

(19) 日本国特許庁(JP)

## (12) 特許公報(B2)

(11) 特許番号

特許第5011278号  
(P5011278)

(45) 発行日 平成24年8月29日(2012.8.29)

(24) 登録日 平成24年6月8日(2012.6.8)

(51) Int.Cl.	F 1		
HO4N 7/01	(2006.01)	HO4N 7/01	Z
HO4N 5/262	(2006.01)	HO4N 5/262	
HO4N 5/928	(2006.01)	HO4N 5/92	J

請求項の数 15 (全 23 頁)

(21) 出願番号	特願2008-509406 (P2008-509406)
(86) (22) 出願日	平成18年4月10日 (2006.4.10)
(65) 公表番号	特表2008-541525 (P2008-541525A)
(43) 公表日	平成20年11月20日 (2008.11.20)
(86) 国際出願番号	PCT/EP2006/061488
(87) 国際公開番号	W02006/117281
(87) 国際公開日	平成18年11月9日 (2006.11.9)
審査請求日	平成21年4月7日 (2009.4.7)
(31) 優先権主張番号	05090138.8
(32) 優先日	平成17年5月4日 (2005.5.4)
(33) 優先権主張国	欧洲特許庁 (EP)

(73) 特許権者	501263810 トムソン ライセンシング Thomson Licensing フランス国, 92130 イッシー レ ムーリノー, ル ジヤンヌ ダルク, 1-5 1-5, rue Jeanne d' A rc, 92130 ISSY LES MOULINEAUX, France
(74) 代理人	100070150 弁理士 伊東 忠彦
(74) 代理人	100091214 弁理士 大貫 進介
(74) 代理人	100107766 弁理士 伊東 忠重

最終頁に続く

(54) 【発明の名称】 24 pオーディオ／ビデオデータストリームを更なる50 iフォーマットのデータアイテムで補充することで24 pオーディオ／ビデオデータストリームをオーサリングする方法及び装置

## (57) 【特許請求の範囲】

## 【請求項 1】

ビデオ信号及び1以上のオーディオ信号を表すオーディオ及びビデオデータストリームをオーサリングする方法であって、

前記ビデオ信号は、24 pフォーマットで示されるプログレッシブフォーマット及び24 Hzのビデオフレーム周波数を有し、

前記オーディオ信号は、オリジナルのオーディオ速度で示される前記ビデオ信号の速度と同期した速度を有し、

当該オーサリング方法は、

24 pのフォーマットでビデオデータをエンコードするステップ、60 Hz又は59.94 Hzのフィールド周波数での再生を容易にするために3-2のプルダウンデータアイテムを付加するステップ、及び前記オリジナルのオーディオ速度に従ってオーディオデータをエンコードするステップを含み、

50 iのフォーマットで示される48 Hzと50 Hzとの間のレンジにある平均のフィールド周波数との前記オーディオ及びビデオ信号の同期した表示を達成するため、ビデオフィールド又はフレームの繰り返し及びオーディオフレームのコマ落ちを、デコードされたビデオ及びオーディオ信号の表示ユニットで容易にするため、前記オーディオ及びビデオデータストリームにおけるビデオ及びオーディオのシーンを自動的に判定し、前記50 iのフォーマットで必要とされるビデオ及びオーディオの表示の同期に関する異なるタイプの前記ビデオ及びオーディオのシーンを自動的にカテゴリ化し、オーディオフレームの

10

20

コマ落ち又はビデオフィールドの繰り返しのそれが生じない重大なビデオ及びオーディオのシーンを判定するステップと、

前記シーンのそれぞれ1つについて、オーディオフレームのコマ落ち及び／又はビデオフィールドの繰り返しのそれまでの最大の数を計算するステップと、

重要ではないシーンについて、オーディオフレームコマ落ち及び／又はビデオフィールドの繰り返しの対応する分布を計算し、隣接する重要ではないシーンにおいて対応する数のオーディオフレームのコマ落ち及び／又はビデオフィールドの繰り返しにより、現在の重要なシーンにより導入されたビデオ／オーディオの遅延を補償することができるかを計算し、補償することができる場合、重要ではないシーンについて対応する数のビデオ繰り返し情報アイテム及び対応する数のオーディオコマ落ち情報アイテムを供給し、前記ビデオ繰り返し情報アイテム及び前記オーディオコマ落ち情報アイテムは前記50iのフォーマットの更なるデータアイテムの一部を形成し、補償することができない場合、前記50iのフォーマットの更なるデータアイテムの他の部分を形成するように、前記現在の重要なシーンについてオーディオデータ又はビデオデータを前記50iのフォーマットで再びエンコードするステップと、

前記50iのフォーマットの更なるデータアイテムを、前記オーディオ及びビデオデータストリームのデータフィールド又は前記オーディオ／ビデオデータストリームのコピーに組み込むステップと、

を含むことを特徴とする方法。

#### 【請求項2】

前記50iのフォーマットの更なるデータアイテムについて、サブタイトルユニットが前記ビデオ又はオーディオデータに揃えられる、

請求項1記載の方法。

#### 【請求項3】

前記カテゴリのタイプは、  
シーンカットが現在存在するか否か、  
オーディオ及びビデオのリップシンクが現在必要とされるか否か、  
少なくとも1つのビデオに関連するタイプ及び少なくとも1つのオーディオに関連するタイプを含み、

前記少なくとも1つのビデオに関連するタイプ及び少なくとも1つのオーディオに関連するタイプは、

現在のビデオコンテンツにおいて動きがあるか否か、  
非常に暗いシーン又は非常に明るいシーンが現在存在するか否か、  
音が現在存在するか否か、  
低いボリューム、又は聞くことのできる雑音、又は音楽なし、或いは一定の周波数及びボリュームが現在存在するか否か、

聞くことのできる雑音なしに調和音の現在の変化が存在するか否か、から選択される、  

請求項1又は2記載の方法。

#### 【請求項4】

前記50iのフォーマットで必要とされるビデオ及びオーディオの表示の同期に関する異なるタイプの前記ビデオ及びオーディオのシーンを自動的にカテゴリ化することは、自動的に発見された候補となるリップシンクのシーンがリップシンクを実際に必要であるか否かに関する確認に応答して実行される、

請求項3記載の方法。

#### 【請求項5】

動きのないシーンタイプで、できるだけ多数の静止画像のフィールドは、ビデオをオーディオと同期させるために繰り返され、

消音されたシーンタイプで、できるだけ多数の消音されたオーディオフレームは、オーディオをビデオと同期させるためにコマ落ちされ、

シーンカットのシーンタイプで、できるだけ多数の最初と最後のシーンカットのビデオ

10

20

30

40

50

フィールドは、オーディオとビデオを同期するために繰り返される、  
請求項 1 乃至 4 の何れか記載の方法。

**【請求項 6】**

前記 50i のフォーマットの更なるデータアイテムは、動きのないシーンタイプ又は低速の動きのシーンタイプで、ビデオフィールドは 24 : 1 のプルダウンを使用することで繰り返され、及び / 又は低いボリュームのシーンで、24 のフレームおきにオーディオフレームがコマ落ちされる、

請求項 1 乃至 5 の何れか記載の方法。

**【請求項 7】**

前記 50i のフォーマットの更なるデータアイテムは、ナビゲーションデータパケットに挿入され、及び / 又は前記オーディオ / ビデオデータストリームの.info ファイルに挿入される、

請求項 1 乃至 6 の何れか記載の方法。

**【請求項 8】**

コンピュータに、請求項 1 乃至 7 の何れか記載の方法を実行させる命令を含むプログラムを記録したコンピュータ読み取り可能な記録媒体。

**【請求項 9】**

ビデオ信号及び 1 以上のオーディオ信号を表すオーディオ及びビデオデータストリームをオーサリングする装置であって、

前記ビデオ信号は、24p フォーマットで示されるプログレッシブフォーマット及び 24 Hz のビデオフレーム周波数を有し、

前記オーディオ信号は、オリジナルのオーディオ速度で示される前記ビデオ信号の速度と同期した速度を有し、

当該オーサリング装置は、

24p のフォーマットでビデオデータをエンコードし、60 Hz 又は 59.94 Hz のフィールド周波数での再生を容易にするために 3 - 2 のプルダウンデータアイテムを付加し、前記オリジナルのオーディオ速度に従ってオーディオデータをエンコードする手段を有し、

当該オーサリング装置は、

50i のフォーマットで示される 48 Hz と 50 Hz との間のレンジにある平均のフィールド周波数との前記オーディオ及びビデオ信号の同期した表示を達成するため、ビデオフィールド又はフレームの繰り返し及びオーディオフレームのコマ落ちを、デコードされたビデオ及びオーディオ信号の表示ユニットで容易にするため、前記オーディオ及びビデオデータストリームにおけるビデオ及びオーディオのシーンを自動的に判定し、前記 50i のフォーマットで必要とされるビデオ及びオーディオの表示の同期に関する異なるタイプの前記ビデオ及びオーディオのシーンを自動的にカテゴリ化し、オーディオフレームのコマ落ち又はビデオフィールドの繰り返しのそれぞれが生じない重大なビデオ及びオーディオのシーンを判定する手段と、

前記シーンのそれぞれ 1 つについて、オーディオフレームのコマ落ち及び / 又はビデオフィールドの繰り返しのそれぞれの最大の数を計算し、重要ではないシーンについて、オーディオフレームのコマ落ち及び / 又はビデオフィールドの繰り返しの対応する分布を計算し、隣接する重要ではないシーンにおいて対応する数のオーディオフレームのコマ落ち及び / 又はビデオフィールドの繰り返しにより、現在の重要なシーンにより導入されたビデオ / オーディオの遅延を補償することができるかを計算する手段と、

補償することができる場合、重要ではないシーンについて対応する数のビデオ繰り返し情報アイテム及び対応する数のオーディオコマ落ち情報アイテムを供給する手段と、前記ビデオ繰り返し情報アイテム及び前記オーディオコマ落ち情報アイテムは前記 50i のフォーマットの更なるデータアイテムの一部を形成し、前記供給する手段は、補償することができない場合、前記 50i のフォーマットの更なるデータアイテムの他の部分を形成するように、前記現在の重要なシーンについてオーディオデータ又はビデオデータを前記 5

10

20

30

40

50

0 i のフォーマットで再びエンコードし、

前記 50 i のフォーマットの更なるデータアイテムを、前記オーディオ及びビデオデータストリームのデータフィールド又は前記オーディオ / ビデオデータストリームのコピーに組み込む手段とを更に有する、

ことを特徴とする装置。

【請求項 10】

前記 50 i のフォーマットの更なるデータアイテムについて、サブタイトルユニットが前記ビデオ又はオーディオデータに挿えられる、

請求項 9 記載の装置。

【請求項 11】

前記カテゴリのタイプは、

シーンカットが現在存在するか否か、

オーディオ及びビデオのリップシンクが現在必要とされるか否か、

少なくとも 1 つのビデオに関連するタイプ及び少なくとも 1 つのオーディオに関連するタイプを含み、

前記少なくとも 1 つのビデオに関連するタイプ及び少なくとも 1 つのオーディオに関連するタイプは、

現在のビデオコンテンツにおいて動きがあるか否か、

非常に暗いシーン又は非常に明るいシーンが現在存在するか否か、

音が現在存在するか否か、

低いボリューム、又は聞くことのできる雑音、又は音楽なし、又は一定の周波数及びボリュームが現在存在するか否か、

聞くことのできる雑音なしに調和音の現在の変化が存在するか否か、から選択される、  
請求項 9 又は 10 記載の装置。

【請求項 12】

前記 50 i のフォーマットで必要とされるビデオ及びオーディオの表示の同期に関する異なるタイプの前記ビデオ及びオーディオのシーンを自動的にカテゴリ化することは、自動的に発見された候補となるリップシンクのシーンがリップシンクを実際に必要であるか否かに関する確認に応答して実行される、

請求項 11 記載の装置。

【請求項 13】

前記判定する手段は、

動きのないシーンタイプで、できるだけ多数の静止画像のフィールドを、ビデオをオーディオと同期させるために繰り返し、

消音されたシーンタイプで、できるだけ多数の消音されたオーディオフレームを、オーディオをビデオと同期させるためにコマ落ちさせ、

シーンカットのシーンタイプで、できるだけ多数の最初と最後のシーンカットのビデオフィールドを、オーディオとビデオを同期させるために繰り返す、

請求項 9 乃至 12 の何れか記載の装置。

【請求項 14】

前記 50 i のフォーマットの更なるデータアイテムは、動きのないシーンタイプ又は低速の動きのシーンタイプで、ビデオフィールドは 24 : 1 のプルダウンを使用することで繰り返され、及び / 又は低いボリュームのシーンで、24 のフレームおきにオーディオフレームがコマ落ちされる、

請求項 9 乃至 13 の何れか記載の装置。

【請求項 15】

前記 50 i のフォーマットの更なるデータアイテムは、ナビゲーションデータパケットに挿入され、及び / 又は前記オーディオ / ビデオデータストリームの.info ファイルに挿入される、

請求項 9 乃至 14 の何れか記載の装置。

10

20

30

40

50

**【発明の詳細な説明】****【技術分野】****【0001】**

本発明は、24p(24Hz)オーディオ/ビデオデータストリームを更なる50iフォーマットのデータアイテムで補充することで24pオーディオ/ビデオデータストリームをオーサリングし、これによりそのデータストリームの24p及び50iフォーマットの再生を可能にするする方法及び装置に関する。

**【背景技術】****【0002】**

映画は、35mmフィルムのための世界的な規格に従い、24Hzフレーム周波数で、10 フィルムシアターにおける投影のためにプログレッシブスキャニングフォーマット(24p)で製作される。投影速度は、数十年以来、毎秒24フルピクチャに固定されている。同じマテリアルの光学トラックに典型的に存在する、付随する音声信号は、同じ速度で再生される。係る再生は、世界における実際の既存のフィルムシアターで実行される。係る映画がたとえばブロードキャストによるといったテレビジョンシステムを介して再生されるとき、状況は異なる。世界における主要なTVシステムは、インターレーススキャニング及び、50Hzフィールド周波数(PAL及びSECAMについて、たとえば欧洲、中国及び豪州において50iで示される)又は60Hzフィールド周波数(NTSCについて、たとえば米国及び日本国において60iで示される)の何れかを使用する。

**【0003】**

60iの国では、TVシステムは、30のフルビデオフレームが毎秒当たり表示されるように構築される。映画の再生は、3:2プルダウンを使用して2又は3のビデオフィールドで交互に1つのフィルムフレームを表示することで、基本的な正しい全体のスピードで実行される。NTSCカラーTVシステムでは、実際のフィールドレートは59.94Hzであり、これは、映画が毎秒23.976フレームで実際にブロードキャストされることを意味し、すなわち1000/1001のファクタによりオリジナルの速度よりも遅い。これは、TV視聴者により一般に知覚されない。20

**【0004】**

50Hzの国では、状況は更に困難である。最も簡単なソリューションは、毎秒25フレームで、すなわち4.2%速く映画を放送することである。この方法により、音声は4.2%高い(ほぼ2分の1トーン高い)ピッチを有するが、典型的に、消費者からの不平はない。映画がAVメディアに記録された場合(たとえばVCRビデオ、CDI及びDVD)、及び係るメディアが表示装置としてTV受像機を使用して再生された場合、一般に同じ原理が当てはまる。メディアに存在する随伴する音声が(たとえばリニアPCM又はAC-3符号化又はDTS符号化された)デジタル形式である場合、再生装置は、オーディオ情報を、一般に問題とはならない4.2%高い速度でデコードして表示する必要があるか、50Hzの国で販売されているDVDビデオ又はCDIディスクは、ビデオ信号が4.2%高い速度で出力される一方で、プレーヤのオーディオデコーダはオーディオ信号の正しい速度及びピッチを自動的に出力するように、既に符号化されているオーディオデータストリームを含む必要がある。30

**【0005】**

要するに、シネマで使用される毎秒24プログレッシブフレーム(24p)、毎秒50インターレースフィールド(50i)及び毎秒60インターレースフィールドといった、AVストリーム(オーディオビデオストリーム)について3つの確立された再生速度が現在存在する。40

**【0006】**

代替は、プレーヤにおけるオーディオデコーディング、ピッチ変換及びオーディオ再符号化を実行することである。しかし、これは、更なるオーディオエンコーディングのために、非常に高価なプレーヤハードウェア及びライセンスコストを必要とする。したがって、オーディオの再符号化は、オーサリングの間に実行される必要がある。これは、様々な50

圧縮されたオーディオストリームの再符号化のために多数のプロセッサパワー（すなわち高価なハードウェア）及び特別なライセンス及びソフトウェアを必要とするためである。したがって、必要とされるオーディオの再符号化は、同じ映画について2つのマスタを作製するための現実の理由である。

#### 【0007】

ディスク製作者が2を超える異なる速度についてレコードを製作した場合、ある速度から他の速度のうちの1つへの変換は最初の選択である。それは、新たな速度で再び映画を製作するために余りに多くの時間とお金を消費するため、オリジナルの映画は、再び使用されず、すなわちオリジナルのレコードは、新たな速度にトランスコードされる。通常、映画製作者は、24pの映画を60iのDVDビデオ及び50iDVDビデオにトランスコードする。10 現代のTV受像機は、同じカラーシステムPAL、NTSC又はSECAMを維持しつつ、異なる再生速度を扱うことができる。たとえば、欧州で現在販売されている多くのTVセットは、再生された60iレコードを表示することができる。しかし、多くの旧式のTVセットは、かかる速度シフトを有するビデオ入力信号を処理することができない。さらに、現代の100Hz TVセットは、60i入力データを受信したときに誤動作する。幾つかの旧式の装置は、カラーなしでカラー60iコンテンツを再生し、他の装置は、安定のピクチャでさえ再生することができない。係る理由のため、たとえばDVDビデオフォーマットは、それらの好適な／サポートされた速度によるディスクを取得するためには50iの国と60iの国との間で区別する。

#### 【0008】

なお、（アナログ）カラーシステムは、それが記録された（MPEG）ビットストリームの一部ではないために問題ではない。適切なカラーシステムは、ビデオコンテンツの伸張の後に適用されるか、エンコードされる。20

#### 【0009】

上述されたように、24pから60iへのトランスコードのために3:2プルダウンが使用される。60iの国における消費者は、フィールドの繰り返しによる僅かにジャーキングするビデオ再生のような結果的な問題を知り、広く受け入れている。30 4.2%の高速の再生速度は、24pから50iへのトランスコードのために使用されている。50iの国における大部分の消費者は、50iのフリッカ及び増加するピッチのような結果的な問題を受け入れている。幾つかの問題は、たとえば洗練された動き補償により100Hz TVセットを使用することで、既に低減されるか、更に除去されている。しかし、全体の再生時間は、シネマにおけるよりも4.2%短い。係る速度の利点は、ビデオ再生がジャークを有さないが、オーディオストリームが高速の再生を実行するためにトランスコードされる必要があることである。したがって、同じ映画のオーディオストリームは、50iディスクと60iディスクの間で異なるが、（MPEG）ビデオストリームは、非常に類似している。ビデオストリームの主要な違いは、60iのビデオストリームが更なる3:2プルダウンフラッグ及び異なるタイムスタンプを含むことである。MPEGは、たとえばMPEG-2ビデオ、MPEG-4ビデオ又はMPEG-4 AVCを意味する。

#### 【0010】

消費者は、それらの国で使用されるトランスコード方法の問題を受け入れるが、他の国の異なる問題を好みない。60iの消費者は、50iのTVセットのフリッカが好きではない（理論的に、これは、100Hz技術を適用することで50iに互換性があるTVセットで現在の60iのTVセットを置き換えることで、又は低速のLCDディスプレイを使用することで克服される）。40

#### 【0011】

50iの消費者は、60iビデオのジャークを好みない（理論的に、これは現在のディスク、プレーヤ及びTVセットを24pの互換性のあるディスク、プレーヤ及びTVセットで置き換えることで克服することができる）。

#### 【0012】

現在のDVDプレーヤは、24pの再生をサポートしていない。圧縮されたデジタルオ50

ーディオは、所望の再生速度について常に符号化される。これは、プレーヤが、圧縮されていないオーディオストリームを外部のデコーダに直接に送出するためである。

#### 【0013】

従来のオーサリング、たとえばD V Dビデオのオーサリングは、通常、2つの主要な手順で実行される。60iの国について、24pフォーマットでビデオをエンコードし、60i再生のために3-2プルダウン機能を追加する。多様なデコーダ及び言語についてデジタルオーディオをエンコードする。サブタイトル(D V Dではサブピクチャ)をエンコードする。メニューを発生する。マスタを作成する。

#### 【0014】

50iの国について、エンコードされたビデオをリストアンプし、3-2プルダウンを除く。たとえばピッチ変換を使用してデジタルオーディオストリームを再エンコードする。サブタイトル(D V Dではサブピクチャ)をエンコードする。メニューを調整する。第二のマスタを作成する。60iディスクは、50iの国における多数のT Vセットで再生可能ではない。50iディスクは、60iの国における大部分のT Vセットで再生可能ではない。

10

#### 【0015】

E P 0 5 0 9 0 0 2 0 . 8は、24pディスクを記述しており、50iの表示のために再生することができ、これにより、ビデオフィールドが適応的に繰り返され、オーディオフレームが適応的にドロップされて、記録された番組の表示に応じてビデオ速度に整合するオーディオ速度を平均して達成する。

20

#### 【発明の開示】

#### 【発明が解決しようとする課題】

#### 【0016】

コンテンツプロバイダは、世界の大部分で再生することができるシングルピクチャ周波数(single-picture-frequency)、シングルオーディオ速度(single-audio-speed)のA Vディスクを提供することを好む。勿論、これは、多様な言語でオーディオを提供する必要のため、及びビデオと共に全てのオーディオ言語を記憶するためにたとえば1つのD V Dディスクで利用可能な十分なメモリがないため、世界的に販売される全てのディスクについて機能しない。

#### 【0017】

30

しかし、H D - D V D又はB l u - r a yのような新たなディスクフォーマットは、A Vコンテンツについて更なるディスクキャパシティを提供する。B l u - r a yレコーダは、2003年以降、日本の市場で販売されている。また、益々市場における新たなT Vセットが更に24pフォーマットのビデオを処理して表示することができる。

#### 【課題を解決するための手段】

#### 【0018】

本発明により解決される問題は、24p(又は48p又は60p)の記憶媒体のマスタリングを提供することであり、そのA Vコンテンツには、世界的に50iフォーマットにおいても主観的なビデオ及びオーディオの高い品質、ビデオフィールドを繰り返すことができ、かつオーディオフレームをドロップすることができるA Vコンテンツにおける最良のロケーションを提供することができる。この問題は、請求項1, 2及び3に開示される方法により解決される。請求項1の方法を利用する装置は、請求項3に開示される。対応する記憶媒体は、請求項11に開示される。対応するオーディオ/ビデオデータストリームは、請求項12に開示される。

40

#### 【0019】

上述されたように、異なるオーディオ「スピード」は、異なる国について必要とされる。今日のA V記憶媒体は、圧縮されたオーディオストリームを使用するため、オーディオストリームの自動的なピッチ変換によるシンプルな速度シフトは可能ではない。プレーヤは、オーディオストリームをデコードし、ピッチ変換を実行し、全てのチャネルについてオーディオストリームをリエンコード又は再圧縮して、適切な標準化されたデジタルオーディオストリームを送出する。

50

ディオストリームを外部のオーディオデコーダに伝達する必要がある。これは、異なるオーディオストリームコードのタイプ（たとえばDTS、AC-3、MPEG）をデコード及びエンコードするための更なるハードウェア要件のため、オーディオの再エンコードのための更なるライセンスコストのため、受け入れることができない高価なディスクプレイヤにつながる。

#### 【0020】

本発明を利用するとき、メディアプレーヤ又はレコーダでは、現在のビデオ及びオーディオコンテンツに依存して、オーディオ信号フレームが適応的にドロップされるか（Aドロップ）及び／又はビデオフィールド又はフレームが適応的に繰り返される（Vリピート）。ドロッピング及びリピーティングは、ドロッピング及びリピーティングのそれぞれが知覚できる信号がそれら共通の表示のためにビデオ信号とオーディオ信号の間の同期を制御するように、ビデオ及びオーディオ信号の現在のコンテンツに依存する。

#### 【0021】

特別のビデオフィールド又はフレームを繰り返すこと及び特別のオーディオフレームをドロップすることは、同期されたAVストリームとなり、これは、24p又は60iの速度で24pコンテンツを処理又は表示できないディスプレイ／ラウドスピーカセット又はモニタ又はTVセットについて互換性がある。

#### 【0022】

これにより、ユーザは、彼のディスプレイ又はTVセットにより可能であるように、最も高い画像及び音声の品質でディスクコンテンツを再生することができる。再生の品質の制限は、ディスクコンテンツのフォーマットに依存しないが、ユーザのディスプレイのタイプに依存する。本発明では、用語「オリジナルオーディオ信号速度」は、オリジナルに意図されたビデオフォーマットの再生速度に整合するオーディオ符号化及びオーディオ再生速度を示す。たとえばオリジナルに意図されたビデオフォーマットの再生速度は、24pフォーマットに対応し、オーディオコーディング及び意図されたオーディオ再生速度又は表示は、24pフォーマットのそれに対応する。

#### 【0023】

マスタリングされるべきディスク又は記憶媒体は、オリジナルの再生速度を容易にするオリジナルの24pビデオ及び適切な符号化されたオーディオデータを含み、更なる情報は、特定のオーディオ信号フレーム又はブロックに関して、及び特定のビデオ信号フレーム又はフィールドに関してメディアに記憶される。

#### 【0024】

オリジナルのAVデータは、24pの互換性のあるディスプレイ／ラウドスピーカ又はTVセットについてコンテンツを再生するために使用され、これにより最良の画質を達成する。

#### 【0025】

AVデータストリームに埋め込まれた公知の3:2のプルダウン情報又はデータアイテムは、たとえば60iのDVDビデオディスクについて適用されるときMPEGビデオストリームにおける特別のフラグといった、60iのディスプレイ／ラウドスピーカ又はTVセットについてAVストリームを再生するために使用される。

#### 【0026】

50iの表示について、24pのビデオは、3:2プルダウンの使用なしにビデオ出力に表示される。幾つかの選択されたビデオフィールドのみがビデオ出力に付加的に表示される。これは、たとえばオリジナル速度よりも4.2%高速であることからオリジナル速度よりも2%高速であることに、ビデオ出力を低下させる。対応するフィールド選択は、オーサリングの間に行われる。

#### 【0027】

以下が存在する。ビデオフィールド／フレームリピートについて適切であって、オーディオフレームドロップについて適切なシーン。ビデオフィールド／フレームリピートについて適切であるが、オーディオフレームドロップについて適切ではないシーン。ビデオフ

10

20

30

40

50

イールド / フレームリピートについて適切ではないが、オーディオフレームドロップについて適切であるシーン。ビデオフィールド / フレームリピートについて適切ではないが、オーディオフレームドロップについても適切ではないシーン。

#### 【 0 0 2 8 】

多くのシーンは、たとえば、バックグランドミュージックとのランドスケープビデオのシーンといった正確な同期を必要としない。係るシーンでは、2つのシーン間のカットは、幾つかのブラックのビデオフレームを繰り返し、オーディオと同期される。

#### 【 0 0 2 9 】

幾つかの選択されたオーディオフレームがドロップされる。これにより、オリジナルの速度からオリジナルの速度よりも2%高速にオーディオ出力をスピードアップする。両方の測定は、平均してオリジナルの24pの映画の速度よりも2%高速である、同期されたビデオ / オーディオの再生につながる。対応するオーディオフレームの選択は、オーサリングの間に行われる。

10

#### 【 0 0 3 0 】

特定のAドロップ及びVリピート情報又はデータアイテムは、50iの表示についてAVストリームに挿入され、これにより、50iのTVセット又はディスプレイ / ラウドスピーカでの24pのAVストリームの結果的な平均の再生速度は、48iと50iの間にあり、そのレンジ内で（又はそのレンジを超えて）瞬間に変化する。

#### 【 0 0 3 1 】

現在のビデオ及びオーディオコンテンツ（すなわちシーンコンテンツ）に依存して、ビデオのスローダウンとオーディオのスピードアップ間のバランスは変化する。たとえば、強い動きをもつシーンのためにビデオのスローダウンは、再生の間に目に見えるジャークにつながる。したがって、オーディオのスピードアップは、係るシーンについて好適である。

20

#### 【 0 0 3 2 】

本発明のオーサリング、たとえばDVDビデオのオーサリングは、単一の主要な手順で実行される。

a) 24pのフォーマットでデジタルビデオをエンコードし、60iの再生について3-2プルダウン機能を付加する。

b) 多様なデコーダ及び言語について、24pフォーマットのスピードに従ってデジタルオーディオをエンコードする。

30

c) サブタイトル（DVDにおいて、サブピクチャ）をエンコードする。

d) メニューを生成する。

e) 50iのコンテンツをオーサリングする。e1) AVストリームにおけるシーンのタイプをカテゴリ化する。e2) AVストリームにおけるビデオの繰り返し / オーディオのドロップを配置又は隠匿する。e3) サブタイトルユニットをオーディオ又はビデオに揃える。e4) 50iの情報アイテムを24p / 60iのディスクのデータ画像（すなわちハードディスクに記憶される）に、すなわちAVストリーム及び.info file filesに組み込む。

f) マスタを作製する。

40

#### 【 0 0 3 3 】

本発明の原理がHD-DVD又はBlu-rayのような標準仕様の一部である場合、50iのディスプレイ及び60iのディスプレイ及び24pのディスプレイと互換性をもつ記憶媒体又はディスクを作製するのを可能にする。これは、コストを低減し（2つの代わりに1つのマスタのみが必要とされ）、50iと60iの国間でディスクの交換を容易にする。

#### 【 0 0 3 4 】

原理的に、本発明の方法は、ビデオ信号及び1以上のオーディオ信号を表すオーディオ / ビデオデータストリームをオーサリング又は発生又は符号化するために適しており、前記ビデオ信号は、24Hzビデオフレーム周波数を有し、前記オーディオ信号はオリジナ

50

ルの速度を有しており、前記オーサリング又は発生又は符号化は、以下のステップを含む。24pのフォーマットでビデオデータをエンコードし、任意に、60Hz又はほぼ60Hzのフィールド周波数で再生を容易にするために3-2のプルダウンデータアイテムを付加し、前記オリジナルの速度に従ってオーディオデータをエンコードするステップ。サブタイトルデータをエンコードし、前記ビデオ、オーディオデータ及びサブタイトルデータのうちの少なくとも1つに関連するメニューを発生するステップ。50iのフォーマットで示される48Hzと50Hzとの間のレンジにある平均のフィールド周波数との前記オーディオ及びビデオの主観的に同期された表示を容易にする更なるデータアイテムを発生するステップ。これにより、前記オーディオ/ビデオデータストリームに、ビデオ信号フィールド又はフレーム繰り返しを制御するための前記ビデオ及びオーディオ信号の表示ユニットで使用されるビデオ繰り返し情報アイテムを含むことができ、オーディオ信号フレームドロップを制御するために前記表示ユニットで使用することができるオーディオドロッピング情報アイテムを含むことができる。

### 【0035】

前記方法は、更なるステップを含む。前記オーディオ/ビデオデータストリームでビデオ及びオーディオシーンを自動的に判定し、前記50iのフォーマットで必要とされるビデオ/オーディオ表示の同期に関して前記ビデオ及びオーディオシーンの異なるタイプを自動的にカテゴリ化するステップ。オーディオフレームのドロップ又はビデオフィールドの繰り返しのそれそれが生じない重大なビデオ及びオーディオシーンを判定するステップ。前記シーンのそれぞれ1つについて、オーディオフレームドロップ及び/又はビデオフィールドリピートのそれぞれの最大の管理可能な数を計算するステップ。重要ではないシーンについて、オーディオフレームドロップ情報アイテム及び/又はビデオフィールドリピート情報アイテムの対応する分布を計算し、隣接する重要ではないシーンにおいて対応する数のオーディオフレームのドロップ及び/又はビデオフィールドのリピートにより、現在の重要なシーンにより導入されたビデオ/オーディオの遅延を補償することができるか否かを計算し、真である場合、重要ではないシーンについて対応する数のビデオ繰り返し情報アイテム及びオーディオドロッピング情報アイテムを提供し、前記ビデオ繰り返し情報アイテム及び前記オーディオドロッピング情報アイテムは前記50iのフォーマットの更なるデータアイテムの一部を形成し、真ではない場合、前記50iのフォーマットで、前記50iフォーマットの更なるデータアイテムの一部を形成するように、前記現在の重要なシーンのためのオーディオデータ又はビデオデータをリエンコードするステップ。前記50iのフォーマットの更なるデータアイテムを、前記オーディオ/ビデオデータストリーム又は前記オーディオ/ビデオデータストリームのコピーの関連するデータフィールドに組み込むステップ。

### 【0036】

原理的に、本発明の方法は、ビデオ信号及び1以上のオーディオ信号を表すオーディオ/ビデオデータストリームをオーサリング又は発生又は符号化するために適しており、前記ビデオ信号は、60Hzのビデオフレーム周波数を有し、前記オーディオ信号はオリジナルの速度を有する。前記オーサリング又は発生又は符号化は、以下のステップを含む。60pのフォーマットでビデオデータをエンコードし、前記オリジナルの速度に従ってオーディオデータをエンコードするステップ。サブタイトルデータをエンコードし、前記ビデオ、オーディオデータ及びサブタイトルデータのうちの少なくとも1つに関連するメニューを発生するステップ。50iのフォーマットで示される48Hzと50Hzとの間のレンジにある平均のフィールド周波数との前記オーディオ及びビデオの主観的に同期された表示を容易にする更なるデータアイテムを発生するステップ。これにより、前記オーディオ/ビデオデータストリームに、ビデオ信号フィールド又はフレームのドロッピングを制御するための前記ビデオ及びオーディオ信号の表示ユニットで使用されるビデオドロッピング情報アイテムを含むことができ、オーディオ信号フレームドロップを制御するために前記表示ユニットで使用することができるオーディオドロッピング情報アイテムを含むことができる。

10

20

30

40

50

## 【0037】

前記方法は、更なるステップを含む。前記オーディオ／ビデオデータストリームでビデオ及びオーディオシーンを自動的に判定し、前記50iのフォーマットで必要とされるビデオ／オーディオ表示の同期に関して前記ビデオ及びオーディオシーンの異なるタイプを自動的にカテゴリ化するステップ。オーディオフレームのドロップ又はビデオフィールド又はフレームのドロップのそれぞれが生じない重大なビデオ及びオーディオシーンを判定するステップ。前記シーンのそれぞれ1つについて、オーディオフレームドロップ及び／又はビデオフィールド又はフレームのドロップのそれぞれの最大の管理可能な数を計算するステップ。重要ではないシーンについて、オーディオフレームドロップ情報アイテム及び／又はビデオフィールド又はフレームドロップ情報アイテムの対応する分布を計算し、隣接する重要ではないシーンにおいて対応する数のオーディオフレームのドロップ及び／又はビデオフィールド又はフレームのドロップにより、現在の重要なシーンにより導入されたビデオ／オーディオの遅延を補償することができるか否かを計算し、真である場合、重要ではないシーンについて対応する数のビデオドロッピング情報アイテム及びオーディオドロッピング情報アイテムを提供し、前記ビデオドロッピング情報アイテム及び前記オーディオドロッピング情報アイテムは前記50iのフォーマットの更なるデータアイテムの一部を形成し、真ではない場合、前記50iのフォーマットで、前記50iフォーマットの更なるデータアイテムの一部を形成するように、前記現在の重要なシーンのためのオーディオデータ又はビデオデータをリエンコードするステップ。前記50iのフォーマットの更なるデータアイテムを、前記オーディオ／ビデオデータストリーム又は前記オーディオ／ビデオデータストリームのコピーの関連するデータフィールドに組み込むステップ。  
。

## 【0038】

原理的に、本発明の方法は、ビデオ信号及び1以上のオーディオ信号を表すオーディオ／ビデオデータストリームをオーサリング又は発生又は符号化するために適しており、前記ビデオ信号は、48Hzのビデオフレーム周波数を有し、1以上の前記オーディオ信号はオリジナルの速度を有する。前記オーサリング又は発生又は符号化は、以下のステップを含む。48pのフォーマットでビデオデータをエンコードし、60Hz又はほぼ60Hzのフレーム周波数での再生を容易にするために3：2プルダウンデータアイテムを付加し、前記オリジナルの速度に従ってオーディオデータをエンコードするステップ。サブタイトルデータをエンコードし、前記ビデオ、オーディオデータ及びサブタイトルデータのうちの少なくとも1つに関連するメニューを発生するステップ。50iのフォーマットで示される48Hzと50Hzとの間のレンジにある平均のフィールド周波数との前記オーディオ及びビデオ信号の主観的に同期された表示を容易にする更なるデータアイテムを発生するステップ。これにより、前記50iフォーマットの表示について前記3：2プルダウンデータアイテムを無視し、前記オーディオ／ビデオデータストリームに、ビデオ信号フィールド又はフレームのドロッピングを制御するための前記ビデオ及びオーディオ信号の表示ユニットで使用されるビデオドロッピング情報アイテムを含むことができ、オーディオ信号フレームドロップを制御するために前記表示ユニットで使用することができるオーディオドロッピング情報アイテムを含むことができる。  
40

## 【0039】

前記方法は、更なるステップを含む。前記オーディオ／ビデオデータストリームでビデオ及びオーディオシーンを自動的に判定し、前記50iのフォーマットで必要とされるビデオ／オーディオ表示の同期に関して前記ビデオ及びオーディオシーンの異なるタイプを自動的にカテゴリ化するステップ。オーディオフレームのドロップ又はビデオフィールドの繰り返しのそれぞれが生じない重大なビデオ及びオーディオシーンを判定するステップ。前記シーンのそれぞれ1つについて、オーディオフレームドロップ及び／又はビデオフィールド又はフレームのドロップのそれぞれの最大の管理可能な数を計算するステップ。重要ではないシーンについて、オーディオフレームドロップ情報アイテム及び／又はビデオフィールド又はフレームドロップ情報アイテムの対応する分布を計算し、隣接する重要  
50

ではないシーンにおいて対応する数のオーディオフレームのドロップ及び／又はビデオフィールド又はフレームのドロップにより、現在の重要なシーンにより導入されたビデオ／オーディオの遅延を補償することができるか否かを計算し、真である場合、重要ではないシーンについて対応する数のビデオドロッピング情報アイテム及びオーディオドロッピング情報アイテムを提供し、前記ビデオドロッピング情報アイテム及び前記オーディオドロッピング情報アイテムは前記 50 i のフォーマットの更なるデータアイテムの一部を形成し、真ではない場合、前記 50 i のフォーマットで、前記 50 i フォーマットの更なるデータアイテムの一部を形成するように、前記現在の重要なシーンのためのオーディオデータ又はビデオデータをリエンコードするステップ。前記 50 i のフォーマットの更なるデータアイテムを、前記オーディオ／ビデオデータストリーム又は前記オーディオ／ビデオデータストリームのコピーの関連するデータフィールドに組み込むステップ。

#### 【0040】

原理的に、本発明の装置は、ビデオ信号及び 1 以上のオーディオ信号を表すオーディオ／ビデオデータストリームをオーサリング又は発生又は符号化するために適しており、前記ビデオ信号は、24 Hz のビデオフレーム周波数を有し、1 以上の前記オーディオ信号はオリジナルの速度を有する。前記オーサリング又は発生又は符号化装置は、以下の手段を含む。24 p のフォーマットでビデオデータをエンコードし、60 Hz 又はほぼ 60 Hz のフィールド周波数での再生を容易にするために 3 - 2 プルダウンデータアイテムを付加し、前記オリジナルの速度に従ってオーディオデータをエンコードし、サブタイトルデータをエンコードし、前記ビデオ、オーディオデータ及びサブタイトルデータのうちの少なくとも 1 つに関連するメニューを発生する手段。50 i のフォーマットで示される 48 Hz と 50 Hz との間のレンジにある平均のフィールド周波数との前記オーディオ及びビデオ信号の主観的に同期された表示を容易にする更なるデータアイテムを発生する手段。これにより、前記オーディオ／ビデオデータストリームに、ビデオ信号フィールド又はフレームの繰り返しを制御するための前記ビデオ及びオーディオ信号の表示ユニットで使用されるビデオドロップを制御するために前記表示ユニットで使用することができるオーディオドロッピング情報アイテムを含むことができる。

#### 【0041】

前記手段は、前記オーディオ／ビデオデータストリームでビデオ及びオーディオシーンを自動的に判定し、前記 50 i のフォーマットで必要とされるビデオ／オーディオ表示の同期に関して前記ビデオ及びオーディオシーンの異なるタイプを自動的にカテゴリー化し、オーディオフレームのドロップ又はビデオフィールドの繰り返しのそれが生じない重大なビデオ及びオーディオシーンを判定する。

#### 【0042】

前記手段は、前記シーンのそれぞれ 1 つについて、オーディオフレームドロップ及び／又はビデオフィールドの繰り返しのそれぞれの最大の管理可能な数を計算する。

#### 【0043】

前記手段は、重要ではないシーンについて、オーディオフレームドロップ情報アイテム及び／又はビデオフィールド繰り返し情報アイテムの対応する分布を計算し、隣接する重要ではないシーンにおいて対応する数のオーディオフレームのドロップ及び／又はビデオフィールドの繰り返しにより、現在の重要なシーンにより導入されたビデオ／オーディオの遅延を補償することができるか否かを計算し、真である場合、重要ではないシーンについて対応する数のビデオドロップ情報をアイテム及びオーディオドロッピング情報アイテムを提供し、前記ビデオドロップ情報をアイテム及び前記オーディオドロッピング情報アイテムは前記 50 i のフォーマットの更なるデータアイテムの一部を形成し、真ではない場合、前記 50 i のフォーマットで、前記 50 i フォーマットの更なるデータアイテムの一部を形成するように、前記現在の重要なシーンのためのオーディオデータ又はビデオデータをリエンコードするステップ。

#### 【0044】

10

20

30

40

50

前記手段は、前記 50 i のフォーマットの更なるデータアイテムを、前記オーディオ / ビデオデータストリームの関連するデータフィールド、又は前記オーディオ / ビデオデータストリームのコピーに組み込む。

#### 【0045】

本発明の有利な更なる実施の形態は、それぞれの従属の請求項で開示される。

本発明の例示的な実施の形態は、添付図面を参照して記載される。

#### 【発明を実施するための最良の形態】

##### 【0046】

図 1 は、図 1 a における公知の表示モード 24 p、図 1 b における 60 i、及び図 1 c における 50 i、並びに図 1 d における本発明に関する 50 i 表示モードを示す。幾つかのオーディオ又はビデオデータの矩形の下にあるハッチングされた領域は、ドロップ / 繰り返しされたオーディオ又はビデオデータの相対的な量を示す。たとえば、図 1 b では、60 i について 3 : 2 プルダウンは、符号化されたビデオフィールドの 25 % を 2 度示している。図 1 d では、選択されたビデオフィールドの繰り返しは約 2 % を追加し、選択されたオーディオフレームのドロップは、約 2 % を占める。10

##### 【0047】

実際に、平均のビデオストリームの再生速度  $V_v$  は 50 i から 48 i の近くにまで減少し (48 i  $V_v$  50 i)、平均のオーディオストリームの再生速度  $V_a$  は 48 i から 50 i の近くにまで増加し (48 i  $V_a$  50 i)、平均として  $V_v$  は  $V_a$  に等しい結果が得られる（しかし、再生の開始と終了で小さな時間のギャップが生じる場合がある）。20  $V_v$  と  $V_a$  は、AV ストリーム内で変化する場合がある。言い換えれば、50 i の TV セット又はディスプレイ / ラウドスピーカでの 24 p の AV ストリームの結果的な平均の再生速度は、48 i と 50 i の間にあり、すなわち平均の再生速度は、0 と 4.2 % の間により増加される。局所的に、短時間の間に、ビデオ又はオーディオ速度は、48 i よりも低速であって、50 i よりも高速となる可能性がある。

##### 【0048】

異なる視点で、図 2 は、60 i の 3 : 2 のプルダウンと 50 i の表示の間の違いを示す。図 2 の上の部分は、それらトップフィールド T 及びボトムフィールド B と共に、I 符号化ピクチャ I、予測された符号化ピクチャ P、及び双方向で予測された符号化ピクチャ B を含む MPEG GOP (グループオブピクチャ) の一部を示す。50 i フォーマットについて、幾つかの（スマートに選択された）ビデオフィールドのみが繰り返され（ハッチングされた領域）、幾つかの（スマートに選択された）オーディオフレーム（たとえばオーディオの 32 ms をそれぞれ表している）のみがドロップされる。30

##### 【0049】

迷惑となる 50 i に関する作用が生じる。

ビデオにおける強い動きの間のビデオフィールドの繰り返しのケース。調和音の間のオーディオフレームのドロッピングのケース。リップシンクのシーンの間のオーディオとビデオの間にある遅延のケース。かかる悩ましい作用は、オーディオフレームのドロップ / ビデオフィールドの繰り返しといった、本発明のスマートな散乱 (smart scattering) により回避される。40

##### 【0050】

図 3 は、スマートに選択されたオーディオフレームのドロップ / ビデオフィールドの繰り返しの幾つかの例を示している。左の部分では、バイオリンの音楽が再生されている間に動きが無いか、又は動きが殆どない。オーディオフレームのドロップはない。12 ピクチャの長さを有する図示された GOP では、I ピクチャに続く最初の B ピクチャでは、トップフィールドが繰り返されている。これにより、次のフィールドの繰り返しが行われて始めて、後続のピクチャは、ボトムフィールドでそれぞれ開始する。中段の部分は、シーンカットを表しており、ここでは、最初のシーンの終了で、ボトムフィールドとしてフィールドが繰り返され、第二のシーンの最初のフィールドは、更なるトップフィールドにより先行される。右及び殆ど静かな部分は、強い動きをもつ更なるシーンを表す。フィール50

ドの繰り返しを予測されないが、更に頻繁なオーディオフレームのドロップが強制される。

**【 0 0 5 1 】**

上記ステップ e ) の 4 つのフェーズ e 1 ) ~ e 4 ) は、以下に更に詳細に記載される。

**【 0 0 5 2 】**

e 1 ) A V スクリーンにおけるシーンのカテゴリ化。このフェーズは、全体のビデオ系列をシーンにカテゴリ化する。それぞれのシーン変化は、ビデオ又はオーディオ表示、若しくはその両者の特別な振る舞いの変化を記述する。

**【 0 0 5 3 】**

シーンのカテゴリは、以下を含む。

10

カテゴリのタイプ：動き、音、シーンカット、リップシンク；

発見されたカテゴリタイプのレベル（それぞれのカテゴリタイプのレベルの種類及び量はタイプに依存する）；

シーンの開始のタイムスタンプ。

**【 0 0 5 4 】**

一般に、

（サブピクチャユニットについて D V D ビデオで実現されるように）新たなシーンは前のシーンを停止する。

1 以上のカテゴリレベルが変化するとき新たなシーンが開始する。

それぞれのシーンは、全てのカテゴリにより記載されるが、異なるレベルである。

20

**【 0 0 5 5 】**

フェーズ e 1 ) は、幾つかの検出プロセスを含んでおり、2つはビデオコンテンツ向けであり、2つはオーディオコンテンツ向けであり、これらは、並列に実行される場合がある。

**【 0 0 5 6 】**

ビデオに関連したカテゴリの検出：

（自動的なプロセスである）シーンにおける動きを検出する。

M P E G 又は V C 1 動きベクトルを使用してシーンを判定する。3 レベル：（1）静止画像、（2）低速の動き、（3）強い調和振動。ブランクのビデオフレームは（頻繁に現れる）静止画像である。

30

非常に暗いか明るいシーンを検出するために輝度の D C 値を使用する（これはフィールドの繰り返しについて理想的であり、すなわち、レベル（2）としてマークされる）。高い解像度のレベル（たとえば 10 レベル）は、ビデオフレームの繰り返しのスムーズな隠匿につながり、すなわち幾つかの動きを含むシーンは、少ない動きを有するシーンよりも少ないビデオフィールドの繰り返しをカバーする。

**【 0 0 5 7 】**

（自動的なプロセスである）シーンカットを検出する。

既に利用可能なソフトウェアはシーンカットを自動的に検出する。3 レベル：（1）シーンの開始、（2）シーンの終了、（3）ノンシーンカットのシーン。レベル（1）のシーンの最初のフレーム、又はレベル（2）のシーンの最後のフレームは、ビデオフィールドの繰り返しを実行するために理想的である。検出されたカットは、現実的なシーンカットである必要はない（すなわち、シーンカットの検出プロセスは、強固にフォールトトレラントである）。ある画像から次の画像への自然発生の変化をもつシーン（たとえばフラッシュ）は、このカテゴリにフィットされる。

40

**【 0 0 5 8 】**

サウンドに関するカテゴリの検出。

音のシーンを検出する。これは、自動的なプロセスであり、サウンドアナライザ（プリグラム）がボリューム及び / 又は周波数の変化を検出する。

レベル（1）：消音されたオーディオ、演説又は音楽の間の短い中断。

レベル（2）：低いボリューム、可聴の雑音、又はノンミュージック、又は非一定の周

50

波数及びボリューム。

レベル(3)：可聴の雑音のない（たとえばバイオリンコンサートといった）調和オーディオの変化。なお、ショートレベル(2)のシーンは、たとえばバイオリンコンサートといった間でさえ現れる。

高い解像度のレベル（たとえば10レベル）は、オーディオフレームのドロップのスムーズな秘匿をもたらす。

#### 【0059】

リップシンクシーンの検出。これは、自動的なシーン候補の検出である。サウンドアライザ（プログラム）は、演説及び／又は音のピークを検出する。人物は、発見された候補のシーンがリップシンクを必要とするか否かを確認する。リップシンク(1)及びリップシンク(2)といった2つのレベルが使用される。高解像度のレベル（たとえば10レベル）は、異なるレベルの許可された遅延を容易にする。たとえば、ドアを急に閉めることは、強い「リップ」の同期を必要とする。しかし、通過する自動車の音について、オーディオとビデオの間の遅延は許容可能である。

#### 【0060】

言い換えれば、カテゴリタイプは、以下に示される。

#### 【0061】

シーンカットが現在存在する／存在しない。

オーディオ及びビデオのリップシンクが現在必要とされる／必要とされない。少なくとも1つのビデオに関連したタイプ及び少なくとも1つのオーディオに関連したタイプ。

現在のビデオコンテンツにおける動き／動きなし。

非常に暗い又は非常に明るいシーンが現在存在する／存在しない。

音が現在存在する／存在しない。

低いボリューム、又は可聴の雑音、又はノンミュージック、又は一定の周波数及びボリュームが現在存在する／存在しない。

可聴雑音のない調和音の現在の変化が存在する／存在しない。

#### 【0062】

シーンカテゴリ化プロセスは、主に自動的なプロセスである。候補となるリップシンクのシーンは、ある人物による確認又は拒否を必要とする。このプロセスは、強固にフルトレントであり、正確な設定が必要ではない。カテゴリ化プロセスにより、カテゴリ及び割り当てられたスタンプからなるリストが得られる。

#### 【0063】

デモンストレーションのムービーディスクのためのカテゴリリストが生成される。図7は、そのカテゴリ化プロセスの例となる抜粋を示している。ビデオフレームの番号は、タイムスタンプとして使用される。B／V／vは、ビデオの動きレベル(1)／(2)／(3)を示す。M／A／aは、オーディオサウンドレベル(1)／(2)／(3)を示す。S／s／-は、シーンカット（シーンの開始／終了／シーンカットなし）を示す。Y／yは、リップシンク／ノンリップシンクを示す。図7aは、管理可能な遅延のレンジの計算を示す。図7bは、重要なシーン(critical scene)及び前遅延(pre-scene)並びに後遅延(post-scene)の管理を示す。図7cは、最適化を示す。図7dは、終了処理を示す。破線の円は、それら前の状態から変化されたエリアを示す。

#### 【0064】

e2) AVストリームにおけるビデオのリピート／オーディオのドロップの配置又は秘匿。このフェーズは、オーディオフレームのドロップ(drop:コマ落ち)及びビデオフィールドのリピート(repetition:繰り返し)を計算するため、先に発見されたカテゴリを使用する。処理は、以下のように実行することができる。

#### 【0065】

図7a及び図4を参照して、それぞれのシーンについて管理可能な遅延のレンジ(min及びmax)のドロップ／リピート)を計算する。図4において、SCENE1及びSCENE3は、SCENE2により生じたオーディオとビデオの間の遅延を補償する。SC

10

20

30

40

50

E N E 1 では、このスタンダロンのシーンについて必要なよりも多くのビデオフィールドの繰り返しが存在する。S C E N E 2 では、ビデオフィールドの繰り返し又はオーディオフレームのコマ落ちのいずれも存在せず、前のシーン及び後ろのシーンは、（たとえば、開始及び終了のクレジットといった）同期を提供する。S C E N E 3 では、このスタンダロンシーンについて必要なよりも多くのビデオフィールドのリピート及びオーディオフレームのドロップが存在する。

#### 【 0 0 6 6 】

図 7 b 及び図 4 を参照して、重要なシーン（“ a v y ” 及び “ a v Y ” ）の前遅延及び後遅延を計算する。

重要なシーンを管理する。隣接するシーンは重要なシーンを補償することができるか？  
真である場合、前遅延及び後遅延を隣接する 1 以上のシーンに配置する。真ではない場合、シーンのオーディオを再びエンコードし、オーディオストリーム又は全体の並列な A V  
ストリームの部分を追加する。

#### 【 0 0 6 7 】

最適化は、できるだけ同期について残りの遅延を低減する。

静止画像：できるだけ多数の静止画像のフィールドであって、ビデオをオーディオと同期させるために必要なフィールドを繰り返す。

消音されたオーディオ：できるだけ多数の消音されたオーディオフレームであって、オーディオをビデオと同期するために必要なフレームをドロップする。

シーンカット：できるだけ多くの最初と最後のシーンカットのビデオフィールドであって、オーディオとビデオを同期させるために必要なフィールドを繰り返す。重要度の低いシーンから重要度のない隣接するシーンに幾つかのドロップ / リピートをシフトする。

#### 【 0 0 6 8 】

終了処理は、全ての残りの未だ同期されていないシーンを補償する。動きがないか又は低い動きのシーンにおいて、“ V ” ビデオフィールドは、たとえば 2 4 : 1 のプルダウンを使用することで繰り返される。低いボリュームのシーンにおいて、“ A ” オーディオフレームは、たとえば 2 5 番目のオーディオフレーム毎にドロップされる。

#### 【 0 0 6 9 】

図 5 は、このフェーズのための例を示す。図 5 は、図 4 に類似しており、シーンを連結する前の遅延を強調し、シーンの同期化された連結を得るために、オーディオのドロップ / ビデオのリピートの使用を強調する。

#### 【 0 0 7 0 】

e 3 ) サブタイトルユニットをオーディオ又はビデオに揃える。

このフェーズは、サブタイトル ( D V D におけるサブピクチャ ) をビデオ又はオーディオに揃える。

サブタイトルは、オーディオと同期される（デフォルト）。

強い遅延をもつシーンにおいてサブタイトルを検出する（オートマチック）。

ある人物は、ビデオアラインメントが好まれるか否かを判定する。

遅延はめったになく、すなわち、幾つかのサブタイトルのみが確認される必要がある。

存在する場合、同じアラインメントを様々な言語の全ての並列なサブタイトルに割り当てる。

#### 【 0 0 7 1 】

e 4 ) 5 0 i の情報を（たとえばハードディスクに記憶される） 2 4 p / 6 0 i のディスクデータ画像、すなわち A V ストリーム及び.info file 又は files に組み込む。

( H D - ) D V D ビデオ：ナビゲーションパック NV\_PCK ( すなわち、この V O B U のための情報のみ ) 。

（符号化フレーム 0 . . . 1 2 の）繰り返されるべきビデオフィールドを一覧にする ( V O B U の平均期間 < 0 . 5 s ) 。

（符号化フレーム 0 . . . X の）ドロップされるべき G O F ( リニア P C M 、グループ

10

20

30

40

50

オーディオフレーム)のオーディオフレームを一覧にする(Xはオーディオのタイプに依存し、V O B Uの平均期間<0.5s)。

(再生の開始について)オーディオとビデオの間の同期オフセットは、この再生がトップ又はボトムフィールドで開始するかを示すことを含む。

サブタイトル(たとえばD V Dにおけるサブピクチャユニット)ビデオに挿えられる場合のフラグ("0"又は"1")。(通常はB C Dフォーマットである)記憶される表示時間は、50iの再生について更に利用可能である。

#### 【0072】

(H D - ) D V D ビデオ : プログラムチェイン P G C

I F O ファイル及びA Vストリームにおいて50iの再生と同様に24p / 60iについての情報アイテムが存在することを示すため、情報ファイルにおいて情報を追加する(ナビゲーション又は上位レベルの情報、D V DではI F O ファイル)。これにより、I F O ファイルにおける上位レイヤが、全てのP G Cが50iの情報を含むことを既に示している場合、この種の情報は使用されなくなる。

#### 【0073】

(たとえばオーディオストリーム又は全体のV O B Uといった)50iのための専用のA Vデータが存在する場合、50i - P G C又は50i - P G Cを追加する。これにより、50i - P G Cは、既存の24p - P G Cのコピーであるが、非常に重要なシーンについてオーディオストリーム又はセルの何れかをリポイント(re-point)する。

#### 【0074】

専用の50i - P G Cの代替として、24p、60i及び50iとしてA Vデータの表示について情報をカバーするため、P G Cの定義を拡張する。

50iの再生について、適切なオーディオストリームを選択するために情報を追加する。

#### 【0075】

正しいセルを選択するために情報を追加する。

#### 【0076】

たとえば全体のプログラムの期間といった(たとえばB C Dフォーマットで)正しい表示時間に関する情報を追加する。

#### 【0077】

なお、全体のプログラムの期間は、50iと24p / 60iの間で異なる。

#### 【0078】

(H D - ) D V D ビデオ : 50iの互換性のフラグをI F O ファイルに設定する。

(H D - ) D V D ビデオ : (通常はB C Dフォーマットである)I F O ファイルにおける全ての記憶された表示時間は、50iの再生のために更に利用可能である。

#### 【0079】

好ましくは、オーディオとビデオの間の同期オフセットに関連する更なる50iの情報アイテム、及びサブピクチャの表示時間、並びに、たとえばビデオフィールドの繰り返し及びオーディオフレームのドロップを示すためのビットリストがG O P (グループオブピクチャ)指向の方式で、D V D又はV O B U (ビデオオブジェクトユニット)指向の方式で本発明の出力A Vストリームに配置される。

#### 【0080】

図10は、関連するオーディオフレームA F R M S A 0 ~ A 1 4と共に、24pのタイムライン24p T Lでの表示順序における、ビデオフレームV F R M S I 0 , P 0 ~ P 2及びB 0 ~ B 7をもつV O B U又はG O Pを示し、これにより、ハッチングされた部分又はグレイの部分は、この現在のV O B Uで符号化される。現在のV O B Uは、前のV O B UであるP V O B Uと後続のV O B UであるN V O B Uの間で配置される。

#### 【0081】

必要な場合、オーディオ表示とビデオ表示の間の開始オフセットは、たとえば映画のチャプターへの直接的なジャンプの後の再生開始といった、表示の開始について適用される

10

20

30

40

50

。図11は、タイムライン24pTL及び625/50のタイムライン625/50TLにおけるビデオとオーディオ間のオフセットAV\_OFFSETの関連する処理を示しており、図11aは、ポジティブなAVオフセットを示し、図11bは、ゼロAVオフセットを示し、図11cは、ネガティブなAVオフセットを示す。図11a,b及びcの上部は、タイムライン24pTLを示し、図11a,b及びcの下部は、タイムライン625/50TLを示す。表示がマークされたVOBで開始する場合、AV\_OFFSETは、オーディオとビデオの間の開始オフセットを記述する。時間スタンプts1は、24pのタイムラインにおける特別のビデオフレームの表示の開始時間である。625/50のタイムラインにおけるこのSVFの表示の開始時間は、オーディオスポットの表示時間と対応する。タイムスタンプts2は、このオーディオスポットの24pのタイムラインにおける対応する時間である。ts1及びts2に対応する符号化されたオーディオフレーム、及びこの符号化されたSVFは、現在のVOB又は後続のVOBであるNVOBで開始する。オーディオフレーム及びこのフレームSVFの両者は、任意の前のVOBのPVOBをデコードすることなしにデコード可能である。図11cでは、特定のビデオフレームSVFは、ネガティブなAVオフセットにより、フレームI0ではなく、フレームP0である。完全なGOPは前にデコードされるが、表示はフレームP0で開始する。

#### 【0082】

図6は、かかる稀なケースで、更なる50iの関連するオーディオストリーム又は全体のAV系列(PGCAV)がAVストリームに挿入することができるよう、重要なシーンを示す。“PGCA”は、プログラムチェインが対応するオーディオストリームを選択するが、同じビデオストリームを使用することを意味する。“PGCAV”は、プログラムチェインが24p/60i又は50iの再生の何れかについてセル/角度を選択することを意味する。

#### 【0083】

図9における本発明のマスタリング装置のブロック図では、DVDビデオ24p/60iのAVストリームについてソース24p/60iAVSを供給する、ビデオ/オーディオ/サブスタイル/メニューエンコーダVASENCが存在する。このAVストリームAVSは、デマルチプレクサDMUXでオーディオストリームASとビデオストリームVSに分割される。ビデオストリームは、オートマチックビデオシーン検出器AUVCDETに入力し、ここでビデオストリームは、ビデオデコーダVDECでデコードされ、オーディオストリームは、オートマチックオーディオシーン検出器AUASCDETに入力し、ここでオーディオストリームは、オーディオデコーダADECでデコードされる。AUVCDETは、動きベクトルを使用するビデオ動き検出ステージVMOTDET、静止画検出ステージSTPICDET及びシーンカット検出ステージSCCTDETを含む。AUASCDETは、調波音のみの検出ステージHAO、オーディオピーク検出ステージAP及び音声検出ステージSPを含む。検出された(結合された又は個別の)ビデオシーン情報アイテムDETVCは、検出されたオーディオシーン情報アイテムDETAS、及び検出されたリップシンクシーン情報アイテムDETLPSSCは、結合手段CMBで結合され、収集されたシーンアイテムは、メモリMEMに記憶される。このメモリは、シーンアイテム、AVドロップ/繰り返しアイテム及びサブタイトルメント(始めはオーディオに揃える)のためのストレージとしての役割を果たす。

#### 【0084】

AVストリームAVSは、1以上のモニタMONに供給される。あるモニタで、ある人物は、オーディオシーン検出器AUASCDETで自動的に検出された候補となるオーディオピークのシーン及び候補となる演説シーンをチェックする。結果的に得られるヒューマンコントロール信号HCTRLL1は、たとえばスイッチSWを使用することで、自動的に検出された候補となるオーディオピークシーン及び候補となるスピーチシーンのどれがリップシンクシーン情報アイテムDETLPSSCを形成するかを判定する。

#### 【0085】

メモリMEMから取り出された(ビデオ及びオーディオ)シーン情報アイテムSCは、

10

20

30

40

50

A V 遅延計算手段 A V D E L C に供給される。結果的に得られる遅延情報アイテムは、ジェネレータ ADD A V D G E N で非常に重要なシーンについて更なる A V データを発生し、モニタ M O N に強い遅延とサブタイトルをもつシーンの形式で遅延最適化手段 D E L O P T を通過するために使用される。ある人物は、サブタイトルが対応するビデオ信号コンテンツにアライメントされるか否かをチェックする。結果的に得られるヒューマンコントロール信号 H C T R L 2 は、サブタイトル / ビデオアライメントについて対応する変化を判定し、結果的に得られる情報アイテムは、メモリ M E M に記憶される。遅延最適化手段 D E L O P T は、メモリ M E M に記憶されている、オーディオフレームのドロップ及びビデオフィールドの繰り返しについて必要とされる情報アイテムを提供する。

## 【 0 0 8 6 】

10

オーディオフレームのドロップとビデオフィールドの繰り返し、サブタイトルアライメント及び新たな表示時間に関する情報は、メモリ手段 M E M から最終的な A V ストリーム変更手段 F A V S M に供給され、I F O ファイル及び更なる A V ストリーム情報を補正するために使用され、これらは D V D ビデオの 2 4 p / 6 0 i の A V ストリームのソース 2 4 p / 6 0 i A V S 又はそのストリームのコピーに更なる A V ストリームデータとして盛り込むことができる。ジェネレータ ADD A V D G E N で発生された非常に重要なシーンについて更なる A V データは、D V D ビデオの 2 4 p / 6 0 i の A V ストリームのソース 2 4 p / 6 0 i A V S 又はそのストリームのコピーへの更なる A V ストリームデータ及び I F O ファイルとして盛り込まれる。

## 【 0 0 8 7 】

20

有利なことに、マージされたオーサリングにより、5 0 i の処理は、同じ記憶媒体のマスタについて更なる処理ステップである。5 0 i のオーサリングは、主に自動的なプロセスであり、検出結果に関する幾つかの確認についてのみ人間の対話が必要とされる。発見されるカテゴリは、非常にフォールトトレラント、すなわち実現するのが簡単である。幾つかの不適切なセット (bad-set) のビデオの繰り返しでさえも殆ど知覚されない。

## 【 0 0 8 8 】

一般に、ディスクメモリの消費は増加しない。稀なケースで、幾つかの更なるメガバイトが必要とされる。

全ての映画のディスクの 9 0 % を超える容量が更なる P G C を必要としないことが予測される。

30

マスタの数の低減により、安価な大量生産が容易になる。

## 【 0 0 8 9 】

望まれる場合、本発明のディスクは、2 4 p / 5 0 i のみについてオーサリングされ、すなわち、6 0 i - P G C を含まない。

2 4 p の高精細 (H D ) から標準の精細 (S D ) へのダウンコンバージョンは、5 0 i の表示についても機能する。

## 【 0 0 9 0 】

40

5 0 i のフォーマットの表示について 6 0 p (H D 又は S D ) のソースコンテンツを変換することも可能である。このタスクのため、アナログの種類の処理が使用される。上述された方法に対する主な違いは、5 0 i のフォーマットの表示を容易にするために繰り返される代わりに、(多くの) ビデオフィールドはドロップされる必要があることである。フィールドの数の低減は、理論的に「逆 3 - 2 のプルダウン」を実行し、残りのビデオフィールドの 5 0 % と 5 4 . 2 % の間で更なるドロップを行うことで達成され、これにより、所望の 5 0 i の論理的な (又は 4 9 i の物理的な) フォーマットの表示を得るために結果的に得られる 4 8 p のフォーマットに本発明を適用することで、全体のビデオフレームがドロップされる。実際に、本発明の原理を使用して、6 0 p から約 4 9 i へのダイレクトビデオコンバージョンが存在するように、固定された比率のプルダウンは存在しないが、適応的なフィールド / フレームのドロッピングが存在する。

## 【 0 0 9 1 】

ソースコンテンツが 4 8 p のフォーマットを有し、対応する 3 - 2 プルダウンデータア

50

アイテムを発生及び加えることで、60pのフォーマットがオーサリングされた場合、平均して（物理的な）49iの表示速度が達成されるように、ビデオフィールドのドロップ及び対応するオーディオフレームのドロップが実行される。これにより、3-2プルダウンは、50iの論理的な（又は49iの物理的な）フォーマット表示についてアクチベートされない（すなわち、対応するフラグが無視される）。ドロップされるビデオフィールドの量は50%と54.2%の間であり、すなわち、第二のフィールド毎にドロップされる必要があり、全体のビデオフレームがドロップされることさえある。

#### 【0092】

この説明において、表現「オーディオフレーム」が使用される。「オーディオフレーム」は、15と50msの間の期間に対応する。オーディオ信号のタイプ（たとえばリニアPCM）が短いオーディオ部分を使用するケースで、係る短いオーディオフレームの適切なグループが代わりに使用される。たとえば、GOFは、リニアPCMのケースで「オーディオフレーム」を表す。

#### 【図面の簡単な説明】

#### 【0093】

【図1】公知の表示モード24p、60i及び50iと本発明に関連する50i表示モードとの比較を示す図である。

【図2】幾つかのスマートに選択されたオーディオドロップ／ビデオ繰り返しのみによる、3:2のプルダウン60i処理と50i処理との間の違いを示す図である。

【図3】スマートに選択されたオーディオフレームのドロップ及びビデオフィールドの繰り返しの例を示す図である。

【図4】重要なシーンである、シーン2により導入されるオーディオとビデオの間の遅延を補償するシーン1及びシーン3を示す図である。

【図5】図4に関する強調される遅延を示す図である。

【図6】更なる50iの関連するオーディオストリーム又は全体のAV系列がストリームに挿入されるような、重要なシーンを示す図である。

【図7】オーディオフレームのドロップ及びビデオフィールドの繰り返しプロセスにおけるステップを示す図である。

【図8】本発明のマスタリングプロセスのフローチャートである。

【図9】本発明のマスタリング装置のブロック図である。

【図10】ビデオフレーム及び関連するオーディオフレームによるVOBU又はGOPを示す図である。

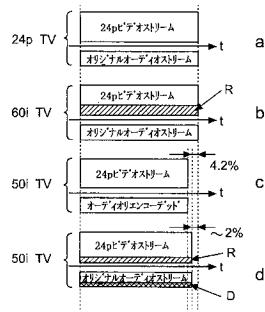
【図11】ビデオとオーディオの間のオフセットを示す図である。

10

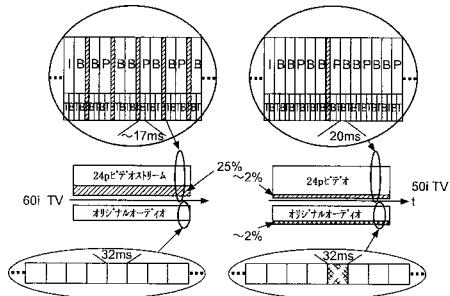
20

30

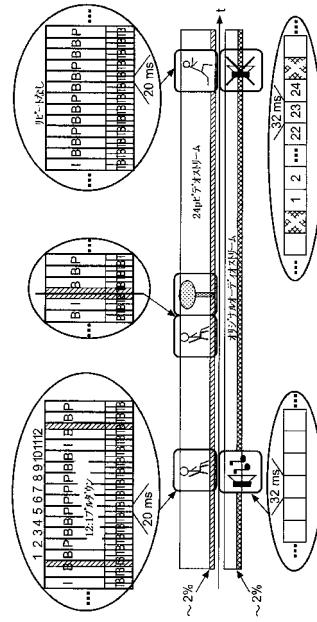
【図1】



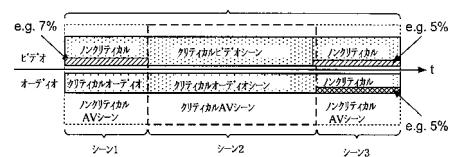
【図2】



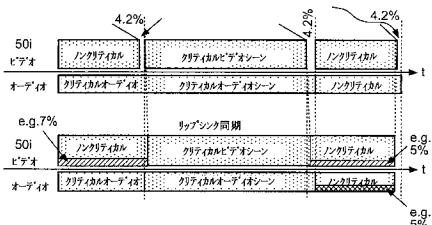
【図3】



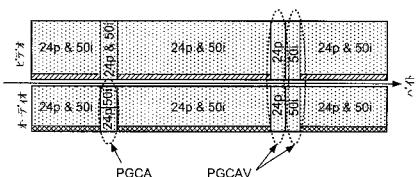
【図4】



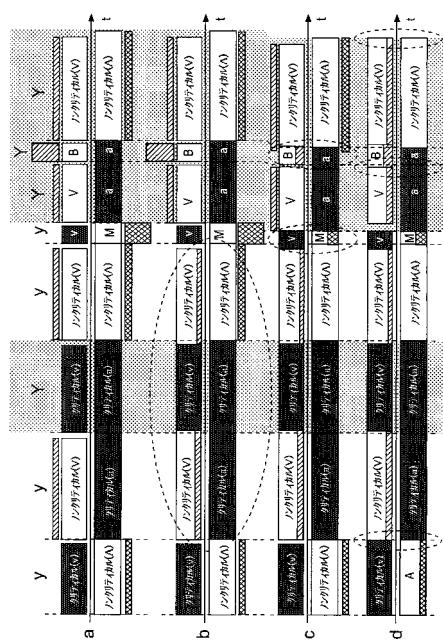
【図5】



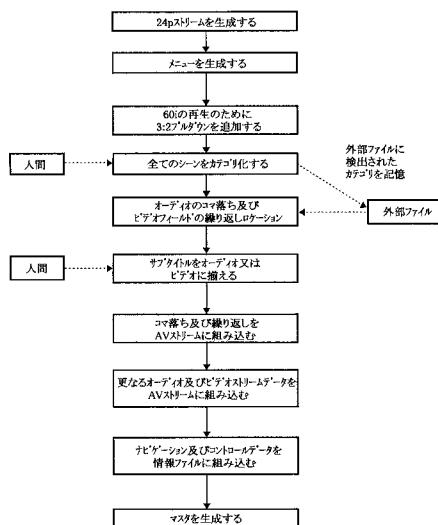
【図6】



【図7】



【 四 8 】



【 図 9 】

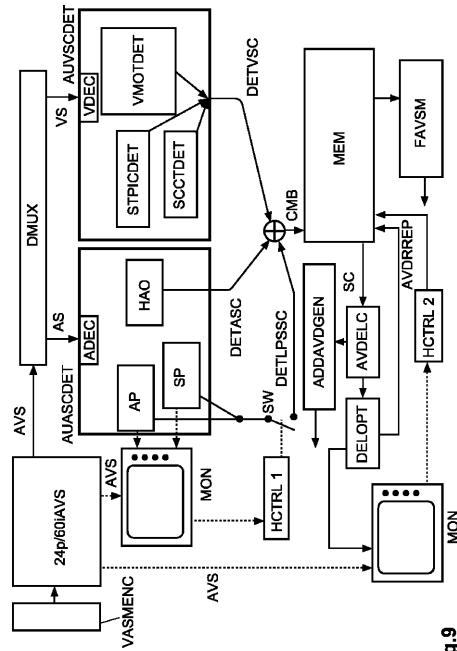
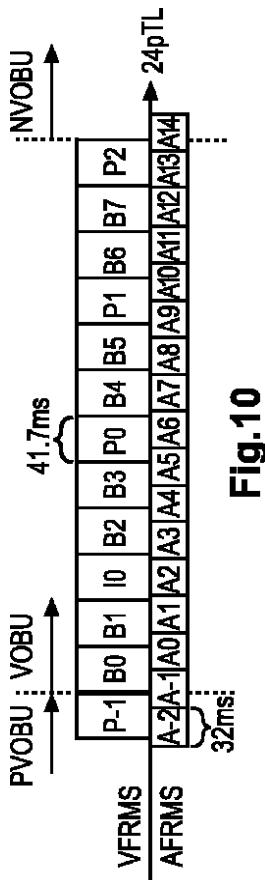


Fig. 6

【 四 1 0 】



**Fig. 10**

【 囮 1 1 】

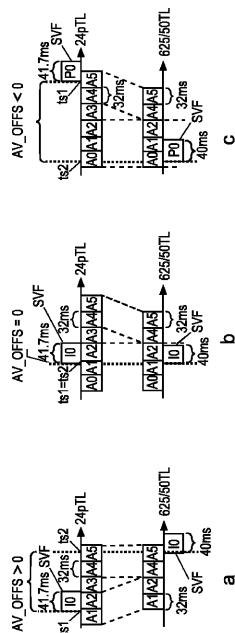


Fig.11

---

フロントページの続き

(72)発明者 ヴィンター , マルコ  
　　ドイツ連邦共和国 , 3 0 6 5 9 ハノーヴァー , シュナイデルミューラー・ヴェーク 19

(72)発明者 ガンドルフ , ディルク  
　　ドイツ連邦共和国 , 3 0 9 5 2 ローネンベルク , ヴァルブリンク 2

(72)発明者 ヘルベル , カルステン  
　　ドイツ連邦共和国 , 3 0 9 7 4 ヴェーニクセン , シュヴァルツェ - ドルン - シュトラーセ 4

(72)発明者 ヘーレントループ , ヨーブスト  
　　ドイツ連邦共和国 , 3 0 1 6 3 ハノーヴァー , ガベルスベルガーシュトラーセ 18

(72)発明者 ヤンセン , ウヴェ  
　　ドイツ連邦共和国 , 3 0 9 2 6 ゼールツェ , ニーダーザクセンシュトラーセ 53

(72)発明者 オステルマン , ラルフ  
　　ドイツ連邦共和国 , 3 0 6 5 7 ハノーヴァー , ニッデナー・ヴェーク 7

(72)発明者 ペーテルス , ハルトムート  
　　ドイツ連邦共和国 , 3 0 8 9 0 バルジンクハウゼン , オーヴェーク 34

(72)発明者 シエフツォフ , アンドレイ  
　　ドイツ連邦共和国 , 3 0 1 6 3 ハノーヴァー , タラフェラシュトラーセ 14

審査官 伊東 和重

(56)参考文献 国際公開第2004/082274 (WO , A1 )  
　　国際公開第2005/074270 (WO , A1 )

(58)調査した分野(Int.Cl. , DB名)

H04N 7/01

H04N 5/262

H04N 5/928

G11B 20/10