(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第3585190号
(P3585190)
(45)発行日　平成16年11月4日(2004.11.4)　(24)登録日　平成16年8月13日(2004.8.13)

(51)Int.Cl.[7]　F I
G10L 13/00　G10L 3/00　S
G10L 21/02　G10L 3/02　F
G10L 21/06

請求項の数 10　(全 8 頁)

(21)出願番号　特願平7-180129
(22)出願日　平成7年7月17日(1995.7.17)
(65)公開番号　特開平9-34483
(43)公開日　平成9年2月7日(1997.2.7)
審査請求日　平成14年7月11日(2002.7.11)

(73)特許権者　000001007
キヤノン株式会社
東京都大田区下丸子３丁目３０番２号
(74)代理人　100077481
弁理士　谷　義一
(74)代理人　100088915
弁理士　阿部　和夫
(72)発明者　野口　利之
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
(72)発明者　大洞　恭則
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　山下　剛史

最終頁に続く

(54)【発明の名称】音声出力装置および方法

(57)【特許請求の範囲】

【請求項１】

少なくとも口画像を含む複数の画像要素を格納する画像データベースから読み出した画像要素を合成して顔画像を生成する生成手段と、

前記生成手段で生成した顔画像を表示するよう制御する表示制御手段と、

前記表示制御手段で表示された顔画像中の口画像の面積を変更する変更手段と、

前記変更手段で変更した口画像の面積に対応づけた音量の音声を出力する出力手段とを備えたことを特徴とする音声出力装置。

【請求項２】

前記画像データベースは、前記口画像の、音量がゼロであることを示す第１の形態と、音量が最大であることを示す第２の形態とを格納し、

前記変更手段は、前記第１の形態、前記第２の形態、または前記第１及び第２の形態を補間して得られた中間形態に口画像を変更することを特徴とする請求項１記載の音声出力装置。

【請求項３】

前記画像データベースは前記口画像の複数の形態を格納し、

前記変更手段は、前記複数の形態から現在表示している形態とは異なる形態を選択して変更することを特徴とする請求項１記載の音声出力装置。

【請求項４】

テキストデータを格納するテキスト保持手段と、

前記テキスト保持手段から読み出した前記テキストを音声に変換する変換手段を更に備え、
前記音声出力手段は、当該変換手段により変換された音声を出力することを特徴とする請求項１記載の音声出力装置。
【請求項５】
少なくとも口画像を含む複数の画像要素を格納する画像データベースから読み出した画像要素を合成して生成された顔画像を表示するよう制御する表示制御ステップと、
前記表示制御ステップで表示された顔画像中の口画像の面積を変更する変更ステップと、
前記変更ステップで変更した口画像の面積に対応づけた音量の音声を出力する出力ステップとを備えたことを特徴とする音声出力方法。
【請求項６】
前記画像データベースは、前記口画像の、音量がゼロであることを示す第１の形態と、音量が最大であることを示す第２の形態とを格納し、
前記変更ステップは、前記第１の形態、前記第２の形態、または前記第１及び第２の形態を補間して得られた中間形態に口画像を変更することを特徴とする請求項５記載の音声出力方法。
【請求項７】
前記画像データベースは前記口画像の複数の形態を格納し、
前記変更ステップは、前記複数の形態から現在表示している形態とは異なる形態を選択して変更することを特徴とする請求項５記載の音声出力方法。
【請求項８】
テキストデータを予めテキスト保持手段に保持しており、
前記テキスト保持手段から読み出した前記テキストを音声に変換する変換ステップを更に備え、
前記出力ステップは、前記変換ステップにより変換された音声を出力することを特徴とする請求項５記載の音声出力方法。
【請求項９】
請求項５乃至８に記載の音声出力方法をコンピュータに実行させるための制御プログラム。
【請求項１０】
請求項９に記載の制御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】
【０００１】
【発明の属する技術分野】
本発明は、グラフィカルユーザインタフェースを用いて音声出力の音量を制御する音声出力装置および方法に関する。
【０００２】
【従来の技術】
従来、コンピュータで音声を出力する場合に、機械的なボリュームスイッチや、キーボードから入力する音量調整コマンドや、画面に表示したスライドバーなどにより、音量を調整している。
【０００３】
【発明が解決しようとする課題】
しかしながら、機械的なボリュームスイッチを設けるた場合は、装置の価格が高くなる。音量調節コマンドを用いた場合は、操作者がそのコマンドを覚えなくてはならない。スライドバーも操作者に対して直感的でなく、特に小児にとっては、操作方法が理解しにくいという課題があった。
【０００４】
そこで本発明は、このような課題を解決し、音声出力の音量を容易に調整することのできる音声出力装置および方法を提供することを目的とする。

【０００５】
【課題を解決するための手段】
このような目的を達成するために、本発明は、少なくとも口画像を含む複数の画像要素を格納する画像データベースから読み出した画像要素を合成して顔画像を生成する生成手段と、
前記生成手段で生成した顔画像を表示するよう制御する表示制御手段と、
前記表示制御手段で表示された顔画像中の口画像の面積を変更する変更手段と、
前記変更手段で変更した口画像の面積に対応づけた音量の音声を出力する出力手段とを備えたことを特徴とする。
【００１０】
また本発明は、少なくとも口画像を含む複数の画像要素を格納する画像データベースから読み出した画像要素を合成して生成された顔画像を表示するよう制御する表示制御ステップと、
前記表示制御ステップで表示された顔画像中の口画像の面積を変更する変更ステップと、
前記変更ステップで変更した口画像の面積に対応づけた音量の音声を出力する出力ステップとを備えたことを特徴とする。
【００２０】
【発明の実施の形態】
以下、図面を参照して本発明の実施例を詳細に説明する。
【００２１】
図１は、本発明の実施例に係る音声出力装置の構成を示すブロック図である。同図において、１はユーザからコマンド及びデータを入力する入力手段であり、例えば、キーボードおよびマウス等のポインティングデバイスから構成される。２は入力手段１からの入力に対応して画像の表示を制御する表示制御手段、３は表示制御手段２から出力された画像を表示する表示手段である。４は表示手段に表示する顔などの画像の要素を格納する画像データベースである。
【００２２】
図２に、画像データベース４に格納した画像要素の一例を示す。画像データベース４は、人の顔の画像の各構成要素としての「目」、「鼻」、「口」、「輪郭」、「髪」の、画像要素ファイルを有する。各画像要素ファイルには、図２に示すように各構成要素のデフォルト画像の形態およびデフォルト画像を変形した変形画像の形態が格納されている。デフォルト画像および変形画像はスケーラブルであり、入力手段１からの入力等に基づいて拡大または縮小することができる。
【００２３】
表示制御手段２は、各画像要素ファイルから必要な画像を取り出し、必要に応じて大きさを変更し、各画像を合成して表示手段３に表示する。また表示制御手段２は、１つの画像ファイルに格納された１つの画像要素の、複数の形態を補間することにより、複数の形態の中間的な画像を形成することができる。
【００２４】
テキスト保持手段５は、音声合成を行う対象となる入力テキストを保存する。入力手段１から音声合成開始コマンドが入力されると、中央制御手段１０は、入力テキスト保持手段６からテキストを読みだし、音声合成手段６に転送する。すると音声合成手段６は、転送されたテキストに基づいて音声を合成する。
【００２５】
中央制御手段１０は、入力手段１から入力された音量指示に従って音量制御手段７を制御する。音量制御手段７は、音声合成手段６から入力された音声を、中央制御手段１０からの制御に従って拡大し、スピーカ８に出力する。
【００２６】
図３は、図１の中央制御手段１０の動作を示すフローチャートである。また図４に、図３のフローチャートに示す動作に従って、表示手段３に表示される顔の画像の例を表す。両

図を用いて本音声出力装置の動作を説明する。ステップ（Ｓと略記する）１０１（図３）では、表示制御手段２を介して、画像データベース４に格納された画像ファイルから「目」、「鼻」、「口」、「輪郭」および「髪」の各画像データ要素のデフォルト画像を読み出し、読み出したデフォルト画像を合成して、画像４１（図４）を表示手段３に表示する。

【００２７】

Ｓ１０２では、表示した顔画像の中の「一部の画像要素」の一例としての、口の部分がポインティングデバイスにより指定されているかどうかを判断する。口の部分は、例えば画像４２（図４）に示すようにマウスのポインタを口の上に移動してマウスのボタンを押すことにより指定することができる。

【００２８】

中央制御装置１０は、ポインタが口の部分にあることおよびマウスボタンが押されていることを条件に、口が指定されたと判断することができる。口の部分が指定されていない場合はＳ１０２に戻る。口の部分が指定されていた場合はＳ１０３に進み、口が中程度に開いた画像を画像データベース４から読み出して、口のデフォルト画像に換えて表示手段に表示する。Ｓ１０３で表示される画像の一例を画像４３（図４）に示す。

【００２９】

Ｓ１０４では、音声を中程度に拡大するように音量制御手段７を設定する。またテキスト保持手段からテキストを読みだし、音声合成手段６に転送する。すると音声合成手段６が音声を合成し、音量制御手段７が音声を中程度に拡大し、スピーカ８から音声が出力される。Ｓ１０５では、音声合成手段６が音声の出力が終了したか否かを判断する。音声の出力が終了していなければＳ１０６に進み、音声の出力が終了していればＳ１１５に進む。

【００３０】

Ｓ１０６では、口の部分がポインティングデバイスにより指定されているか否かを検出する。口の部分が指定されていない場合は、Ｓ１０５に戻る。Ｓ１０６で口の部分が指定されていると、口の画像を大きくする指示が入力されているか否かを判断する（Ｓ１０７）。口の画像は、マウスボタンを押したままポインタを口の中心から遠ざけることにより拡大できる。Ｓ１０７で口の画像を大きくする操作でなかった場合はＳ１１０に進む。

【００３１】

Ｓ１０７で口の画像を大きくする操作であった場合は、マウスの移動量に応じて、音量制御手段７を用いて音声の増幅率を高める（Ｓ１０８）。Ｓ１０９では、中程度に口が開いた画像と、口が大きく開いた画像とを補間することにより、マウスの移動量に応じて開かれた口の画像を形成して表示手段３に表示する。但し、多様な大きさに開いた口の画像を予め用意し、マウスの移動量に応じて適切な画像を読み出して表示手段３に表示しても良い。Ｓ１０９により表示される画像の一例を画像４４（図４）に示す。

【００３２】

Ｓ１１０では、口の画像を小さくする指示が入力されたか否かを判断する。口の画像は、マウスボタンを押したままポインタを口の中心へ近づける方向に移動させることにより縮小することができる。口を小さくする指示が入力された場合は、音声制御手段による音量の増幅率を、マウスの移動量に応じて小さくする（Ｓ１１１）。また表示制御手段２を介して画像データベース４から、小さな口の画像を読み出し、表示手段３に表示して（Ｓ１１２）、Ｓ１０５に戻る。

【００３３】

例えば画像４５（図４）のようにポインタが口画像から離れており、Ｓ１１０で口を小さくする指示が入力されていなかった場合は、音声の出力を停止する指示が入力されたか否かを判断する（Ｓ１１３）。音声は、例えば表示手段３に表示した顔画像の中の口画像をクリックすることにより停止できる。この場合中央制御手段１０は、Ｓ１１３で、ポインタが口画像の上にあって、かつマウスのボタンが押されたことを条件として、音声の出力を停止する指示が入力されたと判断する。

【００３４】

Ｓ１１３で出力の停止が指示されていなければＳ１０５に戻る。Ｓ１１３で出力を停止する指示が入力されていれば音声合成手段６による音声の合成を停止し（Ｓ１１４）、口画像ファイルの中からの閉じたデフォルト画像を画像データベース４の「口」画像ファイルから読み出し、既に表示している顔画像の口の部分と置き換えて表示する（Ｓ１１５）。Ｓ１１５により表示される画像を画像４６（図４）に示す。Ｓ１１５の後に、Ｓ１０２に戻る。

【００３５】

（その他）

実施例１では、ポインテインゲデバイスであるマウスを用いて口の画像を指示したが、画像の「指定手段」は、表示手段に表示された口の画像を指定できるものであれば足り、例えばペン、デジタイザなどの他の入力デバイスであってもよい。

【００３６】

図４ではポインタの一例として矢印を表示したが、ポインタは表示手段に表示された口が指定されていることを表示できれば足り、例えば図５に示す様なマイクの形のポインタを用いても良い。

【００３７】

実施例１では、音声を出力し始めると口が開いた画像に切り替えたが、口の画像の変化の方法は、これに限定されるものではない。例えば、口の部分をアニメーションのように動画として動かしてもよい。この場合、音声の大きさに応じて口の部分の動きを大きくしても良い。

【００３８】

実施例１では、ポインタで口の画像をクリックすることにより音声の合成を開始したが、音声として出力したいテキストを表わすアイコンを、口の部分に移動することにより音声の出力を開始してもよい。

【００３９】

実施例１では、音声をスピーカに出力したが、音声の「出力手段」はスピーカに限定されるものではなく、例えは音声記録デッキに出力することにより、記録テープに録昔させてもよい。

【００４０】

実施例１では、ポインティングデバイスを用いて「口」の画像を指定し、拡大し、または縮小したが、ポインティングデバイスに換えて、キーボードにより拡大又は縮小しても良い。

【００４１】

実施例１では口の画像をクリックすることにより音声を停止したが、例えば画面に表示した「停止」ボタンをクリックすることにより音声を停止しても良い。

【００４２】

本発明は、複数の装置から構成されるシステムにも単体の装置にも適用できる。また、本発明はシステム或は装置にプログラムを読み込むことによっても実施できる。そのようなプログラムを記載した記録媒体に、本出願による特許権の特許法第１０１条の効力が及ぶことは言うまでもない。

【００４３】

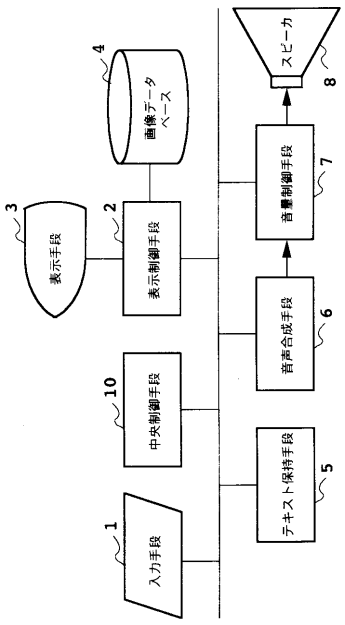
【発明の効果】

以上の説明から明らかなように、請求項１又は請求項５に記載の発明によれば、出力する音声の音量を口画像の面積に対応づけたので、ユーザにとって音量の制御が理解しやすく、制御が容易になる。特に、面積が最小の場合に音声の出力を停止することにより、音声出力の開始及び停止も容易に制御することができる。

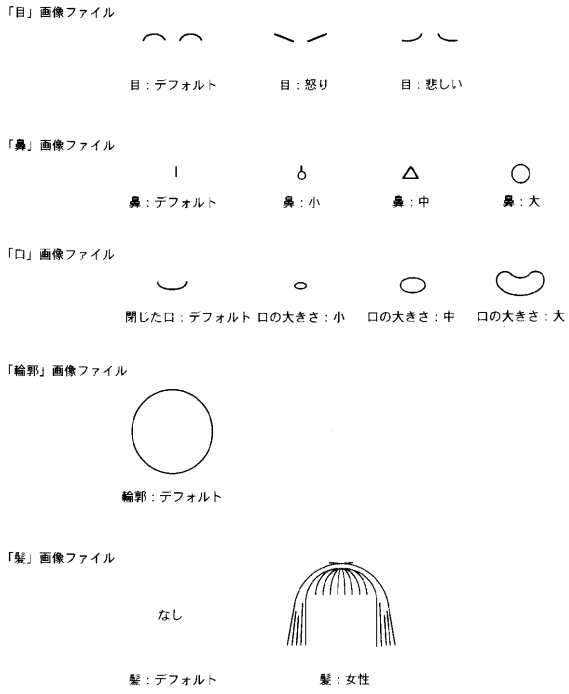
【００４４】

請求項２又は請求項６に記載の発明によれば、表示する全ての画像の形態を予め格納するのではなく、第１の形態と第２の形態との中間形態を補間して出力するので、全ての形態を予め格納した場合と比較して、必要なデータの格納容量を小さくすることができる。

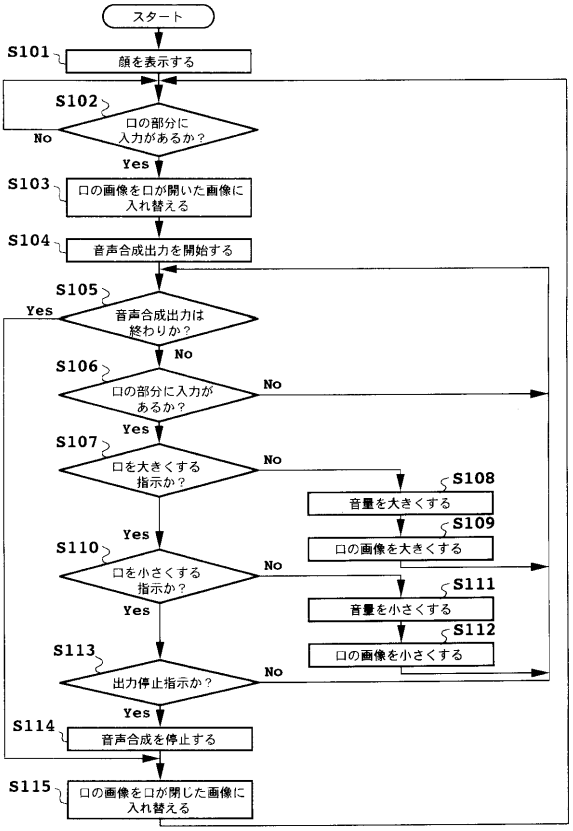
【００４５】
請求項３又は請求項７に記載の発明によれば、表示すべき複数の形態を予め格納しているので、表示する画像の面積を容易に変更することができる。
【００４７】
請求項４又は請求項８に記載の発明によれば、テキストの読み出し時の音量を、理解しやすいインタフェースで容易に制御することができる。
【図面の簡単な説明】
【図１】本発明の一実施例にかかる音声出力装置の構成を示すブロック図である。
【図２】画像データベースに格納された各データファイルの画像の例を示す説明図である。
【図３】本発明の一実施例にかかる音声出力装置の処理手順を示すフローチャートである。
【図４】図３のフローチャートに示す動作により表示される画像を説明する説明図である。
【図５】本発明の一実施例により表示される画像を説明する説明図である。
【符号の説明】
１　入力手段
２　表示制御手段
３　表示手段
４　画像データベース
５　テキスト保持手段
６　音声合成手段
７　音量制御手段
８　スピーカ
１０　中央制御手段
【図１】



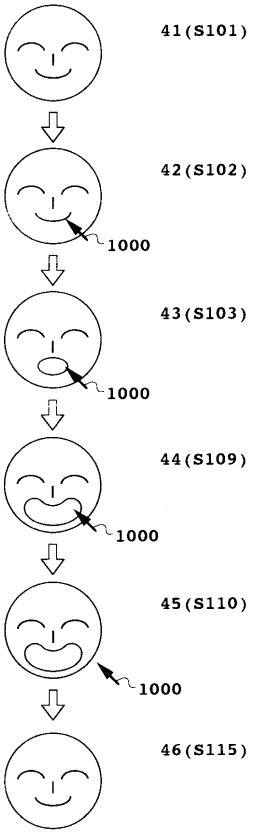
【図２】



【図３】


スタート
S101
顔を表示する
S102
口の部分に入力があるか？
No
Yes
S103
口の画像を口が開いた画像に入れ替える
S104
音声合成出力を開始する
S105
音声合成出力は終わりか？
Yes
No
S106
口の部分に入力があるか？
No
Yes
S107
口を大きくする指示か？
No
S108
音量を大きくする
S109
口の画像を大きくする
Yes
S110
口を小さくする指示か？
No
S111
音量を小さくする
S112
口の画像を小さくする
Yes
S113
出力停止指示か？
No
Yes
S114
音声合成を停止する
S115
口の画像を口が閉じた画像に入れ替える

【図４】


41(S101)
42(S102)
1000
43(S103)
1000
44(S109)
1000
45(S110)
1000
46(S115)

【図５】



フロントページの続き

(56)参考文献 実開平０３－０１０６１５（ＪＰ，Ｕ）
特開平０４－２３７２８８（ＪＰ，Ａ）
特開平２－２３４２８５（ＪＰ，Ａ）
特開平６－９６１８８（ＪＰ，Ａ）
特開平７－４４７２７（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.[7]，ＤＢ名)
G10L 13/00,21/00-21/06
G06F 3/16
G09F 27/00
G09G 5/00