

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4765274号  
(P4765274)

(45) 発行日 平成23年9月7日(2011.9.7)

(24) 登録日 平成23年6月24日(2011.6.24)

(51) Int.Cl.		F I		
<b>G 1 0 L</b>	<b>13/08</b>	<b>(2006.01)</b>	<b>G 1 0 L</b>	<b>13/08</b> <b>1 1 0 Z</b>
<b>G 1 0 L</b>	<b>13/00</b>	<b>(2006.01)</b>	<b>G 1 0 L</b>	<b>13/08</b> <b>1 2 4</b>
			<b>G 1 0 L</b>	<b>13/08</b> <b>1 3 2</b>
			<b>G 1 0 L</b>	<b>13/00</b> <b>1 0 0 B</b>

請求項の数 6 (全 28 頁)

(21) 出願番号	特願2004-202320 (P2004-202320)	(73) 特許権者	000002185
(22) 出願日	平成16年7月8日(2004.7.8)		ソニー株式会社
(65) 公開番号	特開2006-23592 (P2006-23592A)		東京都港区港南1丁目7番1号
(43) 公開日	平成18年1月26日(2006.1.26)	(74) 代理人	100067736
審査請求日	平成19年7月6日(2007.7.6)		弁理士 小池 晃
		(74) 代理人	100086335
			弁理士 田村 榮一
		(74) 代理人	100096677
			弁理士 伊賀 誠司
		(72) 発明者	熊倉 恵理香
			東京都品川区北品川6丁目7番35号 ソ
			ニー株式会社内
		(72) 発明者	岡田 俊二
			東京都品川区北品川6丁目7番35号 ソ
			ニー株式会社内

最終頁に続く

(54) 【発明の名称】 音声合成装置及び音声合成方法

(57) 【特許請求の範囲】

【請求項1】

画像又は音響データに付随する情報を読み出す情報読み出し手段と、

上記情報読み出し手段によって読み出された情報に基づいて文を生成する文生成手段と、

上記文生成手段によって生成された文を音声合成して出力する音声合成手段とを有し、

上記文生成手段は、出力する文の内容を決定するために上記情報読み出し手段によって読み出された情報に基づく文字列を解析する文字列解析手段と、上記文字列解析手段にて解析された文字列から単語を抽出する単語抽出手段と、文を生成するために用いるテンプレートを記憶しているテンプレート記憶手段とを更に有し、

上記文生成手段は、上記単語抽出手段によって抽出された単語を含む文字列を、上記テンプレート記憶手段から選択したテンプレートが要求する文構造決定用のルールに当てはまるように変形し、上記変形した文字列を上記テンプレートに当てはめて文を生成する音声合成装置。

【請求項2】

概念辞書を更に有し、

上記文生成手段は、上記テンプレート内の概念を示す文字列に対して、上記単語抽出手段によって抽出した単語から上記概念を示す文字列に応じた単語を上記概念辞書から選出し、上記テンプレートに当てはめて文を生成する請求項1記載の音声合成装置。

【請求項3】

10

20

上記画像又は音響データに付随する情報は、ユーザ登録メモ、格納フォルダ名、又は再生頻度を含み、

上記音声合成手段は、上記画像又は音響データに付随する情報に応じて、合成音声の韻律、話速、音量又は声色を変更する請求項 2 記載の音声合成装置。

【請求項 4】

上記音声合成手段は、上記ユーザ登録メモを音声合成するときは柔らかなイントネーションに韻律を変更する請求項 3 記載の音声合成装置。

【請求項 5】

上記音声合成手段は、上記画像又は音響データが記録された時間に応じて、韻律、話速、音量又は声色を変更する請求項 3 記載の音声合成装置。

10

【請求項 6】

画像又は音響データに付随する情報を読み出す情報読み出し工程と、

上記情報読み出し工程によって読み出された情報に基づいて文を生成する文生成工程と、

上記文生成工程によって生成された文を音声合成して出力する音声合成工程とを有し、

上記文生成工程は、出力する文の内容を決定するために上記情報読み出し工程によって読み出された情報に基づく文字列を解析する文字列解析工程と、上記文字列解析工程にて解析された文字列から単語を抽出する単語抽出工程と、文を生成するために用いるテンプレートを記憶しているテンプレート記憶手段からテンプレートを選択するテンプレート選択工程とを更に有し、

20

上記文生成工程では、上記単語抽出工程によって抽出された単語を含む文字列を、上記テンプレート記憶手段から選択したテンプレートが要求する文構造決定用のルールに当てはまるように変形し、上記変形した文字列を上記テンプレートに当てはめて文を生成する音声合成方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は音声合成装置及び音声合成方法に関し、特に画像又は音響データに付随する情報に基づいて文を生成し、音声合成する音声合成装置及び音声合成方法に関する。

【背景技術】

30

【0002】

近年、デジタルスチルカメラ、携帯情報端末（PDA）等では、着脱型メモリであるリムーバブルメモリに画像、動画及び音楽ファイル等の画像ファイル及び画像に関連するファイルをデジタルカメラファイルフォーマットに基づいて記録し、上記機器間やプリンタ等の関連機器間で簡単にファイルを交換するようになってきた。

【0003】

例えば、デジタルスチルカメラで撮影し、リムーバブルメディアに上記デジタルカメラフォーマットに基づいて記録した画像ファイルに関連機器等で呼び出そうとする場合には、画像検索が行われる。また、デジタルスチルカメラで後日、目的の画像ファイルを呼び出そうとするときにも画像検索が行われる。

40

【0004】

一般的に、画像検索方法、画像検索装置においては、原画像に対応して画像情報検索のデータベースに登録された検索キーを使い、あるいは見出し検索用の縮小画像を配列表示し、フォーカス選択して、縮小画像と原画像とのリンク構造を使用して原画像を表示鑑賞することが以下に記す特許文献などにより様々な形態で行われている。

【0005】

特許文献 1 には、プレゼンテーション装置に使用する、縮小見出し画像と付随データの行列表示技術が開示されている。特許文献 2 には、電子画像の媒体記録再生装置に使用した場合の見出し画像を記録、再生利用する技術が開示されている。特許文献 3 には、電子画像表示装置で選択された縮小画像に枠フォーカス表示付けて選択状態表示を知らせる技

50

術が開示されている。特許文献4には、検索した画像データに追加属性情報を登録する技術が開示されている。特許文献5には、配列表示された縮小画像の一つを選択して、それを起点に後続の縮小画像集合を配列表示する技術が開示されている。特許文献6には、配列表示された見出し縮小画像で画像検索する静止画動画の画像検索装置があるシステム例の開示がある。特許文献7には、情報サービス受信装置および情報サービス受信方法が開示されている。特許文献8には、「デジタルカメラ」メモ리카ードの識別名をつけるという技術が開示されている。また、非特許文献1には、データベースの検索のために必要な複数の項目およびデータを日常会話に近い自然語文表現でデータベースを検索できれば操作性が非常に向上することが開示されている。

【0006】

【特許文献1】特開昭61-235981号公報

【特許文献2】特開昭62-248389号公報

【特許文献3】特許第2530612号公報

【特許文献4】特許第03192613号公報

【特許文献5】特開平02-90876号公報

【特許文献6】特許第3081304号公報

【特許文献7】特開平10-91499号公報

【特許文献8】特開2001-169225号公報

【非特許文献1】「日立評論」VOL.69, No.3(1987-3)の第23~27ページ「コンピュータのための自然語インターフェース」

【発明の開示】

【発明が解決しようとする課題】

【0007】

しかしながら、前述の如く、小型化する携帯型画像表示装置において検索画像を閲覧する場合には、装置付属の表示画面サイズがますます小型化され、さらには見出し画像を多数配置するとその微細な情報を認識判別することが非常に困難になってきている。また、登録キーを表示しようにも、表示画面自体が小さいので文字も大きくできず、文字数も多数を同時に表示できず、求める画像を的確に選択するのに困難なものになってきていた。

【0008】

また、小型携帯型画像表示機器は、その記録内容のパーソナルコンピュータPCとの流通性から、小型携帯型画像表示機器専用のメディアファイルフォーマットを使用している。

【0009】

本発明は、上記実情に鑑みてなされたものであり、画像又は音響データに付随する情報である、ユーザ登録メモ、格納フォルダ名、記録時の情報、再生頻度、又は好感度情報を読み出し、読み出した情報に基づいて文を生成し、生成した文を音声合成して音声として出力することにより、ユーザフレンドリに音声合成によって画像の検索を可能とする音声合成装置及び音声合成方法の提供を目的とする。

【課題を解決するための手段】

【0010】

本発明に係る音声合成装置は、上記課題を解決するために、画像又は音響データに付随する情報を読み出す情報読み出し手段と、上記情報読み出し手段によって読み出された情報に基づいて文を生成する文生成手段と、上記文生成手段によって生成された文を音声合成して出力する音声合成手段とを有し、上記文生成手段は、出力する文の内容を決定するために上記情報読み出し手段によって読み出された情報に基づく文字列を解析する文字列解析手段と、上記文字列解析手段にて解析された文字列から単語を抽出する単語抽出手段と、文を生成するために用いるテンプレートを記憶しているテンプレート記憶手段とを更に有し、上記文生成手段は、上記単語抽出手段によって抽出された単語を含む文字列を、上記テンプレート記憶手段から選択したテンプレートが要求する文構造決定用のルールに当てはまるように変形し、上記変形した文字列を上記テンプレートに当てはめて文を生成

10

20

30

40

50

する。

【 0 0 1 1 】

画像又は音響データから情報読み出し部が必要な情報を抽出し、その情報を元に文生成部がガイド文を生成し、生成されたガイド文をテキスト音声合成部が音声合成し、音声を出力する。

【 0 0 1 2 】

本発明に係る音声合成方法は、上記課題を解決するために、画像又は音響データに付随する情報を読み出す情報読み出し工程と、上記情報読み出し工程によって読み出された情報に基づいて文を生成する文生成工程と、上記文生成工程によって生成された文を音声合成して出力する音声合成工程とを有し、上記文生成工程は、出力する文の内容を決定するために上記情報読み出し工程によって読み出された情報に基づく文字列を解析する文字列解析工程と、上記文字列解析工程にて解析された文字列から単語を抽出する単語抽出工程と、文を生成するために用いるテンプレートを記憶しているテンプレート記憶手段からテンプレートを選択するテンプレート選択工程とを更に有し、上記文生成工程では、上記単語抽出工程によって抽出された単語を含む文字列を、上記テンプレート記憶手段から選択したテンプレートが要求する文構造決定用のルールに当てはまるように変形し、上記変形した文字列を上記テンプレートに当てはめて文を生成する。

10

【 発明の効果 】

【 0 0 1 3 】

本発明の音声合成装置及び音声合成方法は、画像又は音響データに付随する情報である、ユーザ登録メモ、格納フォルダ名、記録時の情報、再生頻度、又は好感度情報を読み出し、読み出した情報に基づいて文を生成し、生成した文を音声合成して音声として出力することができる。このため、本発明の音声合成装置を携帯型の画像音声記録再生装置に適用することで、音声によって案内される情報によって、画像音声記録再生装置によって記録/再生された画像を検索することができる。

20

【 0 0 1 4 】

本発明の音声合成装置及び音声合成方法を携帯型の画像音声記録再生装置に適用すると、以下のような効果も得られる。先ず、画像説明または楽曲ファイル対応説明のための説明音声のオーディオデータファイルを別途記録する必要がなくなり、撮影枚数が数桁以上、例えば数千枚以上に増加しても対応する説明音声オーディオファイルを記録することなくテキスト音声合成によるファイル説明のガイド文音声出力が可能になる。

30

【 0 0 1 5 】

また、書き込み媒体が光ディスクのようなメディア上のシークに時間のかかる場合に、対応する説明音声のオーディオファイルをファイルシステムで検索する制御が不要となり、ファイル数が増大していても説明音声へのアクセス時間が増加しない。

【 0 0 1 6 】

また、選択表示が縮小画像に付加表示処理された状態で、この選択された縮小画像と対応する画像付加情報コードがテキスト音声合成インターフェース手段に出力されるので、携帯型機器の表示画面がさらに縮小されて縮小画像の内容が認識しにくい場合でも個別の縮小画像が選択されることによって、その選択された縮小画像と対応する画像付加情報が説明音声として認識できる。

40

【 0 0 1 7 】

また、別途音声説明オーディオファイルの記録必要がなくなるので、書き込み回数の減少により、書き込み回数に上制限のあるメモリや一括して書き込み消去をおこなう必要のある、たとえばフラッシュメモリを使用するメモリカードの場合、メディアがさらに長く使用できることとなる。

【 0 0 1 8 】

また、音声合成によるので、文字表示することにより小さな画面を隠したり表示妨害することなく、情報を省略する必要なく音声にて聞いて把握することができる。

【 0 0 1 9 】

50

複数言語にも対応できる。複数ユーザ年齢にも対応できる。複数再生場所にも対応できる。複数記録場所にも対応できる。

【発明を実施するための最良の形態】

【0020】

以下、本発明を実施するための最良の形態を説明する。本実施の形態は、携帯型撮影画像メディアフォーマットとして広く使用されている、交換可能メディア画像ファイルフォーマット(Exchangeable Image File Formqt: Exif、イグジフ)、特にその中で静止画デジタルカメラに広く採用され普及している、DCF (Design Rule for Camera File)システムを用いて撮影画像や音声をリムーバブルメディアに記録すると共に上記リムーバブルメディアから撮影画像や音声を再生する携帯型の画像音声記録再生装置に適用される音声合成装置である。

10

【0021】

音声合成装置は、携帯型画像音声記録再生装置が上記DCFシステムにしたがってリムーバブルメディアに記録した撮影メディアフォーマットに記載された情報を用いて文を生成し、生成した文を音声合成して出力する。ここでいう、撮影メディアフォーマットに記載された情報とは、画像又は音響データに付随する情報のことで、ユーザ登録メモ、画像データを格納する格納フォルダ名、記録時の情報、再生頻度、又は好感度情報がある。

【0022】

図1は音声合成装置1のブロック図である。この音声合成装置1は、画像又は音響データに付随する情報を読み出す情報読み出し部2と、情報読み出し部2によって読み出された情報に基づいて文を生成する文生成部3と、文生成部3によって生成された文を音声合成して出力するテキスト音声合成部4とによって構成される。

20

【0023】

図2は音声合成装置の全体的な処理手順を示すフローチャートである。画像又は音響データから情報読み出し部2が必要な情報を抽出し(ステップS1)、その情報を元に文生成部がガイド文を生成し(ステップS2)、生成されたガイド文をテキスト音声合成部が音声合成し(ステップS3)、音声を出力する。

【0024】

情報読み出し部2は、上述したように、例えば、ユーザ登録メモ、格納フォルダ名、記録時の情報、再生頻度、又は好感度情報からなる、画像又は音響データに付随する情報を読み出す。以下に、情報読み出し部2の上記付随する情報の読み出しについて説明する。図3は交換可能メディア画像ファイルフォーマットであるExifファイルの画像ファイル構成例である。図4はExifファイルのオーディオファイル構成例である。

30

【0025】

Exifファイルの画像ファイル構成例では、図3に示すように、SOI (Start Of Image)によりJPEGストリームの先頭が示される。SOIの直後にはAPP1 (Application marker segment 1)が挿入される。APP1にはExif情報が記載される。このExif情報については詳細を後述する。APP1に続いてJPEGテーブルが記載される。量子化テーブル、ハフマンテーブル、フレームヘッド、スキャンヘッドなどが含まれる。次に、JPEG基本原画像データが続く。そして、EOI (End Of Image)でJPEGデータの終端が示される。

40

【0026】

APP1に記載されるExif情報は、APP1マーカー、APP1のデータ長、Exifの識別コード(Exif IFD)及び付属情報本体から構成される。これら全てを含むAPP1の大きさは、JPEG規格により64kByteを越えない。付属情報は、最大二つのIFD (0thIFD、1stIFD)を記録できる。0thIFDには、圧縮されている画像(主画像)に関する付属情報を記録する。1stIFDには、サムネイル画像(JPEG圧縮基本サムネイルデータ)を記録する。

【0027】

Exif IFDは、Exif固有の付属情報を記録するためのタグの集まりである。Exifバージョンタグ、画像データ特性タグ、画像構造タグ、ユーザコメントタグ、関連ファイル情報タグ、日時タグ、撮影条件タグ、IFDポインタ互換性からなる。

50

## 【 0 0 2 8 】

バージョンタグは、本Exif規格での対応バージョンを示す。このフィールドが存在しなければ、本Exif規格に準拠していないと判断される。本Exif規格に準拠する場合には、4ByteのASCII " 0210 " を記録する。

## 【 0 0 2 9 】

画像データ特性タグは、色空間を示す情報である。本規格では、PCモニタによる観察条件から色空間を規定したsRGB(=1)を使用する。

## 【 0 0 3 0 】

画像構造タグは、実効画像幅 (PixelXDimension)、実効画像高さ (PixelYDimension)、各コンポーネントの意味 (ComponentsConfiguration)、画像圧縮モード (CompressedBitsPerPixel) からなる。実効画像幅は、圧縮データ固有の情報である。圧縮ファイルを記録する場合には、パディングデータの有無、リスタートマーカの有無に関わらず、有意な画像の幅をこの実効画像幅タグに必ず記録する。非圧縮ファイルでは書かない。実効画像高さも圧縮データ固有の情報である。圧縮ファイルを記録する場合には、パディングデータの有無、リスタートマーカの有無に関わらず、有意な画像の高さをこの実効画像高さタグに必ず記録する。非圧縮ファイルでは書かない。各コンポーネントの意味も圧縮データ固有の情報である。各コンポーネントのチャンネルを、第1コンポーネントから第4コンポーネントの順に示す。圧縮画像モードも圧縮データ固有の情報である。画像圧縮時に設定された圧縮モードを単位bit/pelで示す。

## 【 0 0 3 1 】

ユーザコメントタグは、メーカーノートと共にユーザ情報に関するタグを構成する。メーカーノートは、Exifライターのメーカーが個別の情報を記入するためのタグである。内容については各メーカーの運用に任せられている。ユーザコメントタグは、イメージディスクリプシオン以外に画像に対してExifユーザがキーワードやコメントをユーザ登録メモとして書き込むためのタグである。

## 【 0 0 3 2 】

ユーザコメントに書かれる文字コードを判別するために、文字コードタイプをタグのデータ領域先頭に8バイト固定で記入し、余った領域にはnullでパディングする。文字コードとしてはASCII、JIS、Unicode、Undefinedがあり、リファレンスとしてITU-TT.501A5、JISX0208-1990、UnicodeStandard、Undefinedがある。Exifのリーダーは、文字コードを判別する機能を必ず持つ。そして、リーダーの能力に応じて表示する。文字コードタイプに続けて、ユーザコメントデータ欄が設けられる。文字コードタイプで識別された文字コードにより、Exifユーザがキーワードやコメントをユーザ登録メモリとして書き込むところである。

## 【 0 0 3 3 】

関連ファイル情報タグは、画像データに関連する音声ファイル名を記録している。関連情報として、相手先であるExif音声ファイルのファイル名と拡張子を一つだけ記録する。Exif画像ファイルとExif音声ファイルの対応関係には、3通りある。一つのExif画像ファイルに対して一つのExif音声ファイルが対応する関係、複数のExif画像ファイルに対して一つのExif音声ファイルが対応する関係、一つのExif画像ファイルに対して複数のExif音声ファイルが対応する関係である。本タグを用いて画像ファイルと音声ファイルを関連づける場合には、必ず関連付けされた音声ファイル側にも関連情報を記載しなければならない。

## 【 0 0 3 4 】

日時に関するタグは、原画像の生成日時DateTimeOriginal、デジタルデータ作成日時DateTimeDigitized、DateTimeのサブセックSubsecTime、DateTimeOriginalのサブセックSubsecTimeOriginal、DateTimeDigitizedのサブセックSubsecTimeDigitizedである。

## 【 0 0 3 5 】

原画像の生成日時DateTimeOriginalは、原画像の生成された日付と時間を示す。デジタルスチルカメラでは、撮影された日付と時間を記載する。フォーマットは " YYYY:MM:DD H

10

20

30

40

50

H:MM:SS”となる。時間は24時間表示し、日付と時間の間に空白文字[20.H]を一つ埋める。文字列の長さは、NULLを含めて20バイトとする。記載が無いときは不明として扱う。

【0036】

デジタルデータ作成日時DateTimeDigitizedは、画像がデジタルデータ化された日付と時間を示す。例えば、デジタルスチルカメラにより撮影され、同時にファイルが記録される場合、原画像の生成日時DateTimeOriginalとデジタルデータ作成日時DateTimeDigitizeは同じ内容となる。フォーマットは”YYYY:MM:DD HH:MM:SS”となる。時間は24時間表示し、日付と時間の間に空白文字[20.H]を一つ埋める。文字列の長さは、NULLを含めて20バイトとする。記載が無いときは不明として扱う。

【0037】

DateTimeのサブセックSubsecTimeは、DateTimeタグに関連して時刻を小数点以下の秒単位まで記録するためのタグである。DateTimeOriginalのサブセックSubsecTimeOriginalは、原画像の生成日時DateTimeOriginalタグに関連して時刻を小数点以下の秒単位まで記録するためのタグである。DateTimeDigitizedのサブセックSubsecTimeDigitizedは、デジタルデータ作成日時DateTimeDigitizedタグに関連して時刻を小数点以下の秒単位まで記録するためのタグである。

【0038】

撮影条件に関するタグは、露出時間ExposureTime、シャッタースピードShutterSpeedValue、絞り値ApertureValue、輝度値BrightnessValue、露光補正值ExposureBiasValue、レンズ最小F値MaxApertureValue、被写体距離SubjectDistance、測光方式MeteringMode、光源LightSource、フラッシュFlash、レンズ焦点距離FocalLength、FナンバFNumber、露出プログラムExposureProgram、スペクトル感度SpectralSensitivity、ISOスピードレートISO Speed Ratings、光電変換関数OECF、フラッシュ強度FlashEnergy、空間周波数応答SpatialFrequencyResponse、焦点面の幅の解像度FocalPlaneXResolution、焦点面の高さの解像度FocalPlaneYResolution、焦点面解像度単位FocalPlaneResolutionUnit、被写体位置SubjectLocation、露出インデックスExposureIndex、センサ方式SensingMethod、ファイルソースFileSource、シーンタイプSceneType、CFAパターンCFAPatternがある。IFDポインタ互換性は、互換性IFDへのポインタを示す。

【0039】

Exifファイルのオーディオファイル構成例は、図4に示すようなデータ構造を持つ。オーディオファイルでは、ストレージ領域を区画するファイルフォーマットは「チャンク」という区画名称で各データ区画を呼んでいる。最初の“RIFF”というチャンクID(ckID)に続いてRIFFサイズ指定(ckSize)、その後“WAVE”タイプ指定があり、フォーマットチャンクでのMPEGタグ、ステレオの指定があつて、WAVEのMP3Audioが登録されていると判定されるオーディオ登録情報データがある。

【0040】

つづいて、ファクトチャンクがあつて、これに続き、付属情報サブチャンクとして、INFOリストチャンクとExifリストチャンクがある。そのあとに、チャンクID(ckID)としてASCII文字列“Data”(64617461)を開始マーカーとしてMPEG Layer3圧縮形式WAVEのオーディオストリーム本体データが続く。

【0041】

INFOリストチャンク、Exifリストチャンク、データチャンクの各チャンクについて説明する。INFOリストチャンクには英数字で、タイトルの英数字、ジャンル英数字、ファイル作成年月日英数字、コメント英数字、アーティスト英数字、著作権英数字がある。

【0042】

Exifリストチャンクは、LISTマーカーから開始して、そのサイズ、“exif”とその規格バージョンが記載され、関連画像ファイルがある場合にはそのファイル名(ポインタは許可されない)が「ファイル名.拡張子」形式で直接記載され、その後、音声ファイル作成年月日として年:月:日:時:分:秒があつて、秒.サブ秒の記載方式でサブ秒まで記載される。

10

20

30

40

50

## 【 0 0 4 3 】

その後、ユーザコメントeucmのデータフィールドがある。このユーザコメントには、eucmのタグ文字につづき、eucmのチャンクサイズ、それにつづきチャンクデータである、ユーザコメントが記載される。このユーザコメントは先頭 8 バイトの文字コードのASCII大文字でのタイプ宣言があり、望ましくはUnicode、日本のパソコン事情でShiftJISを使用する場合はUndefinedが指定される。このユーザコメント欄には、2 バイトコード文字でのユーザ文字としての、アーティストや曲名、ファイル作成日時、さび部分の歌詞などが文字列登録されている。これが本実施の形態ではユーザ登録メモ情報として使用される。

## 【 0 0 4 4 】

情報読み出し部 2 は、上記のような画像ファイルやオーディオファイルと音声合成に必要な情報との対応表を保持している。図 5 は情報読み出しのための対応表の具体例である。

## 【 0 0 4 5 】

合成に必要な情報のデータラベル、Exifファイルのどこを参照すればよいかのポインタ、Exifファイルにおける表現形式が対応付けされている。例えば、格納フォルダ名は、Exifリストチャンクから「文字列」を参照する。ファイル名は、Exifリストチャンクから「ファイル名・拡張子」を参照する。撮影時間は、Exifリストチャンクから「年：月：日：時：分：秒」を参照する。撮影の場所は、ユーザコメントeucmから「文字列」を参照する。再生頻度は、ユーザコメントeucmから「数字」を参照する。好感度情報は、ユーザコメントeucmから「文字列」を参照する。ユーザ登録メモは、ユーザコメントeucmから「文字列」を参照する。そして、情報読み出し部において、上記対応表を用いて、ファイルの上記場所から、本実施の形態のファイル属性データの各情報データが参照され、文字列に変換されて読み出される。

## 【 0 0 4 6 】

図 6 が情報読み出し部 2 によって読み出されたデータの例である。情報読み出し部 2 は、上記情報読み出しのための対応表を用いて、格納フォルダ名から“家族/犬/”を、ファイル名から“モモの散歩.jpg”を、撮影の時間から“2002/06/27”、“09:52:15”を、撮影の場所から“近所の公園”を参照し、各々文字列に変換して読み出す。また、再生頻度から0を、好感度情報から5を、ユーザ登録メモから“モモちゃんが初めて散歩に行った。”を参照し、各々文字列に変換して読み出す。情報読み出し部 2 によって読み出された、上記図 6 に示したデータは、文生成部 3 に送られる。

## 【 0 0 4 7 】

図 7 は文生成部 3 の構成図である。文生成部 3 は、出力内容を決定する出力内容決定部 5 と、出力内容決定部 5 で決定された出力内容に合う文を生成する文出力部 6 とから構成される。このような構成により、文生成部 3 は、図 6 に示したようなデータをもとに、出力内容決定部 5 で、出力言語、文の内容、文の数を決定し、それに合うテンプレートを選択し、文出力部 6 では上記テンプレートに基づいて文を生成する。

## 【 0 0 4 8 】

図 8 は出力内容決定部 5 の構成図である。出力内容決定部 5 は、図 6 に示したようなデータから得られる情報を元に、生成する文の内容を決定するために、文字列を解析する文字列解析部 7 a と、文字列解析部 7 a にて解析された文字列から単語を抽出する単語抽出部 7 b と、出力言語を決定する出力言語決定部 1 0 と、文内容を決定する文内容決定部 1 1 と、文の数を決定する文の数決定部 1 2 と、テンプレートデータベース 1 4 から好適なテンプレートを選択するテンプレート選択部 1 3 を備える。

## 【 0 0 4 9 】

文字列解析部 7 a は、情報読み出し部 2 が読み出した上記図 6 に示したデータに言語的な情報、例えば品詞情報を付与する。文字列は、情報読み出し部 2 において、単語に分割されているので、文字列解析部 7 a は、各々の単語をルールファイル 8、辞書ファイル 9 と照らし合わせることによって、言語的な情報を得ることができる。

10

20

30

40

50



## 【 0 0 5 0 】

単語抽出部 7 b は、文字列解析部 7 a にて解析された文字列から単語を抽出する。ここで、抽出された単語は、出力言語決定部 1 0、文内容決定部 1 1 及び文の数決定部 1 2 に渡される。

## 【 0 0 5 1 】

出力言語決定部 1 0 は、出力する言語の種類を決定するためのものであり、例えばユーザ指定、画像音声記録装置の起動時の言語モード、ユーザ登録メモの言語という順番で優先度を付ける。もちろん、出力言語決定部 1 0 は、言語判定のためのサブモジュールを有する。言語判定サブモジュールは、文字列を受け取り、その文字列が何語であるかを判定する。

10

## 【 0 0 5 2 】

言語判定は、他の自然言語処理アプリケーションで用いられている、どの言語判定手法を用いて良い。例えば、入力文字列の文字コード、字種によって判定することができるし、文字列の並びの特徴を言語判定のためのルールとしてもよい。また、内部に辞書データを持ち、入力文字列を辞書引きすることによって言語を判定することもできる。

## 【 0 0 5 3 】

文内容決定部 1 1 は、ガイド文としてどのような文を生成するかを決定する。ユーザ登録メモ、ファイル名、撮影時間、撮影場所のデータから、文字列解析部 7 a で解析された文字列中、或いは単語抽出部 7 b から得られた単語データを基に文生成を行う際、どの情報に関する文であるかを決定する。また、文内容決定部 1 1 は、再生情報、好感度情報、感性キー情報、アクセス頻度キー情報などをもとに、画像に対するユーザの捉え方に関するガイド文を生成するかどうかを決定する。ユーザの捉え方に関するガイド文を生成することを決定したときには、テキスト音声合成の制御を行なう。また、アプリケーション使用時間帯情報を基にユーザに対して挨拶文を作成するかどうかを決定するようにしてもよい。もちろん、ユーザ指定情報に基づいて文の内容を決定してもよい。

20

## 【 0 0 5 4 】

文の数決定部 1 2 は、ユーザ登録メモ、ファイル名、撮影時間、撮影場所のデータから、文字列解析部 7 a で解析された文字列中、或いは単語抽出部 7 b から得られた単語データを基に文の数を決定する。画像記録のモード情報がファイル名等から判明すれば、連続スライドショーモードであるか否かを判定でき、さらにExifIFD日付タグのサブセコンドタグから連続する画像が何枚であるがわかるので、その枚数に応じて、出力する文の分数を決定することができる。また、文の数決定部 1 2 は、好感度情報の数に合わせて、文を追加するようにしてもよい。

30

## 【 0 0 5 5 】

テンプレート選択部 1 3 は、文内容決定部 1 1 で決定された内容で、文の数決定部 1 2 で指定された文の数分の、出力言語決定部 1 0 からの指示に従った言語のテンプレートを選択する。作成する文がユーザ登録メモを基にした文であれば、テンプレート選択部 1 3 は、ユーザ登録メモ用のテンプレートを選択する。また、撮影年月日についての文を生成するのであれば、そのためのテンプレートを選択する。

## 【 0 0 5 6 】

また、テンプレート選択部 1 3 は、好感度キーによる好感度情報にしたがって、テンプレートを選擇することもできる。例えば、テンプレートデータベース 1 4 に「これは、<名詞>です。」、「これは、お気に入りの<名詞>です。」という 2 種類のデータが記録されていたとする。もし、好感度キーが、その画像に対する好感度がよいことを示しているならば、テンプレート選択部 1 3 は、「これはお気に入りの<名詞>です。」というテンプレートの方を選択する。好感度がよいことをしめしていなければ、他方のテンプレートを選択する。

40

## 【 0 0 5 7 】

また、テンプレート選択部 1 3 は、どんなユーザがアプリケーションを使用しているかの情報を基に、テンプレートを選択することができる。例えば、ユーザが 1 0 代の若者で

50

ある場合、テンプレート選択部 13 は、「これは、<名詞>だよ。」というようなくだけた口調のテンプレートを選択することもできる。

【0058】

図9は出力内容決定部5の処理手順を示すフローチャートである。先ず、概略的に説明すると、文字列解析部7aにより文字列を解析すると共に単語抽出部7bにより単語を抽出し(ステップS11)、出力言語決定部10により出力言語を決定し(ステップS12)、文内容決定部11により文内容を決定し(ステップS13)、文の数決定部12により文の数を決定し(ステップS14)、テンプレート選択部13によりテンプレートを選択する(ステップS15)。出力する文の数だけテンプレートを選択したか否かの判定(ステップS16)によりYESを判定するまで後、各テンプレートは文出力部6に送られる。文出力部6は、各テンプレートに基づいて文を生成する。

10

【0059】

以下には、出力内容決定部5の処理手順を詳細に説明する。文字列解析部7aはステップS1にて文字列を解析し、後段の処理に必要な情報を付加する。本実施の形態では、文字列解析部7aの解析内容を、入力文字列に対し、形態素境界と各形態素の品詞を求める形態素解析であると仮定する。このため、文字列解析部7aは、解析のためのルールファイル8と辞書ファイル9を用いる。

【0060】

ルールファイル8には、形態素の並びに関する規則(文法や接続マトリックス、形態素Nグラムなど)と辞書に登録されていない形態素に必要な情報を付与する規則とが書かれている。辞書ファイル9には、表記、品詞、読みなど、登録されている形態素に関する情報が納められている。

20

【0061】

形態素境界と品詞の判定は、最長一致法または分割数最小法のような経験則による判定手法を用いても良いし、文法的接続可能性や、統計的な言語モデルを使用しての解析を行っても良い。形態素解析の手法は本実施の形態においては特に特定されず、たとえば、フリーで公開されている形態素解析システム茶釜(<http://chasen.aist-nara.ac.jp/hiki/ChaSen/>)などをそのまま用いることが出来る。撮影時間や好感度情報など、値に数値を持つものは、図5のフォーマットにしたがって、数値が取り出される。

【0062】

文字列解析部7aによる文字列の解析の結果、図6のデータは図10のように解析結果が追加されて、出力される。すなわち、格納フォルダ名の“家族/犬/”には、“家族”-[名詞]、“犬”-[名詞]という解析結果が追加されて出力される。また、ファイル名の“モモの散歩.jpg”には、“モモ”-[名詞]、“の”-[格助詞]、“散歩”-[名詞]という解析結果が追加されて出力される。またファイル名拡張子は“jpg”と解析される。撮影の時間の“2002/06/27”は「2002」、「06」、「27」と、また“09:52:15”は「09」、「52」、「15」と解析されて数値が取り出される。また、撮影の場所の“近所の公園”には、“近所”-[名詞]、“の”-[格助詞]、“公園”-[名詞]という解析結果が追加されて、出力される。また、再生頻度“0”からは「0」が、好感度情報“5”からは「5」が取り出される。また、ユーザ登録メモの“モモちゃんが初めて散歩に行った。”には、“モモ”-[名詞]、“ちゃん”-[接尾辞]、“が”-[格助詞]、“初めて”-[副詞]、“散歩”-[名詞]、“に”-[格助詞]、“行”-[動詞]、“っ”-[語尾]、“た”-[助動詞]、“。”-[句点]という解析結果が追加されて出力される。

30

40

【0063】

また、ステップS11では、単語抽出部7bが図10の解析結果から単語を抽出する。ここで、単語抽出部7bによって抽出された単語は、テンプレートデータベース14からテンプレート選択部13によって選択されたテンプレートに当てはめられる。

【0064】

次に、ステップS12において出力言語決定部10により、何語の文を生成するかが決

50

定される。出力言語決定部 10 は、データ (図 10) のユーザ登録メモ欄に値が入っているかどうかをチェックする。ユーザ登録メモ欄に値が入っていれば、ユーザ登録メモの言語を出力言語とする。ユーザ登録メモ欄に値が入っていないときは、あらかじめデフォルト言語を決めておき、デフォルト言語を出力言語とする。本実施の形態では、デフォルト言語を日本語と仮定して、説明を進める。

**【 0065 】**

ユーザ登録メモがある場合、ユーザ登録メモが何語で書かれているかを判定する必要がある。本実施の形態にあって、出力言語決定部 10 は、言語ごとの辞書を複数保持しており、言語ごとにユーザ登録メモの形態素が何語辞書に載っていたかをカウントし、最も単語数が多かった辞書の言語を出力言語に決定する。

10

**【 0066 】**

次に、ステップ S 13 においてどのような内容の文を生成するかを文内容決定部 11 が決定する。文内容決定部 11 には、単語抽出部 7b から得られたデータのうち、どれをもとに文を生成するか、また、どれをもとにテキスト音声合成の際の制御情報に用いるかの情報を持たせておく。本実施の形態では、ユーザ登録メモ、ファイル名、撮影時間、撮影場所のデータをもとに文生成を行い、どの情報に関する文であるかや、再生情報、好感度情報をもとにテキスト音声合成の制御を行なうとする。

**【 0067 】**

文内容決定部 11 では、データのラベル名 (格納フォルダ名、ファイル名、ファイル名拡張子...) ごとにその値の有無をチェックする。文生成に用いられるデータに値がある場合、その情報と実際の値をテンプレート選択部 13 に出力する。

20

**【 0068 】**

次に、ステップ S 14 において、文の数決定部 12 が内容ごとに文の数を決定する。デフォルト文数は 1 にしておき、文の数決定ルールにマッチする場合のみ、出力する文の数を変更する。文の数決定ルールとしては、ユーザ登録メモはその文数の文を出力する、好感度情報が 5 以上の場合にコメント文を 1 文追加する、画像が連写モードで撮影された場合に連写の数の文を生成する、などが挙げられる。このルールはアプリケーション設計者が自由に決めてよい。

**【 0069 】**

本実施の形態のデータでは、ユーザ登録メモは 1 つの文のみから構成されているので、ユーザ登録メモに関する文は 1 文だが、好感度情報が 5 なので、コメント文が 1 文追加される。

30

**【 0070 】**

次に、ステップ S 15 においてテンプレート選択部 13 が、文内容決定部 11 で決定された内容に従い、文の数決定部 12 で指定された文の数のテンプレートを選択する。テンプレートの言語は、出力言語決定部 10 からの指示に従う。

**【 0071 】**

図 11 は本実施の形態で用いるテンプレートの具体例である。本実施の形態では、入力データに、画像ファイルであることを示す拡張子 jpg、撮影時間、撮影場所が記載されているので、テンプレート選択部 13 において、テンプレートデータベース 14 から「この画像は<年>の<月日>に<場所>で撮影しました。」というテンプレートが選ばれる。次に、ユーザ登録メモがあるので、「<ユーザ登録メモ>」というテンプレートが選ばれる。さらに、コメント文が 1 文追加されることになっており、好感度が高いので、コメント文のうち、ポジティブなものが選択され、「素敵な画像ですね。」が選ばれる。<>で囲んである部分は、後段の処理でデータから値が代入される。<>の中に "," で区切って複数の項目が記載されている場合は、それらはテンプレート適用の制約条件となり、すべての項目を満たしていなければ、そのテンプレートを適用することは出来ない。

40

**【 0072 】**

出力内容決定部 5 は、テンプレート選択部 13 にて選択されたテンプレートを文出力部 6 に出力する。

50

## 【 0 0 7 3 】

次に文出力部 6 の説明をする。図 1 2 は文出力部 6 の構成図である。文出力部 6 は、出力内容決定部 5 からの入力文字列である単語列を変形する単語変形部 1 5 と、入力文の構造を判定する文構造判定部 1 7 と、上記図 1 0 に示したデータから必要な値を抽出してテンプレートに当てはめるテンプレート適用部 1 8 とから構成される。

## 【 0 0 7 4 】

単語変形部 1 5 は、出力内容決定部 5 から得られたテンプレートに、図 1 0 の出力されるデータから単語抽出部 7 b が抽出して得られた文字列を当てはめる際に、テンプレートに当てはまるように、入力文字列を変形する。このとき、単語変形部 1 5 は、概念辞書（シソーラス）ファイル 1 6 を参照して入力文字列を他の語に置き換えてもよい。つまり、単語変形部 1 5 は、単語抽出部 7 b によって抽出された単語を、テンプレートに当てはまるように変形するか、或いは概念辞書ファイル 1 6 を参照して得た他の語に置き換える。概念辞書は、単語間の概念的な関係を表す辞書である。文字列を同じ概念の語、または上位概念を持つ語などで置き換えるのに用いられる。

## 【 0 0 7 5 】

たとえば、「<ユーザ登録メモ>,<文>」というテンプレートがテンプレート選択部 1 3 によって選択されているとする。このテンプレートは、ユーザ登録メモが文という構造を持っていることを要求している。

## 【 0 0 7 6 】

そこで、文構造判定部 1 7 は、ユーザ登録メモが文であるかどうかをチェックする。文構造判定部 1 7 は、図 1 3 の文構造決定用のルールを有しており、入力がどのような構造を持っているかを調べることが出来る。本実施の形態のルールでは、入力形態素の並びがルールの右辺と合致すれば、左辺のシンボルで代入することができる。最後まで代入して、得られたシンボルが入力の構造となる。この場合、入力文は文と判定される。

## 【 0 0 7 7 】

図 1 3 において、入力形態素が“名詞”、或いは“名詞 接尾辞”であれば[名詞句1]で代入できる。また、“名詞 各助詞”、或いは“名詞 接尾辞 各助詞”であれば[名詞句2]で代入できる。また、“名詞句1”、“名詞句2”、“名詞句2 名詞句1”、“名詞句2 名詞句2”、“名詞句2 名詞句2 名詞句1”、“名詞句2 名詞句2 名詞句2”であれば[名詞句]に代入できる。また、入力形態素が“動詞”、“動詞 語尾”、“動詞 助動詞”、“動詞 語尾 助動詞”であれば[動詞句]で代入できる。また、“副詞”であれば[修飾句]で代入できる。そして、“名詞句 動詞句”、“名詞句 助動詞”、“修飾句、名詞句 動詞句”、“名詞句 修飾句、動詞句”、“名詞句、修飾句 名詞句 動詞句”であれば[文]で代入できる。

## 【 0 0 7 8 】

もし、ユーザ登録メモが名詞句であると判定された場合、テンプレート「<ユーザ登録メモ>,<文>」は使うことが出来ない。ユーザ登録メモの文字列を変形する必要がある。たとえば、ユーザ登録メモの内容が、「モモちゃんの散歩の様子」であった場合、最後に「です。」をつけることによって、文にすることが出来、テンプレートを適用できるようになる。また、「これはお気に入りの<ファイル名>,<名詞句>です。」というテンプレートの場合、ファイル名の文字列「モモの散歩」は名詞句なので、単語変形部 1 5 は特に何もしない。

## 【 0 0 7 9 】

また、「これは、<“色”>の<“花”>です。」のようなテンプレートが選択されているとき、テンプレートの適用に概念辞書を用いる。上記テンプレートにおいて、“ ”に囲まれた文字列は、それが概念であることを示す。テンプレートのはじめの部分には、色を表す名詞を当てはめることができる。2 番目の部分には、花という概念を持つ語が入る。

## 【 0 0 8 0 】

例えば、ユーザ登録メモから得られた文字列が、「赤」と、「チューリップ」であった

と仮定する。出力内容決定部 5 の文字列解析部 7 a で、「赤」と、「チューリップ」が共に名詞であるという情報が得られる。また、概念辞書ファイル 16 を用いると、赤の上位概念が色であり、チューリップの上位概念が花であることが判明する。

【 0 0 8 1 】

これらの解析結果を用いることで、「赤」、「チューリップ」という文字列を上記テンプレートに当てはめることができることが判定される。その結果、「これは赤のチューリップです。」という文が生成される。

【 0 0 8 2 】

テンプレート適用部 18 は、テンプレート選択部 13 によって選択されたテンプレートを用いて文を生成する。具体的に、テンプレート適用部 18 は、図 10 に示したデータから、必要な値を抽出し、あるいは単語変形部 25 により、変形又は置き換えた語をテンプレートに当てはめる。「この画像は<年>の<月日>に<場所>で撮影しました。」というテンプレートの場合、データから撮影時間の年月日と、撮影場所を抽出し、テンプレートに当てはめて、「この画像は、2002年の6月27日に近所の公園で撮影しました。」という文を得る。

10

【 0 0 8 3 】

次にテキスト音声合成部 4 の説明をする。テキスト音声合成部 4 は、文生成部 3 にて生成されたテキストを音声波形に変換するためのモジュールである。図 14 は、テキスト音声合成部 4 の構成を示したものである。テキスト音声合成部 4 は、入力されたテキストを音声記号列に変換するための言語処理部 21 と、言語処理部 21 からの音声記号列を韻律データに変換するための韻律生成部 24 と、韻律生成部 24 からの韻律データを元に音声波形を生成するための波形生成部 27 とによって構成されている。

20

【 0 0 8 4 】

言語処理部 21 は、辞書ファイル 22 と、ルールファイル 23 とを有し、入力されたテキストを音声記号列に変換する。韻律生成部 24 は、ルールファイル 25 と、データファイル 26 とを有し、言語処理部 21 からの音声信号列を韻律データに変換する。波形生成部 27 は、素片ファイル 28 を有し、韻律生成部 24 からの韻律データを基に音声波形を生成する。

【 0 0 8 5 】

概略的に動作を説明すると、テキスト音声合成部 4 は、入力されたテキストを言語処理部 21 により辞書ファイル 22、ルールファイル 23 を参照して音声信号列に変換し、韻律生成部 24 によりルールファイル 25、データファイル 26 を参照して上記音声信号列を韻律データに変換し、波形生成部 27 により上記韻律データを基に素片データベース 28 を参照して音声波形を生成する。

30

【 0 0 8 6 】

以下、言語処理部 21、韻律生成部 24 及び波形生成部 27 の詳細な構成及び動作について図 15 ~ 図 19 を参照しながら説明する。

【 0 0 8 7 】

図 15 は、言語処理部 21 の構成を示したものである。言語処理部 21 は、入力されたテキストを音声記号列に変更するためのものであり、入力されたテキストを解析し、それぞれの形態素の情報を得るためのテキスト解析部 29 と、形態素情報を元に音声記号を生成するための韻律情報生成部 30 とからなる。

40

【 0 0 8 8 】

テキスト解析部 29 は、ルールファイル 30 と、辞書ファイル 31 に接続している。ルールファイル 30 には、形態素の並びに関する規則(文法や接続マトリックス、形態素Nグラムなど)と辞書に登録されていない形態素に必要な情報を付与する規則とが書かれている。辞書ファイル 31 には、表記、品詞、読みなど、登録されている形態素に関する情報が納められている。

【 0 0 8 9 】

テキスト音声合成部 4 に入力されたテキストは、言語処理部 21 にあるテキスト解析部

50

29に渡される。入力されるテキストは1文の場合も、複数文からなる場合もありうるが、本実施の形態では、説明を容易にするため1文として、話を進める。

【0090】

入力された文は、テキスト解析部29において、ルールファイル30と辞書ファイル31とを用いて形態素に分割され、辞書ファイル31からそれぞれの形態素に関する情報を得る。辞書ファイル31に登録されていない形態素については、ルールファイル30を用いて、必要な形態素情報を生成する。テキスト音声合成部4の言語処理部21内のテキスト解析部29は、文生成部3の出力内容決定部5の文字列解析部7と同じものを用いることが出来る。このようにして得られた形態素情報は、韻律情報生成部32へと送られる。

【0091】

韻律情報生成部32は、形態素情報を解析し韻律情報を得るためのルールファイル33を有する。ルールファイル33には、形態素情報をもとにフレーズングや読みの変更を行なうための規則が納められている。また、入力された各形態素について、韻律生成のためのさらに詳しい情報を得るため、テキスト解析部29で用いた辞書ファイル31とも接続されている。韻律生成のための情報とは、例えば、複数の形態素が一緒になって複合語を形成する際に、読みがどのように変化するかの情報や、アクセント核位置の移動パターンの情報などである。

【0092】

本実施の形態では、形態素に関する情報も韻律生成のための情報も1つの辞書ファイル31に納めるといった構成を取っているが、それぞれを別の辞書ファイルとして保持し、テキスト解析部29には形態素に関する情報のみが納められた辞書ファイルが接続されており、韻律情報生成部32には韻律生成のための情報のみが納められた辞書ファイルが接続されているという構成で同等の処理を行なうことができる。

【0093】

韻律情報生成部32では、テキスト解析部29から受け取った形態素情報をもとに、入力文に対し、ルールを用いて韻律情報を付与するという処理を行なう。韻律情報とは、言語によって異なるが、日本語であれば、例えば、アクセント核位置、アクセントの強さ、ポーズ位置、ポーズの長さ、読みの変化などが挙げられる。韻律情報は、装置の設計によって異なるので、上記情報を全て含む必要はなく、上記情報に限定する必要もない。

【0094】

韻律情報生成部32での処理によって得られる韻律情報は、音声記号列で表現される。音声記号列とは、入力テキストを発音記号、アクセント記号、ポーズ記号、音調記号などを用いて表現したものである。発音記号やアクセント記号では、IPA(国際音声字母、International Phonetic Alphabet)やSAMPAなどが広く知られている。また、ポーズ記号や音調記号については、ToBI(Tone and Break Indices)が有名である。言語処理部21は、得られた音声記号列を韻律生成部24へと出力する。

【0095】

図16は韻律生成部24の構成を示したものである。韻律生成部24は、入力された音声記号列を韻律データに変換するためのものである。韻律生成部24は、各音素の継続時間長を求めるための継続時間長決定部34と、各音素の基本周波数を求めるための基本周波数決定部39とによって構成されている。

【0096】

継続時間長決定部34は、言語処理部21からの音声信号列を音素に関する解析データファイル36を用いて解析する音声信号列解析部35と、音声信号列解析部35にて解析されたそれぞれの音素の継続時間長を、継続時間長を決めるためのルールファイル38を用いて決定するルール適用部37とを有する。解析データファイル36に格納されている、音素に関する解析データとは、例えば、それぞれの音素が母音かどうかやそれぞれの音素のデフォルトの継続時間長などを指す。ルールファイル38に格納されている、継続時間長を決めるためのルールとは、どのような音韻環境の時にデフォルト継続時間長を伸縮させるかを定めるための規則のことである。音韻環境とは、その音素が母音であるかどうか

10

20

30

40

50

か、前後の音素が何であるか、音節の何番目の音素であるか、アクセント核位置がどうかなどを指す。

【 0 0 9 7 】

継続時間長決定部 3 4 は、音声信号列解析部 3 5 にて解析データファイル 3 6 を用いて言語処理部 2 1 からの音声信号列を解析し、ルール適用部 3 7 にてルールファイル 3 8 を用いてそれぞれの音素の継続時間長を決定する。つまり、ルール適用部 3 7 にてルールを適用するのに必要な情報は入力された音声記号列を音声信号列解析部 3 5 にて解析することで得られる。継続時間長決定部 3 4 にて決定された継続時間長の情報は基本周波数決定部 3 9 へと送られる。

【 0 0 9 8 】

基本周波数決定部 3 9 は、言語処理部 2 1 からの音声信号列と継続時間長決定部 3 4 にて決定された継続時間長の情報から韻律データを生成する。このため、基本周波数決定部 3 9 は、上記音声信号列の基本周波数パターンを、基本周波数パターンに関する解析データファイル 4 1 を用いて解析してルール適用に必要な情報を得る音声信号列解析部 4 0 と、音声信号列解析部 4 0 にて得られた情報を基に、パターン選択用ルールファイル 4 3 を用いて最適なパターンを選択するルール適用部 4 2 と、ルール適用部 4 2 にて決定された最適なパターンを継続時間長決定部 3 4 にて既に決定されたそれぞれの音素の継続時間長に合わせて変形して音素毎の基本周波数を決定する時間伸縮部 4 4 とを備える。

【 0 0 9 9 】

解析データの持ち方とルールデータの選択方法は装置の設計によって、様々な実現形態が考えられるが、本実施の形態では、その音素が属する単語のアクセント型や、アクセントの強さ、係り受けなどの文構造などを基準にルールの適用可能性是非を判定する手法を用いる。

【 0 1 0 0 】

上記基本周波数パターンのデータは、複数の音素にまたがる基本周波数の値を保持しておく。解析データファイルは、ルールで選択できるように、アクセント型や音調など予測要因毎にパターンが分類してある。

【 0 1 0 1 】

基本周波数決定部 3 9 は、入力された音声記号列を音声信号列解析部 4 0 により上記解析データファイル 4 1 を参照して解析し、ルール適用に必要な情報を得る。得られた情報を元に、ルール適用部 4 2 がパターン選択用ルールファイル 4 3 を参照して、最適なパターンを選択する。ルール適用部 4 2 にて選択されたパターンは、継続時間長決定部 3 4 ですでに決定されたそれぞれの音素の継続時間長にあわせて、時間伸縮部 4 4 にて変形され、音素毎の基本周波数が決定される。継続時間長決定部 3 4 で決定された音素毎の継続時間長と基本周波数決定部 3 9 で決定された音素毎の基本周波数は、韻律データとして韻律生成部 2 4 から、波形生成部 2 7 へ送られる。

【 0 1 0 2 】

韻律データの一部を図 1 7 及び図 1 8 に示す。韻律データの前半は図 1 7 の継続時間長 (duration) のデータ、後半は図 1 8 の基本周波数 (pitch) のデータである。図 1 7 において、継続時間長のデータは、一番左が音声開始時点からのサンプル数、次が音素列、一番右が音素ごとの継続時間長になっている。図 1 8 において基本周波数のデータは、左が音声開始時点からのサンプル数、右が基本周波数の値である。

【 0 1 0 3 】

図 1 9 は、波形生成部 2 7 の構成を示したものである。波形生成部 2 7 は、韻律生成部 2 4 の出力である韻律データから音声波形を生成する。このため、本実施の形態で用いられる波形生成部 2 7 は、入力された韻律データを音素列情報、音素時間長情報、ピッチパターン情報に分配する韻律データ分配部 5 1 と、入力された音素列情報に含まれる音素列を選択し、選択した音素列に対応する音響特長パラメータを音声データファイル 5 3 から順次読み出して出力する素片選択部 5 2 と、音素列によって時間軸上に並べられた音響特長パラメータを入力とし音素時間情報によってそれぞれの音素の長さに等しくなるように

10

20

30

40

50

パラメータを時間軸上で伸縮するパラメータ補正部 5 4 と、パラメータ補正部 5 4 から出力される音響特長パラメータの系列、また、韻律データ分配部 5 1 からのピッチパターン情報に基づいてピッチを変更しながら音声信号波形を合成して出力する波形組み立て部 5 5 とにより構成される。

【 0 1 0 4 】

素片選択部 5 2 が有している音声データファイル 5 3 に格納されている音声データとは、既存の規則音声合成装置に使用されるものと同様のもので、たとえばケプストラム係数などの音声の音響的な特長を示すパラメータなどである。また、素片のひとつひとつの長さは、CV、CVC、その他の合成単位に応じた単位、あるいはコーパススペースの合成方式では特に合成単位を固定せずに保存する場合もある。

10

【 0 1 0 5 】

この波形生成部 2 7 の概略的な動作は以下ようになる。入力された韻律データは、韻律データ分配部 5 1 によって音素列情報、音素時間長情報、ピッチパターン情報に分けられ出力される。音素列情報は素片選択部 5 2 に送られる。素片選択部 5 2 では、入力された音素列情報に基づいて音声データファイル 5 3 を参照して上記音素情報に含まれる音素列を選択して、選択した音素列に対応する音響特長パラメータを順次読み出して出力する。パラメータ補正部 5 4 は、音素列によって時間軸上に並べられた音響特長パラメータを入力とし、音素時間情報によってそれぞれの音素の長さに等しくなるようにパラメータを時間軸上で伸縮する。また、素片の接続部でのパラメータの不整合を避けるためにパラメータが滑らかに変化するように音響特長パラメータを補正する。また、パラメータを変更することにより、それぞれの音素の強さを変更する。波形組み立て部 5 5 は、パラメータ補正部 5 4 から出力される音響特長パラメータの系列に基づいて、また、韻律データ分配部 5 1 からのピッチパターン情報に基づいてピッチを変更しながら音声信号波形を合成して出力する。

20

【 0 1 0 6 】

このように本実施の形態の音声合成装置 1 は、Exifファイルの画像ファイルや、オーディオファイルから画像又は音響データに付随する情報である、ユーザ登録メモ、格納フォルダ名、記録時の情報、再生頻度、又は好感度情報を読み出し、読み出した情報に基づいて文を生成し、生成した文を音声合成して音声として出力することができる。このため、本発明の音声合成装置 1 を画像音声記録再生装置に適用すれば、音声によって案内される情報によって、画像音声記録再生装置によって記録/再生された画像を検索することができる。

30

【 0 1 0 7 】

なお、本発明の音声合成装置によれば、出力音声の話速、音量、声の高さ、声色、韻律的特徴を変更することができる。以下に、テキスト音声合成における出力音声変更の手法について説明する。

【 0 1 0 8 】

それぞれの変更は以下の手法にて行われる。

話速 = 継続時間長の伸縮

音量 = パラメータの補正、変更

声の高さ = 基本周波数の値を変更

声色 = 音声データの変更

韻律的特長 = 音素ごとの継続時間長、パラメータ、基本周波数の変更

韻律的特長とは、たとえば、はつらつとしたイントネーション、強調イントネーション、暗いイントネーションなど、音声を聞いたときの印象に関するものである。

40

【 0 1 0 9 】

はつらつとしたイントネーションは、高くて大き目の音で、少し速めに合成することによって実現することが出来る。強調イントネーションは、高くて大き目の音で、少しゆっくり目に合成することによって実現することが出来る。このように、合成する音の高さ、長さ、強さを変更することで、異なる韻律的特徴を実現することが可能となる。このよう

50



に、本発明では、画像又は音響データに付随する情報によって、出力される合成音の話速、音量、声の高さ、声色、韻律的特徴を変更する。

【0110】

図20は出力する合成音を変更可能な音声合成装置60の構成図である。出力する合成音を変更可能な音声合成装置60は、図1に構成を示した音声合成装置1に加え、音声合成制御部61を有する。

【0111】

音声合成制御部61は情報読み出し部2にて抽出された情報と、文生成にて生成された文とをもとに合成音をどのように変化させるかを決定し、テキスト音声合成部4をコントロールするためのモジュールである。

10

【0112】

また、音声合成制御部61は、効果音のデータベースを有し、合成音と同期させて、効果音を出力することも可能である。たとえば、音声合成制御部61には、文生成部3で生成された文が、ユーザ登録メモをもとに生成されたものである場合、柔らかなイントネーションで音声合成するというルールが保持されているとする。

【0113】

音声合成制御部61は、情報読み出し部2と文生成部3から得られる情報によって、テキスト音声合成部4が処理しようとしている文がユーザ登録メモから生成されたものであるかどうかを判断する。もし、文がユーザ登録メモから生成されたものであった場合、音声合成制御部61は韻律データと音響特徴パラメータ、音声データを変更する。たとえば、柔らかなイントネーションは、ゆっくり目で穏やかな声によって実現されるとすると、音声合成制御部61は、韻律データの継続時間長の値をそれぞれ同じ割合で伸長させる。また、テキスト音声合成部4の波形生成部27が用いる音声データを、穏やかな声のデータをもとに生成された音声データと入れ替える。そして、音響特徴パラメータの音量に関する部分の値を同じ割合で小さくする。これにより、出力される合成音は柔らかなイントネーションを持つようになる。

20

【0114】

同様にして、再生頻度によって、強調イントネーションを実現したり、好感度情報によってはつらつとしたイントネーションを実現したりすることが出来る。また、画像ないし音響データが記録された時間に応じて、合成音と同期させて、効果音を出力することも出来る。

30

【0115】

音声合成装置1、音声合成装置60は、携帯型の画像音声記録再生装置に適用できる。図21は、音声合成装置60を内蔵した携帯型画像音声記録再生装置の構成を示す図である。

【0116】

図21において、画像音声記録再生装置70は、被写体をレンズ71を通して撮像するカメラ撮像部72と、カメラ撮像部72で撮像された被写体画像に記録処理や再生処理を施すと共に、画像圧縮処理や画像伸長処理を施す画像記録再生部73と、画像記録再生部73で記録処理や再生処理が施された画像を後述するシステム制御部74による制御の基に表示する画像表示部75とを備える。

40

【0117】

カメラ撮像部72は、撮像した画像を画像記録再生部73に入力する。画像記録再生部73は、カメラ撮像部72で撮像された撮影画像を圧縮してファイルとして撮影画像ファイルデータベース76に記録するための記録処理、ファイルを撮影画像ファイルデータベース76から読み出して伸長する再生処理を行う。また、画像記録再生部73は、必要に応じて音声信号をファイルフォーマットを使用して撮影画像ファイルデータベース76に画像信号と共に入力する。画像表示部75は、好ましくは画像又はアイコン表示の可能な平板表示器であり、画像ファイルに対応してそのファイル内の画像データや見出し用縮小画像であるサムネイル画像を表示する。また、撮影原画像も表示できる。また、文字入力

50

メニュー表示機能も有している。

【0118】

また、この画像音声記録再生装置70は、画像記録再生部73により記録処理が施された各原画像に対応して付加される情報や撮影メディアフォーマットに記載された情報をシステム制御部74の制御に基づいてファイルデータとして格納している撮影画像ファイルデータベース76と、使用者により操作される操作キー部78と、操作キー部78からの操作入力信号、特に画像選択情報データ、登録文字列情報データ、連続再生スタート画像指定データなどを受けてシステム制御部74に送る登録キー制御部77とを備える。

【0119】

撮影画像ファイルデータベース76は、書き込み可能な不揮発性メモリまたは書き込み消去可能なメモリメディアであって、例えばメモリカード、カード型HDD、リムーバブル光ディスク等のリムーバブルメディアからなり、DCFファイルシステム等のリムーバブルメディアフォーマットに準拠して読み出しが可能とされる。また、ユーザ登録文字情報がリムーバブルメディアフォーマットに準拠して入力できることが好ましい。

【0120】

また、この画像音声記録再生装置70は、画像記録再生部73、画像表示部75、撮影画像ファイルデータベース76、登録キー制御部77に接続されて各部を制御するシステム制御部74を備える。システム制御部74は、記録再生制御機能部81、カメラ状態制御機能部82、ファイルシステム制御機能部83、ユーザキー設定機能部84を有する。

【0121】

また、システム制御部74内の上記カメラ状態機能部82、ファイルシステム機能部83及びユーザキー設定機能部84は、音声合成装置60の情報読み出し部2を構成する。つまり、情報読み出し部2は、システム制御部74内のカメラ状態機能部82、ファイルシステム機能部83及びユーザキー設定機能部84を介して、撮影画像ファイルデータベースから上記画像又は音響データに付随する情報を読み出す。情報読み出し部2で読み出された情報は、音声合成部60を構成する文生成部3に供給される。文生成部3は、読み出した情報を基にテキストを作成する。テキスト音声合成部4は、文生成部3にて作成されたテキストを音声に合成して出力する。音声合成制御部61は、テキスト音声合成部4における音声合成をコントロールする。

【0122】

情報読み出し部2を構成する、カメラ情報制御機能部82は、デジタルカメラの状態に関する情報を文生成部3に供給する。デジタルカメラの状態とは、使用されているアプリケーションが何語モードで起動されているかや、その時点での何月日、時間、呼び出しメニューのモードなどを指す。

【0123】

ファイルシステム制御部83は、画像に関する情報を文生成部3に供給する。画像に関する情報とは、その画像の日時、ファイル拡張子、その画像へのアクセス頻度、その画像が格納されているフォルダの名前、ユーザコメント、感性キー情報などを指す。

【0124】

ユーザキー設定制御機能部84は、出力音声について使用者から指定された情報を供給する。出力音声についてユーザから指定された情報とは、出力音声の言語、ボリューム、スピード、声の高さなどを指す。

【0125】

音声合成部60は、上述したように、Exifファイルの画像ファイルや、オーディオファイルから画像又は音響データに付随する情報である、ユーザ登録メモ、格納フォルダ名、記録時の情報、再生頻度、又は好感度情報を読み出し、読み出した情報に基づいて文を生成し、生成した文を音声合成して音声として出力することができる。また、音声合成部60は、音声合成制御部61を備えているので、文がユーザ登録メモから生成されたものであった場合、韻律データと音響特徴パラメータ、音声データを変更する。これにより、柔らかなイントネーションを合成音に持たせることができる。同様にして、再生頻度によ

10

20

30

40

50

て、強調イントネーションを実現したり、好感度情報によってはつらつとしたイントネーションを実現したりすることが出来る。また、画像ないし音響データが記録された時間に応じて、合成音と同期させて、効果音を出力することも出来る。

【0126】

このような音声合成部60を備えることにより、画像音声記録再生装置70は、音声によって案内される情報を基に、記録/再生された画像をユーザに簡単に検索させることができる。

【0127】

なお、本発明に係る音声合成装置は、デジタルカメラによって撮影されて記録された画像を検索する画像検索装置として用いられてもよい。この画像検索装置は、デジタルカメラから得られる情報を基にテキストを生成する文生成手段と、上記文生成手段によって生成されたテキストを基に音声を合成するテキスト音声合成手段とを備え、上記文生成手段によって生成されたテキストによる文を読み上げる。

10

【0128】

この画像検索装置にあって、上記文生成手段は、デジタルカメラの撮影メディアフォーマットに記載された情報、すなわち画像又は音響データに付随する情報を基にテキストを生成する。また、文生成手段は、ユーザ登録メモを基にテキストを生成する。また、上記文生成手段は、画像データを格納するフォルダ名を基にテキストを生成する。また、上記文生成手段は、上記デジタルカメラから得られる情報に基づいて生成する文の数を変える。また、上記文生成手段は、画像へのアクセス頻度に基づいて生成する文の数を変えてもよい。また、好感度キーに基づいて生成する文の数を変えてもよい。

20

【0129】

また、上記文生成手段は、ユーザ登録メモから単語抽出し、テンプレートに当てはめて文生成を行う。このとき、上記文生成手段は、ユーザ登録メモに記載された文字列を順に、テンプレートの空欄に当てはめて文生成を行なう。また、上記文生成手段は、ユーザ登録メモに記載された文字列を解析し、単語を抜き出して、テンプレートの空欄に当てはめて文生成を行う。また、上記文生成手段は、ユーザ登録メモに記載された文字列をテンプレートに当てはまるように変形し、テンプレートの空欄に当てはめて文生成を行う。また、上記文生成手段は、ユーザ登録メモに記載された文字列を解析し、単語を抜き出して、テンプレートに当てはまるように変形し、テンプレートの空欄に当てはめて文生成を行う。

30

【0130】

また、上記文生成手段は、ユーザ登録メモに記載された単語を用いて、概念辞書から別の単語を選出し、テンプレートに当てはめて文生成を行なう。また、上記文生成手段は、ユーザ登録メモに記載された単語を用いて、概念辞書から同じ意味情報を持つ単語を選出し、テンプレートに当てはめて文生成を行なう。また、上記文生成手段は、ユーザ登録メモに記載された単語を用いて、概念辞書からその単語の上位概念の単語を選出し、テンプレートに当てはめて文生成を行なう。

【0131】

また、上記文生成手段は、画像データを格納するフォルダ名をテンプレートに当てはめて文生成を行なってもよい。また、上記文生成手段は、撮影情報データから時間情報を抽出し、テンプレートに当てはめて文生成を行ってもよい。また、上記文生成手段は、デジタルカメラから得られる情報の種類によって、使用するテンプレートを選択する。また、上記文生成手段は、デジタルカメラから得られる情報として、ユーザ別アクセス頻度を使用してもよい。また、上記文生成手段は、デジタルカメラから得られる情報として、好感度キーに基づいた好感度情報を使用してもよい。また、上記文生成手段は、デジタルカメラから得られる情報として、ファイルの拡張子によって、使用するテンプレートを選択してもよい。

40

【0132】

また、上記文生成手段は、検索対象の他の画像ファイルに付随する情報を用いて、使用

50

するテンプレートを選択してもよい。検索対象の他の画像ファイルに付随する情報として、画像が撮影された時間を用い、使用するテンプレートを選択する。上記文生成手段は、検索対象の他の画像ファイルに付随する情報として、ユーザ登録メモに記載されている単語を用い、使用するテンプレートを選択してもよい。また、上記文生成手段は、ユーザ登録メモから単語を抽出し、メモに記載された単語の情報として、単語の表記、品詞、意味情報のいずれか1つ以上を用いて、テンプレートを選択する。

【0133】

上記画像検索装置にあって、上記テキスト音声合成手段は、デジタルカメラから得られる言語の情報によって、出力される合成音声の言語を変更する。また、上記テキスト音声合成手段は、ユーザによって指定された言語を合成音の言語として出力してもよい。また、上記テキスト音声合成手段は、デジタルカメラのモードから使用言語を選択してもよい。また、上記テキスト音声合成手段は、ユーザ登録メモに含まれる文字列を解析し、言語判定をする。また、上記テキスト音声合成手段は、決定された言語が、上記文生成手段によって生成された文の言語と異なる場合、生成された文を決定された言語の文に変換する機能を有する。

10

【0134】

また、上記テキスト音声合成手段は、デジタルカメラから得られる情報の内容によって、上記文生成手段によって生成された文を出力するときの文体を変更する。また、上記テキスト音声合成手段は、デジタルカメラから得られる情報として、ユーザ登録メモの内容、撮影情報データ、撮影時間、撮影場所、画像へのアクセス頻度、好感度キーによる好感度情報、呼び出しモードのメニューのうちいずれか1つ以上を用いることによって、上記文生成手段によって生成された文を出力するときの文体を変更する。上記テキスト音声合成手段は、装置が使用される季節や時間帯、日によって、上記文生成手段によって生成された文を出力するときの文体を変更する。

20

【0135】

また、上記テキスト音声合成手段は、デジタルカメラから得られる情報の内容によって、上記文生成手段によって生成された文を出力するときの合成音声の韻律を切り替える。また、上記テキスト音声合成手段は、韻律として、基本周波数と音韻継続時間長のいずれか1つ以上を切り替える。

【0136】

また、上記テキスト音声合成手段は、デジタルカメラから得られる情報として、ユーザ登録メモの内容、撮影情報データ、撮影時間、撮影場所、画像へのアクセス頻度、好感度キーによる好感度情報、呼び出しモードのメニューのうちいずれか1つ以上を用いることによって、上記文生成手段によって生成された文を出力するときの合成音声の韻律を切り替える。また、上記テキスト音声合成手段は、装置が使用される季節や時間帯、日によって、上記文生成手段によって生成された文を出力するときの合成音声の韻律を変更する。

30

【0137】

また、上記テキスト音声合成手段は、デジタルカメラから得られる情報の内容によって、上記文生成手段によって生成された文を出力するときの合成音声の話速を決定する。また、上記テキスト音声合成手段は、デジタルカメラから得られる情報として、ユーザ登録メモの内容、撮影情報データ、撮影時間、撮影場所、画像へのアクセス頻度、好感度キー、呼び出しモードのメニューのうちいずれか1つ以上を用いることによって、上記文生成手段によって生成された文を出力するときの合成音声の話速を決定する。上記テキスト音声合成手段は、装置が使用される季節や時間帯、日によって、上記文生成手段によって生成された文を出力するときの合成音声の話速を決定する。

40

【0138】

また、上記テキスト音声合成手段は、デジタルカメラから得られる情報の内容によって、上記文生成手段によって生成された文を出力するときの合成音声の音量レベルを決定する。また、上記テキスト音声合成手段は、デジタルカメラから得られる情報として、ユーザ登録メモの内容、撮影情報データ、撮影時間、撮影場所、画像へのアクセス頻度、好感

50

度キー、呼び出しモードのメニューのうちいずれか1つ以上を用いることによって、上記文生成手段によって生成された文を出力するときの合成音声の音量レベルを決定する。また、上記テキスト音声合成手段は、装置が使用される季節や時間帯、日によって、上記文生成手段によって生成された文を出力するときの合成音声の音量レベルを決定する。

【0139】

また、上記テキスト音声合成手段は、デジタルカメラから得られる情報の内容によって、上記文生成手段によって生成された文を出力するときの合成音声の声色を決定する。上記テキスト音声合成手段は、デジタルカメラから得られる情報として、ユーザ登録メモの内容、撮影情報データ、撮影時間、撮影場所、画像へのアクセス頻度、好感度キー、呼び出しモードのメニューのうちいずれか1つ以上を用いることによって、上記文生成手段によって生成された文を出力するときの合成音声の声色を決定する。上記テキスト音声合成手段は、装置が使用される季節や時間帯、日によって、上記文生成手段によって生成された文を出力するときの合成音声の声色を決定する。

10

【0140】

また、上記テキスト音声合成手段は、デジタルカメラから得られる情報の内容によって、上記文生成手段によって生成された文を出力するときの合成音声に効果音を付与するか否かを決定する。また、上記テキスト音声合成手段は、デジタルカメラから得られる情報として、ユーザ登録メモの内容、撮影情報データ、撮影時間、撮影場所、画像へのアクセス頻度、好感度キー、呼び出しモードのメニューのうちいずれか1つ以上を用いることによって、上記文生成手段によって生成された文を出力するときの合成音声に効果音を付与するか否かを決定する。また、上記テキスト音声合成手段は、装置が使用される季節や時間帯、日によって、上記文生成手段によって生成された文を出力するときの合成音声に効果音を付与するか否かを決定する。また、上記テキスト音声合成手段は、デジタルカメラから得られる情報の内容によって、上記文生成手段によって生成された文を出力するときの合成音声に付与する効果音の種類を決定する。また、上記テキスト音声合成手段は、装置が使用される季節や時間帯、日によって、上記文生成手段によって生成された文を出力するときの合成音声に付与する効果音の種類を決定する。

20

【図面の簡単な説明】

【0141】

【図1】音声合成装置のブロック図である。

30

【図2】音声合成装置の全体的な処理手順を示すフローチャートである。

【図3】Exifファイルの画像ファイル構成例を示す図である。

【図4】Exifファイルのオーディオファイル構成例を示す図である。

【図5】情報読み出し部が保持している情報読み出しのための対応表を示す図である。

【図6】情報読み出し部によって読み出されたデータの例を示す図である。

【図7】文生成部の構成を示すブロック図である。

【図8】出力内容決定部の構成を示すブロック図である。

【図9】出力内容決定部の処理手順を示すフローチャートである。

【図10】文字列解析部によって解析されて出力されるデータの例を示す図である。

【図11】テンプレートの具体例を示す図である。

40

【図12】文生成部内のさらに文生成部の構成を示すブロック図である。

【図13】文構造判定部が有する構造決定用ルールの具体例を示す図である。

【図14】テキスト音声合成部の構成を示すブロック図である。

【図15】言語処理部の構成を示すブロック図である。

【図16】韻律生成部の構成を示すブロック図である。

【図17】韻律データの前半部の継続時間長データを示す図である。

【図18】韻律データの後半部の基本周波数データを示す図である。

【図19】波形生成部の構成を示すブロック図である。

【図20】音声合成制御部を備えた音声合成装置の構成を示すブロック図である。

【図21】音声合成装置を備えた画像音声記録再生装置の構成を示すブロック図である。

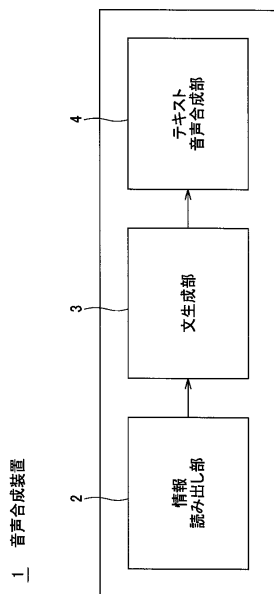
50

【符号の説明】

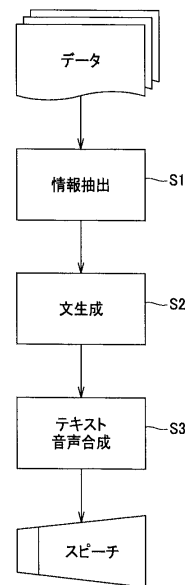
【0142】

- 1 音声合成装置、 2 情報読み出し部、 3 文生成部、 4 テキスト音声合成装置、
- 5 出力内容決定部、 7 文字列解析部、 10 出力言語決定部、 11 文内容決定部、
- 12 文の数決定部、 13 テンプレート選択部、 14 テンプレートデータベース

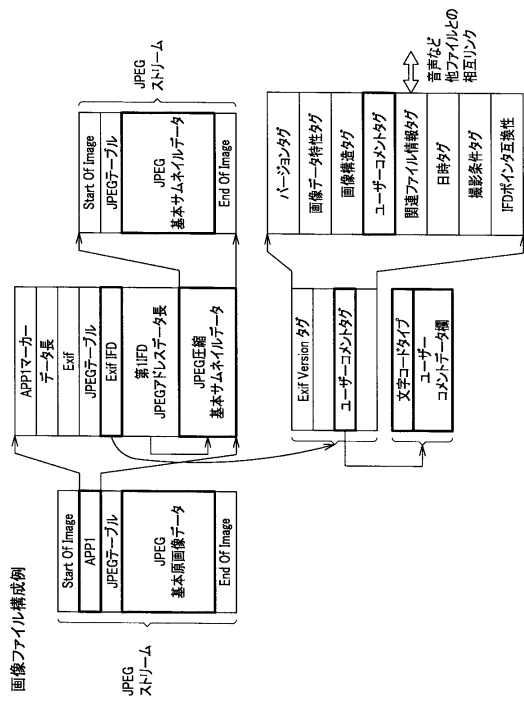
【図1】



【図2】



【 図 3 】

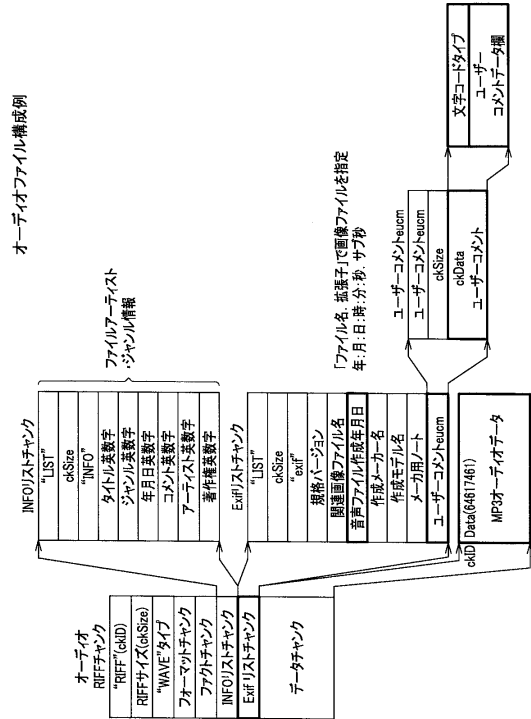


【 図 5 】

情報読み出しのための対応表

データラベル	ファイルへのポインタ	ファイルでの形式
格納フォルダ名	Exifリストチャック	文字列
ファイル名	Exifリストチャック	ファイル名、拡張子
撮影の時間	Exifリストチャック	年:月:日:時:分:秒
撮影の場所	ユーザーコメント eucom	文字列
再生頻度	ユーザーコメント eucom	数字
好感度情報	ユーザーコメント eucom	文字列
ユーザー登録メモ	ユーザーコメント eucom	文字列

【 図 4 】

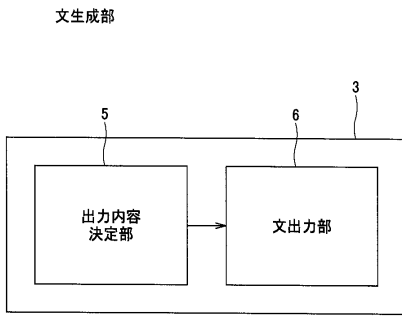


【 図 6 】

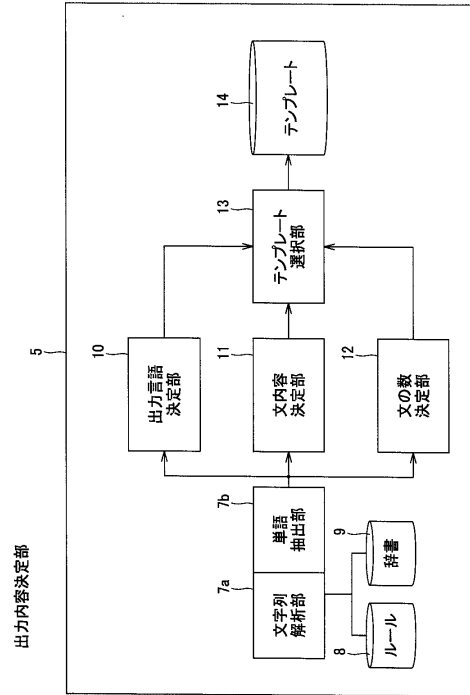
読み出されたデータの例

格納フォルダ名 = "家族/犬"
ファイル名 = "モモの散歩.jpg"
撮影の時間 = "2002/06/27", "09:52:15"
撮影の場所 = "近所の公園"
再生頻度 = 0
好感度情報 = 5
ユーザー登録メモ = "モモちゃんが初めて散歩に行った。"

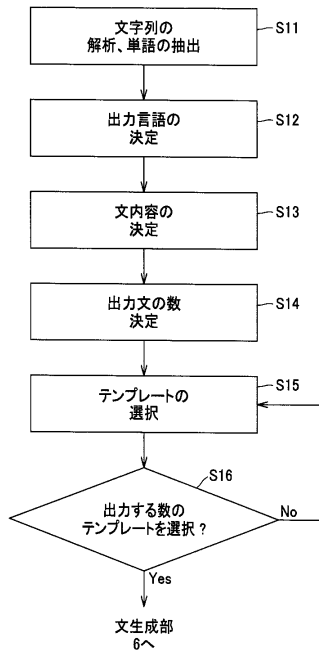
【図7】



【図8】



【図9】



【図10】

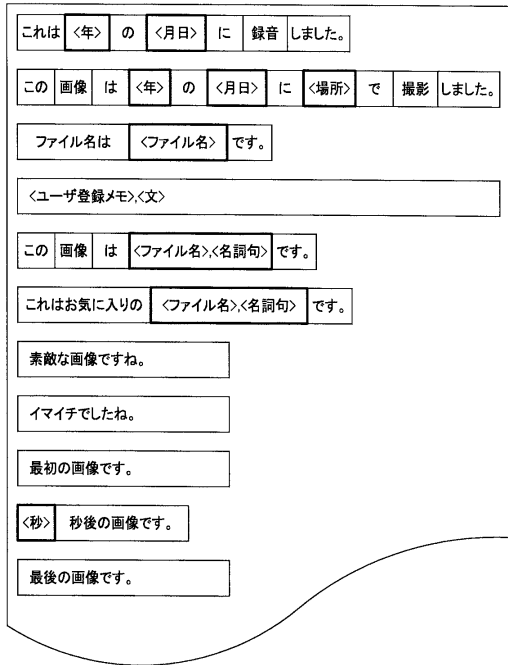
出力されるデータの例

格納フォルダ名	家族	名詞
	犬	名詞
ファイル名	モモ	名詞
	の	格助詞
	散歩	名詞
ファイル名拡張子	jpg	
撮影時間	2002	
	06	
	27	
	09	
	52	
	15	
撮影の場所	近所	名詞
	の	格助詞
	公園	名詞
再生頻度	0	
好感度情報	5	

ユーザ登録メモ	モモ	名詞
	ちゃん	接尾辞
	が	格助詞
	初めて	副詞
	散歩	名詞
	に	格助詞
	行	動詞
	っ	語尾
	た	助動詞
	。	句点

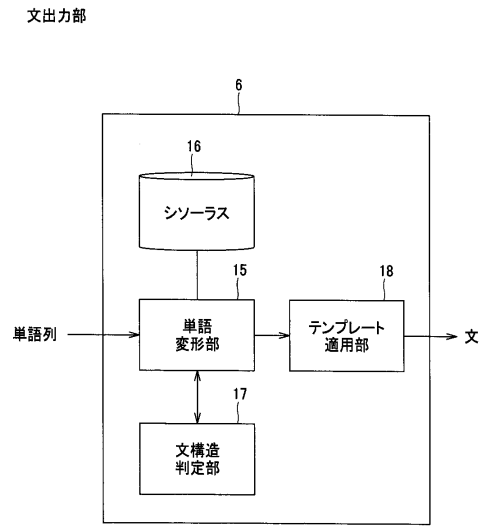


【図 1 1】

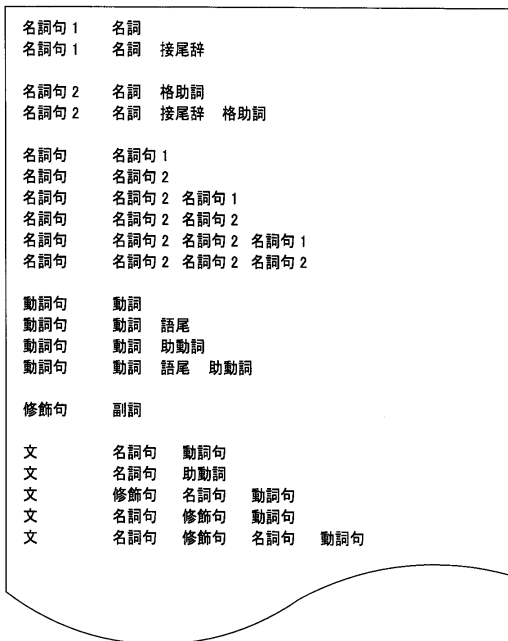


テンプレートの例

【図 1 2】

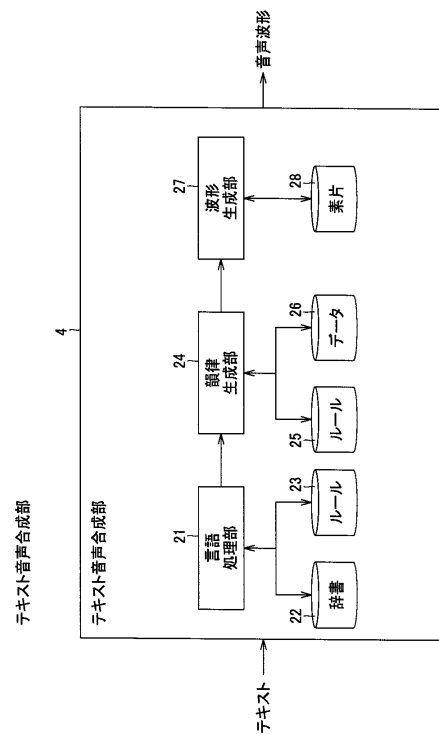


【図 1 3】

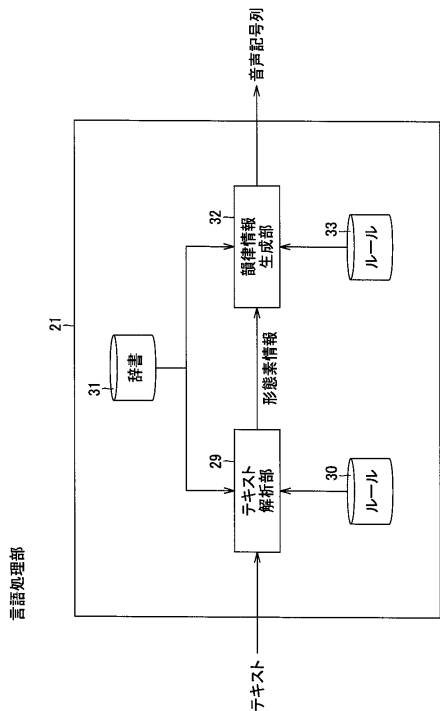


文構造決定用ルール

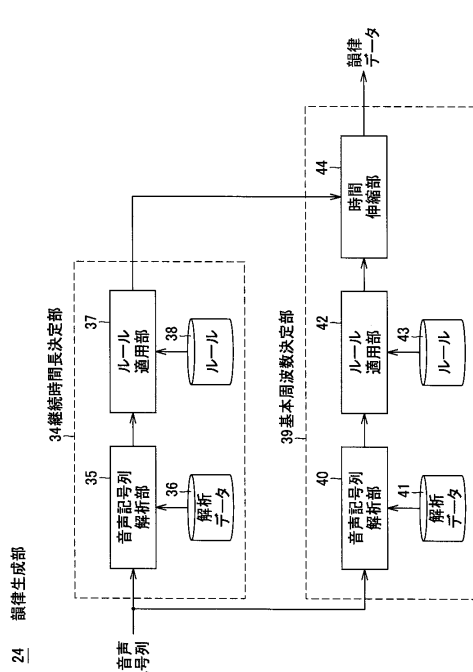
【図 1 4】



【図15】



【図16】



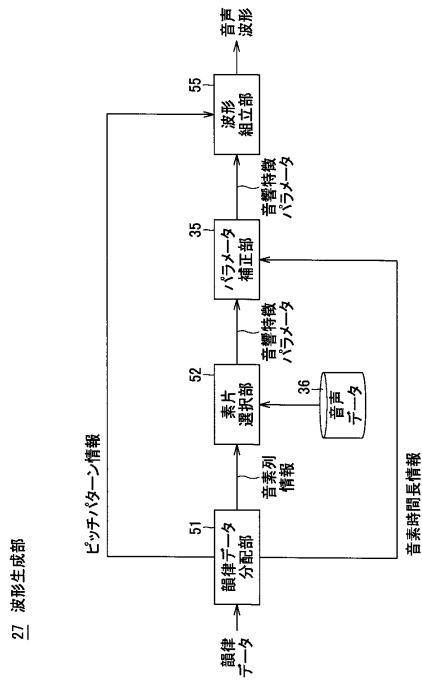
【図17】

[duration]			
0	pau	41580	
41580	s	1815	
43395	o	1375	
44770	r	440	
45210	e	1760	
46970	k	1705	
48675	a	1100	
49775	r	495	
50270	a	1815	
52800	w	2530	
54175	a	3135	
57310	pau	10560	
67870	y	1100	
68970	a	1155	
70125	m	770	
70895	a	1870	
72165	u	1485	
74250	u	1320	
75570	ch,j	2475	
78045	k	1485	
175175	o	1595	
176770	m	1485	
178255	o	1485	
179740	d	1045	
180785	o	1320	
182105	r	605	
182710	e	1320	
184030	n	1045	
185075	a	1595	
186670	k	1485	
188155	u	825	
188980	n	990	
189970	a	1650	
191620	r	715	
192335	i	1375	
193710	m	1045	
194755	a	1925	
196680	sh,i	2585	
198265	t	2035	
201300	a	880	
202180	pau	1000	

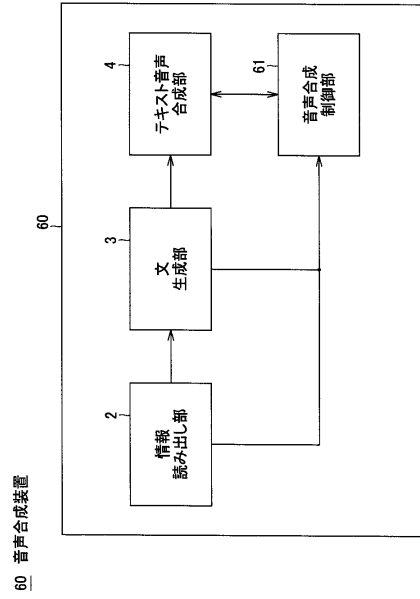
【図18】

[pitch]		
43293		238.906311
43513		234.910721
43734		230.767242
43954		227.435181
44175		225.769623
44395		226.314301
44616		229.292206
44836		234.609772
45057		241.770844
45277		249.844055
112309		159.380371
112530		159.833099
112750		161.753784
112971		164.357941
113191		166.815460
113412		168.570435
113632		169.459976
113853		169.640289
114073		169.417862
114294		169.083038
114514		168.809052
194776		138.374146
194997		138.937317
195217		139.687164
195438		140.680725
195658		141.870010
195879		143.029205
196099		143.828522
196320		144.030182
196540		143.651123
196761		142.945404
202794		122.000000

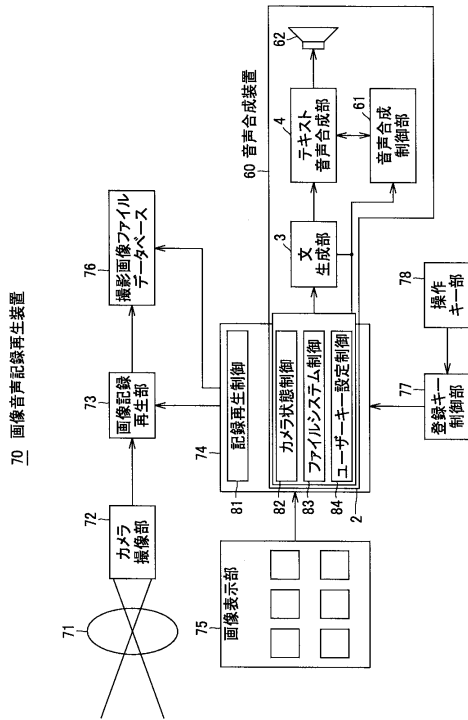
【図19】



【図20】



【図21】



---

フロントページの続き

審査官 柴垣 俊男

- (56)参考文献 特開2002-175176(JP,A)  
特開2002-175094(JP,A)  
特開2002-163277(JP,A)  
特開2002-207494(JP,A)  
特開2001-296882(JP,A)  
特開平07-072888(JP,A)

- (58)調査した分野(Int.Cl., DB名)  
G10L 13/00 - 13/08  
JSTPlus(JDreamII)