

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2010-136067

(P2010-136067A)

(43) 公開日 平成22年6月17日(2010.6.17)

(51) Int.Cl.	F I	テーマコード (参考)
HO4N 7/173 (2006.01)	HO4N 7/173 630	5C023
HO4N 5/445 (2006.01)	HO4N 5/445 Z	5C025
HO4N 5/265 (2006.01)	HO4N 5/265	5C053
HO4N 5/278 (2006.01)	HO4N 5/278	5C164
HO4N 5/93 (2006.01)	HO4N 5/93 Z	
審査請求 未請求 請求項の数 5 O L (全 18 頁) 最終頁に続く		

(21) 出願番号 特願2008-309578 (P2008-309578)
 (22) 出願日 平成20年12月4日 (2008.12.4)

(71) 出願人 000002185
 ソニー株式会社
 東京都港区港南1丁目7番1号
 (74) 代理人 100082131
 弁理士 稲本 義雄
 (74) 代理人 100121131
 弁理士 西川 孝
 (72) 発明者 浅野 康治
 東京都港区港南1丁目7番1号 ソニー株式会社内
 (72) 発明者 小林 賢一郎
 東京都港区港南1丁目7番1号 ソニー株式会社内

最終頁に続く

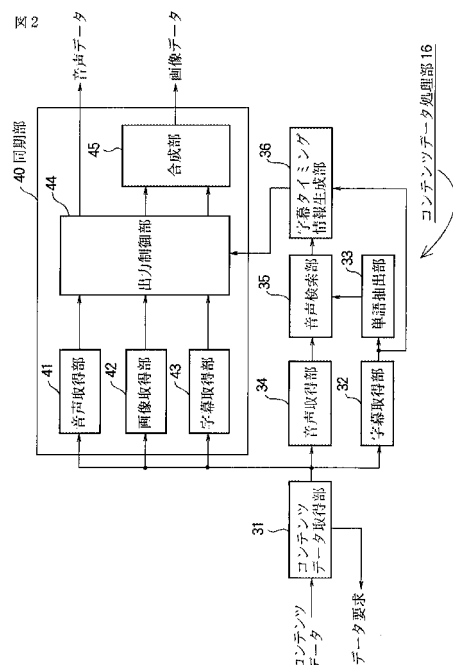
(54) 【発明の名称】 データ処理装置、データ処理方法、及び、プログラム

(57) 【要約】

【課題】コンテンツの画像と字幕とを、容易に、同期して表示等する。

【解決手段】単語抽出部33は、画像データ、音声データ、及び、字幕データを含むコンテンツデータのうちの字幕データから、検索対象単語を抽出し、音声検索部35が、音声データから、検索対象単語を発話した音声データを検索することにより、画像データにおける、検索対象単語の発話のタイミングを表すタイミング情報を取得し、字幕タイピング情報生成部36が、検索対象単語を含む字幕データの字幕識別情報と、検索対象単語の発話のタイミング情報とを対応付けた字幕タイピング情報を生成する。一方、出力制御部44は、字幕タイピング情報に従い、字幕データと、その字幕データの字幕識別情報と対応付けられているタイミング情報が表すタイミングの画像データとを同期させて出力する。本発明は、コンテンツを再生するレコーダ等に適用できる。

【選択図】図2



【特許請求の範囲】**【請求項 1】**

画像データ、音声データ、及び、字幕データを含むコンテンツデータのうちの前記字幕データから、検索の対象とする文字列を抽出し、

前記コンテンツデータのうちの前記音声データから、前記文字列を発話した音声データを検索し、前記画像データにおける、前記文字列の発話のタイミングを表すタイミング情報を取得し、

前記文字列を含む前記字幕データを識別する識別情報と、前記文字列の発話のタイミングを表すタイミング情報とを対応付けた字幕タイミング情報を生成する

ことにより得られる前記字幕タイミング情報に従い、前記字幕データと、その字幕データの識別情報と対応付けられているタイミング情報が表すタイミングの画像データとを同期させて出力する出力制御手段を備える

データ処理装置。

【請求項 2】

前記コンテンツデータのうちの前記字幕データから、検索の対象とする文字列を抽出する抽出手段と、

前記コンテンツデータのうちの前記音声データから、前記文字列を発話した音声データを検索し、前記画像データにおける、前記文字列の発話のタイミングを表すタイミング情報を取得する検索手段と、

前記文字列を含む前記字幕データを識別する識別情報と、前記文字列の発話のタイミングを表すタイミング情報とを対応付けた字幕タイミング情報を生成する生成手段と

をさらに備える請求項 1 に記載のデータ処理装置。

【請求項 3】

前記コンテンツデータが、前記字幕データを表示する表示時刻を表す表示時刻情報をさらに含む場合、

前記検索手段は、前記文字列を発話した音声データの検索を、前記表示時刻情報が表す表示時刻のタイミング付近の音声データに限定して行う

請求項 2 に記載のデータ処理装置。

【請求項 4】

データ処理装置が、

画像データ、音声データ、及び、字幕データを含むコンテンツデータのうちの前記字幕データから、検索の対象とする文字列を抽出し、

前記コンテンツデータのうちの前記音声データから、前記文字列を発話した音声データを検索し、前記画像データにおける、前記文字列の発話のタイミングを表すタイミング情報を取得し、

前記文字列を含む前記字幕データを識別する識別情報と、前記文字列の発話のタイミングを表すタイミング情報とを対応付けた字幕タイミング情報を生成する

ことにより得られる前記字幕タイミング情報に従い、前記字幕データと、その字幕データの識別情報と対応付けられているタイミング情報が表すタイミングの画像データとを同期させて出力するステップを含む

データ処理方法。

【請求項 5】

画像データ、音声データ、及び、字幕データを含むコンテンツデータのうちの前記字幕データから、検索の対象とする文字列を抽出し、

前記コンテンツデータのうちの前記音声データから、前記文字列を発話した音声データを検索し、前記画像データにおける、前記文字列の発話のタイミングを表すタイミング情報を取得し、

前記文字列を含む前記字幕データを識別する識別情報と、前記文字列の発話のタイミングを表すタイミング情報とを対応付けた字幕タイミング情報を生成する

ことにより得られる前記字幕タイミング情報に従い、前記字幕データと、その字幕デー

10

20

30

40

50

タの識別情報と対応付けられているタイミング情報が表すタイミングの画像データとを同期させて出力する出力制御手段

として、コンピュータを機能させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、データ処理装置、データ処理方法、及び、プログラムに関し、特に、例えば、テレビジョン放送の番組等のコンテンツの画像と字幕とを、容易に、同期して表示等することができるようにするデータ処理装置、データ処理方法、及び、プログラムに関する。

10

【背景技術】

【0002】

例えば、テレビジョン放送等では、聴覚に障害がある視聴者が、コンテンツとしての番組の画像の内容を把握することができるように、クロズドキャプション(closed captioning)等といった形で、テレビジョン放送の放送信号に、字幕データが含まれる。

【0003】

テレビジョン放送の放送信号を受信するTV(テレビジョン受像機)やレコーダ等の受信装置では、字幕データが、番組の画像データに重畳され、これにより、番組の画像とともに、字幕が表示される。なお、一般に、字幕は、非表示(closed)にすることができる。

20

【0004】

ところで、現在のテレビジョン放送において、字幕は、必ずしも、画像と同期して表示されとは限らない。すなわち、字幕は、その字幕に対応する画像(字幕が内容を説明している画像)からずれて表示されることがある。

【0005】

特に、例えば、生放送の番組については、その場で、オペレータが、字幕となる文字をタイプするため、画像に対応する字幕が、その画像の表示から、数秒から数十秒程度遅れて表示されることがある。この場合、画像の内容と、その画像に重畳されている字幕が説明する内容とが整合(一致)しないため、視聴者が混乱すること、あるいは違和感を感じることがある。

30

【0006】

上述のような字幕が遅れて表示される字幕の遅延を補正する方法としては、例えば、画像データに対する、字幕データの遅延時間を、テレビジョン放送を行う放送局で計算し、その遅延時間を、受信装置に送信する方法がある(例えば、特許文献1を参照)。この場合、受信装置では、放送局から送信されてくる遅延時間に従い、字幕が、画像に同期して表示される。

【0007】

【特許文献1】特開2006-211636号公報

【発明の開示】

【発明が解決しようとする課題】

【0008】

上述のように、遅延時間を、放送局から受信装置に送信するには、例えば、遅延時間の送信に関して、放送規格を策定(変更)し、放送局、及び、受信装置の両方で、その放送規格に対応する必要がある。

40

【0009】

しかしながら、放送規格の策定には、煩雑な手続が必要である。また、放送局、及び、受信装置の両方で、放送規格に対応するのも煩雑であり、容易ではない。

【0010】

本発明は、このような状況に鑑みてなされたものであり、コンテンツの画像と字幕とを、容易に、同期して表示すること、すなわち、例えば、放送規格の変更や放送局の対応なしで、コンテンツの画像と字幕とを、同期して表示することができるようにするものであ

50

る。

【課題を解決するための手段】

【0011】

本発明の一側面のデータ処理装置、又は、プログラムは、画像データ、音声データ、及び、字幕データを含むコンテンツデータのうちの前記字幕データから、検索の対象とする文字列を抽出し、前記コンテンツデータのうちの前記音声データから、前記文字列を発話した音声データを検索し、前記画像データにおける、前記文字列の発話のタイミングを表すタイミング情報を取得し、前記文字列を含む前記字幕データを識別する識別情報と、前記文字列の発話のタイミングを表すタイミング情報とを対応付けた字幕タイミング情報を生成することにより得られる前記字幕タイミング情報に従い、前記字幕データと、その字幕データの識別情報と対応付けられているタイミング情報が表すタイミングの画像データとを同期させて出力する出力制御手段を備えるデータ処理装置、又は、データ処理装置として、コンピュータを機能させるためのプログラムである。

10

【0012】

本発明の一側面のデータ処理方法は、データ処理装置が、画像データ、音声データ、及び、字幕データを含むコンテンツデータのうちの前記字幕データから、検索の対象とする文字列を抽出し、前記コンテンツデータのうちの前記音声データから、前記文字列を発話した音声データを検索し、前記画像データにおける、前記文字列の発話のタイミングを表すタイミング情報を取得し、前記文字列を含む前記字幕データを識別する識別情報と、前記文字列の発話のタイミングを表すタイミング情報とを対応付けた字幕タイミング情報を生成することにより得られる前記字幕タイミング情報に従い、前記字幕データと、その字幕データの識別情報と対応付けられているタイミング情報が表すタイミングの画像データとを同期させて出力するステップを含むデータ処理方法である。

20

【0013】

以上のような一側面においては、画像データ、音声データ、及び、字幕データを含むコンテンツデータのうちの前記字幕データから、検索の対象とする文字列を抽出し、前記コンテンツデータのうちの前記音声データから、前記文字列を発話した音声データを検索し、前記画像データにおける、前記文字列の発話のタイミングを表すタイミング情報を取得し、前記文字列を含む前記字幕データを識別する識別情報と、前記文字列の発話のタイミングを表すタイミング情報とを対応付けた字幕タイミング情報を生成することにより得られる前記字幕タイミング情報に従い、前記字幕データと、その字幕データの識別情報と対応付けられているタイミング情報が表すタイミングの画像データとが同期して出力される。

30

【0014】

なお、データ処理装置は、独立した装置であっても良いし、1つの装置を構成している内部ブロックであっても良い。

【0015】

また、プログラムは、伝送媒体を介して伝送することにより、又は、記録媒体に記録して、提供することができる。

【発明の効果】

40

【0016】

本発明の一側面によれば、コンテンツの画像と字幕とを、容易に、同期して表示することができる。

【発明を実施するための最良の形態】

【0017】

[本発明を適用したレコーダの一実施の形態の構成例]

【0018】

図1は、本発明を適用したレコーダの一実施の形態の構成例を示すブロック図である。

【0019】

図1において、レコーダは、例えば、HD(Hard Disk)レコーダ等であり、チューナ11

50

、デマルチプレクサ 1 2、記録再生制御部 1 3、記録媒体 1 4、デコーダ 1 5、及び、コンテンツデータ処理部 1 6 等から構成される。

【 0 0 2 0 】

チューナ 1 1 は、例えば、ディジタル放送の放送局（図示せず）から送信されてくる、テレビジョン放送の番組の送信信号を受信して復調することにより、コンテンツデータを含む、例えば、TS(Transport Stream)を取得して、デマルチプレクサ 1 2 に供給する。

【 0 0 2 1 】

ここで、図 1 のレコーダでは、上述のように、ディジタル放送の送信信号からコンテンツデータ（を含むTS）を取得することができる他、LAN(Local Area Network)やインターネット等のネットワークから、コンテンツデータを取得することができる。

10

【 0 0 2 2 】

デマルチプレクサ 1 2 は、チューナ 1 1 からのTSから、EPG(Electronic Program Guide)データを含む、必要な制御用のデータ（PSI/SI(Program Specific Information/ Service Information)）を抽出し、記録再生制御部 1 3 に供給する。

【 0 0 2 3 】

また、デマルチプレクサ 1 2 は、ユーザ（視聴者）の操作等に応じて、チューナ 1 1 からのTSから、所定の番組のコンテンツデータを含むTSパケットを抽出し、記録再生制御部 1 3 に供給する。

【 0 0 2 4 】

記録再生制御部 1 3 は、記録媒体 1 4 に対するデータの記録と再生（読み出し）を制御する。

20

【 0 0 2 5 】

すなわち、記録再生制御部 1 3 は、デマルチプレクサ 1 2 からのEPGデータやコンテンツデータ（を含むTSパケット）等を、記録媒体 1 4 に記録させる。また、記録再生制御部 1 3 は、ユーザの操作や、コンテンツデータ処理部 1 6 からの要求等に応じて、記録媒体 1 4 からコンテンツデータ等を再生し（読み出し）、デコーダ 1 5 に供給する。

【 0 0 2 6 】

記録媒体 1 4 は、例えば、HD等の磁気ディスク、DVD(Digital Versatile Disc)やBlu-rayディスク等の光ディスク、フラッシュメモリ等の半導体メモリ等であり、記録再生制御部 1 3 の制御に従い、データを記録（記憶）する。

30

【 0 0 2 7 】

デコーダ 1 5 は、記録再生制御部 1 4 からのコンテンツデータを含むTSパケットを、例えば、MPEG(Moving Picture Experts Group)方式に従ってデコードし、その結果得られる画像データ、音声データ、及び、クローズドキャプション等の字幕データを含むコンテンツデータを、コンテンツデータ処理部 1 6 に供給する。

【 0 0 2 8 】

コンテンツデータ処理部 1 6 は、記録再生制御部 1 3 に対して、必要なコンテンツデータを要求する。また、コンテンツデータ処理部 1 6 は、デコーダ 1 5 から供給されるコンテンツデータに含まれる画像データ、音声データ、及び字幕データを処理し、字幕データを、対応する画像データに重畳（合成）することで得られる合成画像データと、付随する音声データとを出力する。

40

【 0 0 2 9 】

以上のように構成されるレコーダでは、チューナ 1 1 が、送信信号を受信して復調することにより、TSを取得して、デマルチプレクサ 1 2 に供給する。

【 0 0 3 0 】

デマルチプレクサ 1 2 は、チューナ 1 1 からのTSから、所定の番組のコンテンツデータを含むTSパケットを抽出し、記録再生制御部 1 3 に供給する。

【 0 0 3 1 】

記録再生制御部 1 3 は、デマルチプレクサ 1 2 からのコンテンツデータを含むTSパケットを、記録媒体 1 4 に記録し、これにより、いわゆる番組の録画が行われる。

50

【 0 0 3 2 】

以上のようにして録画が行われた番組の再生時には、記録再生制御部 1 3 が、記録媒体 1 4 からコンテンツデータを含むTSパケットを再生し（読み出し）、デコーダ 1 5 に供給する。

【 0 0 3 3 】

デコーダ 1 5 は、記録再生制御部 1 4 からのコンテンツデータを含むTSパケットをデコードし、その結果得られる、画像データ、音声データ、及び字幕データを含むコンテンツデータを、コンテンツデータ処理部 1 6 に供給する。

【 0 0 3 4 】

コンテンツデータ処理部 1 6 は、デコーダ 1 5 から供給されるコンテンツデータに含まれる画像データ、音声データ、及び字幕データを処理し、字幕データを、対応する画像データに重畳した合成画像データと、付随する音声データとを出力する。

【 0 0 3 5 】

コンテンツデータ処理部 1 6 が出力する合成画像データと音声データは、図示せぬモニタに供給される。モニタでは、合成画像データに対応する合成画像、すなわち、番組の画像に、その画像の内容を説明する字幕が重畳された画像が表示されるとともに、音声データに対応する音声出力される。

【 0 0 3 6 】

なお、字幕の表示（字幕の重畳）は、例えば、ユーザの操作等に応じて、オン又はオフにすることができる。

【 0 0 3 7 】

[コンテンツデータ処理部 1 6 の構成例]

【 0 0 3 8 】

図 2 は、図 1 のコンテンツデータ処理部 1 6 の構成例を示している。

【 0 0 3 9 】

図 2 において、コンテンツデータ処理部 1 6 は、デコーダ 1 5（図 1）から供給される、画像データ、音声データ、及び、字幕データを含むコンテンツデータを処理するデータ処理装置である。コンテンツデータ処理部 1 6 は、コンテンツデータ取得部 3 1、字幕取得部 3 2、単語抽出部 3 3、音声取得部 3 4、音声検索部 3 5、字幕タイミング情報生成部 3 6、及び、同期部 4 0 等から構成される。

【 0 0 4 0 】

コンテンツデータ取得部 3 1 は、記録再生制御部 1 3（図 1）に対して、コンテンツデータを要求し、その要求に応じて、記録再生制御部 1 3 が、記録媒体 1 4 から再生し、デコーダ 1 5 を介してコンテンツデータ取得部 3 1 に供給するコンテンツデータを取得する。

【 0 0 4 1 】

また、コンテンツデータ取得部 3 1 は、例えば、ユーザの操作等に応じて、記録再生制御部 1 3 が、記録媒体 1 4 から再生し、デコーダ 1 5 を介してコンテンツデータ取得部 3 1 に供給するコンテンツデータを取得する。

【 0 0 4 2 】

コンテンツデータ取得部 3 1 は、以上のようにして取得したコンテンツデータを、必要に応じて、字幕取得部 3 2、音声取得部 3 4、及び、同期部 4 0 に供給する。

【 0 0 4 3 】

字幕取得部 3 2 は、コンテンツデータ取得部 3 1 から供給されるコンテンツデータから字幕データを、例えば、ひとまとまりの字幕等の単位で取得（分離）し、単語抽出部 3 3、及び、字幕タイミング情報生成部 3 6 に供給する。

【 0 0 4 4 】

単語抽出部 3 3 は、字幕取得部 3 2 からの字幕データとしての、例えば、テキストデータから、検索の対象とする文字列としての、例えば、1 以上の単語（列）を、形態素解析等の自然言語処理技術を用いて抽出し、音声検索部 3 5 に供給する。

10

20

30

40

50

【 0 0 4 5 】

ここで、単語抽出部 3 3 が、字幕データから、検索の対象とする文字列として抽出する単語を、検索対象単語ともいう。

【 0 0 4 6 】

また、単語抽出部 3 3 は、字幕データから検索対象単語を抽出する他、その検索対象単語を含む字幕データに対応する字幕を表示する表示時刻を表す表示時刻情報を、必要に応じて取得し、検索対象単語とともに、音声検索部 3 5 に供給する。

【 0 0 4 7 】

すなわち、コンテンツデータ取得部 3 1 から字幕取得部 3 2 に供給されるコンテンツデータには、例えば、字幕データの表示時刻情報が含まれる。字幕データ取得部 3 2 は、コンテンツデータ取得部 3 1 からコンテンツデータから、字幕データの他、表示時刻情報も取得して、単語抽出部 3 3 に供給する。

10

【 0 0 4 8 】

単語抽出部 3 3 は、字幕取得部 3 2 から表示時刻情報が供給される場合、字幕データから抽出した検索対象単語とともに、その検索対象単語を含む字幕データの表示時刻情報を、音声検索部 3 5 に供給する。

【 0 0 4 9 】

音声取得部 3 4 は、コンテンツデータ取得部 3 1 から供給されるコンテンツデータから、音声データを取得し、音声検索部 3 5 に供給する。

【 0 0 5 0 】

20

音声検索部 3 5 は、音声取得部 3 4 から供給される音声データから、単語抽出部 3 3 から供給される検索対象単語を発話した音声データを検索し、その音声データが付随する画像データにおける、検索対象単語の発話のタイミングを表すタイミング情報を取得する。

【 0 0 5 1 】

ここで、タイミング情報としては、例えば、コンテンツデータ取得部 3 1 が取得したコンテンツデータに対応するコンテンツとしての番組の先頭のタイミングを基準とする相対時刻を表すタイムコード等を採用することができる。

【 0 0 5 2 】

また、音声検索部 3 5 では、単語抽出部 3 3 から表示時刻情報が供給される場合には、検索対象単語を発話した音声データの検索を、表示時刻情報が表す表示時刻のタイミング付近の音声データに限定して行うことができる。

30

【 0 0 5 3 】

検索対象単語を発話した音声データの検索を、コンテンツデータ取得部 3 1 が取得したコンテンツデータに含まれる音声データの全区間ではなく、表示時刻情報が表す表示時刻のタイミング付近の音声データに限定して行うことにより、音声データの検索の精度の向上、検索に要する処理量の低減、及び、検索の処理の高速化等を図ることができる。

【 0 0 5 4 】

字幕タイミング情報生成部 3 6 は、字幕取得部 3 2 から供給される字幕データを識別する字幕識別情報と、その字幕データから抽出された検索対象単語を発話した音声データの検索によって音声検索部 3 5 で得られた、検索対象単語の発話のタイミングを表すタイミング情報とを対応付けた字幕タイミング情報を生成する。

40

【 0 0 5 5 】

さらに、字幕タイミング情報生成部 3 6 は、コンテンツデータ取得部 3 1 が取得したコンテンツデータを識別するコンテンツ識別情報に、そのコンテンツデータから生成された字幕タイミング情報を対応付けて記憶する。

【 0 0 5 6 】

字幕タイミング情報生成部 3 6 に記憶された字幕タイミング情報は、同期部 4 0 の出力制御部 4 4 によって、必要に応じて参照される。

【 0 0 5 7 】

なお、字幕データを識別する字幕識別情報としては、例えば、その字幕データそのもの

50

を採用することができる。また、コンテンツデータを識別するコンテンツ識別情報としては、例えば、そのコンテンツデータに対応するコンテンツとしての番組のタイトル、又は、放送日時及びチャンネル等を採用することができる。

【0058】

その他、コンテンツデータに対して、ユニークなID(Identification)を付与し、そのIDを、コンテンツ識別情報として採用することができる。字幕識別情報についても、同様である。但し、字幕識別情報は、例えば、字幕データを引数として求められるハッシュ値等のように、字幕データそれ自体から求められる情報である必要がある。

【0059】

同期部40は、コンテンツデータ取得部31から供給されるコンテンツデータに対して、そのコンテンツデータに含まれる画像データと、その画像データに対応する字幕データとを同期して表示するための処理を施すことにより、字幕データを、対応する画像データに重畳した合成画像データを得て、付随する音声データとともに出力する。

10

【0060】

すなわち、同期部40は、音声取得部41、画像取得部42、字幕取得部43、出力制御部44、及び、合成部45等から構成される。

【0061】

音声取得部41、画像取得部42、及び、字幕取得部43には、コンテンツデータ取得部31からコンテンツデータが供給される。

【0062】

20

音声取得部41は、コンテンツデータ取得部31から供給されるコンテンツデータから、音声データを取得し、出力制御部44に供給する。

【0063】

画像取得部42は、コンテンツデータ取得部31から供給されるコンテンツデータから、画像データを取得し、出力制御部44に供給する。

【0064】

字幕取得部43は、コンテンツデータ取得部31から供給されるコンテンツデータから、字幕データを取得し、出力制御部44に供給する。

【0065】

出力制御部44は、字幕タイミング情報生成部36に記憶された字幕タイミング情報に従い、字幕データ取得部43から供給される字幕データと、その字幕データの字幕識別情報と対応付けられているタイミング情報が表すタイミングの画像データとを同期させ、その画像データに付随する音声データとともに出力する。

30

【0066】

すなわち、出力制御部44は、画像取得部42から供給される画像データと、音声取得部41から供給される、その画像データに付随する音声データとを、必要に応じて記憶することにより、一定時間だけ遅延して出力する。

【0067】

また、出力制御部44は、字幕タイミング情報生成部36に記憶された字幕タイミング情報のうちの、コンテンツデータ取得部31で取得されたコンテンツデータのコンテンツ識別情報と対応付けられている字幕タイミング情報すべてを、注目する注目字幕タイミング情報として参照する。

40

【0068】

さらに、出力制御部44は、注目字幕タイミング情報において、字幕取得部43から供給される字幕データの字幕識別情報と対応付けられているタイミング情報を検出し、そのタイミング情報が表すタイミングの画像データ、つまり、字幕取得部43から供給される字幕データに対応する字幕が内容を説明する画像データに同期して、字幕取得部43から供給される字幕データを出力する。

【0069】

なお、出力制御部44が出力する音声データは、図示せぬモニタに供給される。

50

【 0 0 7 0 】

また、出力制御部 4 4 が出力する画像データと、その画像データの内容を説明する字幕の字幕データとは、合成部 4 5 に供給される。

【 0 0 7 1 】

合成部 4 5 は、出力制御部 4 4 からの画像データに、同じく出力制御部 4 4 からの字幕データを重畳（合成）することにより、合成画像データを生成して出力する。

【 0 0 7 2 】

合成部 4 5 が出力する合成画像データは、出力制御部 4 4 が出力する、その合成画像データに付随する音声データとともに、図示せぬモニタに供給される。

【 0 0 7 3 】

以上のように構成されるコンテンツデータ処理部 1 6 では、字幕タイミング情報取得処理と、同期再生処理とが行われる。

【 0 0 7 4 】

字幕タイミング情報取得処理では、例えば、記録媒体 1 4 に記録されたコンテンツデータ（を含むTSパケット）から、字幕タイミング情報が生成されること等により取得される。また、同期再生処理では、字幕タイミング情報取得処理で取得された字幕タイミング情報に従い、コンテンツデータに含まれる画像データと字幕データとを同期させ、その画像データに付随する音声データとともに出力する処理が行われる。

【 0 0 7 5 】

[字幕タイミング情報取得処理の説明]

【 0 0 7 6 】

図 3 を参照して、図 2 のコンテンツデータ処理部 1 6 が行う字幕タイミング情報取得処理について説明する。

【 0 0 7 7 】

字幕タイミング情報取得処理は、例えば、周期的、又は非周期的なタイミング等の任意のタイミングで開始することができる。

【 0 0 7 8 】

字幕タイミング情報取得処理では、ステップ S 1 1 において、コンテンツデータ取得部 3 1 が、記録媒体 1 4（図 1）に記録されたコンテンツデータのうちの、まだ、字幕タイミング情報を取得していないコンテンツデータの 1 つを取得する。

【 0 0 7 9 】

すなわち、ステップ S 1 1 では、コンテンツデータ取得部 3 1 は、記録再生制御部 1 3（図 1）に対して、コンテンツデータを要求する。そして、コンテンツデータ取得部 3 1 は、その要求に応じて、記録再生制御部 1 3 が、記録媒体 1 4 から再生し、デコーダ 1 5 を介してコンテンツデータ取得部 3 1 に供給するコンテンツデータを取得する。

【 0 0 8 0 】

さらに、コンテンツデータ取得部 3 1 は、記録再生制御部 1 3 からデコーダ 1 5 を介して取得したコンテンツデータを、字幕取得部 3 2、及び、音声取得部 3 4 に供給し、処理は、ステップ S 1 1 からステップ S 1 2 に進む。

【 0 0 8 1 】

ステップ S 1 2 では、字幕取得部 3 2 は、コンテンツデータ取得部 3 1 から供給されるコンテンツデータから、すべての字幕データを取得し、単語抽出部 3 3、及び、字幕タイミング情報生成部 3 6 に供給する。さらに、ステップ S 1 2 では、字幕取得部 3 2 は、コンテンツデータ取得部 3 1 から供給されるコンテンツデータから取得した各字幕データの表示時刻情報を、そのコンテンツデータから取得し、単語抽出部 3 3 に供給する。

【 0 0 8 2 】

また、ステップ S 1 2 では、音声取得部 3 4 が、コンテンツデータ取得部 3 1 から供給されるコンテンツデータから、音声データ（音声の波形データ）を取得し、音声検索部 3 5 に供給する。

【 0 0 8 3 】

その後、処理は、ステップ S 1 2 からステップ S 1 3 に進み、単語抽出部 3 3 が、字幕取得部 3 2 からの各字幕データから、検索の対象とする文字列としての、検索対象単語を抽出する。さらに、単語抽出部 3 3 は、各字幕データから抽出した検索対象単語を、字幕取得部 3 2 からの表示時刻情報（検索対象単語が抽出された字幕の表示時刻情報）と対応付けて登録した検索対象リストを作成する。そして、単語抽出部 3 3 は、検索対象リストを、音声検索部 3 5 に供給して、処理は、ステップ S 1 3 からステップ S 1 4 に進む。

【0084】

ここで、単語抽出部 3 3 は、例えば、字幕データとしてのテキストデータについて、形態素解析を行うことで、そのテキストデータを単語単位に分割するとともに、各単語の読み（音韻）を取得する。さらに、単語抽出部 3 3 は、後述するステップ S 1 6 において、注目単語となる検索対象単語を発話した音声データの検索を精度良く行うこと等ができるようにするため、例えば、テキストデータを分割して得られる単語のうちの、読みの最も長い単語や、固有名詞、所定の文字数以上の複数の連続する単語（単語列）等を、検索対象単語として選択（抽出）する。

【0085】

ステップ S 1 4 では、音声検索部 3 5 が、単語抽出部 3 3 からの検索対象リストに、検索対象単語が登録されているかどうかを判定する。

【0086】

ステップ S 1 4 において、検索対象リストに、検索対象単語が登録されていると判定された場合、処理は、ステップ S 1 5 に進み、音声検索部 3 5 は、検索対象リストに登録されている検索対象単語のうちの 1 つを、注目単語として選択する。

【0087】

その後、処理は、ステップ S 1 5 からステップ S 1 6 に進み、音声検索部 3 5 は、音声取得部 3 4 からの音声データから、注目単語を発話した音声データを検索する。そして、音声検索部 3 5 は、その検索結果に基づき、直前のステップ S 1 1 で取得されたコンテンツデータに含まれる画像データにおける、注目単語の発話のタイミングを認識する。さらに、音声検索部 3 5 は、注目単語の発話のタイミングを表すタイミング情報を生成することにより取得し、字幕タイミング情報生成部 3 6 に供給し、処理は、ステップ S 1 6 からステップ S 1 7 に進む。

【0088】

ここで、音声検索部 3 5 は、注目単語を発話した音声データの検索を、例えば、いわゆるキーワードスポッティングや、音声取得部 3 4 からの音声データの音素、及び、音素の位置をインデクスとして作成しておき、注目単語を構成する音素の系列を、そのインデクスから探し出す方法等を利用して行う。

【0089】

また、音声検索部 3 5 では、注目単語を発話した音声データの検索を、単語抽出部 3 3 からの検索対象リストにおいて、注目単語に対応付けられている表示時刻情報が表す表示時刻のタイミング付近（例えば、表示時刻情報が表す表示時刻のタイミングを含み、注目単語の読みの長さに比例する時間）の音声データに限定して行うことができる。

【0090】

ステップ S 1 7 では、字幕タイミング情報生成部 3 6 は、字幕取得部 3 2 から供給される各字幕データを、その字幕データの字幕識別情報として、音声検索部 3 5 からの注目単語の発話のタイミングを表すタイミング情報と、その注目単語が抽出された字幕データの字幕識別情報とを対応付けた字幕タイミング情報を生成する。

【0091】

さらに、字幕タイミング情報生成部 3 6 は、直前のステップ S 1 1 で取得されたコンテンツデータのコンテンツ識別情報に、注目単語について生成された字幕タイミング情報を対応付けて記憶し、処理は、ステップ S 1 7 からステップ S 1 8 に進む。

【0092】

ステップ S 1 8 では、音声検索部 3 5 が、単語抽出部 3 3 からの検索対象リストから、

10

20

30

40

50

注目単語と、その注目単語に対応付けられている表示時刻情報とを削除する。

【 0 0 9 3 】

その後、処理は、ステップ S 1 8 からステップ S 1 4 に戻り、以下、同様の処理が繰り返される。

【 0 0 9 4 】

そして、ステップ S 1 4 において、検索対象リストに、検索対象単語が登録されていないと判定された場合、すなわち、直前のステップ S 1 1 で取得されたコンテンツデータに含まれるすべての字幕データについて、字幕タイミング情報が生成された場合、字幕タイミング情報取得処理は、終了する。

【 0 0 9 5 】

[同期再生処理の説明]

【 0 0 9 6 】

図 4 を参照して、図 2 のコンテンツデータ処理部 1 6 が行う同期再生処理について説明する。

【 0 0 9 7 】

同期再生処理では、ステップ S 3 1 において、再生の対象とするコンテンツデータの指定が、例えば、ユーザがレコーダを操作すること等によって行われたかどうかを、コンテンツデータ取得部 3 1 が判定する。

【 0 0 9 8 】

ステップ S 3 1 において、再生の対象とするコンテンツデータの指定が行われていないと判定された場合、処理は、ステップ S 3 1 に戻る。

【 0 0 9 9 】

また、ステップ S 3 1 において、再生の対象とするコンテンツデータの指定が行われたと判定された場合、処理は、ステップ S 3 2 に進み、コンテンツデータ取得部 3 1 は、コンテンツデータの取得を開始して、処理は、ステップ S 3 3 に進む。

【 0 1 0 0 】

すなわち、再生の対象とするコンテンツデータの指定が、ユーザがレコーダを操作すること等によって行われた場合、記録再生制御部 1 3 (図 1) は、その指定がされたコンテンツデータ (以下、指定コンテンツデータともいう) を、記録媒体 1 4 から再生し、デコーダ 1 5 を介して、コンテンツデータ処理部 1 6 に供給する。

【 0 1 0 1 】

コンテンツデータ取得部 3 1 は、以上のようにしてコンテンツデータ処理部 1 6 に供給される指定コンテンツデータの取得を開始する。

【 0 1 0 2 】

コンテンツデータ取得部 3 1 が取得したコンテンツデータは、同期部 4 0 の音声取得部 4 1、画像取得部 4 2、及び、字幕取得部 4 3 に供給される。

【 0 1 0 3 】

ステップ S 3 3 では、コンテンツデータ取得部 3 1 が、コンテンツデータの再生を終了するかどうかを判定する。

【 0 1 0 4 】

ステップ S 3 3 において、コンテンツデータの再生を終了しないと判定された場合、すなわち、例えば、ユーザが、コンテンツの再生を終了することを指示する操作等を行っていない場合、処理は、ステップ S 3 4 に進み、指定コンテンツから、音声データ、画像データ、及び、字幕データが取得され、処理は、ステップ S 3 5 に進む。

【 0 1 0 5 】

すなわち、ステップ S 3 4 では、音声取得部 4 1 が、コンテンツデータ取得部 3 1 からの指定コンテンツデータから、音声データを取得し、出力制御部 4 4 に供給する。

【 0 1 0 6 】

さらに、画像取得部 4 2 が、コンテンツデータ取得部 3 1 からの指定コンテンツデータから、画像データを取得し、出力制御部 4 4 に供給する。

10

20

30

40

50

【 0 1 0 7 】

また、字幕取得部 4 3 が、コンテンツデータ取得部 3 1 からの指定コンテンツデータから、字幕データを取得し、出力制御部 4 4 に供給する。

【 0 1 0 8 】

ステップ S 3 5 では、出力制御部 4 4 は、字幕タイミング情報生成部 3 6 に記憶された字幕タイミング情報のうちの、コンテンツデータ取得部 3 1 で取得された指定コンテンツデータのコンテンツ識別情報と対応付けられている字幕タイミング情報すべてを、注目する注目字幕タイミング情報として参照する。

【 0 1 0 9 】

さらに、出力制御部 4 4 は、字幕取得部 4 3 からの字幕データを、字幕識別情報として、注目字幕タイミング情報において、その字幕識別情報と対応付けられているタイミング情報を検出することにより取得する。

10

【 0 1 1 0 】

そして、処理は、ステップ S 3 5 からステップ S 3 6 に進み、出力制御部 4 4 は、注目字幕タイミング情報から取得した字幕タイミング情報に従い、音声取得部 4 1 からの音声データ、画像取得部 4 2 からの画像データ、及び、字幕取得部 4 3 からの字幕データの出力を制御する。

【 0 1 1 1 】

すなわち、出力制御部 4 4 は、画像取得部 4 2 から供給される画像データと、音声取得部 4 1 から供給される、その画像データに付随する音声データとを、必要に応じて遅延して、又は遅延せずに出力する。

20

【 0 1 1 2 】

さらに、出力制御部 4 4 は、注目字幕タイミング情報から取得した字幕タイミング情報が表すタイミングの画像データ、つまり、字幕取得部 4 3 から供給される字幕データに対応する字幕が内容を説明する画像データに同期して、字幕取得部 4 3 からの字幕データを出力する。すなわち、出力制御部 4 4 は、注目字幕タイミング情報から取得した字幕タイミング情報が表すタイミングの画像データの出力と同時に、字幕取得部 4 3 から供給される字幕データの出力を開始する。

【 0 1 1 3 】

出力制御部 4 4 が出力する画像データと（その画像データの内容を説明する字幕の）字幕データとは、合成部 4 5 に供給される。

30

【 0 1 1 4 】

その後、処理は、ステップ S 3 6 からステップ S 3 7 に進み、合成部 4 5 が、出力制御部 4 4 からの画像データに、同じく出力制御部 4 4 からの字幕データを重畳（合成）して、合成画像データを生成して出力する。

【 0 1 1 5 】

合成部 4 5 が出力する合成画像データは、出力制御部 4 4 が出力する、その合成画像データに付随する音声データとともに、図示せぬモニタに供給される。

【 0 1 1 6 】

これにより、モニタでは、合成画像データに対応する合成画像、すなわち、番組の画像に、その画像の内容を説明する字幕が重畳された画像が表示されるとともに、音声データに対応する音声出力される。

40

【 0 1 1 7 】

その後、処理は、ステップ S 3 7 からステップ S 3 3 に戻り、以下、同様の処理が繰り返される。

【 0 1 1 8 】

そして、ステップ S 3 3 において、コンテンツデータの再生を終了すると判定された場合、すなわち、例えば、ユーザが、コンテンツの再生を終了することを指示する操作等を行った場合、同期再生処理は、終了する。

【 0 1 1 9 】

50

以上のように、コンテンツデータ処理部 16 では、字幕タイミング情報取得処理（図 3）において、単語抽出部 33 が、画像データ、音声データ、及び、字幕データを含むコンテンツデータのうちの字幕データから、検索の対象とする文字列としての検索対象単語を抽出し、音声検索部 35 が、コンテンツデータのうちの音声データから、検索対象単語を発話した音声データを検索することにより、画像データにおける、検索対象単語の発話のタイミングを表すタイミング情報を取得し、字幕タイミング情報生成部 36 が、検索対象単語を含む字幕データを識別する字幕識別情報と、検索対象単語の発話のタイミングを表すタイミング情報とを対応付けた字幕タイミング情報を生成する。

【0120】

さらに、コンテンツデータ処理部 16 では、同期再生処理（図 4）において、出力制御部 44 が、字幕タイミング情報に従い、字幕データと、その字幕データの字幕識別情報と対応付けられているタイミング情報が表すタイミングの画像データとを同期させて出力する。

10

【0121】

したがって、コンテンツの画像と字幕とを、容易に、同期して表示することができる。すなわち、例えば、放送規格の変更や放送局の対応なしで、コンテンツの画像に、その画像の内容を説明する字幕を重畳して表示することが可能となる。

【0122】

その結果、字幕が、その字幕が内容を説明している画像からずれて表示されることが解消され、視聴者が混乱することや、違和感を感じることを防止することができる。

20

【0123】

なお、単語抽出部 33 において、字幕データから抽出する、検索の対象とする文字列としては、1つの単語の他、例えば、分節や文等の複数の単語列を採用することが可能である。

【0124】

また、本実施の形態では、コンテンツデータ処理部 16 において、記録媒体 14 に記録されたコンテンツデータから、字幕タイミング情報を生成することにより取得することとしたが、定義タイミング情報は、その他、例えば、インターネット上のサーバ等から取得することができる。すなわち、字幕タイミング情報は、インターネット上のサーバで生成して提供することができ、この場合、コンテンツデータ処理部 16 は、字幕タイミング情報を、インターネット上のサーバからダウンロードすることにより取得することができる。

30

【0125】

字幕タイミング情報を、インターネット上のサーバからダウンロードすることにより取得する場合、コンテンツデータ処理部 16 は、字幕タイミング情報を作成するためのブロックである字幕取得部 32、単語抽出部 33、音声取得部 34、音声検索部 35、及び、字幕タイミング情報生成部 36 を設けずに構成することができる。

【0126】

さらに、本実施の形態では、字幕タイミング情報取得処理（図 3）、及び、同期再生処理（図 4）を、記録媒体 14 に記録（録画）されたコンテンツデータを対象に行うこととしたが、字幕タイミング情報取得処理、及び、同期再生処理は、記録媒体 14 に記録されたコンテンツデータの他、リアルタイムで放送されている番組のコンテンツデータを対象に行うことが可能である。

40

【0127】

リアルタイムで放送されている番組のコンテンツデータを対象に、字幕タイミング情報取得処理、及び、同期再生処理を行う場合には、コンテンツデータ処理部 16 では、コンテンツデータが、字幕取得部 32、音声取得部 34、並びに、同期部 40 の音声取得部 41、画像取得部 42、及び字幕取得部 43 に供給され、字幕タイミング情報取得処理（図 3）、及び、同期再生処理（図 4）が、並列で行われる。

【0128】

50

そして、この場合、出力制御部 4 4 (図 2) は、同期再生処理として、少なくとも、1 つの字幕データ (ひとまとまりの字幕の字幕データ) について、字幕タイミング情報を取得するのに必要な時間等の所定の時間だけ、音声取得部 4 1 からの音声データと、画像取得部 4 2 からの画像データとを遅延して出力する。

【0129】

一方、字幕取得部 3 2、単語抽出部 3 3、音声取得部 3 4、音声検索部 3 5、及び、字幕タイミング情報生成部 3 6 は、字幕タイミング情報取得処理として、字幕取得部 3 2 で取得された最新の字幕データについて、字幕タイミング情報を生成し、出力制御部 4 4 に供給する。

【0130】

出力制御部 4 4 は、最新の字幕タイミング情報に従い、字幕取得部 4 3 から供給される最新の字幕データを、対応する画像データに同期させて出力する。

【0131】

以上、本発明をレコーダに適用した場合について説明したが、本発明は、レコーダの他、テレビジョン放送の番組を受信する TV その他の、コンテンツを再生する装置に適用可能である。

【0132】

次に、コンテンツデータ処理部 1 6 による上述した一連の処理は、ハードウェアにより行うこともできるし、ソフトウェアにより行うこともできる。一連の処理をソフトウェアによって行う場合には、そのソフトウェアを構成するプログラムが、汎用のコンピュータ等にインストールされる。

【0133】

そこで、図 5 は、上述した一連の処理を実行するプログラムがインストールされるコンピュータの一実施の形態の構成例を示している。

【0134】

プログラムは、コンピュータに内蔵されている記録媒体としてのハードディスク 1 0 5 や ROM 1 0 3 に予め記録しておくことができる。

【0135】

あるいはまた、プログラムは、フレキシブルディスク、CD-ROM (Compact Disc Read Only Memory)、MO (Magneto Optical) ディスク、DVD (Digital Versatile Disc)、磁気ディスク、半導体メモリなどのリムーバブル記録媒体 1 1 1 に、一時的あるいは永続的に格納 (記録) しておくことができる。このようなリムーバブル記録媒体 1 1 1 は、いわゆるパッケージソフトウェアとして提供することができる。

【0136】

なお、プログラムは、上述したようなリムーバブル記録媒体 1 1 1 からコンピュータにインストールする他、ダウンロードサイトから、デジタル衛星放送用の人工衛星を介して、コンピュータに無線で転送したり、LAN (Local Area Network)、インターネットといったネットワークを介して、コンピュータに有線で転送し、コンピュータでは、そのようにして転送されてくるプログラムを、通信部 1 0 8 で受信し、内蔵するハードディスク 1 0 5 にインストールすることができる。

【0137】

コンピュータは、CPU (Central Processing Unit) 1 0 2 を内蔵している。CPU 1 0 2 には、バス 1 0 1 を介して、入出力インタフェース 1 1 0 が接続されており、CPU 1 0 2 は、入出力インタフェース 1 1 0 を介して、ユーザによって、キーボードや、マウス、マイク等で構成される入力部 1 0 7 が操作等されることにより指令が入力されると、それに従って、ROM (Read Only Memory) 1 0 3 に格納されているプログラムを実行する。あるいは、また、CPU 1 0 2 は、ハードディスク 1 0 5 に格納されているプログラム、衛星若しくはネットワークから転送され、通信部 1 0 8 で受信されてハードディスク 1 0 5 にインストールされたプログラム、またはドライブ 1 0 9 に装着されたリムーバブル記録媒体 1 1 1 から読み出されてハードディスク 1 0 5 にインストールされたプログラムを、RAM (Rand

10

20

30

40

50

om Access Memory) 1 0 4 にロードして実行する。これにより、CPU 1 0 2 は、上述したフローチャートにしたがった処理、あるいは上述したブロック図の構成により行われる処理を行う。そして、CPU 1 0 2 は、その処理結果を、必要に応じて、例えば、入出力インタフェース 1 1 0 を介して、LCD(Liquid Crystal Display)やスピーカ等で構成される出力部 1 0 6 から出力、あるいは、通信部 1 0 8 から送信、さらには、ハードディスク 1 0 5 に記録等させる。

【0138】

ここで、本明細書において、コンピュータに各種の処理を行わせるためのプログラムを記述する処理ステップは、必ずしもフローチャートとして記載された順序に沿って時系列に処理する必要はなく、並列的あるいは個別に実行される処理（例えば、並列処理あるいはオブジェクトによる処理）も含むものである。

10

【0139】

また、プログラムは、1のコンピュータにより処理されるものであっても良いし、複数のコンピュータによって分散処理されるものであっても良い。さらに、プログラムは、遠方のコンピュータに転送されて実行されるものであっても良い。

【0140】

なお、本発明の実施の形態は、上述した実施の形態に限定されるものではなく、本発明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。

【図面の簡単な説明】

【0141】

20

【図1】本発明を適用したレコーダの一実施の形態構成例を示すブロック図である。

【図2】コンテンツデータ処理部16の構成例を示すブロック図である。

【図3】字幕タイミング情報取得処理を説明するフローチャートである。

【図4】同期再生処理を説明するフローチャートである。

【図5】本発明を適用したコンピュータの一実施の形態の構成例を示すブロック図である。

【符号の説明】

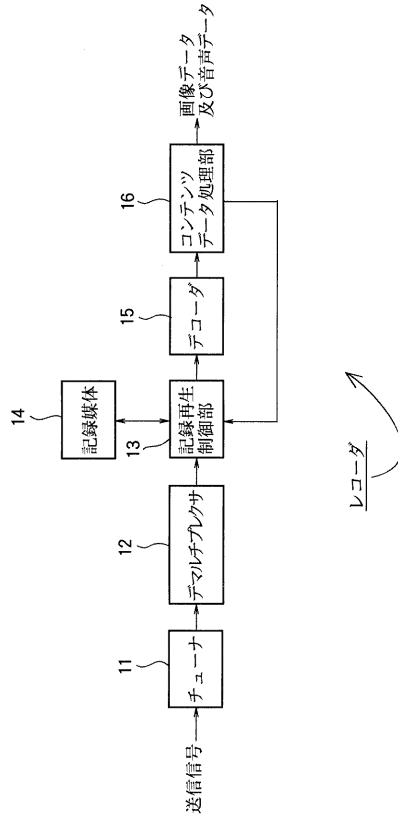
【0142】

1 1 チューナ, 1 2 デマルチプレクサ, 1 3 記録再生制御部, 1 4 記録媒体, 1 5 デコーダ, 1 6 コンテンツデータ処理部, 3 1 コンテンツデータ取得部, 3 2 字幕取得部, 3 3 単語抽出部, 3 4 音声取得部, 3 5 音声検索部, 3 6 字幕タイミング情報生成部, 4 0 同期部, 4 1 音声取得部, 4 2 画像取得部, 4 3 字幕取得部, 4 4 出力制御部, 4 5 合成部, 1 0 1 バス, 1 0 2 CPU, 1 0 3 ROM, 1 0 4 RAM, 1 0 5 ハードディスク, 1 0 6 出力部, 1 0 7 入力部, 1 0 8 通信部, 1 0 9 ドライブ, 1 1 0 入出力インタフェース, 1 1 1 リムーバブル記録媒体

30

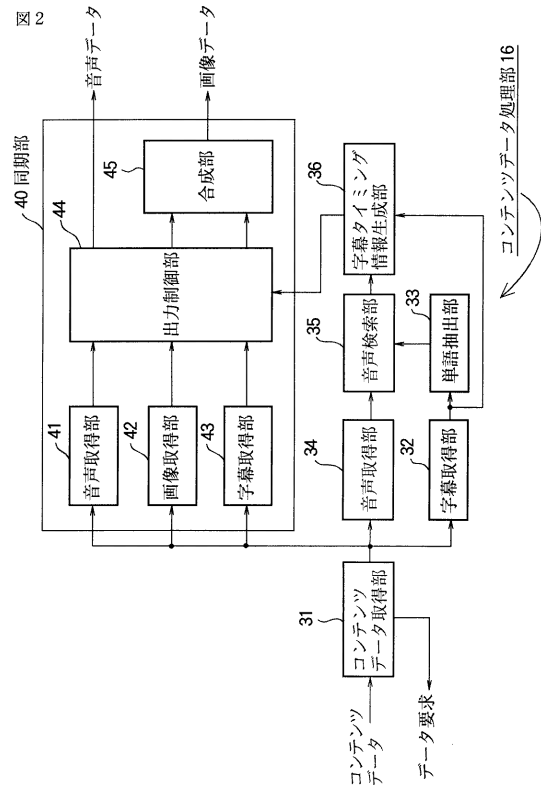
【図 1】

図 1



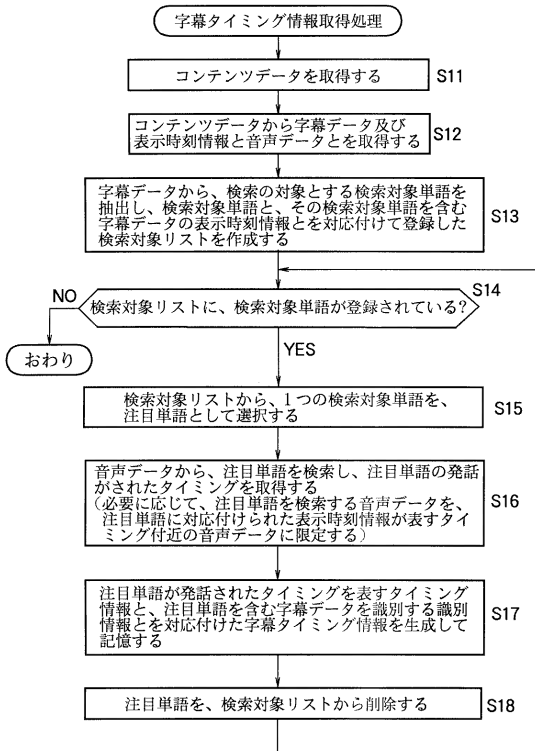
【図 2】

図 2



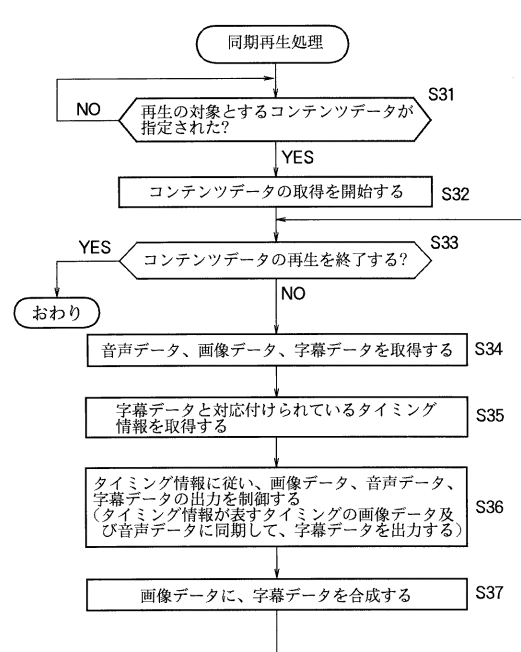
【図 3】

図 3



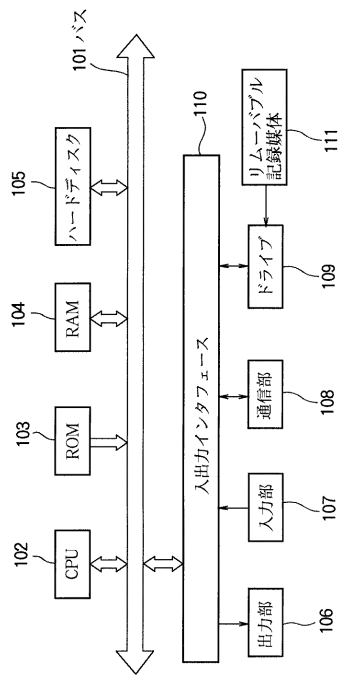
【図 4】

図 4



【図 5】

図 5



フロントページの続き

(51)Int.Cl. F I テーマコード(参考)
H 0 4 N 5/91 (2006.01) H 0 4 N 5/91 Z

(72)発明者 山本 則行
東京都港区港南 1 丁目 7 番 1 号 ソニー株式会社内

F ターム(参考) 5C023 AA40 BA11 BA16 CA01 CA04 CA05
5C025 BA27 BA30 CA09 CB10
5C053 GB06 GB12 HA01 JA22
5C164 MA06S MA07S PA44 UB05P UB08P UB10S