

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5910379号  
(P5910379)

(45) 発行日 平成28年4月27日(2016.4.27)

(24) 登録日 平成28年4月8日(2016.4.8)

(51) Int.Cl.		F I			
HO4N	5/91	(2006.01)	HO4N	5/91	N
HO4N	5/765	(2006.01)	HO4N	5/91	L
G11B	20/10	(2006.01)	HO4N	5/91	Z
G11B	27/02	(2006.01)	G11B	20/10	G
			G11B	27/02	A

請求項の数 12 (全 27 頁)

(21) 出願番号	特願2012-156201 (P2012-156201)	(73) 特許権者	000002185
(22) 出願日	平成24年7月12日(2012.7.12)		ソニー株式会社
(65) 公開番号	特開2014-22758 (P2014-22758A)		東京都港区港南1丁目7番1号
(43) 公開日	平成26年2月3日(2014.2.3)	(74) 代理人	100093241
審査請求日	平成27年1月6日(2015.1.6)		弁理士 官田 正昭
		(74) 代理人	100101801
			弁理士 山田 英治
		(74) 代理人	100095496
			弁理士 佐々木 榮二
		(74) 代理人	100086531
			弁理士 澤田 俊夫
		(74) 代理人	110000763
			特許業務法人大同特許事務所

最終頁に続く

(54) 【発明の名称】 情報処理装置、情報処理方法、表示制御装置および表示制御方法

(57) 【特許請求の範囲】

【請求項1】

コンテンツを構成する音声データの言語解析に基づいて、上記コンテンツを構成する映像データのチャプタ毎に、コンテンツ関連付け情報を取得する情報取得部と、

上記取得されたコンテンツ関連付け情報を出力する情報出力部とを備え、

上記コンテンツ関連付け情報は、

チャプタ内の文章の文字列の情報と、該文字列における対応文字位置の情報が付加された各音声特徴区間の情報を含む

情報処理装置。

【請求項2】

上記コンテンツ関連付け情報は、

上記コンテンツを構成する映像データの解析で検出された映像の動き情報をさらに含む

請求項1に記載の情報処理装置。

【請求項3】

上記情報出力部は、

上記取得されたコンテンツ関連付け情報を上記コンテンツと関連付けて記録メディアに記録する

請求項1に記載の情報処理装置。

【請求項4】

上記情報出力部は、

上記取得されたコンテンツ関連付け情報を上記コンテンツの送信単位に合わせてネットワークに送信する

請求項 1 に記載の情報処理装置。

【請求項 5】

コンテンツを構成する音声データの言語解析に基づいて、上記コンテンツを構成する映像データのチャプタ毎に、コンテンツ関連付け情報を取得する情報取得ステップと、

上記取得されたコンテンツ関連付け情報を出力する情報出力ステップとを有し、

上記コンテンツ関連付け情報は、

チャプタ内の文章の文字列の情報と、該文字列における対応文字位置の情報が付加された各音声特徴区間の情報を含む

情報処理方法。

【請求項 6】

コンテンツを構成する音声データの言語解析に基づいて上記コンテンツを構成する映像データのチャプタ毎に得られたコンテンツ関連付け情報を取得する情報取得部と、

上記取得されたコンテンツ関連付け情報に基づいて、上記コンテンツに関する表示を制御する表示制御部とを備え、

上記コンテンツ関連付け情報は、

チャプタ内の文章の文字列の情報と、該文字列における対応文字位置の情報が付加された各音声特徴区間の情報を含む

表示制御装置。

【請求項 7】

上記表示制御部は、

上記コンテンツ関連付け情報に基づいて、上記コンテンツの有用度の高い第 1 の区間と有用度の低い第 2 の区間を特定し、

上記コンテンツのタイムラインが上記第 1 の区間と上記第 2 の区間とが識別可能な状態で表示されるように制御する

請求項 6 に記載の表示制御装置。

【請求項 8】

上記表示制御部は、

上記タイムラインの所定区間の選択に応じて、上記コンテンツ関連付け情報に含まれる文字列情報を利用して、該所定区間の音声に対応する文字列が表示されるように制御する

請求項 7 に記載の表示制御装置。

【請求項 9】

上記表示制御部は、

上記所定区間の音声に対応する文字列と共に、前後の区間の音声に対応した文字列も表示され、

上記所定区間の音声に対応する文字列がハイライト表示されるように制御する

請求項 8 に記載の表示制御装置。

【請求項 10】

上記ハイライト表示は、他の区間とは表示色を異ならせることによるハイライト表示である

請求項 9 に記載の表示制御装置。

【請求項 11】

上記表示制御部は、

上記コンテンツ関連付け情報に含まれる文字列情報で示される文字列の文節毎の区間に分割された上記コンテンツのタイムラインが表示されると共に、

該タイムラインの各区間に対応して、該各区間の音声に対応した文字列が表示されるように制御する

請求項 6 に記載の表示制御装置。

【請求項 12】

10

20

30

40

50

コンテンツを構成する音声データの言語解析に基づいて上記コンテンツを構成する映像データのチャプタ毎に得られたコンテンツ関連付け情報を取得する情報取得ステップと、上記取得されたコンテンツ関連付け情報に基づいて、上記コンテンツに関する表示を制御する表示制御ステップとを有し、上記コンテンツ関連付け情報は、チャプタ内の文章の文字列の情報と、該文字列における対応文字位置の情報が付加された各音声特徴区間の情報を含む表示制御方法。

【発明の詳細な説明】

10

【技術分野】

【0001】

本技術は、情報処理装置、情報処理方法、表示制御装置および表示制御方法に関し、特に、音声を含むコンテンツを処理する情報処理装置等に関する。

【背景技術】

【0002】

従来、ユーザの嗜好を考慮、類推して、コンテンツ群から面白いと思われる部分のみを抽出して再生する方法が存在する。万人が面白いと思う箇所を特定する方法では、どうしても抽出対象となるコンテンツが少なくなる。そのため、スライドショー等で用いるには不足分のコンテンツを補う必要がある。また、ユーザの嗜好を考慮する場合は、ユーザの嗜好を特定するためのデータベースをどのように作るかという問題がある。

20

【0003】

例えば、特許文献1において、スピーチ等を収録したコンテンツの中から自動的にチャプタを生成し再生するという技術が提案されている。この技術は、音の波形解析から、音声の塊を判定し、音声の塊単位でチャプタジャンプを実施することで、視聴するというものである。また、例えば、特許文献2において、音解析から編集ポイントを自動判定する技術が提案されている。この技術は、一定時間無音区間が続く場合に該当区間を編集ポイントとして記憶しておき、後の編集時にカット編集するものである。

【先行技術文献】

【特許文献】

30

【0004】

【特許文献1】国際公開第2009/025155号

【特許文献2】特開平10-84526号公報

【発明の概要】

【発明が解決しようとする課題】

【0005】

特許文献1で提案されている技術では、スピーチの内容を聴く上で不要な区間を特定することができず、コンテンツ中の有意な箇所のみを切り出して再生することまではできない。また、音声の塊を音の振幅によって判定するため、語尾が間延びしてしまった区間等を判定することができてない。

40

【0006】

また、特許文献2で提案されている技術では、会話やスピーチ時に発生する冗長な区間として考えられる、長すぎる無音区間を特定することはできる。しかし、音声以外で発生している音や、音声ではあるが会話やスピーチの内容を聴く上で、意味をなさない音のみの区間を特定することができない。一般に、会話の最中には、なるべく無意味な区間がないようにするため、音声検出される区間においても有意な区間が無効な区間かを判定する方法が望まれる

【0007】

本技術の目的は、編集ポイント特定のための良好な情報の提供を実現することにある。

【課題を解決するための手段】

50

## 【0008】

本技術の概念は、  
音声を含むコンテンツの言語解析に基づいて、該コンテンツの編集ポイントを特定するための情報を取得する情報取得部と、  
上記取得された情報を出力する情報出力部とを備える  
情報処理装置にある。

## 【0009】

本技術において、情報取得部により、音声を含むコンテンツの言語解析に基づいて、このコンテンツの編集ポイントを特定するための情報が取得される。例えば、この編集ポイントを特定するための情報は、編集ポイントの近傍期間における言語解析による言語情報を含む、ようにされてもよい。この場合、例えば、言語情報は、音声の発音情報、音声の文字列情報などを含む、ようにされてもよい。

10

## 【0010】

情報出力部により、取得されたコンテンツの編集ポイントを特定するための情報が出力される。例えば、情報出力部は、この情報をコンテンツと関連付けて記録メディアに記録する、ようにされてもよい。また、例えば、情報出力部は、この情報をコンテンツの送信単位に合わせてネットワークに送信する、ようにされてもよい。

## 【0011】

このように本技術においては、音声を含むコンテンツに対して言語解析に基づいてそのコンテンツの編集ポイントを特定するための情報を取得して出力するものであり、編集ポイント特定のための良好な情報の提供が可能となる。

20

## 【0012】

なお、本技術において、例えば、情報出力部は、編集ポイントの近傍期間における、コンテンツに含まれる映像の動き情報をさらに出力する、ようにされてもよい。このように映像の動き情報をさらに出力することで、編集ポイントを特定するための情報として、さらに有用な情報の提供が可能となる。

## 【0013】

また、本技術の他の概念は、  
音声を含むコンテンツの言語解析に基づいて得られた、該コンテンツの編集ポイントを特定するための情報を取得する情報取得部と、  
上記取得された情報に基づいて、上記コンテンツに関する表示を制御する表示制御部とを備える  
表示制御装置にある。

30

## 【0014】

本技術において、情報取得部により、音声を含むコンテンツの言語解析に基づいて得られた、このコンテンツの編集ポイントを特定するための情報が取得される。例えば、この編集ポイントを特定するための情報は、編集ポイントの近傍期間における言語解析による言語情報を含む、ようにされてもよい。この場合、例えば、言語情報は、音声の発音情報、音声の文字列情報などを含む、ようにされてもよい。そして、表示制御部により、取得された情報に基づいて、コンテンツに関する表示が制御される。

40

## 【0015】

例えば、表示制御部は、編集ポイントを特定するための情報に基づいて、コンテンツの有用度の高い第1の区間と有用度の低い第2の区間を特定し、コンテンツのタイムラインが第1の区間と第2の区間とが識別可能な状態で表示されるように制御する、ようにされてもよい。

## 【0016】

この場合、例えば、表示制御部は、タイムラインの所定区間の選択に応じて、言語情報に含まれる文字列情報を利用して、該所定区間の音声に対応する文字列が表示されるように制御する、ようにされてもよい。このとき、例えば、表示制御部は、所定区間の音声に対応する文字列と共に、前後の区間の音声に対応した文字列も表示され、所定区間の音声

50

に対応する文字列がハイライト表示されるように制御する、ようにされてもよい。ここで、ハイライト表示は、例えば、他の区間とは表示色を異ならせることによるハイライト表示などである。

【0017】

また、例えば、表示制御部は、言語情報に含まれる文字列情報で示される文節毎の区間に分割されたコンテンツのタイムラインが表示されると共に、このタイムラインの各区間に対応して、各区間の音声に対応した文字列が表示されるように制御する、ようにされてもよい。

【0018】

このように本技術においては、音声を含むコンテンツの言語解析に基づいて得られた、このコンテンツの編集ポイントを特定するための情報に基づいて、コンテンツに関する表示を制御するものである。そのため、コンテンツに関する表示、例えば編集のためのタイムライン等の表示を良好に行うことが可能となる。

【発明の効果】

【0019】

本技術によれば、編集ポイント特定のための良好な情報の提供を実現できる。

【図面の簡単な説明】

【0020】

【図1】実施の形態としての情報処理装置の構成例を示すブロック図である。

【図2】音声特徴検出部の処理手順の一例を示すフローチャートである。

【図3】テキスト変換部の処理手順の一例を示すフローチャートである。

【図4】言語解析結果記録フォーマット変換部が作成するコンテンツ関連付け構造体、文字列構造体を示す図である。

【図5】文字列、音声特徴区間（スプリット）および映像データ（コンテンツ）との対応関係の一例を模式的に示す図である。

【図6】実施の形態としての編集装置の構成例を示すブロック図である。

【図7】言語解析情報リスト生成部における文章毎のサブチャプタ登録の処理手順の一例を示すフローチャートである。

【図8】音声特徴区間の有用度の高低を判定する処理手順の一例を示すフローチャートである。

【図9】テキスト情報取得部における文字列データの取得の処理手順の一例を示すフローチャートである。

【図10】表示部に表示される編集画面の表示例を示す図である。

【図11】ユーザがマウス操作等でタイムラインの所定サブチャプタ区間の選択を行った場合に、その所定サブチャプタ区間およびその前後の区間の音声に対応した文字列が表示されることを説明するための図である。

【図12】言語解析情報リスト生成部における文章毎のサブチャプタ登録の処理手順の一例を示すフローチャートである。

【図13】各音声特徴区間（スプリット）毎に分割されたタイムラインが表示され、さらに各音声特徴区間に対応させてその区間の音声に対応した文字列が表示されることを示す図である。

【図14】テーブル生成部におけるテーブル生成の処理手順の一例を示すフローチャートである。

【図15】単語選択による映像検索システムを説明するための図である。

【図16】ネットワークへのライブストリームを実施する場合における情報処理装置の構成例を示すブロック図である。

【図17】ネットワーク送信を行う場合におけるコンテンツ関連付け構造体の一例を示す図である。

【図18】情報処理装置の他の構成例を示すブロック図である。

【図19】解析情報統合部で作成するコンテンツ関連付け構造体およびそれに含まれる文

10

20

30

40

50

字列構造体、動き構造体などを示す図である。

【図 20】編集装置の他の構成例を示すブロック図である。

【図 21】音声特徴区間の有用度の高低を判定する処理手順の他の例を示すフローチャートである。

【図 22】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】

【0021】

以下、発明を実施するための形態（以下、「実施の形態」とする）について説明する。なお、説明を以下の順序で行う。

1. 実施の形態

10

2. 変形例

【0022】

< 1. 実施の形態 >

[ 情報処理装置の構成例 ]

図 1 は、情報処理装置 10 の構成例を示している。この情報処理装置 10 は、映像入力部 101 と、音声入力部 102 と、動画多重化部 103 と、言語解析部 104 と、言語マッチングデータベース 105 と、動画 / 言語解析情報関連付け部 106 と、記録メディア 107 を有している。

【0023】

映像入力部 101 は、コンテンツを構成する映像データを入力する部分である。音声入力部 102 は、上述の映像入力部 101 に入力される映像データに対応した、コンテンツを構成する音声データを入力する部分である。動画多重化部 103 は、映像入力部 101 に入力された映像データと音声入力部 102 に入力された音声データを多重化して多重化データを生成する。

20

【0024】

言語解析部 104 は、音声特徴検出部 104a、テキスト変換部 104b および言語解析結果記録フォーマット変換部 104c を有している。音声特徴検出部 104a は、音声入力部 102 に入力された音声データを解析して、音声特徴区間（スプリット）を検出する。音声特徴検出部 104a は、音声特徴区間を検出する毎に、音声特徴の種別を示す発音情報と、開始および終了の時間情報を出力する。

30

【0025】

音声特徴区間は、例えば、（a）母音の長音（う～ん、え～、あ～、など）区間、（b）破裂音、強いアクセントで始まる区間、（c）擬音表現の区間、（d）音声の高低の変化区間、等である。（a）の母音の長音区間は、話している中では比較的無意味な区間である可能性が高く、編集時にはカットしたい区間である。（b）の破裂音、強いアクセントで始まる区間は、話をしている人が強調したいところ、驚いたところ等であって、編集時には採用したい区間である。

【0026】

（c）の擬音表現の区間は、話をしている人が何かを説明したいが適切な言葉を思い出せないか、あるいはその何かを強調したいと思ったところであって、編集時には採用したい区間である。（d）の音の高低の変化区間に関して、低への変化区間はより説明的な部分であることが多く、高への変化区間は、感情的表現が行われていることが多く、いずれも編集時には採用したい区間である。

40

【0027】

図 2 のフローチャートは、音声特徴検出部 104a の処理手順の一例を示している。まず、音声特徴検出部 104a は、ステップ S T 1 において、音声データの入力を受け付ける。次に、音声特徴検出部 104a は、ステップ S T 2 において、音声解析を行う。次に、音声特徴検出部 104a は、ステップ S T 3 において、音声特徴区間の開始から終了まで進んだ状態にあるか否か、つまり音声特徴区間が検出されたか否かを判断する。

【0028】

50

音声特徴区間が検出されていないとき、音声特徴検出部 104a は、ステップ S T 1 に戻って、上述したと同様の処理を繰り返す。一方、音声特徴区間が検出されるとき、音声特徴検出部 104a は、ステップ S T 4 において、音声特徴区間の種別を特定し、その種別情報と、開始および終了の時間情報を、検出された音声特徴区間の情報として出力する。その後、音声特徴検出部 104a は、ステップ S T 1 に戻って、次の音声特徴区間の検出処理に移る。

【0029】

テキスト変換部 104b は、音声入力部 102 に入力された音声データを、言語マッチングデータベース 105 を用いて言語解析し、動画のチャプタ毎に、それに含まれる文章毎の文字列を抽出する。テキスト変換部 104b は、文字列を検出する毎に、文字列データと、開始および終了の時間情報を出力する。

10

【0030】

図 3 のフローチャートは、テキスト変換部 104b の処理手順の一例を示している。まず、テキスト変換部 104b は、ステップ S T 1 1 において、音声データの入力を受け付ける。次に、テキスト変換部 104b は、ステップ S T 1 2 において、会話開始、すなわち文章開始であるか否かを判断する。文章開始でないとき、テキスト変換部 104b は、ステップ S T 1 1 に戻って、上述したと同様の処理を繰り返す。

【0031】

文章開始であるとき、テキスト変換部 104b は、ステップ S T 1 3 において、音声データに対して、言語解析を施し、テキスト化の処理を行う。そして、テキスト変換部 104b は、ステップ S T 1 4 において、文章の区切れ目（音声の区切れ目）を検知したか否かを判断する。区切れ目を検知していないとき、テキスト変換部 104b は、区切れ目を検知するまで、ステップ S T 1 5 において音声データを入力し、ステップ S T 1 3 において言語解析を施し、テキスト化の処理を行う。

20

【0032】

文章の区切れ目を検知するとき、テキスト変換部 104b は、ステップ S T 1 6 において、文章の情報、すなわち、文字列データと、開始および終了の時間情報を出力する。その後、テキスト変換部 104b は、ステップ S T 1 1 に戻って、次の文章の言語解析によるテキスト化の処理に移る。

【0033】

図 1 に戻って、言語解析結果記録フォーマット変換部 104c は、音声特徴検出部 104a からの各音声特徴区間の情報と、テキスト変換部 104b からの各文章の情報を、記録フォーマットに変換する。

30

【0034】

ここで、言語解析結果記録フォーマット変換部 104c は、チャプタ毎に、図 4 ( a ) に示すような、コンテンツ関連付け構造体を作成する。「関連コンテンツを特定する ID」は、コンテンツのチャプタ識別情報である。「解析文字列のリスト」の部分に、チャプタに含まれる各文章の、図 4 ( b ) に示すような、文字列構造体が含まれる。

【0035】

この文字列構造体には、文章の文字列データ（図示の例では“ X X X X A A A A A C C C C C B B B B B ” のデータ）が含まれ、また、この文字列の開始時間および終了時間の情報が含まれる。ここで、開始時間は、文字列の発音が開始された時間を示し、終了時間は文字列の発音が終了した時間を示す。開始時間および終了時間は、コンテンツを構成する映像データおよび音声データとの対応関係を示すものとなる。つまり、開始時間および終了時間は、この文字列がコンテンツのどの区間に対応するかを示し、編集時における切り出しのために必要となる。

40

【0036】

また、この文字列構造体には、文章内の各音声特徴区間（スプリット）の情報が含まれる。1つの音声特徴区間の情報は、開始文字位置、開始時間および発音情報からなる。開始文字位置は、文字列中の区切れ目位置を示す。この開始文字位置は、文字列と音声特徴

50

区間とを対応させるために、文字列の何番目という形で記述される。例えば、図示の例において、文字列中の「A A A A」に対応する音声特徴区間における開始文字位置は、5文字目である。

【0037】

開始時間は、区切れ目位置の時間を示し、コンテンツを構成する映像データおよび音声データとの対応関係を示す。発音情報は、音声特徴の種別を示す。音声特徴の種別は、上述したように、母音の長音区間、破裂音、強いアクセントで始まる区間、擬音表現の区間、音声の高低の変化区間などである。

【0038】

図5は、文字列、音声特徴区間(スプリット)および映像データ(コンテンツ)との対応関係の一例を模式的に示している。図5(a)は文字列を示し、図5(b)は映像データの各フレームをフィルム形式で模式的に表している。この例では、“XXXXA A A A A C C C C C C B B B B B”の文字列に、「split[0]」-「split[3]」の4個の音声特徴区間(スプリット)が含まれている。

10

【0039】

図1に戻って、動画/言語解析情報関連付け部106は、動画多重化部103で得られる映像データおよび音声データの多重化データと、言語解析部104で得られるコンテンツ関連付け構造体とを、関連付けして、記録メディア107に記録する。ここで、関連付けは、例えば、以下の(a)~(d)のような形式で記録メディアに記録することで行うことができる。

20

【0040】

(a)チャプタ毎に、チャプタの多重化データの末尾にそれに対応するコンテンツ関連付け構造体(図4参照)を付加して、記録する。(b)チャプタ毎に、チャプタの多重化データと、それに対応するコンテンツ関連付け構造体とを、同名ファイル、拡張子違いで、記録する。(c)チャプタ毎に、チャプタの多重化データと、それに対応するコンテンツ関連付け構造体を任意のファイル名で記録するが、それらを関連付けするための別データベースを作成しておく。

【0041】

(d)チャプタ毎に、チャプタの多重化データと、それに対応するコンテンツ関連付け構造体を任意のファイル名で記録するが、各チャプタのコンテンツ関連付け構造体の中に、対応する多重化データを特定する情報、例えばUUI D(Universally Unique Identifier)を埋め込んでおく。

30

【0042】

図1に示す情報処理装置10の動作を説明する。映像入力部101に入力される映像データは、動画多重化部103に供給される。また、音声入力部102に入力される音声データは、動画多重化部103に供給される。動画多重化部103では、映像データと音声データとが多重化され、多重化データが得られる。

【0043】

また、音声データ入力部102に入力される音声データは、言語解析部104の音声特徴検出部104aおよびテキスト変換部104bに供給される。音声特徴検出部104aでは、音声データが解析されて、音声特徴区間(スプリット)が検出される。この音声特徴検出部では、音声特徴区間が検出される毎に、音声特徴区間の種別を示す発音情報と、開始および終了の時間情報が出力される。

40

【0044】

また、テキスト変換部104bでは、音声データが、言語マッチングデータベース105が用いられて言語解析され、動画のチャプタ毎に、それに含まれる文章毎の文字列が検出される。このテキスト変換部104では、文章毎の文字列が検出される毎に、文字列データと、開始および終了の時間情報が出力される。

【0045】

言語解析部104において、音声特徴検出部104aの出力情報およびテキスト変換部

50



10 bの出力情報は、言語解析結果記録フォーマット変換部104cに供給される。このフォーマット変換部104cでは、音声特徴検出部104aからの各音声特徴区間の情報と、テキスト変換部104bからの各文章の情報が、記録フォーマットに変換される。

【0046】

すなわち、このフォーマット変換部104cでは、チャプタ毎に、コンテンツ関連付け構造体が作成される(図4(a)参照)。このコンテンツ関連付け構造体には、チャプタに含まれる各文章の文字列構造体が含まれる(図4(b)参照)。この文字列構造体には、文章の文字列データが含まれ、また、この文字列の開始時間および終了時間の情報が含まれる。また、この文字列構造体には、文章内の各音声特徴区間(スプリット)の情報が含まれる。ここで、1つの音声特徴区間の情報には、開始文字位置、開始時間および発音

10

【0047】

動画多重化部103で得られる映像データおよび音声データの多重化データは、動画/言語解析情報関連付け部106に供給される。また、この関連付け部106には、言語解析部104で得られる記録フォーマット化された各文章の情報(コンテンツ関連付け構造体)が供給される。この関連付け部106では、映像データおよび音声データの多重化データに対して、各文章の情報が関連付けられて、記録メディア107に記録することが行われる。

【0048】

上述したように、図1に示す情報処理装置10においては、コンテンツを構成する音声データに対して言語解析が行われて、チャプタ毎に、コンテンツ関連付け構造体が作成される。このコンテンツ関連付け構造体には、音声データの解析による音声特徴区間(スプリット)の情報が含まれると共に、テキスト変換による文字列データ等が含まれる。そして、コンテンツを構成する映像データおよび音声データの多重化データと共に、コンテンツ関連付け構造体が関連付けられて、記録メディア107に記録される。そのため、編集ポイント特定のための良好な情報の提供が可能となる。

20

【0049】

[編集装置の構成例]

図6は、編集装置20の構成例を示している。この編集装置20は、上述の図1に示す情報処理装置10で記録メディア107に記録された各チャプタの多重化データおよびコンテンツ関連付け構造体を取り扱う。この編集装置20は、動画読み込み部201と、言語解析情報読み込み部202と、動画/言語解析情報関連付け部203と、言語解析情報リスト生成部204を有している。また、この編集装置20は、タイムライン(チャプタ)生成部205と、表示制御部206と、テキスト情報取得部207と、表示部208を有している。

30

【0050】

動画読み込み部201は、記録メディア107から、編集に係る所定のチャプタの映像データおよび音声データの多重化データを読み込む。言語解析情報読み込み部202は、記録メディア107から、動画読み込み部201で読み込まれた所定チャプタの映像データおよび音声データの多重化データに対応したコンテンツ関連付け構造体(図4参照)を読み込む。この際、動画/言語解析情報関連付け部203は、チャプタの識別情報あるいはファイル情報などの関連付け情報を、言語解析情報読み込み部202に送る。

40

【0051】

言語解析情報リスト生成部204は、言語解析情報読み込み部202で読み込まれたコンテンツ関連付け構造体を取り込み、言語解析情報リストを生成する。すなわち、このリストには、所定のチャプタに含まれる文章毎の文字列データ、開始時間および終了時間が登録される。また、このリストには、その文章毎の、各音声特徴区間(スプリット)の開始文字位置、開始時間および発音情報が登録される。

【0052】

さらに、このリストには、その文章毎に、各音声特徴区間(スプリット)の有用度の高

50

低が判定され、有用度の高低で区分けした区間、つまりサブチャプタが登録される。このサブチャプタの登録では、そのサブチャプタの識別情報と、そのサブチャプタの開始位置、終了位置の情報とが関連付けられて記憶される。

【 0 0 5 3 】

図7のフローチャートは、リスト生成部204における、文章毎のサブチャプタ登録の処理手順の一例を示している。リスト生成部204は、ステップST31で処理を開始し、その後に、ステップST32の処理に移る。このステップST32において、リスト生成部204は、処理対象の文字列構造体を取り出す。

【 0 0 5 4 】

次に、リスト生成部204は、ステップST33において、最初の音声特徴区間（スプリット）の情報を処理対象の音声特徴区間として取り出す。そして、リスト生成部204は、ステップST34において、その音声特徴区間が有用度の高い区間であるか低い区間であるかを判定する。

【 0 0 5 5 】

次に、リスト生成部204は、ステップST35において、区間切り替わりか否かを判定する。区間切り替わりでないとき、リスト生成部204は、ステップST33に戻って、次の音声特徴区間（スプリット）の情報を処理対象の音声特徴区間として取り出し、その処理に移る。一方、区間切り替わりであるとき、リスト生成部204は、ステップST34において、切り替わり前までの区間についてサブチャプタの登録を行う。

【 0 0 5 6 】

その後、リスト生成部204は、ステップST33に戻って、次の音声特徴区間（スプリット）の情報を処理対象の音声特徴区間として取り出し、その処理に移る。なお、次の音声特徴区間がないときは、サブチャプタ登録がされていないそこまでの区間についてサブチャプタ登録を行って、処理を終了する。

【 0 0 5 7 】

図8は、図7のフローチャートにおけるステップST34の区間判定の処理手順の一例を示している。リスト生成部204は、ステップST41において、処理を開始し、その後に、ステップST42の処理に移る。このステップST42において、リスト生成部204は、処理対象の音声特徴区間が母音の長音区間であるか否かを判定する。

【 0 0 5 8 】

母音の長音区間でないとき、リスト生成部204は、ステップST43において、処理対象の音声特徴区間の開始文字列が辞書に存在するか否かを判定する。辞書に存在するとき、リスト生成部204は、ステップST44において、処理対象の音声特徴区間は有用度の高い区間と判定する。その後、リスト生成部204は、ステップST45において、処理を終了して戻る。

【 0 0 5 9 】

ステップST43で辞書に存在しないとき、リスト生成部204は、ステップST46において、処理対象の音声特徴区間が破裂音で始まる区間か否かを判定する。破裂音で始まる区間であるとき、リスト生成部204は、ステップST44において、処理対象の音声特徴区間は有用度の高い区間と判定する。その後、リスト生成部204は、ステップST45において、処理を終了して戻る。

【 0 0 6 0 】

ステップST42で処理対象の音声特徴区間が母音の長音区間であるとき、あるいはステップST46で処理対象の音声特徴区間が破裂音で始まる区間でないとき、リスト生成部204は、ステップST47において、処理対象の音声特徴区間は有用度の低い区間と判定する。その後、リスト生成部204は、ステップST45において、処理を終了する。

【 0 0 6 1 】

図6に戻って、タイムライン生成部205は、動画読み込み部201で読み込まれた、編集に係る所定のチャプタの映像データおよび音声データのタイムラインを作成する。こ

10

20

30

40

50

のタイムラインの作成は、表示制御部 206 の制御のもとで行われる。表示制御部 206 は、上述の言語解析情報リストに登録されるサブチャプタ登録に基づいて、有用度の高い区間と有用度の低い区間とが識別可能な状態でタイムラインが作成されるように、タイムライン生成部 205 の動作を制御する。

#### 【0062】

テキスト情報取得部 207 は、タイムラインに対応させて表示する文字列データを、言語解析情報リストに基づいて取得する。このテキスト情報の取得は、表示制御部 206 の制御のもとで行われる。表示制御部 206 は、ユーザによるタイムラインの所定サブチャプタ区間の選択情報に基づいて、その所定サブチャプタ区間およびその前後のサブチャプタ区間の音声に対応した文字列データが取得されるように、テキスト情報取得部 207 の動作を制御する。なお、ユーザは、所定サブチャプタ区間の選択を、図示しないマウス等のポインティングデバイスを操作することで行う。

10

#### 【0063】

図 9 のフローチャートは、テキスト情報取得部 207 における文字列データの取得の処理手順の一例を示している。テキスト情報取得部 207 は、ステップ ST51 において、ユーザによるサブチャプタの選択処理があるとき、処理を開始する。

#### 【0064】

次に、テキスト情報処理部 207 は、ステップ ST52 において、選択されたサブチャプタの開始位置、終了位置の情報を取得する。この場合、テキスト情報処理部 207 は、表示制御部 206 から与えられる選択されたサブチャプタの識別情報に基づき、言語解析情報リスト生成部 204 で生成された言語解析情報リストから、そのサブチャプタの開始位置、終了位置の情報を取得する。

20

#### 【0065】

次に、テキスト情報処理部 207 は、ステップ ST53 において、サブチャプタ区間およびその前後の区間の文字列データを取得する。この場合、ステップ ST52 で取得されたそのサブチャプタの開始位置、終了位置の情報に基づき、言語解析情報リストの対応する音声特徴区間（スプリット）の情報を参照して、必要な文字列データを取得する。

#### 【0066】

図 6 に戻って、表示部 208 は、LCD (Liquid Crystal Display) 等のディスプレイを用いて構成されており、編集画面を表示する。この表示部 208 は、表示制御部 206 の制御のもと、タイムライン生成部 205 で生成されたタイムラインを表示し、さらに、テキスト情報取得部 207 で取得された文字列を表示する。

30

#### 【0067】

図 10 は、表示部 208 に表示される編集画面の表示例を示している。画面の下部に水平方向に延びる映像タイムライン VTL および音声タイムライン ATL が並べて表示されている。また、画面の右上部にプレビュー画面領域 PVS が設けられている。プレビュー画面領域には、例えば、ユーザが映像タイムライン VTL 上で指定した位置における静止画、あるいはその位置からの動画が表示される。タイムライン VTL, ATL は、図示のように、有用度の高いサブチャプタ区間 HS と有用度の低いサブチャプタ区間 LS とが、例えば明度、色相、彩度、模様などにより識別可能な状態で表示されている。

40

#### 【0068】

ユーザがマウス操作等でタイムラインの所定サブチャプタ区間の選択を行った場合、図 11 に示すように、この所定サブチャプタ区間およびその前後の区間の音声に対応した文字列が表示される。この場合、例えば、全ての文字列を同じ状態で表示されてもよいが、この実施の形態において、例えば、所定サブチャプタ区間の音声に対応した文字列は、その前後の区間の音声の文字列に対して、表示色を異ならせる等の方法で、ハイライト表示される。これにより、所定サブチャプタ区間の文字列とその他の区間の文字列を、視覚的に容易に区別可能となる。

#### 【0069】

図 6 に示す編集装置 20 の動作を説明する。動画読み込み部 201 では、記録メディア

50

107から、編集に係る所定のチャプタの映像データおよび音声データの多重化データが読み込まれる。また、言語解析情報読み込み部202では、記録メディア107から、動画読み込み部201で読み込まれた所定チャプタの映像データおよび音声データの多重化データに対応したコンテンツ関連付け構造体を読み込まれる(図4参照)。

【0070】

言語解析情報読み込み部202で読み込まれたコンテンツ関連付け構造体は、言語解析情報リスト生成部204に供給される。このリスト生成部204では、コンテンツ関連付け構造体に基づいて、言語解析情報リストが生成される。このリストには、所定のチャプタに含まれる文章毎の文字列データ、開始時間および終了時間が登録される。また、このリストには、その文章毎の、各音声特徴区間(スプリット)の開始文字位置、開始時間および発音情報が登録される。

10

【0071】

さらに、このリストには、その文章毎に、各音声特徴区間(スプリット)の有用度の高低が判定され、有用度の高低で区分けした区間、つまりサブチャプタが登録される。このサブチャプタの登録では、そのサブチャプタの識別情報と、そのサブチャプタの開始位置、終了位置の情報とが関連付けされて記憶される。

【0072】

動画読み込み部201で読み込まれた、編集に係る所定のチャプタの映像データおよび音声データは、タイムライン生成部205に供給される。このタイムライン生成部205では、編集に係る所定のチャプタの映像データおよび音声データのタイムラインが作成される。この場合、表示制御部206の制御のもと、言語解析情報リストに登録されるサブチャプタ登録に基づいて、有用度の高い区間と有用度の低い区間とが識別可能な状態でタイムラインが作成される。

20

【0073】

タイムライン生成部205で生成されたタイムラインの情報は、表示部208に供給される。表示部208には、表示制御部206の制御のもと、映像タイムラインVTLおよび音声タイムラインATLが表示される。この場合、タイムラインVTL, ATLは、有用度の高い区間HSと有用度の低い区間LSとが識別可能な状態で表示される(図10参照)。

【0074】

また、言語解析情報リスト生成部204で生成される言語解析情報リストは、テキスト情報取得部207に供給される。また、このテキスト情報取得部207には、ユーザによるタイムラインの所定サブチャプタ区間の選択に応じて、表示制御部206から、その所定サブチャプタ区間の識別情報が供給される。

30

【0075】

テキスト情報取得部207では、タイムラインに対応させて表示する文字列データとして、言語解析情報リストに基づいて、ユーザの選択に係る所定サブチャプタ区間およびその前後のサブチャプタ区間の音声に対応した文字列データが取得される。この文字列データは、表示部208に供給される。表示部208の編集画面には、ユーザの選択に係る所定サブチャプタ区間およびその前後のサブチャプタ区間の音声に対応した文字列が表示される(図11参照)。

40

【0076】

上述したように、図6に示す編集装置20においては、記録メディア107から編集に係る所定チャプタの映像データおよび音声データの多重化データが読み込まれるとき、それに関連付けて記録されているコンテンツ関連付け構造体の読み込みが行われ、言語解析情報リストが生成される。このコンテンツ関連付け構造体には、音声データの解析による音声特徴区間(スプリット)の情報が含まれると共に、テキスト変換による文字列データ等が含まれる。

【0077】

多重化データに基づいて映像や音声のタイムラインが作成されるとき、言語解析情報リ

50

ストに登録されるサブチャプタ登録に基づいて、有用度の高い区間と有用度の低い区間とが識別可能な状態でタイムラインが作成される。そのため、編集画面において、映像タイムラインV T Lおよび音声タイムラインA T Lは、有用度の高い区間H Sと有用度の低い区間L Sとが識別可能な状態で表示される。そのため、ユーザは、両区間の境界を編集ポイントとして活用でき、適切な編集を行うことができる。

【 0 0 7 8 】

また、ユーザがタイムライン上でマウス等を操作して所定サブチャプタを選択するとき、言語解析情報リストに基づいて、ユーザの選択に係る所定サブチャプタ区間およびその前後のサブチャプタ区間の音声に対応した文字列データが取得される。そのため、編集画面には、ユーザの選択に係る所定サブチャプタ区間およびその前後のサブチャプタ区間の音声に対応した文字列が表示される。そのため、ユーザは、コンテンツを再生することなく、ある程度のコンテンツの認識が可能となり、編集を効率的かつ効果的に行うことができる。

10

【 0 0 7 9 】

< 2 . 変形例 >

「サブチャプタ登録の他の例」

なお、上述実施の形態においては、編集装置20の言語解析情報リスト生成部204では、音声特徴区間(スプリット)毎に有用度の高低判定を行って、有用度の高低で区別した区間をサブチャプタとして登録するものである。そして、表示部208には、有用度の高い区間H Sと、有用度の低い区間L Sに分割された、映像、音声のタイムラインが表示されるものである。

20

【 0 0 8 0 】

しかし、文節毎の各区間、つまり各音声特徴区間(スプリット)をサブチャプタとして登録し、表示部208には、文節毎に分割された映像、音声のタイムラインを表示し、さらに、各文節に対応して文字列を表示するようにすることも考えられる。

【 0 0 8 1 】

図12のフローチャートは、言語解析情報リスト生成部204(図6参照)における、文章毎のサブチャプタ登録の処理手順の一例を示している。リスト生成部204は、ステップS T 6 1で処理を開始し、その後、ステップS T 6 2の処理に移る。このステップS T 6 2において、リスト生成部204は、処理対象の文章の文字列構造体を取り出す。

30

【 0 0 8 2 】

次に、リスト生成部204は、ステップS T 6 3において、ステップS T 6 2で取り出された文字列構造体に含まれる最初の音声特徴区間(スプリット)の情報を取り出す。そして、リスト生成部204は、ステップS T 6 4において、その音声特徴区間をサブチャプタとして登録する。このサブチャプタの登録では、そのサブチャプタの識別情報と、そのサブチャプタの開始位置、終了位置の情報とが関連付けられて記憶される。

【 0 0 8 3 】

その後、リスト生成部204は、ステップS T 6 3に戻って、次の音声特徴区間(スプリット)の情報を処理対象の音声特徴区間として取り出し、その処理に移る。なお、次の音声特徴区間がないときは、処理を終了する。

40

【 0 0 8 4 】

タイムライン生成部205では、編集に係る所定のチャプタの映像データおよび音声データのタイムラインを作成する際に、言語解析情報リストに登録されるサブチャプタ登録に基づいて、音声特徴区間(スプリット)毎に分割されたタイムラインが作成される。また、テキスト情報取得部207では、言語解析情報リストから、各音声特徴区間(スプリット)の音声に対応した文字列データがそれぞれ取得される。

【 0 0 8 5 】

そのため、表示部208には、タイムライン生成部205で生成されたタイムラインの情報により、図13に示すように、各音声特徴区間(スプリット)毎に分割されたタイムラインV T L , A T Lが表示される。また、表示部208には、テキスト情報取得部20

50

7で取得された文字列データに基づいて、各音声特徴区間（スプリット）に対応させて、その区間の音声に対応した文字列が表示される。

【0086】

「単語選択による映像検索システム」

また、上述していないが、編集装置20において、言語解析情報読み込み部202で読み込まれる編集に係る所定チャプタのコンテンツ関連付け構造体に基づいて、文字列を構成する単語と、その単語が属する音声特徴区間（スプリット）との対応関係を示すテーブルを作成できる。そして、このテーブルを用いて、選択された単語が存在する音声特徴区間（スプリット）に対応した映像を検索するシステムを構成できる。

【0087】

図14のフローチャートは、テーブル生成部（図6には図示していない）におけるテーブル生成の処理手順の一例を示している。テーブル生成部は、ステップST71において、処理を開始する。その後、テーブル作成部は、ステップST72において、記録メディア107から編集に係る所定ピクチャの映像データおよび音声データの多重化データに関連付けられたコンテンツ関連付け構造体を読み込む（図4参照）。

【0088】

次に、テーブル生成部は、ステップST73において、コンテンツ関連付け構造体に含まれる各文字列に対応した文字列構造体を取り出す。そして、テーブル生成部は、ステップST74において、各文字列構造体に含まれる文字列から単語を抽出する。この場合、擬音も単語として含める。そして、テーブル生成部は、ステップST75において、ステップST74で抽出した単語と、その単語が属する音声特徴区間（スプリット）との対応関係を、テーブルに登録する。

【0089】

ステップST74の単語抽出処理と、ステップST75のテーブル登録処理を、全ての単語について行うまで繰り返す。全ての単語について処理が終了したとき、テーブル生成部は、処理を終了する。

【0090】

図15は、上述のように生成されたテーブルを用いた検索例を示している。例えば、表示部208に表示される編集画面に、例えば、図15(a)に示すように、テーブルに登録された単語が、「A」から始まる単語、「B」から始まる単語、・・・等のように分類されて表示される。

【0091】

ユーザがその中から所定の単語を選択するとき、テーブルが参照されて、その単語が属する音声特徴区間（スプリット）が取得され、その区間の所定フレーム、例えば、最初と最後のフレームの画像が、例えば、図15(b)に示すように、編集画面のプレビュー画面領域PVSに表示される。なお、プレビュー画面領域PVSに、該当する区間の動画が表示されるようにしてもよい。また、これと共に、タイムライン上に該当区間が明示されるようにしてもよい。

【0092】

「ネットワークへの送信」

また、上述実施の形態において、情報処理装置10（図1参照）の動画/言語解析情報関連付け部106は、コンテンツを構成する映像データおよび音声データの多重化データと言語解析情報（コンテンツ関連付け構造体）とを記録メディア107に関連付けて記録する例を示した。

【0093】

しかし、ネットワークへのライブストリームを実施する場合には、コンテンツを構成する映像データおよび音声データの多重化データに言語解析情報を関連付けて送信することになる。その場合、多重化データは、例えば4秒とかの送信単位（バッファリング単位）で送信されるが、参照性をよくするために、言語解析情報もその送信単位で区切られたものとする。

10

20

30

40

50

## 【 0 0 9 4 】

図 1 6 は、ネットワークへのライブストリームを実施する場合における情報処理装置 1 0 A の構成例を示している。この図 1 6 において、図 1 と対応する部分には同一符号を付し、適宜、その詳細説明は省略する。

## 【 0 0 9 5 】

言語解析部 1 0 4 の言語解析結果記録フォーマット変換部 1 0 4 cA は、コンテンツを構成する映像データおよび音声データの多重化データの送信単位毎に、図 1 7 に示すような、コンテンツ関連付け構造体を作成する。「関連コンテンツを特定する ID」は、コンテンツのチャプタ識別情報である。

## 【 0 0 9 6 】

また、このコンテンツ関連付け構造体には、「参考テキスト」、「開始時間」、「継続時間」、「発音情報」などの情報が含まれる。「参考テキスト」の情報は、送信単位区間中に発生された言葉の文字列データである。「開始時間」は、区切れ目位置の開始時間位置を示す。「継続時間」は、区切れ目位置の継続時間を示し、送信単位区間に対応した最長継続時間以下の時間である。「発音情報」は、音声特徴の種別を示す。音声特徴の種別は、上述したように、母音の長音区間、破裂音、強いアクセントで始まる区間、擬音表現の区間、音声の高低の変化区間などである。

## 【 0 0 9 7 】

動画 / 言語解析情報関連付け部 1 0 6 A は、動画多重化部 1 0 3 で得られた多重化データを、送信単位で、順次ネットワーク 1 0 8 に送信する。その際、動画 / 言語解析情報関連付け部 1 0 6 A は、送信単位毎に、言語解析部 1 0 4 で生成されたコンテンツ関連付け構造体（図 1 7 参照）を付加して送信する。

## 【 0 0 9 8 】

なお、図 1 6 に示す情報処理装置 1 0 A においては、破線矢印で示すように、言語マッチングデータベース 1 0 5 は、ネットワーク（クラウド）1 0 8 に存在するものを利用することもできる。

## 【 0 0 9 9 】

「動き解析情報の付加」

また、上述実施の形態においては、映像データおよび音声データの多重化データに言語解析情報のみ関連付けて記録あるいは送信する例を示した。しかし、映像データから取得される動き解析情報を、さらに、関連付けて記録あるいは送信することも考えられる。これにより、編集ポイントを特定するための情報として、さらに有用な情報の提供が可能となる。

## 【 0 1 0 0 】

図 1 8 は、情報処理装置 1 0 B の構成例を示している。この図 1 8 において、図 1 と対応する部分には同一符号を付し、適宜、その詳細説明は省略する。この情報処理装置 1 0 B は、映像入力部 1 0 1 と、音声入力部 1 0 2 と、動画多重化部 1 0 3 と、言語解析部 1 0 4 と、言語マッチングデータベース 1 0 5 を有している。また、この情報処理装置 1 0 B は、動き特徴検出部 1 1 1 と、動き解析結果記録フォーマット変換部 1 1 2 と、解析情報統合部 1 1 3 と、動画 / 解析情報関連付け部 1 0 6 B と、記録メディア 1 0 7 を有している。

## 【 0 1 0 1 】

言語解析部 1 0 4 は、音声特徴検出部 1 0 4 a、テキスト変換部 1 0 4 b および言語解析結果記録フォーマット変換部 1 0 4 cB を有している。記録フォーマット変換部 1 0 4 cB は、チャプタに含まれる文章毎に、音声特徴検出部 1 0 4 a からの各音声特徴区間の情報と、テキスト変換部 1 0 4 b からの各文章の情報に基づいて、図 1 9 ( d ) に示すような、文字列構造体を作成する。

## 【 0 1 0 2 】

この文字列構造体には、図 1 9 ( b ) に示すような、解析情報構造体が含まれる。この解析情報構造体は、「構造種別」、「開始時間」および「終了時間」の情報を持っている

10

20

30

40

50

。文字列構造体に含まれる解析情報構造体にあつては、「構造種別」の情報は文字列構造であることを示し、開始時間は文字列の発音が開始された時間を示し、終了時間は文字列の発音が終了した時間を示す。

**【 0 1 0 3 】**

また、文字列構造体には、文章の文字列データ（図示の例では“ X X X X A A A A C C C C C B B B B ”のデータ）が含まれる。また、この文字列構造体には、文章内の各音声特徴区間（スプリット）の情報が含まれる。1つの音声特徴区間の情報は、開始文字位置、開始時間および発音情報からなる。開始文字位置は、文字列中の区切れ目位置を示す。この開始文字位置は、文字列と音声特徴区間とを対応させるために、文字列の何番目という形で記述される。例えば、図示の例において、文字列中の「 A A A A 」に対応する音声特徴区間における開始文字位置は、5文字目である。

10

**【 0 1 0 4 】**

動き特徴検出部 1 1 1 は、映像入力部 1 0 1 に入力された映像データを解析して、動き特徴を検出する。ここで、検出される動き特徴には、フォーカス位置移動、撮影方向変化などがある。動き特徴検出部 1 1 1 は、動き特徴区間毎に、動き特徴の種別と、開始および終了の時間情報を出力する。

**【 0 1 0 5 】**

動き解析結果記録フォーマット変換部 1 1 2 は、動き特徴区間毎に、動き特徴検出部 1 1 1 からの動き特徴区間の情報に基づいて、図 1 9 ( d ) に示すような、動き構造体を作成する。この動き構造体には、図 1 9 ( b ) に示すような、解析情報構造体が含まれると共に、動き特徴の種別、例えばフォーカス位置移動、撮影方向変化などを示す「動き種別」の情報が含まれる。

20

**【 0 1 0 6 】**

解析情報構造体は、「構造種別」、「開始時間」および「終了時間」の情報を持っている。動き構造体に含まれる解析情報構造体にあつては、「構造種別」の情報は動き構造であることを示し、開始時間は動き特徴区間の開始時間を示し、終了時間は動き特徴区間の終了時間を示す。

**【 0 1 0 7 】**

解析情報統合部 1 1 3 は、チャプタ毎に、言語解析部 1 0 4 の言語解析結果記録フォーマット変換部 1 0 4 cB で作成された文字列構造体と、動き解析結果記録フォーマット変換部 1 1 2 で作成された動き構造体とを統合し、図 1 9 ( a ) に示すような、コンテンツ関連付け構造体を作成する。「関連コンテンツを特定する ID」は、コンテンツのチャプタ識別情報である。「解析情報のリスト」の部分に、各文章の文字列構造体と、各動き特徴区間の動き構造体が含まれる。

30

**【 0 1 0 8 】**

動画／解析情報関連付け部 1 0 6 B は、動画多重化部 1 0 3 で得られる映像データおよび音声データの多重化データと、解析情報統合部 1 1 3 で得られるコンテンツ関連付け構造体とを、関連付けして、記録メディア 1 0 7 に記録する。図 1 8 に示す情報処理装置 1 0 B のその他は、図 1 に示す情報処理装置 1 0 と同様に構成され、同様に動作する。

**【 0 1 0 9 】**

図 2 0 は、編集装置 2 0 B の構成例を示している。この図 2 0 において、図 6 と対応する部分には同一符号を付し、その詳細説明は省略する。この編集装置 2 0 B は、上述の図 1 8 に示す情報処理装置 1 0 B で記録メディア 1 0 7 に記録された各チャプタの多重化データおよびコンテンツ関連付け構造体を取り扱う。

40

**【 0 1 1 0 】**

この編集装置 2 0 B は、動画読み込み部 2 0 1 と、言語解析情報読み込み部 2 0 2 B と、動画／解析情報関連付け部 2 0 3 B と、解析情報リスト生成部 2 0 4 B を有している。また、この編集装置 2 0 B は、タイムライン（チャプタ）生成部 2 0 5 と、表示制御部 2 0 6 と、テキスト情報取得部 2 0 7 と、表示部 2 0 8 を有している。

**【 0 1 1 1 】**

50



解析情報読み込み部 202B は、記録メディア 107 から、動画読み込み部 201 で読み込まれた所定チャプタの映像データおよび音声データの多重化データに対応したコンテンツ関連付け構造体（図 19 参照）を読み込む。この際、動画/解析情報関連付け部 203B は、チャプタの識別情報あるいはファイル情報などの関連付け情報を、解析情報読み込み部 202A に送る。

【0112】

解析情報リスト生成部 204B は、解析情報読み込み部 202B で読み込まれたコンテンツ関連付け構造体を取り込み、解析情報リストを生成する。すなわち、このリストには、所定のチャプタに含まれる文章毎の文字列データ、開始時間および終了時間が登録される。また、このリストには、その文章毎の、各音声特徴区間（スプリット）の開始文字位置、開始時間および発音情報が登録される。また、このリストには、所定のチャプタに含まれる動き特徴区間毎の種別情報、開始時間および終了時間が登録される。

10

【0113】

さらに、このリストには、文章毎に、各音声特徴区間（スプリット）の有用度の高低が判定され、有用度の高低で分けした区間、つまりサブチャプタが登録される。このサブチャプタの登録では、そのサブチャプタの識別情報と、そのサブチャプタの開始位置、終了位置の情報とが関連付けされて記憶される。

【0114】

リスト生成部 204B における、文章毎のサブチャプタ登録の処理は、上述の図 6 の編集装置 20 と同様に、図 7 のフローチャートに沿って行われる。ただし、ステップ ST 34 の区間判定処理は、図 8 のフローチャートの代わりに、ここでは、図 21 のフローチャートに示すように、言語解析情報だけでなく、動き解析情報をも使用されて、行われる。

20

【0115】

リスト生成部 204B は、ステップ ST 81 において、処理を開始し、その後に、ステップ ST 82 の処理に移る。このステップ ST 82 において、リスト生成部 204B は、処理対象の音声特徴区間が母音の長音区間であるか否かを判定する。

【0116】

母音の長音区間でないとき、リスト生成部 204B は、ステップ ST 83 において、処理対象の音声特徴区間の開始文字列が辞書に存在するか否かを判定する。辞書に存在するとき、リスト生成部 204B は、ステップ ST 84 において、処理対象の音声特徴区間は有用度の高い区間と判定する。その後、リスト生成部 204B は、ステップ ST 85 において、処理を終了して戻る。

30

【0117】

ステップ ST 83 で辞書に存在しないとき、リスト生成部 204B は、ステップ ST 86 において、処理対象の音声特徴区間が破裂音で始まる区間か否かを判定する。破裂音で始まる区間であるとき、リスト生成部 204B は、ステップ ST 84 において、処理対象の音声特徴区間は有用度の高い区間と判定する。その後、リスト生成部 204B は、ステップ ST 85 において、処理を終了して戻る。

【0118】

ステップ ST 82 で処理対象の音声特徴区間が母音の長音区間であるとき、あるいはステップ ST 86 で処理対象の音声特徴区間が破裂音で始まる区間でないとき、リスト生成部 204B は、ステップ ST 87 において、近傍の動き情報では有用か否かを判定する。例えば、リスト生成部 204B は、近傍にフォーカス位置移動、あるいは撮影方向変化の動き特徴区間が存在するとき、有用であると判定する。

40

【0119】

有用と判定するとき、リスト生成部 204B は、ステップ ST 84 において、処理対象の音声特徴区間は有用度の高い区間と判定する。その後、リスト生成部 204B は、ステップ ST 85 において、処理を終了して戻る。一方、有用でないとき判定するとき、リスト生成部 204B は、ステップ ST 87 において、処理対象の音声特徴区間は有用度の低い区間と判定する。その後、リスト生成部 204B は、ステップ ST 85 において、処理を

50

終了する。

【 0 1 2 0 】

図 2 0 に戻って、タイムライン生成部 2 0 5 は、動画読み込み部 2 0 1 で読み込まれた、編集に係る所定のチャプタの映像データおよび音声データのタイムラインを作成する。このタイムラインの作成は、表示制御部 2 0 6 の制御のもとで行われる。表示制御部 2 0 6 は、上述の解析情報リストに登録されるサブチャプタ登録に基づいて、有用度の高い区間と有用度の低い区間とが識別可能な状態でタイムラインが作成されるように、タイムライン生成部 2 0 5 の動作を制御する。

【 0 1 2 1 】

テキスト情報取得部 2 0 7 は、タイムラインに対応させて表示する文字列データを、解析情報リストに基づいて取得する。このテキスト情報の取得は、表示制御部 2 0 6 の制御のもとで行われる。表示制御部 2 0 6 は、ユーザによるタイムラインの所定サブチャプタ区間の選択情報に基づいて、その所定サブチャプタ区間およびその前後のサブチャプタ区間の音声に対応した文字列データが取得されるように、テキスト情報取得部 2 0 7 の動作を制御する。なお、ユーザは、所定サブチャプタ区間の選択を、図示しないマウス等のポインティングデバイスを操作することで行う。

10

【 0 1 2 2 】

表示部 2 0 8 は、LCD (Liquid Crystal Display) 等のディスプレイを用いて構成されており、編集画面を表示する。この表示部 2 0 8 は、表示制御部 2 0 6 の制御のもと、タイムライン生成部 2 0 5 で生成されたタイムラインを表示し、さらに、テキスト情報取得部 2 0 7 で取得された文字列を表示する。図 2 0 に示す編集装置 2 0 B のその他は、図 6 に示す編集装置 2 0 と同様に構成され、同様に動作する。

20

【 0 1 2 3 】

「コンテンツ関連付け構造体による再生制御」

また、上述実施の形態においては、チャプタ毎に、映像データおよび音声データの多重化データに関連付けされているコンテンツ関連付け構造体に含まれる情報に基づいて、編集画面の表示制御を行う例を示した。しかし、コンテンツ関連付け構造体に基づいて、コンテンツのハイライト再生時における制御を行うことも考えられる。例えば、コンテンツ関連付け構造体に基づいて有用度の低い区間と判定された区間を、不要区間として予め排除した上で演出できる。そのため、より効率的にハイライトシーンを検出可能となる。

30

【 0 1 2 4 】

[ 本技術を適用したコンピュータの説明 ]

上述した情報処理装置、編集装置における一連の処理は、ハードウェアにより行うこともでき、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で行う場合には、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。

【 0 1 2 5 】

図 2 2 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構成例を示している。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記憶部 3 0 8 や ROM (Read Only Memory) 3 0 2 に予め記録しておくことができる。

40

【 0 1 2 6 】

また、プログラムは、リムーバブルメディア 3 1 1 に格納 (記録) しておくことができる。このようにリムーバブルメディア 3 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。ここで、リムーバブルメディア 3 1 1 としては、例えば、フレキシブルディスク、CD-ROM (Compact Disc Read Only Memory)、MO (MagnetoOptical) ディスク、DVD (Digital Versatile Disc)、磁気ディスク、半導体メモリなどがある。

【 0 1 2 7 】

なお、プログラムは、上述したようなリムーバブルメディア 3 1 1 からドライブ 3 1 0 を介してコンピュータにインストールする他、通信網や放送網を介して、コンピュータに

50

ダウンロードし、内蔵する記憶部 308 にインストールすることもできる。すなわち、プログラムは、例えば、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、あるいは、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することができる。

【0128】

コンピュータは、CPU (Central Processing Unit) 301 を内蔵しており、CPU 301 には、バス 304 を介して、入出力インタフェース 305 が接続されている。CPU 301 は、入出力インタフェース 305 を介して、ユーザによって、入力部 306 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 302 に格納されているプログラムを実行する。あるいは、CPU 301 は、記憶部 308 に格納されたプログラムを、RAM (Random Access Memory) 303 にロードして実行する。

10

【0129】

これにより、CPU 301 は、上述したフローチャートに従った処理、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 301 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 305 を介して、出力部 307 から出力、あるいは、通信部 309 から送信、さらには、記憶部 308 に記録等させる。なお、入力部 306 は、キーボードや、マウス、マイク等で構成される。また、出力部 307 は、LCD (Liquid Crystal Display) やスピーカ等で構成される。

【0130】

ここで、本明細書において、コンピュータがプログラムに従って行う処理は、必ずしもフローチャートとして記載された順序に従って時系列に行われる必要はない。すなわち、コンピュータがプログラムに従って行う処理は、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り処理) も含む。また、プログラムは、一のコンピュータ (プロセッサ) 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よい。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よい。

20

【0131】

また、本技術は、以下のような構成を取ることにもできる。

(1) 音声を含むコンテンツの言語解析に基づいて、該コンテンツの編集ポイントを特定するための情報を取得する情報取得部と、

30

上記取得された情報を出力する情報出力部とを備える  
情報処理装置。

(2) 上記編集ポイントを特定するための情報は、上記編集ポイントの近傍期間における上記言語解析による言語情報を含む

前記(1)に記載の情報処理装置。

(3) 上記言語情報は、上記音声の発音情報を含む

前記(2)に記載の情報処理装置。

(4) 上記言語情報は、上記音声の文字列情報を含む

前記(2)または(3)に記載の情報処理装置。

(5) 上記情報出力部は、

上記編集ポイントの近傍期間における、上記コンテンツに含まれる映像の動き情報をさらに出力する

40

前記(1)から(4)のいずれかに記載の情報処理装置。

(6) 上記情報出力部は、

上記取得された情報を上記コンテンツと関連付けて記録メディアに記録する

前記(1)から(5)のいずれかに記載の情報処理装置。

(7) 上記情報出力部は、

上記取得された情報を上記コンテンツの送信単位に合わせてネットワークに送信する

前記(1)から(5)のいずれかに記載の情報処理装置。

(8) 音声を含むコンテンツの言語解析に基づいて、該コンテンツの編集ポイントを特

50

定するための情報を取得するステップと、

上記取得された情報を出力する情報出力ステップとを備える  
情報処理方法。

(9) 音声を含むコンテンツの言語解析に基づいて得られた、該コンテンツの編集ポイントを特定するための情報を取得する情報取得部と、

上記取得された情報に基づいて、上記コンテンツに関する表示を制御する表示制御部とを備える

表示制御装置。

(10) 上記編集ポイントを特定するための情報は、上記編集ポイントの近傍期間における上記言語解析による言語情報を含む

前記(9)に記載の表示制御装置。

(11) 上記表示制御部は、

上記編集ポイントを特定するための情報に基づいて、上記コンテンツの有用度の高い第1の区間と有用度の低い第2の区間を特定し、

上記コンテンツのタイムラインが上記第1の区間と上記第2の区間とが識別可能な状態で表示されるように制御する

請求項(9)または(10)に記載の表示制御装置。

(12) 上記表示制御部は、

上記タイムラインの所定区間の選択に応じて、上記言語情報に含まれる文字列情報を利用して、該所定区間の音声に対応する文字列が表示されるように制御する

前記(11)に記載の表示制御装置。

(13) 上記表示制御部は、

上記所定区間の音声に対応する文字列と共に、前後の区間の音声に対応した文字列も表示され、

上記所定区間の音声に対応する文字列がハイライト表示されるように制御する

前記(12)に記載の表示制御装置。

(14) 上記ハイライト表示は、他の区間とは表示色を異ならせることによるハイライト表示である

前記(13)に記載の表示制御装置。

(15) 上記表示制御部は、

上記言語情報に含まれる文字列情報で示される文節毎の区間に分割された上記コンテンツのタイムラインが表示されると共に、

該タイムラインの各区間に対応して、該各区間の音声に対応した文字列が表示されるように制御する

前記(10)に記載の表示制御装置。

(16) 音声を含むコンテンツの言語解析に基づいて得られた、該コンテンツの編集ポイントを特定するための情報を取得する情報取得ステップと、

上記取得された情報に基づいて、上記コンテンツに関する表示を制御する表示制御ステップとを備える

表示制御方法。

【符号の説明】

【0132】

10, 10A, 10B・・・情報処理装置

20, 20B・・・編集装置

101・・・映像入力部

102・・・音声入力部

103・・・動画多重化部

104・・・言語解析部

104a・・・音声特徴検出部

104b・・・テキスト変換部

10

20

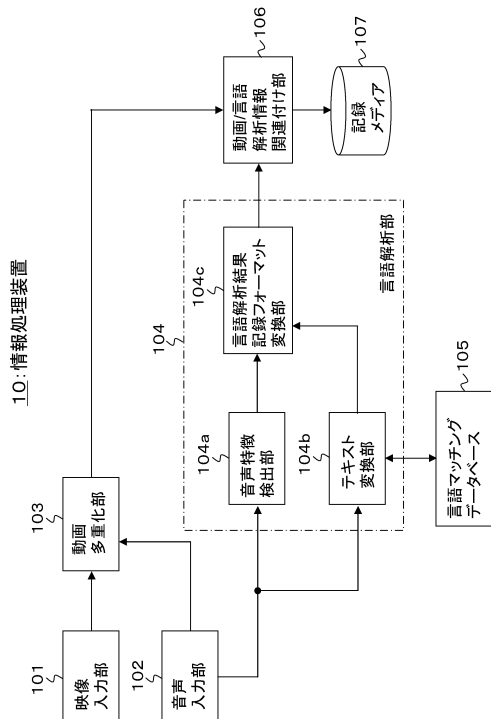
30

40

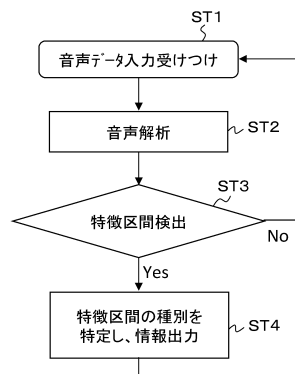
50

- 1 0 4 c , 1 0 4 cA , 1 0 4 cB . . . 言語解析結果記録フォーマット変換部
- 1 0 5 . . . 言語マッチングデータベース
- 1 0 6 . . . 動画 / 言語解析情報関連付け部
- 1 0 7 . . . 記録メディア
- 1 0 8 . . . ネットワーク
- 1 1 1 . . . 動き特徴検出部
- 1 1 2 . . . 動き解析結果記録フォーマット変換部
- 1 1 3 . . . 解析情報統合部
- 2 0 1 . . . 動画読み込み部
- 2 0 2 . . . 言語解析情報読み込み部
- 2 0 2 A . . . 解析情報読み込み部
- 2 0 3 . . . 動画 / 言語解析情報関連付け部
- 2 0 3 A . . . 動画 / 解析情報関連付け部
- 2 0 4 . . . 言語解析情報リスト生成部
- 2 0 4 A . . . 解析情報リスト生成部
- 2 0 5 . . . タイムライン生成部
- 2 0 6 . . . 表示制御部
- 2 0 7 . . . テキスト情報取得部
- 2 0 8 . . . 表示部

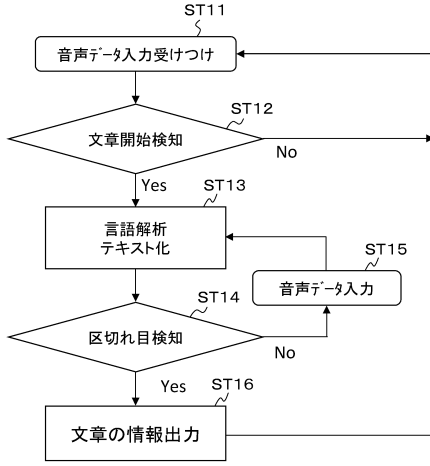
【 図 1 】



【 図 2 】



【図3】



【図4】

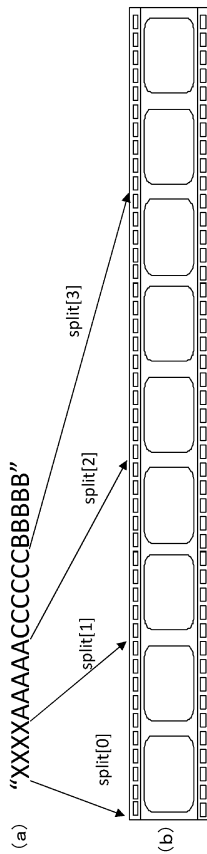
(a)コンテンツ関連付け構造体

```
struct コンテンツ関連付け構造体 {
  関連コンテンツを特定するID
  解析文字列のリスト
}
```

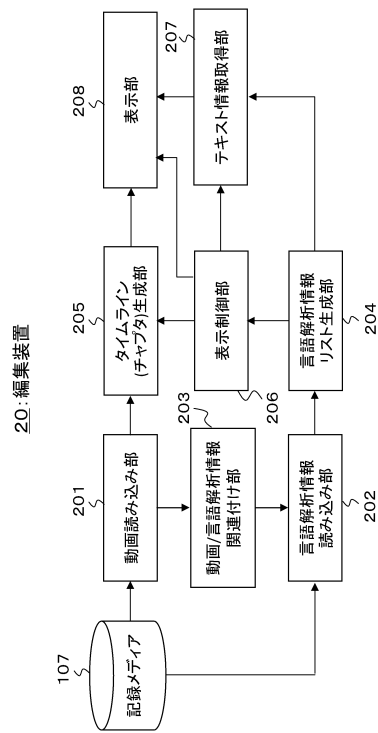
(b)文字列構造体

```
struct 文字列構造体 {
  "XXXXXXXXAAACCCCCBBB" // 文字列
  開始時間 // 文字列の発音が開始された時間
  終了時間 // 文字列の発音が終了した時間
  struct {
    開始文字位置 // 文字列中の区切れ目位置
    開始時間 // 区切れ目位置の時間位置
    発音情報 // 発音特徴の種類
  } split[n]
}
```

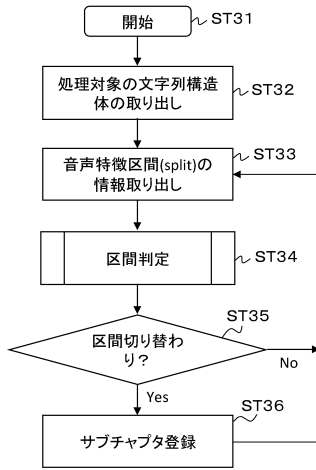
【図5】



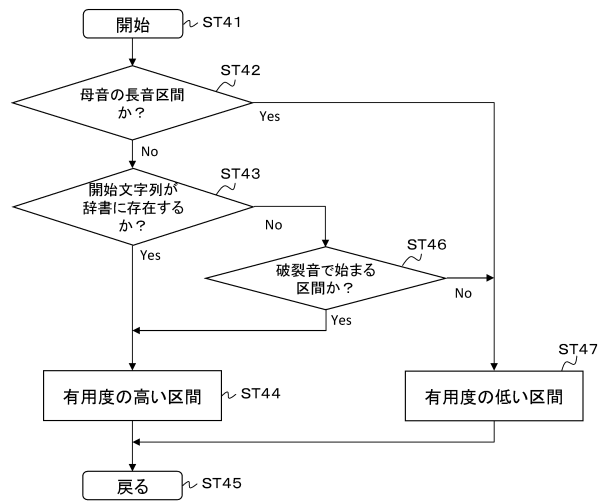
【図6】



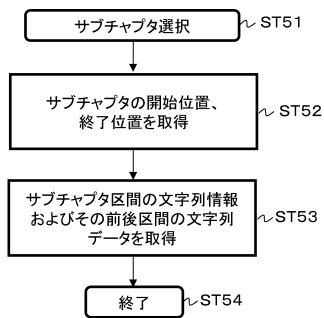
【図7】



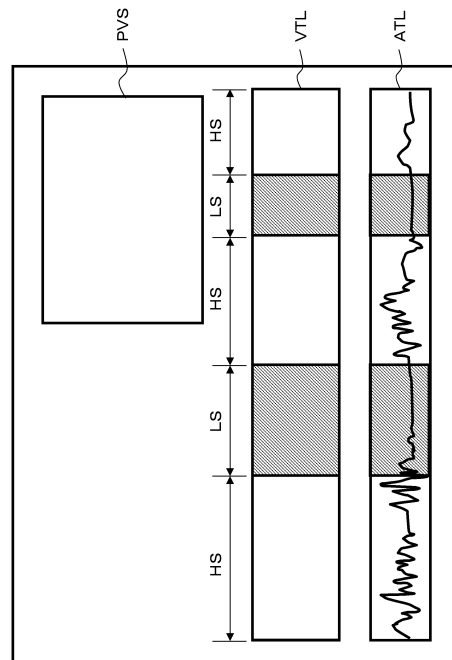
【図8】



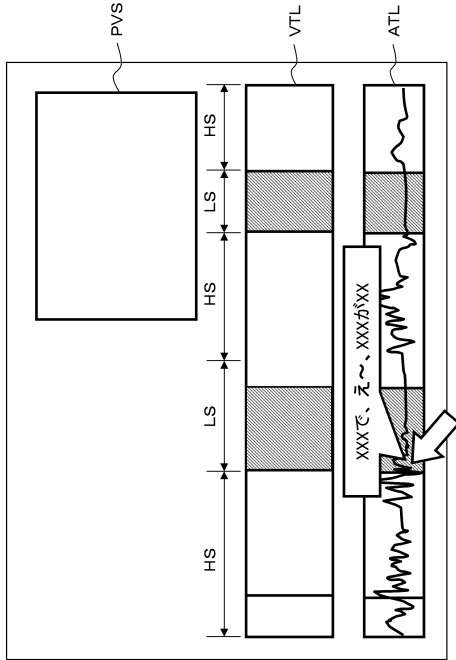
【図9】



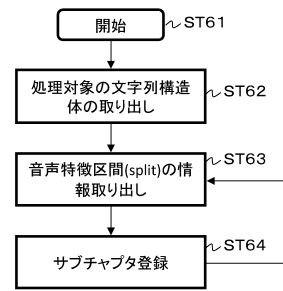
【図10】



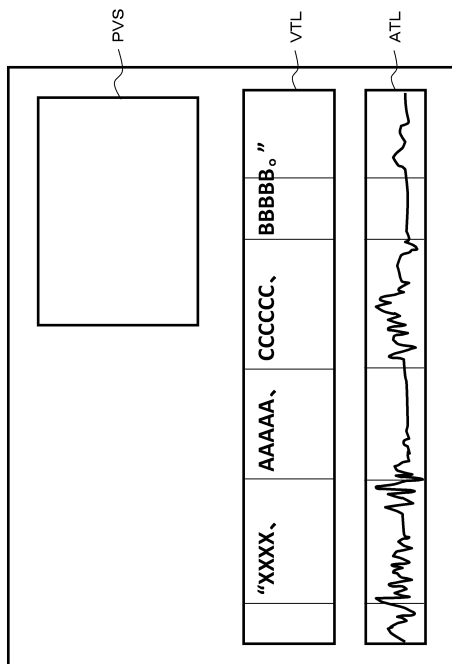
【図 1 1】



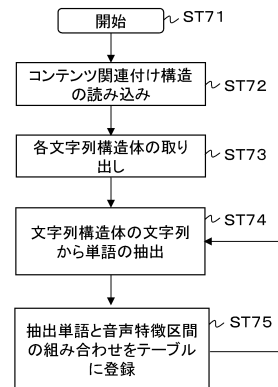
【図 1 2】



【図 1 3】

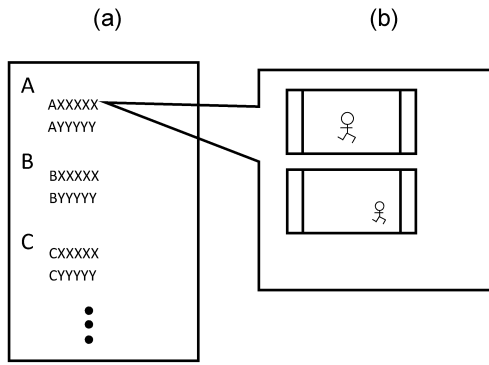


【図 1 4】

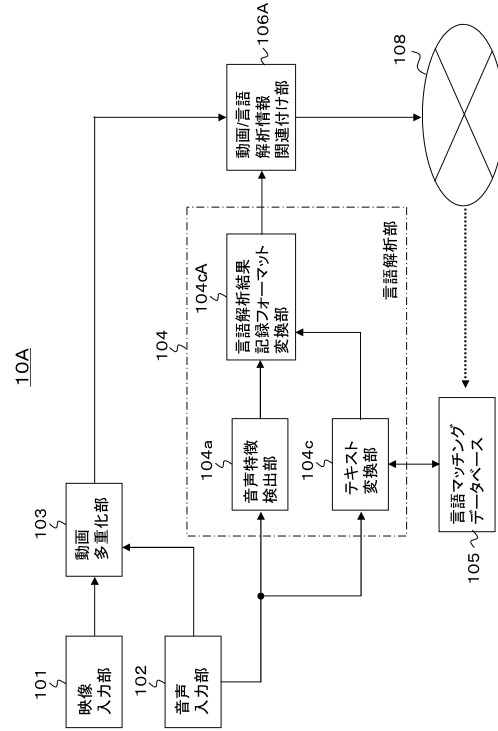




【図15】



【図16】



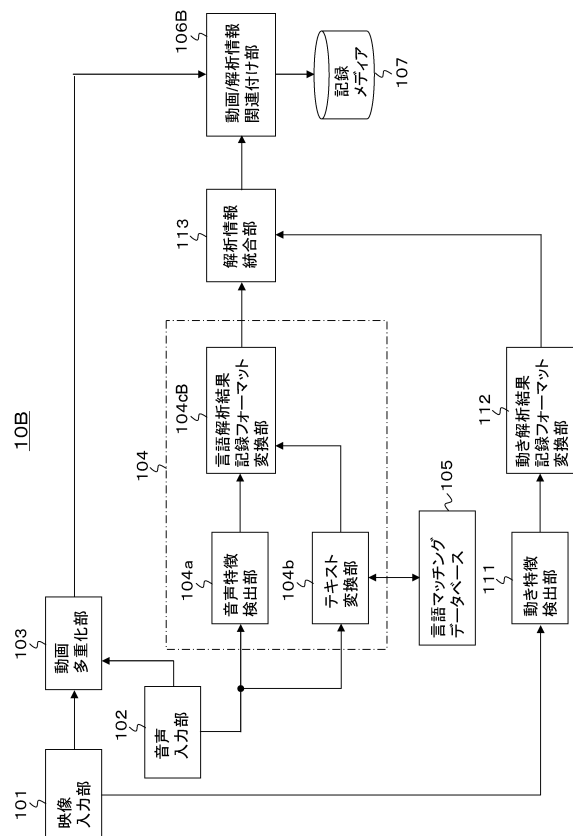
【図17】

```

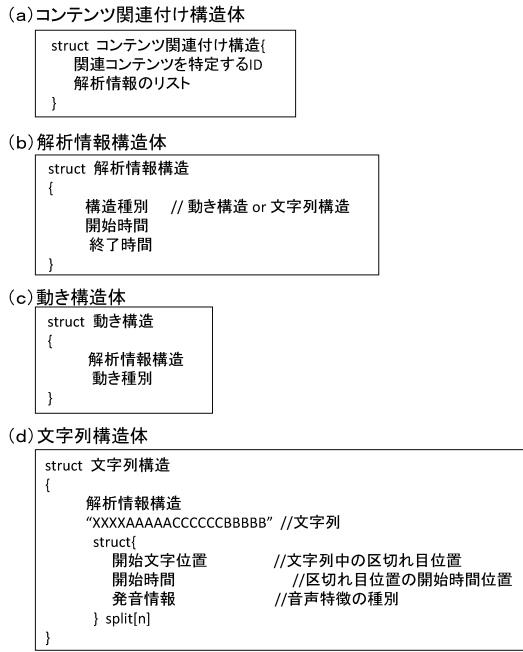
struct {
  関連コンテンツを特定するID
  参考テキスト //区間中に発声された言葉
  開始時間 //区切れ目位置の開始時間位置
  継続時間 //区切れ目位置の継続時間
  発音情報 //音声特徴の種類
}

```

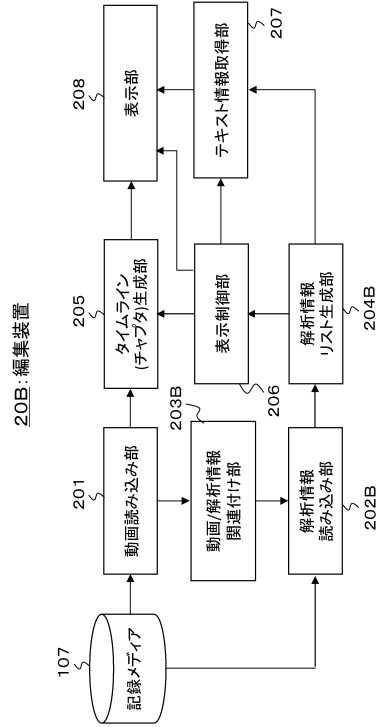
【図18】



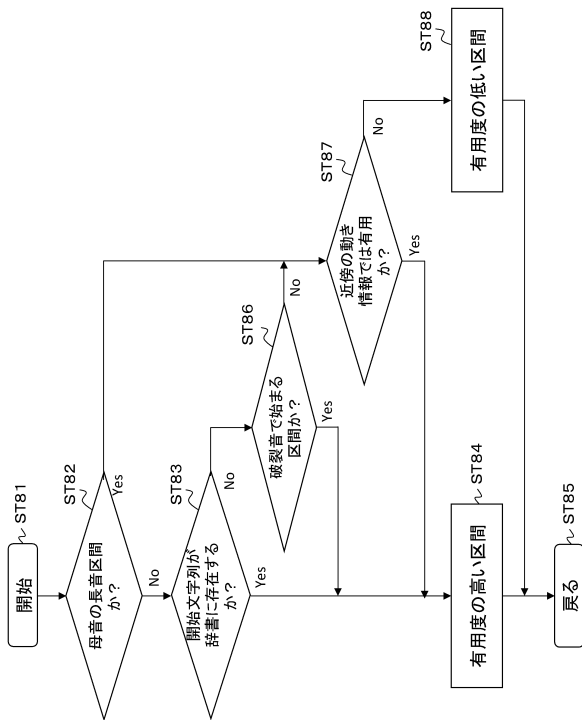
【図19】



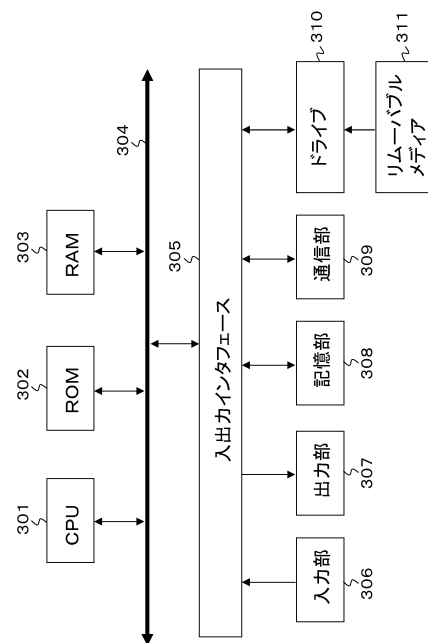
【図20】



【図21】



【図22】



---

フロントページの続き

(72)発明者 桑原 立  
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 堀 洋介

(56)参考文献 特開2004-363643(JP,A)  
特開2003-255979(JP,A)  
特開2012-084979(JP,A)  
特開2009-163644(JP,A)  
特開2005-333381(JP,A)  
特開2006-140707(JP,A)  
特開2003-224774(JP,A)  
特開2003-230094(JP,A)  
米国特許出願公開第2011/0317984(US,A1)

(58)調査した分野(Int.Cl., DB名)

H04N 5/91  
G11B 20/10  
G11B 27/02  
H04N 5/765