

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第5097007号
(P5097007)

(45) 発行日 平成24年12月12日 (2012.12.12)

(24) 登録日 平成24年9月28日 (2012.9.28)

(51) Int.Cl.	F I
G 1 O L 13/08 (2006.01)	G 1 O L 13/08 1 6 O
G 1 O L 13/06 (2006.01)	G 1 O L 13/06 2 4 O C

請求項の数 5 (全 16 頁)

(21) 出願番号	特願2008-134655 (P2008-134655)	(73) 特許権者	000001007
(22) 出願日	平成20年5月22日 (2008.5.22)		キヤノン株式会社
(65) 公開番号	特開2009-37214 (P2009-37214A)		東京都大田区下丸子3丁目30番2号
(43) 公開日	平成21年2月19日 (2009.2.19)	(74) 代理人	100076428
審査請求日	平成23年4月8日 (2011.4.8)		弁理士 大塚 康德
(31) 優先権主張番号	特願2007-182555 (P2007-182555)	(74) 代理人	100112508
(32) 優先日	平成19年7月11日 (2007.7.11)		弁理士 高柳 司郎
(33) 優先権主張国	日本国 (JP)	(74) 代理人	100115071
			弁理士 大塚 康弘
		(74) 代理人	100116894
			弁理士 木村 秀二
		(74) 代理人	100130409
			弁理士 下山 治
		(74) 代理人	100134175
			弁理士 永川 行光

最終頁に続く

(54) 【発明の名称】 音声処理装置及び方法

(57) 【特許請求の範囲】

【請求項 1】

複数の語または句で構成される文を、録音再生方式または規則合成方式を用いて再生可能な音声処理装置であって、

再生する文を構成する複数の語または句のそれぞれが、録音再生方式で再生される語または句か、規則合成方式で再生される語または句かを特定する特定手段と、

前記複数の語または句のそれぞれを、前記特定手段により特定された再生方式を用いて第1の配置順序で再生する場合に、録音再生方式を用いた再生と規則合成方式を用いた再生とが切り替わる反転回数が所定回数未満であれば前記第1の配置順序による再生を選択し、前記反転回数が前記所定回数以上であれば、前記第1の配置順序に対して文構成として自然さを低下させた複数の配置順序のうち、前記反転回数が前記所定回数未満となる配置順序による再生を選択する選択手段と、

前記複数の語または句のそれぞれを、前記特定手段により特定された再生方式を用いて、前記選択手段により選択された配置順序で再生する再生手段と

を備えることを特徴とする音声処理装置。

【請求項 2】

前記反転回数は、録音再生方式を用いた再生から規則合成方式を用いた再生へ切り替わる回数と、規則合成方式を用いた再生から録音再生方式を用いた再生へ切り替わる回数との和に相当することを特徴とする請求項1に記載の音声処理装置。

【請求項 3】

複数の語または句で構成される文を、録音再生方式または規則合成方式を用いて再生可能な音声処理装置によって実行される音声処理方法であって、

特定手段が、再生する文を構成する複数の語または句のそれぞれが、録音再生方式で再生される語または句か、規則合成方式で再生される語または句かを特定する特定工程と、

選択手段が、前記複数の語または句のそれぞれを、前記特定工程で特定された再生方式を用いて第１の配置順序で再生する場合に、録音再生方式を用いた再生と規則合成方式を用いた再生とが切り替わる反転回数が所定回数未満であれば前記第１の配置順序による再生を選択し、前記反転回数が前記所定回数以上であれば、前記第１の配置順序に対して文構成として自然さを低下させた複数の配置順序のうち、前記反転回数が前記所定回数未満となる配置順序による再生を選択する選択工程と、

10

再生手段が、前記複数の語または句のそれぞれを、前記特定工程で特定された再生方式を用いて、前記選択工程で選択された配置順序で再生する再生工程とを有することを特徴とする音声処理方法。

【請求項４】

請求項３に記載の音声処理方法の各工程をコンピュータに実行させるためのプログラム。

【請求項５】

請求項４に記載のプログラムを記憶したコンピュータ読み取り可能な記憶媒体。

【発明の詳細な説明】

【技術分野】

20

【０００１】

本発明は、音声処理装置及び方法に関する。

【背景技術】

【０００２】

音声合成方式には、録音再生方式と規則合成方式がある。録音再生方式は音質が高いが、定型文にしか使えない。規則合成方式はあらゆる文に使えるが、録音再生方式よりも音質は劣り、読み間違いも起こる。

【０００３】

従来、音声合成によりガイダンス音声を出力する音声処理装置において、録音再生方式と規則合成方式を併用する方法が存在する（特許文献１）。

30

【０００４】

【特許文献１】特開平９－９７０９４号公報

【発明の開示】

【発明が解決しようとする課題】

【０００５】

しかしながら、上記の従来技術では、１つのガイダンス音声の中で録音再生方式と規則合成方式の切り替えを頻繁に行うと、両者の音質の差のため、ガイダンスが聞きにくくなるという課題があった。

【０００６】

そこで、本発明は、録音再生方式と規則合成方式とを切り替えながら音声合成を行う音声処理装置において、合成音声の聴覚的な自然性の向上させることを目的とする。

40

【課題を解決するための手段】

【０００７】

上記目的を達成するため、本発明の一側面によれば、複数の語または句で構成される文を、録音再生方式または規則合成方式を用いて再生可能な音声処理装置であって、再生する文を構成する複数の語または句のそれぞれが、録音再生方式で再生される語または句か、規則合成方式で再生される語または句かを特定する特定手段と、前記複数の語または句のそれぞれを、前記特定手段により特定された再生方式を用いて第１の配置順序で再生する場合に、録音再生方式を用いた再生と規則合成方式を用いた再生とが切り替わる反転回数に基づいて、前記複数の語または句のそれぞれを前記第１の配置順序で再生するか、該

50

第１の配置順序とは異なる配置順序で再生するかを選択する選択手段と、前記複数の語または句のそれぞれを、前記特定手段により特定された再生方式を用いて、前記選択手段により選択された配置順序で再生する再生手段とを備えることを特徴とする音声処理装置が提供される。

【発明の効果】

【０００８】

本発明によれば、録音再生方式と規則合成方式とを切り替えながら音声合成を行う音声処理装置において、合成音声の聴覚的な自然性を向上させることができる。

【発明を実施するための最良の形態】

【０００９】

10

以下、図面を参照して本発明の好適な実施形態について詳細に説明する。なお、本発明は以下の実施形態に限定されるものではなく、本発明の実施に有利な具体例を示すにすぎない。また、以下の実施形態の中で説明されている特徴の組み合わせの全てが本発明の課題解決手段として必須のものであるとは限らない。

【００１０】

以下の実施形態では、ＦＡＸ機能を有する画像形成装置に本発明を適用した例を用いて説明する。

【００１１】

図１Ａは、本発明の音声処理装置が適用される画像形成装置のハードウェア構成の概略を示すブロック図である。

20

【００１２】

２０１はＣＰＵ（中央処理装置）であり、システム制御部として装置全体の動作を制御している。２０２はＲＯＭであり、制御プログラムを格納する。具体的には、後述する音声処理を行うための音声処理プログラム、画像の符号処理等を行うための画像処理プログラムを格納している。２０３はＲＡＭであり、ＣＰＵ２０１のワークエリアを提供し、各種データ等を格納するために用いられる。

【００１３】

２０４Ａはマイクロホン等の音声入力装置、２０４Ｂはスピーカ等の音声出力装置である。

【００１４】

30

２０５はスキャナ部であり、画像データを読み取り、バイナリデータに変換する機能を有するデバイスである。２０６はプリンタ部であり、画像データを、記録紙に出力するプリンタ機能を有するものである。

【００１５】

２０７はファクシミリ通信制御部であり、電話回線等の外部回線を介して遠隔的に設置されたファクシミリ装置とファクシミリ通信を行うためのインタフェースである。２０８は操作部であり、オペレータによる操作を行うためのものである。具体的にはテンキー等の操作ボタン、タッチパネル等が含まれる。

【００１６】

２０９は画像／音声処理部である。具体的には、ＤＳＰ等のハードウェアチップから構成され、画像処理、音声処理における積和演算等を高速に実行する。

40

【００１７】

２１０はネットワーク通信制御部であり、ネットワーク回線へのインタフェースとしての機能を有し、プリントジョブの受信やインターネットＦＡＸ送受信を実行するために用いられる。２１１はハードディスク装置（ＨＤＤ）であり、後述するアドレス帳や音声データ等を保持する。

【００１８】

図１Ｂは、上述の画像形成装置において実現される音声処理装置の機能構成を示すブロック図である。

【００１９】

50

エントリ取得部101は、少なくとも表記とその読みとその音声を登録可能なエントリを取得する。エントリ（語または句など）は、HDD211に構成されるエントリ保持部106によって保持される。

【0020】

エントリ保持部106は、例えば、図4に示すようなデータ構造のアドレス帳を構成するエントリの集合を保持している。各エントリには、ユーザの操作に関連付けられる、表記、その表記の読み、その読みの音声、電話番号、FAX番号、E-mailアドレスが登録可能となっている。

【0021】

エントリに登録する音声は、音声入力装置204Aを介してエントリの内容を発声し録音したものである。図4で音声の欄にあるw2001、w2002などの記号は、音声を取り出すための音声インデックスである。

【0022】

登録情報判断部102は、エントリ取得部101が取得したエントリに音声に登録されているか否かを判断する。

【0023】

ガイダンス選択部103は、エントリ取得部101が取得したエントリに合わせて、HDD211に構成されるガイダンス保持部107が保持するガイダンスを選択する。エントリに音声に登録されている場合は後述のガイダンス1を選択する。エントリに音声に登録されていない場合は後述のガイダンス2を選択する。ガイダンス保持部107は、IDを用いてガイダンスを管理している。各IDに対してガイダンス1（第1ガイダンス）とガイダンス2（第2ガイダンス）を保持する。また、各ガイダンスは、メッセージの内容が固定されている固定部分の他に、ユーザの操作に応じたメッセージが挿入されることを示す可変部分を含む。

【0024】

ガイダンス保持部107が保持するガイダンスの一例を図5に示す。ガイダンス中の<\$name>の部分が可変部分であり、その他の部分が固定部分である。IDが1のガイダンスは、FAX機能が選択されてFAXの送り先を確認するときに用いる。IDが2のガイダンスは、メール機能が選択されてメールの送り先を確認するときに用いる。

【0025】

図5に示されるように、ガイダンス1とガイダンス2は同義の内容を表すが、異なった表現を用いている。即ち、2つのガイダンスは語または句の配列順序が異なる。具体的には、ガイダンス1は、「FAXを」、「へ送ります。」という固定部分を有し、その中間に可変部分が位置する。一方、ガイダンス2は、ガイダンス1の可変部分を固定部分の末尾に位置させている。この場合、可変部分の直前には当該可変部分を説明する語または句が配置されることになる。図5の例では「送り先は、」の句が可変部分の直前に位置している。

【0026】

ガイダンス作成部104は、ガイダンス選択部103が選択したガイダンスに、エントリ取得部101が取得したエントリの情報を挿入し、最終的に出力するガイダンスを作成する。

【0027】

音声合成部105は、録音再生方式と規則合成方式とを選択的に切り替えながら音声合成を行うことが可能であり、ガイダンス作成部104が作成したガイダンスの合成音声を音声出力装置204Bを介して発生する。具体的には、ガイダンス中の固定部分、音声に登録されたエントリの部分には録音再生方式を用いる。音声に登録されていないエントリ（語または句）の部分には規則合成方式を用いる。

【0028】

HDD211に構成される基本合成単位辞書108は、ガイダンスの固定部に含まれる語または句に関する情報を保持する。また、少なくとも表記とその音声を取り出すための音声インデックスを保持する。図6に例を示す。ここで、読点「、」の音声インデックスw1006は、300ミリ秒の無音を指しているとする。また、区点「。」の音声インデックスw1007は、400ミリ秒の無音を指しているとする。

10

20

30

40

50

【 0 0 2 9 】

HDD211に構成される低水準合成単位辞書109は、規則合成に必要な音声インデックスを保持する。音声の単位は、音素、ダイフォン、モーラなどである。図7にモーラを単位とした場合の低水準合成単位辞書109の例を示す。

【 0 0 3 0 】

HDD211に構成される音声データベース110は、エントリ保持部106、基本合成単位辞書108、低水準合成単位辞書109が保持する音声インデックスに対応する音声をまとめて保持する。

【 0 0 3 1 】

図2は、本実施形態における音声処理装置の動作を説明するフローチャートである。このフローチャートに対応するプログラムは例えば音声処理プログラムに含まれ、CPU 201によって実行される。ここでは、FAX機能を持つ画像形成装置に上記構成の音声処理装置を適用した例を用いて説明する。さらに、FAXの送り先を確認するためのガイダンスを出力する場合について説明する。

10

【 0 0 3 2 】

まずステップS201で、ユーザが操作部208を介してFAX送信の準備をする。例えば、FAX送信のメニューを選択し、原稿を画像形成装置にセットする。

【 0 0 3 3 】

次に、ステップS202で、ユーザはアドレス帳を開き、所望の送り先を選択する。アドレス帳の例は図4に示すようなものである。

20

【 0 0 3 4 】

ステップS203で、エントリ取得部101は、ユーザが選択した送り先のエントリを取得する。

【 0 0 3 5 】

ステップS204で、登録情報判断部102が、ステップS203で取得されたエントリに音声登録されているか否かを判断する。例えば、図4のアドレス帳で、「佐藤」のエントリには音声登録されているが、「田中」のエントリには音声登録されていない。エントリに音声登録されている場合はステップS205へ進み、登録されていない場合はステップS207へ進む。

【 0 0 3 6 】

ステップS205で、ガイダンス選択部103が、ガイダンス保持部107からガイダンス1を選択する。ところで、出力するガイダンスは、FAXの送り先を確認するためのガイダンスである。これは図5でIDが1のガイダンスである。よって選択されるガイダンスは「FAXを<\$name>へ送ります。」となる。

30

【 0 0 3 7 】

ステップS206で、ガイダンス作成部104が、ステップS205で選択されたガイダンス1の可変部分に、ステップS203で取得されたエントリの情報をタグにして挿入する。タグには音声インデックスを登録する。

【 0 0 3 8 】

例えば、ステップS203で取得されたエントリが図4の「佐藤」であるとする。すると、作成されるガイダンスは「FAXを<音声=w2001;>へ送ります。」となる。ここで、<音声=w2001;>の部分がタグである。タグは<>で囲み、「項目名=値;」の形で情報を登録するものとする。

40

【 0 0 3 9 】

一方ステップS207では、ガイダンス選択部103が、ガイダンス保持部107からガイダンス2を選択する。ステップS205と同様、図5でIDが1のガイダンスを選択する。よって、選択されるガイダンスは、「FAXを送ります。送り先は、<\$name>。」となる。

【 0 0 4 0 】

ステップS208で、登録情報判断部102が、S203で取得されたエントリに読みが登録されているか否かを判断する。例えば、図4のアドレス帳で、「田中」のエントリには読みが

50

登録されているが、「鈴木」のエントリには読みが登録されていない。エントリに読みが登録されている場合はステップS209へ進み、登録されていない場合はステップS210へ進む。

【0041】

ステップS209では、ガイドンス作成部104が、ステップS207で選択されたガイドンス2の可変部分に、ステップS203で取得されたエントリの情報をタグにして挿入する。タグには読みを登録する。例えば、ステップS203で取得されたエントリを図4の「田中」とする。すると、作成されるガイドンスは「FAXを送ります。送り先は、＜読み=タナカ；＞。」となる。

【0042】

一方ステップS210では、ガイドンス作成部104が、ステップS207で選択されたガイドンス2の可変部分に、ステップS203で取得されたエントリの情報をタグにして挿入する。タグには表記を登録する。例えば、ステップS203で取得されたエントリを図4の「鈴木」とする。すると、作成されるガイドンスは「FAXを送ります。送り先は、＜表記=鈴木；＞。」となる。

【0043】

ステップS211で、音声合成部105が、ステップS206またはS209またはS210で作成されたガイドンスを音声出力する。

【0044】

ステップS212で、ユーザがステップS211で出力された音声ガイドンスを聞き、FAXの送り先が正しいか否かを判断する。正しい場合はステップS213へ進む。正しくない場合はステップS202へ戻り、送り先を選択しなおす。

【0045】

ステップS213で、画像形成装置がFAXを送信して処理を終了する。

【0046】

図3は、本実施形態における音声合成部105の処理手順を説明するフローチャートである。

【0047】

ステップS301で、音声出力するガイドンスを取得する。このガイドンスは、ガイドンス作成部104がステップS206またはS209またはS210で作成したものである。

【0048】

次にステップS302で、基本合成単位辞書108を用いて、ガイドンスを基本合成単位に分割する。また、元々ガイドンスに挿入されていたタグはそのまま基本合成単位とする。この分割は公知の形態素解析技術を利用可能である。例えば、基本合成単位辞書の表記とガイドンスとのマッチングを行い、左最長一致の基準に従って分割する。

【0049】

図6の基本合成単位辞書を用いて、ガイドンス「FAXを送ります。送り先は、＜読み=タナカ；＞。」を分割した結果を図8に示す。ガイドンスは7個の基本合成単位に分割される。また、元々ガイドンスに挿入されていたタグ＜読み=タナカ；＞は、そのまま基本合成単位となる。

【0050】

ステップS303で、分割された基本合成単位をタグへ置換する。タグには表記と音声インデックスを登録する。また、元々ガイドンスに挿入されていたタグはそのまま残す。例えば、基本合成単位「FAXを」をタグ＜表記=FAXを；音声=w1001；＞へ置換する。図8の基本合成単位をタグへ置換した結果を図9に示す。

【0051】

ステップS304で、変数iに1を設定する。また、変数nにタグの数を設定する。図9の例で、タグの数は7である。

【0052】

ステップS305で、iがn以下か否かを判断する。iがn以下の場合はステップS306へ進む。

10

20

30

40

50

iがnより大きい場合は処理を終了する。

【 0 0 5 3 】

ステップS306で、i番目のタグに音声インデックスが登録されているか否かを判断する。音声インデックスが登録されている場合はステップS307へ進む。音声インデックスが登録されていない場合はステップS308へ進む。図9の例では、6番目のタグには音声インデックスが登録されていないが、その他のタグには音声インデックスが登録されている。

【 0 0 5 4 】

ステップS307で、i番目のタグに登録されている音声インデックスを用いて音声を取り出す。取り出した音声を再生する。これは、録音再生方式の音声合成である（第1音声合成）。

10

【 0 0 5 5 】

一方のS308では、i番目のタグに読みが登録されているか否かを判断する。読みが登録されている場合はステップS310へ進む。読みが登録されていない場合はステップS309へ進む。

【 0 0 5 6 】

ステップS309で、i番目のタグに読み付けを行う。まず、i番目のタグに登録されている表記を取り出す。次に、取り出した表記の読みを推定する。この処理には未知語の読み付け技術が利用可能である。最後に、推定した読みをi番目のタグに登録する。例えば、タグ<表記=鈴木;>の表記「鈴木」から読み「スズキ」を推定したとする。すると、タグは<表記=鈴木;読み=スズキ;>となる。もっとも、未知語の読み付け技術は誤りを含みうる。例えば、表記「鈴木」から、誤った読み「リンボク」を推定する可能性もある。

20

【 0 0 5 7 】

ステップS310で、i番目のタグに登録されている読みを取り出す。次に、規則合成を用いて、取り出した読みから音声合成を行い出力する（第2音声合成）。

【 0 0 5 8 】

ステップS311で、変数iの値を1増やして、ステップS305に戻る。

【 0 0 5 9 】

以上説明したように、音声に登録されていないエントリを取得するとガイダンス2を選択する。そして、固定部分を録音再生方式を用いて出力し、可変部分を規則合成方式を用いて出力する。ところで、ガイダンス2は可変部分がガイダンスの最後にある。これにより、録音再生の部分と規則合成の部分を分離した出力が可能となる。上述したように、音声に登録されていないエントリ（語または句）をガイダンス1（第1の文法）で再生する場合より、ガイダンス2（第2の文法）で再生するほうが、録音再生方式で再生される語または句と規則合成方式で再生される語または句との切替わりの回数が減少する場合がある。即ち、本実施の形態の効果として、上記切替わりの回数を減少させることができると言えるだろう。以上により、録音再生方式による出力音と規則合成方式による出力音の品質の差によるガイダンスの聞きにくさを低減することが可能となる。

30

【 0 0 6 0 】

また、上述したガイダンス2の文法においては、可変部分の前に可変部分を説明する語がある。ユーザはこの可変部分を説明する語を前もって聞くことにより、可変部分の内容（情報の種類）が推測しやすくなる。これは規則合成で出力される可変部分の聞きやすさにつながる。

40

【 0 0 6 1 】

なお、エントリに登録される読みにはアクセント情報が付与されていてもよい。この場合、ステップS309でアクセント情報付きの読みを推定する。また、ステップS310で規則合成の入力は、アクセント情報付きの読みとなる。

【 0 0 6 2 】

また、ステップS310で、読みを低水準合成単位に分割し、低水準合成単位の音声をそのまま再生してもよい。例えば、読み「スズキ」を分割した結果は、<モーラ=ス;音声=w0165;><モーラ=ズ;音声=w0160;><モーラ=キ;音声=w0210;>となる。この結果をステッ

50

ブS307の録音再生で出力する。ただし、「鈴木」に対して音声登録されている場合に比べ、出力される音質は悪くなる。

【0063】

また、ガイダンス2の可変部分に、「さん」「です」などの短い付属語を加えてもよい。具体的には、「FAXを送ります。送り先は、<\$name>さん。」、「FAXを送ります。送り先は、<\$name>です。」、「FAXを送ります。送り先は、<\$name>さんです。」などとしてもよい。つまり、可変部分をガイダンスの最後の文節または句または語に置く。

【0064】

上述の実施形態では、FAX機能を有する画像形成装置に本発明の音声処理装置を適用した例を説明したが、本発明はこれに限られない。音声合成機能を有する情報処理装置であれば、同様に適用できることは言うまでもない。

10

【0065】

以上で説明している音声処理装置は、複数の語または句で構成される文を、録音再生方式または規則合成方式を用いて再生可能な音声処理装置であり、次の処理を行うものである。まず、再生する文を構成する複数の語または句の夫々が、録音再生方式で再生される語または句か、規則合成方式で再生される語または句かを特定する。次に、上記複数の語または句のそれぞれを、特定された再生方式を用いて第1の配置順序で再生する場合に、録音再生方式を用いた再生と規則合成方式を用いた再生とが切り替わる回数（反転回数）に基づいて、上記複数の語または句のそれぞれを第1の配置順序（第1の文法）で再生するか、第1の配置順序とは異なる配置順序（第1の文法とは異なる文法）で再生するかを選択する。上記の処理においては、同義の文を異なる文法で表現する場合に、必ずしも全ての語句を一致させることを主眼には置いてはいない。

20

【0066】

上記音声処理装置の特徴は、録音再生方式を用いた再生と規則合成方式を用いた再生とが頻繁に切り替わることによる、聴覚的な聞き難さを解消しようとするものである。そのために文法（別の表現をするならば、文を構成する語または句の配置順序）を異ならせるものである。

【0067】

以上では、解りやすくするために、録音再生方式を用いた再生と規則合成方式を用いた再生とが切り替わる回数（反転回数）が最大で2回になるような短文を用いた簡単な例を説明した。この場合、録音再生方式を用いた再生と規則合成方式を用いた再生とが切り替わる回数が2回の場合（録音再生方式から規則合成方式に変わり、規則合成方式から録音再生方式に変わる場合）に、この切替わり回数を1回に減少させるという単純な制御になる。

30

【0068】

しかしながら、録音再生方式を用いた再生と規則合成方式を用いた再生とが切り替わる回数（反転回数）が最大で2回を超えるような長文においては、上述したような2種類のガイダンスの切り替えでは十分とは言えない。

【0069】

このような長文を想定した場合には、上記切り替わる回数が許容範囲を超えるかどうかを基準として、ガイダンス1（第1の文法（第1の配置順序））と他のガイダンス（第1とは異なる1つ以上の文法（第2の配置順序））を選択することが有効である。

40

【0070】

以下の説明は、上記音声処理装置が長文にも対応可能であることを追加的に説明するものである。

【0071】

以下では、図10、図11を用いて、1つのガイダンス中に、上述した可変部分（録音再生方式と規則合成方式が選択的に適用される部分）が2箇所含まれる例について説明する。

【0072】

50

図11は、ガイダンス保持部107が保持するガイダンスの一例である。ガイダンス1からガイダンス4までの「文の構成（単語の配列順）としての聞き易さ」の関係は、ガイダンス1>ガイダンス2=ガイダンス3>ガイダンス4の関係であるとする。即ち、もし、再生される文の全ての語句が録音再生方式である場合には、ガイダンス1を用いて再生される音声が一番聞き取り易く、ガイダンス4を用いて再生される音声が一番聞き取り難い。また、ガイダンス2とガイダンス3の聞き取り易さは同様である。また、ガイダンス中の<\$title>と<\$name>の部分が可変部分である。IDが1のガイダンスは、原稿をスキャンして電子メールで送信する機能が選択されて、送り先と原稿のタイトルを確認するときに用いる。

【0073】

10

図10は、本実施形態における音声処理装置の動作を説明するフローチャートである。

【0074】

まずステップS1001で、ユーザが操作部208を介して電子メール送信の準備をする。例えば、電子メール送信のメニューを選択し、原稿を画像形成装置にセットする。

【0075】

次に、ステップS1002で、ユーザはアドレス帳を開き、所望の送り先を選択する。これはステップS202と同じ処理である。

【0076】

ステップS1003で、エントリ取得部101は、ユーザが選択した送り先のエントリを取得する。これはステップS203と同じ処理である。

20

【0077】

ステップS1004で、ユーザがセットした原稿のタイトルを取得する。例えば、スキャナ部205が、原稿を読み込み、その結果にOCRを行い、タイトルを取得する。

【0078】

ステップS1005で、ガイダンス1を基本合成単位に分割し、その後、タグへ変換する。ガイダンス1の<\$name>には、ステップS1003で取得したエントリをタグにして挿入する。例えば図4の「佐藤」を取得したとする。ガイダンス1の<\$title>には、ステップS1004で取得したタイトルを挿入する。例えば「週報」を取得したとする。以上の例によると、ガイダンス1は「スキャン原稿、週報を電子メールで<音声=w2001;>へ送ります。」となる。

30

【0079】

基本合成単位への分割はステップS302と同じ処理である。ただし、ガイダンス1に、基本合成単位辞書108に含まれない文字列がある場合は、<表記=;>のタグを用いる。例えば「週報を」が基本合成単位辞書108に含まれていない場合、<表記=週報を;>とする。タグへの変換はステップS303と同じ処理である。タグへ変換した結果の例を図12に示す。基本合成単位辞書108として図6を用いる。また、「スキャン原稿」の音声インデックスがw1010であるとする。

【0080】

ステップS1006で、音声合成部105がガイダンス1を音声出力するときに、再生が録音再生方式を用いた再生と規則合成方式を用いた再生との間で切り替わる回数（反転回数）を計算する。この回数は、録音再生方式を用いた再生から規則合成方式を用いた再生へ切り替わる回数と、規則合成方式を用いた再生から録音再生方式を用いた再生へ切り替わる回数との和に相当する。タグに音声インデックスが登録されている場合、録音再生方式を用いる。タグに音声インデックスが登録されていない場合、規則合成方式を用いる。

40

【0081】

図12の例を用いて具体的に説明する。IDが3のタグは、音声インデックスが登録されていないので、規則合成方式を用いる。その他のタグは、音声インデックスが登録されているので、録音再生方式を用いる。IDが3のタグの前で、録音再生方式から規則合成方式へ切り替わる。IDが3のタグの後ろで、規則合成方式から録音再生方式へ切り替わ

50

る。よって、切り替わる回数は2回である。

【0082】

ステップS1007で、録音再生方式と規則合成方式とが切り替わる回数が、所定数(N回)より小さいか否かを判定する。Nは予め定めた定数である。所定数未満の場合(YES)は、ステップS1015へ進み、所定数以上の場合(NO)はステップS1008へ進む。例えばN=2である場合、図12の例では、ステップS1008へ進む。

【0083】

ステップS1008からステップS1010の処理は、ガイダンス1の代わりにガイダンス2を用いることを除いて、ステップS1005からステップS1007の処理と同じである。

10

【0084】

ステップS1011からステップS1013の処理は、ガイダンス1の代わりにガイダンス3を用いることを除いて、ステップS1005からステップS1007の処理と同じである。

【0085】

ステップS1014の処理は、ガイダンス1の代わりにガイダンス4を用いることを除いて、ステップS1005の処理と同じである。

【0086】

ステップS1015で、ステップS1005、ステップS1008、ステップS1011、またはステップS1014で置換されたタグを基に音声出力を行う。具体的な処理は、図3のステップS304からステップS311の処理と同じである。

20

【0087】

ステップS1003では、エントリとして「佐藤」を取得し、ステップS1004でタイトルとして「週報」を取得した例について、ステップS1008以降の処理を説明する。

【0088】

ステップS1008では、ガイダンス2が「スキャン原稿、週報を電子メールで送ります。送り先は、<音声=w2001;>。」となる。タグに変換した結果の例を図13に示す。IDが3のタグの前後で、録音再生方式と規則合成方式の切り替えが起こり、切り替わる回数は2回となる。よって、ステップS1010において、切り替わり回数(2)はN(2)よりも小さくないので(NO)、ステップS1011へ進む。

30

【0089】

ステップS1011で、ガイダンス2が「スキャン原稿を電子メールで<音声=w2001;>へ送ります。タイトルは、週報。」となる。タグに変換した結果の例を図14に示す。「タイトルは」の音声インデックスがw1011であるとする。IDが9のタグの前後で、録音再生方式と規則合成方式の切り替えが起こる。ただし、IDが10のタグは、400ミリ秒の無音であり、そのあとにもタグはない。つまり、IDが9のタグの後には音声がない。このように、後ろに音声がない場合は切り替え回数として数えないことにする。よって、この場合の切り替わり回数は1回となる。そして、ステップS1013の判定において、切り替わり回数が2回よりも小さくなっているため(YES)、ステップS1015へ進む。ステップS1015で、ガイダンス3を音声出力する。

40

【0090】

上述したN=2の場合は、例えば「ユーザとしては、2回以上の切り替わりを許容できない」という意図が有る。よって、図10の工程においては、文の構成(語句の配列順)として自然なガイダンス1からガイダンス3まで順番に、ガイダンス2回以上の切り替わりが発生しないガイダンスが見つかるまで判定を続けることになる。しかしながら、もし、各判定(S1007,S1010,S1013)で所望の切り替わり回数未満のガイダンスが見つからなければ、最終的にガイダンス4を選ぶことになる。このガイダンス4は、各可変部の最後に無音部分を配置することにより、「<\$name>と<\$title>の両方ともに規則合成方式で再生されるような場合などに最も切り替わり回数(反転回数)が少なくなる」という性

50

質を備える。

【0091】

以上の実施の形態によれば、できるだけ文の構成（単語の配列順）としての聞き易く、かつ、ユーザの許容範囲の切り替わり回数（反転回数）において再生できるガイダンスを、ユーザーに提供できる。

【0092】

（他の実施形態）

以上、本発明の実施形態を詳述したが、本発明は、複数の機器から構成されるシステムに適用してもよいし、また、一つの機器からなる装置に適用してもよい。

【0093】

なお、本発明は、前述した実施形態の各機能を実現するプログラムを、システム又は装置に直接又は遠隔から供給し、そのシステム又は装置に含まれるコンピュータがその供給されたプログラムコードを読み出して実行することによっても達成される。

【0094】

従って、本発明の機能・処理をコンピュータで実現するために、そのコンピュータにインストールされるプログラムコード自体も本発明を実現するものである。つまり、上記機能・処理を実現するためのコンピュータプログラム自体も本発明の一つである。

【0095】

その場合、プログラムの機能を有していれば、オブジェクトコード、インタプリタにより実行されるプログラム、OSに供給するスクリプトデータ等、プログラムの形態を問わない。

【0096】

プログラムを供給するためのコンピュータ読み取り可能な記録媒体としては、例えば、フレキシブルディスク、ハードディスク、光ディスク、光磁気ディスク、MO、CD-ROM、CD-R、CD-RWなどがある。また、記録媒体としては、磁気テープ、不揮発性のメモリカード、ROM、DVD（DVD-ROM、DVD-R）などもある。

【0097】

また、プログラムは、クライアントコンピュータのブラウザを用いてインターネットのホームページからダウンロードしてもよい。すなわち、ホームページから本発明のコンピュータプログラムそのもの、若しくは圧縮され自動インストール機能を含むファイルをハードディスク等の記録媒体にダウンロードしてもよい。また、本発明のプログラムを構成するプログラムコードを複数のファイルに分割し、それぞれのファイルを異なるホームページからダウンロードする形態も考えられる。つまり、本発明の機能・処理をコンピュータで実現するためのプログラムファイルを複数のユーザに対してダウンロードさせるWWWサーバも、本発明の構成要件となる場合がある。

【0098】

また、本発明のプログラムを暗号化してCD-ROM等のコンピュータ読み取り可能な記憶媒体に格納してユーザに配布してもよい。この場合、所定条件をクリアしたユーザにのみ、インターネットを介してホームページから暗号化を解く鍵情報をダウンロードさせ、その鍵情報で暗号化されたプログラムを復号して実行し、プログラムをコンピュータにインストールしてもよい。

【0099】

また、コンピュータが、読み出したプログラムを実行することによって、前述した実施形態の機能が実現されてもよい。なお、そのプログラムの指示に基づき、コンピュータ上で稼動しているOSなどが、実際の処理の一部又は全部を行ってもよい。もちろん、この場合も、前述した実施形態の機能が実現され得る。

【0100】

さらに、記録媒体から読み出されたプログラムが、コンピュータに挿入された機能拡張ボードやコンピュータに接続された機能拡張ユニットに備わるメモリに書き込まれてもよい。そのプログラムの指示に基づき、その機能拡張ボードや機能拡張ユニットに備わるC

10

20

30

40

50

P Uなどが実際の処理の一部又は全部を行ってもよい。このようにして、前述した実施形態の機能が実現されることもある。

【図面の簡単な説明】

【 0 1 0 1 】

【図 1 A】実施形態に係る画像形成装置のハードウェア構成例を示すブロック図である。

【図 1 B】実施形態における音声処理装置の機能構成を示すブロック図である。

【図 2】実施形態における音声処理装置の動作例を説明するフローチャートである。

【図 3】実施形態における音声合成部の処理手順を説明するフローチャートである。

【図 4】実施形態におけるエントリ保持部が保持するアドレス帳の構造例を示す図である。

10

【図 5】実施形態におけるガイダンス保持部が保持するガイダンスの例を示す図である。

【図 6】実施形態における基本合成単位辞書の例を示す図である。

【図 7】実施形態における低水準合成単位辞書の例を示す図である。

【図 8】実施形態におけるガイダンスの基本合成単位への分割例を示す図である。

【図 9】実施形態における分割した基本合成単位のタグへの置換例を示す図である。

【図 1 0】実施形態における音声処理装置の動作例を説明するフローチャートである。

【図 1 1】実施形態におけるガイダンス保持部が保持するガイダンスの例を示す図である。

【図 1 2】実施形態における分割した基本合成単位のタグへの置換例を示す図である。

【図 1 3】実施形態における分割した基本合成単位のタグへの置換例を示す図である。

20

【図 1 4】実施形態における分割した基本合成単位のタグへの置換例を示す図である。

【符号の説明】

【 0 1 0 2 】

101 エントリ取得部

102 登録情報判断部

103 ガイダンス選択部

104 ガイダンス作成部

105 音声合成部

106 エントリ保持部

107 ガイダンス保持部

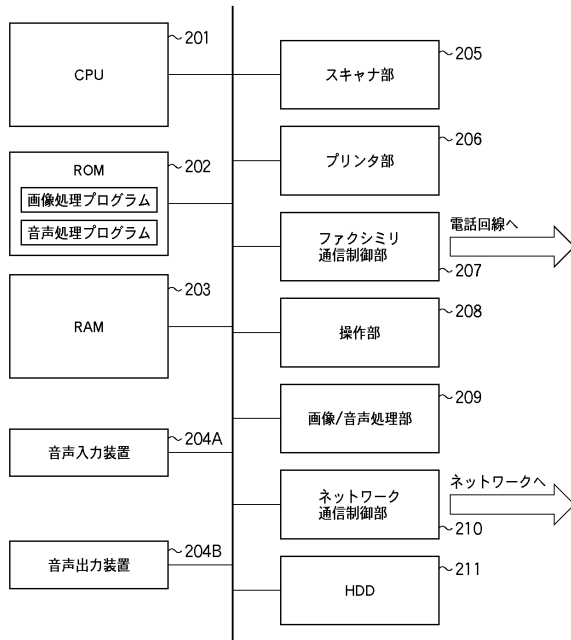
108 基本合成単位辞書

109 低水準合成単位辞書

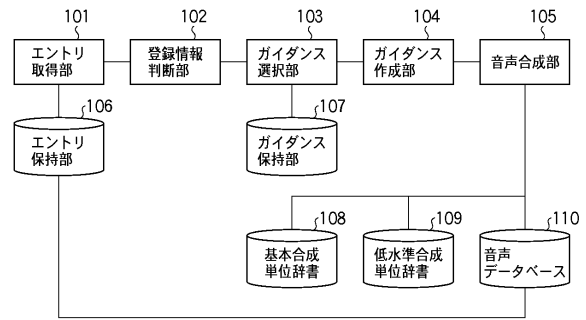
110 音声データベース

30

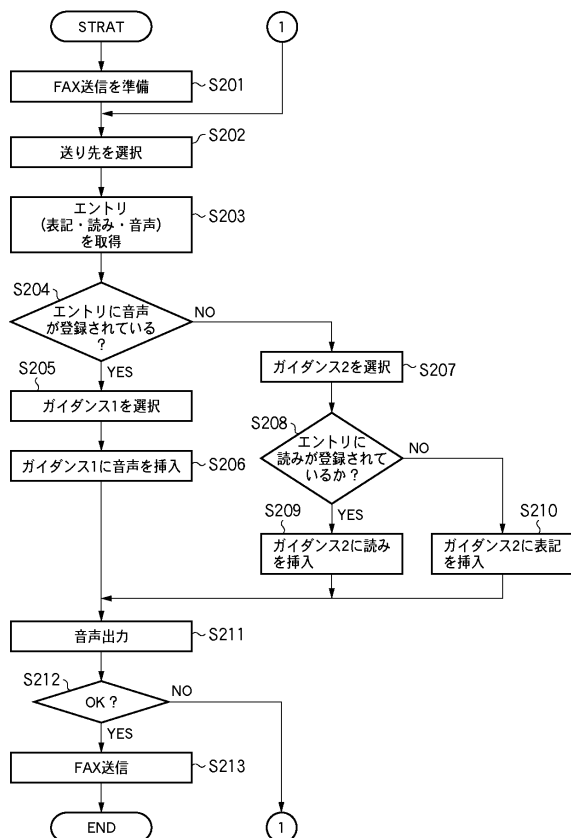
【図 1 A】



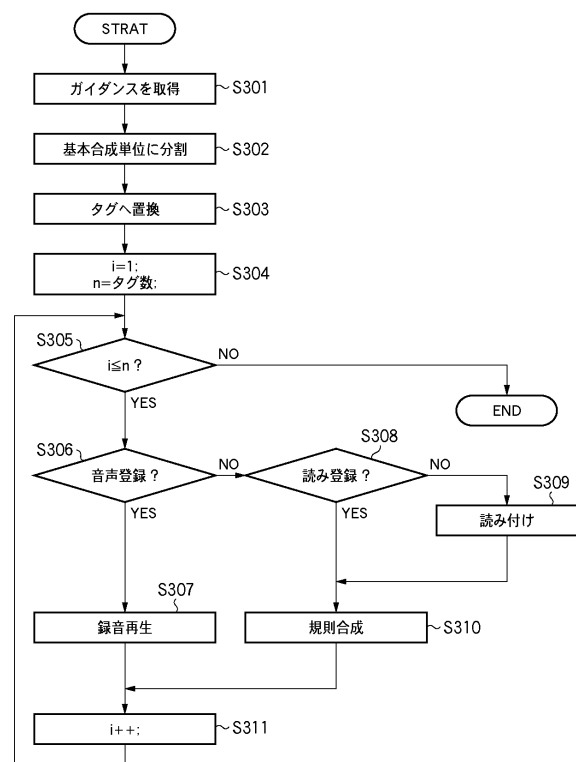
【図 1 B】



【図 2】



【図 3】



【図 4】

表記	読み	音声	電話	FAX	E-mail
佐藤	サトウ	w2001	111-1111	111-1112	sato@aaa.com
鈴木			222-1111	222-1112	suzuki@bbb.com
田中	タナカ		333-1111	333-1112	tanaka@ccc.com
Boyle		w2002	444-1111	444-1112	boyle@ccc.com
Smith	スミス	w2003	555-1111	555-1112	smith@eee.com
..

【図 5】

ID	ガイドス1	ガイドス2
1	FAXを<\$name>へ送ります。	FAXを送ります。送り先は、<\$name>。
2	スキャン原稿を電子メールで<\$name>へ送ります。	スキャン原稿を電子メールで送ります。送り先は、<\$name>。
..

【図 6】

表記	音声
FAXを	w1001
電子メールで	w1002
へ	w1003
送ります	w1004
スキャン原稿を	w1005
、	w1006
。	w1007
送り先は	w1008
....

【図 9】

ID	タグ
1	<表記=FAXを: 音声=w1001:>
2	<表記=送ります: 音声=w1004:>
3	<表記=、: 音声=w1007:>
4	<表記=送り先は: 音声=w1008:>
5	<表記=、: 音声=w1006:>
6	<読み=タナカ:>
7	<表記=、: 音声=w1007:>

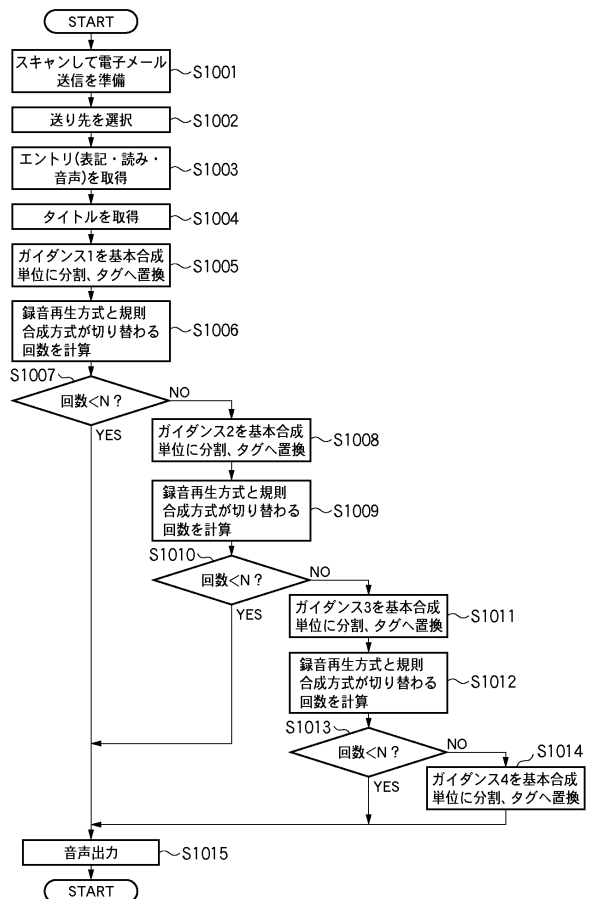
【図 7】

モーラ	音声
ア	w0226
イ	w0223
ウ	w0221
エ	w0215
オ	w0213
カ	w0212
....
ス	w0165
....

【図 8】

ID	基本合成単位
1	FAXを
2	送ります
3	、
4	送り先は
5	、
6	<読み=タナカ>
7	。

【図 10】



【図 1 1】

ID	ガイドンス1	ガイドンス2	ガイドンス3	ガイドンス4
1	スキャン原稿、 <\$title> を電子メールで <\$name> へ送ります。	スキャン原稿、 <\$title> を電子メールで 送ります。 送り先は、 <\$name>、	スキャン原稿を 電子メールで <\$name>へ 送ります。 タイトルは、 <\$title>、	スキャン原稿を 電子メールで 送ります。 送り先は、 <\$name>、 タイトルは、 <\$title>、
...

【図 1 4】

ID	タグ
1	<表記=スキャン原稿を；音声=w1005；>
2	<表記=電子メールで；音声=w1002；>
3	<音声=w2001；>
4	<表記=へ；音声=w1003；>
5	<表記=送ります；音声=w1004；>
6	<表記=、；音声=w1007；>
7	<表記=タイトルは；音声=w1011；>
8	<表記=、；音声=w1006；>
9	<表記=週報；>
10	<表記=、；音声=w1007；>

【図 1 2】

ID	タグ
1	<表記=スキャン原稿；音声=w1010；>
2	<表記=、；音声=w1006；>
3	<表記=週報を；>
4	<表記=電子メールで；音声=w1002；>
5	<音声=w2001；>
6	<表記=へ；音声=w1003；>
7	<表記=送ります；音声=w1004；>
8	<表記=、；音声=w1007；>

【図 1 3】

ID	タグ
1	<表記=スキャン原稿；音声=w1010；>
2	<表記=、；音声=w1006；>
3	<表記=週報を；>
4	<表記=電子メールで；音声=w1002；>
5	<表記=送ります；音声=w1004；>
6	<表記=、；音声=w1007；>
7	<表記=送り先は；音声=w1008；>
8	<表記=、；音声=w1006；>
9	<音声=w2001；>
10	<表記=、；音声=w1007；>

フロントページの続き

(72)発明者 相澤 道雄
東京都大田区下丸子3丁目30番2号 キヤノン株式会社内

審査官 菊地 陽一

(56)参考文献 特開平07-210194(JP,A)
特開平09-034490(JP,A)

(58)調査した分野(Int.Cl., DB名)
G10L 13/08
G10L 13/06