

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第3712227号  
(P3712227)

(45) 発行日 平成17年11月2日(2005.11.2)

(24) 登録日 平成17年8月26日(2005.8.26)

(51) Int. Cl.<sup>7</sup>

F I

G 1 0 L 13/06

G 1 0 L 5/04

F

G 0 6 F 3/16

G 0 6 F 3/16

3 3 0 K

G 1 0 L 13/08

G 1 0 L 3/00

H

請求項の数 5 (全 15 頁)

(21) 出願番号	特願2000-5380 (P2000-5380)	(73) 特許権者	000005326
(22) 出願日	平成12年1月14日 (2000.1.14)		本田技研工業株式会社
(65) 公開番号	特開2001-195080 (P2001-195080A)		東京都港区南青山二丁目1番1号
(43) 公開日	平成13年7月19日 (2001.7.19)	(74) 代理人	100064414
審査請求日	平成12年12月22日 (2000.12.22)		弁理士 磯野 道造
審査番号	不服2003-13115 (P2003-13115/J1)	(72) 発明者	京光 達哉
審査請求日	平成15年7月10日 (2003.7.10)		埼玉県和光市中央1丁目4番1号
			株式会社本田技術研究所
			内
		(72) 発明者	小島 康一
			埼玉県和光市中央1丁目4番1号
			株式会社本田技術研究所
			内

最終頁に続く

(54) 【発明の名称】 音声合成装置、音声合成方法におけるデータ作成方法、及び音声合成方法

(57) 【特許請求の範囲】

【請求項1】

送信者が作成したデータを送信する送信側装置と、該送信側装置から受信したデータを音声合成して読み上げる受信側装置とからなる音声合成装置において、

前記送信側装置には、音素の集合データとしての音声データを複数記憶する音声記憶手段と、

送信者の音声の特徴を抽出し、前記音声記憶手段の各音声データと音素ごとに比較して類似度が最大となる所定の音声データを決定するとともに、前記送信者の音声に近似させるため前記決定した所定の音声データを修正する特徴量修正データを作成する処理手段と、

前記所定の音声データを指定する指定データ及び前記特徴量修正データを送信する送信手段とを設け、

前記受信側装置には、前記送信側装置から送信された前記各データを受信する受信手段と、

前記送信側装置と同じ内容の音声記憶手段と、

前記受信した指定データに基づいて前記音声記憶手段から該当する音声データを読み込み、前記受信した特徴量修正データに基づいて前記読み込んだ音声データを修正した修正音声データを作成するとともに、前記修正音声データに基づいてデータを読み上げて音声合成を行う処理手段とを設けたこと、

を特徴とする音声合成装置。

## 【請求項 2】

前記送信側装置には、さらにテキストデータを入力する入力手段と、前記テキストデータを送信する送信手段を設け、

前記受信側装置の処理手段は、受信した前記テキストデータに基づいてアクセント位置又はポーズ長を含む韻律パラメータを予測した音声インデックスデータを作成すると共に、前記修正音声データと前記音声インデックスデータとに基づいて音声合成を行うこと、を特徴とする請求項 1 に記載の音声合成装置。

## 【請求項 3】

前記送信側装置には、さらにテキストデータを入力する入力手段を設け、

前記送信側装置の処理手段は、前記テキストデータに基づいてアクセント位置又はポーズ長を含む韻律パラメータを予測した音声インデックスデータを作成し、

前記送信手段は、前記音声インデックスデータも送信するとともに、

前記受信側装置の処理手段は、前記修正音声データと受信した前記音声インデックスデータとに基づいて音声合成を行うこと、

を特徴とする請求項 1 に記載の音声合成装置。

## 【請求項 4】

音素の集合データとしての音声データを複数記憶する音声記憶手段の中から音声合成側で使用すべき音声データを指定する指定データと、この指定データにより指定した音声データを修正する特徴量修正データを音声合成側に送信し、

前記音声記憶手段と同じ内容の音声記憶手段を備える前記音声合成側で、前記送信された指定データと前記特徴量修正データに基づいて音声合成する音声合成方法におけるデータ作成方法であって、

入力音声に対して音素ごとの特徴量を抽出し、

この特徴量と前記音声記憶手段の各音声データとを音素ごとに比較して、前記入力音声と類似度が最大となる所定の音声データを決定すると共に、この所定の音声データを指定する指定データを決定し、

さらに、この決定した所定の音声データを前記入力音声により近似するように当該音声データを修正する特徴量修正データを作成すること、

を特徴とする音声合成方法におけるデータ作成方法。

## 【請求項 5】

音素の集合データとしての音声データを複数記憶する音声記憶手段の中から音声合成に使用すべき音声データを読み出し、この読み出した音声データを修正し、この修正した修正音声データに基づいて音声合成の対象となるデータを読み上げるようにした音声合成方法であって、

入力音声に対して音素ごとの特徴量を抽出し、この特徴量と前記音声記憶手段と同じ内容の送信側の音声記憶手段の各音声データとを音素ごとに比較することで、前記入力音声と類似度が最大となる所定の音声データを指定する指定データと、この指定データにより指定される音声データを前記入力音声により近似させるための特徴量修正データと、を受信し、

この受信した指定データを用いて前記音声記憶手段の中から音声合成の対象となる音声データを読み出し、

このように読み出した音声データに対して、前記受信した特徴量修正データで、前記読み出した音声データが前記入力音声により近似するように修正し、

このように修正した修正音声データに基づいて、前記音声合成の対象となるデータを読み上げるようにしたこと、

を特徴とする音声合成方法。

## 【発明の詳細な説明】

## 【0001】

## 【発明の属する技術分野】

本発明は、音声合成方法に関し、殊に、電子メールなどで送られてくる電子データとして

10

20

30

40

50

のテキストデータを、このテキストデータの作成者などの音声で音声合成して読み上げる音声合成方法に関する。

【0002】

【従来の技術】

近年インターネットの普及により、電子メールでの文書のやり取りが頻繁に行われるようになった。通常電子メールでの文書のやり取りはテキストベースで行われ、電子メールの受信者はパソコン（パーソナルコンピュータ）やPDA（Personal Data Assistants）の画面上で電子メールの内容を目視により確認する。しかし、この電子メールの内容を送信者の音声で聞きたい、あるいは自己の音声で受信者に聞かせたいという要望がある。音声は、性別はもとより、年齢、体格、風貌や性格などにより異なり、人間の特徴をよく表すからである。つまり、受信者にとって音声での電子メールの内容確認は、まだ会ったことのない電子メールの送信者の具体的な人物像が思い浮かんできたり、電子メールの文書の行間では表せない送信者の感情を読み取ることができたりする。また、送信者にとっても同様であり、例えば、文書だけでは表せない行間を表現して受信者に伝えることができたりする。

10

【0003】

ところで、自己の音声をデジタル記録して音声ファイルを作成し、これを電子メールで相手方に送信することもできる。例えば、wav形式で自己の音声をデジタル記録して、これを電子メールの添付ファイルとして送信する方法である。wav形式の音声ファイルは、Windows（マイクロソフト社の登録商標）をOS（Operating System）とするパソコンなどで一般的に使用されるものであり、動作が軽く音質が大変よいなどの特徴がある。しかし、wav形式の音声ファイルは、ほんの数秒（2～3秒）の音声を記録するだけで100kB（キロバイト）以上のデータサイズになってしまう。比較の仕方によっても異なるが、同じ内容の文書を保存・記録するとして、wav形式の音声ファイルは、txt形式の文書ファイル（テキストデータ）に比べて数千倍のデータサイズを必要とする（場合によっては差が一万倍以上になることもある）。したがって、パソコンのハードディスクの容量やフロッピーディスクの容量など及び電子メールの送受信に要する時間などから、wav形式の音声ファイルを電子メールに添付することは非現実的であり、ほとんど行われていない。これは、Windowsばかりでなく他のOSにおける音声ファイル（例えばAIFF形式）でも同様である。

20

30

【0004】

最近、音声データを圧縮する手段としてMP3という音声データ圧縮規格が使用されるようになり、MP3に関連する商品が多数販売されるようになった。このMP3は「MPEG-1 Audio Layer-3」の略であり、人間にとって聞くことのできない周波数帯域の音声データ（つまり不要な部分の音声データ）をカットすることなどにより、CD並みの音質を確保しつつ、もとの音声ファイルの約1/10にデータサイズを圧縮することができる。しかし、MP3規格で音声データを圧縮しても約1/10がやっとであり、まだテキストデータに比べてデータサイズはかなり大きく、送信者側のパソコンや受信者側のパソコンのハードウェアに対する負担が大きいと共に、電子メールの送受信に時間を要する。

40

また、音声ファイルに共通していえることであるが、テキストファイルのように簡単に音声ファイルを編集することができない。したがって、音声ファイル作成の際、つまり音声入力する際に、言いよどんだり言い間違えたりなどの失敗が許されない。

【0005】

一方、近年、テキストデータを機械的に読み上げる音声合成に関する研究が盛んに行なわれている。このうち音声データベースを用いたコーパスベース（Corpus Base）の音声合成では、テキストデータをより人間に近い発音に音声合成するため、以下の方法をとっている。

【0006】

〔手順1〕 先ず、読み上げ用テキストデータの形態素解析を行なう。これは、読み上

50

げ用テキストデータがどのような単語とどのような助詞・助動詞・活用語尾で構成されるかを構文解析することである。具体的には、「僕は日本人です。」というテキストを、/僕/は/日本人/です。/という正しい構文解析を行なうことを指す。ちなみに、「日本人」という単語は、“日本/人”というように、“日本”と“人”という構文解析も可能である。この手順1は、それを間違いなく“日本人”という単語であるということを認識する過程である。

#### 【0007】

〔手順2〕 形態素解析で得られた各単語・構文情報をもとに、音素の選択・各単語のアクセントの位置・イントネーション・ポーズなどを決定する。この過程では、例えば、“日本人”という単語を音声合成するための音素を選択し、さらにその選択された音素のピッチやアクセントを決定する。“日本人”という言葉構成する“Ni”という発音でも、文字にすると“にほんじん”や“ニックネーム”というふうに、単語が違っていてもテキストは“に”と変わらないが、その音素を考えた場合、単語によって、また、後に続く発音や単語のアクセントによって音素は異なる。また、音素の長さ・高さ・強さも異なる。

10

つまり、この手順2で、各単語・構文情報をもとに、音声合成が自然に聞こえるための各音素について最適な長さ・高さ・強さを決定し、さらに音素系列における最適なイントネーション・ポーズなどを決定する。

#### 【0008】

〔手順3〕 手順2で得られた各データをインデックス化し、音声合成装置にこのインデックス化したデータを引き渡して音声合成を行ない、テキストデータを読み上げる。

20

#### 【0009】

このようにすることで、男・女といったように、予め決められた特定の音声での音声合成を行い、テキストデータを読み上げること（音声出力すること）が可能になっている。したがって、wav形式やMP3形式の音声データを送受信するのと違い、データサイズを少なくすることも可能である。また、編集も容易である。

#### 【0010】

##### 【発明が解決しようとする課題】

しかしながら、人間の音声は犯罪捜査に使われるように千差万別であり、男女のみの二者択一的な音声選択では、電子メールの送信者の具体的な人物像が湧いてこないし、テキストデータには表されていない行間を読み取ることもできない。また、電子メールの送信者にとっては、自己の人物像を受信者に湧かせたり、テキストデータでは表せない行間を受信者に読み取らせることができない。さらには、二者択一的な音声選択では面白みに欠ける。これは、フロッピーディスクなどの記憶媒体で、テキストベースの電子データをやり取りする場合も同様である。

30

#### 【0011】

従って、電子メールなどの送信者（テキストデータの作成者）の具体的な人物像などが湧き、あるいは、テキストデータでは表せない行間を読み取ることなどができ、さらには、多様な音声を楽しむことなどができ、しかも、電子データのデータサイズを少なくして送受信して音声合成することができる方が望まれる。

40

そこで、本発明は、上記課題を解決することを主たる目的とする。

#### 【0012】

##### 【課題を解決するための手段】

本発明の音声合成装置（請求項1）では、送信側装置からは、送信者の音声と類似度が最大になる音声データを受信側装置の音声記憶手段（音声データベース）の中から指定する指定データと、前記指定データにより指定される音声データに修正を施す特徴量修正データを送信する。一方、音声合成用の受信側装置では、これらデータを受信すると、音声記憶手段から受信した指定データに基づいて対応する音声データを読み出し、このように読み出した音声データに受信した特徴量修正データで修正を施して修正音声データを作成し、音声合成を行う。

50

この構成においては、送信者の音声に特徴が類似する所定の音声データを指定する指定データで、音声記憶手段に記憶されている音声データが読み出される。そして、この読み出した音声データに対して、より前記音声に近似するように音声データの修正を行う。このため、送信者の音声（入力音声）により近い音声で音声合成し、データを読み上げることが可能になる。

ちなみに、例えば実施形態の「音声データベース」は、請求項の「音声記憶手段」に相当する。また、実施形態の「音声規範データ作成部 1 1 a」は、類似度が最大となる所定のデータの選択、及び特徴量修正データの作成を行う請求項の「処理手段」に相当する。また、実施形態の「音声合成部 2 1 a」は、修正音声データの作成、及び音声合成を行う請求項の「処理手段」に相当する。また、実施形態の「DB指定データ」は請求項の音声データを指定する「指定データ」に相当する。また、実施形態の「テキストデータ」及び「音声インデックスデータ」は、請求項における「修正音声データに基づいてデータを読み上げ...」の「データ」に相当する。

10

#### 【0013】

また、本発明は、音声記憶手段の中から音声合成側で使用すべき音声データを指定する指定データと、この指定データにより指定した音声データを修正する特徴量修正データを音声合成側に送信し、送信側と同じ内容の音声記憶手段を備える音声合成側で受信したデータ（指定データ、特徴量修正データ）に基づいて音声合成の対象となるデータ（実施形態におけるテキストデータ、音声インデックスデータ）を読み上げるようにした音声合成方法におけるデータ作成方法であり、入力音声に対して類似度が最大となる所定の音声データを決定すると共に、この所定の音声データを指定する指定データを決定し、さらに、この決定した所定の音声データを前記入力音声により近似するように当該音声データを修正する特徴量修正データを作成することを特徴とする（請求項 4）。

20

また、本発明は、送信側の音声記憶手段と同じ内容の音声記憶手段を備える音声合成側の装置を用い、受信した指定データと特徴量修正データに基づいて音声合成の対象となるデータを読み上げるようにした音声合成方法である（請求項 5）。

この構成（請求項 4・請求項 5）においては、前記した音声合成装置（請求項 1～請求項 3）と同様に入力音声に類似度が最大となる音声データが音声合成側で読み出されるように指定することができる。かつ、この類似度が最大となる音声データに対して入力音声により近似するように特徴量修正データに修正を施すことができる。このため、入力された音声（入力音声、送信者の音声）により近い音声で音声合成を行うことが可能になる。

30

#### 【0014】

##### 【発明の実施の形態】

以下、本発明の音声合成装置、音声合成方法におけるデータ作成方法、及び音声合成方法の実施の形態を、図面を参照して詳細に説明する。なお、本発明の実施の形態は、（1）1 指定データ（以下「DB指定データ」という）及び音素の特徴量修正データ（以下「特徴量修正データ」という）を 2 音声インデックスデータとは別に持つ第 1 の実施形態と、（2） 1 DB指定データ及び特徴量修正データを 2 音声インデックスデータと一緒に持つ第 2 の実施形態の、2 つの実施形態に分けて説明する。

#### 【0015】

##### 第 1 の実施形態；

先ず、第 1 の実施形態の説明を行う。本実施形態では、電子データとして、電子メールを送受信する。

図 1 は音声合成方法の送信者側及び受信者側で行う処理を、図 2 は音声規範データの作成を、図 3 は受信者側で行う処理を、図 4 はハードウェアの構成をそれぞれ示す。

#### 【0016】

##### ハードウェア構成

本実施形態の音声合成方法は、図 4 に示すように、すくなくとも送信者側装置（送信側装置、送信側）10、受信者側装置（受信側装置、受信側、音声合成側）20 及び電子データ転送手段 30 をもって実施される。

40

50

## 【 0 0 1 7 】

電子メールの送信者に係る送信者側装置 1 0 は、中央処理装置 1 1、外部記憶装置 1 2、入出力装置 1 3、D S U (Digital Service Unit) 1 4 がバス 1 5 に接続された構成を有する。中央処理装置 1 1 は、送信者側装置 1 0 を統括的に制御すると共に、入力された音声から音声合成の規範となる音声規範データ M D を作成する音声規範データ作成部 1 1 a (音声規範データ作成エンジン) 及びテキストデータ作成部 1 1 b を有する。外部記憶装置 1 2 は、各種データやプログラムを格納すると共に、音声規範データ M D を作成するための音声データ D n を多数記憶した音声データベース D B を有する。音声データ D n のそれぞれは、図 2 などに示すように音素の集合データである。入出力装置 1 3 には、テキストデータ T X を入力するためのキーボードや音声入力を行うためのサウンドシステムなどが、I / O 装置を介して接続されている(図示外)。また、D S U 1 4 は、I S D N などのデジタル回線のユーザ側に設置する電子データの送受信装置である。この送信者側装置 1 0 は、携帯可能あるいは車両設置可能なものであってもよい。

10

## 【 0 0 1 8 】

電子メールの受信者に係る受信者側装置 2 0 は、中央処理装置 2 1、外部記憶装置 2 2、入出力装置 2 3、D S U 2 4 がバス 2 5 に接続された構成を有する。中央処理装置 2 1 は、受信者側装置 2 0 を統括的に制御する他、送信者側装置 1 0 から送られてくる音声規範データ M D に基づいてテキストデータ T X を音声合成する音声合成部 2 1 a (音声合成エンジン) を有する。外部記憶装置 2 2 は、各種データやプログラムを格納すると共に、音声規範データ M D に基づいて音声合成を行うための音声データ D n (音素の集合データ) を多数記憶した音声データベース D B を有する。入出力装置 2 3 には、キーボードや合成された音声を出力するサウンドシステムなどが、I / O 装置を介して接続されている(図示外)。また、D S U 2 4 は、I S D N などのデジタル回線のユーザ側に設置する電子データの送受信装置である。なお、受信者側装置 2 0 の音声データベース D B と送信者側装置 1 0 の音声データベース D B は、同じ内容のものである。この受信者側装置 2 0 は、送信者側装置 1 0 と同様に携帯可能あるいは車両設置可能なものであってもよい。

20

## 【 0 0 1 9 】

電子データ転送手段 3 0 は、ここでは I S D N などのデジタル通信回線である。なお、電子データ転送手段 3 0 は、アナログの公衆電話回線などでもよい。有線か無線かは問わない。また、フロッピーディスクなどの電子データ記憶媒体を宅配や郵送する電子データ転送手段であってもよい。あるいは、インターネット上のサイトから電子データが配信(ダウンロード)される態様や、雑誌などで配布される態様など、メディアを介した態様も電子データ転送手段 3 0 に含まれる。いずれの態様も、本発明における「送信」及び「受信」に該当する。

30

## 【 0 0 2 0 】

## 音声合成方法

本実施形態の音声合成方法は、図 1 に示すように、電子メールの送信者側で行う処理と受信者側で行う処理とに分けられる。

送信者側では、音声規範データの作成(S 1)、テキストデータの作成(S 2)、そして、作成された音声規範データ及びテキストデータの送信が行われる(S 3、電子メールの送信)。なお、ステップ S 1 の音声規範データの作成とステップ S 2 のテキストデータの作成は、その順序が逆でもよい。一方、受信者側では、音声規範データ及びテキストデータの受信(S 5、電子メールの受信)、音声データの読み込み及び修正(S 6)、音声インデックスデータの作成(S 7)、そして、音声合成(S 8)が行われる。なお、ステップ S 6 の音声データの読み込み及び修正とステップ S 7 の音声インデックスデータの作成は、その順序が逆でもよい。また、受信者側では、音声規範データ M D の記憶が、必要に応じて行われる。

40

以下、送信者側の処理と受信者側の処理とに分けて詳細に説明する。

## 【 0 0 2 1 】

〔送信者側の処理〕

50

第 1 の実施形態の音声合成方法では、電子メールの送信者は、図 4 で例示した送信者側装置 10 を用いて、音声規範データ MD 及びテキストデータ TX を作成し、これを電子メールで受信者に向けて送信する。なお、本実施形態では、電子メールの送信者が、送信者自身に似せた声で音声合成させる音声規範データ MD を作成するものとする。

ちなみに、音声規範データ MD は、DB 指定データ  $x_n$  及び特徴量修正データ FD からなる (図 3 参照)。特徴量修正データ FD は、特許請求の範囲における音素の変換データに該当する。

#### 【0022】

(1) 音声規範データの作成；  まず、送信者は自分の声を送信者側装置 10 に入出力装置 13 から音声入力する (S11)。音声入力するのは、音素系列に偏りのない文書 (定型文) もしくは 50 音である。入力した音声は、音声規範データ作成部 11a で特徴抽出され、入力音声の音素ごとの特徴量  $i$  が抽出される (S12)。ちなみに、本実施形態の特徴量  $i$  は、基本周波数とスペクトル包絡からなる。ここで、スペクトル包絡は音声スペクトルの概形である。なお、特徴量としては、次の表に示すようなものを使用することもできる。

#### 【0023】

#### 【表 1】

音声合成に関する特徴量

	基本属性	スペクトル上の特徴	特徴量
音韻性	音色	周波数軸上の緩やかな起伏のパターン	ホルマント周波数 ケプストラム低次成分 線形予測係数
	有声・無声	線スペクトルの有無 低域・高域エネルギー比	最大自己相関値 零交差数 ケプストラム最大値
韻律性	強さ	スペクトル密度の積分	短時間平均振幅 (電力)

#### 【0024】

次に、音声規範データ作成部 11a は、特徴抽出した入力音声の音素ごとの特徴量  $i$  と音声データベース DB が記憶する音声データ  $D_n$  ( $n=1 \cdot 2 \cdot \dots \cdot n$ ) の音素ごとの特徴量の類似度を計算し、全体を比較する (S13)。そして、入力音声と類似度が最大である特定の音声データ  $D_{xn}$  を、音声データベース DB の中から決定する (S14)。以後この音声データ  $D_{xn}$  を標準音声データ  $D_s$  とする。ちなみに、音声データベース DB の中から標準音声データ  $D_s$  を特定するためのデータ (アドレスなど) が、DB 指定データ  $x_n$  になる。なお、受信者側装置 20 と送信者側装置 10 の音声データベース DB は同じものである。

#### 【0025】

続いて、音声規範データ作成部 11a は、標準音声データ  $D_s$  が持つ各音素に対して、先に得られた各音素ごとの特徴量  $i$  を基に、特徴量修正データ FD を作成する (S15)。この特徴量修正データ FD と DB 指定データ  $x_n$  とで、送信者の声の情報である音声規範データ MD が構成される。

#### 【0026】

ここで、特徴量修正データ FD の作成方法をさらに説明する。

1 基本周波数の修正；  音素  $k$  の特徴量である基本周波数  $f_k$  と入力された音声の基本周波数  $f$  の差が範囲内  $|f_k - f| < a$  で最小となるような  $f_k$  を求め、 $f_k$  を基本周波数  $f_k$  として置き換える。つまり、ある範囲内 (範囲  $a$ ) において、 $f_k$  を  $f$  に近似させるため、あるいは等しくするため、 $f_k$  を計算する ( $f_k$  は音素  $k$  の基本周波数  $f_k$  と入力音声の基本周波数  $f$  の差になる)。ここで範囲  $a$  は、 $|f_k - f|$  の数値が大きくなりすぎないようにするための閾値でもある。この閾値により修正した  $f_k$  ( $= f_k$ ) が修正す

10

20

30

40

50

る前の  $f_k$  から大きくかけ離れた周波数にならず、修正する前の  $f_k$  のもつ特徴を維持する。ちなみに、修正する前の  $f_k$  が “あ” と聞こえるための基本周波数であるとする、修正した  $f_k (= f_{k'})$  を使用した音声 “あ” と聞こえる限界の範囲 ( 閾値 ) が  $a$  である。

【 0 0 2 7 】

2 スペクトル包絡の修正； 音素  $k$  のスペクトル包絡を表す特徴量を入力音声の特徴量に近似させる。ただし、音素  $k$  のスペクトル包絡の修正量は、所定の閾値範囲内とする。

【 0 0 2 8 】

( 2 ) テキストデータの作成； 音声合成の対象となるテキストデータ  $T_X$  は、入出力装置 1 3 から文字をキー入力し、テキストデータ作成部 1 1 b で作成される ( 図 1 ( a ) の S 2 ) 。ちなみに、テキストデータ  $T_X$  の内容は 「 こんにちは... 」 である。 10

【 0 0 2 9 】

( 3 ) 音声規範データ及びテキストデータの送信； 作成された音声規範データ  $M_D$  及びテキストデータ  $T_X$  は、電子メールにより受信者宛てに送信される ( 図 1 ( a ) の S 3 ) 。このうち音声規範データ  $M_D$  は、電子メールの添付ファイルとして送信される。なお、送信される音声規範データ  $M_D$  やテキストデータ  $T_X$  は、文字や数字のデータであり、データサイズは  $w_a_v_e$  形式や  $M_P3$  形式の音声データなどに比べて遥かに小さい。したがって、送信者側及び受信者側のハードウェアに与える負担が小さいと共に、送受信に要する時間を小さくすることができる。 20

【 0 0 3 0 】

〔 受信者側の処理 〕

第 1 の実施形態の音声合成方法では、電子メールの受信者は、図 4 で例示した受信者側装置 2 0 を用いて、次のように音声合成を行う ( 図 3 参照 ) 。

【 0 0 3 1 】

( 1 ) 標準音声データの読み込み・修正； 受信者側装置 2 0 の  $D_S U 2 4$  は、電子データ転送手段 3 0 を介して電子メールを受信する ( S 5 1 ) 。次に、受信者側装置 2 0 の音声合成部 2 1 a は、電子メールに添付された音声規範データ  $M_D$  のうち、 $D_B$  指定データ  $x_n$  に基づいて音声データベース  $D_B$  を検索し、該当する音声データ  $D_{x_n}$  ( つまり標準音声データ  $D_s$  ) を読み込む。続いて、音声規範データ  $M_D$  のうちの特徴量修正データ  $F_D$  に基づいて、読み込んだ標準音声データ  $D_s$  に修正を施し修正音声データ  $D_{s'}$  を作成する ( S 5 2 ) 。修正音声データ  $D_{s'}$  は、送信者の音声の特徴が反映され、送信者の音声に近似した音声 ( 音素 ) になっている。 30

【 0 0 3 2 】

( 2 ) 音声インデックスデータの作成； 受信した電子メールのうち 「 こんにちは... 」 という内容のテキストデータ  $T_X$  について、音声合成部 2 1 a でテキスト解析及び音律予測を行い、音声インデックスデータ  $I_D$  を作成する ( S 5 3 ) 。音声インデックスデータ  $I_D$  は、図 3 の吹き出しに示す構造を有する。なお、テキスト解析は、音声合成して出力したいテキストデータのテキスト解析を行い、べた書きのテキストを単語ごと ( つまり形態素ごと ) に分割して、アクセントの位置やポーズ長などを決定する。また、音律予測は、前記テキスト解析のデータにもとづいて、音声合成して出力したい音素系列の韻律パラメータを音素ごとに予測する。 40

ちなみに、音声インデックスデータ  $I_D$  に記述された音素 「  $k, o, n, n, i, ch, i, w, a, \dots$  」 は、音素指定データである。

【 0 0 3 3 】

( 3 ) 音声合成； 前記 ( 1 ) で作成した修正音声データ  $D_{s'}$  と前記 ( 2 ) で作成した音声インデックスデータ  $I_D$  に基づいて音声合成を行う ( S 5 4 ) 。具体的には、修正音声データ  $D_{s'}$  の修正済み音素を用いて、音声インデックスデータ  $I_D$  の音素指定データの順序どおりに音素系列を組み上げ (  $k, o, n, n, i, ch, i, w, a, \dots$  ) 、その系列に基づいて音声合成を行いテキストデータ  $T_X$  を読み上げる。このようにすることで、受信者は、送信 50



者の音声（送信者の音声と同等あるいは近似した音声）で電子メールの内容を聞くことができる。

【0034】

なお、音声規範データMDは個人に特有のものであるが、この音声規範データMDを外部記憶装置22に記憶しておけば、次回、同じ送信者から電子メールを受信する際に、音声規範データMDが添付されていなくとも、電子メールの内容をこの送信者の声で音声合成して読み上げることができる。この際、送信者のe-mailアドレスや電話番号などのユニークな番号をアドレスとして、音声規範データMDを保存しておくのがよい。同じ送信者からの電子メールに対して、当該送信者用の音声規範データMDを即座に呼び出して、電子メールの内容を音声合成して読み上げることができるからである。

10

【0035】

このように音声合成することにより、受信者は、テキストデータでは表せない行間を読み取ることなどことができ、かつ送受信される電子データのデータサイズも音声ファイルに比べてはるかに少なくすることができる。また、音声規範データを受信者側装置に記憶することにより、同じ送信者が再度電子メールを送る際には、テキストデータ（音声インデックスデータ）を送信するだけでよいので、送受信される電子データのデータサイズをさらに少なくすることができる。さらに、送信者は、音声規範データを作成する際に、テキストデータの内容そのものを読み上げる必要がなく、定型文分や50音を読み上げればよいだけである。したがって、オフィス空間で音声規範データを作成しても、テキストデータの内容そのものを読み上げるのと異なり、周囲に違和感を与えない。加えて、一度作成した音声規範データは、繰り返し使用することができるので、電子メール送信のたびに定型文や50音を読み上げる必要はない。ましてや、音声ファイルを電子メールで送信するのと異なり、テキストデータの内容が異なるような場合でも、その都度テキストデータの内容そのものを読み上げる必要がない。さらに、音声ファイルと異なり、テキストファイルの修正・編集は極めて容易である。また、受信者は、電子メールの内容確認が、自動車の運転などの他の作業と並行して行なえる。

20

【0036】

なお、上記実施形態では、音声インデックスデータの作成を受信者側で行ったが、送信者側で行って電子メールに添付するようにしてもよい。送信者側で音声インデックスデータを作成することにより、送信者が自由に音素の長さ・高さ・強さなどを編集することができ、より自分の意図した音声（アクセントやイントネーションなど）を合成させることが可能になるからである。このように音声インデックスデータを送信者の側で作成すると共に、電子メールに添付して送信する場合は、電子メールには音声合成の対象となるテキストデータが存在しなくともよい。音声インデックスデータの音素指定データの順序どおりに音素系列を組み上げて、音声合成することができるからである。この場合、音声インデックスデータはテキストデータに由来するものであり、この音声インデックスデータが特許請求の範囲の記載におけるテキストデータに該当する。

30

【0037】

第2の実施形態；

次に、第2の実施形態の音声合成方法の説明を行う。なお、第1の実施形態と同一の部材・要素などについては、第1の実施形態で使用した図面を参酌すると共に同一の符号を付し、その説明を省略する。ここで、図5は音声合成方法の送信者側及び受信者側で行う処理を、図6は音声合成方法における各データの間係を、図7は受信者側で行う処理をそれぞれ示す。

40

この第2の実施形態では、音声インデックスデータを送信者側で作成すると共に、音声インデックスデータと音声規範データを併記した書式の併合音声規範データを、電子メールで受信者側に送信する。

【0038】

ハードウェア構成

第2の実施形態の音声合成方法は、第1の実施形態に使用されるハードウェアをそのまま

50

適用することができるので、ハードウェア構成の説明を省略する。但し、第2の実施形態では、音声インデックスデータIDを送信者側装置10で作成する。したがって、送信者側装置10の音声規範データ作成部21aは、音声インデックスデータIDの作成が可能なものである。

#### 【0039】

##### 音声合成方法

第2の実施形態の音声合成方法も第1の実施形態の音声合成方法と同様、電子メールの送信者側で行う処理と受信者側で行う処理とに分けられる。

送信者側では、音声規範データの作成(S101)、テキストデータの作成(S102)、テキストデータにもとづいての音声インデックスデータの作成(S103)、音声規範データと音声インデックスデータの併合、つまり併合音声規範データの作成が行われ(S104)、そして、併合音声規範データの送信が行われる(S105、電子メールの送信)。一方、受信者側では、併合音声規範データの受信(S106、電子メールの受信)、音声データの読み込み及び修正(S107)、そして、音声合成が行われる(S8)。

以下、送信者側の処理と受信者側の処理とに分けて詳細に説明する。

#### 【0040】

##### 〔送信者の処理〕

(1) 音声規範データの作成、(2) テキストデータの作成、(3) 音声インデックスデータの作成を、それぞれ行う。それぞれの作成手順は、第1の実施形態で説明したのと同じであるので、その説明を省略する。なお、前記の通り、音声インデックスデータIDは、送信者側で作成する。この点第1の実施形態と異なる。

#### 【0041】

(4) 併合音声規範データの作成； 音声規範データMDと音声インデックスデータIDを併合して併合音声規範データCMDを作成する。併合音声規範データCMDは、図6(c)に示すように、DB指定データxnと音声インデックスデータIDの内容全部とこの音声インデックスデータIDの音素指定データに対応した特徴量修正データFDを有する。

#### 【0042】

(5) 併合音声規範データの送信； 作成された併合音声規範データCMDは、電子メールにより受信者宛てに送信される(図5(a)のS105)。送信される併合音声規範データCMDは、文字や数字のデータでありデータサイズは、wave形式やMP3形式の音声データなどに比べて遥かに小さい。したがって、送信者側及び受信者側のハードウェアに与える負担が小さいと共に、送受信に要する時間を小さくすることができる。なお、電子メールには、テキストデータTXは不用である。併合音声規範データCMDが有する音素指定データに基づいて音素系列を組み上げ、音声合成することができるからである。但し、受信者が目視で電子メールの内容を確認できるように、テキストデータTXを電子メールに記載してもよい。

#### 【0043】

##### 〔受信者側の処理〕

第2の実施形態の音声合成方法では、電子メールの受信者は、図4で示した受信者側装置20を用いて、次のように音声合成を行う(図7参照)。

#### 【0044】

(1) 標準音声データの読み込み・修正； 受信者側装置20のDSU24は、電子データ転送手段30を介して電子メールを受信する(S151)。次に、受信者側装置20の音声合成部21aは、電子メールに添付された併合音声規範データCMDのうち、DB指定データxnに基づいて音声データベースDBを検索し、該当する音声データDxn(つまり標準音声データDs)を読み込む。続いて、併合音声規範データCMDのうち、特徴量修正データFDに基づいて読み込んだ標準音声データDsに修正を施し修正音声データDs'を作成する(S152)。修正音声データDs'は、送信者の音声の特徴が反映され、送信者の音声に近似した音声(音素)になっている。

10

20

30

40

50

## 【 0 0 4 5 】

( 2 ) 音声合成 ; 作成した修正音声データ D s ' と併合音声規範データ C M D の音声インデックスデータ I D に由来する部分のデータに基づいて音声合成を行う ( S 1 5 3 ) 。具体的には、修正音声データ D s ' の修正済み音素を用いて、併合音声規範データ C M D の音素指定データの順序どおりに音素系列を組み上げ ( k , o , n , n , i , c h , i , w , a . . . ) 、その系列に基づいて音声合成を行い送信者がステップ S 1 0 2 で作成したテキストデータ T X を読み上げる。このようにすることで、受信者は、送信者の音声 ( 送信者の音声と同等あるいは近似した音声 ) で電子メールの内容を聞くことができる。

## 【 0 0 4 6 】

このように音声合成することにより、第 1 の実施形態と同様にテキストデータでは表せない行間を読み取ることなどができ、かつ送受信される電子データのデータサイズも音声ファイルを送受信するのに比べて少なくすることができる。また、特徴量修正データを全ての音素に対して準備する必要がないので、第 1 の実施形態よりもデータサイズを小さくすることも可能である。なお、本実施形態において、受信者に送られる併合音声規範データのうち、音声インデックスデータに由来する部分が、特許請求の範囲におけるテキストデータに該当する。

## 【 0 0 4 7 】

以上説明した本発明は、上記した実施の形態に限定されることなく広く変更実施することができる。

例えば、似せる声は送信者自身のものでなくと、家族や友人、あるいは他人のものでもよい。受信者にとって聞いたことのない声であっても、多様な音声で電子データの内容 ( テキストデータ ) を聞くことができ、面白みを味わうことができるからである。

## 【 0 0 4 8 】

## 【 発明の効果 】

本発明 ( 音声合成装置、音声合成方法におけるデータ作成方法、及び音声合成方法 ) によれば、送信者などの声で電子メールの内容を聞くことができるので、受信者には送信者などの具体的人物像が湧いてきたり、テキストデータでは表せない行間を読み取ったりすることができる。また、受信者は、多様な声で面白味を味わいながらテキストデータの内容を聞くことができる。しかも、音声ファイルを送受信するのと異なり、データサイズを小さくすることができる。さらには、音声ファイルを送受信するのと異なり、送信者は、テキストデータの内容そのものを読み上げる必要がない。加えて、指定データ及び特徴量修正データ ( 音声規範データ ) は、電子データとして構成されるため再利用が可能である。かつ、指定データ及び特徴量修正データは、テキストデータの内容に影響されないので、異なるテキストデータを作成する場合においても、そのテキストデータに合わせて指定データ及び特徴量修正データを再度作成する必要がない。

## 【 0 0 4 9 】

特に、請求項 2 に記載の発明によれば、受信側装置で音声インデックスデータを作成するので、送信側装置での負荷を低減することができる。また、請求項 3 に記載の発明によれば、送信側装置で音声インデックスデータを作成するので、受信側装置での負荷を低減することができる。また、本発明によれば、送受信されるデータサイズを確実に小さくすることができると共に、確実に送信者が意図する音声で受信者側に音声合成させてテキストデータの内容を聞かせることができる。また、請求項 4 の発明によれば、音声合成側で使用する指定データ、特徴量修正データを適切に決定及び作成することができる。また、請求項 5 に記載の発明によれば、指定データ、特徴量修正データに基づいて、より入力音声に近似させて音声合成を行い、読み上げ対象のデータを読み上げることができる。

## 【 図面の簡単な説明 】

【 図 1 】 第 1 の実施形態の音声合成方法の処理を示す流れ図であり、( a ) は送信者側の処理を、( b ) は受信者側の処理を示す。

【 図 2 】 第 1 の実施形態の音声合成方法の送信者側で行う音声規範データの作成をデータの構成と共に示す流れ図である。

10

20

30

40

50

【図3】 第1の実施形態の音声合成方法の受信者側で行う処理をデータの構成と共に示す流れ図である。

【図4】 第1の実施形態の音声合成方法が適用されるハードウェアの構成を示すブロック図である。

【図5】 第2の実施形態の音声合成方法の処理を示す流れ図であり、(a)は送信者側の処理を、(b)は受信者側の処理を示す。

【図6】 第2の実施形態の音声合成方法における各データの間関係を説明する図であり、(a)は音声インデックスデータの構成を、(b)は音声規範データの構成を、(c)は併合音声規範データの構成を示す。

【図7】 第2の実施形態の音声合成方法の受信者側で行う処理をデータの構成と共に示す図である。

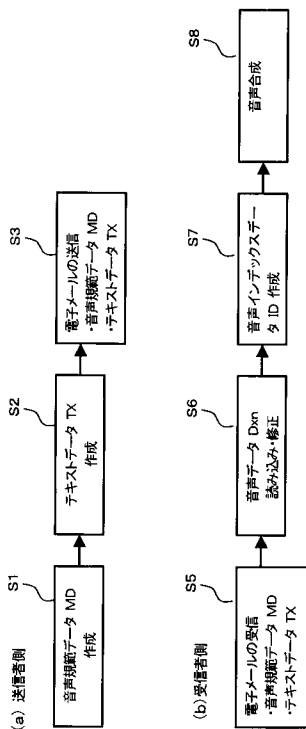
10

【符号の説明】

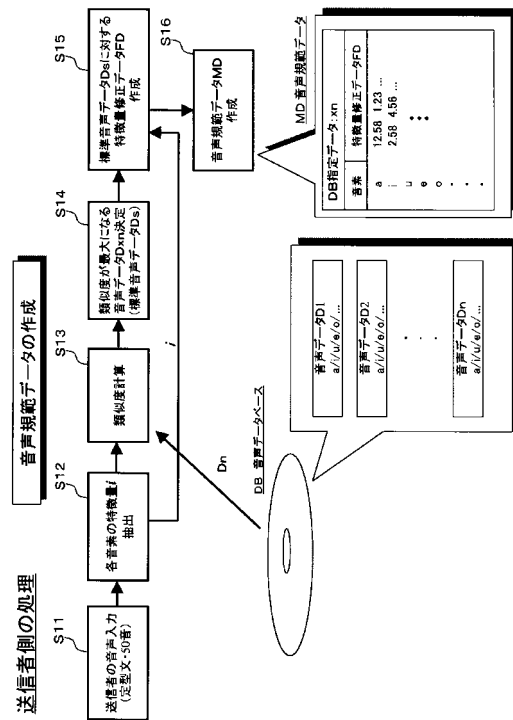
- 10 ... 送信者側装置(送信者)
- 20 ... 受信者側装置(受信者)
- MD ... 音声規範データ
- TX ... テキストデータ
- DB ... 音声データベース
- Dn ... 音声データ
- Ds・Dxn ... 標準音声データ(特定の音声データ)
- xn ... DB指定データ(音声データベース指定データ)
- FD ... 特徴量変換データ(音素の変換データ)

20

【図1】

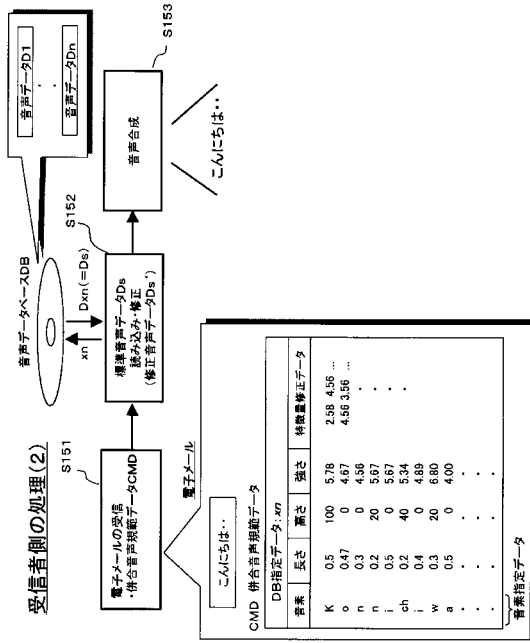


【図2】





【図7】



フロントページの続き

合議体

審判長 原 光明

審判官 清水 正一

審判官 堀井 啓明

(56)参考文献 特開平5 - 260082 (JP, A)

特開平9 - 50286 (JP, A)

特開平7 - 319495 (JP, A)