は、合計のオーディオ信号を再びエンコードしてビットストリームを形成し、このビットストリームを第1の出力150-1において第1の参加者の会議端末160へもたらず。

【0039】

同様に、第2のエンコーダ140-2及び第3のエンコーダ140-3も、それぞれ第2の加算器130-2及び第3の加算器130-3から受信される時間領域の合計のオーディオ信号をエンコードし、エンコード済みのデータを第2の出力150-2及び第3の出力150-3をそれぞれ介してそれぞれの参加者へと送り返す。

30

【0040】

実際のミキシングを実行するために、オーディオ信号が完全にデコードされ、非圧縮の形態で合計される。その後、クリッピング作用(すなわち、許容される値の範囲の超過)を防止するために、それぞれの出力信号を圧縮することによってレベル調節を任意に行うことができる。クリッピングは、単独のサンプル値が許される値の範囲を過ぎて上昇又は下降し、該当の値が切り落とされる(クリップされる)場合に生じうる。例えばCDの場合に使用されているような16ビットの量子化の場合には、サンプル値ごとに、-32768から32767の間の整数値の範囲が利用可能である。

40

【0041】

信号について生じうるオーバーステアリング又はアンダーステアリングに対処するために、圧縮アルゴリズムが使用される。これらのアルゴリズムは、サンプル値を許容可能な値の範囲に保つために、特定のしきい値を超える展開又は特定のしきい値を下回る展開を制限する。

【0042】

図1に示したような会議システム100などの会議システムにおいて、オーディオデータのコーディングを行う場合に、最も容易に実現できる方法にて非エンコード状態でミキシングを実行するために、いくつかの欠点が容認される。さらに、エンコード後のオーデ

50

ィオ信号のデータレートは、送信される周波数のより狭い範囲へとさらに制限される。なぜならば、ナイキスト - シャノンのサンプリング定理によれば、帯域幅が狭いほど、可能なサンプリング周波数が低くなり、したがって少ないデータしか許されないからである。ナイキスト - シャノンのサンプリング定理によれば、サンプリング周波数が、標本化される信号の帯域幅に依存して決まり、帯域幅の（少なくとも）2倍の大きさでなければならない。

#### 【 0 0 4 3 】

国際電気通信連合（ITU）及びその電気通信標準化部門（ITU-T）が、マルチメディア会議システムのためのいくつかの規格を開発済みである。H.320が、ISDN用の標準の会議プロトコルである。H.323は、パケットベースのネットワーク（TCP/IP）用の標準の会議システムを定めている。H.324は、アナログ電話網及び無線電気通信システムのための会議システムを定めている。

10

#### 【 0 0 4 4 】

これらの規格においては、信号の送信だけでなく、オーディオ信号のエンコード及び処理も定められている。会議の運営は、1つ以上のサーバ（いわゆるH.231規格による多地点制御ユニット（MCU））によって担当される。多地点制御ユニットは、複数の参加者のビデオ及びオーディオデータの処理及び配布も担当する。

#### 【 0 0 4 5 】

これを達成するために、多地点制御ユニットは、各々の参加者へと、他のすべての参加者のオーディオデータを含んでいるミックス後の出力信号又は得られた信号を送信し、その信号をそれぞれの参加者へもたらす。図1は、会議システム100のブロック図だけでなく、そのような会議の状況における信号の流れも示している。

20

#### 【 0 0 4 6 】

H.323及びH.320規格の枠組みにおいては、クラスG.7xxのオーディオコーデックがそれぞれの会議システムにおいて機能するように定義されている。規格G.711は、ケーブルでつながれた電話システムにおけるISDN伝送に使用される。8kHzのサンプリング周波数において、G.711規格は、300から3400Hzの間のオーディオ帯域幅をカバーし、8ビットの（量子化）深度において64Kbit/sのビットレートを必要とする。このコーディングは、わずか0.125msというきわめて短い遅延しか生じない $\mu$ -Law又はA-Lawと呼ばれる単純な対数コーディングによって形成される。

30

#### 【 0 0 4 7 】

G.722規格は、より広い50から7000Hzのオーディオ帯域幅を16kHzのサンプリング周波数でエンコードする。結果として、このコーデックは、48、56、又は64Kbit/sのビットレートのより狭帯域のG.7xxオーディオコーデックと比べ、より良好な品質を、1.5msという遅延で実現する。さらに、より低いビットレートでも同等なスピーチ品質を提供する2つのさらなる発展、すなわちG.722.1及びG.722.2が存在する。G.722.2は、25msの遅延において、6.6kbit/sから23.85kbit/sの間のビットレートの選択を可能にしている。

#### 【 0 0 4 8 】

ボイスオーバーIP通信（VoIP）とも称されるIP電話通信の場合においては、G.729規格が典型的に使用される。このコーデックは、スピーチに最適化されており、後の合成のための分析済みのスピーチパラメータの組をエラー信号とともに送信する。結果として、G.729は、G.711規格と比べた場合に、同等のサンプルレート及びオーディオ帯域幅において約8kbit/sの大幅に良好なコーディングを実現する。しかしながら、より複雑なアルゴリズムゆえ、約15msの遅延が生じる。

40

#### 【 0 0 4 9 】

欠点として、G.7.xxコーデックは、スピーチのエンコードに最適化されており、狭い周波数帯域幅の他に、スピーチ付きの音楽又は純粋な音楽のコーディング時に大きな問題を示す。

50

## 【 0 0 5 0 】

したがって、図 1 に示したような会議システム 1 0 0 は、スピーチ信号を伝送及び処理する場合には容認できる品質のために使用することができるが、スピーチに最適化された遅延の少ないコーデックを使用した場合、一般的なオーディオ信号を満足には処理できない。

## 【 0 0 5 1 】

換言すると、例えば音楽を有するオーディオ信号などの一般的なオーディオ信号を処理するために、スピーチ信号のコーディング及びデコーディングのためのコーデックを使用すると、品質に関して満足できる結果はもたらされない。図 1 に示したような会議システム 1 0 0 の枠組みにおいて、一般的なオーディオ信号をエンコード及びデコードするためのオーディオコーデックを使用することで、品質を改善することが可能である。しかしながら、図 2 による文脈においてさらに詳しく概説されるように、そのような会議システムにおいて一般的なオーディオコーデックを使用することは、1 つだけ挙げるのであれば遅延の増加など、さらなる望ましくない影響につながりかねない。

## 【 0 0 5 2 】

しかしながら、図 2 をさらに詳しく説明する前に、本明細書において、それぞれの対象が、或る実施の形態もしくは図において又は複数の実施の形態もしくは図において、2 回以上現れる場合に、そのような対象が同じ又は類似の参照符号で指し示されることに注意すべきである。同じ又は類似の参照符号によって指し示された対象は、そのようでないとき明示的又は黙示的に示されない限りは、例えばそれらの回路、プログラミング、特徴、又は他のパラメータに関して、同様又は同一の方法で実施することが可能である。したがって、図面のいくつかの実施の形態に現れ、同じ又は類似の参照符号によって指し示されている対象は、同じ仕様、パラメータ、及び特徴を有するように実施することが可能である。当然ながら、例えば、境界条件もしくはパラメータが、図から図へと変化し、又は実施の形態から実施の形態へと変化する場合には、別の符号を使用したりそれに適応した符号を使用したりすることも可能である。

## 【 0 0 5 3 】

さらに、以下においては、対象のグループ又は種類（個々の対象ではなくて）を指し示すために、集約的な参照符号が使用される。図 1 の枠組みにおいて、これがすでに行われており、例えば、第 1 の入力を入力 1 1 0 - 1 と称し、第 2 の入力を入力 1 1 0 - 2 と称し、第 3 の入力を入力 1 1 0 - 3 と称する一方で、これらの入力、集約的な参照符号 1 1 0 のみで述べられている。換言すると、そのようでないとき明示的に示されない限りは、集約的な参照符号によって指し示される対象について述べている明細書の部分は、そのような集約的な参照符号に対応する個別の参照符号を有している他の対象にも関係しうる。

## 【 0 0 5 4 】

これは、同じ又は類似の参照符号で指し示された対象についても当てはまるため、両方の処置が、明細書の短縮ならびに明細書に開示の実施の形態のより明確かつ簡潔な様相での説明に役立つ。

## 【 0 0 5 5 】

図 2 は、さらなる会議システム 1 0 0 のブロック図を会議端末 1 6 0 とともに示しており、どちらも図 1 に示した会議システム及び会議端末に類似している。図 2 に示した会議システム 1 0 0 も、図 1 に示した会議システム 1 0 0 と比べて同じように相互接続された入力 1 1 0、デコーダ 1 2 0、加算器 1 3 0、エンコーダ 1 4 0 及び出力 1 5 0 を備えている。図 2 に示した会議端末 1 6 0 も、やはりエンコーダ 1 7 0 及びデコーダ 1 8 0 を備えている。したがって、図 1 に示した会議システム 1 0 0 の説明が参照される。

## 【 0 0 5 6 】

しかしながら、図 2 に示した会議システム 1 0 0 及び図 2 に示した会議端末 1 6 0 は、一般的なオーディオコーデック（コーダ - デコーダ）を使用するように構成されている。結果として、各々のエンコーダ 1 4 0、1 7 0 が、時間 / 周波数変換器 1 9 0 を量子化器

10

20

30

40

50

/コーダ-200の手前に接続してなる直列接続を備えている。時間/周波数変換器190は図2では「T/F」としても示されており、量子化器/コーダ-200は図2では「Q/C」と標記されている。

【0057】

各々のデコーダ120、180は、図2では「Q/C<sup>-1</sup>」と称されているデコーダ/逆量子化器210を、図2では「T/F<sup>-1</sup>」と称されている周波数/時間変換器220に直列に接続して備えている。単に簡潔さの目的のために、時間/周波数変換器190、量子化器/コーダ-200、デコーダ/逆量子化器210及び周波数/時間変換器220が、エンコーダ140-3及びデコーダ120-3の場合においてのみ、そのように標記されている。しかしながら、以下の説明は、他のそのような構成要素にも関する。

10

【0058】

説明をエンコーダ140又はエンコーダ170などのエンコーダから始めると、時間/周波数変換器190へともたらされたオーディオ信号が、変換器190によって時間領域から周波数領域又は周波数関連の領域へと変換される。その後、変換後のオーディオデータが、時間/周波数変換器190によって生成されたスペクトル表現にて、ビットストリームを形成すべく量子化及びエンコードされ、次いでこのビットストリームが、例えばエンコーダ140の場合には、会議システム100の出力150へもたらされる。

【0059】

デコーダ120又はデコーダ180などのデコーダに関しては、デコーダへもたらされたビットストリームが、最初にオーディオ信号の少なくとも一部分のスペクトル表現を形成すべくデコード及び逆量子化され、次いでこれが、周波数/時間変換器220によって再び時間領域へと変換される。

20

【0060】

したがって、時間/周波数変換器190ならびに逆要素である周波数/時間変換器220はそれぞれ、もたらされたオーディオ信号の少なくとも一部分のスペクトル表現を生成するように構成され、かつ、そのスペクトル表現を時間領域のオーディオ信号の該当部分へと再び変換するように構成されている。

【0061】

オーディオ信号を時間領域から周波数領域へ変換し、再び周波数領域から時間領域へと変換するプロセスにおいて、ずれが生じる可能性があり、すなわち再建、再現、又はデコードされたオーディオ信号が元のオーディオ信号又は原始オーディオ信号から相違する可能性がある。量子化エンコーダ200及び再コーダ-210の枠組みにおいて実行される量子化及び逆量子化の追加の工程によって、さらなるアーチファクトが加えられる可能性がある。換言すると、元のオーディオ信号と再生されたオーディオ信号が互いに相違する可能性がある。

30

【0062】

時間/周波数変換器190及び周波数/時間変換器220を、例えば、MDCT（修正離散余弦変換）、MDST（修正離散正弦変換）、FFTベースの変換器（FFT＝高速フーリエ変換）、又は他のフーリエベースの変換器に基づいて実現することができる。量子化器/コーダ-200及びデコーダ/逆量子化器210の枠組みにおける量子化及び逆量子化を、例えば直線量子化、対数量子化、又は他のより複雑な量子化アルゴリズム（例えば、人間の聴覚の特性をより具体的に考慮するなど）に基づいて実現することができる。量子化器/コーダ-200及びデコーダ/逆量子化器210のエンコーダ及びデコーダ部分は、例えば、ハフマンコーディング又はハフマンデコーディングの仕組みを使用することによって機能することができる。

40

【0063】

しかしながら、より複雑な時間/周波数190及び周波数/時間変換器220、ならびにより複雑な量子化器/コーダ-200及びデコーダ/逆量子化器210も、ここに記載されるような種々の実施の形態及びシステムにおいて、例えばエンコーダ140、170としてのAAC-ELDエンコーダ及びデコーダ120、180としてのAAC-ELD

50

デコーダの一部として使用することができ、又はそのようなエンコーダ及びデコーダを形成するものとして使用することができる。

【 0 0 6 4 】

言うまでもないが、会議システム 1 0 0 及び会議端末 1 6 0 の枠組みにおいて、エンコーダ 1 7 0、1 4 0 及びデコーダ 1 8 0、1 2 0 を同一又は少なくとも互換性のあるものとして実現することを推奨できる。

【 0 0 6 5 】

一般的なオーディオ信号のコーディング及びデコーディングの仕組みに基づく図 2 に示したような会議システム 1 0 0 も、オーディオ信号の実際のミキシングを時間領域において実行する。加算器 1 3 0 に、再現された時間領域のオーディオ信号がもたらされ、重畳が実行されて、時間領域のミックス信号が次のエンコーダ 1 4 0 の時間/周波数変換器 1 9 0 へともたらされる。したがって、この会議システムも、やはりデコーダ 1 2 0 及びエンコーダ 1 4 0 の直列接続を備えており、したがって図 1 及び 2 に示したような会議システム 1 0 0 は、典型的に「タンデム・コーディング・システム」と称される。

【 0 0 6 6 】

タンデム・コーディング・システムは、高度な複雑さという欠点を示すことがしばしばである。ミキシングの複雑さは、使用されるデコーダ及びエンコーダの複雑さに強く依存し、いくつかのオーディオ入力及びオーディオ出力信号の場合に大幅に増大しうる。さらに、エンコーディング及びデコーディングの仕組みの大部分がロスのないものではないという事実によって、図 1 及び 2 に示した会議システム 1 0 0 に使用されるようなタンデムコーディングの仕組みは、典型的には、品質への悪影響につながる。

【 0 0 6 7 】

さらなる欠点として、デコーディング及びエンコーディングの繰り返しの工程が、会議システム 1 0 0 の入力 1 1 0 と出力 1 5 0 との間のエンドツーエンド遅延とも称される全体としての遅延も拡大する。使用されるデコーダ及びエンコーダの初期の遅延に応じて、会議システム 1 0 0 そのものが、会議システムの枠組みにおける使用を不安にさせないまでも魅力のないものにし、さらには不可能にさせかねないレベルにまで、遅延を増大させる可能性がある。多くの場合、約 5 0 m s の遅延が、参加者が会話において容認できる最大の遅延であると考えられる。

【 0 0 6 8 】

遅延の主たる原因として、時間/周波数変換器 1 9 0 及び周波数/時間変換器 2 2 0 が会議システム 1 0 0 のエンドツーエンド遅延の原因であり、さらなる遅延が会議端末 1 6 0 によって加わる。さらなる構成要素、すなわち量子化器/コーダー 2 0 0 及びデコーダ/逆量子化器 2 1 0 によって引き起こされる遅延は、これらの部品が時間/周波数変換器 1 9 0 及び周波数/時間変換器 2 2 0 と比べてはるかに高い周波数で動作できるため、あまり重要でない。時間/周波数変換器 1 9 0 及び周波数/時間変換器 2 2 0 の大部分はブロック動作又はフレーム動作であり、すなわち多くの場合に、ブロックのフレーム長を有するバッファ又はメモリを満たすために必要な時間に等しい時間量としての最小遅延を考慮に入れなければならない。しかしながら、この時間が、典型的には数 k H z から数十 k H z の範囲にあるサンプリング周波数によって大きく左右される一方で、量子化器/コーダー 2 0 0 及びデコーダ/逆量子化器 2 1 0 の動作速度は、主として下層のシステムのクロック周波数によって決定される。これは、典型的には、少なくとも 2、3 又は 4 桁以上大きい。

【 0 0 6 9 】

したがって、一般的なオーディオ信号コーデックを使用する会議システムにおいては、いわゆるビットストリームミキシング技術が導入されている。ビットストリームミキシング法は、例えば、上述の欠点の少なくとも一部を回避可能にし、タンデムコーディングによって導入される M P E G - 4 A A C - E L D コーデックに基づいて実現される。

【 0 0 7 0 】

しかしながら、原理的に、図 2 に示したような会議システム 1 0 0 を、G . 7 x x コー

10

20

30

40

50

デック系列の上述したスピーチベースのコードと比べて同様のビットレート及び大幅に広い周波数帯域幅を有するMPEG-4 AAC-ELDコーデックに基づいて実現してもよいことに、注意すべきである。これは、すべての信号種について大幅に良好なオーディオ品質を、大幅に高いビットレートという犠牲を払って達成可能であることも、ただちに意味する。MPEG-4 AAC-ELDは、G.7xxコーデックの遅延の範囲にある遅延を提供するが、これを図2に示したような会議システムの枠組みにおいて実施することは、現実的な会議システム100をもたらさない可能性がある。以下で、図3に関して、上述のいわゆるビットストリームミキシングに基づくより現実的なシステムを概説する。

#### 【0071】

単に簡潔さの目的のために、以下では主としてMPEG-4 AAC-ELDコーデックならびにそのデータストリーム及びビットストリームにのみ注目することに注意すべきである。しかしながら、他のエンコーダ及びデコーダも、図3に例示及び図示されるような会議システム100の環境において使用することができる。

#### 【0072】

図3は、図2の文脈において説明したように、ビットストリームミキシングの原理に従って動作する会議システム100を会議端末160とともに示したブロック図である。会議システム100そのものは、図2に示した会議システム100の簡略版である。より正確には、図2の会議システム100のデコーダ120が、図3に示されているように、デコーダ/逆量子化器210-1、210-2、210-3、・・・によって置き換えられている。換言すると、図2及び3に示した会議システム100を比べたとき、デコーダ120の周波数/時間変換器220が取り除かれている。同様に、図2の会議システム100のエンコーダ140が、量子化器/コーダ200-1、200-2、200-3によって置き換えられている。したがって、図2及び3に示した会議システム100を比べたとき、エンコーダ140の時間/周波数変換器190が取り除かれている。

#### 【0073】

結果として、加算器130は、もはや時間領域で動作するのではなく、周波数/時間変換器220及び時間/周波数変換器190がないため、周波数又は周波数関連の領域で動作する。

#### 【0074】

例えば、MPEG-4 AAC-ELDコーデックの場合には、会議端末160にのみ存在する時間/周波数変換器190及び周波数/時間変換器220がMDCT変換に基づいている。したがって、会議システム100の内部において、ミキサー130が直接的にMDCT周波数表現のオーディオ信号の処理に寄与する。

#### 【0075】

図2に示した会議システム100の場合に、変換器190、220が遅延の主たる原因を呈するため、これらの変換器190、220を取り除くことによって、遅延が大幅に少なくなる。さらに、会議システム100の内部の2つの変換器190、220によって持ち込まれる複雑さも、大幅に軽減される。例えば、MPEG-2 AACデコーダの場合には、周波数/時間変換器220の枠組みにおいて実行される逆MDCT変換が、全体としての複雑さの約20%の原因である。MPEG-4変換器も同様の変換に基づいているため、周波数/時間変換器220だけを会議システム100から取り除くことによって、全体としての複雑さへの無視できぬ寄与を取り除くことが可能である。

#### 【0076】

MDCT領域又は他の周波数領域におけるオーディオ信号のミキシングは、MDCT変換の場合又は同様のフーリエベースの変換の場合に、これらの変換が線形変換であるがゆえに可能である。したがって、変換が、数学的な加算性という特性を有しており、すなわち

$$f(x + y) = f(x) + f(y) \quad (1)$$

10

20

30

40

50

であり、数学的な同次性という性質を有しており、すなわち

$$f(a \cdot x) = a \cdot f(x) \quad (2)$$

であり、ここで  $f(x)$  は変換関数であり、 $x$  及び  $y$  はその適切な引数であり、 $a$  は実数値又は虚数値の定数である。

【0077】

MDC T変換又は他のフーリエベースの変換の両方の特徴が、時間領域におけるミキシングと同様のそれぞれの周波数領域におけるミキシングを可能にしている。したがって、すべての計算を、スペクトル値に基づいて同様に上手く実行することができる。時間領域へのデータの変換は不要である。

10

【0078】

いくつかの状況においては、さらなる条件が満たされなければならないかもしれない。すべての関連のスペクトルデータが、すべての関連のスペクトル成分についてのミキシングプロセスの際に、それらの時間インデックスに関して同じでなければならない。これが、変換の際にいわゆるブロックスイッチング技法が使用され、したがって会議端末160のエンコーダが特定の条件に応じて種々のブロック長の間で自由に切り換わることができる場合には、最終的に満たされない可能性がある。ブロックスイッチングは、ミックスされるべきデータが同じウィンドウで処理されている場合を除き、異なるブロック長及び対応するMDC Tウィンドウ長の間での切り替えゆえに、時間領域において個々のスペクトル値をサンプルへと一意に割り当てることをできなくする可能性がある。分散した会議端末160を有する一般的なシステムにおいては、これが最終的に保証されない可能性があるため、複雑な補間が必要となり、結果としてさらなる遅延及び複雑さが生じる可能性がある。結果として、最終的に、ブロック長の切り替えに基づくビットストリームのミキシングプロセスを実施しないことが推奨されるかもしれない。

20

【0079】

対照的に、AAC - ELDコーデックは、ただ1つのブロック長に基づいており、したがって、ミキシングをより容易に実現できるよう、周波数データの上述の割り当て又は同期をより容易に保証することができる。図3に示した会議システム100は、換言すると、ミキシングを変換領域又は周波数領域において実行することができるシステムである。

30

【0080】

上述のように、図2に示した会議システム100において変換器190、220によって持ち込まれる追加の遅延を除くために、会議端末160において使用されるコーデックは、固定の長さ及び形状のウィンドウを使用する。これは、上述のミキシングプロセスを、オーディオストリームを時間領域へ再変換することなく直接的に実施できるようにする。この手法は、追加で持ち込まれるアルゴリズム的な遅延の大きさを抑えることを可能にする。さらに、デコーダにおける逆変換の工程及びエンコーダにおける順変換の工程が存在しないため、複雑さも低下する。

【0081】

しかしながら、図3に示したような会議システム100の枠組みにおいても、加算器130によるミキシングの後で、オーディオデータの逆量子化が必要になる可能性があり、これがさらなる量子化ノイズを持ち込む可能性がある。この追加の量子化ノイズは、例えば、会議システム100へもたらされる種々のオーディオ信号の種々の量子化工程に起因して生じうる。結果として、例えば量子化の段階の数がすでに制限されているきわめて低いビットレートの伝送の場合に、周波数領域又は変換領域における2つのオーディオ信号のミキシングのプロセスが、生成される信号に望ましくない追加の量のノイズ又は他のひずみを引き起こす可能性がある。

40

【0082】

複数の入力データストリームのミキシングのための装置の形態の本発明による第1の実施の形態を説明する前に、図4に関して、データストリーム又はビットストリームを、そ

50



こに含まれるデータとともに簡単に説明する。

【 0 0 8 3 】

図 4 は、スペクトル領域のオーディオデータの少なくとも 1 つ（多くの場合、2 つ以上）のフレーム 2 6 0 を含んでいるビットストリーム又はデータストリーム 2 5 0 を概略的に示している。より正確には、図 4 が、スペクトル領域のオーディオデータの 3 つのフレーム 2 6 0 - 1、2 6 0 - 2 及び 2 6 0 - 3 を示している。さらに、データストリーム 2 5 0 は、例えばオーディオデータのエンコードの方法を知らせる制御値、他の制御値、又は時間インデックスもしくは他の関連データに関する情報など、付加的情報又は付加的情報のブロック 2 7 0 を含むことができる。当然ながら、図 4 に示したようなデータストリーム 2 5 0 はさらなるフレームをさらに含むことができ、又はフレーム 2 6 0 が、2 チャネル以上のオーディオデータを含んでもよい。例えば、ステレオオーディオ信号の場合に、各々のフレーム 2 6 0 が、例えば左チャンネルからのオーディオデータ、右チャンネルからのオーディオデータ、右及び左チャンネルの両方から導出されたオーディオデータ、又は上述のデータの任意の組み合わせを含むことができる。

10

【 0 0 8 4 】

したがって、図 4 は、データストリーム 2 5 0 が、スペクトル領域のオーディオデータのフレームだけでなく、追加の制御情報、制御値、ステータス値、ステータス情報、プロトコル関連の値（例えば、チェックサム）なども含んでよいことを示している。

【 0 0 8 5 】

図 5 は、例えばデータストリーム 2 5 0 のフレーム 2 6 0 に含まれているようなスペクトル成分に関する（スペクトル）情報を概略的に示している。より正確には、図 5 は、フレーム 2 6 0 のただ 1 つのチャンネルのスペクトル領域の情報の簡単な図を示している。スペクトル領域において、オーディオデータのフレームを、例えば周波数  $f$  の関数としての強度値  $I$  に関して記述することができる。例えばデジタルシステムなどの離散的なシステムにおいては、周波数分解能も離散的であり、したがってスペクトル情報は、典型的には、個々の周波数、狭い帯域又はサブ帯域など、特定のスペクトル成分についてのみ存在する。サブ帯域だけでなく、個々の周波数又は狭い帯域もスペクトル成分と称される。

20

【 0 0 8 6 】

図 5 は、6 個の別々の周波数 3 0 0 - 1、・・・、3 0 0 - 6、及び周波数帯域又はサブ帯域 3 1 0（図 5 に示した事例では、4 つの別々の周波数を含んでいる。）について、強度分布を概略的に示している。個々の周波数又はこれらの周波数に対応する狭い帯域 3 0 0 と、サブ帯域又は周波数帯 3 1 0 との両方が、スペクトル成分を形成しており、このスペクトル成分に関して、フレームがスペクトル領域のオーディオデータに関する情報を含んでいる。

30

【 0 0 8 7 】

サブ帯域 3 1 0 に関する情報は、例えば、全体としての強度又は平均強度値であってよい。振幅、それぞれのスペクトル成分そのもののエネルギー、又はエネルギーもしくは振幅から導出される他の値など、強度又は他のエネルギー関連の値のほかに、位相情報及び他の情報もフレームに含まれることができ、したがって、これらの情報もスペクトル成分に関する情報と考えることができる。

40

【 0 0 8 8 】

本発明による実施の形態の動作原理は、ミキシングが、到来するすべてのストリームが信号の時間領域への逆変換、ミキシング及び再演コーディングを含むデコードを受けるという意味での単刀直入な方法で行われるようなものではない。

【 0 0 8 9 】

本発明による実施の形態は、それぞれのコーデックの周波数領域で行われるミキシングに基づいている。考えられるコーデックは、AAC - ELD コーデック又は一様な変換ウィンドウを有する任意の他のコーデックであってよい。そのような場合、それぞれのデータをミックスできるようにするための時間 / 周波数変換は不要である。本発明の実施の形態による態様は、量子化の刻みのサイズ及び他のパラメータなどのすべてのビットストリ

50

ームパラメータへのアクセスが可能であり、これらのパラメータをミックス済みの出力ビットストリームの生成に使用することができるという事実を利用する。

【0090】

本発明の実施の形態による態様は、スペクトル成分に関するスペクトル線又はスペクトル情報のミキシングを、ソースとなる原始スペクトル線又は原始スペクトル情報の重み付け和によって実行できるという事実を利用する。重み付け係数は、ゼロもしくは1であってよく、又は原理的には、両者の間の任意の値であってよい。ゼロという値は、ソースが無関係として取り扱われ、まったく使用されないことを意味する。帯域又はスケール係数帯域などの線のグループが、本発明による実施の形態の場合に、同じ重み付け係数を使用することができる。しかしながら、すでに示したように、重み付け係数（例えば、ゼロ及び1の分布）を、1つの入力データストリームの1つのフレームの複数のスペクトル成分について変化させることができる。さらに、本発明の実施の形態による態様は、スペクトル情報のミキシング時にゼロ又は1の重み付け係数をもっぱら使用するようには決して要求されない。いくつかの状況下では、入力データストリームのフレームのただ1つではなくて複数の全体的なスペクトル情報について、それぞれの重み付け係数を、ゼロ又は1とは異なるものとすることができる。

10

【0091】

1つの特定の事例は、1つのソース（入力データストリーム510）のすべての帯域又はスペクトル成分が1という係数に設定され、他のソースの係数がすべてゼロに設定される事例である。この場合、1人の参加者の完全な入力ビットストリームが、ミキシング後の最終的なビットストリームとして同一にコピーされる。重み付け係数を、フレーム毎の方法で計算することができるが、フレームの長い方のグループ又は並びに基づいて計算又は決定することも可能である。当然ながら、そのようなフレームの並びの内部又は単一のフレームの内部でも、上述のように、異なるスペクトル成分について重み付け係数を変えてもよい。重み付け係数を、本発明の実施の形態によるいくつかの態様において、心理音響モデルの結果に従って計算又は決定することができる。

20

【0092】

心理音響モデル又は該当のモジュールが、一部の入力ストリームのみが含まれてエネルギー値 $E_f$ をもたらししているミックス信号と、エネルギー値 $E_c$ を有する完全なミックス信号との間のエネルギー比 $r(n)$ を計算することができる。次いで、エネルギー比 $r(n)$ が、 $E_c$ によって除算された $E_f$ の対数の20倍として計算される。

30

【0093】

この比が十分に大きい場合、あまり支配的でないチャンネルが、支配的なチャンネルによってマスクされていると考えることができる。したがって、無関係の削減が処理され、すなわち、まったく顕著でなく、1という重み付け係数に属するストリームだけが含められ、他のすべてのストリーム（1つのスペクトル成分の少なくとも1つのスペクトル情報）が破棄される。換言すると、これらは、ゼロという重み付け係数に属している。

【0094】

より具体的には、これを、

$$E_c = \sum_{n=1}^N E_n \quad (3)$$

40

及び

$$E_{f^{(n)}} = \sum_{\substack{i=1 \\ i \neq n}}^N E_i \quad (4)$$

に従って達成でき、比 $r(n)$ が、

$$r(n) = 20 \cdot \log \frac{E_{f^{(n)}}}{E_c} \quad (5)$$

50

に従って計算され、ここで  $n$  は、入力データストリームの添え字であり、 $N$  は、全入力データストリーム又は関連の入力データストリームの数である。比  $r(n)$  が十分に大きい場合、入力データストリーム 510 のあまり支配的でないチャンネル又はあまり支配的でないフレームが支配的なチャンネル又はフレームによってマスクされていると考えることができる。したがって、無関係の削減を処理することができ、すなわち、ストリームのうちのとにかく顕著なスペクトル成分だけが含められる一方で、他のストリームは破棄される。

【0095】

式(3)から(5)の枠組みにおいて考慮すべきエネルギー値を、例えば、それぞれの強度値の平方を計算することによって、強度値から導出することができる。スペクトル成分に関する情報が他の値を含んでもよい場合には、同様の計算を、フレームに含まれた情報の形態に応じて実行することができる。例えば、複素値情報の場合には、スペクトル成分に関する情報を構成している個々の値の実数部分及び虚数部分の絶対値の計算を、実行しなければならないかもしれない。

10

【0096】

個々の周波数とは別に、式(3)から(5)による心理音響モジュールの適用のために、式(3)及び(4)における合計は、2つ以上の周波数を含むことができる。換言すると、式(3)及び(4)において、それぞれのエネルギー値  $E_n$  を、複数の個々の周波数に対応する全体としてのエネルギー値、すなわち周波数帯のエネルギーによって置き換えることができ、より一般的な言葉にすれば、1つ以上のスペクトル成分に関する1つ又は複数のスペクトル情報で置き換えることができる。

20

【0097】

例えば、AAC-ELDは、人間の聴覚系が同時に取り扱う周波数のグループと同様に、帯域ごとの方法でスペクトル線に作用するため、無関係さの推定又は心理音響モデルを同様の方法で実行することができる。この方法で心理音響モデルを適用することによって、必要であればただ1つの周波数帯域だけの信号の一部を除去又は置換することができる。

【0098】

心理音響的調査が示しているように、信号を他の信号によりマスキングすることは、それぞれの信号の種類に依存する。無関係さの判断のための最小しきい値として、最悪の場合の筋書きを適用することができる。例えば、ノイズを正弦曲線又は他の別個かつ明確な音によってマスキングするためには、21から28 dBの差が典型的には必要である。約28.5 dBのしきい値が良好な置換結果をもたらすことが、試験によって示されている。この値を、検討対象の実際の周波数帯も考慮に入れて、最終的に改善することができる。

30

【0099】

したがって、式(5)による値  $r(n)$  が -28.5 dB よりも大きいことを、検討対象の1つ以上のスペクトル成分に基づく心理音響的評価及び無関係性の評価に関して無関係であると考えることができる。異なるスペクトル成分について異なる値を使用することができる。検討対象のフレームに関する入力データストリームの心理音響的無関係性の指標として、10 dB から40 dB、20 dB から30 dB、あるいは25 dB から30 dBのしきい値を使用することが、有用であると考えられる。

40

【0100】

逆量子化の工程の数が少なくなるがゆえに、タンデムコーディングの影響があまり生じず、あるいはまったく生じないという利点が生じうる。各々の量子化段階が、追加の量子化ノイズの軽減について大きな障害となるため、複数の入力データストリームをミキシングするための装置の形態の本発明による実施の形態を使用することによって、オーディオ信号の全体としての品質を改善することができる。これは、出力データストリームが、決定された入力ストリーム又はその一部のフレームの量子化レベルの分布と比べた量子化レベルの分布が維持されるように生成される場合に当てはまるであろう。

【0101】

50

図6Aは、第1の入力データストリーム510-1及び第2の入力データストリーム510-2のフレームをミキシングするための装置500の簡単なブロック図を示している。装置500は、出力データストリーム530を生成するように構成された処理ユニット520を備えている。より正確には、装置500及び処理ユニット520が、第1の入力データストリーム510-1の第1のフレーム540-1及び第2の入力データストリーム510-2の第2のフレーム540-2に基づいて、出力データストリーム530に含まれる出力フレーム550を生成するように構成されている。

【0102】

第1のフレーム540-1及び第2のフレーム540-2の両者がそれぞれ、第1及び第2のオーディオ信号に関するスペクトル情報を含んでいる。スペクトル情報は、スペクトルの下方部分及び該当のスペクトルの上方部分へと分割され、スペクトルの上方部分は時間/周波数格子分解能でのエネルギー又はエネルギー関連の値に関するSBRデータによって記述されている。スペクトルの下方部分及び上方部分は、SBRパラメータのうちの1つであるいわゆるクロスオーバー周波数において互いに分けられている。スペクトルの下方部分は、それぞれのフレーム540の内側のスペクトル値に関して記述されている。図6Aにおいて、これが、スペクトル情報560の概略表現によって概略的に示されている。スペクトル情報560は、図6Bにおける文脈においてさらに詳しく後述される。

10

【0103】

当然ながら、装置500の形態の本発明による実施の形態を、入力データストリーム510のフレーム540の並びの場合に、類似又は同じ時間インデックスに対応するフレーム540だけが比較及び決定において考慮されるように実施することを推奨できるであろう。

20

【0104】

出力フレーム550も、やはり図6Aに概略的に示されている同様のスペクトル情報表現560を含んでいる。したがって、出力フレーム550も、出力クロスオーバー周波数において互いに接する出力スペクトルの上方部分及び出力スペクトルの下方部分を有する同様のスペクトル情報表現560を含んでいる。入力データストリーム510のフレーム540と同様に、出力フレーム550の出力スペクトルの下方部分も、出力スペクトル値に関して記述される一方で、スペクトルの上方部分(より高い部分)は、出力時間/周波数格子分解能でのエネルギー値を含んでいるSBRデータに関して記述される。

30

【0105】

上述のように、処理ユニット520は、上述のような出力フレームを生成及び出力するように構成されている。一般的な場合において、第1のフレーム540-1の第1のクロスオーバー周波数及び第2のフレーム540-2の第2のクロスオーバー周波数が、異なっていることに注意すべきである。結果として、処理ユニットは、第1のクロスオーバー周波数、第2のクロスオーバー周波数及び出力クロスオーバー周波数のうちの最小の値を下回る周波数に対応する出力スペクトルデータを、第1及び第2のスペクトルデータに基づいてスペクトル領域において直接生成するように構成される。これは、例えば、同じスペクトル成分に対応するそれぞれのスペクトル情報の加算又は線形結合によって達成できる。

40

【0106】

さらに、処理ユニット520は、出力フレーム550の出力スペクトルの上方部分を記述する出力SBRデータを、SBR領域において第1のフレーム540-1の第1のSBRデータ及び第2のフレーム540-2の第1のSBRデータを処理することによって生成するように、さらに構成されている。これは、図9Aから9Eに関してさらに詳しく説明される。

【0107】

やはりさらに詳しく後述されるように、処理ユニット520を、上記定義の最小値と最大値との間の周波数領域について、第1及び第2のスペクトルデータの少なくとも一方からの少なくとも1つのSBR値を評価し、出力SBRデータの対応するSBR値を少なく

50

ともこの評価によるSBR値に基づいて生成するように構成することができる。これは、例えば、検討対象のスペクトル成分の周波数及び考慮事項が、関係する最大のクロスオーバー周波数よりも低い、その最小値よりも高い場合に当てはまるであろう。

【0108】

そのような状況においては、入力フレーム540のうちの少なくとも1つが、それぞれのスペクトルの下方部分の一部としてスペクトル値を含む一方で、それぞれのスペクトル成分が出力クロスオーバー周波数よりも上方にあるために、出力フレームがSBRデータを期待することが生じうる。換言すると、関係するクロスオーバー周波数のうちの最小値と関係するクロスオーバー周波数値のうちの最大値との間のこの中間的な周波数領域において、スペクトルのうちの1つの下方部分からのスペクトルデータに基づいて、対応するSBRデータを推定しなければならないことが生じうる。そのとき、考慮中のスペクトル成分に対応する出力SBRデータは、少なくとも推定されたSBRデータに基づく。これを本発明の実施の形態に従ってどのように実行できるのかについて、さらに詳しい説明が、図9Aから9Eの文脈において後述される。

10

【0109】

他方で、上記定義の中間周波数領域にある関連のスペクトル成分又は周波数について、出力フレーム550がスペクトル値を期待することが、それぞれのスペクトル成分が出力スペクトルの下方部分に属するがゆえに生じうる。しかしながら、入力フレーム540のうちの1つが、該当のスペクトル成分についてSBRデータしか含んでいない可能性がある。この場合、SBRデータにもとづき、さらに随意により考慮対象の入力フレームのスペクトルの下方部分のスペクトル情報又はその少なくとも一部分に基づいて、該当のスペクトル情報を推定することが推奨されうる。換言すると、SBRデータに基づくスペクトルデータの推定も、いくつかの状況のもとで必要になる可能性がある。次いで、推定されたスペクトル値にもとづき、それをスペクトル領域において直接的に処理することによって、それぞれのスペクトル成分の対応するスペクトル値を決定又は入手することができる。

20

【0110】

しかしながら、本発明の実施の形態による装置500の処理及び動作ならびにSBR全般のより良好な理解を促進するために、図6Bが、SBRデータを用いるスペクトル情報のより詳細な表現560を示している。

30

【0111】

本明細書の冒頭部分で概説したとおり、SBRツール又はSBRモジュールは、典型的には、基本的なMP EG-4エンコーダ又はデコーダに隣接する別個のエンコーダ又はデコーダとして機能する。SBRツールは、やはり線形変換を呈する直交ミラーフィルタバンク(QMF)の使用に基づいている。

【0112】

SBRツールは、記述された周波数データの正確なデコーディングを促進するために、MP EGエンコーダのデータストリーム又はビットストリームに、自身の固有の情報及びデータ(SBRパラメータ)を保存する。情報は、SBRツールに関してフレーム格子又は時間/周波数格子分解能として記述される。時間/周波数格子は、現在のフレーム540、550のみに関するデータを含んでいる。

40

【0113】

図6Bは単一のフレーム540、550のそのような時間/周波数格子を概略的に示している。横座標が時間軸であり、縦座標は周波数軸である。

【0114】

自身の周波数 $f$ に関して表示されたスペクトルが、すでに示したように、前もって定められるクロスオーバー周波数( $f_x$ )570によって、下方部分580及び上方部分又はより高い部分590へと分けられる。スペクトルの下方部分580が、典型的には、利用可能な最低の周波数(例えば、0Hz)からクロスオーバー周波数570まで広がり、スペクトルの上方部分590は、クロスオーバー周波数570で始まり、典型的には、図6

50

Bにおいて線600によって示されているようにクロスオーバー周波数の2倍(2f<sub>x</sub>)で終わる。

【0115】

スペクトルの下方部分580は、典型的には、斜線の領域としてのスペクトルデータ又はスペクトル値610によって記述される。なぜならば、多くのフレームベースのコーデック及びそれらの時間/周波数変換器において、オーディオデータのそれぞれのフレームが完全に周波数領域へ変換されており、スペクトルデータ610が、典型的には、明示的なフレーム内部の時間従属性を含まないからである。結果として、スペクトルの下方部分580に関して、スペクトルデータ610を、図6Bに示したそのような時間時間/周波数座標系に完全に正しくは表示できない可能性がある。

10

【0116】

しかしながら、上述のように、SBRツールが、スペクトルの少なくとも上方部分590を、各々のサブ帯域信号が時間従属性又は時間分解能を含んでいる複数のサブ帯域へと分けるQMF時間/周波数変換に基づいて機能する。換言すると、SBRツールによって実行されるようなサブ帯域領域への変換が、「ミックスされた時間及び周波数表現」を生成する。

【0117】

本明細書の冒頭部分で概説したとおり、スペクトルの上方部分590が下方部分580との顕著な類似性を有し、すなわち顕著な相関を有するという仮定にもとづき、SBRツールは、上方部分590のスペクトル成分の周波数へとコピーされるスペクトルの下方部分580のスペクトルデータの振幅の操作を周波数に関して記述するために、エネルギー関連の値又はエネルギー値を導出することができる。したがって、下方部分580からのスペクトル情報を上方部分590の周波数へとコピーし、それらのそれぞれの振幅を修正することによって、スペクトルデータの上方部分590が、ツールの名前によって示唆されるとおり復元される。

20

【0118】

スペクトルの下方部分580の時間分解能が、例えば位相情報又は他のパラメータを含むことによって本質的に存在する一方で、スペクトルの上方部分590のサブ帯域の記述は、時間分解能への直接のアクセスを可能にする。

【0119】

SBRツールは、SBRフレーム長及び基礎をなすエンコーダのフレーム長が互換性をもち、SBRツール及び基礎をなすエンコーダ又はデコーダのどちらもブロック切り替え技法を使用しない場合に、各々のSBRフレームについてフレーム540、550と同一の数の時間スロットを含むSBRパラメータを生成する。この境界条件は、例えば、MPEG-4 AAC-ELDコーデックによって満足される。

30

【0120】

時間スロットは、SBRモジュールのフレーム540、550の時間アクセスを小さな等間隔の時間領域に分割する。各々のSBRフレームにおけるこれらの時間領域の数は、それぞれのフレームのエンコーディングに先立って決定される。MPEG-4 AAC-ELDコーデックの文脈において使用されるSBRツールは、16の時間スロットへと設定される。

40

【0121】

次いで、これらの時間スロットが、1つ以上のエンベロープを形成するために組み合わせられる。エンベロープは、グループへと形成された少なくとも2つ以上の時間スロットを含んでいる。各々のエンベロープが、特定の数のSBR周波数データに組み合わせられている。フレーム格子に、時間スロットに関する数及び長さが、各々のエンベロープとともに保存される。

【0122】

図6Bに示したスペクトル情報560の単純化した表現が、第1のエンベロープ620-1及び第2のエンベロープ620-2を示している。原理的には、エンベロープ620

50

を自由に定めることができ、MPEG-4 AAC-ELDコーデックの枠組みにおいて、2つ未満の時間スロットの長さを有するようにすることもできるが、SBRフレームは、わずか2つの種類、すなわちFIXFIXクラス及びLD\_TRANクラスのいずれかに属する。結果として、原理的にはエンベロープに関する時間スペクトルの任意の分布が可能であるが、以下では、主としてMPEG-4 AAC-ELDコーデックに言及し、その実施を主に説明する。

#### 【0123】

FIXFIXクラスは、16の利用可能な時間スロットを、いくつかの等しい長さのエンベロープ（例えば、それぞれ16個、8個、4個の時間スロットを含んでいる1つ、2つ、4つのエンベロープ）へ分割し、一方、LD\_TRANクラスは、2つ又は3つのエンベロープを含んでおり、そのうちの1つが正確に2つのスロットを含んでいる。正確に2つの時間スロットを含んでいるエンベロープは、オーディオ信号の過渡を含んでおり、換言すると、きわめて大きくて突然な音など、オーディオ信号の急激な変化を含んでいる。この過渡の前後の時間スロットは、それぞれのエンベロープが十分に長い限りにおいて、最大2つのさらなるエンベロープを含むことができる。

10

#### 【0124】

換言すると、SBRモジュールは、フレームをエンベロープへ動的に分割できるため、より精密な周波数分解能をもってオーディオ信号の過渡に反応することができる。現在のフレームに過渡が存在する場合、SBRエンコーダはフレームを適切なエンベロープ構造へと分割する。上述のように、フレーム分割は、SBRとともにAAC-ELDの場合において標準化されており、可変のTRANPOSによって特徴付けられるように時間スロットに関する過渡の位置に依存する。

20

#### 【0125】

過渡が存在する場合にSBRエンコーダによって選択されるSBRフレームクラス、すなわちLD\_TRANクラスは、典型的には3つのエンベロープを含んでいる。最初のエンベロープは、ゼロからTRANPOS-1までの時間スロットインデックスを有するフレームの最初から過渡の位置まで含んでおり、過渡は、TRANPOSからTRANPOS+2までの時間スロットインデックスを有する正確に2つの時間スロットを含むエンベロープによって囲まれている。3番目のエンベロープは、TRANPOS+3からTRANPOS+16までのインデックスを有する以後のすべての時間スロットを含んでいる。しかしながら、SBRによるAAC-ELDコーデックのエンベロープの最小長さは、フレーム境界の近くに過渡を有するフレームが2つのエンベロープへのみ分割されるよう、2つの時間スロットへと制限される。

30

#### 【0126】

図6Bにおいて、2つのエンベロープ620-1、620-2が、等しい長さであり、2つのエンベロープを有するFIXFIX SBRフレームクラスに属している状況が示されている。したがって、各々のエンベロープは、8つの時間スロットの長さを有している。

#### 【0127】

各々のエンベロープに属する周波数分解能は、各々のエンベロープについて計算されそれに関して保存されるべきエネルギー値又はSBRエネルギー値の数を決定する。AAC-ELDコーデックの文脈におけるSBRツールは、高分解能と低分解能の間で切り替えることができる。低い分解能のエンベロープと比べたとき、より高い分解能のエンベロープの場合。低い分解能のエンベロープと比べたとき、より高い分解能のエンベロープの場合には、このエンベロープについてより精密な周波数分解能を可能にするために2倍もの多さのエネルギー値が使用される。高い分解能又は低い分解能のエンベロープの周波数値の数は、ビットレート、サンプリング周波数及び他のパラメータなどのエンコーダパラメータに依存して決まる。MPEG-4 AAC-ELDコーデックの場合には、SBRツールが、きわめて頻繁に、高い分解能のエンベロープにおいて16から14の値を使用する。したがって、低い分解能のエンベロープにおいては、エネルギー値の数は、多くの場

40

50

合に、エンベロープ1つにつき7から8の間の範囲にある。

【0128】

図6Bは、2つのエンベロープ620-1、620-2の各々について、各々が1つのエネルギー又はエネルギー関連のSBR値を表わしている6つの時間/周波数領域630-1a、・・・、630-1f、630-2a、・・・、630-2fを示している。単に簡潔さの目的のために、そのような標記は、2つのエンベロープ620-1、620-2の各々の時間/周波数領域630のうちの3つになされている。さらに、同じ理由のため、2つのエンベロープ620-1、620-2の時間/周波数領域630の周波数分布は、同一に選択されている。当然ながら、これは、多数の可能性のうちの1つの可能性を示しているにすぎない。より正確には、時間/周波数領域630は、各々のエンベロープ620について個別に分布してもよい。したがって、エンベロープ620の間の切り替え時に、スペクトル又はスペクトルの上方部分590を同じ分布へと分割する必要は決してない。また、時間/周波数領域630の数も、やはり上述のように検討対象のエンベロープ620に良好に依存できることに、注意すべきである。

10

【0129】

さらに、追加のSBRデータとして、ノイズ関連のエネルギー値及び正弦曲線関連のエネルギー値も、各々のエンベロープ620に含まれることができる。これらの追加の値は、単に簡潔さの目的のために、図示されていない。ノイズ関連の値は、所定のノイズ源のそれぞれの時間/周波数領域630のエネルギー値に関するエネルギー値を記述し、一方、正弦曲線のエネルギー値は所定の周波数及びそれぞれの時間/周波数領域のそれに等しいエネルギー値を有する正弦振動に関する。典型的には、ノイズ関連又は正弦曲線関連の値のうちの2つ又は3つをエンベロープ620ごとに含むことができる。しかしながら、より少数又はより多数を含むこともできる。

20

【0130】

図7は、図6Aをベースとする本発明の実施の形態による装置500のより詳細なさらなるブロック図を示している。したがって、図6Aの説明が参照される。

【0131】

上述の図6Bのスペクトル情報及び表現560の検討が示しているように、本発明による実施の形態について、出力フレーム550のための新たなフレーム格子を生成するために、最初にフレーム格子を分析することを推奨することができる。結果として、処理ユニット520は、2つの入力データストリーム510-1、510-2がもたらされる分析部640を備えている。処理ユニット520は、入力データストリーム510又は分析部640の出力が接続されるスペクトルミキサー650をさらに備えている。さらに、処理ユニット520は、やはり入力データストリーム510又は分析部640の出力へと接続されるSBRミキサー660をさらに備えている。さらに、処理ユニット520は、やはり2つの入力データストリーム510及び/又は分析部640へと接続され、分析ずみデータ及び/又は入力データストリーム(フレーム540を含む。)を受信する推定部670を備えている。実際の実施例に応じて、推定部670はスペクトルミキサー650又はSBRミキサー660のうちの少なくとも1つへ接続され、それらのうちの少なくとも1つの出力へ、関係するクロスオーバー周波数のうちの最大値と最小値との間の前もって定められた中間領域の周波数について推定したSBR値又は推定したスペクトル値を供給することができる。

30

40

【0132】

SBRミキサー660及びスペクトルミキサー650は、出力フレーム550を含む出力データストリーム530を生成及び出力するミキサー680へと接続される。

【0133】

動作の態様に関しては、分析部640はフレーム540を分析してそこに含まれるフレーム格子を割り出し、新たなフレーム格子(例えばクロスオーバー周波数を含む。)を生成するように構成されている。スペクトルミキサー650が、関係するクロスオーバー周波数のうちの最小値を下回る周波数又はスペクトル成分について、フレーム540のスペ

50



クトル値又はスペクトル情報をスペクトル領域においてミックスするように構成され、一方、SBRミキサー660は、それぞれのSBRデータをSBR領域においてミックスするように同様に構成されている。推定部670は、必要であれば、先に述べた最大値及び最小値の間の中間周波数領域について、2つのミキサー650、660のそれぞれに、これらのミキサーがこの中間周波数領域においても動作できるように、スペクトル領域又はSBR領域の適切なデータを供給する。次いで、ミキサー680が、2つのミキサー650、660から受信されるスペクトルデータ及びSBRデータを編集して出力フレーム550を形成及び生成する。

#### 【0134】

本発明による実施の形態を、例えば、会議システム（例えば、3名以上の参加者を有するテレビ/ビデオ会議システム）の枠組みにおいて使用することができる。そのような会議システムは、時間-周波数変換の工程及び再エンコーディングの工程を省略できるため、時間領域のミキシングに比べて複雑さが少ないという利点を提供することができる。さらに、フィルターバンク遅延が存在しないため、時間領域におけるミキシングに比べて、これらの構成要素によって引き起こされるさらなる遅延が存在しない。

#### 【0135】

しかしながら、本発明による実施の形態を、聴覚雑音置換（PNS）、時間雑音整形（TNS）、及び種々の態様のステレオコーディングなどのモジュールを含むより複雑な用途においても使用することができる。そのような実施の形態が、図8を参照してさらに詳しく説明される。

#### 【0136】

図8は、処理ユニット520を備えている複数の入力データストリームのミキシングのための装置500の概略のブロック図である。より正確には、図8は、入力データストリーム（ビットストリーム）にエンコードされたきわめてさまざまなオーディオ信号を処理することができるきわめて柔軟な装置500を示している。したがって、後述される構成要素のうちの一部は、すべての環境及び本発明によるすべての実施の形態の枠組みにおいて実施される必要は必ずしもない随意による構成要素である。

#### 【0137】

処理ユニット520は、処理ユニット520によって処理されるべき入力データストリーム又はコード済みのオーディオビットストリームの各々について、ビットストリームデコーダ700を備えている。単に簡素化のために、図8には、2つのビットストリームデコーダ700-1、700-2だけが図示されている。当然ながら、処理すべき入力データストリームの数に応じて、より多数のビットストリームデコーダ700を実装することができ、あるいは例えばビットストリームデコーダ700が2つ以上の入力データストリームを順に処理できる場合には、より少数のビットストリームデコーダ700を実装することができる。

#### 【0138】

ビットストリームデコーダ700-1ならびに他のビットストリームデコーダ700-2、・・・の各々は、信号を受信し、受信した信号を処理し、ビットストリームに含まれるデータを分離及び抽出するように構成されたビットストリーム読み取り部710を備えている。例えば、ビットストリーム読み取り部710を、到着するデータを内部クロックに同期させるように構成でき、到着するビットストリームを適切なフレームへと分けるようにさらに構成することができる。

#### 【0139】

さらに、ビットストリームデコーダ700は、ビットストリーム読み取り部710の出力へ接続されてビットストリーム読み取り部710から分離済みのデータを受信するハフマンデコーダ720を備えている。ハフマンデコーダ720の出力が、逆量子化器とも称されるデクオンタイザー730へ接続されている。ハフマンデコーダ720の後方に接続されたデクオンタイザー730に、スケーラー740が続いている。ハフマンデコーダ720、デクオンタイザー730及びスケーラー740が第1のユニット750を形成して

10

20

30

40

50

おり、第1のユニット750の出力において、それぞれの入力データストリームのオーディオ信号の少なくとも一部が、参加者(図8には図示されていない)のエンコーダが機能する周波数領域又は周波数関連領域において入手可能である。

【0140】

さらに、ビットストリームデコーダ700は、データに関して第1のユニット750の後ろに接続された第2のユニット760を備えている。第2のユニット760はステレオデコーダ770(M/Sモジュール)を備えており、ステレオデコーダ770の後ろにPNSデコーダが接続されている。TNSデコーダ790が、データに関してPNSデコーダ780に後続しており、PNSデコーダ780及びステレオデコーダ770とともに第2のユニット760を形成する。

10

【0141】

オーディオデータの上述の流れとは別に、ビットストリームデコーダ700は、制御データに関する種々のモジュール間の複数の接続をさらに備えている。より正確には、ビットストリーム読み取り部710が、適切な制御データを受け取るためにハフマンデコーダ720にも接続されている。さらに、ハフマンデコーダ720は、スケーラー740へスケーリング情報を伝えるために、スケーラー740へ直接に接続されている。ステレオデコーダ770、PNSデコーダ780及びTNSデコーダ790も、それぞれ適切な制御データを受け取るためにビットストリーム読み取り部710へ接続されている。

【0142】

処理ユニット520は、ミキシングユニット800をさらに備えており、次いでミキシングユニット800が、入力に関してビットストリームデコーダ700へ接続されたスペクトルミキサー810を備えている。スペクトルミキサー810は、例えば、周波数領域において実際のミキシングを実行するための1つ以上の加算器を備えることができる。さらに、スペクトルミキサー810は、ビットストリームデコーダ700によってもたらされるスペクトル情報の任意の線形結合を可能にするための乗算器をさらに備えることができる。

20

【0143】

さらに、ミキシングユニット800は、データに関してスペクトルミキサー810の出力へ接続された最適化モジュール820を備えている。しかしながら、最適化モジュール820は、スペクトルミキサー810に制御情報をもたらすために、スペクトルミキサー810にも接続されている。データに関して、最適化モジュール820は、ミキシングユニット800の出力を呈している。

30

【0144】

ミキシングユニット800は、種々のビットストリームデコーダ700のビットストリーム読み取り部710の出力へ直接に接続されたSBRミキサー830をさらに備えている。SBRミキサー830の出力がミキシングユニット800のもう1つの出力を形成している。

【0145】

さらに処理ユニット520は、ミキシングユニット800へ接続されたビットストリームエンコーダ850を備えている。ビットストリームエンコーダ850は、TNSエンコーダ870、PNSエンコーダ880及びステレオエンコーダ890をこの順に直列接続して備えている第3のユニット860を備えている。したがって、第3のユニット860は、ビットストリームデコーダ700の第1のユニット750の逆のユニットを形成している。

40

【0146】

ビットストリームエンコーダ850は、第4のユニット900をさらに備えており、第4のユニット900は、第4のユニットの入力と出力との間で直列接続を形成しているスケーラー910、クオンタイザー920及びハフマンコーダー930を備えている。従って、第4のユニット900は、第1のユニット750の逆のモジュールを形成している。したがって、スケーラー910は、ハフマンコーダー930に該当の制御データをもたら

50

すために、ハフマンコーダ－９３０にも直接に接続されている。

【０１４７】

また、ビットストリームエンコーダ８５０は、ハフマンコーダ－９３０の出力へ接続されたビットストリームライタ９４０を備えている。さらに、ビットストリームライタ９４０は、ＴＮＳエンコーダ８７０、ＰＮＳエンコーダ８８０、ステレオエンコーダ８９０及びハフマンコーダ－９３０から制御データ及び情報を受信するために、これらのモジュールにも接続されている。ビットストリームライタ９４０の出力が処理ユニット５２０及び装置５００の出力を形成している。

【０１４８】

さらに、ビットストリームエンコーダ８５０は、ミキシングユニット８００の出力へ接続された心理音響モジュール９５０を備えている。ビットストリームエンコーダ８５０は、第３のユニット８６０のモジュールへ、例えば第３のユニット８６０のユニットの枠組みにおいてミキシングユニット８００によって出力されるオーディオ信号をエンコードするためにどれを使用できるのかを知らせる適切な制御情報を供給するように構成されている。

10

【０１４９】

したがって、原理的には、第３のユニット８６０の入力までの第２のユニット７６０の出力において、送信側に使用されるエンコーダによって定められるとおり、スペクトル領域のオーディオ信号の処理が可能である。しかしながら、すでに示したように、完全なデコーディング、逆量子化、デスクーリング及びさらなる処理工程は、例えば１つの入力データストリームのフレームのスペクトル情報が支配的である場合に、最終的には必要でないかもしれない。そこで、本発明の実施の形態によれば、それぞれのスペクトル成分のスペクトル情報の少なくとも一部が、出力データストリームの該当のフレームのスペクトル成分へとコピーされる。

20

【０１５０】

そのような処理を可能にするために、装置５００及び処理ユニット５２０は、最適化されたデータ交換のためのさらなる信号線を備えている。図８に示した実施の形態においてそのような処理を可能にするために、ハフマンデコーダ７２０の出力、ならびにスケラ－７４０、ステレオデコーダ７７０及びＰＮＳデコーダ７８０の出力が、他のビットストリーム読み取り部７１０のそれぞれの構成要素とともに、それぞれの処理のためにミキシングユニット８００の最適化モジュール８２０へと接続されている。

30

【０１５１】

それぞれの処理の後のビットストリームエンコーダ８５０の内部の対応するデータの流れを容易にするために、最適化されたデータの流れのための対応するデータ線も実装されている。より正確には、最適化モジュール８２０の出力が、ＰＮＳエンコーダ７８０の入力、ステレオエンコーダ８９０、第４のユニット９００及びスケラ－９１０の入力、ならびにハフマンコーダ９３０への入力へ接続されている。さらに、最適化モジュール８２０の出力がビットストリームライタ９４０へも直接に接続されている。

【０１５２】

すでに示したように、上述のようなモジュールのほぼすべてが、本発明による実施の形態において必ずしも実施する必要がない随意によるモジュールである。例えば、ただ１つのチャンネルしか含んでいないオーディオデータストリームの場合には、ステレオコーディングユニット８９０及びステレオデコーディングユニット７７０を省略することができる。したがって、ＰＮＳベースでない信号を処理すべき場合には、該当のＰＮＳデコーダ７８０及びＰＮＳエンコーダ８８０を省略することができる。ＴＮＳモジュール７９０、８７０も、処理される信号及び出力される信号がＴＮＳデータにもとづかない場合には省略することが可能である。第１のユニット７５０及び第４のユニット９００の内部において、逆量子化器７３０、スケラ－７４０、量子化器９２０及びスケラ－９１０も最終的に省略することが可能である。したがって、これらのモジュールも、随意による構成要素と考えられる。

40

50

## 【 0 1 5 3 】

ハフマンデコーダ 7 2 0 及びハフマンエンコーダ 9 3 0 を、別のアルゴリズムを使用して別の方法で実現してもよく、あるいは完全に省略してもよい。

## 【 0 1 5 4 】

装置 5 0 0 及び装置 5 0 0 に含まれる処理ユニット 5 2 0 の動作の態様に関して、到着する入力データストリームが、最初にビットストリーム読み取り部 7 1 0 によって読み取られ、適切な情報片へと分けられる。ハフマンデコーディングの後、得られたスペクトル情報を、最終的にデクオンタイザ 7 3 0 によって逆量子化し、スケーラ 7 4 0 によって適切にスケーリングすることができる。

## 【 0 1 5 5 】

その後、入力データストリームに含まれる制御情報に依存して、入力データストリーム内にエンコードされたオーディオ信号を、ステレオデコーダ 7 7 0 の枠組みにおける 2 つ以上のチャンネルのオーディオ信号へと分解することができる。例えば、オーディオ信号が中央チャンネル ( M ) 及び横チャンネル ( S ) を含んでいる場合には、対応する左チャンネル及び右チャンネルデータを、中央及び横チャンネルデータを互いに加算及び減算することによって得ることができる。多くの実施例において、中央チャンネルが左チャンネル及び右チャンネルのオーディオデータの合計に比例し、横チャンネルは左チャンネル ( L ) 及び右チャンネル ( R ) の間の差に比例している。実施例に応じて、上述のチャンネルを、クリッピング作用を防止するために係数  $1 / 2$  を考慮しつつ加算及び / 又は減算することができる。一般的に言うと、種々のチャンネルを線形結合によって処理して、それぞれのチャンネルをもたすこと

10

20

## 【 0 1 5 6 】

換言すると、ステレオデコーダ 7 7 0 の後、オーディオデータを、適切であれば、2 つの個別のチャンネルへと分解することができる。当然ながら、逆のデコーディングもステレオデコーダ 7 7 0 によって実行することができる。例えば、ビットストリーム読み取り部 7 1 0 によって受信されるオーディオ信号が左及び右チャンネルを含んでいる場合、ステレオデコーダ 7 7 0 は、適切な中央及び横チャンネルデータを同様に上手く計算又は決定することができる。

## 【 0 1 5 7 】

装置 5 0 0 の実施例だけでなく、それぞれの入力データストリームをもたす参加者のエンコーダの実施例にも応じて、それぞれのデータストリームは、PNS パラメータ ( PNS = 聴覚雑音置換 ) を含むことができる。PNS は、人間の耳が、帯域又は個々の周波数などの限られた周波数範囲又はスペクトル成分のノイズ状の音を、合成的に生成されたノイズからほとんど区別することができないという事実に基づいている。したがって、PNS は、オーディオ信号の実際のノイズ状の寄与を、それぞれのスペクトル成分へ合成的に導入されるべきノイズのレベルを示しており、実際のオーディオ信号を度外視しているエネルギー値で置き換える。換言すると、PNS デコーダ 7 8 0 は、1 つ以上のスペクトル成分において、入力データストリームに含まれる PNS パラメータに基づいて実際のノイズ状のオーディオ信号の寄与を再生することができる。

30

## 【 0 1 5 8 】

TNS デコーダ 7 9 0 及び TNS エンコーダ 8 7 0 に関しては、それぞれのオーディオ信号を、送信側で動作している TNS モジュールに関して、変更されていないバージョンへ再変換しなければならないかもしれない。時間雑音整形 ( TNS ) は、オーディオ信号のフレームにおける過渡状の信号の場合に存在しうる量子化ノイズによって引き起こされるプレエコーアーチファクトを軽減するための手段である。この過渡に対処するために、少なくとも 1 つの適応予測フィルターが、スペクトルの低い側、スペクトルの高い側、又はスペクトルの両側から出発してスペクトル情報へと加えられる。予測フィルターの長さ及び周波数範囲は、それぞれのフィルターの適用先に合わせる事ができる。

40

## 【 0 1 5 9 】

換言すると、TNS モジュールの動作は、1 つ以上の適応 IIR フィルター ( IIR =

50

無限インパルス応答)を計算することに基づいており、予測及び実際のオーディオ信号の間の差を記述する誤差信号を予測フィルターのフィルター係数とともにエンコード及び送信することによる。結果として、残る誤差信号の振幅を減らすために周波数領域において予測フィルターを適用することによって過渡状の信号に対処する(その結果、過渡状のオーディオ信号を直接エンコードする場合に比べて、より少ない量子化の刻みを使用しつつ、同様の量子化ノイズでエンコードすることができる。)ことによって、送信器のデータストリームのビットレートを維持しつつ、オーディオ品質を高めることを可能にできる。

【0160】

TNSの用途に関しては、使用されるコーデックによって決定されるスペクトル領域での「純粋な」表現に達するために入力データストリームのTNS部分をデコードするためにTNSデコーダ790の機能を使用することを、いくつかの環境において推奨できるかもしれない。TNSデコーダ790の機能のこの応用は、心理音響モデル(例えば、心理音響モジュール950において適用される。)の推定をTNSパラメータに含まれる予測フィルターのフィルター係数に基づいて推定することがすでにできない場合に、有用かもしれない。これは、少なくとも1つの入力データストリームがTNSを使用しているが、他の入力データストリームがTNSを使用していない場合に、特に重要かもしれない。

【0161】

処理ユニットが、入力データストリームのフレームの比較にもとづき、TNSを使用している入力データストリームのフレームからのスペクトル情報を使用すべきと判断する場合、TNSパラメータを、出力データのフレームのために使用することができる。もし、例えば互換性がないという理由で、出力データストリームの受け手がTNSデータをデコードできない場合、誤差信号のそれぞれのスペクトルデータ及びさらなるTNSパラメータをコピーせず、TNS関連データから再現されたデータを処理してスペクトル領域の情報を得、TNSエンコーダ870を使用しないことが有用かもしれない。これは、図8に示した構成要素又はモジュールの一部を、本発明によるさまざまな実施の形態において実装する必要が必ずしもないことを、やはり示している。

【0162】

PN Sデータを比較する少なくとも1つのオーディオ入力ストリームの場合にも、同様の方法を当てはめることができる。もし、入力データストリームのスペクトル成分についてのフレームの比較から、1つの入力データストリームが、その現在のフレーム及びそれぞれのスペクトル成分又はスペクトル成分に関して支配的であることが明らかになる場合、それぞれのPN Sパラメータ(すなわち、それぞれのエネルギー値)を、出力フレームのそれぞれのスペクトル成分へ直接コピーしてもよい。しかしながら、もし、受け手がPN Sパラメータを受け付けることができない場合には、スペクトル情報を、それぞれのエネルギー値によって示されるとおりの適切なエネルギーレベルを有するノイズを生成することによって、それぞれのスペクトル成分についてPN Sパラメータから再現することができる。次いで、ノイズデータを、スペクトル領域において相応に処理することができる。

【0163】

すでに概説したように、送信されるデータも、先に説明した機能を実行するSBRミキサー830によって後に処理されるSBRデータを含んでいる。

【0164】

SBRは、2つのステレオチャネルのコーディングを可能にし、左チャネル及び右チャネルを別々にコーディングすること、ならびに左チャネル及び右チャネルを結合チャネル(C)に関してコーディングすることを可能にするため、本発明の実施の形態によれば、それぞれのSBRパラメータ又はその少なくとも一部分の処理は、SBRパラメータのC要素を、決定及び送信すべきSBRパラメータの左及び右の両要素へコピーすること、又はその反対を含むことができる。

【0165】

さらに、本発明の実施の形態による種々の態様において、入力データストリームが、1

10

20

30

40

50

つ又は2つの個別のチャンネルをそれぞれ含むモノラル及びステレオの両方のオーディオ信号を含む可能性があるため、モノラルからステレオへのアップミックス又はステレオからモノラルへのダウンミックスを、入力データストリームのフレームを処理して出力データストリームの出力フレームを生成する枠組みにおいて、さらに実行することができる。

【0166】

先の説明を、TNSパラメータに関して示したので、再量子化を防止するために、それぞれのTNSパラメータを支配的な入力データストリームからのフレーム全体のスペクトル情報とともに出力データストリームへと処理することが望ましいかもしれない。

【0167】

PNSベースのスペクトル情報の場合には、基礎をなすスペクトル成分をデコードすることなく個々のエネルギー値を処理することが、実行可能な方法かもしれない。さらに、この処理による場合には、複数の入力データストリームのフレームの支配的なスペクトル成分からのそれぞれのPNSパラメータだけが、出力データストリームの出力フレームの対応するスペクトル成分に対して、追加の量子化ノイズを持ち込むことなく生じる。

【0168】

上記概説のとおり、本発明による実施の形態は、複数の入力データストリームのフレームを比較した後、かつ比較に基づいて、出力データストリームの出力フレームのスペクトル成分について、正確に1つのデータストリームがスペクトル情報の源であるべきと判断した後で、スペクトル成分に関するスペクトル情報を単純にコピーすることを含むこともできる。

【0169】

心理音響モジュール950の枠組みにおいて実行される置換アルゴリズムが、ただ1つの有効成分を有するスペクトル成分を特定するために、得られる信号の基礎をなすスペクトル成分(例えば、周波数帯域)に関するスペクトル情報の各々を調べる。これらの帯域について、入力ビットストリームのそれぞれの入力データストリームの量子化された値を、特定のスペクトル成分についてそれぞれのスペクトルデータを再エンコード又は再量子化することなくエンコーダからコピーすることができる。いくつかの状況下では、すべての量子化されたデータを、ただ1つの有効な入力信号から取得して、出力ビットストリーム又は出力データストリームを形成することができ、したがって装置500に関して、入力データストリームのロスのないコーディングを実現できる。

【0170】

さらに、エンコーダの内部の心理音響分析などの処理工程を省略することが可能になるかもしれない。これは、基本的に、特定の状況下において1つのビットストリームから他のビットストリームへのデータのコピーだけを実行すればよいため、エンコーディング処理の短縮を可能にし、計算の複雑さの軽減を可能にする。

【0171】

例えば、PNSの場合に、PNSでコードされた帯域のノイズ係数を出力データストリームの1つから出力データストリームへとコピーすることができるため、置換を実行することができる。PNSパラメータがスペクトル成分に特有であり、すなわち換言すると、互いに独立したきわめて良好な近似であるため、個々のスペクトル成分を適切なPNSパラメータで置き換えることが可能である。

【0172】

しかしながら、上述のアルゴリズムの2つの積極的な適用が、聴取体験の低下又は望ましくない品質の低下につながる可能性がある。したがって、置換を、個々のスペクトル成分に関して、スペクトル情報よりもむしろ、個々のフレームに限ることが望ましいかもしれない。そのような動作の態様においては、無関係さの推定又は無関係さの判断、ならびに置換の分析を、不変のままに実行することができる。しかしながら、置換を、この動作の態様において、有効なフレーム内のスペクトル成分のすべて又は少なくともかなりの数が置換可能である場合に限って実行することができる。

【0173】

10

20

30

40

50

この結果、置換の数がより少なくなるかもしれないが、スペクトル情報の内部強度を、いくつかの状況において改善でき、さらにわずかに改善された品質をもたらすことができる。

**【 0 1 7 4 】**

本発明の実施の形態による S B R ミキシングに戻ると、図 8 に示した装置 5 0 0 の追加の随意による構成要素を除外し、S B R の動作原理及び S B R データのミキシングを、以下でさらに詳しく説明する。

**【 0 1 7 5 】**

すでに概説したとおり、S B R ツールは、線形変換を呈する Q M F (直交ミラーフィルタバンク)を使用する。結果として、スペクトルデータ 6 1 0 (図 6 B を参照)を直接的にスペクトル領域において処理できるだけでなく、そのスペクトルの上方部分 5 9 0 (図 6 B を参照)の時間/周波数領域 6 3 0 の各々に関するエネルギー値を処理することもできる。しかしながら、すでに示したように、ミキシングに先立って、関係する時間/周波数格子を最初に調節することが望ましく、いくつかの場合には必要であるかもしれない。

**【 0 1 7 6 】**

原理的には、完全に新しい時間/周波数格子を生成することが可能であるが、以下では、1つのソースにおいて生じている時間/周波数格子が出力フレーム 5 5 0 の時間/周波数格子として使用される状況を説明する。時間/周波数格子のうちのどれを使用することができるかの決定は、例えば心理音響的な考慮に基づくことができる。例えば、格子のうちの1つが過渡を含んでいる場合、この過渡を含んでいるか、又はこの過渡に適合する時間/周波数格子を使用することが望ましいかもしれない。なぜならば、人間の聴覚系のマスキング効果ゆえに、聞き取り可能なアーチファクトが、この特定の格子から外れる場合に、最終的に導入される可能性があるからである。例えば、過渡を有する2つ以上のフレームを本発明の実施の形態による装置 5 0 0 によって処理すべき場合、これらのうちの最先の過渡に適合する時間/周波数格子を選択することが望ましいかもしれない。やはり、マスキング効果ゆえに、最先のアタックを含んでいる格子を選択することが、心理音響的な考慮にもとづき、好ましい選択となりうる。

**【 0 1 7 7 】**

しかしながら、これらの状況のもとでも、他の時間/周波数格子を計算又は選択してもよいことを、指摘しておかなければならない。

**【 0 1 7 8 】**

したがって、S B R フレーム格子をミキシングするとき、いくつかの場合には、フレーム 5 4 0 に含まれる1つ以上の過渡の存在及び位置を分析及び判断することが望ましい。これに加え、あるいはこれに代えて、それぞれのフレーム 5 4 0 の S B R データのフレーム格子を評価し、フレーム格子そのものがそれぞれの過渡に適合しているか否か、あるいはそれぞれの過渡の存在を示しているか否かを検証することによって、これを達成することも可能である。例えば、A A C E L D コーデックの場合に、L D \_ T R A N フレームクラスの使用が、過渡が存在することを示すことができる。このクラスは、T R A N S P O S E 変数も含んでいるため、図 7 に示されるように、時間スロットに関する過渡の位置も、分析部 6 4 0 に知られる。

**【 0 1 7 9 】**

しかしながら、他の S B R フレームクラスである F I X F I X も使用できるため、出力フレーム 5 5 0 の時間/周波数格子を生成するとき、さまざまな群が生じることができる。

**【 0 1 8 0 】**

例えば、過渡を有さないフレーム又は同じ過渡の位置を有するフレームが生じることができる。フレームが過渡を含まない場合、フレーム全体に広がるただ1つのエンベロープだけを有するエンベロープ構造を使用することさえ可能であるかもしれない。また、エンベロープの数が同一である場合に、基本的なフレーム構造をコピーすることが可能である

10

20

30

40

50

。1つのフレームに含まれるエンベロープの数が、他のフレームのエンベロープの数の整数倍である場合には、細かい方のエンベロープ分布を使用することも可能である。

【0181】

同様に、すべてのフレーム540が同じ位置に過渡を含んでいる場合には、時間/周波数格子を、2つの格子のいずれかからコピーすることができる。

【0182】

過渡を有さないただ1つのエンベロープのフレーム及び過渡を有するフレームのミキシングの場合には、過渡を含んでいるフレームのフレーム構造をコピーすることができる。この場合、それぞれのデータをミキシングするときに新たな過渡がもたらされることがないことを、安全に保証することができる。多くの場合、すでに存在する過渡だけが、増幅され、あるいは減衰させられるであろう。

10

【0183】

異なる過渡の位置のフレームが関係する場合には、フレームの各々が、基礎をなす時間スロットに対して異なる位置に過渡を含んでいる。この場合、過渡の位置に基づく適切な分布が望ましい。多くの状況において、最初の過渡の位置が問題とされる。なぜならば、プレエコー効果及び他の問題は、最初の過渡の残効によってマスクされる可能性が高いからである。したがって、この状況においては、フレーム格子を最初の過渡の位置に合わせることが適切かもしれない。

【0184】

フレームに対するエンベロープの分布を決定した後で、個々のエンベロープの周波数分解能を決定することができる。新たなエンベロープの分解能として、典型的には、入力エンベロープの最高の分解能が使用される。例えば、分析されるエンベロープのうちの1つの分解能が高い場合、出力フレームも、自身の周波数に関して高い分解能を有するエンベロープを含む。

20

【0185】

この状況をさらに詳しく説明するために、特に2つの入力データストリーム510-1、510-2の入力フレーム540-1、540-2が異なるクロスオーバー周波数を有する場合に、図9A及び9Bが、2つの入力フレーム540-1、540-2のそれぞれについて、図6Bに示したようなそれぞれの表現を示している。図6Bのきわめて詳細な説明ゆえに、図9A及び9Bの説明を、ここでは省略することができる。さらに、図9Aに示したフレーム540-1は、図6Bに示したフレームと同一である。すでに説明したように、クロスオーバー周波数570の上方に複数の時間/周波数領域630を有している2つの等しい長さのエンベロープ620-1、620-2を含んでいる。

30

【0186】

図9Bに概略的に示したとおりの第2のフレーム540-2は、いくつかの態様に関して、図9Aに示したフレームから相違する。フレーム格子が同じ長さではない3つのエンベロープ620-1、620-2及び620-3を含んでいるという事実の他に、時間/周波数領域630に関する周波数分解能及びクロスオーバー周波数570も、図9Aに示したものと異なっている。図9Bに示した例では、クロスオーバー周波数570が図9Aのフレーム540-1のクロスオーバー周波数よりも高い。結果として、スペクトル590の上方部分が、図9Aに示したフレーム540-1のスペクトルの上方部分よりも相応に大きい。

40

【0187】

図9A及び9Bに示したフレーム540がAAC ELDコーデックによってもたらされているという仮定にもとづき、フレーム540-2のフレーム格子が3つの異なる長さのエンベロープ620を含んでいるという事実が、3つのエンベロープ620のうち2番目が過渡を含んでいるという結論につながる。したがって、2番目のフレーム540-2のフレーム格子が、少なくともその時間における分布に関して、出力フレーム550に選択すべき分解能である。

【0188】

50



しかしながら、図9Cに示されるように、異なるクロスオーバー周波数570がここでは使用されているという事実に起因して、追加の課題が生じる。より具体的には、図9Cは、2つのフレーム540-1、540-2をそれらのスペクトル情報の表現560に関して一緒に示した重ね合わせの状況を図示している。図9Aに示した第1のフレーム540のクロスオーバー周波数570-1(クロスオーバー周波数 $f \times 1$ )及び図9Bに示した第2のフレーム540-2のより高いクロスオーバー周波数570-2(クロスオーバー周波数 $f \times 2$ )だけを考えると、中間の周波数範囲1000については、第1のフレーム540-1からのSBRデータのみが利用可能であり、第2のフレーム540-1からのスペクトルデータ610だけが利用可能である。換言すると、中間周波数範囲1000の範囲内の周波数のスペクトル成分については、ミキシングの手順が、図7に示した推定部670によってもたらされるような推定によるSBR又は推定によるスペクトルデータに頼る。

10

## 【0189】

図9Cに示した状況においては、2つのクロスオーバー周波数570-1、570-2によって周波数に関して囲まれた中間周波数範囲1000が、推定部670及び処理ユニット520が機能する周波数範囲を呈する。この周波数範囲1000においては、SBRデータが第1のフレーム540-1からのみ入手可能であり、一方、この周波数範囲において、第2のフレーム540-2からはスペクトル情報又はスペクトル値のみが入手可能である。結果として、中間周波数範囲1000の周波数又はスペクトル成分が出力のクロスオーバー周波数よりも上か下かに応じて、SBR値又はスペクトル値を評価し、推定された値をフレーム540-1、540-2の一方からの元々の値とSBR領域又はスペクトル領域においてミキシングしなければならない。

20

## 【0190】

図9Dは、出力フレームのクロスオーバー周波数が、2つのクロスオーバー周波数570-1、570-2の低い方に等しい状況を示している。結果として、出力クロスオーバー周波数570-3( $f \times 0$ )は、第1のクロスオーバー周波数570-1( $f \times 1$ )に等しく、これがエンコード後のスペクトルの上方部分を、たった今述べたクロスオーバー周波数の2倍に抑えている。

## 【0191】

先に決定された時間分解能又はそのエンベロープ分布に基づいて時間/周波数格子の周波数分解能をコピー又は再決定することで、中間周波数範囲1000(図9Cを参照)の出力SBRデータが、これらの周波数についての第2のフレーム540-2のスペクトルデータ610から対応するSBRデータを推定することによって決定される。

30

## 【0192】

この推定は、第2のクロスオーバー周波数570-2を上回る周波数についてのSBRデータを考慮して、その周波数範囲の第2のフレーム540-2のスペクトルデータ610に基づいて実行することができる。これは、時間分解能又はエンベロープ分布に関して、第2のクロスオーバー周波数570-2の周囲の周波数が最も等しく影響されている可能性が高いという仮定に基づいている。したがって、中間周波数範囲1000のSBRデータの推定を、例えば、SBRデータによって記述される最も細かい時間及び周波数分解能について、各々のスペクトル成分についてのスペクトル情報に基づいてそれぞれのエネルギー値を計算し、第2のフレーム540-2のSBRデータのエンベロープによって示されるとおりの振幅の時間展開に基づいてそれぞれに減衰又は増幅を加えることによって達成できる。

40

## 【0193】

その後、平滑化フィルター又は他のフィルター処理を加えることによって、推定によるエネルギー値が、出力フレーム550のために決定された時間/周波数格子の時間/周波数領域630へと写像される。図9Dに示したような技術的解決策は、例えば、より低いビットレートにおいて興味深いかもしれない。到着するすべてのストリームのうちの最低のSBRクロスオーバー周波数が、出力フレームのSBRクロスオーバー周波数として

50

使用され、SBRエネルギー値が、コアコーダー（クロスオーバー周波数までで動作する。）とSBRコーダー（クロスオーバー周波数よりも上で動作する。）との間のすき間の周波数範囲1000について、スペクトル情報又はスペクトル係数から推定される。推定は、例えばMDC T（修正離散余弦変換）又はLDF B（少遅延フィルターバンク）スペクトル係数から導出できる幅広くさまざまなスペクトル情報に基づいて実行することができる。さらに、コアコーダーとSBR部分との間のすき間を閉じるために、平滑化フィルターを適用することができる。

【0194】

さらに、この技術的解決策を、例えば64 k b i t / sを含んでいる高いビットレートのストリームを、例えば32 k b i t / sしか含まないより低いビットストリームへと下げるためにも使用できることに、注意すべきである。そのような技術的解決策の実装が望ましいと考えられる状況は、例えば、ミキシングユニットへの低いデータレートの接続（例えば、モデムによるダイヤルイン接続などによって確立される。）しか持たない参加者に、ビットストリームを供給する状況である。

【0195】

異なるクロスオーバー周波数の別の例が、図9Eに示されている。

【0196】

図9Eは、2つのクロスオーバー周波数570-1、570-2の高い方が、出力クロスオーバー周波数570-3として使用される場合を示している。したがって、出力フレーム550は、出力クロスオーバー周波数まではスペクトル情報610を含んでおり、出力クロスオーバー周波数よりも上では、クロスオーバー周波数570-3の典型的には2倍の周波数まで、対応するSBRデータを含んでいる。しかしながら、この状況は、中間周波数範囲1000（図9Cを参照）のスペクトルデータをどのように回復するのかという問題を生じさせる。中間周波数範囲1000の第1のフレーム540-1のSBRデータにもとづき、時間/周波数格子の時間分解能又はエンベロップ分布を決定し、出力クロスオーバー周波数570-3を上回る周波数について時間/周波数格子の周波数分解能を少なくとも部分的にコピー又は決定した後で、スペクトルデータが、処理ユニット520及び推定部670によって推定される。これは、第1のクロスオーバー周波数570-1（図9Aを参照）よりも下のスペクトル情報610の一部又はすべてを随意により考慮しつつ、第1のフレーム540-1の前記周波数範囲1000のSBRデータに基づいてスペクトル情報を部分的に再現することによって達成できる。換言すると、欠けているスペクトル情報の推定を、中間周波数範囲1000の周波数へと少なくとも部分的にSBRデコーダの再現アルゴリズムを適用することによって、SBRデータ及びスペクトルの下方部分580の対応するスペクトル情報からスペクトル情報をスペクトル的に複製することで、達成することができる。

【0197】

例えば周波数領域へと部分的なSBRデコーディング又は再現を適用することによって中間周波数範囲のスペクトル情報を推定した後で、得られた推定によるスペクトル情報を、例えば線形結合を適用することによってスペクトル領域で第2のフレーム540-2のスペクトル情報に直接的にミックスすることができる。

【0198】

クロスオーバー周波数よりも上の周波数又はスペクトル成分のスペクトル情報の再現又は複製は、逆フィルターとも称される。この文脈において、追加の高調波及び追加のノイズエネルギー値も、中間周波数範囲1000の周波数又は成分のそれぞれのスペクトル情報を推定するとき考慮されうることに、注意すべきである。

【0199】

この技術的解決策は、例えば、装置500へと接続された参加者又はミキシングユニットがより高いビットレートを自由に利用できる場合に興味深いかもしれない。パッチ又はコピーアルゴリズムを、スペクトル領域のスペクトル情報（例えば、MDC T又はLDF Bスペクトル係数）へと、これらを該当のクロスオーバー周波数によって分けられたコア

10

20

30

40

50

コーダーとSBR部分との間のすき間を閉じるべく下方の帯域から上方の帯域へとコピーするために適用することができる。これらのコピー係数は、SBRペイロードに格納されたエネルギーパラメータに従って弱められる。

【0200】

図9D及び9Eに示したような両方の筋書きにおいて、最も低いクロスオーバー周波数よりも下方のスペクトル情報を、スペクトル領域において直接的に処理できる一方で、最も高いクロスオーバー周波数を上回るSBRデータは、SBR領域において直接的に処理することができる。典型的には関係するクロスオーバー周波数の最小値の2倍を上回るSBRデータによって記述されるとおりの最高の周波数のうちの最低値を超えるきわめて高い周波数については、出力フレーム550のクロスオーバー周波数に応じて、異なる手法を適用することができる。原理的には、図9Eに示したように、関係するクロスオーバー周波数のうちの最大のクロスオーバー周波数を出力のクロスオーバー周波数570-3として使用する場合、最高の周波数についてのSBRデータは、主として第2のフレーム540-2のみのSBRデータに基づく。さらなる選択肢として、これらの値を、前記クロスオーバー周波数を下回る周波数についてSBRエネルギー値を線形結合させる枠組みにおいて適用される正規化係数又は減衰係数によって弱めることができる。入手可能なクロスオーバー周波数のうちの最小のクロスオーバー周波数が出力のクロスオーバー周波数として使用される図9Dに示したような状況においては、第2のフレーム540-2の該当のSBRデータを無視することができる。

10

【0201】

当然ながら、本発明による実施の形態が、決して2つの入力データストリームに限られず、3つ以上の入力データストリームを含む複数の入力データストリームへと容易に拡張できることに、注意すべきである。そのような場合、上述の手法を、入力データストリームに鑑みて使用される実際のクロスオーバー周波数に応じて、種々の入力データストリームへと容易に適合させることができる。例えば、この入力データストリーム及びその入力データストリームに含まれるフレームのクロスオーバー周波数が、出力フレーム550の出力クロスオーバー周波数よりも高い場合、図9Dの文脈において説明したとおりのアルゴリズムを適用することができる。反対に、該当のクロスオーバー周波数がより低い場合には、図9Eの文脈において説明したアルゴリズム及びプロセスを、この入力データストリームへと適用することができる。SBRデータ又はスペクトル情報の実際のミキシングは、それぞれのデータのうちの2つ以上が合計されるという意味である。

20

30

【0202】

さらには、出力クロスオーバー周波数570-3を任意に選択できることに注意すべきである。出力クロスオーバー周波数が、入力データストリームのクロスオーバー周波数のいずれかと同一である必要は決してない。例えば、図9D及び9Eの文脈において説明したような状況において、クロスオーバー周波数は、入力データストリーム510の両方のクロスオーバー周波数570-1、570-2の間、下方、又は上方にあってよい。出力フレーム550のクロスオーバー周波数を自由に選択できる場合、スペクトルデータ及びSBRデータの推定に関して、上述したアルゴリズムのすべてを実行することが望ましいかもしれない。

40

【0203】

他方で、本発明によるいくつかの実施の形態を、常に最小又は最大のクロスオーバー周波数を使用するように実現してもよい。そのような場合、上述のような機能のすべてを実装する必要はないかもしれない。例えば、常に最低のクロスオーバー周波数が使用される場合には、推定部670は、典型的には、スペクトル情報を推定できる必要はなく、SBRデータだけを推定できればよい。したがって、スペクトルデータを推定する機能を、ここでは最終的に無効にすることができる。対照的に、本発明による実施の形態が、常に最高の出力クロスオーバー周波数を使用するように実現する場合には、SBRデータを推定できるという推定部670の機能は不要であるかもしれず、したがって省略可能である。

50

## 【0204】

本発明による実施の形態は、多チャンネルのダウンミックス又は多チャンネルのアップミックスの構成要素をさらに備えることができ、例えば一部の参加者がステレオ又は他の多チャンネルのストリームを送信でき、一部の参加者がモノラルのストリームしか送信できない場合に、ステレオダウンミックス又はステレオアップミックスの構成要素を備えることができる。この場合、入力データストリームに含まれるチャンネルの数に関して対応するアップミックス又はダウンミックスを実行することが望ましいであろう。アップミキシング又はダウンミキシングによってストリームのいくつかを処理し、入力ストリームのパラメータに一致するミックスされたビットストリームをもたらすことが望ましいであろう。これは、モノラルのストリームを送信する参加者が、応答においてやはりモノラルのストリームを受信することを所望できることを意味する。結果として、他の参加者からのステレオ又は他の多チャンネルオーディオデータを、モノラルのストリームへと変換しなければならないかもしれないが、その逆も然りである。

10

## 【0205】

これは、実施上の制約及び他の境界条件に応じて、例えば、本発明の実施の形態による複数の装置を用意することによって達成でき、あるいはすべての入力データストリームをただ1つの装置に基づいて処理し、到着するデータストリームを装置による処理に先立ってダウンミックス又はアップミックスし、処理後に参加者の端末の要件に一致するようにダウンミックス又はアップミックスすることによって達成できる。

20

## 【0206】

SBRは、ステレオチャンネルのコーディングの2つの態様も可能にする。1つの動作の態様が、左右のチャンネル(LR)を別々に取り扱う一方で、第2の動作の態様は、結合チャンネル(C)について機能する。LRエンコードの要素及びCエンコードの要素をミキシングするためには、LRエンコードの要素をC要素へ写像するか、あるいはその逆を行わなければならない。どちらのコーディング法を使用すべきかの実際の決定は、あらかじめ定めることができ、あるいはエネルギー消費、計算の複雑さ、などの条件を考慮して行うことができ、あるいは別個の処理の妥当性に関する心理音響的な推定に基づくことができる。

## 【0207】

すでに指摘したように、実際のSBRエネルギー関連データのミキシングを、それぞれのエネルギー値の線形結合によってSBR領域において達成することができる。これは、式

30

$$E(n) = \sum_{k=1}^N a_k \cdot E_k(n) \quad (6)$$

に従って達成でき、ここで $a_k$ は、重み係数であり、 $E_k(n)$ は、 $n$ によって示される時間/周波数格子における位置に対応する入力データストリーム $k$ のエネルギー値である。 $E(n)$ は、同じ添え字 $n$ に対応する該当のSBRエネルギー値である。 $N$ は、入力データストリームの数であり、図9A及び9Eに示した例では、2に等しい。

## 【0208】

40

係数 $a_k$ を、正規化ならびに出力フレーム550の各々の時間/周波数領域630及び対応するそれぞれの入力フレーム540の時間/周波数領域630の重なり合いに関する重み付けを実行するために使用することができる。例えば、出力フレーム550の2つの時間/周波数領域630及び該当の入力フレーム540が、出力フレーム550の考慮対象の時間/周波数領域630の50%が入力フレーム540の対応する時間/周波数領域630によって構成されているという意味で、50%の範囲の互いの重なり合いを有する場合、 $0.5 (= 50\%)$ という値を、それぞれのオーディオ入力ストリームとそこに含まれる入力フレーム540との関連性を示す全体としての利得係数によって乗算することができる。

## 【0209】

50

より一般的に表現すると、係数  $a_k$  の各々を、

$$a_k = \sum_{i=1}^M r_{ik} \cdot g \quad (7)$$

によって定義することができ、ここで  $r_{ik}$  は、入力フレーム 540 及び出力フレーム 550 のそれぞれの 2 つの時間 / 周波数領域 630  $i$  及び  $k$  の重なり合いの領域を示す値である。  $M$  は、入力フレーム 540 のすべての時間 / 周波数領域 630 の数であり、  $g$  は、例えばミキシングプロセスが値の許容範囲をオーバーシュート又はアンダーシュートするという結果を防止するために、  $1/N$  に等しくてよい全体的な正規化係数である。係数  $r_{ik}$  は、0 から 1 の間の範囲であってよく、0 は、2 つの時間 / 周波数領域 630 がまったく重なり合わないことを示し、1 という値は、入力フレーム 540 の時間 / 周波数領域 630 が、出力フレーム 550 の該当の時間 / 周波数領域 630 に完全に含まれることを意味する。

10

#### 【0210】

しかしながら、複数の入力フレーム 540 のフレーム格子が等しいことも生じうる。この場合、フレーム格子を、入力フレーム 540 のうちの 1 つから出力フレーム 550 へとコピーすることができる。したがって、関連の SBR エネルギー値のミキシングを、きわめて容易に実行することができる。この場合には、出力値を合計及び正規化することによって、対応するスペクトル情報（例えば、MDC T 値）のミキシングと同様に、対応する周波数値を合計することができる。

20

#### 【0211】

しかしながら、周波数に関して、時間 / 周波数領域 630 の数が、それぞれのエンベロープの分解能に応じて変化する可能性があるため、低いエンベロープの高いエンベロープへの写像、及びその反対を実施することが望ましいかもしれない。

#### 【0212】

図 10 が、これを、例えば 8 つの時間 / 周波数領域 630 -  $l$  及び 16 個の対応する時間周波数領域 630 -  $h$  を含んでいる高いエンベロープについて示している。すでに概説したように、低い分解能のエンベロープは、典型的には、より高い分解能のエンベロープと比べたときに半分の数の周波数データしか含んでおらず、単純な照合を、図 10 に示したとおりに確立することができる。低い分解能のエンベロープを高い分解能のエンベロープへと写像するとき、低い分解能のエンベロープの時間 / 周波数領域 630 -  $l$  の各々が、より高い分解能のエンベロープの 2 つの対応する時間周波数領域 630 -  $h$  へと写像される。

30

#### 【0213】

現実の状況に応じて、例えば正規化に関して、0.5 というさらなる係数を使用することが、ミックス後の SBR エネルギー値のオーバーシュートを防止するために望ましいかもしれない。写像が逆のやりかたで行われる場合には、2 つの隣り合う時間周波数領域 630 -  $h$  を、算術平均を割り出すことによって平均し、低い分解能のエンベロープの 1 つの時間 / 周波数領域 630 -  $l$  を得ることができる。

#### 【0214】

換言すると、式 (7) に関し、最初の状況においては、係数  $r_{ik}$  が 0 又は 1 のいずれかであり、一方、係数  $g$  は 0.5 に等しく、第 2 の場合には、係数  $g$  を 1 に設定でき、一方、係数  $r_{ik}$  が 0 又は 0.5 のいずれかとすることができる。

40

#### 【0215】

しかしながら、係数  $g$  を、ミックスされる入力データストリームの数を考慮する追加の正規化係数を含めることによって、さらに変更しなければならないかもしれない。すべての入力信号のエネルギー値をミックスするために、それらが合計され、スペクトルミキシングの手順の際に加えられる正規化係数によって随意により乗算される。この追加の正規化係数を、式 (7) の係数  $g$  を決定するとき、やはり最終的に考慮に入れなければならないかもしれない。結果として、これが、ベースコーデックのスペクトル係数のスケール

50

係数がSBRエネルギー値の値の許容範囲に一致することを最終的に保証することができる。

【0216】

本発明による実施の形態は、当然ながら、その実施に関してさまざまであってよい。これまでの実施の形態においては、ハフマンデコーディング及びエンコーディングを、単一エントロピーエンコーディングの仕組みとして説明したが、他のエントロピーエンコーディングの仕組みも使用可能である。さらには、エントロピーエンコーダ又はエントロピーデコーダを実装することは、決して必須ではない。したがって、これまでの実施の形態の説明は、主としてACC-ELDコーデックに集中していたが、他のコーデックも、参加者側での入力データストリームの供給及び出力データストリームのデコードに使用することができ、例えば、ブロック長の切り替えを有さないシングルウィンドウに基づく任意のコーデックを使用することが可能である。

10

【0217】

図8に示した実施の形態についての先の説明も示しているとおり、そこで説明されたモジュールは、必須ではない。例えば、本発明の実施の形態による装置を、フレームのスペクトル情報について動作することによって単純に実現することができる。

【0218】

さらに、本発明による実施の形態を、さまざまな異なる方法で実現できることに、注意すべきである。例えば、複数の入力データストリームのミキシングのための装置500及びその処理ユニット520を、抵抗器、トランジスタ、インダクターなどのディスクリートな電気及び電子デバイスに基づいて実現することができる。さらに、本発明による実施の形態を、集積回路のみに基づいて、例えばSOCs(SOC=システム・オン・チップ)、CPU(CPU=中央演算ユニット)及びGPU(GPU=グラフィック処理ユニット)などのプロセッサ、ならびに特定用途向け集積回路(AASIC)などの他の集積回路(IC)の形態で実現することもできる。

20

【0219】

さらに、ディスクリートな実施例の一部又は集積回路の一部である電気デバイスを、本発明の実施の形態による装置の実現の全体において、異なる目的及び異なる機能のために使用することに注意すべきである。当然ながら、集積回路及びディスクリートな回路に基づく回路の組み合わせも、本発明による実施の形態を実現するために使用することができる。

30

【0220】

プロセッサをベースに、本発明による実施の形態を、コンピュータープログラム、ソフトウェアプログラム、又はプロセッサ上で実行されるプログラムに基づいて実現することも可能である。

【0221】

換言すると、本発明の方法の実施の形態の特定の実現の要件に応じて、本発明の方法の実施の形態を、ハードウェア又はソフトウェアにて実現することができる。実現を、電子的に読み取ることができる信号(本発明の方法の実施の形態が実行されるように、プログラマブルなコンピューター又はプロセッサと協働する。)が保存されてなるデジタル記憶媒体(特に、ディスク、CD、又はDVD)を使用して行うことができる。したがって、一般に、本発明の実施の形態は、プログラムコードを機械によって読み取り可能な担体に保存して有しているコンピュータープログラム製品であり、そのようなプログラムコードは、コンピュータープログラム製品がコンピューター又はプロセッサ上で実行されるときに、本発明の方法の実施の形態を実行するように動作することができる。したがって、さらに換言すると、本発明の方法の実施の形態は、コンピュータープログラムに関し、そのようなコンピュータープログラムが、コンピューター又はプロセッサ上で実行されたときに本発明の方法の実施の形態の少なくとも1つを実行するプログラムコードを有している。プロセッサを、コンピューター、チップカード、スマートカード、特定用途向け集積回路、システム・オン・チップ(SOC)又は集積回路(IC)によって形成する

40

50

ことができる。

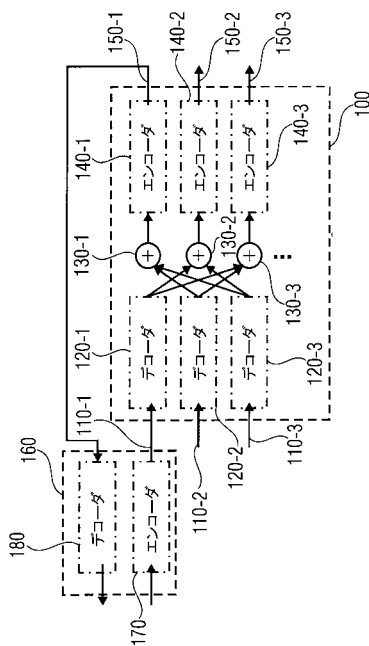
【符号の説明】

【0222】

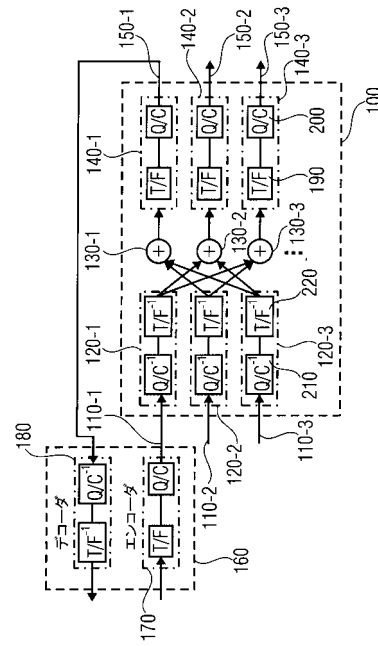
100	会議システム	
110	入力	
120	デコーダ	
130	加算器	
140	エンコーダ	
150	出力	
160	会議端末	10
170	エンコーダ	
180	デコーダ	
190	時間/周波数変換器	
200	量子化器/コーダー	
210	デコーダ/逆量子化器	
220	周波数/時間変換器	
250	データストリーム	
260	フレーム	
270	さらなる情報のブロック	
300	周波数	20
310	周波数帯域	
500	装置	
510	入力データストリーム	
520	処理ユニット	
530	出力データストリーム	
540	フレーム	
550	出力フレーム	
560	スペクトル情報表現	
570	クロスオーバー周波数	
580	スペクトルの下方部分	30
590	スペクトルの上方部分	
600	線	
610	スペクトルデータ	
620	エンベロープ	
630	時間/周波数領域	
640	分析部	
650	スペクトルミキサー	
660	SBRミキサー	
670	推定部	
680	ミキサー	40
700	ビットストリームデコーダ	
710	ビットストリーム読み取り部	
720	ハフマンコーダー	
730	デクオンタイザー	
740	スケーラー	
750	第1のユニット	
760	第2のユニット	
770	ステレオデコーダ	
780	PNSデコーダ	
790	TNSデコーダ	50

- 8 0 0 ミキシングユニット
- 8 1 0 スペクトルミキサー
- 8 2 0 最適化モジュール
- 8 3 0 SBRミキサー
- 8 5 0 ビットストリームエンコーダ
- 8 6 0 第3のユニット
- 8 7 0 TNSエンコーダ
- 8 8 0 PNSエンコーダ
- 8 9 0 ステレオエンコーダ
- 9 0 0 第4のユニット
- 9 1 0 スケーラー
- 9 2 0 量子化器
- 9 3 0 ハフマンコーダー
- 9 4 0 ビットストリームライタ
- 9 5 0 心理音響モジュール
- 1 0 0 0 中間周波数範囲

【図1】

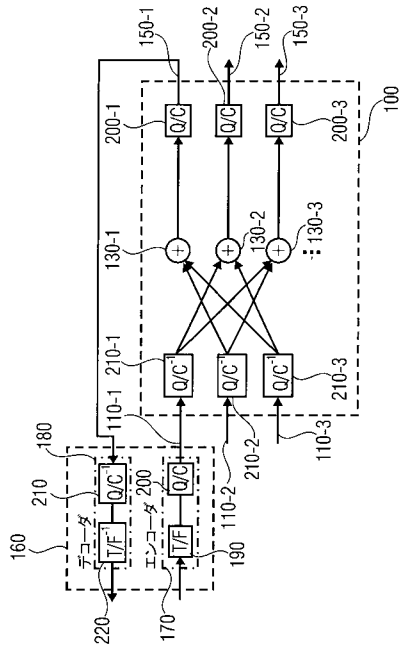


【図2】

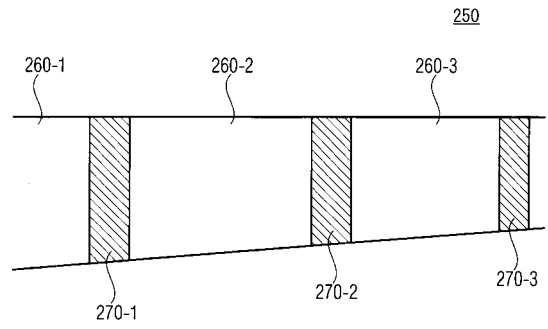




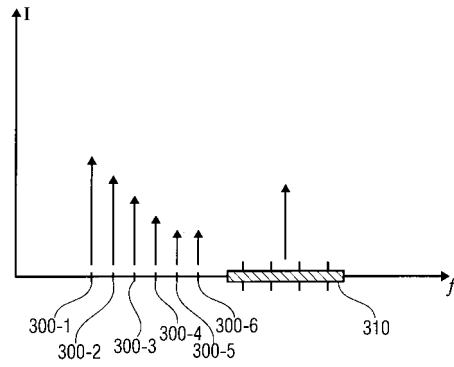
【図3】



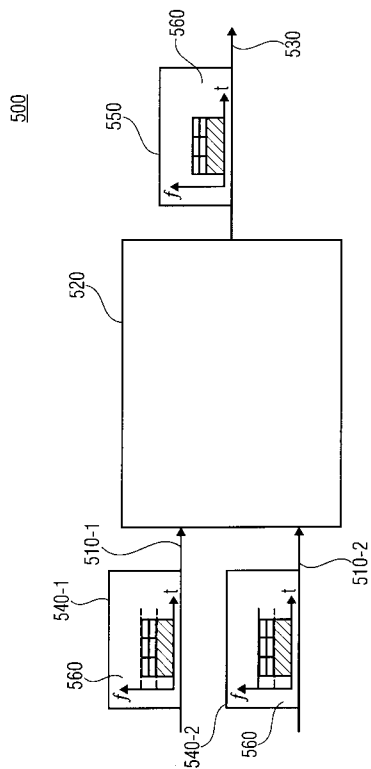
【図4】



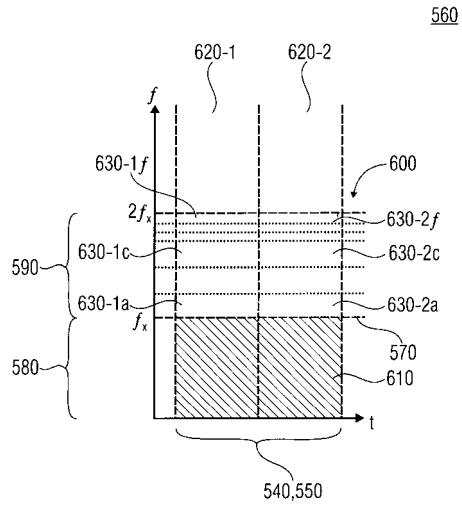
【図5】



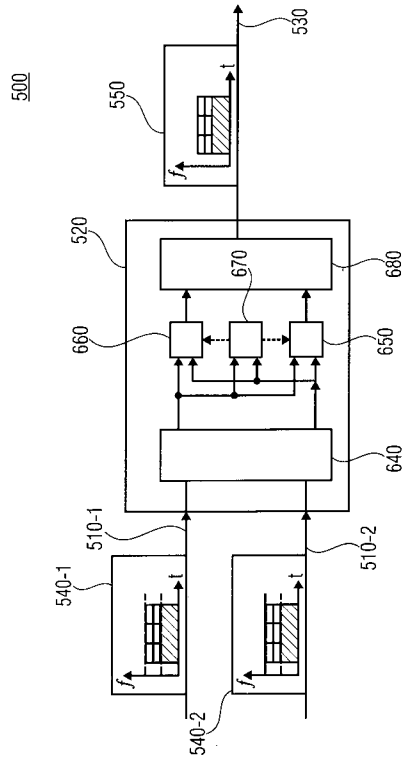
【図6A】



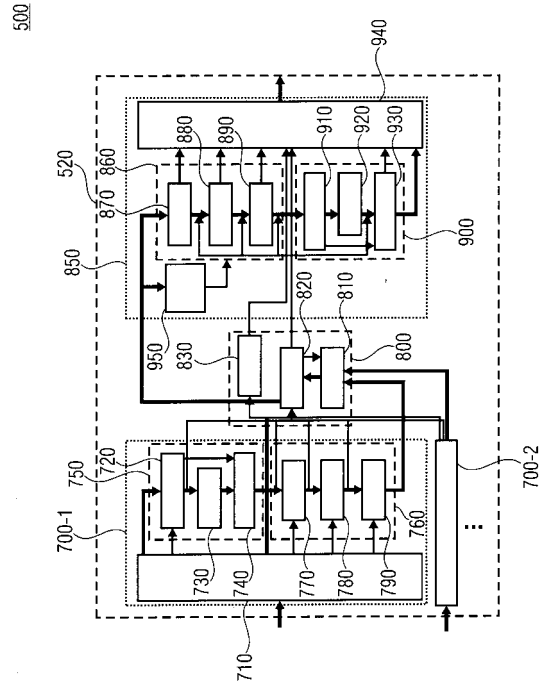
【図6B】



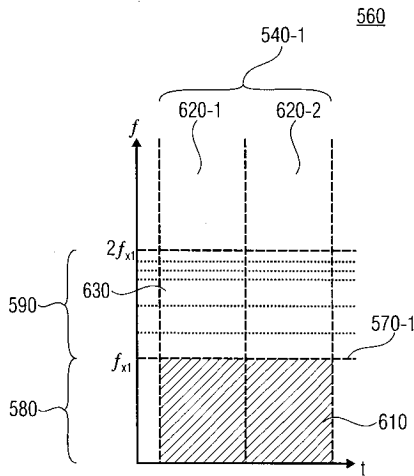
【 図 7 】



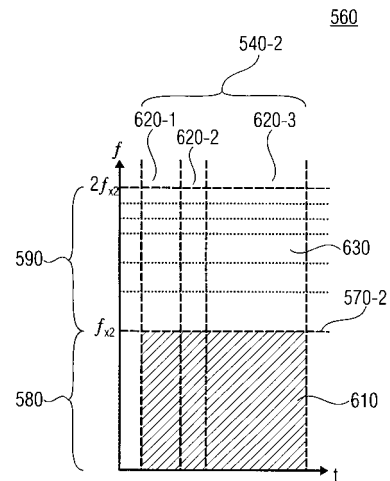
【 図 8 】



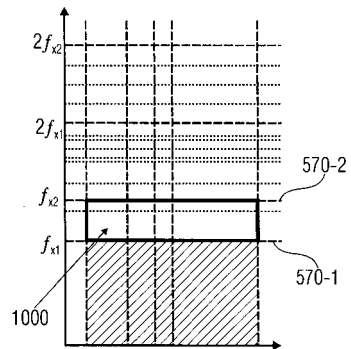
【 図 9 A 】



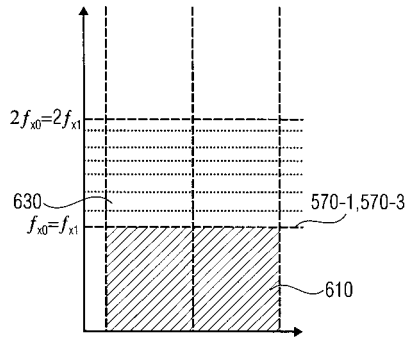
【 図 9 B 】



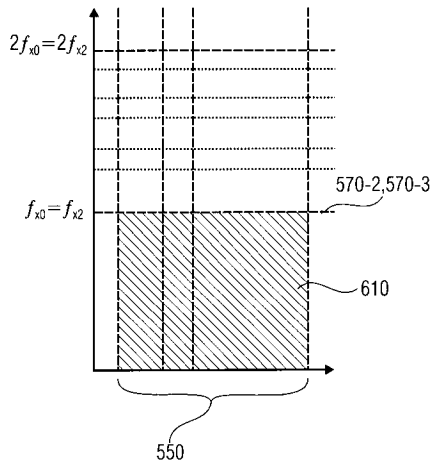
【 図 9 C 】



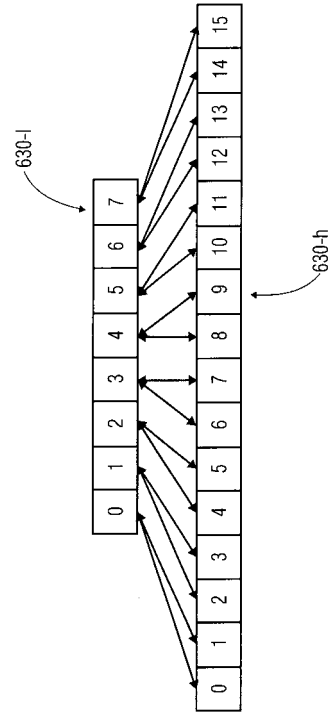
【 図 9 D 】



【 図 9 E 】



【 図 1 0 】



---

フロントページの続き

(72)発明者 マンフレッド・ルツキー

ドイツ連邦共和国 9 0 4 2 7 ニュルンベルグ ハイน์リヒ・フォン・ブレンタノ・シュトラ  
セ 9

(72)発明者 マルクス・ムルツラス

ドイツ連邦共和国 9 0 4 6 9 ニュルンベルグ エツラウブヴェグ 7

審査官 五貫 昭一

(56)参考文献 特表2007-514962(JP,A)

(58)調査した分野(Int.Cl., DB名)

G 1 0 L 1 9 / 0 0

H 0 4 R 3 / 0 0