

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号
特許第6349977号
(P6349977)

(45) 発行日 平成30年7月4日 (2018.7.4)

(24) 登録日 平成30年6月15日 (2018.6.15)

(51) Int.Cl.

F I

HO 4 N 21/44 (2011.01)

HO 4 N 21/44

G 1 O L 25/57 (2013.01)

G 1 O L 25/57

請求項の数 20 (全 51 頁)

(21) 出願番号	特願2014-115406 (P2014-115406)	(73) 特許権者	000002185
(22) 出願日	平成26年6月4日 (2014.6.4)		ソニー株式会社
(65) 公開番号	特開2015-149705 (P2015-149705A)		東京都港区港南1丁目7番1号
(43) 公開日	平成27年8月20日 (2015.8.20)	(74) 代理人	100121131
審査請求日	平成29年1月27日 (2017.1.27)		弁理士 西川 孝
(31) 優先権主張番号	特願2013-218267 (P2013-218267)	(74) 代理人	100082131
(32) 優先日	平成25年10月21日 (2013.10.21)		弁理士 稲本 義雄
(33) 優先権主張国	日本国 (JP)	(72) 発明者	高橋 秀介
(31) 優先権主張番号	特願2013-257508 (P2013-257508)		東京都港区港南1丁目7番1号 ソニー株
(32) 優先日	平成25年12月12日 (2013.12.12)		式会社内
(33) 優先権主張国	日本国 (JP)	(72) 発明者	井上 晃
(31) 優先権主張番号	特願2014-831 (P2014-831)		東京都港区港南1丁目7番1号 ソニー株
(32) 優先日	平成26年1月7日 (2014.1.7)		式会社内
(33) 優先権主張国	日本国 (JP)		

最終頁に続く

(54) 【発明の名称】 情報処理装置および方法、並びにプログラム

(57) 【特許請求の範囲】

【請求項 1】

第1のコンテンツの音声信号から特徴量を抽出する特徴量計算部と、
前記第1のコンテンツと時間同期関係を有する第2のコンテンツに対して同期がとれた
状態で取得された前記特徴量と、前記特徴量計算部により抽出された前記特徴量とを比較
することで、前記第2のコンテンツを前記第1のコンテンツと同期して再生するための音
声特徴量に基づく同期補正情報を生成する同期計算部と
を備え、
前記同期計算部は、取得された前記特徴量と、前記特徴量計算部により抽出された前記
特徴量とのフレームレートが一致するように、取得された前記特徴量、または前記特徴量
計算部により抽出された前記特徴量の少なくとも一方に対してフレームレート変換を行っ
てから前記特徴量を比較する

情報処理装置。

【請求項 2】

前記特徴量計算部は、再生された前記第1のコンテンツの音声を受音することで得られ
た前記音声信号から前記特徴量を抽出する

請求項1に記載の情報処理装置。

【請求項 3】

前記第2のコンテンツと、前記第2のコンテンツに同期がとれた状態で前記第2のコン
テンツに対応付けられている前記特徴量とを取得する第1の入力部をさらに備える

請求項 1 または請求項 2 に記載の情報処理装置。

【請求項 4】

前記第 2 のコンテンツおよび前記特徴量は、前記第 1 のコンテンツとの到着時間差が考慮されたタイミングで前記情報処理装置に送信される

請求項 3 に記載の情報処理装置。

【請求項 5】

前記第 1 の入力部は、前記第 2 のコンテンツおよび前記特徴量の送信を要求し、その要求に応じて送信されてきた前記第 2 のコンテンツおよび前記特徴量を受信する

請求項 3 に記載の情報処理装置。

【請求項 6】

前記第 1 のコンテンツの送信を要求し、その要求に応じて送信されてきた前記第 1 のコンテンツを受信する第 2 の入力部をさらに備える

請求項 5 に記載の情報処理装置。

【請求項 7】

前記特徴量計算部は、再生された 1 つの前記第 1 のコンテンツについて、前記音声信号から前記特徴量を抽出し、

前記同期計算部は、前記第 2 のコンテンツと対応付けられている複数の前記第 1 のコンテンツの前記特徴量のそれぞれと、前記特徴量計算部により抽出された前記特徴量とを比較することで、再生された前記第 1 のコンテンツを特定し、特定された前記第 1 のコンテンツと前記第 2 のコンテンツを同期して再生するための前記音声特徴量に基づく同期補正情報を生成する

請求項 2 乃至請求項 6 の何れか一項に記載の情報処理装置。

【請求項 8】

前記第 2 のコンテンツの再生を制御する再生処理部をさらに備える

請求項 2 乃至請求項 7 の何れか一項に記載の情報処理装置。

【請求項 9】

前記再生処理部は、前記音声特徴量に基づく同期補正情報に基づいて前記第 2 のコンテンツの再生位置を補正する

請求項 8 に記載の情報処理装置。

【請求項 10】

前記第 1 のコンテンツの提示時刻情報を取得する取得部と、

前記第 1 のコンテンツの前記提示時刻情報と、前記第 2 のコンテンツの前記提示時刻情報とを比較して、提示時刻情報に基づく同期補正情報を生成する比較部と

をさらに備え、

前記同期計算部は、取得された前記特徴量の系列のうちの前記提示時刻情報に基づく同期補正情報により示される範囲に含まれる前記特徴量と、前記特徴量計算部により抽出された前記特徴量とを比較して前記音声特徴量に基づく同期補正情報を生成する

請求項 1 乃至請求項 9 の何れか一項に記載の情報処理装置。

【請求項 11】

第 1 のコンテンツの音声信号から特徴量を抽出する特徴量計算ステップと、

前記第 1 のコンテンツと時間同期関係を有する第 2 のコンテンツに対して同期がとれた状態で取得された前記特徴量と、前記特徴量計算ステップの処理により抽出された前記特徴量とを比較することで、前記第 2 のコンテンツを前記第 1 のコンテンツと同期して再生するための音声特徴量に基づく同期補正情報を生成する同期計算ステップと

を含み、

前記同期計算ステップにおいて、取得された前記特徴量と、前記特徴量計算ステップの処理により抽出された前記特徴量とのフレームレートが一致するように、取得された前記特徴量、または前記特徴量計算ステップの処理により抽出された前記特徴量の少なくとも一方に対してフレームレート変換を行ってから前記特徴量を比較する

情報処理方法。

10

20

30

40

50

【請求項 1 2】

第 1 のコンテンツの音声信号から特徴量を抽出する特徴量計算ステップと、

前記第 1 のコンテンツと時間同期関係を有する第 2 のコンテンツに対して同期がとれた状態で取得された前記特徴量と、前記特徴量計算ステップの処理により抽出された前記特徴量とを比較することで、前記第 2 のコンテンツを前記第 1 のコンテンツと同期して再生するための音声特徴量に基づく同期補正情報を生成する同期計算ステップと

を含む処理をコンピュータに実行させ、

前記同期計算ステップにおいて、取得された前記特徴量と、前記特徴量計算ステップの処理により抽出された前記特徴量とのフレームレートが一致するように、取得された前記特徴量、または前記特徴量計算ステップの処理により抽出された前記特徴量の少なくとも一方に対してフレームレート変換を行ってから前記特徴量を比較する

10

プログラム。

【請求項 1 3】

第 1 のコンテンツの音声信号から特徴量を抽出し、抽出された前記特徴量をダウンサンプルする特徴量計算部と、

前記第 1 のコンテンツに対して時間同期関係を有する第 2 のコンテンツと、前記第 2 のコンテンツに対して同期がとれた状態で、前記第 2 のコンテンツに対応付けられた、前記ダウンサンプルされた前記特徴量とを出力する第 1 の出力部と

を備える情報処理装置。

20

【請求項 1 4】

前記第 1 のコンテンツを出力する第 2 の出力部をさらに備える

請求項 1 3 に記載の情報処理装置。

【請求項 1 5】

前記第 1 の出力部は、前記第 1 のコンテンツとの到着時間差が考慮されたタイミングで前記第 2 のコンテンツおよび前記ダウンサンプルされた前記特徴量を出力する

請求項 1 4 に記載の情報処理装置。

【請求項 1 6】

前記第 1 の出力部は、前記第 2 のコンテンツおよび前記ダウンサンプルされた前記特徴量の送信が要求された場合、その要求に応じて前記第 2 のコンテンツおよび前記ダウンサンプルされた前記特徴量を出力する

30

請求項 1 4 に記載の情報処理装置。

【請求項 1 7】

前記第 2 の出力部は、前記第 1 のコンテンツの送信が要求された場合、その要求に応じて前記第 1 のコンテンツを出力する

請求項 1 6 に記載の情報処理装置。

【請求項 1 8】

前記特徴量計算部は、複数の前記第 1 のコンテンツについて、前記音声信号から前記特徴量を抽出し、

前記第 1 の出力部は、複数の前記第 1 のコンテンツの前記ダウンサンプルされた前記特徴量を前記第 2 のコンテンツに対応付けて出力する

40

請求項 1 3 乃至請求項 1 7 の何れか一項に記載の情報処理装置。

【請求項 1 9】

第 1 のコンテンツの音声信号から特徴量を抽出し、抽出された前記特徴量をダウンサンプルする特徴量計算ステップと、

前記第 1 のコンテンツに対して時間同期関係を有する第 2 のコンテンツと、前記第 2 のコンテンツに対して同期がとれた状態で、前記第 2 のコンテンツに対応付けられた、前記ダウンサンプルされた前記特徴量とを出力する出力ステップと

を含む情報処理方法。

【請求項 2 0】

第 1 のコンテンツの音声信号から特徴量を抽出し、抽出された前記特徴量をダウンサン

50

プルする特徴量計算ステップと、

前記第 1 のコンテンツに対して時間同期関係を有する第 2 のコンテンツと、前記第 2 のコンテンツに対して同期がとれた状態で、前記第 2 のコンテンツに対応付けられた、前記ダウンサンプルされた前記特徴量とを出力する出力ステップと

を含む処理をコンピュータに実行させるプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本技術は情報処理装置および方法、並びにプログラムに関し、特に、異なる経路で取得した複数のコンテンツを同期させることができるようにした情報処理装置および方法、並びにプログラムに関する。

10

【背景技術】

【0002】

近年、多機能型携帯電話機やタブレット型端末装置など、ネットワーク接続を前提とし、様々なメディアコンテンツを再生できる機器が増加してきている。さらに、従来から存在しているテレビジョン受像機なども含め、ネットワーク機能を用いた複数機器の連携による利活用が求められている。

【0003】

例えば複数機器の連携では、以下に示す(A1)乃至(A4)などの時間同期関係を有する複数のメディアコンテンツを、放送やインターネットなどを通じて複数機器で受信し、それらのコンテンツを同期して再生するというアプリケーションプログラムが想定される。

20

【0004】

(A1)メインの映像・音声コンテンツに対する外国語音声コンテンツ、解説音声コンテンツ、クローズドキャプションと文字情報

(A2)ある楽曲を楽器毎に演奏、撮影した複数の映像・音声コンテンツ

(A3)1つのシーンを複数の角度から撮影した映像・音声コンテンツ

(A4)メインの映像・音声コンテンツとその高解像版の映像・音声コンテンツ

【0005】

このような複数のコンテンツは、再生時には同期が保たれた状態で再生させる必要がある。例えば、複数のコンテンツを同期させる技術として、異なる複数の撮影装置で同時に撮影された各コンテンツから特徴量を抽出し、それらの特徴量の類似度を計算することで、複数のコンテンツを同期させる技術が開示されている(例えば、特許文献1参照)。

30

【先行技術文献】

【特許文献】

【0006】

【特許文献1】特開2013-174765号公報

【発明の概要】

【発明が解決しようとする課題】

【0007】

ところで、実際には、上記のようなメディアコンテンツを、複数の機器がそれぞれ異なる経路で受信しようとする場合、伝送遅延や、送出および受信の処理遅延、受信機器の動作クロックの違いなどの要因により、同期を保ってコンテンツを再生することが困難である。また、特許文献1に記載の技術では、同期をとって再生しようとするコンテンツ同士が類似した特徴を有していない場合には、それらのコンテンツの同期をとることができなかった。

40

【0008】

本技術は、このような状況に鑑みてなされたものであり、異なる経路で取得した複数のコンテンツを同期させることができるようにするものである。

【課題を解決するための手段】

50

【 0 0 0 9 】

本技術の第 1 の側面の情報処理装置は、第 1 のコンテンツの音声信号から特徴量を抽出する特徴量計算部と、前記第 1 のコンテンツと時間同期関係を有する第 2 のコンテンツに対して同期がとれた状態で取得された前記特徴量と、前記特徴量計算部により抽出された前記特徴量とを比較することで、前記第 2 のコンテンツを前記第 1 のコンテンツと同期して再生するための音声特徴量に基づく同期補正情報を生成する同期計算部とを備え、前記同期計算部は、取得された前記特徴量と、前記特徴量計算部により抽出された前記特徴量とのフレームレートが一致するように、取得された前記特徴量、または前記特徴量計算部により抽出された前記特徴量の少なくとも一方に対してフレームレート変換を行ってから前記特徴量を比較する。

10

【 0 0 1 0 】

前記特徴量計算部には、再生された前記第 1 のコンテンツの音声を收音することで得られた前記音声信号から前記特徴量を抽出させることができる。

【 0 0 1 1 】

情報処理装置には、前記第 2 のコンテンツと、前記第 2 のコンテンツに同期がとれた状態で前記第 2 のコンテンツに対応付けられている前記特徴量とを取得する第 1 の入力部をさらに設けることができる。

【 0 0 1 2 】

前記第 2 のコンテンツおよび前記特徴量は、前記第 1 のコンテンツとの到着時間差が考慮されたタイミングで前記情報処理装置に送信されるようにすることができる。

20

【 0 0 1 3 】

前記第 1 の入力部には、前記第 2 のコンテンツおよび前記特徴量の送信を要求させ、その要求に応じて送信されてきた前記第 2 のコンテンツおよび前記特徴量を受信させることができる。

【 0 0 1 4 】

情報処理装置には、前記第 1 のコンテンツの送信を要求し、その要求に応じて送信されてきた前記第 1 のコンテンツを受信する第 2 の入力部をさらに設けることができる。

【 0 0 1 5 】

前記特徴量計算部には、再生された 1 つの前記第 1 のコンテンツについて、前記音声信号から前記特徴量を抽出させ、前記同期計算部には、前記第 2 のコンテンツと対応付けられている複数の前記第 1 のコンテンツの前記特徴量のそれぞれと、前記特徴量計算部により抽出された前記特徴量とを比較することで、再生された前記第 1 のコンテンツを特定させ、特定された前記第 1 のコンテンツと前記第 2 のコンテンツを同期して再生するための前記音声特徴量に基づく同期補正情報を生成させることができる。

30

【 0 0 1 6 】

情報処理装置には、前記第 2 のコンテンツの再生を制御する再生処理部をさらに設けることができる。

【 0 0 1 7 】

前記再生処理部には、前記音声特徴量に基づく同期補正情報に基づいて前記第 2 のコンテンツの再生位置を補正させることができる。

40

【 0 0 1 8 】

情報処理装置には、前記第 1 のコンテンツの提示時刻情報を取得する取得部と、前記第 1 のコンテンツの前記提示時刻情報と、前記第 2 のコンテンツの前記提示時刻情報とを比較して、提示時刻情報に基づく同期補正情報を生成する比較部とをさらに設け、前記同期計算部には、取得された前記特徴量の系列のうちの前記提示時刻情報に基づく同期補正情報により示される範囲に含まれる前記特徴量と、前記特徴量計算部により抽出された前記特徴量とを比較して前記音声特徴量に基づく同期補正情報を生成させることができる。

【 0 0 2 0 】

本技術の第 1 の側面の情報処理方法またはプログラムは、第 1 のコンテンツの音声信号から特徴量を抽出する特徴量計算ステップと、前記第 1 のコンテンツと時間同期関係を有

50

する第2のコンテンツに対して同期がとれた状態で取得された前記特徴量と、前記特徴量計算ステップの処理により抽出された前記特徴量とを比較することで、前記第2のコンテンツを前記第1のコンテンツと同期して再生するための音声特徴量に基づく同期補正情報を生成する同期計算ステップとを含み、前記同期計算ステップにおいて、取得された前記特徴量と、前記特徴量計算ステップの処理により抽出された前記特徴量とのフレームレートが一致するように、取得された前記特徴量、または前記特徴量計算ステップの処理により抽出された前記特徴量の少なくとも一方に対してフレームレート変換を行ってから前記特徴量を比較する。

【0021】

本技術の第1の側面においては、第1のコンテンツの音声信号から特徴量が抽出され、前記第1のコンテンツと時間同期関係を有する第2のコンテンツに対して同期がとれた状態で取得された前記特徴量と、抽出された前記特徴量とを比較することで、前記第2のコンテンツを前記第1のコンテンツと同期して再生するための音声特徴量に基づく同期補正情報が生成される。また、前記特徴量の比較時には、取得された前記特徴量と、抽出された前記特徴量とのフレームレートが一致するように、取得された前記特徴量、または抽出された前記特徴量の少なくとも一方に対してフレームレート変換が行われてから前記特徴量が比較される。

【0022】

本技術の第2の側面の情報処理装置は、第1のコンテンツの音声信号から特徴量を抽出し、抽出された前記特徴量をダウンサンプルする特徴量計算部と、前記第1のコンテンツに対して時間同期関係を有する第2のコンテンツと、前記第2のコンテンツに対して同期がとれた状態で、前記第2のコンテンツに対応付けられた、前記ダウンサンプルされた前記特徴量とを出力する第1の出力部とを備える。

【0023】

情報処理装置には、前記第1のコンテンツを出力する第2の出力部をさらに設けることができる。

【0024】

前記第1の出力部には、前記第1のコンテンツとの到着時間差が考慮されたタイミングで前記第2のコンテンツおよび前記ダウンサンプルされた前記特徴量を出力させることができる。

【0025】

前記第1の出力部には、前記第2のコンテンツおよび前記ダウンサンプルされた前記特徴量の送信が要求された場合、その要求に応じて前記第2のコンテンツおよび前記ダウンサンプルされた前記特徴量を出力させることができる。

【0026】

前記第2の出力部には、前記第1のコンテンツの送信が要求された場合、その要求に応じて前記第1のコンテンツを出力させることができる。

【0027】

前記特徴量計算部には、複数の前記第1のコンテンツについて、前記音声信号から前記特徴量を抽出させ、前記第1の出力部には、複数の前記第1のコンテンツの前記ダウンサンプルされた前記特徴量を前記第2のコンテンツに対応付けて出力させることができる。

【0029】

本技術の第2の側面の情報処理方法またはプログラムは、第1のコンテンツの音声信号から特徴量を抽出し、抽出された前記特徴量をダウンサンプルする特徴量計算ステップと、前記第1のコンテンツに対して時間同期関係を有する第2のコンテンツと、前記第2のコンテンツに対して同期がとれた状態で、前記第2のコンテンツに対応付けられた、前記ダウンサンプルされた前記特徴量とを出力する出力ステップとを含む。

【0030】

本技術の第2の側面においては、第1のコンテンツの音声信号から特徴量が抽出され、抽出された前記特徴量がダウンサンプルされ、前記第1のコンテンツに対して時間同期関

10

20

30

40

50

係を有する第2のコンテンツと、前記第2のコンテンツに対して同期がとれた状態で、前記第2のコンテンツに対応付けられた、前記ダウンサンプルされた前記特徴量とが出力される。

【発明の効果】

【0031】

本技術の第1の側面および第2の側面によれば、異なる経路で取得した複数のコンテンツを同期させることができる。

【0032】

なお、ここに記載された効果は必ずしも限定されるものではなく、本開示中に記載された何れかの効果であってもよい。

10

【図面の簡単な説明】

【0033】

【図1】提供装置の構成例を示す図である。

【図2】音声同期用特徴量計算部の構成例を示す図である。

【図3】音声同期用特徴量のダウンサンプルについて説明する図である。

【図4】コンテンツ再生システムの構成例を示す図である。

【図5】音声同期用特徴量計算部の構成例を示す図である。

【図6】同期計算部の構成例を示す図である。

【図7】音声同期用特徴量の同期計算について説明する図である。

【図8】音声同期用特徴量の同期計算について説明する図である。

20

【図9】音声同期用特徴量の同期計算について説明する図である。

【図10】類似度の計算の対象とするブロックについて説明する図である。

【図11】類似度の計算について説明する図である。

【図12】送信処理を説明するフローチャートである。

【図13】サブチャンネル信号と音声同期用特徴量の多重化について説明する図である。

【図14】音声同期用特徴量算出処理を説明するフローチャートである。

【図15】メインコンテンツ再生処理を説明するフローチャートである。

【図16】サブコンテンツ再生処理を説明するフローチャートである。

【図17】音声同期用特徴量算出処理を説明するフローチャートである。

【図18】同期補正情報生成処理を説明するフローチャートである。

30

【図19】本技術の適用例を示す図である。

【図20】本技術の適用例を示す図である。

【図21】本技術の適用例を示す図である。

【図22】提供装置の構成例を示す図である。

【図23】コンテンツ再生システムの構成例を示す図である。

【図24】送信処理を説明するフローチャートである。

【図25】メインコンテンツ再生処理を説明するフローチャートである。

【図26】サブコンテンツ再生処理を説明するフローチャートである。

【図27】同期補正情報生成処理を説明するフローチャートである。

【図28】類似度の計算の対象とするブロックについて説明する図である。

40

【図29】コンピュータの構成例を示す図である。

【発明を実施するための形態】

【0034】

以下、図面を参照して、本技術を適用した実施の形態について説明する。

【0035】

第1の実施の形態

本技術の特徴

まず、本技術の特徴について説明する。

【0036】

本技術は、特に以下の特徴B1乃至特徴B6を有している。

50

【 0 0 3 7 】

(特徴 B 1)

本技術によれば、内容の異なる複数のメディアコンテンツを異なる伝送路経由で送信し、異なる複数の機器で受信する際に音声を使って自動同期を行う下記の構成を有する方法および装置を実現することができる。

(1) メディアコンテンツは映像、音声、画像、文字情報などを多重化したデータストリームとされている

(2) 伝送対象とする複数メディアコンテンツは、上述した (A 1) 乃至 (A 4) に示した例のように時間同期関係を有する

(3) 送出対象とする複数のメディアコンテンツのうちの少なくとも 1 つをメインチャンネル信号と定め、その音声信号から音声同期用特徴量を算出し、システムが規定する伝送フォーマットによりメインチャンネル信号からメイン送出信号を生成する

(4) 残りの各メディアコンテンツ (サブチャンネル信号) とメインチャンネル信号の時間同期関係が符合するように、システムが規定する伝送フォーマットによりメインチャンネル信号の音声同期用特徴量とサブチャンネル信号の多重化処理を行い、サブ送出信号を生成する

(5) メイン送出信号を受信するメイン受信機器では、メインチャンネル信号の再生時において、その音声信号をスピーカなどにより出力する

(6) メインチャンネル信号の音声同期用特徴量を含むサブ送出信号を受信するサブ受信機器は、メイン受信機器がスピーカ出力したメインチャンネル信号の音声をマイクロホンなどにより収音して、音声同期用特徴量を計算し、受信したメインチャンネル信号の音声同期用特徴量との自動同期計算を行い、音声特徴量に基づく同期補正情報 (時間差情報) を算出する

(7) 上記音声特徴量に基づく同期補正情報に基づき、サブ受信機器は受信したサブチャンネル信号に対してメインチャンネル信号との同期補正処理を行い再生する

【 0 0 3 8 】

なお、上記の (1) のデータストリームの伝送としては、放送波、インターネットなどのネットワークにおけるメディアコンテンツの伝送を想定し、多重化データストリームが占有する論理伝送路を伝送路と呼ぶこととする。

【 0 0 3 9 】

また、上記にある「音声同期用特徴量の算出」と「自動同期計算」は、例えば特開 2 0 1 3 - 1 7 4 7 6 5 号公報に記載されている技術により実現される。なお、音声同期用特徴量を伝送前にダウンサンプルしたり、音声同期用特徴量を利用した自動同期計算時に、必要に応じて音声同期用特徴量のフレームレート変換を行うようにすることもできる。

【 0 0 4 0 】

このような技術を用いることにより、メインチャンネル信号の音声をサブ受信機器が収音する際に、雑音や騒音があるような悪環境であってもロバストに自動同期計算を行うことができる。なお、必ずしもこの技術を用いる必要はない。

【 0 0 4 1 】

(特徴 B 2)

上記の (特徴 B 1) において、送出側システムがメイン送出信号とサブ送出信号を各々メイン受信機器、サブ受信機器に対して一方的に送出する。

【 0 0 4 2 】

この場合、サブ送出信号はメイン送出信号に先んじて送出する必要がある。

【 0 0 4 3 】

(特徴 B 3)

上記の (特徴 B 1) において、送出側システムがメイン送出信号をメイン受信機器に対して一方的に送出し、サブ受信機器は自身のタイミングによりネットワーク経由などでサブ送出信号を取得して自動同期計算を行い、サブチャンネル信号の同期再生を行う。

【 0 0 4 4 】

この構成の利点としては、ネットワークの伝送遅延などを加味してサブ受信機器が自身の都合によりサブ送出信号の取得を制御できる。

【 0 0 4 5 】

(特徴 B 4)

上記の (特徴 B 1) において、メイン受信機器は、自身のタイミングによりネットワーク経由などでメイン送出信号を取得して、メインチャンネル信号の再生を行い、サブ受信機器も自身のタイミングによりネットワーク経由などでサブ送出信号を取得して自動同期計算を行い、サブチャンネル信号の同期再生を行う。

【 0 0 4 6 】

この構成の利点としては、ネットワークの伝送遅延などを加味してサブ受信機器が自身の都合によりサブ送出信号の取得を制御できる。

10

【 0 0 4 7 】

(特徴 B 5)

上記の (特徴 B 1) において、メインチャンネル信号の音声信号が複数系統ある。

【 0 0 4 8 】

例えば複数系統のメインチャンネル信号は、2か国語放送の主音声と副音声などとされる。全ての系統の音声信号に対して音声同期用特徴量が算出され、サブチャンネル信号と多重化されて送出される。サブ受信機器では、収音した音声と受信した全ての音声同期用特徴量との同期計算を行う際に、メインチャンネル信号のどの音声か再生されているかが判別される。メイン受信機器が出力する音声信号の切り替えも上記同期計算により検出される。

20

【 0 0 4 9 】

(特徴 B 6)

上記の (特徴 B 1) において、サブ受信機器における自動同期計算において「同期ずれ」を検出し、サブ受信機器側でリアルタイム補正処理を行う。

【 0 0 5 0 】

メイン受信機器とサブ受信機器は独立して動作するため、オーディオクロックが異なり同期ずれが発生する。そこで、その同期ずれを検出して補正することで、同期を保ったまま複数のコンテンツを再生することが可能となる。

【 0 0 5 1 】

30

提供装置の構成例

次に、本技術を適用した具体的な実施の形態について説明する。

【 0 0 5 2 】

まず、上述した (A 1) 乃至 (A 4) に示した例のように時間同期関係を有するコンテンツを提供する提供装置の構成例について説明する。

【 0 0 5 3 】

図1は、提供装置の構成例を示す図である。この提供装置11には、主となるコンテンツ(以下、メインコンテンツと称する)を再生するための信号であるメインチャンネル信号と、メインコンテンツと内容的な関連性を有するコンテンツ(以下、サブコンテンツと称する)を再生するための信号であるサブチャンネル信号とが供給される。

40

【 0 0 5 4 】

ここで、メインコンテンツとサブコンテンツは、少なくとも映像と音声の何れかからなり、互いに時間同期関係を有するコンテンツである。つまり、再生時には、メインコンテンツとサブコンテンツが同期した状態で再生されることが望ましい。

【 0 0 5 5 】

なお、以下では、メインコンテンツおよびサブコンテンツは、それぞれ映像を再生する画像信号と、その画像信号に付随する音声信号とからなるものとして説明を続ける。したがって、この例ではメインチャンネル信号とサブチャンネル信号は、それぞれ画像信号と音声信号から構成されている。

【 0 0 5 6 】

50

提供装置 11 は、変換部 21、出力部 22、音声同期用特徴量計算部 23、多重化処理部 24、および出力部 25 を有している。

【0057】

変換部 21 は、供給されたメインチャンネル信号を、所定の放送規格などで定められたフォーマットに変換し、その結果得られたメイン送出信号を出力部 22 に供給する。出力部 22 は、変換部 21 から供給されたメイン送出信号を、例えば放送波により送信したり、インターネットなどの通信網を介して送信したりする。

【0058】

音声同期用特徴量計算部 23 は、供給されたメインチャンネル信号を構成する音声信号から、音声同期用特徴量を抽出し、多重化処理部 24 に供給する。ここで、音声同期用特徴量は、メインコンテンツとサブコンテンツとの再生時に、サブコンテンツをメインコンテンツに同期させて再生させるために用いられる特徴量である。

【0059】

多重化処理部 24 は、供給されたメインチャンネル信号を用いて、音声同期用特徴量計算部 23 からの音声同期用特徴量と、供給されたサブチャンネル信号との時間同期関係を調整する。すなわち、提供装置 11 では、予めメインチャンネル信号とサブチャンネル信号とは同期がとれた状態となっているので、多重化処理部 24 は、メインチャンネル信号を用いて、時間同期関係において音声同期用特徴量とサブチャンネル信号とが同期した状態で、音声同期用特徴量をサブチャンネル信号に対応付ける。例えば MPEG-4 System において、オーディオ信号、ビデオ信号などはそれぞれ 1 つのメディアオブジェクト (ES (Elementary Stream)) として扱われ、多重化される。ES を分割して得られる Access Unit (AU) と呼ばれる最小単位毎に時間属性が定義されるので、音声同期用特徴量も時間属性情報を有する 1 つのメディアオブジェクトとして扱うことにより、サブチャンネル信号であるメディアオブジェクトと容易に多重化を行うことができる。

【0060】

また、多重化処理部 24 は、時間的に同期がとれた状態で音声同期用特徴量とサブチャンネル信号とを多重化した後、必要に応じてフォーマット変換を行って、その結果得られたサブ送出信号を出力部 25 に供給する。

【0061】

出力部 25 は、多重化処理部 24 から供給されたサブ送出信号を、例えば放送波により送信したり、インターネットなどの通信網を介して送信したりする。ここで、メイン送出信号とサブ送出信号とは、互いに異なる伝送路を介してコンテンツの再生側のシステムに送信される。

【0062】

なお、図 1 の例では、提供装置 11 は、1 つの装置により構成されているが、提供装置 11 が複数の装置により構成されてもよいし、各処理がクラウドコンピューティングにより実行されるようにしてもよい。

【0063】

音声同期用特徴量計算部の構成例

また、図 1 に示した音声同期用特徴量計算部 23 は、より詳細には例えば図 2 に示すように構成される。

【0064】

音声同期用特徴量計算部 23 は、周波数帯域分割部 51、周期性検出部 52 - 1 乃至周期性検出部 52 - 4、周期性強度検出部 53 - 1 乃至周期性強度検出部 53 - 4、周期性情報統合部 54、ピーク検出部 55、およびダウンサンプル部 56 を有している。

【0065】

周波数帯域分割部 51 は、供給されたメインチャンネル信号を構成する音声信号を、窓関数を用いて数 10 msec 乃至 100 msec 程度の時間区間に分割する。

【0066】

ここで、周波数帯域分割部 51 からピーク検出部 55 までの間で行われる処理は、1 つ

10

20

30

40

50

の時間区間に対して行われる。このとき窓関数を適用する時間位置を数msec乃至100msec程度後にシフトさせることで時間方向に連続した複数の時間区間（時間フレーム）を得ることができる。これに対してダウンサンプル部56では、連続した複数の時間区間の結果が1つに統合され、統合後の新たな時間区間に対する特徴量が算出される。

【0067】

周波数帯域分割部51は、複数のバンドパスフィルタを用いて、時間区間ごとの音声信号を4つの周波数帯域に分割し、各周波数帯域の音声信号を、周期性検出部52-1乃至周期性検出部52-4のそれぞれに供給する。

【0068】

なお、バンドパスフィルタとしては、例えばオクターブバンドフィルタなどの、高い周波数ほど通過周波数帯域幅が広がるフィルタを用いると効果的である。

【0069】

周期性検出部52-1乃至周期性検出部52-4は、周波数帯域分割部51から供給された所定の周波数帯域の時間区間ごとの音声信号の自己相関関数を計算することにより、時間区間ごとの周期性を表す周期性情報を抽出する。

【0070】

なお、ここでは、周期性情報として、インデックスが b である周波数帯域の、インデックスが τ である時間遅れの音声信号の自己相関関数 $x(b, \tau)$ そのものを用いるが、自己相関関数 $x(b, \tau)$ を $x(b, 0)$ で除算した値を用いることもできる。また、自己相関関数 $x(b, \tau)$ の計算方法としては、所定の周波数帯域の音声信号に対して離散フーリエ変換を行うことにより得られるスペクトルのピークを用いた方法などを用いることができる。

【0071】

周期性検出部52-1乃至周期性検出部52-4は、抽出された時間区間ごとの周期性情報を周期性強度検出部53-1乃至周期性強度検出部53-4と周期性情報統合部54に供給する。なお、以下、周期性検出部52-1乃至周期性検出部52-4を特に区別する必要がない場合、単に周期性検出部52と称することとする。

【0072】

周期性強度検出部53-1乃至周期性強度検出部53-4は、周期性検出部52-1乃至周期性検出部52-4から供給される時間区間ごとの周期性情報に基づいて、時間区間ごとの周期性の強度を計算する。具体的には、 $\tau = 0$ 近傍以外の τ における周期性情報である自己相関関数 $x(b, \tau)$ の最大値が周期性の強度として計算される。この周期性の強度が大きいほど、処理対象の音声信号の周期性が強く、周期性の強度が小さいほど、処理対象の音声信号の周期性がノイズらしくなる。

【0073】

周期性強度検出部53-1乃至周期性強度検出部53-4は、時間区間ごとの周期性の強度を、閾値を超えたか否かにより2値化し、時間区間ごとの周期性強度情報とする。すなわち、時間区間ごとに、周期性の強度が所定の閾値を超えた場合、周期性強度情報は1とされ、周期性の強度が所定の閾値以下である場合、周期性強度情報は0とされる。周期性強度検出部53-1乃至周期性強度検出部53-4は、時間区間ごとの周期性強度情報を周期性情報統合部54に供給する。

【0074】

なお、以下、周期性強度検出部53-1乃至周期性強度検出部53-4を特に区別する必要がない場合、単に周期性強度検出部53とも称する。

【0075】

周期性情報統合部54は、周期性検出部52から供給された時間区間ごとの周期性情報と、周期性強度検出部53から供給された時間区間ごとの周期性強度情報とに基づいて、時間区間ごとの周期性情報を統合する周期性統合処理を行う。具体的には、周期性情報統合部54は、次式(1)を用いて時間区間ごとに周期性情報である自己相関関数 $x(b, \tau)$ の総和を求める。

10

20

30

40

50

【 0 0 7 6 】

【 数 1 】

$$s(\tau) = \frac{1}{N_p} \sum_{b=1}^{N_b} x(b, \tau) \cdot p(b) \quad \dots (1)$$

【 0 0 7 7 】

なお、式 (1) において、 N_b は周波数帯域の総数を表し、 $p(b)$ は周期性強度情報を表す。また、 N_p は周期性強度情報 $p(b)$ が 1 となる周波数帯域の数を表す。

【 0 0 7 8 】

周期性情報統合部 5 4 は、周期性統合処理の結果得られる時間区間ごとの周期性情報の総和 $S()$ をピーク検出部 5 5 に供給する。

【 0 0 7 9 】

ピーク検出部 5 5 は、時間区間ごとに、周期性情報統合部 5 4 から供給された周期性情報の総和 $S()$ に対してピーク検出を行い、ピーク位置 p の値が 1 となり、ピーク位置 p 以外の値が 0 となるピーク情報 $P()$ を生成する。ピーク検出方法としては、例えば、周期性情報の総和 $S()$ の微分値が正から負に変わるときのインデックス i を、ピーク位置 p として検出する方法がある。

【 0 0 8 0 】

なお、ピーク検出部 5 5 は、ピーク位置 p の周期性情報の総和 $S(p)$ が、所定の閾値より小さい場合、そのピーク位置 p のピーク情報 $P(p)$ を 0 とするようにしてもよい。これにより、ピーク情報 $P(p)$ のノイズを低減することができる。また、ピーク情報は、周期性情報の総和 $S()$ そのものであってもよい。

【 0 0 8 1 】

ピーク検出部 5 5 は、時間区間ごとのピーク情報 $P()$ を、時間区間ごとの音声同期用特徴量の時系列データとしてダウンサンプル部 5 6 に供給する。

【 0 0 8 2 】

ダウンサンプル部 5 6 は、ピーク検出部 5 5 から供給された複数の時間区間の音声同期用特徴量、つまり複数の時間区間のピーク情報 $P()$ を新たな 1 つの時間区間の情報として統合し、最終的な音声同期用特徴量としてのピーク情報 $P'_{i}()$ を生成する。換言すれば、ダウンサンプル部 5 6 は、ピーク情報 $P()$ をダウンサンプルすることでピーク情報 $P'_{i}()$ を生成する。

【 0 0 8 3 】

なお、 $P'_{i}()$ において i は時間遅れを示すインデックスであり、 i は時間区間を示すインデックスである。ダウンサンプル部 5 6 は、このようにして得られた時間区間ごとのピーク情報 $P'_{i}()$ を、時間区間ごとの音声同期用特徴量の時系列データとして多重化処理部 2 4 に供給する。

【 0 0 8 4 】

ここで、図 3 を参照してピーク情報 $P'_{i}()$ の生成について説明する。なお、図 3 において、縦軸は時間遅れを示すインデックス i を示しており、横軸は時間、すなわち時間区間を示すインデックス i を示している。

【 0 0 8 5 】

この例では図中、上側にはピーク情報 $P()$ の系列が示されており、図中、下側にはピーク情報 $P'_{i}()$ の系列が示されている。特に、図 3 では時間遅れが i であり、インデックス i により特定される時間区間のピーク情報 $P()$ が $P'_{i}()$ で表されている。また、各四角形は 1 つの時間区間のピーク情報を表している。特に、白色の四角形は、その四角形により表されるピーク情報が 0 であることを表しており、黒色の四角形は、その四角形により表されるピーク情報が 1 であることを表している。

【 0 0 8 6 】

図中、上側に示されるピーク情報 $P'_{i}()$ の時間区間の長さは 8msec とされている。

10

20

30

40

50

つまり、ピーク情報 P_i () は8msecの時間間隔で算出されている。そして、ここでは時間遅れ が同じであり、時間方向(時間区間方向)に隣接する4つのピーク情報 P_i () が1つに統合され、1つのピーク情報 P'_i () とされる。したがって、1つのピーク情報 P'_i () の時間区間は32msecとなる。

【0087】

例えばダウンサンプル部56は、次式(2)を計算することでピーク情報 P_i () を統合(ダウンサンプル)し、ピーク情報 P'_i () とする。

【0088】

【数2】

$$P'_i(\tau) = P_{4i}(\tau) \cup P_{4i+1}(\tau) \cup P_{4i+2}(\tau) \cup P_{4i+3}(\tau) \cdots (2)$$

10

【0089】

この式(2)の計算では、統合される4つの連続するピーク情報 P_i () のうち、1つでも値が「1」であるピーク情報 P_i () がある場合、統合により得られたピーク情報 P'_i () の値は「1」とされる。逆に、統合される4つの連続するピーク情報 P_i () の値が全て「0」である場合に、統合により得られたピーク情報 P'_i () の値は「0」とされる。

【0090】

このように時間区間方向に連続して並ぶピーク情報 P_i () の論理和を求めてダウンサンプルを行うことで、時間方向に並ぶピーク情報の系列に含まれているピーク位置に関する情報がダウンサンプルにより除去されてしまうことがなくなる。これにより、ダウンサンプル後においても、時間遅れ方向においてピーク位置に関する情報がどのように遷移していくかを保持することが可能となる。

20

【0091】

例えば、ピーク情報 P_i () をダウンサンプルする際に、単純に複数の時間区間のピーク情報 P_i () のうちの1つのピーク情報 P_i () の値を、ダウンサンプル後のピーク情報 P'_i () の値として採用した場合、情報量が欠落し、同期計算の精度が低下する。つまり性能が劣化する。

【0092】

具体的には、例えば4つの時間区間にわたってピーク位置が遷移している場合、単純にピーク情報 P_i () を間引いてピーク情報 P'_i () とすると、遷移途中の1つの時間区間のピーク情報 P_i () のみが最終的な特徴量として採用され、ピーク位置が遷移した情報は失われてしまう。

30

【0093】

一方、上述したようにダウンサンプルの際に複数の時間区間のピーク情報 P_i () に基づいて適切な値をピーク情報 P'_i () として出力する手法では、ダウンサンプル後の時間区間の中で遷移が発生しているという情報を、ダウンサンプル後の1時間区間に保持することができる。結果として、ダウンサンプル処理を行っても検出性能を保持することができる。

【0094】

40

しかも、このようなダウンサンプル処理を行うことで、音声同期用特徴量を伝送する際の伝送量を削減することができる。また計算済み音声同期用特徴量をメモリやストレージに保持する際に、必要な容量を削減することができる。

【0095】

さらに、2つの音声同期用特徴量間の同期処理を行う際の演算量を削減することができる。同期処理は入力特徴量の長さがn倍になると、その演算量が n^2 倍となるため、ダウンサンプル処理の効果は大きい。一方、単純に間引き処理を行っただけでは、同期の検出性能が劣化してしまうため、ダウンサンプル部56によるダウンサンプル方法のように、必要な情報を保持したままダウンサンプルを行う処理が必要となる。

【0096】

50

なお、図 3 では、音声同期用特徴量としてのピーク情報を 1/4 にダウンサンプルする例について説明したが、1/2 や 1/8 など、他のどのようなレートで変換（ダウンサンプル）することも可能である。

【0097】

また、ピーク情報のダウンサンプルの際には、上述した式（2）の計算方法以外の方法を用いることもできる。

【0098】

例えば、4 つの時間区間のうちの 2 つの時間区間以上、ピーク情報 $P_i(\quad)$ の値が「1」である場合にダウンサンプル後のピーク情報 $P'_i(\quad)$ の値を「1」としてもよい。また、3 つの時間区間以上、ピーク情報 $P_i(\quad)$ の値が「1」である場合にダウンサンプル後のピーク情報 $P'_i(\quad)$ の値を「1」としてもよいし、4 つの全ての時間区間のピーク情報 $P_i(\quad)$ の値が「1」である場合にダウンサンプル後のピーク情報 $P'_i(\quad)$ の値を「1」としてもよい

10

【0099】

さらに、ダウンサンプル前の 4 つの時間区間において 2 つの時間区間以上、連続してピーク情報 $P_i(\quad)$ の値が「1」である場合にダウンサンプル後のピーク情報 $P'_i(\quad)$ の値を「1」としてもよいし、3 つの時間区間以上、連続してピーク情報 $P_i(\quad)$ の値が「1」である場合にダウンサンプル後のピーク情報 $P'_i(\quad)$ の値を「1」としてもよい。

【0100】

20

また、以上においては時間軸方向（時間区間方向）にピーク情報 $P_i(\quad)$ をダウンサンプルする方法について説明したが、ピーク情報 $P_i(\quad)$ を時間遅れ 方向にダウンサンプルするようにしてもよい。

【0101】

そのような場合、ダウンサンプル部 56 は例えば次式（3）を計算することでピーク情報 $P_i(\quad)$ をダウンサンプルし、ピーク情報 $P'_i(\quad)$ とする。

【0102】

【数 3】

$$P'_i(\tau) = P_i(4\tau) \cup P_i(4\tau+1) \cup P_i(4\tau+2) \cup P_i(4\tau+3) \dots (3)$$

30

【0103】

式（3）の計算では、時間遅れ 方向に連続して並び、同じ時間区間の 4 つのピーク情報 $P_i(\quad)$ が統合されて 1 つのピーク情報 $P'_i(\quad)$ とされる。

【0104】

このとき、統合される 4 つの連続するピーク情報 $P_i(\quad)$ のうち、1 つでも値が「1」であるピーク情報 $P_i(\quad)$ がある場合、統合により得られたピーク情報 $P'_i(\quad)$ の値は「1」とされる。逆に、統合される 4 つの連続するピーク情報 $P_i(\quad)$ の値が全て「0」である場合に、統合により得られたピーク情報 $P'_i(\quad)$ の値は「0」とされる。

40

【0105】

さらに、ピーク情報 $P_i(\quad)$ を時間区間 i 方向および時間遅れ 方向の両方向にダウンサンプルするようにしてもよい。

【0106】

そのような場合、ダウンサンプル部 56 は例えば次式（4）を計算することでピーク情報 $P_i(\quad)$ をダウンサンプルし、ピーク情報 $P'_i(\quad)$ とする。

【0107】

【数 4】

$$P_i'(\tau) = P_{2i}(2\tau) \cup P_{2i+1}(2\tau) \cup P_{2i}(2\tau+1) \cup P_{2i+1}(2\tau+1) \dots (4)$$

【0108】

式(4)の計算では、時間区間*i*方向に連続して並ぶ、同じ時間遅れの2つのピーク情報 $P_i(\quad)$ と、それらの2つのピーク情報 $P_i(\quad)$ に対して時間遅れ 方向に隣接して並ぶ2つのピーク情報 $P_i(\quad)$ とからなる合計4つのピーク情報 $P_i(\quad)$ が統合されて1つのピーク情報 $P'_i(\quad)$ とされる。

【0109】

このとき、統合される4つのピーク情報 $P_i(\quad)$ のうち、1つでも値が「1」であるピーク情報 $P_i(\quad)$ がある場合、統合により得られたピーク情報 $P'_i(\quad)$ の値は「1」とされる。逆に、統合される4つのピーク情報 $P_i(\quad)$ の値が全て「0」である場合に、統合により得られたピーク情報 $P'_i(\quad)$ の値は「0」とされる。

【0110】

以上のようにしてダウンサンプル部56は、ピーク情報 $P(\quad)$ をダウンサンプルしてピーク情報 $P'_i(\quad)$ を求めると、得られた新たな時間区間ごとのピーク情報 $P'_i(\quad)$ を、時間区間ごとの音声同期用特徴量の時系列データとして多重化処理部24に供給する。

【0111】

コンテンツ再生システムの構成例

次に、提供装置11から送信されるメイン送出信号とサブ送出信号を、それぞれメイン受信信号およびサブ受信信号として受信してメインコンテンツとサブコンテンツを再生するコンテンツ再生システムの構成について説明する。このようなコンテンツ再生システムは、例えば図4に示すように構成される。

【0112】

図4に示すコンテンツ再生システムは、メイン受信機器81、表示部82、スピーカ83、マイクロホン84、サブ受信機器85、表示部86、およびスピーカ87を有している。なお、ここでは、コンテンツ再生システムが複数の装置から構成される場合を例として示しているが、コンテンツ再生システムは1つの装置から構成されるようにしてもよい。

【0113】

メイン受信機器81は、提供装置11から送信されたメイン受信信号を受信し、メイン受信信号から得られるメインコンテンツの再生を制御する。

【0114】

メイン受信機器81は、入力部111および再生処理部112を備えている。

【0115】

入力部111は、提供装置11から送信されたメイン送出信号を、メイン受信信号として受信して再生処理部112に供給する。再生処理部112は、入力部111から供給されたメイン受信信号に含まれているメインコンテンツの画像信号と音声信号を抽出し、画像信号を表示部82に供給して再生させるとともに、音声信号をスピーカ83に供給して再生させる。すなわち、再生処理部112は、メインコンテンツの再生を制御する。

【0116】

表示部82は、例えば液晶表示装置などからなり、再生処理部112から供給された画像信号に基づいて、メインコンテンツの画像(映像)を表示させる。スピーカ83は、音声再生装置であり、再生処理部112から供給された音声信号に基づいて、メインコンテンツの音声を出力する。

【0117】

マイクロホン84は、スピーカ83から出力されたメインコンテンツの音声を收音し、その結果得られた音声信号をサブ受信機器85に供給する。

【0118】

サブ受信機器85は、提供装置11から送信されたサブ送出信号を、サブ受信信号として受信し、サブ受信信号から得られるサブコンテンツの再生を制御する。

【0119】

サブ受信機器85は、音声同期用特徴量計算部121、バッファ122、入力部123、分離処理部124、バッファ125、同期計算部126、および再生処理部127を備えている。

【0120】

音声同期用特徴量計算部121は、マイクロホン84から供給された音声信号から、音声同期用特徴量を算出し、バッファ122に供給する。バッファ122は、音声同期用特徴量計算部121から供給された音声同期用特徴量を一時的に記録する。

10

【0121】

入力部123は、提供装置11から送信されたサブ受信信号を受信して分離処理部124に供給する。分離処理部124は、入力部123から供給されたサブ受信信号を、音声同期用特徴量とサブチャンネル信号とに分離させて、バッファ125に供給する。バッファ125は、分離処理部124から供給された音声同期用特徴量とサブチャンネル信号を一時的に記録する。

【0122】

同期計算部126は、バッファ122に記録されている音声同期用特徴量と、バッファ125に記録されている音声同期用特徴量とに基づいて、メインコンテンツとサブコンテンツとを同期させるための音声特徴量に基づく同期補正情報を生成し、再生処理部127に供給する。すなわち、同期計算部126は、収音して得られた音声信号から抽出した音声同期用特徴量と、サブ受信信号に含まれている音声同期用特徴量とのマッチング処理により、メインコンテンツとサブコンテンツとの再生時刻のずれを検出し、そのずれを示す音声特徴量に基づく同期補正情報を生成する。

20

【0123】

再生処理部127は、同期計算部126から供給された同期補正情報に基づいて、バッファ125に記録されているサブチャンネル信号の再生タイミング(時刻)を補正し、サブチャンネル信号としての画像信号と音声信号を、それぞれ表示部86およびスピーカ87に供給する。すなわち、再生処理部127は、サブコンテンツの再生を制御する。例えばMPEG-4 Systemを用いて、音声同期用特徴量を1つのメディアオブジェクトとして扱い、サブチャンネル信号のメディアオブジェクトと同期および多重化している場合に、各メディアオブジェクトの最小単位であるAccess Unit(AU)にはそれぞれ時間属性が定義されているので、上記の同期補正情報からサブチャンネル信号のメディアオブジェクトの適切な再生タイミング(時刻)を算出することができる。

30

【0124】

表示部86は、例えば液晶表示装置などからなり、再生処理部127から供給された画像信号に基づいて、サブコンテンツの画像(映像)を表示させる。スピーカ87は、音声再生装置であり、再生処理部127から供給された音声信号に基づいて、サブコンテンツの音声を出力する。

40

【0125】

音声同期用特徴量計算部の構成例

また、図4に示した音声同期用特徴量計算部121は、より詳細には例えば図5に示すように構成される。

【0126】

音声同期用特徴量計算部121は、周波数帯域分割部151、周期性検出部152-1乃至周期性検出部152-4、周期性強度検出部153-1乃至周期性強度検出部153-4、周期性情報統合部154、およびピーク検出部155を有している。

【0127】

なお、これらの周波数帯域分割部151乃至ピーク検出部155は、図2に示した周波

50

数帯域分割部 5 1 乃至ピーク検出部 5 5 と同様であるので、その説明は省略する。但し、周波数帯域分割部 1 5 1 と周波数帯域分割部 5 1 とでは、窓関数のシフト時間を異なった値に設定することも可能である。例えばサブ受信機器 8 5 の演算リソースが豊富な場合に、周波数帯域分割部 1 5 1 において、より短いシフト時間を用いることで、より細かい粒度での音声同期用特徴量の抽出が可能になる。

【 0 1 2 8 】

また、以下、周期性検出部 1 5 2 - 1 乃至周期性検出部 1 5 2 - 4 を特に区別する必要のない場合、単に周期性検出部 1 5 2 とも称し、周期性強度検出部 1 5 3 - 1 乃至周期性強度検出部 1 5 3 - 4 を特に区別する必要のない場合、周期性強度検出部 1 5 3 とも称する。

10

【 0 1 2 9 】

同期計算部の構成例

さらに、図 4 に示した同期計算部 1 2 6 は、より詳細には例えば図 6 に示すように構成される。

【 0 1 3 0 】

図 6 の同期計算部 1 2 6 は、フレームレート変換部 1 8 1、フレームレート変換部 1 8 2、ブロック統合部 1 8 3、ブロック統合部 1 8 4、類似度計算部 1 8 5、および最適パス検索部 1 8 6 を有している。

【 0 1 3 1 】

フレームレート変換部 1 8 1 はバッファ 1 2 2 からメインコンテンツの時間区間ごとの音声同期用特徴量の時系列データを読み出し、音声同期用特徴量のフレームレートを変換してブロック統合部 1 8 3 に供給する。ここでいうフレームレートとは、音声同期用特徴量の時系列データにおける単位時間当たりの時間区間数、つまり時間区間の長さをいう。

20

【 0 1 3 2 】

フレームレート変換部 1 8 2 はバッファ 1 2 5 からメインコンテンツの時間区間ごとの音声同期用特徴量の時系列データを読み出し、音声同期用特徴量のフレームレートを変換してブロック統合部 1 8 4 に供給する。

【 0 1 3 3 】

バッファ 1 2 2 およびバッファ 1 2 5 にそれぞれ保持されている音声同期用特徴量はフレームレート、つまり時間区間の長さが異なる場合がある。

30

【 0 1 3 4 】

例えば提供装置 1 1 から提供されるサブコンテンツ（サブ送出信号）の転送ビットレートを削減するため、サブ送出信号に含まれている音声同期用特徴量が低レートに設定されている一方で、マイクロホン 8 4 で收音された音声から計算される音声同期用特徴量は、伝送の必要がないため高いレートに設定されている場合が考えられる。

【 0 1 3 5 】

このような場合、例えば図 7 に示すように、マイクロホン 8 4 で收音された音声から計算される音声同期用特徴量をダウンサンプル部 5 6 と同じ手法を用いてダウンサンプルすることが考えられる。なお、図 7 において各矢印 Q 1 1 乃至矢印 Q 1 4 に示される音声同期用特徴量としてのピーク情報の縦軸は時間遅れを示しており、横軸は時間区間 i を示している。また、1 つの四角形は 1 つの時間区間におけるピーク情報を表している。

40

【 0 1 3 6 】

この例では、提供装置 1 1 側では、矢印 Q 1 1 に示すように音声同期用特徴量としてのピーク情報が求められた後、そのピーク情報がダウンサンプルされ、矢印 Q 1 2 に示す、より時間区間が長いピーク情報とされてサブ受信機器 8 5 へと伝送される。ここでは、8msec の時間区間のピーク情報が、32msec の時間区間のピーク情報へとフレームレート変換（ダウンサンプル）されている。

【 0 1 3 7 】

一方、サブ受信機器 8 5 の音声同期用特徴量計算部 1 2 1 では、メイン受信機器 8 1 で再生されたメインコンテンツの音声を收音して得られた音声信号から音声同期用特徴量が

50

算出され、その結果、矢印 Q 1 3 に示すピーク情報が音声同期用特徴量として得られる。ここでは矢印 Q 1 3 に示すピーク情報は、8msecの時間区間ごとに算出されている。

【 0 1 3 8 】

このようにして音声同期用特徴量計算部 1 2 1 で得られた音声同期用特徴量と、提供装置 1 1 から受信した音声同期用特徴量とでは時間区間の長さ、つまりフレームレートが異なる。そこで、フレームレート変換部 1 8 1 は、それらの音声同期用特徴量のフレームレートが一致するように、音声同期用特徴量計算部 1 2 1 で得られた音声同期用特徴量に対してフレームレート変換としてダウンサンプルを行い、矢印 Q 1 4 に示す音声同期用特徴量としてのピーク情報を得る。矢印 Q 1 4 に示す音声同期用特徴量は、32msecの時間区間のピーク情報となっている。

10

【 0 1 3 9 】

このようにしてフレームレート（時間区間の長さ）が揃えられた後、音声同期用特徴量を用いられて同期計算が行われる。このようにサブ受信機器 8 5 側において音声同期用特徴量のダウンサンプルを行うことで、任意のフレームレート（ビットレート）に対応することができる。

【 0 1 4 0 】

また、サブ受信機器 8 5 に伝送されてくる音声同期用特徴量は高レートであるが、マイクロホン 8 4 で収音された音声から計算される音声同期用特徴量は低いレートになるケースもある。例えばサブ受信機器 8 5 の演算リソースが潤沢ではなく、音声同期用特徴量の計算に必要な演算量を削減するために、フレームシフト量を大きくするケースなどである。

20

【 0 1 4 1 】

そのような場合、例えば図 8 の矢印 Q 2 1 に示す、サブ送出信号に含まれている音声同期用特徴量のフレームレートがダウンサンプル部 5 6 と同じ手法が用いられてフレームレート変換部 1 8 2 によってダウンサンプルされ、矢印 Q 2 2 に示す音声同期用特徴量を得られる。なお、図 8 において各矢印 Q 2 1 乃至矢印 Q 2 3 に示される音声同期用特徴量としてのピーク情報の縦軸は時間遅れを示しており、横軸は時間区間 i を示している。また、1つの四角形は1つの時間区間におけるピーク情報を表している。

【 0 1 4 2 】

この例では、8msecの時間区間のピーク情報が、32msecの時間区間のピーク情報へとフレームレート変換（ダウンサンプル）されている。

30

【 0 1 4 3 】

また、サブ受信機器 8 5 の音声同期用特徴量計算部 1 2 1 では、メイン受信機器 8 1 で再生されたメインコンテンツの音声を収音して得られた音声信号から音声同期用特徴量が算出され、その結果、矢印 Q 2 3 に示すピーク情報が音声同期用特徴量として得られる。ここでは矢印 Q 2 3 に示すピーク情報は、32msecの時間区間ごとに算出されている。

【 0 1 4 4 】

このようにサブ送出信号に含まれている音声同期用特徴量をダウンサンプルして、サブ送出信号に含まれている音声同期用特徴量のフレームレートと、サブ受信機器 8 5 で算出される音声同期用特徴量のフレームレートを一致させるようにしてもよい。

40

【 0 1 4 5 】

さらに、上記説明ではよりフレームレートが高い音声同期用特徴量をダウンサンプルすることでフレームレートを一致させたが、よりフレームレートが低い音声同期用特徴量をアップサンプルすることでフレームレートを一致させてもよい。

【 0 1 4 6 】

そのような場合、例えば図 9 に示すようにサブ送出信号に含まれている音声同期用特徴量のフレームレートがフレームレート変換部 1 8 2 によってアップサンプルされる。なお、図 9 において各矢印 Q 3 1 乃至矢印 Q 3 4 に示される音声同期用特徴量としてのピーク情報の縦軸は時間遅れを示しており、横軸は時間区間 i を示している。また、1つの四角形は1つの時間区間におけるピーク情報を表している。

50

【 0 1 4 7 】

この例では、提供装置 1 1 側では、矢印 Q 3 1 に示すように音声同期用特徴量としてのピーク情報が求められた後、ダウンサンプルが行われて矢印 Q 3 2 に示す、より時間区間が長いピーク情報とされてサブ受信機器 8 5 へと伝送される。ここでは、8msecの時間区間のピーク情報が、32msecの時間区間のピーク情報へとフレームレート変換（ダウンサンプル）されている。

【 0 1 4 8 】

一方、サブ受信機器 8 5 の音声同期用特徴量計算部 1 2 1 では、メイン受信機器 8 1 で再生されたメインコンテンツの音声を收音して得られた音声信号から音声同期用特徴量が算出され、その結果、矢印 Q 3 3 に示すピーク情報が音声同期用特徴量として得られる。ここでは矢印 Q 3 3 に示すピーク情報は、8msecの時間区間ごとに算出されている。

10

【 0 1 4 9 】

この例では、音声同期用特徴量計算部 1 2 1 により算出された音声同期用特徴量と、提供装置 1 1 から受信した音声同期用特徴量とではフレームレートが一致していない状態となっている。

【 0 1 5 0 】

そこで、フレームレート変換部 1 8 2 は、提供装置 1 1 から受信した音声同期用特徴量としてのピーク情報をアップサンプルし、矢印 Q 3 4 に示す8msecの時間区間のピーク情報を算出することで、同期計算に用いる音声同期用特徴量の時間同期の粒度を揃える。例えばフレームレート変換部 1 8 2 は、次式（ 5 ）を計算することで、ピーク情報をアップ

20

【 0 1 5 1 】

【 数 5 】

$$\begin{aligned} P_{4i}(\tau) &= P_i'(\tau) \\ P_{4i+1}(\tau) &= P_i'(\tau) \\ P_{4i+2}(\tau) &= P_i'(\tau) \\ P_{4i+3}(\tau) &= P_i'(\tau) \quad \dots (5) \end{aligned}$$

30

【 0 1 5 2 】

式（ 5 ）の計算では、時間遅れ が同じであり、時間方向（時間区間方向）に隣接するアップサンプル後の 4 つのピーク情報 $P_i(\quad)$ と同じ位置関係にあるアップサンプル前の 1 つのピーク情報 $P'_i(\quad)$ の値が、そのままアップサンプル後の 4 つの各ピーク情報 $P_i(\quad)$ の値とされている。

【 0 1 5 3 】

このように同期計算に用いる音声同期用特徴量を、より高いフレームレートに合わせて適宜アップサンプルすることで、疑似的に高分解能な同期精度を実現することができる。

【 0 1 5 4 】

さらに、サブ受信機器 8 5 での演算リソース削減のため、サブ送出信号に含まれている音声同期用特徴量と、音声同期用特徴量計算部 1 2 1 で算出された音声同期用特徴量との両方をダウンサンプルすることも可能である。

40

【 0 1 5 5 】

以上のようにフレームレート変換部 1 8 1 およびフレームレート変換部 1 8 2 を有することで、異なるフレームレートの音声同期用特徴量間の同期を行うことができるようになる。また、演算リソースや伝送帯域などに応じて様々なフレームレートを指定できるようになり、システムの柔軟性を高めることができる。

【 0 1 5 6 】

図 6 の説明に戻り、ブロック統合部 1 8 3 は、フレームレート変換部 1 8 1 からメインコンテンツの時間区間ごとの音声同期用特徴量の時系列データの供給を受け、連続した複

50

数（例えば64個）の時間区間を1ブロックとして、ブロック単位で統合する。ブロック統合部183は、ブロック単位の音声同期用特徴量の時系列データを類似度計算部185に供給する。

【0157】

ブロック統合部184は、フレームレート変換部182からメインコンテンツの時間区間ごとの音声同期用特徴量の時系列データの供給を受け、連続した複数（例えば64個）の時間区間を1ブロックとして、ブロック単位で統合する。ブロック統合部184は、ブロック単位の音声同期用特徴量の時系列データを類似度計算部185に供給する。

【0158】

なお、ブロックを構成する複数の時間区間は、連続していなくてもよい。例えば、複数の偶数番目の時間区間を1ブロックとしたり、複数の奇数番目の時間区間を1ブロックとしたりすることもできる。この場合、時間区間ごとの音声同期用特徴量の時系列データに対して間引き処理を行うことができるので、演算量を削減することができる。

【0159】

類似度計算部185は、ブロック統合部183とブロック統合部184のそれぞれから供給されたブロック単位の音声同期用特徴量の時系列データ同士の類似度を計算し、各ブロック間の類似度を表す類似度マトリックスを生成する。類似度計算部185は、類似度マトリックスを最適パス検索部186に供給する。

【0160】

最適パス検索部186は、類似度計算部185から供給された類似度マトリックスから最適な類似度のパスを検索し、そのパス上の類似度に対応する2つのブロックの時間差を表す情報を音声特徴量に基づく同期補正情報として生成する。そして、最適パス検索部186は、音声特徴量に基づく同期補正情報を再生処理部127に供給する。

【0161】

以上のように、同期計算部126は、音声信号のピッチ情報に基づいて、音声特徴量に基づく同期補正情報を生成する。したがって、音声信号ごとに異なるノイズが含まれる場合などにおいても、ロバストに音声特徴量に基づく同期補正情報を生成することができる。

【0162】

すなわち、人間は、周波数特性を有する複数の音を聞いた場合、共通成分として、同一の基本周波数を有する音、つまりピッチが同一である音を知覚することにより、ノイズ音が含まれている場合であっても、共通成分を容易に知覚することができる。本技術は、このことを考慮して、ピッチ情報に基づいて同期補正情報を生成することにより、ノイズ音に対してロバストに同期補正情報を生成する。

【0163】

類似度の計算と最適な類似度のパスの検索について

ここで、類似度の計算と最適な類似度のパスの検索について説明する。

【0164】

図10は、類似度の計算の対象とするブロックを説明する図である。

【0165】

なお、図10において*i*は、音声同期用特徴量計算部121で得られた音声同期用特徴量のブロックのインデックスであり、*j*は、サブ受信信号に含まれている音声同期用特徴量のブロックのインデックスである。なお、より詳細には、これらの音声同期用特徴量は、適宜、フレームレート変換部181やフレームレート変換部182によりフレームレート変換されるが、ここでは説明を簡単にするため、フレームレート変換は行われたいものとして類似度の計算についての説明を続ける。

【0166】

また、 $X(i)$ は、音声同期用特徴量計算部121で得られた音声同期用特徴量のうちのインデックス*i*のブロックの音声同期用特徴量の時系列データを表し、 $Y(j)$ は、サブ受信信号に含まれている音声同期用特徴量のうちのインデックス*j*のブロックの音声同期用特徴

10

20

30

40

50

量の時系列データを表す。

【 0 1 6 7 】

図 1 0 に示すように、類似度の計算の対象は、 n 個の $X(i)$ のそれぞれと、 m 個の $Y(j)$ のそれぞれとの、 $n \times m$ 個の組み合わせである。

【 0 1 6 8 】

図 1 1 は、類似度の計算方法を説明する図である。

【 0 1 6 9 】

なお、図 1 1 のマトリックスでは、横軸が、ブロック内の時間区間の先頭からの個数を表すブロック内時間区間番号を表し、縦軸がインデックスを表している。また、白色の正方形は、対応するブロック内時間区間番号の時間区間のインデックスの音声同期用特
10
徴量の時系列データ $P(\quad)$ が 0 であることを表し、黒色の正方形は、その時系列データ $P(\quad)$ が 1 であることを表している。さらに、図 1 1 の例では、ブロックを構成する時間区間の個数が 4 個であり、 \quad が 0 乃至 3 であるものとする。

【 0 1 7 0 】

図 1 1 に示すように、 $X(i)$ と $Y(j)$ の類似度を計算する場合、まず、 $X(i)$ と $Y(j)$ の論理積 $X(i) \cap Y(j)$ が計算され、次に、 $X(i)$ と $Y(j)$ の論理和 $X(i) \cup Y(j)$ が計算される。例えば、図
20
1 1 に示すように、9 個の 0 と 7 個の 1 からなる $X(i)$ と $Y(j)$ の類似度を計算する場合、まず、12 個の 0 と 4 個の 1 からなる論理積 $X(i) \cap Y(j)$ が計算され、6 個の 0 と 10 個の 1 からなる論理和 $X(i) \cup Y(j)$ が計算される。

【 0 1 7 1 】

そして、次式 (6) により、論理積 $X(i) \cap Y(j)$ の 1 の数である $\text{Number}(X(i) \cap Y(j))$ と
20
論理和の 1 の数である $\text{Number}(X(i) \cup Y(j))$ に基づいて、 $X(i)$ と $Y(j)$ の類似度 $A(i, j)$ が計算される。

【 0 1 7 2 】

【数 6】

$$A(i, j) = \text{Number}(X(i) \cap Y(j)) / \text{Number}(X(i) \cup Y(j)) \quad \dots (6)$$

【 0 1 7 3 】

図 1 1 の例では、 $\text{Number}(X(i) \cap Y(j))$ が 4 であり、 $\text{Number}(X(i) \cup Y(j))$ が 10 である
30
ので、類似度 $A(i, j)$ は 0.4 となる。

【 0 1 7 4 】

なお、音声同期用特徴量の時系列データとして、周期性情報の総和 $S(\quad)$ を採用した場合
40
には、類似度の計算方法としてコサイン距離を用いて類似度を計算する方法などを採用することができる。

【 0 1 7 5 】

また、類似度マトリックスは、例えば横軸がインデックス j とされ、縦軸がインデックス i とされた、インデックス i とインデックス j に対応する各点の類似度 $A(i, j)$ を示す情報である。

【 0 1 7 6 】

最適パス検索部 186 は、動的計画法を用いて、類似度マトリックス上のパスの類似度の積算値が最大となるパスを最適な類似度のパスとして検索する。最適パス検索部 186
40
は、最適な類似度のパス上の類似度に対応するインデックスの差分 $i - j$ を、音声特徴量に基づく同期補正情報として生成する。

【 0 1 7 7 】

送信処理の説明

続いて、提供装置 11 の動作について説明する。

【 0 1 7 8 】

提供装置 11 は、互いに時間同期がとれているメインチャンネル信号とサブチャンネル
50
信号が供給されると、送信処理を行って、メイン送出信号およびサブ送出信号を送信する

。以下、図 1 2 のフローチャートを参照して、提供装置 1 1 による送信処理について説明する。

【 0 1 7 9 】

ステップ S 1 1 において、音声同期用特徴量計算部 2 3 は、音声同期用特徴量算出処理を行って、供給されたメインチャンネル信号を構成する音声信号から、音声同期用特徴量を計算し、多重化処理部 2 4 に供給する。

【 0 1 8 0 】

なお、音声同期用特徴量算出処理の詳細は後述する。

【 0 1 8 1 】

ステップ S 1 2 において、変換部 2 1 は、供給されたメインチャンネル信号を、システムが規定する所定の伝送フォーマットの信号に変換することでメイン送出信号を生成し、得られたメイン送出信号を出力部 2 2 に供給する。

【 0 1 8 2 】

ステップ S 1 3 において、出力部 2 2 は、変換部 2 1 から供給されたメイン送出信号を送信する。

【 0 1 8 3 】

ステップ S 1 4 において、多重化処理部 2 4 は、音声同期用特徴量とサブチャンネル信号との多重化処理を行い、その結果得られたサブ送出信号を出力部 2 5 に供給する。

【 0 1 8 4 】

例えば多重化処理部 2 4 は、供給されたメインチャンネル信号を用いて、音声同期用特徴量計算部 2 3 からの音声同期用特徴量と、供給されたサブチャンネル信号との時間同期関係が符合するように、システムが規定する伝送フォーマットにより音声同期用特徴量とサブチャンネル信号を多重化する。

【 0 1 8 5 】

これにより、例えば図 1 3 に示すサブ送出信号が得られる。

【 0 1 8 6 】

図 1 3 の例では、サブ送出信号としてのビットストリームにおける区間 T 1 1 と区間 T 1 2 には、それぞれ 1 フレーム分の画像信号、音声信号、および音声同期用特徴量が含まれている。

【 0 1 8 7 】

例えば、区間 T 1 1 に含まれる画像信号と音声信号は、1 フレーム分のサブチャンネル信号であり、区間 T 1 1 に含まれる音声同期用特徴量は、そのサブチャンネル信号に時間的に対応するフレームのメインチャンネル信号から抽出された、音声同期用特徴量である。このように、サブ送出信号では、同じフレームのサブチャンネル信号と音声同期用特徴量とが対応付けられて多重化されており、サブ送出信号の受信側では、各フレームのサブチャンネル信号に対応付けられた音声同期用特徴量が特定できるようになされている。

【 0 1 8 8 】

図 1 2 のフローチャートの説明に戻り、ステップ S 1 5 において、出力部 2 5 は、多重化処理部 2 4 から供給されたサブ送出信号を送信し、送信処理は終了する。

【 0 1 8 9 】

以上のようにして、提供装置 1 1 は、メインチャンネル信号から得られた音声同期用特徴量と、サブチャンネル信号とを対応付けて多重化することでサブ送出信号を生成し、サブ送出信号とメイン送出信号を送信する。

【 0 1 9 0 】

このようにサブチャンネル信号に音声同期用特徴量に対応付けて送信することで、受信側においては、メインチャンネル信号とサブチャンネル信号を、異なる伝送路を介して複数の異なる機器で受信した場合においても、音声同期用特徴量を用いて、メインコンテンツとサブコンテンツを、同期を保って再生することができるようになる。

【 0 1 9 1 】

音声同期用特徴量算出処理の説明

10

20

30

40

50

次に、図 1 4 のフローチャートを参照して、図 1 2 のステップ S 1 1 の処理に対応する音声同期用特徴量算出処理について説明する。

【 0 1 9 2 】

ステップ S 4 1 において、周波数帯域分割部 5 1 は、供給された音声信号を、窓関数を用いて、数 10 msec 乃至 100 msec 程度の時間区間に分割する。

【 0 1 9 3 】

ステップ S 4 2 において、周波数帯域分割部 5 1 は、複数のバンドパスフィルタを用いて、時間区間ごとの音声信号を 4 つの周波数帯域に分割する。周波数帯域分割部 5 1 は、各周波数帯域の音声信号を、周期性検出部 5 2 - 1 乃至周期性検出部 5 2 - 4 のそれぞれに供給する。

10

【 0 1 9 4 】

ステップ S 4 3 において、周期性検出部 5 2 は、周波数帯域分割部 5 1 から供給された所定の周波数帯域の時間区間ごとの音声信号の自己相関関数 $x(b, \quad)$ を計算することにより、時間区間ごとの周期性情報を抽出し、周期性強度検出部 5 3 および周期性情報統合部 5 4 に供給する。なお、ステップ S 4 3 の処理は、周期性検出部 5 2 ごとに行われる。

【 0 1 9 5 】

ステップ S 4 4 において、周期性強度検出部 5 3 は、周期性検出部 5 2 から供給された時間区間ごとの周期性情報に基づいて、時間区間ごとの周期性の強度を計算する。そして、周期性強度検出部 5 3 は、時間区間ごとの周期性の強度を、閾値を超えたかどうかで 2 値化することにより、時間区間ごとの周期性強度情報を生成し、周期性情報統合部 5 4 に供給する。なお、ステップ S 4 4 の処理は、周期性強度検出部 5 3 ごとに行われる。

20

【 0 1 9 6 】

ステップ S 4 5 において、周期性情報統合部 5 4 は、周期性検出部 5 2 から供給された時間区間ごとの周期性情報と、周期性強度検出部 5 3 から供給される時間区間ごとの周期性強度情報とに基づいて、上述した式 (1) を用いて周期性統合処理を行う。周期性情報統合部 5 4 は、周期性統合処理の結果得られる時間区間ごとの周期性情報の総和 $S(\quad)$ をピーク検出部 5 5 に供給する。

【 0 1 9 7 】

ステップ S 4 6 において、ピーク検出部 5 5 は、時間区間ごとに、周期性情報統合部 5 4 から供給された周期性情報の総和 $S(\quad)$ に対してピーク検出を行い、ピーク情報 $P(\quad)$ を生成し、ダウンサンプル部 5 6 に供給する。

30

【 0 1 9 8 】

ステップ S 4 7 において、ダウンサンプル部 5 6 は、ピーク検出部 5 5 から供給された複数の時間区間におけるピーク情報 $P(\quad)$ を 1 つの時間区間に統合することで、ピーク情報のダウンサンプル処理を行う。

【 0 1 9 9 】

ダウンサンプル部 5 6 は、このようにして得られた時間区間ごとのピーク情報を、時間区間ごとの音声同期用特徴量の時系列データとして多重化処理部 2 4 に供給し、音声同期用特徴量算出処理は終了する。音声同期用特徴量算出処理が終了すると、その後、処理は図 1 2 のステップ S 1 2 へと進む。

40

【 0 2 0 0 】

音声同期用特徴量計算部 2 3 では、以上のようにして周期性情報に基づいて音声同期用特徴量を算出するので、音声同期用特徴量をロバストに生成することができる。

【 0 2 0 1 】

メインコンテンツ再生処理の説明

また、提供装置 1 1 からメイン送出信号が送信されると、コンテンツ再生システムは、そのメイン送出信号を、メイン受信信号として取得して、メインコンテンツを再生する。以下、図 1 5 のフローチャートを参照して、コンテンツ再生システムによるメインコンテンツ再生処理について説明する。

50

【 0 2 0 2 】

ステップ S 7 1 において、入力部 1 1 1 は、メイン受信信号を取得して再生処理部 1 1 2 に供給する。例えば入力部 1 1 1 は、提供装置 1 1 から送信されたメイン受信信号を受信することで、メイン受信信号を取得する。

【 0 2 0 3 】

ステップ S 7 2 において、再生処理部 1 1 2 は、入力部 1 1 1 から供給されたメイン受信信号に基づいてメインコンテンツを再生させ、メインコンテンツ再生処理は終了する。

【 0 2 0 4 】

例えば、再生処理部 1 1 2 は、メイン受信信号から、メインコンテンツの画像信号と音声信号を抽出し、画像信号を表示部 8 2 に供給して再生させるとともに、音声信号をスピーカ 8 3 に供給して再生させる。これにより、メインコンテンツが再生される。

【 0 2 0 5 】

以上のようにして、コンテンツ再生システムは、メイン受信信号を取得してメインコンテンツを再生する。

【 0 2 0 6 】

サブコンテンツ再生処理の説明

また、メインコンテンツの再生と同期して、コンテンツ再生システムは、サブ受信信号を取得して、サブコンテンツを再生する。以下、図 1 6 のフローチャートを参照して、コンテンツ再生システムによるサブコンテンツ再生処理について説明する。

【 0 2 0 7 】

ステップ S 1 0 1 において、入力部 1 2 3 は、サブ受信信号を取得して分離処理部 1 2 4 に供給する。例えば入力部 1 2 3 は、提供装置 1 1 から送信されたサブ送出信号を、サブ受信信号として受信することで、サブ受信信号を取得する。

【 0 2 0 8 】

ステップ S 1 0 2 において、分離処理部 1 2 4 は、入力部 1 2 3 から供給されたサブ受信信号を、サブチャンネル信号と音声同期用特徴量とに分離させ、分離されたサブチャンネル信号と音声同期用特徴量をバッファ 1 2 5 に供給して記録させる。

【 0 2 0 9 】

ステップ S 1 0 3 において、マイクロホン 8 4 は、スピーカ 8 3 から出力されたメインコンテンツの音声を收音し、その結果得られた音声信号を音声同期用特徴量計算部 1 2 1 に供給する。例えばステップ S 1 0 3 では、図 1 5 のステップ S 7 2 の処理で再生されたメインコンテンツの音声收音される。

【 0 2 1 0 】

ステップ S 1 0 4 において、音声同期用特徴量計算部 1 2 1 は、音声同期用特徴量算出処理を行って、マイクロホン 8 4 から供給された音声信号から音声同期用特徴量を計算し、バッファ 1 2 2 に供給して記録させる。

【 0 2 1 1 】

なお、音声同期用特徴量算出処理として、図 1 7 のフローチャートに示すステップ S 1 3 1 乃至ステップ S 1 3 6 の処理が行われるが、これらの処理は図 1 4 のステップ S 4 1 乃至ステップ S 4 6 の処理と同様であるので、その説明は省略する。但し、図 1 7 に示す音声同期用特徴量算出処理では、マイクロホン 8 4 から供給された音声信号から音声同期用特徴量が計算され、バッファ 1 2 2 に蓄積される。また、音声同期用特徴量計算部 1 2 1 では、ピーク検出部 1 5 5 で得られたピーク情報が音声同期用特徴量とされる。

【 0 2 1 2 】

図 1 6 のフローチャートの説明に戻り、ステップ S 1 0 5 において、同期計算部 1 2 6 は、同期補正情報生成処理を行って、音声特徴量に基づく同期補正情報を生成し、再生処理部 1 2 7 に供給する。なお、同期補正情報生成処理の詳細は後述するが、この処理では、バッファ 1 2 2 に記録されている音声同期用特徴量と、バッファ 1 2 5 に記録されている音声同期用特徴量とを比較することで、メインコンテンツとサブコンテンツとを同期させるための音声特徴量に基づく同期補正情報が生成される。

【 0 2 1 3 】

ステップ S 1 0 6 において、再生処理部 1 2 7 は、同期計算部 1 2 6 から供給された音声特徴量に基づく同期補正情報に基づいて、バッファ 1 2 5 に記録されているサブチャンネル信号の再生タイミングを補正し、補正後のサブチャンネル信号に基づいてサブコンテンツを再生させる。

【 0 2 1 4 】

すなわち、再生処理部 1 2 7 は、サブチャンネル信号を構成する画像信号と音声信号を、音声特徴量に基づく同期補正情報により示される時間だけ遅くまたは早く表示部 8 6 とスピーカ 8 7 に供給し、再生させる。換言すれば、音声特徴量に基づく同期補正情報から特定される、現在時刻において再生されているメインコンテンツの部分と対応する再生時刻のサブコンテンツの部分が再生される。

10

【 0 2 1 5 】

例えば、サブコンテンツをメインコンテンツと同期させるための再生位置の調整（補正）は、サブコンテンツやメインコンテンツの無音区間で行われる。

【 0 2 1 6 】

表示部 8 6 は、再生処理部 1 2 7 から供給された画像信号に基づいて、サブコンテンツの画像を表示し、スピーカ 8 7 は、再生処理部 1 2 7 から供給された音声信号に基づいて、サブコンテンツの音声を出力する。

【 0 2 1 7 】

このようにして、メインコンテンツと同期してサブコンテンツが再生されると、サブコンテンツ再生処理は終了する。

20

【 0 2 1 8 】

以上のようにして、コンテンツ再生システムは、再生されているメインコンテンツの音声を收音して得られた音声信号から音声同期用特徴量を計算し、得られた音声同期用特徴量と、サブ受信信号に含まれている音声同期用特徴量とを用いて音声特徴量に基づく同期補正情報を計算する。また、コンテンツ再生システムは、得られた同期補正情報を用いてサブコンテンツを、メインコンテンツと同期させて再生する。

【 0 2 1 9 】

このように、收音して得られた音声信号から抽出された音声同期用特徴量と、サブ受信信号に含まれている音声同期用特徴量とを用いて音声特徴量に基づく同期補正情報を計算することで、メインコンテンツとサブコンテンツとの伝送経路が異なる場合であっても、それらのコンテンツを同期して再生することができる。

30

【 0 2 2 0 】

なお、この例では、音声同期用特徴量の同期計算、つまりマッチング処理は、毎フレーム行われるが、音声同期用特徴量の同期計算は、必ずしも時間的に連続して行われる必要はなく、間欠的に行われるようにしてもよい。但し、同期計算を連続的に行った方がサブコンテンツの再生時刻（再生位置）の補正時に、違和感なく補正を行うことができる。

【 0 2 2 1 】

同期補正情報生成処理の説明

さらに、図 1 8 のフローチャートを参照して、図 1 6 のステップ S 1 0 5 の処理に対応する同期補正情報生成処理について説明する。

40

【 0 2 2 2 】

ステップ S 1 6 1 において、フレームレート変換部 1 8 1 およびフレームレート変換部 1 8 2 は、必要に応じてフレームレート変換処理を行う。

【 0 2 2 3 】

すなわち、フレームレート変換部 1 8 1 は、バッファ 1 2 2 からメインコンテンツの時間区間ごとの音声同期用特徴量の時系列データを読み出して、必要に応じて音声同期用特徴量をフレームレート変換、つまりダウンサンプルし、ブロック統合部 1 8 3 に供給する。また、フレームレート変換部 1 8 2 は、バッファ 1 2 5 からメインコンテンツの時間区間ごとの音声同期用特徴量の時系列データを読み出して、必要に応じて音声同期用特徴量

50

をフレームレート変換、つまりダウンサンプルまたはアップサンプルし、ブロック統合部 184 に供給する。

【0224】

ステップ S162 において、ブロック統合部 183 およびブロック統合部 184 は、音声同期用特徴量の時系列データを統合する。

【0225】

具体的には、ブロック統合部 183 は、フレームレート変換部 181 からメインコンテンツの時間区間ごとの音声同期用特徴量の時系列データの供給を受ける。そして、ブロック統合部 183 は、連続した複数（例えば64個）の時間区間を 1 ブロックとして、ブロック単位で、供給された時間区間ごとの音声同期用特徴量の時系列データを統合し、類似度計算部 185 に供給する。

10

【0226】

また、ブロック統合部 184 は、フレームレート変換部 182 からメインコンテンツの時間区間ごとの音声同期用特徴量の時系列データの供給を受ける。そして、ブロック統合部 184 は、連続した複数（例えば64個）の時間区間を 1 ブロックとして、ブロック単位で、供給された時間区間ごとの音声同期用特徴量の時系列データを統合し、類似度計算部 185 に供給する。

【0227】

ステップ S163 において、類似度計算部 185 は、ブロック統合部 183 とブロック統合部 184 のそれぞれから供給されたブロック単位の音声同期用特徴量の時系列データ同士の類似度を計算し、各ブロック間の類似度を表す類似度マトリックスを生成する。類似度計算部 185 は、類似度マトリックスを最適パス検索部 186 に供給する。

20

【0228】

ステップ S164 において、最適パス検索部 186 は、類似度計算部 185 から供給された類似度マトリックスから最適な類似度のパスを検索し、音声特徴量に基づく同期補正情報を生成する。そして、最適パス検索部 186 は、音声特徴量に基づく同期補正情報を再生処理部 127 に供給して、同期補正情報生成処理は終了する。

【0229】

以上のようにして、コンテンツ再生システムは、周期性情報に基づいて音声特徴量に基づく同期補正情報を生成するので、同期補正情報をロバストに生成することができる。

30

【0230】

なお、以上においては、メインコンテンツが 1 つである場合について説明したが、メインコンテンツが複数あってもよい。

【0231】

そのような場合、提供装置 11 の音声同期用特徴量計算部 23 は、複数のメインコンテンツごとに音声同期用特徴量を計算し、多重化処理部 24 は、1 つのサブコンテンツのサブチャンネル信号と、複数のメインコンテンツの音声同期用特徴量とを多重化し、サブ送出信号とする。また、出力部 22 は、複数のメインコンテンツのメインチャンネル信号から得られたメイン送出信号を送信する。

【0232】

40

さらに、この場合、図 4 に示したコンテンツ再生システムでは、再生処理部 112 は、複数のメインコンテンツのうちの 1 つを選択して再生する。また、入力部 123 は、1 つのサブチャンネル信号に対して、複数のメインコンテンツの音声同期用特徴量に対応付けられているサブ受信信号を受信する。

【0233】

そして、同期計算部 126 は、入力部 123 で取得された各メインコンテンツの音声同期用特徴量と、音声同期用特徴量計算部 121 で得られた音声同期用特徴量とを比較して類似度を計算し、スピーカ 83 で再生されているメインコンテンツを特定する。例えば、音声同期用特徴量とのマッチングの結果、最も類似度の高い音声同期用特徴量のメインコンテンツが、再生されているメインコンテンツであるとされる。

50

【0234】

再生されているメインコンテンツが特定されると、特定されたメインコンテンツの音声同期用特徴量について得られた同期補正情報に基づいて、サブコンテンツの再生位置が補正される。すなわち、同期計算部126は、特定されたメインコンテンツと、サブコンテンツとを同期させるための音声特徴量に基づく同期補正情報を生成する。

【0235】

本技術の適用例1

また、以上において説明した本技術は、様々な形態のシステムに適用することができる。

【0236】

例えば、本技術は図19に示すシステムに適用可能である。

【0237】

図19に示すシステムでは、例えば放送局などの提供装置211が、図1の提供装置11に対応する。提供装置211は、メインコンテンツとサブコンテンツとを提供する。

【0238】

この例では、提供装置211は、メインコンテンツのメイン送出信号を、例えば放送波により放送することで、メイン受信機器212にメイン送出信号を送信する。そして、メイン受信機器212は、放送波により送信されたメイン送出信号を、メイン受信信号として受信してメインコンテンツを再生する。このとき、メイン受信機器212は、メインコンテンツの音声を、メイン受信機器212に備えられたスピーカ213から出力する。

【0239】

したがって、この例ではメイン受信機器212は、図4に示したメイン受信機器81、表示部82、およびスピーカ83から構成されることになる。この場合、入力部111が、放送波により放送されたメイン受信信号を受信する。また、スピーカ213が、図4のスピーカ83に対応することになる。

【0240】

例えば、メイン受信機器212は、テレビジョン受像機などとされ、ユーザはメイン受信機器212で再生されるメインコンテンツを視聴する。

【0241】

一方、提供装置211からは、サブ送出信号も送信される。この例では提供装置211はサブ送出信号を、例えばインターネットなどの通信網214を介して、ストリーミング配信等によりサブ受信機器215に送信する。ここでは、サブ送信信号は、いわゆるプッシュ型の通信により送信される。

【0242】

また、サブ受信機器215は、例えばタブレット型の端末装置などからなり、通信網214を介して送信されてきたサブ送信信号を、サブ受信信号として受信して、サブコンテンツを再生する。すなわち、サブ受信機器215は、内蔵する表示部にサブコンテンツの画像を表示させるとともに、内蔵するスピーカからサブコンテンツの音声を出力させる。

【0243】

このとき、サブ受信機器215は、スピーカ213から出力されたメインコンテンツの音声を收音して音声同期用特徴量を計算し、得られた音声同期用特徴量と、サブ受信信号に含まれている音声同期用特徴量とを用いて音声特徴量に基づく同期補正情報を生成する。そして、サブ受信機器215は、音声特徴量に基づく同期補正情報を用いてサブコンテンツを、メインコンテンツと同期させて再生させる。

【0244】

これにより、メイン受信機器212で再生されるメインコンテンツと、サブ受信機器215で再生されるサブコンテンツとが同期した状態で再生されることになり、ユーザは、適宜、サブコンテンツを見聞きしながら、メインコンテンツを視聴することができる。つまり、サブコンテンツを、例えばメインコンテンツの補助情報として活用しながら、メインコンテンツを楽しむことができる。

10

20

30

40

50

【 0 2 4 5 】

この例では、サブコンテンツのサブチャンネル信号は、例えばメインコンテンツの映像とは別アングルの映像の画像信号や、メインコンテンツに対するコメンタリー音声の音声信号、メインコンテンツに関連する文字情報などとされる。

【 0 2 4 6 】

この場合、サブ受信機器 2 1 5 は、例えば図 4 に示したマイクロホン 8 4、サブ受信機器 8 5、表示部 8 6、およびスピーカ 8 7 から構成されることになる。したがって、入力部 1 2 3 は、通信網 2 1 4 を介して送信されてきたサブ送信信号を、サブ受信信号として受信することになる。

【 0 2 4 7 】

以上のように、図 1 9 の例では、プッシュ型の通信で、互いに異なる伝送経路で送信されたメインコンテンツとサブコンテンツを、受信側において簡単かつ高精度に同期させて再生することができる。なお、この例では、サブ送出信号は、メイン送出信号に先んじて送出される必要がある。すなわち、メイン送出信号のメイン受信機器 2 1 2 への到着時刻と、サブ送出信号のサブ受信機器 2 1 5 への到着時刻の差（到着時間差）を考慮した時間差で、メイン送出信号とサブ送出信号の送信が行われる必要がある。

【 0 2 4 8 】

本技術の適用例 2

また、本技術は、例えば図 2 0 に示すシステムにも適用可能である。なお、図 2 0 において、図 1 9 における場合と対応する部分には同一の符号を付してあり、その説明は省略する。

【 0 2 4 9 】

図 2 0 の例では、図 1 9 の例と同様に、提供装置 2 1 1 から放送波により、つまりプッシュ型の通信によりメイン送出信号がメイン受信機器 2 1 2 に送信される。

【 0 2 5 0 】

これに対して、サブ送出信号は、サーバ 2 4 1 により通信網 2 1 4 を介してサブ受信機器 2 1 5 に送信される。なお、サーバ 2 4 1 は、何らかの方法により、予めサブ送出信号を提供装置 2 1 1 等から取得して記録している。

【 0 2 5 1 】

この例では、サブ送出信号は、いわゆるプル型の通信により送信される。したがって、サーバ 2 4 1 は、サブ受信機器 2 1 5 からサブ送出信号の送信要求があったとき、通信網 2 1 4 を介して、サブ送出信号をサブ受信機器 2 1 5 に送信する。

【 0 2 5 2 】

すなわち、サブ受信機器 2 1 5 に対応する図 4 のサブ受信機器 8 5 の入力部 1 2 3 は、サーバ 2 4 1 にサブ送出信号の送信要求を送信するとともに、その送信要求に応じてサーバ 2 4 1 から送信されてきたサブ送出信号を、サブ受信信号として受信する。

【 0 2 5 3 】

この場合、サブ受信機器 2 1 5 は、メインコンテンツの放送前に予めサブ送出信号を受信して記録しておくことができる。したがって、予めサブ送出信号を受信して記録しておけば、メインコンテンツの放送時に、通信網 2 1 4 の状態等によってサブコンテンツをメインコンテンツと同期して再生させることができないなどの事態を防止することができる。

【 0 2 5 4 】

サブ受信機器 2 1 5 は、メイン受信機器 2 1 2 でのメインコンテンツの再生が開始されると、スピーカ 2 1 3 から出力されたメインコンテンツの音声を收音して音声同期用特徴量を計算する。そして、サブ受信機器 2 1 5 は、得られた音声同期用特徴量と、サブ受信信号に含まれている音声同期用特徴量とを用いて音声特徴量に基づく同期補正情報を生成し、同期補正情報を用いてサブコンテンツを、メインコンテンツと同期させて再生させる。

【 0 2 5 5 】

このように図 2 0 の例では、サブ受信機器 2 1 5 が自身に都合のよいタイミングでサブ受信信号を取得することができる。

【 0 2 5 6 】

本技術の適用例 3

また、本技術は、例えば図 2 1 に示すシステムにも適用可能である。なお、図 2 1 において、図 2 0 における場合と対応する部分には同一の符号を付してあり、その説明は省略する。

【 0 2 5 7 】

図 2 1 の例では、メインコンテンツ、つまりメイン送出信号は、サーバ 2 4 1 とは異なるサーバ 2 7 1 により提供される。すなわち、サーバ 2 7 1 はメイン受信機器 2 1 2 から要求があったとき、通信網 2 7 2 を介して、記録しているメイン送出信号をメイン受信機器 2 1 2 に送信する。つまり、この例ではメイン送出信号はプル型の通信により送信される。

10

【 0 2 5 8 】

具体的には、メイン受信機器 2 1 2 に対応する図 4 のメイン受信機器 8 1 の入力部 1 1 は、サーバ 2 7 1 にメイン送出信号の送信要求を送信するとともに、その送信要求に応じてサーバ 2 7 1 から送信されてきたメイン送出信号を、メイン受信信号として受信する。

【 0 2 5 9 】

この場合、メイン受信機器 2 1 2 は、予めメイン送出信号を受信して記録しておくことができる。したがって、予めメイン送出信号を受信して記録しておけば、メインコンテンツの再生時に通信網 2 7 2 の状態等によってメインコンテンツの再生が途中で途切れたり、停止したりするなどの事態を防止することができる。

20

【 0 2 6 0 】

また、サブ送出信号は図 2 0 の例と同様に、サーバ 2 4 1 によってプル型の通信により送信される。

【 0 2 6 1 】

サブ受信機器 2 1 5 は、メイン受信機器 2 1 2 でのメインコンテンツの再生が開始されると、スピーカ 2 1 3 から出力されたメインコンテンツの音声を収音して音声同期用特徴量を計算する。そして、サブ受信機器 2 1 5 は、得られた音声同期用特徴量と、サブ受信信号に含まれている音声同期用特徴量とを用いて音声特徴量に基づく同期補正情報を生成し、同期補正情報を用いてサブコンテンツを、メインコンテンツと同期させて再生させる。

30

【 0 2 6 2 】

このように図 2 1 の例では、メイン受信機器 2 1 2 とサブ受信機器 2 1 5 が、それぞれ自身に都合のよいタイミングでメイン受信信号とサブ受信信号を取得することができる。

【 0 2 6 3 】

なお、仮に通信網 2 7 2 が通信網 2 1 4 と同一の通信網であったとしても、メイン送信信号とサブ送信信号の送信タイミングや受信する機器等が異なれば、通常、これらのメイン送信信号とサブ送信信号の伝送経路は異なる経路となる。

40

【 0 2 6 4 】

第 2 の実施の形態

本技術の特徴

ところで、上述した (A 1) 乃至 (A 4) に示した例のように時間同期関係を有する複数のメディアコンテンツを、放送や IP (Internet Protocol) 網などを通じて複数機器で受信し、受信したメディアコンテンツを同期して再生するというアプリケーションプログラムが想定される。

【 0 2 6 5 】

このような機能性の実現のために、Hybridcastのように放送でコンテンツを配信すると同時に、IP 網により個別に追加コンテンツを配信し、コンテンツ受信機では、放送により

50

配信されたコンテンツと、IP網により配信された追加コンテンツを時間的に同期させて同時に出力するという放送通信連携サービスに向けたシステムの研究開発がされている。

【0266】

例えば、Hybridcastについては「松村欣司、鹿喰善明、Michael J Evans, 「インターネット配信情報との連動による放送番組パーソナライズシステムの検討」、映像情報メディア学会年次大会講演予稿集、2009年8月26日、p. 3 - 8」(以下、非特許文献1とも称する)に記載されている。

【0267】

また、「日本放送協会, 「HybridcastTMの概要と技術」, NHK技研R&D, no.124, p.10-17, 2010年11月, 日本放送出版協会, <http://www.nhk.or.jp/str1/publica/rd/rd124/PDF/P10-17.pdf>」(以下、非特許文献2とも称する)や、「日本放送協会, 「HybridcastTMを支える技術」, NHK技研R&D, no.133, p.20-27, 2012年5月, 日本放送出版協会, <http://www.nhk.or.jp/str1/publica/rd/rd133/PDF/P20-27.pdf>」(以下、非特許文献3とも称する)などにもHybridcastについて記載されている。

【0268】

Hybridcastでは、放送ストリームの基準クロック(PCR(Program Clock Reference))に基づく提示時間情報(PTS(Presentation Time Stamp))を付加した追加コンテンツを放送コンテンツの送出と同時に、あるいは少し先んじてストリーミング配信し、受信機で、通信コンテンツの遅延と変動を吸収するために十分な量のバッファを持ち、放送コンテンツを遅らせ、両者のタイムスタンプを比較することで同期をとることを基本原理としている。

【0269】

例えば、非特許文献2によれば、両受信機が同一機器内にある試作環境において1映像フレーム内(33ms)程度の精度で同期がとれることが確認できている。

【0270】

追加コンテンツを受信する機器は、IP網に無線接続されるスマートホンやタブレット型のパーソナルコンピュータといった放送コンテンツの受信機と独立な機器でもよい。そのような場合には、放送コンテンツ受信機は、追加コンテンツを受信する機器に対して、提示時刻情報(タイムスタンプ)を提供する必要がある。これは通常IP網を介して連携される。

【0271】

また、放送でなくともIP網などのネットワーク経由のみで複数コンテンツを配信して、協定世界時(UTC(Coordinated Universal Time))を基準クロックとしてタイムスタンプを付加し、受信機側で同期を行い、出力するシステムの実現も容易に想像できる。

【0272】

実際、上記のような放送通信連携サービスを独立した受信機で利用する場合、タイムスタンプの比較による方法では、以下の2つの要因により厳密な同期をとることが困難である。

【0273】

まず、第1に、放送コンテンツ受信機と追加コンテンツ受信機は独立した電子機器である以上、システムクロックに差異があり、時間の経過とともに同期ずれが発生する。

【0274】

また、第2に、ユーザはテレビジョン受像機などの放送コンテンツ受信機からある程度距離をおき、スマートホンやタブレット型パーソナルコンピュータなどの追加コンテンツ受信機を手元に持ち、IP網経由で配信される追加コンテンツを楽しむという使用形態が想定される。この使用形態で放送コンテンツ、および追加コンテンツに音声信号が含まれる場合、ユーザの視聴位置で厳密な同期を取ることが困難になる。

【0275】

例えば、ユーザが放送コンテンツ受信機から10m離れている場合、放送コンテンツ受信機から出力された音声信号がユーザ位置に到達するには $10(\text{m})/340(\text{m/s}) = \text{約}30(\text{ms})$ の時間

10

20

30

40

50

を要することになる。ここで、音速は約340(m/s)である。

【0276】

また、第1の実施の形態では、放送コンテンツ受信機が出力する音声を追加コンテンツ受信機が收音し、音声同期用特徴量を計算して、IP網で配信されてくる放送コンテンツの音声同期用特徴量と同期計算を行う手法となっている。しかし、IP網の伝送遅延や、ゆらぎなどが大きい場合には、広範囲にわたり同期位置のサーチを行う必要があり、処理量が多くなってしまう。

【0277】

そこで、上述した提供装置とコンテンツ再生システムが、以下の特徴B11乃至特徴B20を有するようにすることで、異なる経路で取得した複数のコンテンツを、さらに少ない処理量で同期させることができるようになる。

10

【0278】

(特徴B11)

メディアコンテンツは映像、音声、画像、文字情報などを多重化したデータストリームとされている。

【0279】

なお、この(特徴B11)のデータストリームの伝送としては、放送波、インターネットなどのネットワークにおけるメディアコンテンツの伝送を想定し、多重化データストリームが占有する論理伝送路を伝送路と呼ぶこととする。

20

【0280】

(特徴B12)

伝送対象とする複数メディアコンテンツは時間同期関係を有する。

【0281】

(特徴B13)

送出対象とする複数のメディアコンテンツのうち少なくとも1つをメインチャンネル信号と定め、残りの各メディアコンテンツをサブチャンネル信号とする。

【0282】

(特徴B14)

基準時刻信号からメインチャンネル信号、およびサブチャンネル信号のそれぞれについて提示時刻情報(PTC)を生成する。

30

【0283】

ここで、基準時刻信号は放送ストリームの基準クロック(PCR)または協定世界時(UTC)などが用いられる。

【0284】

(特徴B15)

メインチャンネル信号の提示時刻情報をメインチャンネル信号と多重化し、メイン送出信号を生成して伝送する。一方、メインチャンネル信号の音声信号から音声同期用特徴量も算出しておく。

【0285】

(特徴B16)

メインチャンネル信号とサブチャンネル信号の時間同期関係が符合するようにし、システムが規定する伝送フォーマットにより、サブチャンネル信号の提示時刻情報とメインチャンネル信号の音声同期用特徴量とサブチャンネル信号の多重化処理を行い、サブ送出信号を生成する。

40

【0286】

(特徴B17)

メイン受信機器はメイン受信信号を取得して分離し、メインチャンネル信号の再生時において、その音声信号に基づく音声をスピーカなどにより出力する。同時にメイン受信機器は、受信したメインチャンネル信号の提示時刻情報を外部より参照したり、取得したりできるよう提示する。

50

【 0 2 8 7 】

例えばメインチャンネル信号の提示時刻情報はソフトウェアのAPI (Application Programming Interface) によりその取得手段が提供され、無線通信によるIP網接続経由などで外部から参照できるようにしておく。

【 0 2 8 8 】

(特徴 B 1 8)

サブ受信機器は、サブ受信信号を取得して分離し、受信したサブチャンネル信号の提示時刻情報とメイン受信機器から取得したメインチャンネル信号の提示時刻情報を比較し、提示時刻情報に基づく同期補正情報を生成する。

【 0 2 8 9 】

(特徴 B 1 9)

サブ受信機器は、メイン受信機器がスピーカから出力したメインチャンネル信号の音声をマイクロホンなどにより収音して、音声同期用特徴量を計算し、(特徴 B 1 8) で生成された提示時刻情報に基づく同期補正情報を考慮して、受信したメインチャンネル信号の音声同期用特徴量との自動同期計算を行い、音声特徴量に基づく同期補正情報 (時間差情報) を算出する。

【 0 2 9 0 】

提示時刻情報の比較で得られる提示時刻情報に基づく同期補正情報から、おおまかな同期位置が分かるので、後段の音声同期用特徴量による自動同期計算処理に要する処理量も少なく済む。

【 0 2 9 1 】

(特徴 B 2 0)

上記音声特徴量に基づく同期補正情報に基づき、サブ受信機器は受信したサブチャンネル信号に対してメインチャンネル信号との同期補正処理を行い再生する。

【 0 2 9 2 】

提供装置の構成例

次に、以上において説明した特徴 B 1 1 乃至特徴 B 2 0 を有する提供装置とコンテンツ再生システムの具体的な実施の形態について説明する。

【 0 2 9 3 】

図 2 2 は、上述した (A 1) 乃至 (A 4) に示した例のように時間同期関係を有するコンテンツを提供する提供装置の構成例を示す図である。なお、図 2 2 において、図 1 における場合と対応する部分には同一の符号を付してあり、その説明は適宜省略する。

【 0 2 9 4 】

提供装置 3 0 1 は、基準時刻信号生成部 3 1 1、多重化処理部 3 1 2、出力部 2 2、音声同期用特徴量計算部 2 3、多重化処理部 2 4、および出力部 2 5 を有している。

【 0 2 9 5 】

提供装置 3 0 1 の構成は、提供装置 1 1 の変換部 2 1 が多重化処理部 3 1 2 に置き換えられ、さらに新たに基準時刻信号生成部 3 1 1 が設けられている点で、提供装置 1 1 の構成と異なっている。

【 0 2 9 6 】

基準時刻信号生成部 3 1 1 は、PCRやUTCに基づいて、メインチャンネル信号とサブチャンネル信号のコンテンツ提示のタイミングを示す提示時刻情報を生成し、多重化処理部 3 1 2 および多重化処理部 2 4 に供給する。例えば、提示時刻情報はPTSなどとされ、この提示時刻情報は再生側において、メインチャンネル信号とサブチャンネル信号の同期をとるために利用される。

【 0 2 9 7 】

多重化処理部 3 1 2 は、供給されたメインチャンネル信号を、所定の放送規格などで定められたフォーマットに変換する。また、多重化処理部 3 1 2 は、フォーマット変換されたメインチャンネル信号と、基準時刻信号生成部 3 1 1 から供給された提示時刻情報とを多重化することでメイン送出信号を生成し、出力部 2 2 に供給する。メイン送出信号に含

10

20

30

40

50

まれている提示時刻情報は、メインチャンネル信号の提示時刻情報である。

【0298】

また、多重化処理部24は、時間的に同期がとれた状態で、音声同期用特徴量計算部23から供給された音声同期用特徴量、供給されたサブチャンネル信号、および基準時刻信号生成部311から供給された提示時刻情報を多重化した後、必要に応じてフォーマット変換を行ってサブ送出信号を生成する。多重化処理部24は、得られたサブ送出信号を出力部25に供給する。サブ送出信号に含まれている提示時刻情報は、サブチャンネル信号の提示時刻情報である。

【0299】

なお、提供装置11における場合と同様に、多重化処理部24がメインチャンネル信号を用いて、音声同期用特徴量、サブチャンネル信号、および提示時刻情報の時間同期関係を調整してもよい。

【0300】

コンテンツ再生システムの構成例

また、提供装置301から送信されるメイン送出信号とサブ送出信号を、それぞれメイン受信信号およびサブ受信信号として受信してメインコンテンツとサブコンテンツを再生するコンテンツ再生システムは、例えば図23に示すように構成される。なお、図23において、図4における場合と対応する部分には同一の符号を付してあり、その説明は適宜省略する。

【0301】

図23に示すコンテンツ再生システムは、メイン受信機器341、表示部82、スピーカ83、マイクロホン84、サブ受信機器342、表示部86、およびスピーカ87を有している。

【0302】

メイン受信機器341は、提供装置301から送信されたメイン受信信号を受信し、メイン受信信号から得られるメインコンテンツの再生を制御する。

【0303】

メイン受信機器341は、入力部111、分離処理部351、提示部352、および再生処理部112を備えている。このメイン受信機器341の構成は、新たに分離処理部351と提示部352が設けられている点で、メイン受信機器81の構成と異なる。

【0304】

分離処理部351は、入力部111から供給されたメイン受信信号を、メインチャンネル信号と、そのメインチャンネル信号の提示時刻情報とに分離し、メインチャンネル信号を再生処理部112に供給するとともに、提示時刻情報を提示部352に供給する。

【0305】

提示部352は、分離処理部351から供給された提示時刻情報を、インターネットなどの有線の通信網や、無線通信網を介してサブ受信機器342に提示する。すなわち、通信相手からの要求に応じて提示時刻情報が送信される。

【0306】

また、サブ受信機器342は、提供装置301から送信されたサブ送出信号を、サブ受信信号として受信し、サブ受信信号から得られるサブコンテンツの再生を制御する。

【0307】

サブ受信機器342は、取得部361、提示時刻情報比較部362、音声同期用特徴量計算部121、バッファ122、入力部123、分離処理部124、バッファ125、同期計算部126、および再生処理部127を備えている。

【0308】

サブ受信機器342の構成は、新たに取得部361、および提示時刻情報比較部362が設けられている点で、サブ受信機器85の構成と異なる。

【0309】

取得部361は、APIなどを利用して、提示部352により提示された提示時刻情報を

10

20

30

40

50

、有線または無線の通信網を介して取得し、提示時刻情報比較部 3 6 2 に供給する。すなわち、取得部 3 6 1 は、提示部 3 5 2 により送信された提示時刻情報を受信する。

【 0 3 1 0 】

分離処理部 1 2 4 は、入力部 1 2 3 から供給されたサブ受信信号を、音声同期用特徴量、サブチャンネル信号、および提示時刻情報に分離させ、提示時刻情報を提示時刻情報比較部 3 6 2 に供給するとともに、音声同期用特徴量およびサブチャンネル信号をバッファ 1 2 5 に供給する。

【 0 3 1 1 】

提示時刻情報比較部 3 6 2 は、分離処理部 1 2 4 から供給された提示時刻情報と、取得部 3 6 1 から供給された提示時刻情報とを比較して、メインチャンネル信号とサブチャンネル信号とを同期させるための提示時刻情報に基づく同期補正情報を生成し、同期計算部 1 2 6 に供給する。

10

【 0 3 1 2 】

この提示時刻情報に基づく同期補正情報は、それ自体でメインチャンネル信号とサブチャンネル信号とのずれを補正し、同期させることができるものである。しかし、この例では、より高精度にそれらの信号を同期させるため、提示時刻情報に基づく同期補正情報は、同期計算部 1 2 6 において、バッファ 1 2 5 から読み出す音声同期用特徴量の範囲を定めるために用いられる。換言すれば、バッファ 1 2 5 に記録されている音声同期用特徴量と、バッファ 1 2 2 に記録されている音声同期用特徴量との大まかな同期をとるために利用される。このように、提示時刻情報に基づく同期補正情報を用いることで、より少ない処理量で音声同期用特徴量のマッチング処理を行うことができるようになる。

20

【 0 3 1 3 】

送信処理の説明

続いて、以上において説明した提供装置 3 0 1 とコンテンツ再生システムの具体的な動作について説明する。

【 0 3 1 4 】

まず、図 2 4 のフローチャートを参照して、提供装置 3 0 1 により行われる送信処理について説明する。

【 0 3 1 5 】

ステップ S 1 9 1 において、基準時刻信号生成部 3 1 1 は、メインチャンネル信号とサブチャンネル信号の提示時刻情報を生成し、多重化処理部 3 1 2 および多重化処理部 2 4 に供給する。

30

【 0 3 1 6 】

ステップ S 1 9 2 において、音声同期用特徴量計算部 2 3 は、音声同期用特徴量算出処理を行って、供給されたメインチャンネル信号を構成する音声信号から、音声同期用特徴量を計算し、多重化処理部 2 4 に供給する。なお、ステップ S 1 9 2 において行われる音声同期用特徴量算出処理は、図 1 4 を参照して説明した音声同期用特徴量算出処理と同様であるので、その説明は省略する。

【 0 3 1 7 】

ステップ S 1 9 3 において、多重化処理部 3 1 2 は、供給されたメインチャンネル信号と、基準時刻信号生成部 3 1 1 から供給された提示時刻情報とを多重化することでメイン送出信号を生成し、出力部 2 2 に供給する。また、このとき多重化処理部 3 1 2 は、必要に応じて、メインチャンネル信号のフォーマット変換を行う。

40

【 0 3 1 8 】

ステップ S 1 9 4 において、出力部 2 2 は、多重化処理部 3 1 2 から供給されたメイン送出信号を送信する。

【 0 3 1 9 】

ステップ S 1 9 5 において、多重化処理部 2 4 は、音声同期用特徴量、サブチャンネル信号、および提示時刻情報を多重化してサブ送出信号を生成し、出力部 2 5 に供給する。

【 0 3 2 0 】

50

すなわち、多重化処理部 2 4 は、音声同期用特徴量計算部 2 3 からの音声同期用特徴量、供給されたサブチャンネル信号、および基準時刻信号生成部 3 1 1 から供給された提示時刻情報を多重化してサブ送出信号とする。

【 0 3 2 1 】

ステップ S 1 9 6 において、出力部 2 5 は、多重化処理部 2 4 から供給されたサブ送出信号を送信し、送信処理は終了する。

【 0 3 2 2 】

以上のようにして、提供装置 3 0 1 は、メインチャンネル信号とサブチャンネル信号とで共通して用いられる提示時刻情報を生成し、提示時刻情報が含まれるメイン送出信号とサブ送出信号を生成する。

10

【 0 3 2 3 】

これにより、コンテンツの再生側において、提示時刻情報を利用して、より少ない処理量で、メインコンテンツとサブコンテンツを同期させることができるようになる。

【 0 3 2 4 】

メインコンテンツ再生処理の説明

また、提供装置 3 0 1 からメイン送出信号が送信されると、コンテンツ再生システムは、そのメイン送出信号を、メイン受信信号として取得して、メインコンテンツを再生する。以下、図 2 5 のフローチャートを参照して、コンテンツ再生システムによるメインコンテンツ再生処理について説明する。

【 0 3 2 5 】

20

ステップ S 2 2 1 において、入力部 1 1 1 は、メイン受信信号を取得して分離処理部 3 5 1 に供給する。例えば入力部 1 1 1 は、提供装置 3 0 1 から送信されたメイン受信信号を受信することで、メイン受信信号を取得する。

【 0 3 2 6 】

ステップ S 2 2 2 において、分離処理部 3 5 1 は、入力部 1 1 1 から供給されたメイン受信信号を、メインチャンネル信号と提示時刻情報とに分離する。分離処理部 3 5 1 は、分離されたメインチャンネル信号を再生処理部 1 1 2 に供給するとともに、提示時刻情報を提示部 3 5 2 に供給する。

【 0 3 2 7 】

ステップ S 2 2 3 において、再生処理部 1 1 2 は、分離処理部 3 5 1 から供給されたメインチャンネル信号に基づいてメインコンテンツを再生させる。なお、ステップ S 2 2 3 では、図 1 5 のステップ S 7 2 の処理と同様の処理が行われる。

30

【 0 3 2 8 】

ステップ S 2 2 4 において、提示部 3 5 2 は、分離処理部 3 5 1 から供給された提示時刻情報を提示して、メインコンテンツ再生処理は終了する。例えば、提示時刻情報は、メインコンテンツの再生と同期した状態で、無線等によりサブ受信機器 3 4 2 に送信される。

【 0 3 2 9 】

以上のようにして、コンテンツ再生システムは、メイン受信信号を取得してメインコンテンツを再生するとともに、メインコンテンツ、すなわちメインチャンネル信号の提示時刻情報の提示を行う。

40

【 0 3 3 0 】

このようにメインコンテンツの再生とともに、そのメインコンテンツの提示時刻情報を提示することで、その提示時刻情報を取得するサブ受信機器 3 4 2 は、より少ない処理量で、音声同期用特徴量を用いた同期計算を行うことができるようになる。

【 0 3 3 1 】

サブコンテンツ再生処理の説明

また、メインコンテンツの再生と同期して、コンテンツ再生システムは、サブ受信信号を取得して、サブコンテンツを再生する。以下、図 2 6 のフローチャートを参照して、コンテンツ再生システムによるサブコンテンツ再生処理について説明する。

50

【 0 3 3 2 】

なお、ステップ S 2 5 1 の処理は、図 1 6 のステップ S 1 0 1 の処理と同様であるので、その説明は省略する。

【 0 3 3 3 】

ステップ S 2 5 2 において、分離処理部 1 2 4 は、入力部 1 2 3 から供給されたサブ受信信号を、サブチャンネル信号、音声同期用特徴量、および提示時刻情報に分離させる。そして分離処理部 1 2 4 は、サブチャンネル信号と音声同期用特徴量をバッファ 1 2 5 に供給して記録させるとともに、サブチャンネル信号の提示時刻情報を提示時刻情報比較部 3 6 2 に供給する。

【 0 3 3 4 】

ステップ S 2 5 3 において、取得部 3 6 1 は、提示部 3 5 2 により送信された提示時刻情報を受信することで、メインチャンネル信号の提示時刻情報を取得し、提示時刻情報比較部 3 6 2 に供給する。

【 0 3 3 5 】

ステップ S 2 5 4 において、提示時刻情報比較部 3 6 2 は、分離処理部 1 2 4 から供給された提示時刻情報と、取得部 3 6 1 から供給された提示時刻情報とを比較して提示時刻情報に基づく同期補正情報を生成し、同期計算部 1 2 6 に供給する。

【 0 3 3 6 】

例えば提示時刻情報に基づく同期補正情報は、バッファ 1 2 5 に時系列に並べられて記録されている各時刻の音声同期用特徴量の系列のうち、同期計算部 1 2 6 での同期計算の対象とされる範囲（以下、探索範囲とも称する）を示す情報とされる。

【 0 3 3 7 】

この探索範囲は、現時点において再生されているメインコンテンツ、つまり取得部 3 6 1 により取得された最新の提示時刻情報と同じ時刻を示しているサブチャンネル信号の提示時刻情報に対応付けられている音声同期用特徴量を含む、所定長の音声同期用特徴量系列とされる。

【 0 3 3 8 】

提示時刻が同じであるメインチャンネル信号とサブチャンネル信号の位置は、互いに同期する信号位置、つまり同時に再生すべき再生位置（フレーム位置）である。したがって、提示時刻情報を比較して、メインチャンネル信号と同じ提示時刻情報を有するサブチャンネル信号の位置を検出することで、再生中のメインコンテンツと大まかに同期がとれたサブコンテンツの再生位置を特定することができる。

【 0 3 3 9 】

提示時刻情報が比較されて提示時刻情報に基づく同期補正情報が生成されると、その後、ステップ S 2 5 5 およびステップ S 2 5 6 の処理が行われるが、これらの処理は図 1 6 のステップ S 1 0 3 およびステップ S 1 0 4 の処理と同様であるので、その説明は省略する。なお、これらの処理では、メインコンテンツの音声が入音され、その音声から音声同期用特徴量が算出される。

【 0 3 4 0 】

ステップ S 2 5 7 において、同期計算部 1 2 6 は、同期補正情報生成処理を行って、音声特徴量に基づく同期補正情報を生成し、再生処理部 1 2 7 に供給する。なお、同期補正情報生成処理の詳細は後述するが、この処理では、提示時刻情報に基づく同期補正情報が用いられて、バッファ 1 2 2 に記録されている音声同期用特徴量と、バッファ 1 2 5 に記録されている音声同期用特徴量とが比較され、音声特徴量に基づく同期補正情報が生成される。

【 0 3 4 1 】

ステップ S 2 5 8 において、再生処理部 1 2 7 は、同期計算部 1 2 6 から供給された音声特徴量に基づく同期補正情報に基づいて、バッファ 1 2 5 に記録されているサブチャンネル信号の再生タイミングを補正し、補正後のサブチャンネル信号に基づいてサブコンテンツを再生させる。ステップ S 2 5 8 では、図 1 6 のステップ S 1 0 6 と同様の処理が行

10

20

30

40

50

われる。

【0342】

このようにして、メインコンテンツと同期してサブコンテンツが再生されると、サブコンテンツ再生処理は終了する。

【0343】

以上のようにして、コンテンツ再生システムは、メインコンテンツの提示時刻情報を取得して、サブ受信信号に含まれているサブコンテンツの提示時刻情報と比較することで、提示時刻情報に基づく同期補正情報を生成する。そして、コンテンツ再生システムは、提示時刻情報に基づく同期補正情報により示される探索範囲に含まれる音声同期用特徴量を対象としてマッチング処理を行い、音声特徴量に基づく同期補正情報を算出する。

10

【0344】

これにより、メインコンテンツとサブコンテンツとの伝送経路が異なる場合であっても、より少ない処理量で同期補正情報を算出し、それらのコンテンツを同期して再生させることができる。

【0345】

すなわち、コンテンツ再生システムでは、サブ受信機器342は、まず提示時刻情報によりメインチャンネル信号とサブチャンネル信号の大まかな同期をとり、さらにメインコンテンツの音声を收音して得られた音声信号から音声同期用特徴量を計算する。

【0346】

そして、サブ受信機器342は、受信したメインチャンネル信号の音声同期用特徴量との自動同期計算を行うことで、サブ受信機器342により再生されるサブコンテンツを視聴するユーザの視聴位置での高精度なコンテンツ同期が可能となる。実際に、提示時刻情報が用いられておおよその同期位置の範囲が絞られているので、音声同期用特徴量による自動同期計算処理に要する処理量も少なくて済む。

20

【0347】

例えば、コンテンツ再生システムにおいて表示部82およびスピーカ83と、表示部86およびスピーカ87とが離れた位置に配置されており、ユーザが表示部86およびスピーカ87の近傍でコンテンツを視聴しているとする。そのような場合、スピーカ83から出力された音声ユーザの視聴位置に到達するまでには、ある程度の時間を要する。

【0348】

したがって、そのような場合には、提示時刻情報を比較するだけでは、ユーザの視聴位置において、メインコンテンツとサブコンテンツの再生を高精度に同期させることは困難である。すなわち、例えばほぼ同じ時刻でスピーカ83とスピーカ87とで、メインコンテンツの音声と、サブコンテンツの音声とがそれぞれ再生されることになるので、メインコンテンツの音声ユーザに到達するまでに時間がかかってしまうと、ユーザには、メインコンテンツの音声とサブコンテンツの音声とがずれて聞こえてしまうことになる。

30

【0349】

これに対して、本技術を適用したコンテンツ再生システムでは、サブ受信機器342に接続され、サブ受信機器342近傍に配置されたマイクロホン84によりメインコンテンツの音声收音されて同期計算が行われる。そのため、コンテンツ再生システムでは、ユーザの視聴位置において同期がとれた状態でメインコンテンツとサブコンテンツを再生することができる。しかも、コンテンツ再生システムでは、提示時刻情報を比較して提示時刻情報に基づく同期補正情報を生成し、マッチング処理の探索範囲を限定することで、より少ない処理量でコンテンツを同期させることができる。

40

【0350】

同期補正情報生成処理の説明

さらに、図27のフローチャートを参照して、図26のステップS257の処理に対応する同期補正情報生成処理について説明する。

【0351】

ステップS281において、フレームレート変換部181およびフレームレート変換部

50

182は、必要に応じてフレームレート変換処理を行う。

【0352】

すなわち、フレームレート変換部181は、バッファ122からメインコンテンツの時間区間ごとの音声同期用特徴量の時系列データを読み出して、必要に応じて音声同期用特徴量をフレームレート変換、つまりダウンサンプルし、ブロック統合部183に供給する。

【0353】

また、フレームレート変換部182は、バッファ125に記録されている音声同期用特徴量の時系列データのうち、提示時刻情報比較部362から供給された提示時刻情報に基づく同期補正情報により示される探索範囲に含まれている時系列データのみを読み出す。

10

【0354】

そして、フレームレート変換部182は、読み出した音声同期用特徴量を必要に応じてフレームレート変換、つまりダウンサンプルまたはアップサンプルし、ブロック統合部184に供給する。

【0355】

ステップS282において、ブロック統合部183およびブロック統合部184は、音声同期用特徴量の時系列データを統合する。

【0356】

具体的には、ブロック統合部183は、フレームレート変換部181からメインコンテンツの時間区間ごとの音声同期用特徴量の時系列データの供給を受ける。そして、ブロック統合部183は、連続した複数（例えば64個）の時間区間を1ブロックとして、ブロック単位で、供給された時間区間ごとの音声同期用特徴量の時系列データを統合し、類似度計算部185に供給する。

20

【0357】

また、ブロック統合部184は、フレームレート変換部182からメインコンテンツの時間区間ごとの音声同期用特徴量の時系列データの供給を受ける。そして、ブロック統合部184は、連続した複数（例えば64個）の時間区間を1ブロックとして、ブロック単位で、供給された時間区間ごとの音声同期用特徴量の時系列データを統合し、類似度計算部185に供給する。

【0358】

30

例えば、図28の矢印A11に示すように、図18のステップS162の処理では、バッファ122に記録されているn個の各ブロック $X(i)$ と、バッファ125に記録されているm個の各ブロック $Y(j)$ とが処理対象とされていた。つまり、探索対象となる音声同期用特徴量のブロックの組み合わせは、 $n \times m$ 通りとされていた。なお、より詳細には、音声同期用特徴量に対して適宜、フレームレート変換が行われるが、図28では説明を簡単にするため、フレームレート変換は行われないものとして説明を続ける。

【0359】

ここで、マッチング処理の対象とされるブロック $Y(j)$ は、バッファ125に記録されている全てのブロック、または十分に広い範囲のブロックとされる。

【0360】

40

なお、図28において、 i は、音声同期用特徴量計算部121で得られた音声同期用特徴量のブロックのインデックスであり、 j は、サブ受信信号に含まれている音声同期用特徴量のブロックのインデックスである。

【0361】

また、 $X(i)$ は、音声同期用特徴量計算部121で得られた音声同期用特徴量のうちのインデックス i のブロックの音声同期用特徴量の時系列データを表し、 $Y(j)$ は、サブ受信信号に含まれている音声同期用特徴量のうちのインデックス j のブロックの音声同期用特徴量の時系列データを表す。

【0362】

一方、ステップS282では、矢印A12に示すように、バッファ125に記録されて

50

いる m 個の各ブロックのうち、提示時刻情報に基づく同期補正情報により示される探索範囲に含まれている m' 個のブロックのみがマッチング処理の対象とされる。すなわち、類似度計算部185での類似度計算の対象とされる。

【0363】

この例では、 PTS_i は提示時刻情報を表しており、この提示時刻情報により示される位置が、現時点で再生されているメインコンテンツの位置となっている。そして、そのメインコンテンツの提示時刻情報と同じ時刻のサブコンテンツの提示時刻情報に対応する位置を含む所定長の範囲、つまり m' 個のブロックからなる範囲が探索範囲とされている。したがって、探索対象となる音声同期用特徴量のブロックの組み合わせは、 $n \times m'$ 通りとなる。

10

【0364】

このように、提示時刻情報を比較して得られる提示時刻情報に基づく同期補正情報を用いれば、マッチング処理の対象とされる音声同期用特徴量の範囲を必要最小限に限定することができるので、類似度計算の探索に要する処理時間を大幅に低減させることができる。

【0365】

図27のフローチャートの説明に戻り、音声同期用特徴量の時系列データが統合されると、その後、処理はステップS283に進む。そして、ステップS283およびステップS284の処理が行われて同期補正情報生成処理は終了するが、これらの処理は図18のステップS163およびステップS164の処理と同様であるので、その説明は省略する。同期補正情報生成処理が終了すると、その後、処理は図26のステップS258へと進む。

20

【0366】

以上のようにして、コンテンツ再生システムは、提示時刻情報に基づく同期補正情報により示される探索範囲の音声同期用特徴量を用いて、音声特徴量に基づく同期補正情報を生成する。これにより、より少ない処理量で、同期補正情報をロバストに生成することができる。

【0367】

また、図22に示した提供装置301と図23に示したコンテンツ再生システムも、図19乃至図21に示した各システムに適用可能である。

30

【0368】

ところで、上述した一連の処理は、ハードウェアにより実行することもできるし、ソフトウェアにより実行することもできる。一連の処理をソフトウェアにより実行する場合には、そのソフトウェアを構成するプログラムが、コンピュータにインストールされる。ここで、コンピュータには、専用のハードウェアに組み込まれているコンピュータや、各種のプログラムをインストールすることで、各種の機能を実行することが可能な、例えば汎用のコンピュータなどが含まれる。

【0369】

図29は、上述した一連の処理をプログラムにより実行するコンピュータのハードウェアの構成例を示すブロック図である。

40

【0370】

コンピュータにおいて、CPU (Central Processing Unit) 501, ROM (Read Only Memory) 502, RAM (Random Access Memory) 503は、バス504により相互に接続されている。

【0371】

バス504には、さらに、入出力インターフェース505が接続されている。入出力インターフェース505には、入力部506、出力部507、記録部508、通信部509、及びドライブ510が接続されている。

【0372】

入力部506は、キーボード、マウス、マイクロホン、撮像素子などよりなる。出力部

50

507は、ディスプレイ、スピーカなどよりなる。記録部508は、ハードディスクや不揮発性のメモリなどよりなる。通信部509は、ネットワークインターフェースなどよりなる。ドライブ510は、磁気ディスク、光ディスク、光磁気ディスク、又は半導体メモリなどのリムーバブルメディア511を駆動する。

【0373】

以上のように構成されるコンピュータでは、CPU501が、例えば、記録部508に記録されているプログラムを、入出力インターフェース505及びバス504を介して、RAM503にロードして実行することにより、上述した一連の処理が行われる。

【0374】

コンピュータ(CPU501)が実行するプログラムは、例えば、パッケージメディア等としてのリムーバブルメディア511に記録して提供することができる。また、プログラムは、ローカルエリアネットワーク、インターネット、デジタル衛星放送といった、有線または無線の伝送媒体を介して提供することができる。

【0375】

コンピュータでは、プログラムは、リムーバブルメディア511をドライブ510に装着することにより、入出力インターフェース505を介して、記録部508にインストールすることができる。また、プログラムは、有線または無線の伝送媒体を介して、通信部509で受信し、記録部508にインストールすることができる。その他、プログラムは、ROM502や記録部508に、あらかじめインストールしておくことができる。

【0376】

なお、コンピュータが実行するプログラムは、本明細書で説明する順序に沿って時系列に処理が行われるプログラムであっても良いし、並列に、あるいは呼び出しが行われたとき等の必要なタイミングで処理が行われるプログラムであっても良い。

【0377】

また、本技術の実施の形態は、上述した実施の形態に限定されるものではなく、本技術の要旨を逸脱しない範囲において種々の変更が可能である。

【0378】

例えば、本技術は、1つの機能をネットワークを介して複数の装置で分担、共同して処理するクラウドコンピューティングの構成をとることができる。

【0379】

また、上述のフローチャートで説明した各ステップは、1つの装置で実行する他、複数の装置で分担して実行することができる。

【0380】

さらに、1つのステップに複数の処理が含まれる場合には、その1つのステップに含まれる複数の処理は、1つの装置で実行する他、複数の装置で分担して実行することができる。

【0381】

また、本明細書中に記載された効果はあくまで例示であって限定されるものではなく、他の効果があってもよい。

【0382】

さらに、本技術は、以下の構成とすることも可能である。

【0383】

(1)

第1のコンテンツの音声信号から特徴量を抽出する特徴量計算部と、

前記第1のコンテンツと時間同期関係を有する第2のコンテンツに対して同期がとれた状態で取得された前記特徴量と、前記特徴量計算部により抽出された前記特徴量とを比較することで、前記第2のコンテンツを前記第1のコンテンツと同期して再生するための音声特徴量に基づく同期補正情報を生成する同期計算部と

を備える情報処理装置。

(2)

10

20

30

40

50

前記特徴量計算部は、再生された前記第 1 のコンテンツの音声を収音することで得られた前記音声信号から前記特徴量を抽出する

(1) に記載の情報処理装置。

(3)

前記第 2 のコンテンツと、前記第 2 のコンテンツに同期がとれた状態で前記第 2 のコンテンツに対応付けられている前記特徴量とを取得する第 1 の入力部をさらに備える

(1) または (2) に記載の情報処理装置。

(4)

前記第 2 のコンテンツおよび前記特徴量は、前記第 1 のコンテンツとの到着時間差が考慮されたタイミングで前記情報処理装置に送信される

(3) に記載の情報処理装置。

(5)

前記第 1 の入力部は、前記第 2 のコンテンツおよび前記特徴量の送信を要求し、その要求に応じて送信されてきた前記第 2 のコンテンツおよび前記特徴量を受信する

(3) に記載の情報処理装置。

(6)

前記第 1 のコンテンツの送信を要求し、その要求に応じて送信されてきた前記第 1 のコンテンツを受信する第 2 の入力部をさらに備える

(5) に記載の情報処理装置。

(7)

前記特徴量計算部は、再生された 1 つの前記第 1 のコンテンツについて、前記音声信号から前記特徴量を抽出し、

前記同期計算部は、前記第 2 のコンテンツと対応付けられている複数の前記第 1 のコンテンツの前記特徴量のそれぞれと、前記特徴量計算部により抽出された前記特徴量とを比較することで、再生された前記第 1 のコンテンツを特定し、特定された前記第 1 のコンテンツと前記第 2 のコンテンツを同期して再生するための前記音声特徴量に基づく同期補正情報を生成する

(2) 乃至 (6) の何れか一項に記載の情報処理装置。

(8)

前記第 2 のコンテンツの再生を制御する再生処理部をさらに備える

(2) 乃至 (7) の何れか一項に記載の情報処理装置。

(9)

前記再生処理部は、前記音声特徴量に基づく同期補正情報に基づいて前記第 2 のコンテンツの再生位置を補正する

(8) に記載の情報処理装置。

(10)

前記第 1 のコンテンツの提示時刻情報を取得する取得部と、

前記第 1 のコンテンツの前記提示時刻情報と、前記第 2 のコンテンツの前記提示時刻情報とを比較して、提示時刻情報に基づく同期補正情報を生成する比較部と

をさらに備え、

前記同期計算部は、取得された前記特徴量の系列のうちの前記提示時刻情報に基づく同期補正情報により示される範囲に含まれる前記特徴量と、前記特徴量計算部により抽出された前記特徴量とを比較して前記音声特徴量に基づく同期補正情報を生成する

(1) 乃至 (9) の何れか一項に記載の情報処理装置。

(11)

前記同期計算部は、取得された前記特徴量と、前記特徴量計算部により抽出された前記特徴量とのフレームレートが一致するように、取得された前記特徴量、または前記特徴量計算部により抽出された前記特徴量の少なくとも一方に対してフレームレート変換を行ってから前記特徴量を比較する

(1) 乃至 (10) の何れか一項に記載の情報処理装置。

10

20

30

40

50

(1 2)

第 1 のコンテンツの音声信号から特徴量を抽出する特徴量計算ステップと、

前記第 1 のコンテンツと時間同期関係を有する第 2 のコンテンツに対して同期がとれた状態で取得された前記特徴量と、前記特徴量計算ステップの処理により抽出された前記特徴量とを比較することで、前記第 2 のコンテンツを前記第 1 のコンテンツと同期して再生するための音声特徴量に基づく同期補正情報を生成する同期計算ステップとを含む情報処理方法。

(1 3)

第 1 のコンテンツの音声信号から特徴量を抽出する特徴量計算ステップと、

前記第 1 のコンテンツと時間同期関係を有する第 2 のコンテンツに対して同期がとれた状態で取得された前記特徴量と、前記特徴量計算ステップの処理により抽出された前記特徴量とを比較することで、前記第 2 のコンテンツを前記第 1 のコンテンツと同期して再生するための音声特徴量に基づく同期補正情報を生成する同期計算ステップとを含む処理をコンピュータに実行させるプログラム。

10

(1 4)

第 1 のコンテンツの音声信号から特徴量を抽出する特徴量計算部と、

前記第 1 のコンテンツに対して時間同期関係を有する第 2 のコンテンツと、前記第 2 のコンテンツに対して同期がとれた状態で、前記第 2 のコンテンツに対応付けられた前記特徴量とを出力する第 1 の出力部と

を備える情報処理装置。

20

(1 5)

前記第 1 のコンテンツを出力する第 2 の出力部をさらに備える

(1 4) に記載の情報処理装置。

(1 6)

前記第 1 の出力部は、前記第 1 のコンテンツとの到着時間差が考慮されたタイミングで前記第 2 のコンテンツおよび前記特徴量を出力する

(1 5) に記載の情報処理装置。

(1 7)

前記第 1 の出力部は、前記第 2 のコンテンツおよび前記特徴量の送信が要求された場合、その要求に応じて前記第 2 のコンテンツおよび前記特徴量を出力する

(1 5) に記載の情報処理装置。

30

(1 8)

前記第 2 の出力部は、前記第 1 のコンテンツの送信が要求された場合、その要求に応じて前記第 1 のコンテンツを出力する

(1 7) に記載の情報処理装置。

(1 9)

前記特徴量計算部は、複数の前記第 1 のコンテンツについて、前記音声信号から前記特徴量を抽出し、

前記第 1 の出力部は、複数の前記第 1 のコンテンツの前記特徴量を前記第 2 のコンテンツに対応付けて出力する

(1 4) 乃至 (1 8) の何れか一項に記載の情報処理装置。

40

(2 0)

前記特徴量計算部は前記特徴量をダウンサンプルし、

前記第 1 の出力部は、前記第 2 のコンテンツと、ダウンサンプルされた前記特徴量とを出力する

(1 4) 乃至 (1 9) の何れか一項に記載の情報処理装置。

(2 1)

第 1 のコンテンツの音声信号から特徴量を抽出する特徴量計算ステップと、

前記第 1 のコンテンツに対して時間同期関係を有する第 2 のコンテンツと、前記第 2 のコンテンツに対して同期がとれた状態で、前記第 2 のコンテンツに対応付けられた前記特

50

微量とを出力する出力ステップと
を含む情報処理方法。

(2 2)

第 1 のコンテンツの音声信号から特徴量を抽出する特徴量計算ステップと、
前記第 1 のコンテンツに対して時間同期関係を有する第 2 のコンテンツと、前記第 2 の
コンテンツに対して同期がとれた状態で、前記第 2 のコンテンツに対応付けられた前記特
微量とを出力する出力ステップと
を含む処理をコンピュータに実行させるプログラム。

【符号の説明】

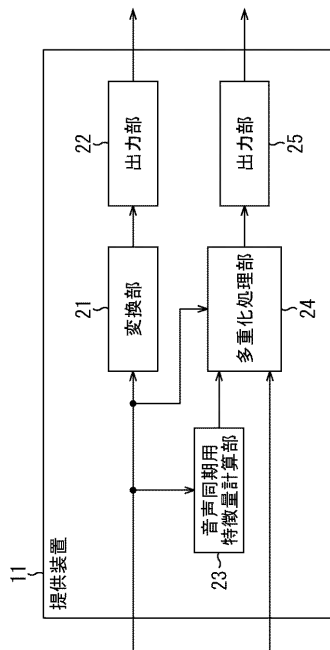
【 0 3 8 4 】

1 1 提供装置, 2 2 出力部, 2 3 音声同期用特徴量計算部, 2 4 多重化
処理部, 2 5 出力部, 8 1 メイン受信機器, 8 5 サブ受信機器, 1 1 1
入力部, 1 1 2 再生処理部, 1 2 1 音声同期用特徴量計算部, 1 2 3 入力部
, 1 2 6 同期計算部, 1 2 7 再生処理部, 3 1 1 基準時刻信号生成部, 3
5 2 提示部, 3 6 1 取得部, 3 6 2 提示時刻情報比較部

10

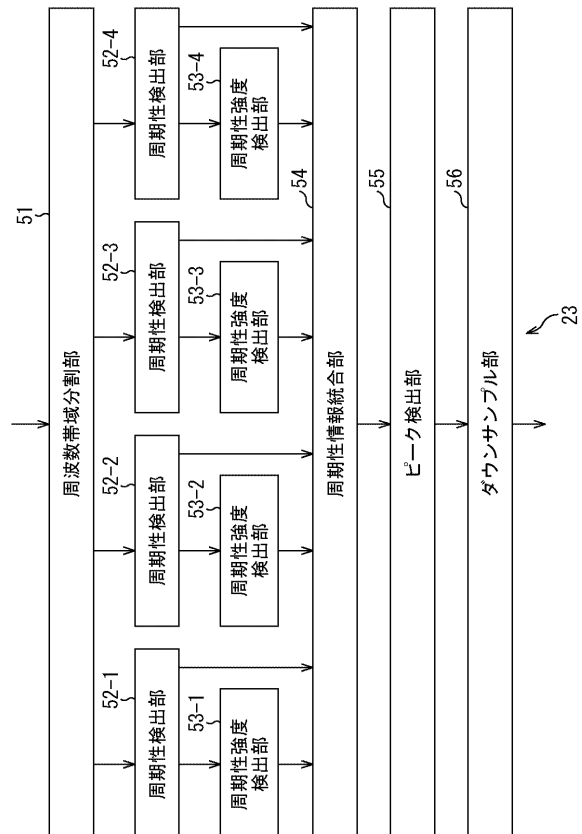
【 図 1 】

図1

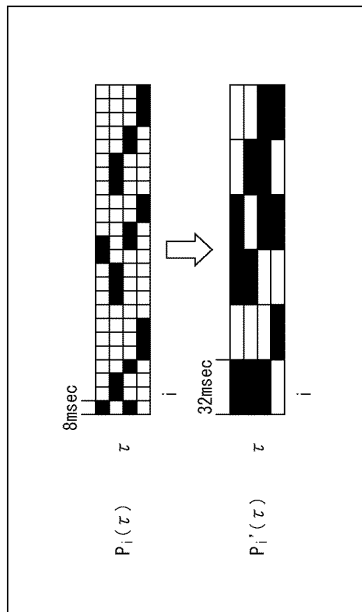


【 図 2 】

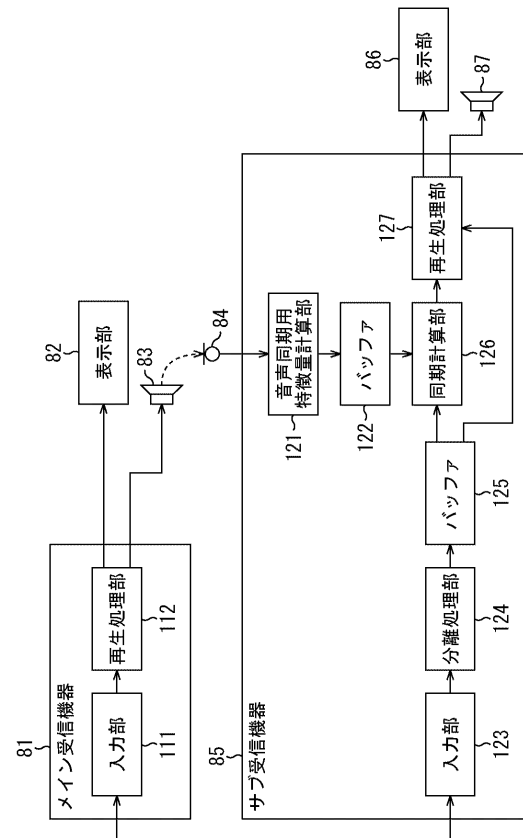
図2



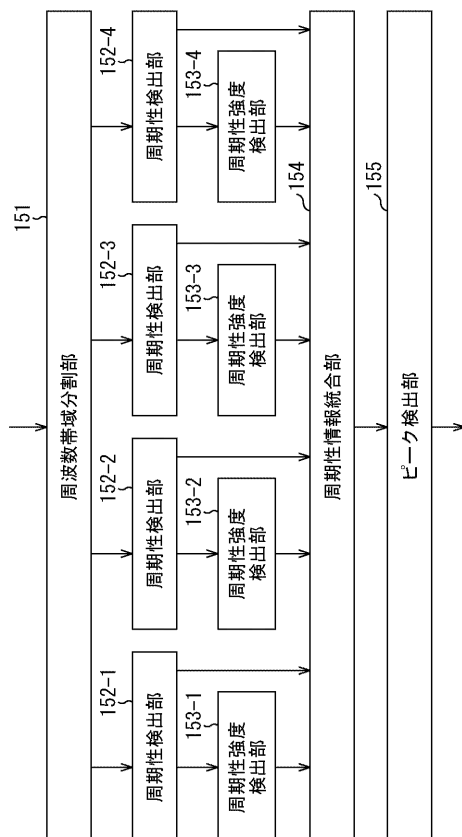
【図 3】
図3



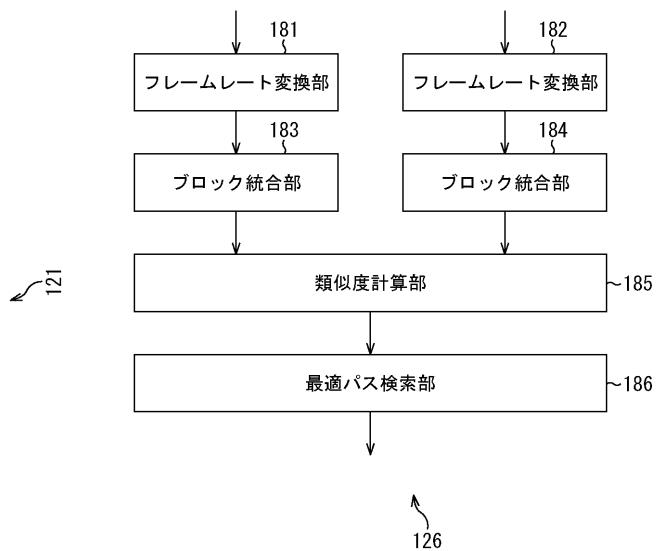
【図 4】
図4



【図 5】
図5

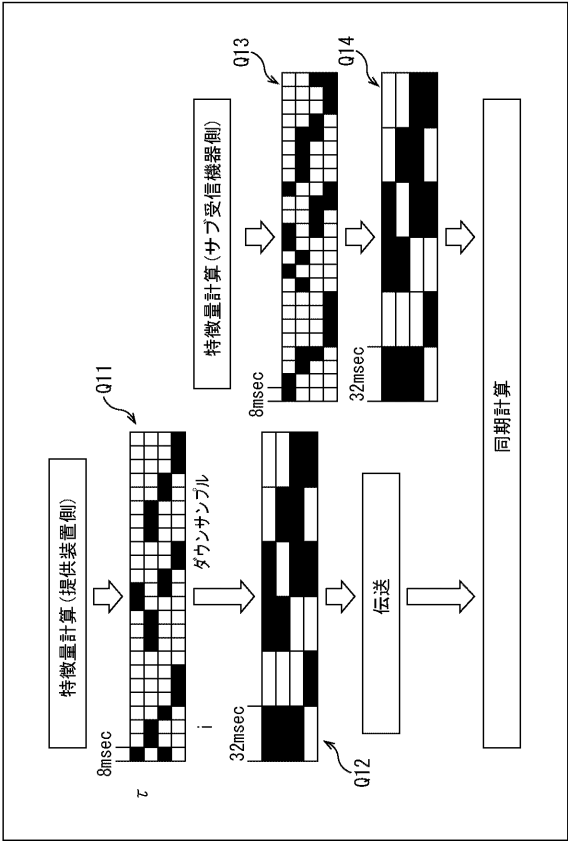


【図 6】
図6



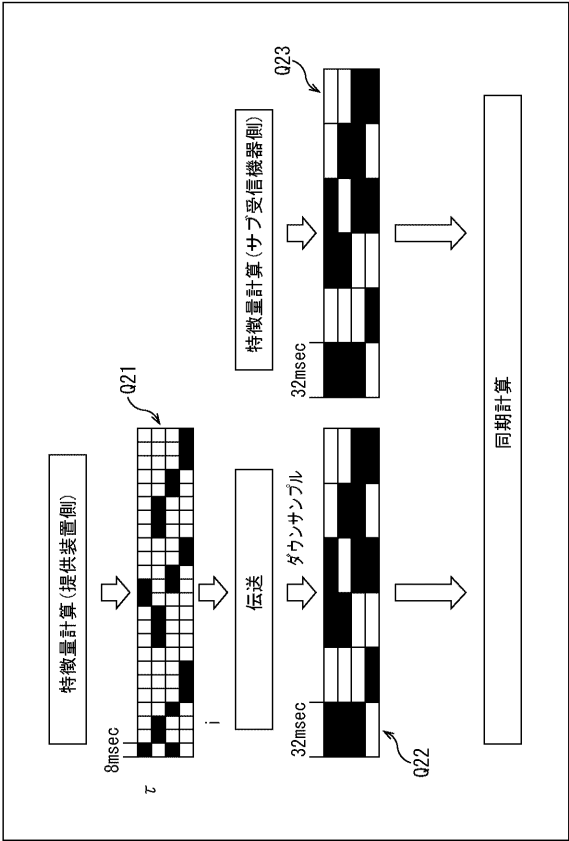
【図 7】

図7



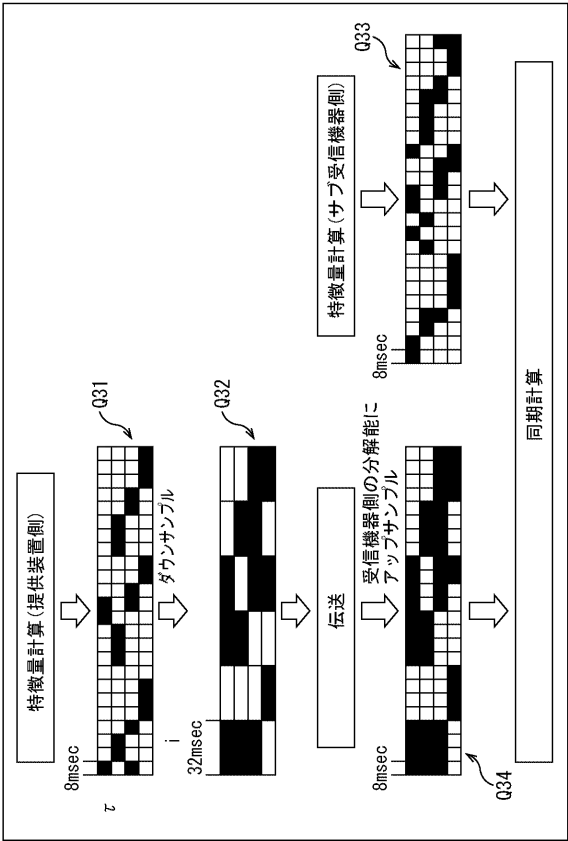
【図 8】

図8



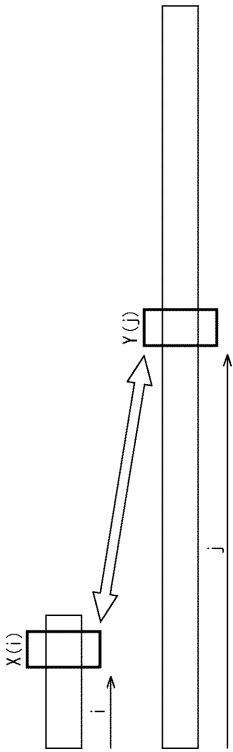
【図 9】

図9



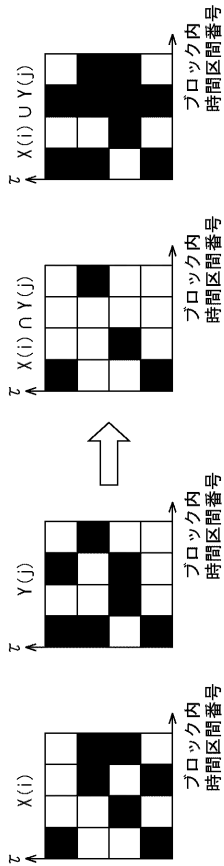
【図 10】

図10



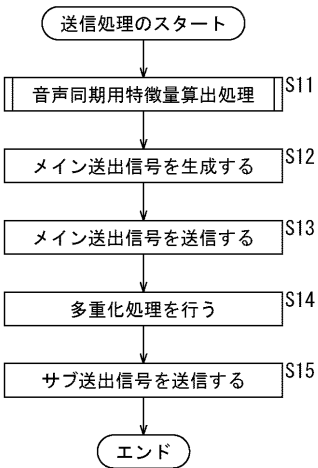
【図 1 1】

図11



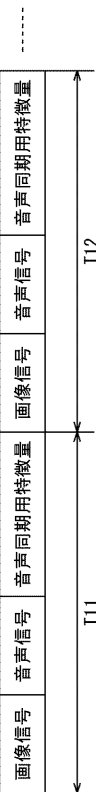
【図 1 2】

図12



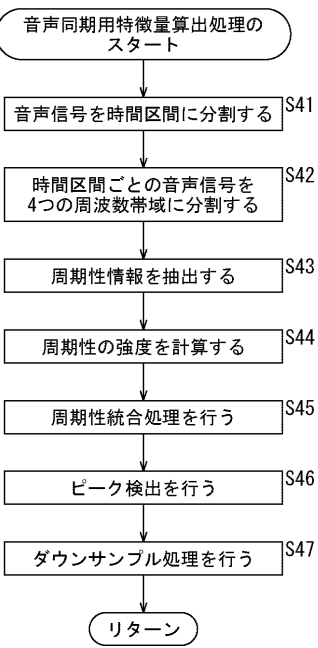
【図 1 3】

図13



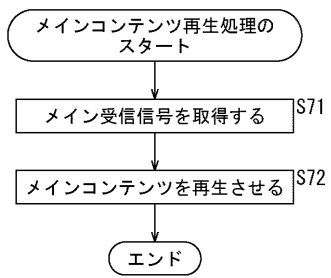
【図 1 4】

図14



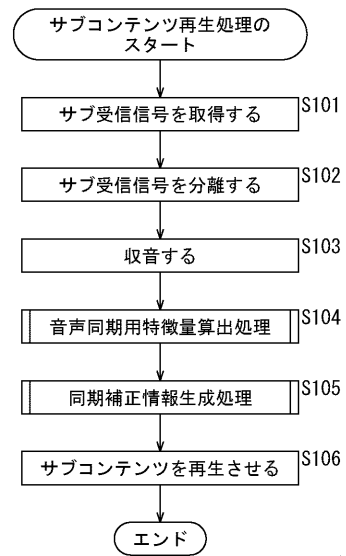
【図 15】

図15



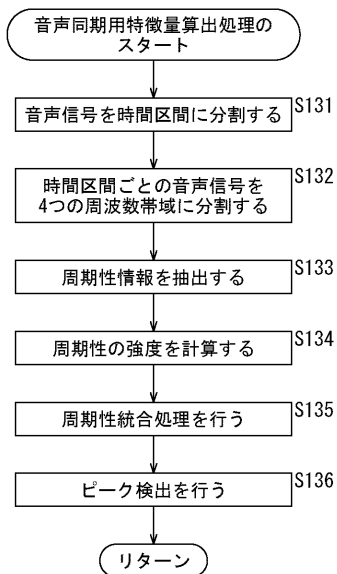
【図 16】

図16



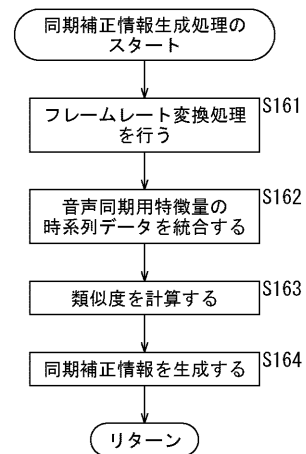
【図 17】

図17

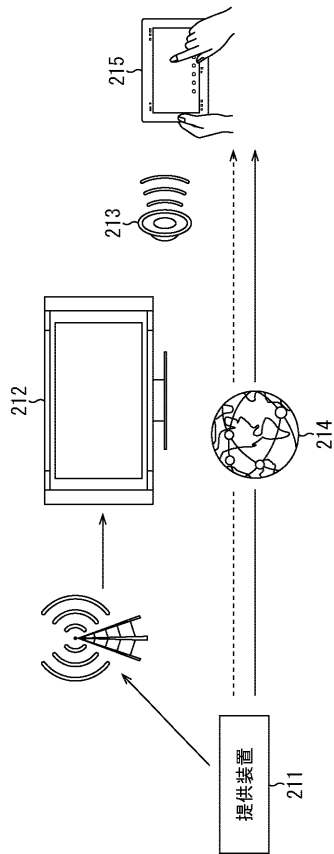


【図 18】

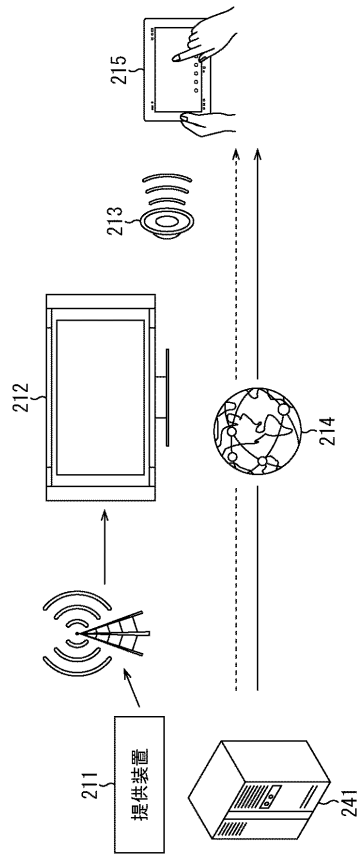
図18



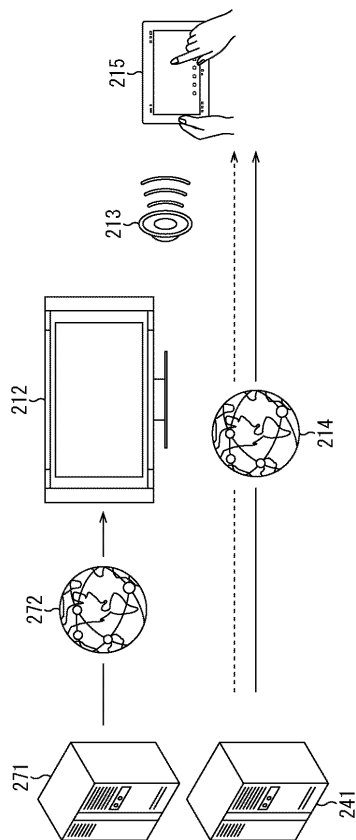
【図 19】
図19



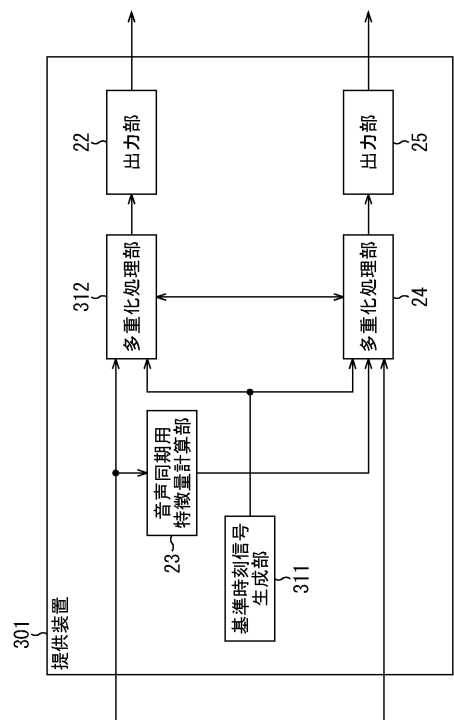
【図 20】
図20



【図 21】
図21

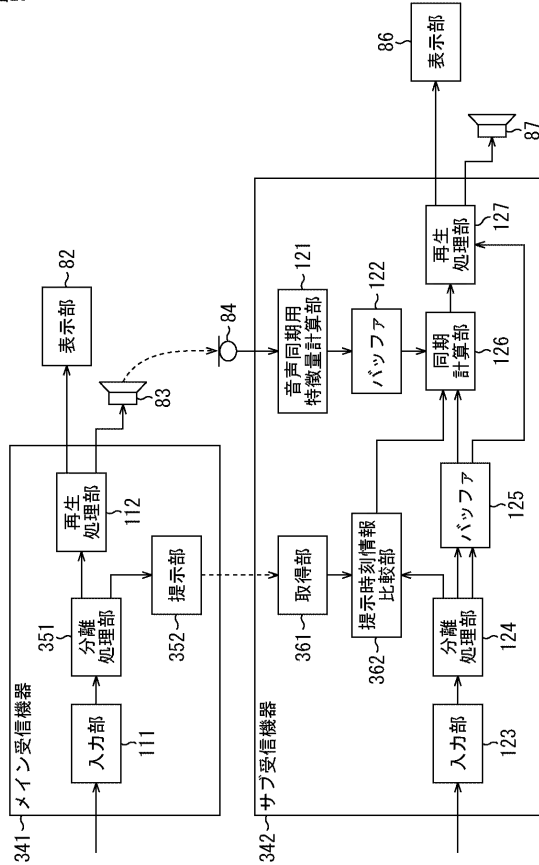


【図 22】
図22



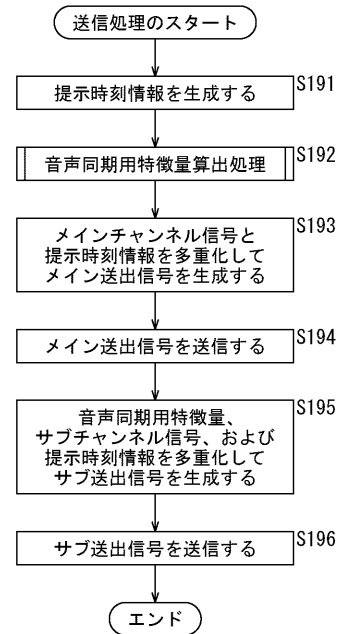
【図 23】

図23



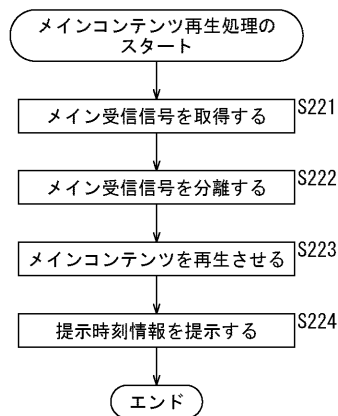
【図 24】

図24



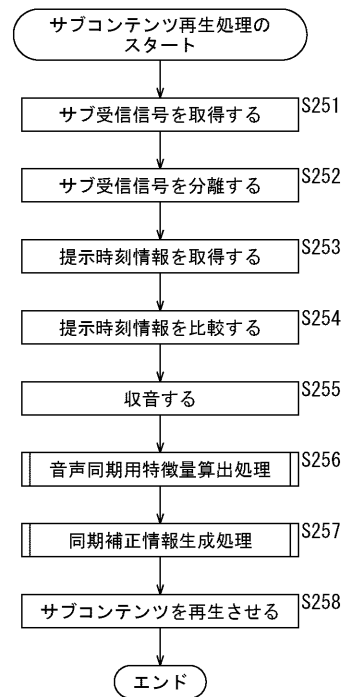
【図 25】

図25



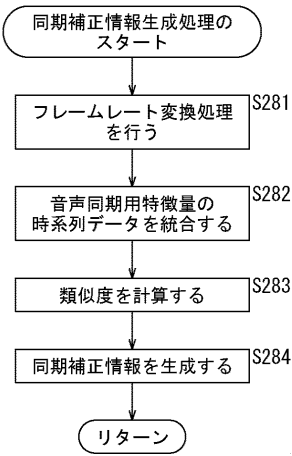
【図 26】

図26



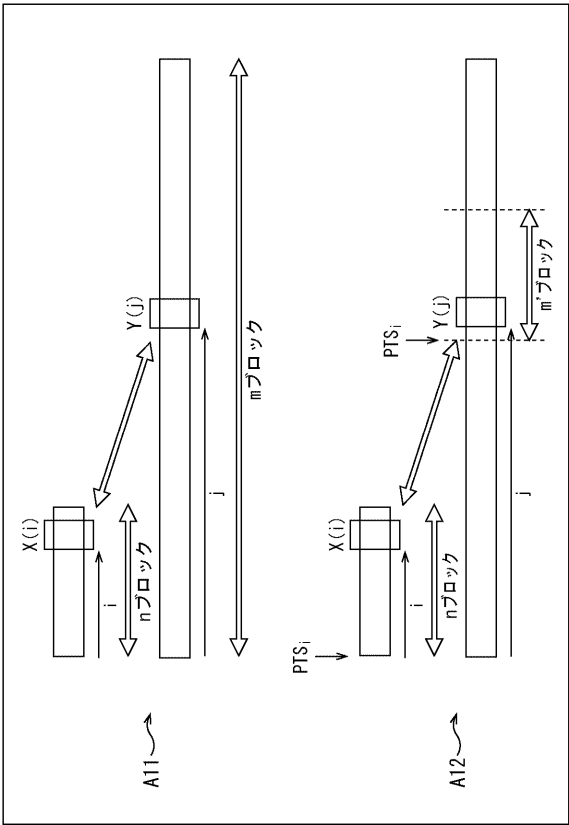
【図 27】

図27



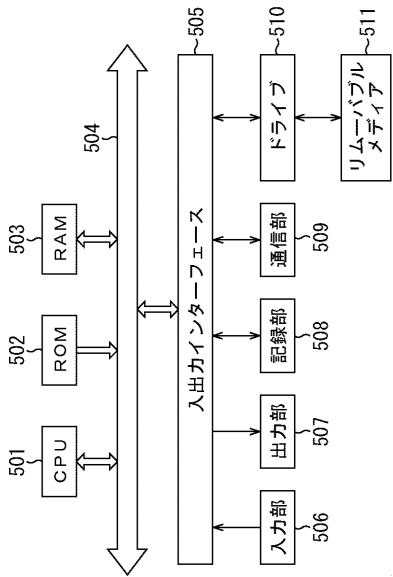
【図 28】

図28



【図 29】

図29



フロントページの続き

(72)発明者 西口 正之
東京都港区港南1丁目7番1号 ソニー株式会社内

審査官 富樫 明

(56)参考文献 特開2012-244413(JP,A)
特開2008-131341(JP,A)
特開2013-115782(JP,A)
米国特許出願公開第2007/0220561(US,A1)
特開2013-085139(JP,A)
特表2012-500584(JP,A)

(58)調査した分野(Int.Cl., DB名)
H04N 21/00 - 21/858
G10L 25/57