(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第4720974号
(P4720974)
(45)発行日　平成23年7月13日(2011.7.13)　(24)登録日　平成23年4月15日(2011.4.15)

(51)Int.Cl.　　F I
G10L 21/04 (2006.01)　　G10L 21/04 12OD
G10L 13/02 (2006.01)　　G10L 13/02 122Z

請求項の数 4　(全 23 頁)

(21)出願番号　特願2004-369108 (P2004-369108)
(22)出願日　平成16年12月21日 (2004.12.21)
(65)公開番号　特開2006-178052 (P2006-178052A)
(43)公開日　平成18年7月6日 (2006.7.6)
審査請求日　平成19年11月16日 (2007.11.16)

（出願人による申告）平成１６年度独立行政法人情報通信研究機構、研究テーマ「超高速知能ネットワーク社会に向けた新しいインタラクション・メディアの研究開発」に関する委託研究、産業活力再生特別措置法第３０条の適用を受ける特許出願

(73)特許権者　393031586
株式会社国際電気通信基礎技術研究所
京都府相楽郡精華町光台二丁目２番地２
(74)代理人　100099933
弁理士　清水　敏
(72)発明者　米澤　朋子
京都府相楽郡精華町光台二丁目２番地２
株式会社国際電気通信基礎技術研究所内
(72)発明者　鈴木　紀子
京都府相楽郡精華町光台二丁目２番地２
株式会社国際電気通信基礎技術研究所内
(72)発明者　小暮　潔
京都府相楽郡精華町光台二丁目２番地２
株式会社国際電気通信基礎技術研究所内

最終頁に続く

(54)【発明の名称】音声発生装置およびそのためのコンピュータプログラム

(57)【特許請求の範囲】
【請求項１】
同じ内容の複数種類の音声に基づいて、声質を変化させながら音声を発生させるための音声発生装置であって、
複数の中間段階の数、及び、シグモイド関数決定のためのパラメータを受付ける受付手段と、
前記受付手段が受付けた前記パラメータにしたがってシグモイド関数を決定するシグモイド関数決定手段と、
前記シグモイド関数決定手段が決定した前記シグモイド関数を用いて、前記受付手段が受付けた数の複数の中間段階におけるモーフィング率をそれぞれ決定する第１のモーフィング率決定手段と、
前記複数種類の音声に対して、前記第１のモーフィング率決定手段が決定した前記複数種類のモーフィング率で音声モーフィングを行なうことによって複数の中間音声を作成するモーフィング実行手段と、
前記複数種類の音声と、前記モーフィング実行手段が作成した前記複数の中間音声とを記憶する音声記憶装置と、
前記複数種類の音声のモーフィング率を指定するためのモーフィング率指定手段と、
前記モーフィング率指定手段によりモーフィング率が指定されたことに応答して、前記音声記憶装置から与えられる、前記複数種類の音声と前記複数の中間音声との中で、指定されたモーフィング率に最も近いモーフィング率の音声を選択する音声選択手段と、

前記音声選択手段から与えられる前記選択された音声を音声信号に変換する音声信号変換手段とを含み、
前記複数種類の音声には、予めそれぞれの声質を特定するラベルが付されており、
前記モーフィング率指定手段は、
前記複数種類の音声の各々について、対応するラベルを表示するラベル表示手段と、
前記複数種類の音声をそれぞれ表す複数の基準点をそれぞれ対応するラベルと関連付けて所定の位置に表示する基準点表示手段と、
前記基準点と所定の関係にある、予め定める領域内にユーザにより配置されたポインタの位置と、前記複数の基準点との間の距離にしたがって、前記複数種類の音声のモーフィング率を決定する第２のモーフィング率決定手段と、を含む、音声発生装置。
【請求項２】
前記モーフィング率指定手段はさらに、
所定空間内における、前記複数種類の音声の各々にそれぞれ対応する前記複数の基準点の位置を決定する基準点決定手段を含む、請求項１に記載の音声発生装置。
【請求項３】
前記音声選択手段による音声発生の基準時刻を定めるタイマと、
前記モーフィング率指定手段によりモーフィング率の指定がされたことに応答して、前記タイマを参照して音声再生の時刻を得て、前記音声選択手段により選択された音声を示す情報とともに音声再生シーケンスとして記録する選択音声記録手段と、
音声再生シーケンスの再生を指示する信号に応答して、当該信号により示される音声再生シーケンスを読出して、当該音声再生シーケンスにより指定される時刻に、当該時刻に指定された音声を選択して前記音声信号変換手段に与える音声再生制御処理手段とをさらに含む、請求項１又は請求項２に記載の音声発生装置。
【請求項４】
コンピュータにより実行されると、請求項１～請求項３のいずれかに記載の音声発生装置として当該コンピュータを動作させる、コンピュータプログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
この発明は音声処理技術に関し、特に、音声に任意の表情付けを容易に行なうことが可能な音声処理技術に関する。
【背景技術】
【０００２】
様々な音声表現においてユーザの所望する表情付けを可能にし、豊かな表情のついた音声を実現するための技術は、様々な用途に利用可能であると思われる。例えば、豊かな表情の歌声の合成などである。そのためには様々な表情付きの音声だけでなく、表情付けの程度が中間的な音声も必要だと考えられる。さらに，それらは自然な音声に近いことが望ましい。
【０００３】
よって、多種多様な音声データを集めることが望まれる。しかし、そのためには所望の表情付けがされた音声を収録する作業が必要であるが、そのような作業は極めて困難である。その結果、音声に対しユーザの所望するような種々の表情付けを容易に行なうことができる従来技術は存在していない。
【非特許文献１】カワハラ、Ｈ．およびマツイ、Ｈ．、「無干渉時間－周波数表現における弾性的知覚的距離尺度に基づく聴覚的モーフィング」、ＩＣＡＳＳＰ’２００３予稿集、第１巻、ｐｐ．２５６－２５９、２００３年（Kawahara, H., and Matsui, H., “AUDITORY MORPHING BASED ON AN ELASTIC PERCEPTUAL DISTANCE METRIC IN AN INTERFERENCE-FREE TIME-FREQUENCY REPRESENTATION,” Proc. ICASSP'2003, vol.I, pp.256-259, 2003.）
【非特許文献２】http://www.wakayama-u.ac.jp/~kawahara/STRAIGHTadv/（高品質音声分

析変換合成システム　ＳＴＲＡＩＧＨＴ）
【発明の開示】
【発明が解決しようとする課題】
【０００４】
以上のように、音声に対しユーザの所望するような種々の表情付けを容易に行なうことができる技術は存在しておらず、そうした技術に対する需要が高まっている。
【０００５】
したがって本発明の主たる目的は、音声の表情付けを自由に変化させながら音声の発生を行なうことができる音声発生装置を提供することである。
【０００６】
本発明の他の目的は、音声の表情付けを自由に、かつ簡単な操作で変化させながら音声の発生を行なうことができる音声発生装置を提供することである。
【０００７】
本発明のさらに他の目的は、多数の音声を収録する必要なく、音声の表情付けを自由に変化させながら音声の発生を行なうことができる音声発生装置を提供することである。
【課題を解決するための手段】
【０００８】
本発明の第１の局面に係る音声発生装置は、同じ内容の複数種類の音声に基づいて、声質を変化させながら音声を発生させるための音声発生装置である。この音声発生装置は、複数種類の音声のモーフィング率を指定するためのモーフィング率指定手段と、モーフィング率指定手段によりモーフィング率が指定されたことに応答して、複数種類の音声と、複数種類の音声に対して複数種類のモーフィング率で音声モーフィングを行なって得た中間音声との中で、指定されたモーフィング率に最も近いモーフィング率の音声を選択するための音声選択手段と、音声選択手段により選択された音声を音声信号に変換するための音声信号変換手段とを含む。
【０００９】
複数種類の音声と、それらの中間音声とを予め準備しておき、モーフィング率指定手段によりモーフィング率が指定されると、それに応答してこれら音声の中で指定されたモーフィング率に最も近いモーフィング率の音声を選択して音声信号に変換する。多くの種類の音声を準備しなくても、中間音声を用いることにより種々の表情付けがされた音声の発生をリアルタイムで行なうことができる。
【００１０】
好ましくは、複数種類の音声には、予めそれぞれの声質を特定するラベルが付されている。そしてモーフィング率指定手段は、複数種類の音声の各々について、対応するラベルを表示するためのラベル表示手段と、複数種類の音声をそれぞれ表す複数の基準点をそれぞれ所定の位置に対応するラベルと関連付けて表示するための基準点表示手段と、基準点と所定の関係にある予め定める領域内にユーザにより配置されたポインタの位置と、複数の基準点との間の距離にしたがって、複数種類の音声のモーフィング率を決定するためのモーフィング率決定手段とを含む。
【００１１】
複数種類の音声に付されたラベルが表示され、さらにそれら複数種類の音声に対応する基準点がラベルと関連付けて表示される。ユーザがそれら基準点と所定の関係にある領域、例えば基準点が二つの場合には基準点を結ぶ直線上、基準点が３つの場合にはそれら３点で囲まれる三角形内にポインタを配置すると、その位置と基準点との間の距離にしたがって、モーフィング率決定手段が音声のモーフィング率を決定する。視覚的で直感的に分りやすい、簡単な操作で音声のモーフィング率を指定することができる。
【００１２】
モーフィング率指定手段は、所定空間内において、複数種類の音声の各々にそれぞれ対応する複数の基準点を決定するための基準点決定手段と、複数の基準点と所定の関係にある、予め定める領域内にユーザにより配置されたポインタの位置と、複数の基準点との間

の距離にしたがって、前記複数種類の音声のモーフィング率を決定するためのモーフィング率決定手段とを含んでもよい。

【００１３】

複数種類の音声に対応する基準点が空間内において決定される。ユーザがそれら基準点と所定の関係にある領域、例えば基準点が４つの場合にはそれら４点で囲まれる三角錐内にポインタを配置すると、その位置と基準点との間の距離にしたがって、モーフィング率決定手段が音声のモーフィング率を決定する。ポインタの配置によってモーフィング率を指定できるため、直感的で分りやすく、簡単な操作で音声のモーフィング率を指定することができる。

【００１４】

さらに好ましくは、音声発生装置は、音声選択手段による音声発生の基準時刻を定めるタイマと、モーフィング率指定手段によりモーフィング率の指定がされたことに応答して、タイマを参照して音声再生の時刻を得て、音声選択手段により選択された音声を示す情報とともに音声再生シーケンスとして記録するための選択音声記録手段と、音声再生シーケンスの再生を指示する信号に応答して、当該信号により示される音声再生シーケンスを読出して、当該音声再生シーケンスにより指定される時刻に、当該時刻に指定された音声を選択して音声信号変換手段に与えるための音声再生制御処理手段とをさらに含む。

【００１５】

音声を選択する操作をしながら音声を発声させると、その操作により選択された音声と、その音声が選択されたときの時刻とが音声再生シーケンスとして記録される。それを音声再生制御手段によって再生することにより、過去に行なった音声再生と同じ順序で、同じ音声を用いた音声再生を再現することができる。

【００１６】

音声発生装置は、複数種類の音声と中間音声とを記憶し、音声選択手段に与えるための音声記憶装置をさらに含んでもよい。

【００１７】

本発明の第２の局面に係るコンピュータプログラムは、コンピュータにより実行されると、上記したいずれかの音声発生装置として当該コンピュータを動作させるものである。したがって、対応する音声発生装置と同様の作用効果を奏することができる。

【発明を実施するための最良の形態】

【００１８】

異なる二つの音声を所望のモーフィング率で混合することにより中間的な声質を有する音声（以下「中間音声」と呼ぶ。）を作成する音声モーフィングと呼ばれる技術が存在する。この音声モーフィングに利用可能な音声分析変換合成ツールとしてＳＴＲＡＩＧＨＴと呼ばれるシステムが知られている（非特許文献１に記載）。

【００１９】

この音声モーフィング技術を用いれば、予め録音した２種類の音声間の中間の声質を持つ音声を作成できる。本実施の形態では、実際の音声データだけではなく、音声モーフィングにより作成した中間音声も使用し、再生中の音声の声質を所望にしたがって変化させる。中間音声を作成する上では、元になる２種類の音声、例えば平坦（ｎｏｒｍａｌ）な音声、暗い（ｄａｒｋ）音声、ねっとりした（ｗｅｔ）音声の中の２種類により、同じ内容のテキストを朗読する、または同じ歌を歌うときの声を録音することが有効である。このとき、テキストの朗読速度、または歌の速さを同じになるようにするとよい。例えば歌の場合であれば予め録音された伴奏にあわせて歌を歌うようにすればよい。また、テキストの朗読の場合には、最初に朗読した音声をヘッドホンで話者に聞かせながら２回目の朗読を行なうようにしてもよい。

【００２０】

なお、本明細書において「表情付け」とは、音声を聞く人が、その音声によりその音声に対して感じる主観的な印象のことをいう。また、本明細書では、そのような印象により表される声の性質を「声質」と呼ぶ。すなわち、本明細書においては「表情付け」と「声

質」とは同じ意味を表す。

【００２１】

なお、以下に記載する実施の形態の説明では、同じ部品には同じ参照符号を付す。それらの名称および機能も同一である。したがってそれらについての詳細な説明は繰返さない。

【００２２】

図１に、本発明の一実施の形態にかかる音声モーフィングシステム２０の概略構成を示す。図１を参照して、この音声モーフィングシステム２０は、予め準備された、基準となる第１の音声３８および第２の音声４０を記憶する基準音声記憶装置２６と、基準音声記憶装置２６に記憶された第１の音声３８および第２の音声４０を利用者により指定されたパラメータを用いて得られる複数種類のモーフィング率をそれぞれ用いてモーフィングすることにより、第１の音声３８と第２の音声４０との間での等間隔の中間音として知覚される９個のモーフィング後音声２８を作成するための音声モーフィング装置２２と、音声モーフィング装置２２により作成されたモーフィング後音声２８を格納するための記憶装置と、音声モーフィング装置２２に対してモーフィング後音声２８として作成される中間音声の数などのパラメータをユーザが入力する際に使用する入出力インタフェース２４とを含む。

【００２３】

ここで「等間隔の中間音として知覚される」とは、聴者の主観的な印象として、発生された音声の声質が、一定の割合で一方の声質から他方の声質に変化していくように感じられることを指す。

【００２４】

音声モーフィング装置２２は、実質的には前述した非特許文献１に記載のＳＴＲＡＩＧＨＴを使用する。

【００２５】

音声モーフィングシステム２０はさらに、モーフィング後音声２８と第１の音声３８および第２の音声４０とを記憶するための音声データ記憶装置３０と、所与のモーフィング率に基づいて、音声データ記憶装置３０に記憶された音声データのうち、与えられたモーフィング率に最も近い音声をリアルタイムで選択し音声信号として再生するためのモーフィング音声選択・再生装置３２と、モーフィング音声選択・再生装置３２に対してモーフィング率などの指示を与えるためにユーザが利用するユーザインタフェース３４と、モーフィング音声選択・再生装置３２により再生された音声信号を音声に変換するためのスピーカシステム３６とを含む。

【００２６】

図１に示す音声モーフィングシステム２０は、一般的にはコンピュータシステムのハードウェアと、当該ハードウェアにより実行されるプログラムとにより実現される。図２にこの音声モーフィングシステム２０を実現するコンピュータシステム５０の外観図を、図３にそのブロック図を、それぞれ示す。

【００２７】

図２を参照して、音声モーフィングシステム２０を実現するコンピュータシステム５０は、マイクロフォン６４と、一組のスピーカ３６と、ＣＤ－ＲＯＭ（Compact Disc Read Only Memory)ドライブ７０およびＦＤ（Flexible Disk)ドライブ７２とを有するコンピュータ７８と、いずれもコンピュータ７８に接続されたＬＣＤ（液晶表示装置）７４とキーボード６６とマウス６８とを含む。

【００２８】

図３はコンピュータ７８のハードウェアのブロック図である。図３を参照して、コンピュータ７８は、ＣＰＵ（Central Processing Unit：中央処理装置）８０と、ＣＰＵ８０に接続されたバス８２と、バス８２に接続された読出専用メモリ（ＲＯＭ)８４と、バス８２に接続されたランダムアクセスメモリ（ＲＡＭ）８６と、バス８２に接続されたハードディスク８８と、ＣＤ－ＲＯＭ９６からデータを読出すＣＤ－ＲＯＭドライブ７０と

、ＦＤ９８からデータを読出したりＦＤ９８にデータを書込んだりするためのＦＤドライブ７２と、バス８２に接続され、マイクロフォン６４とスピーカ３６とが接続されるサウンドボード９０と、バス８２に接続され、ローカルエリアネットワーク（ＬＡＮ）等のネットワーク上でのデータコミュニケーション能力を提供するネットワークボード９２と、ビデオキャプチャボード９４とを含む。

【００２９】

図２、図３に示すコンピュータシステム５０上で所定の制御構造を有するソフトウェアを動作させることにより、図１に示す音声モーフィングシステム２０を実現できる。

【００３０】

図４は、図１に示す音声モーフィング装置２２のブロック図である。図４を参照して、音声モーフィング装置２２は、入出力インタフェース２４を介して操作者から音声モーフィングの特徴を定めるパラメータを受取るためのパラメータ入力部１００と、パラメータ入力部１００により受取られたパラメータにしたがい、モーフィング後音声が知覚的に等間隔に並ぶように二つの音声のモーフィング率を定める際に用いるシグモイド関数を決定するためのシグモイド関数決定部１０２とを含む。

【００３１】

ところで、音声モーフィングにおいては、音声のモーフィング率をどのように決めるかが問題となる。最も簡単な方法として、二つの音声の混合割合を一定割合ずつ増減させていく方法が考えられる。例えば第１の段階として第１の音声を９０％と第２の音声を１０％、第２の段階として第１の音声を８０％と第２の音声を２０％、のように一定の差分でモーフィング率を変えていく方法である。

【００３２】

しかし、本発明の発明者らは、このように一定の差分でモーフィング率を変化させた場合、実際に聴者にとっては一定の割合で音声が変化していくようには感じられないことを実験により確認した。さらに実験により、以下に説明するようにシグモイド関数を用いてモーフィング率を決定すると、聴者にとって一定の割合で音声が変化していくように感じられることが分った。以下、本実施の形態で使用する、シグモイド関数を用いたモーフィング率の決定の仕方を説明する。

【００３３】

図５に、シグモイド関数のグラフ１３０を示す。シグモイド関数は、一般的に以下の式により定義される。

【００３４】

【数１】

$$f(x) = 1/(1+\exp(-ax+b))+c$$

例えばこの式のパラメータａ、ｂおよびｃをパラメータ入力部１００を介して操作者から受取ることにより、シグモイド関数決定部１０２は任意のシグモイド関数を定義できる。

【００３５】

こうして決定されたシグモイド関数１３０のグラフのうち、中央の変曲点を含んだ左右対称な部分をモーフィング率の決定に用いる。まず、その横軸を中間音声の段階数に合わせて等分し、それぞれに段階番号を割当てる。図５には、その例として０～１０までを示す。割当てられた０～１０までの数字のうち、左端の「０」は第１の音声が１００％、第２の音声が０％（すなわち第１の音声のみ）の音声を示す。右端の「１０」は、第１の音声が０％、第２の音声が１００％（すなわち第２の音声のみ）の音声を示す。中間の１～９までは、それぞれ中間の音声の段階を示す。段階０における第２の音声のモーフィング率を０％、段階１０における第２の音声のモーフィング率を１００％として、縦軸にモーフィング率をとることができる。

【００３６】

図５に示すシグモイド関数１３０について、上記した各中間段階（１～９）での値を求

め、その値をその段階での第２の音声のモーフィング率とする。シグモイド関数の曲線は中央に変曲点があるため、上記した図５に示す例では段階５における第２の音声のモーフィング率は５０％となる。しかしそれ以外の点では、モーフィング率は段階番号に対し非線形に変化する。

【００３７】

図４に示す音声モーフィング装置２２は、このようにして求めた各段階でのモーフィング率の値１３２を用いて、各段階での音声モーフィングを行なう。このため、音声モーフィング装置２２はさらにモーフィング率決定部１０４を含む。

【００３８】

音声モーフィング装置２２はさらに、モーフィング率決定部１０４により決定された複数のモーフィング率の全てについて、予め準備された第１の音声３８と第２の音声４０との間の音声モーフィングを行なう繰返し制御を実行するための繰返し制御部１０６と、繰返し制御部１０６による制御にしたがい、第１の音声３８と第２の音声４０とを繰返し制御部１０６から指定されるモーフィング率で混合してモーフィング後の音声を生成するためのモーフィング実行部１０８と、モーフィング実行部１０８により作成されたモーフィング後音声２８を、繰返し制御部１０６による繰返し制御にしたがって異なる名称を付したファイルとして、予め準備された音声記憶装置１２０内に保存するための保存処理部１１０とを含む。音声記憶装置１２０としては、図３に示すハードディスクを用いることができる。

【００３９】

音声モーフィング装置２２は、後述するようにコンピュータハードウェアと、コンピュータにより実行されるプログラムとにより実現される。図６に、繰返し制御部１０６、モーフィング実行部１０８、および保存処理部１１０を実現するためのプログラムの制御構造をフローチャート形式で示す。なお、図４に示すモーフィング率決定部１０４により決定された９つの中間段階のモーフィング率をｒ（ｋ）（ｋ＝１～９）とする。

【００４０】

図６を参照して、このプログラムは、起動されるとまず初期設定を行なう（ステップ１４０）。続いてステップ１４２において、繰返し制御変数ｋに０を代入する。ステップ１４４で繰返し制御変数ｋに１を加算する。ステップ１４６で繰返し制御変数ｋの値が予め設定された中間段階数（本実施の形態では９）を越えたか否かを判定する。越えていれば処理を終了する。越えていなければステップ１４８に進む。

【００４１】

ステップ１４８では、モーフィング率ｒ（ｋ）で第１の音声と第２の音声とをＳＴＲＡＩＧＨＴを用いて音声モーフィングする。ステップ１５０では、得られたモーフィング後の音声を「morph_k.wav」（ｋは１～９までの数字）というファイル名で音声記憶装置１２０に保存する。制御はステップ１４４に戻る。

【００４２】

こうして、ｋ＝１～９まで音声モーフィングと保存とを繰返すことにより、図４の音声記憶装置１２０には、９段階のモーフィング後音声２８が記憶される。なお、第１の音声３８および第２の音声４０も予め音声記憶装置１２０に記憶しておくことにより、音声記憶装置１２０には後述するモーフィング音声選択・再生装置３２において使用する音声資源が全て記憶されることになる。

【００４３】

図７に、そのモーフィング音声選択・再生装置３２の機能的ブロック図を示す。図７を参照して、モーフィング音声選択・再生装置３２は、モーフィング音声選択・再生装置３２をコンピュータにより実現するように予め準備されたプログラム１６０と、ユーザインタフェース３４を用いたモーフィング音声の発生に関する入出力を行なうために、プログラム１６０に基づいて入出力画面１６４を作成し、モニタ７４に表示させるための表示作成部１６２と、入出力画面１６４に対しユーザがユーザインタフェース３４を用いて何らかの操作を行なった際、その操作によりシステム内に発生するイベントを検知し、プログ

ラム１６０内のオブジェクトのうち、適切なものに当該イベントを振分けるためのイベント検知部１６６とを含む。表示作成部１６２およびイベント検知部１６６としての基本的な機能は、コンピュータのオペレーティングシステム（ＯＳ）により提供される。

【００４４】

モーフィング音声選択・再生装置３２には、第１の音声３８、第２の音声４０、およびモーフィング後音声２８を記憶し音声発生のためにモーフィング音声選択・再生装置３２に与えるための、図４に示すものと同じ音声記憶装置１２０と、表示作成部１６２により発生される音声信号を音声に変換するためのスピーカ３６と、表示作成部１６２により作成される音声再生シーケンスファイルを記憶するための記憶装置１７８とが接続される。記憶装置１７８は、音声記憶装置１２０と同様、図３に示すハードディスク８８により実現できる。

【００４５】

表示作成部１６２は、音声記憶装置１２０に記憶された第１の音声３８および第２の音声４０に付されたラベルを読出し、入出力画面１６４に表示することでそれぞれの音声の声質をユーザに対し提示することができる。

【００４６】

図８に入出力画面１６４の例を示す。図８上段を参照して、入出力画面１６４は、それぞれモーフィング対象となる第１の音声３８および第２の音声４０に付されたラベルを表示するためのラベルテキスト領域２１０および２１２と、音声を中間段階のいずれに設定するかをユーザが指定するように準備されたスライダ２１４と、音声発生の開始および停止をそれぞれ指示する際にユーザが使用するための開始ボタン２１６および停止ボタン２１８とを含む。

【００４７】

図８に示す例において、ラベルテキスト領域２１０には「ｎｏｒｍａｌ」（特に特徴をもたない、中立的な平坦な音）、ラベルテキスト領域２１２には「ｗｅｔ」（鼻にかかったような「ねっとり」した声色）というラベルがそれぞれ表示されている。これら以外にも例えば「ｄａｒｋ」（母音が全体的に後舌母音に近くなるような「暗い」音）、子守唄のような「ｗｈｉｓｐｅｒ」（ささやき声）などのラベルが考えられるが、一般的にはラベルは使用者（または録音者）の主観に基づいて自由に音声ファイルに付しておけばよい。

【００４８】

スライダ２１４は、スライダ目盛２４２と、スライダつまみ２４０とを含む。スライダつまみ２４０を例えば図８の下段に示すようにマウスポインタ２４４でドラッグすることにより、スライダつまみ２４０がスライダ目盛２４２上を移動する。スライダ２１４は、スライダ目盛２４２上のスライダつまみ２４０の位置に対応する値をリアルタイムで検知し、属性値として保持する。同時にスライダ２１４が操作されたというイベントをイベント検知部１６６に与える機能を持つ。

【００４９】

開始ボタン２１６が押され、音声発生が開始された後にユーザがスライダつまみ２４０をスライドさせることにより、音声の発生中に、その音声を第１の音声から第２の音声まで、中間段階を含めて１１種類の音にリアルタイムで変更させることができる。モーフィング音声選択・再生装置３２は、音声発生時のユーザ操作による声質変更のシーケンスを記憶し、記憶装置１７８（図７参照）にファイルとして出力する機能を持つ。

【００５０】

再び図８を参照して、入出力画面１６４はさらに、記憶装置１７８に記憶された声質変更のシーケンスファイルからシーケンスを読出して当該シーケンスにしたがって声質を変更して所定の音声を再生する際にユーザが使用する再生ボタン２２０を含む。

【００５１】

プログラム１６０は、図８に示すスライダ２１４の実体を構成するスライダオブジェクト１９０と、それぞれ開始ボタン２１６、停止ボタン２１８、および再生ボタン２２０の

実体を構成する開始ボタンオブジェクト１９２、停止ボタンオブジェクト１９４、および再生ボタンオブジェクト１９６とを含む。

【００５２】

これら各オブジェクトについて、プログラム１６０の実行開始とともにそれぞれインスタンスが生成されて入出力画面１６４の作成、イベント検知部１６６による適切なメソッドの実行、および各インスタンスに付随する属性値の取得と記憶などが実行される。

【００５３】

以下、プログラム１６０に含まれるプログラムコードをコンピュータで実行することにより実現される機能を、それぞれ機能ブロックとして説明する。

【００５４】

すなわち、モーフィング音声選択・再生装置３２はさらに、プログラム１６０により実現される機能ブロックとして、イベント検知部１６６により開始ボタン２１６の押下というイベントが検知されたことに応答して計時を開始し、停止ボタン２１８の押下というイベントが検知されたことに応答して計時を終了するタイマ１８４と、イベント検知部１６６によりスライダオブジェクト１９０の操作に関するイベントが検知されたことに応答して、スライダオブジェクト１９０からその属性値としてスライダ値（区間［０，１０］の間の整数）を読取り、音声記憶装置１２０に記憶された音声（第１の音声３８、第２の音声４０、および９種類の中間音声）のうちいずれを使用するかを決定し、音声ファイルの名称を出力するための音声選択処理部１６８とを含む。

【００５５】

プログラム１６０により実現される機能ブロックはさらに、第１および第２の二つの入力を持ち、第１の入力が音声選択処理部１６８の出力を受けるように接続され、図８に示す開始ボタン２１６が押されたときには第１の入力を、再生ボタン２２０が押されたときには第２の入力を、それぞれ選択し、選択された入力に与えられた信号を出力するためのセレクタ１７０と、セレクタ１７０の出力を受け、音声記憶装置１２０に記憶された音声ファイル（第１の音声３８、第２の音声４０、および９種類の中間音声）のうち、セレクタ１７０の出力により指定される音声ファイルを読出してタイマ１８４により指定される再生位置から音声信号への変換を開始しスピーカ３６に与えるための音声発生処理部１７２とを含む。

【００５６】

プログラム１６０により実現される機能ブロックはさらに、開始ボタン２１６の押下というイベントが検知されたことに応答して、音声発生処理部１７２による音声の発生を開始させるための開始指示部１７３と、開始指示部１７３からの音声発生の開始指示と、音声選択処理部１６８による音声選択処理とに応答して、そのときの音声選択処理部１６８の出力をタイマ１８４の計時値とともにシーケンスとして記録するための選択音声記録部１７４と、停止ボタン２１８の押下というイベントに応答して、選択音声記録部１７４により記録されている再生シーケンスをファイルとして記憶装置１７８に保存させるためのシーケンス保存部１７６とを含む。

【００５７】

プログラム１６０により実現される機能ブロックはさらに、再生ボタン２２０の押下というイベントに応答して、記憶装置１７８に保存されている１または複数の再生シーケンスファイルのいずれかをユーザに選択させるためのファイル選択処理部１８０と、ファイル選択処理部１８０により再生シーケンスファイルが選択されると、タイマ１８４を起動し、ファイル選択処理部１８０により選択された再生シーケンスファイルを記憶装置１７８から読出して、タイマ１８４の計時に基づいて、再生開始後、選択された再生シーケンスにより指定された時刻になると再生シーケンスにより指定された音声ファイル名をセレクタ１７０の第２の入力に与えることにより、再生シーケンスに基づく音声再生を制御するための音声再生制御処理部１８２とを含む。

【００５８】

［動作］

図１～図８を参照して、上記した音声モーフィングシステム２０は以下のように動作する。動作は大きく３つのフェーズに分けられる。第１のフェーズでは、予め準備された、互いの声質の異なる第１の音声３８と第２の音声４０とから９個の中間段階のモーフィング後音声２８を作成する。なお、これに先立ち、同じ話者により、声質（音声の表情）を変えて同一の文章を読んだり同一の歌を歌ったりすることによって二つの音声を収録しておき、これらをそれぞれ第１の音声３８および第２の音声４０として保存しておく。また、第１の音声３８および第２の音声４０の音声ファイルには、付属情報として声質を示すラベルを付しておく。

【００５９】

第２のフェーズでは、このようにして作成されたモーフィング後音声２８と、最初に準備された第１の音声３８および第２の音声４０とを用い、声質を自由に変更しながらこれら音声の発生を行なう。このとき、再生シーケンスが記憶装置１７８に保存される。第３のフェーズでは、記憶装置１７８にファイルとして保存された再生シーケンスを読出し、その再生シーケンスにしたがって音声を選択し発生させることにより、再生シーケンスを再現する。以下、各フェーズでの音声モーフィングシステム２０の動作を説明する。

【００６０】

－音声モーフィング－

図４を参照して、パラメータ入力部１００は、入出力インタフェース２４を用いてユーザから、中間段階の数と、シグモイド関数決定のためのパラメータとを受取る。パラメータ入力部１００は、このパラメータをシグモイド関数決定部１０２に与える。

【００６１】

シグモイド関数決定部１０２は、与えられたパラメータにしたがってシグモイド関数を決定する。決定されたシグモイド関数に関する情報はモーフィング率決定部１０４に与えられる。

【００６２】

モーフィング率決定部１０４は、このシグモイド関数を用い、図５を参照して説明した方法にしたがって、各中間段階におけるモーフィング率ｒ（ｋ）（ｋ＝１～９）を決定する。モーフィング率決定部１０４は、決定したモーフィング率を、中間段階数とともに繰返し制御部１０６に与える。

【００６３】

繰返し制御部１０６は、指定されたモーフィング率で中間段階の音声を音声モーフィングにより作成するようモーフィング実行部１０８を制御し、作成させる。繰返し制御部１０６はまた、このようにして作成された９つの中間段階のモーフィング後音声２８を音声記憶装置１２０に格納するよう保存処理部１１０を動作させる。このとき、保存処理部１１０は、各中間音声のファイルに前述したとおり「morph_k.wav」というファイル名を付す。音声記憶装置１２０には、予めモーフィングに用いられる第１の音声３８および第２の音声４０も準備されているものとする。

【００６４】

以上で第１のフェーズの動作は完了である。音声記憶装置１２０を音声データ記憶装置３０としてモーフィング音声選択・再生装置３２に接続することにより、モーフィング音声選択・再生装置３２による、中間音声を用いて表情を変化させた音声の発生が可能になる。

【００６５】

－中間音声を用いた任意の声質の音声の発生－

図７を参照して、第２のフェーズにおいてはモーフィング音声選択・再生装置３２は以下のように動作する。モーフィング音声選択・再生装置３２の各機能ブロックは、予めプログラム１６０の形で準備されている。

【００６６】

ユーザがモーフィング音声を用いた音声の発生をするようにモーフィング音声選択・再生装置３２に指示すると、図７および図８に示す入出力画面１６４が表示作成部１６２に

より作成され、モニタ７４上に表示される。このとき、表示作成部１６２は、音声記憶装置１２０に記憶された第１の音声３８および第２の音声４０に付されたラベルを読込み、それぞれラベルテキスト領域２１０および２１２に表示する。この表示により、ユーザは二つの音声の声質がどのようなものであるかを知ることができる。

【００６７】

ユーザは、まずスライダ２１４のスライダつまみ２４０をマウスポインタ２４４によりスライドさせることで音声発生開始時の中間音声の段階を指定する。この操作によりイベントが発生し、イベント検知部１６６はこのイベントをスライダオブジェクト１９０に与える。スライダオブジェクト１９０は、属性値として保持しているスライダつまみ２４０の位置を示す値を音声選択処理部１６８に与える。音声選択処理部１６８は、この値に基づいて、音声記憶装置１２０に記憶された音声ファイルのうちどの音声ファイルを選択するかを決定し、セレクタ１７０に与える。

【００６８】

ユーザが開始ボタン２１６を押下すると開始ボタン２１６の押下イベントが発生する。イベント検知部１６６はこのイベントを開始ボタンオブジェクト１９２に与える。開始ボタンオブジェクト１９２は、このイベントに応答し開始指示部１７３を制御してタイマ１８４を起動する。セレクタ１７０は、第１の入力、すなわち音声選択処理部１６８から与えられたファイル名を選択して音声発生処理部１７２に与える。開始指示部１７３は、音声発生処理部１７２に対して音声の発生の開始を指示する。

【００６９】

音声発生処理部１７２は、音声記憶装置１２０から、セレクタ１７０を介して音声選択処理部１６８から与えられたファイル名に対応するファイルを読出し、タイマ１８４の計時にしたがった位置からタイマ１８４の計時に同期して再生を開始する。発生された音声信号はスピーカ３６に与えられ、音声に変換される。

【００７０】

一方、選択音声記録部１７４は、開始ボタン２１６が押下されたことに応答して、タイマ１８４の計時が０のときの音声選択処理部１６８の出力をタイマ１８４の値とともに記録する。

【００７１】

このようにして音声ファイルが再生されている途中でユーザが図８に示すスライダつまみ２４０を操作して、スライダつまみ２４０を別の位置に移動させたものとする。このイベントはイベント検知部１６６により検知され、スライダオブジェクト１９０に与えられる。スライダオブジェクト１９０はこのイベントに応答して、属性値として保持しているスライダつまみ２４０の位置を示す値を音声選択処理部１６８に与える。音声選択処理部１６８は、この値に対応する音声ファイル名を選択し、セレクタ１７０に与える。セレクタ１７０はこの値を音声発生処理部１７２に与えるので、音声発生処理部１７２は指定された音声ファイルを新たに読出し、タイマ１８４により示される位置からタイマ１８４に同期して音声の再生を開始する。

【００７２】

また、音声選択処理部１６８が新たなファイル名を選択したことに応答し、選択音声記録部１７４はその新たなファイル名と、そのときのタイマ１８４の値とを組にして追加して記録する。

【００７３】

図９の上段に、ユーザによるスライダつまみ２４０の操作の例を時系列で示す。図９上段では、縦軸にスライダつまみ２４０の位置をスライダ目盛の値で、横軸に時間を、それぞれ示す。図９上段に示すように、スライダの位置は時間的にある軌跡２５０を描く。これがモーフィング後音声を用いた再生シーケンスを示す。この再生シーケンスを選択音声記録部１７４によって記録しておけば、同じ再生シーケンスを再現することができる。

【００７４】

ただし、スライダつまみ２４０の位置はスライド目盛の中間となることもあり得る。そ

うした場合には、スライダつまみ２４０の位置に最も近い目盛を選択し、その目盛に対応する音声ファイルを選択する。したがって、なだらかな線を描くスライダつまみ２４０の軌跡２５０は、図９下段に示すようにある中間音声から次の中間音声に、音声の種類としては不連続な形で、ただし時間的には連続して、再現されることになる。ただし、このままでは音声のつなぎ目で「プツン」という雑音が入る。そこで、こうしたつなぎ目では、先の音声を徐々にフェードアウトし、後の音声を徐々にフェードインする形で音声を混合することで雑音の発生を防止する。

【００７５】

中間段階の数としてある程度大きい値を選択しておけば、図９下段に示すような形で音声を再生しても、聴者には違和感を与えない。また、各中間音声のモーフィング率は、シグモイド関数を用いてできるだけ等しい間隔で相違した音声となるように設定されている。したがって、聴者には、こうして生じる声質の変化は、滑らかでかつ自然に感じられることになる。

【００７６】

停止ボタン２１８が押されると、そのイベントはイベント検知部１６６により検知され、停止ボタンオブジェクト１９４に与えられる。停止ボタンオブジェクト１９４は、このイベントに応答して選択音声記録部１７４による音声シーケンスの記録を中止させる。またタイマ１８４も停止させる。さらに停止ボタンオブジェクト１９４は、シーケンス保存部１７６に指示を与え、選択音声記録部１７４により記録された再生シーケンスを記憶装置１７８に保存させる。このときの保存名は、ダイアログボックスを開いてユーザに指定させる。

【００７７】

以上が声質を変化させて音声を発生させる際のモーフィング音声選択・再生装置３２の動作である。

【００７８】

記憶装置１７８に保存された再生シーケンスに基づいて音声発生を行なう際には、モーフィング音声選択・再生装置３２は以下のように動作する。

【００７９】

再生シーケンスに基づく音声発生を行なう際には、ユーザは再生ボタン２２０（図８参照）を押下する。このイベントはイベント検知部１６６により検知される。イベント検知部１６６はこのイベントを再生ボタンオブジェクト１９６に与える。

【００８０】

再生ボタンオブジェクト１９６は、イベントが与えられたことに応答してファイル選択処理部１８０を起動する。ファイル選択処理部１８０は、記憶装置１７８に保存された各ファイルのファイル名を読出し、ファイル選択処理部１８０にダイアログボックスとしてファイル選択ダイアログを表示する。ユーザが所望のファイルを選択すると、ファイル選択処理部１８０は選択されたファイル名を音声再生制御処理部１８２に与える。

【００８１】

音声再生制御処理部１８２はタイマ１８４を起動させる。音声再生制御処理部１８２はさらに、選択された再生ファイルを読込み、最初に選択されていた音声ファイルを指定する信号をセレクタ１７０に与え、同時に音声発生処理部１７２を起動する。セレクタ１７０は、第２の入力を選択して音声発生処理部１７２に与える。

【００８２】

音声発生処理部１７２は、セレクタ１７０を介して音声再生制御処理部１８２から与えられた音声ファイルを音声記憶装置１２０から読出し、タイマ１８４の計時にしたがって再生を開始する。

【００８３】

音声再生制御処理部１８２は、再生ファイル中の音声ファイル名とタイマ計時との組のうち、タイマ計時の値をタイマ１８４による計時と常に照合し、タイマ１８４の計時と一致するタイマ計時を持つものがあればそのタイマ計時と組になっている音声ファイル名を

セレクタ１７０に与える。したがって音声発生処理部１７２は、この新たな音声ファイル名により指定される音声ファイルを音声記憶装置１２０から読出し、タイマ１８４の計時にしたがって再生を行なう。

【００８４】

こうして、ファイル選択処理部１８０によって指定された再生ファイルによる再生シーケンスにしたがって、音声発生処理部１７２が音声記憶装置１２０中の音声を随時切替えながら音声発生を行なう。

【００８５】

以上のように本実施の形態に係る音声モーフィングシステム２０によれば、所望の音声を全て収録しなくても、第１の音声３８および第２の音声４０の中間音声を音声モーフィングで準備し、さらにユーザの選択にしたがってリアルタイムで音声をそれらの中で切替えながら音声発生を行なうことができる。中間音声の間の相違が一定に知覚されるように予め音声のモーフィング率を決めて中間音声を作成しているので、音声発生の途中で音声の切替えを行なっても不自然には感じられない。また、リアルタイムで作成した音声シーケンスを記録しておくことで、いつでも同じ再生シーケンスを再現できる。

【００８６】

また、このように音声モーフィングを使用して中間段階の音声を作成し、それらを切替えて音声発生を行なうと、中間段階の音声については第１の音声３８および第２の音声４０の声質に応じ、それらの中間の声質を表すものとして知覚される。

【００８７】

したがって、本実施の形態に係る音声モーフィングシステム２０によれば、多数の音声を収録しなくても、任意の時刻にユーザが選択した声質を用いて音声を発生させることで、豊かな表情を持つ音声の発生が可能になる。

【００８８】

なお、上記実施の形態ではパーソナルコンピュータのユーザインタフェースを使用してスライダを実現したが、本発明はそのような実施の形態には限定されない。例えばシンセサイザ等に組み込む形で、シンセサイザのスライダを用いたコントロールを行なっても良い。

【００８９】

また、予め音声ファイルに付されていたラベルを変更したいとユーザが考えることもあるので、ラベルテキスト領域２１０をテキスト入力可能な領域とし、ラベルをユーザが変更可能にしてもよい。例えば、図８において「ｎｏｒｍａｌ」（平坦）とラベルが表示されている音声が、ユーザには「ｄａｒｋ」（暗い）と感じられることもある。そうした場合には、図１０に示すようにラベルテキスト領域２１０に「ｄａｒｋ」と入力して音声ファイルとともに保存しておくことにより、次にこの音声ファイルを使用する場合には「ｄａｒｋ」というラベルがラベルテキスト領域２１０に表示される。

【００９０】

なお、図８に示すように「ｎｏｒｍａｌ」というラベルを持つ音声と、「ｗｅｔ」というラベルを持つ音声との間で表情付けを変化させるということは、特に特徴をもたない平坦な（中立的な）音声に対し、「ｗｅｔ」な（ねっとりした）表情の強度を変化させながら付加させることであると考えることができる。これに対し図１０に示すように「ｄａｒｋ」というラベルを持つ音声と、「ｗｅｔ」というラベルを持つ音声との間で表情付けを変化させるということは、「ｄａｒｋ」という表情付けと、「ｗｅｔ」という表情付けとの間での、表情付けの種類を変化させる、ということであると考えることができる。

【００９１】

［変形例］

上記した実施の形態では、２種類の音声の間で音声モーフィングを行なって得た中間音声を用いた。しかし本発明はそのように２種類の音声の間でのモーフィングには限定されない。例えば３種類以上の音声の間でのモーフィングを行なうことも可能である。モーフィング自体はＳＴＲＡＩＧＨＴを使用して行なうことができる。問題は、３種類以上の音

声の間でのモーフィング率を定める方法である。

【００９２】

図１１を参照して、３種類の音声の間での音声モーフィングを行なう際のモーフィング率の決定の方法について説明する。今、３種類の音声Ａ、ＢおよびＣの間でのモーフィングを行なうものとする。図１１に示すように、これら３つの音声に対応する頂点２６０、２６２および２６４を有する三角形を考える。

【００９３】

この三角形の各辺を所定数に分割し、各辺と並行な線で分割点同士を結ぶことにより、図１１においてメッシュ２７０を作成できる。このメッシュ２７０を構成する各点に対応したモーフィング音声は以下のようにして作成できる。

【００９４】

例えば音声Ａおよび音声Ｂの間での各分割点に対応する中間音声は、上記した実施の形態での方法と同様の音声モーフィングで行なうことができる。音声Ａおよび音声Ｃの間、音声Ｂおよび音声Ｃの間での音声モーフィングもそれぞれ行なうことができる。さらに、メッシュ２７０の各交点（例えば交点２７２）での中間音声は、その交点を通る任意の線の両端（例えば点２７４、２７６）の中間音声を、その両端からその交点までの距離の比に応じたモーフィング率でモーフィングすることにより作成できる。したがって、メッシュ２７０の各点に対応する中間段階の音声を全て作成できる。

【００９５】

音声発生時には、上記したメッシュ２７０を有する三角形をコンピュータモニタ上に表示し、メッシュ２７０中の交点をポインタにより指定する。具体的には、ポインタの座標を調べ、メッシュ２７０の交点のうちポインタにより表される点に最も近い座標を持つ交点に対応する中間音声を選択すればよい。例えば図１２を参照して、ポインタ２８０がメッシュの３つの交点２９０、２９２および２９４で形成される三角形の内部にあるものとする。このときには、ポインタ２８０の位置と各交点２９０、２９２および２９４との間の距離ｄ１、ｄ２およびｄ３を調べ、距離がもっとも小さくなる点を選択する。

【００９６】

このような中間音声の発生方法は、元となる音声が３種類の場合だけでなく、４種類以上の場合にも同様に適用できる。

【００９７】

さらに、音声発生時には、上記のように作成したメッシュの二つの交点に対応する中間音声の間でさらに音声モーフィングを行なうようにしてもよい。この場合の例を図１３に示す。図１３を参照して、３種類の音声Ａ、ＢおよびＣに対応する３つの頂点２６０、２６２および２６４を有する三角形を考える。その中に、上記した方法と同様にしてメッシュ２７０を作成する。このメッシュ中の任意の交点、例えば二つの交点３１０と３１２とを選択し、この二つの交点を結ぶ線分３１４を任意の数に分割することにより、交点３１０と３１２とに対応する中間段階の音声の間の中間音声をモーフィングにより作成できる。こうして作成した音声を発生させるときには、実施の形態で説明したのと同様の方法を利用できる。

【００９８】

また、例えば元となる音声が３種類の場合には、磁気センサ、光学センサ、画像処理技術を用いた物体検出など、対象物の位置を３次元的な座標系中で検出できるシステムを用い、音声のモーフィング率を指定することができる。例えば予め３種類の音声に対し種々のモーフィング率で中間音声を作成しておく。音声発生時には、図１４に示すように空間にｘｙｚ座標系を設定する。利用者は、所定の３次元ポインタで空間内の１点３２２を指定する。この１点３２２に対し、座標値（Ｘ，Ｙ，Ｚ）が定まる。その座標値（Ｘ，Ｙ，Ｚ）に応じたモーフィング率（Ｘ／（Ｘ＋Ｙ＋Ｚ），Ｙ／（Ｘ＋Ｙ＋Ｚ），Ｚ／（Ｘ＋Ｙ＋Ｚ））に最も近いモーフィング率の中間音声を選択して音声を発生させる。

【００９９】

このように３次元的な座標指定によって発生させる音声を切替えることにより、ダンス

などのパフォーマンスとそれに伴う歌などとを連動させることができる。もちろん、次元数は３次元に限定されず、４次元以上の任意の次元数を用いることも可能である。

【０１００】

なお、本実施の形態による音声の表情付けの変化は、リアルタイムで実行できる。また、音声としては、予め一連の発話を別々の種類の表情付けがされた音声で実際に朗読したり歌ったりした場合だけではなく、予め別々の種類の表情付けがされた合成音声を準備しておいてもよい。したがって、本実施の形態による音声の表情付けを次のような場合にも利用できる。

【０１０１】

例えば、声を出すことができない人が自動プレゼンテーションにより合成音声を発する場合を考える。この場合には、予め複数種類の声質により表情付けがされた合成音声を作成しておき、合成音声によるプレゼンテーションでは、リアルタイムで音声の表情付けを変化させることができる。例えば強調したいところでは張りのある声質の音声でプレゼンテーションし、重要でないところはぼそぼそとした声質の音声でプレゼンテーションさせることができる。すなわち、観客の反応を見ながら最適と思われる声質で自動プレゼンテーションの音声を発生させることができ、プレゼンテーションをより効果的にすることができる。

【０１０２】

また、ディスクジョッキーのパフォーマンスにおいて、歌声を用いたモーフィングにより、歌唱にリアルタイムで変化する表情付けを行なうこともできる。例えばバックミュージックの方が歌よりも重要な場合には歌声は平坦な（表情のない）ものとし、歌詞に注目してほしいときには「ねっとりした」表情付け音声にモーフィングし、ゆっくり歌い終わるときには次第に「ささやき」に表情付けをモーフィングし、など、同一歌唱曲中で連続的に音声モーフィングを行なうことができる。こうすることで、ディスクジョッキーが思う通りの表情付けで歌唱を再生させることができる。

【０１０３】

３つ以上の音声のモーフィング率を指定する方法は、上記したようにモニタ上のポインタまたは３次元的なポインタだけでなく、最初に説明したのと同様、スライダによって指定することもできる。その場合のモニタ表示例を図１５に示す。

【０１０４】

図１５を参照して、モニタ表示３４０には、３つの音声に対応するスライダ３５０、３５２および３５４を表示する。これらスライダ３５０、３５２および３５４の左端には、各音声に付されたラベルを表示するラベルテキスト領域３７０、３７２および３７４が表示される。スライダ３５０、３５２および３５４のつまみ３６０、３６２、３６４を左右にスライドさせることにより、各音声のモーフィング率を調整できる。なおこの場合、スライダ目盛の数値そのものはモーフィング率に対応しない。モーフィング率は、３つのスライダのつまみが指す目盛の値の合計を基準（１００％）とし、合計に対する各スライダの目盛の値の率で定めるようにすればよい。

【０１０５】

さらに、平坦な音声を基準として、任意の表情付けを行ないながら音声を発生させる場合にも本発明を適用することができる。平坦な音声も含めて音声の種類が４種類の場合について、図１６～図１８を参照して音声の選択方法について説明する。

【０１０６】

図１６を参照して、原点３８０を持つ立体座標系を考える。原点をｎｏｒｍａｌというラベルを持つ音声に割当て、３軸をそれぞれ３種類の音声Ａ、Ｂ、およびＣに割当てる。これら３軸をそれぞれ音声Ａ軸、音声Ｂ軸、および音声Ｃ軸と呼ぶことにする。

【０１０７】

図１７を参照して、例えば原点３８０と音声Ａ軸との間の中間表情付け３９０、原点３８０と音声Ｂ軸との間の中間表情付け３９２、および原点３８０と音声Ｃ軸との間の中間表情付け３９４を、それぞれ第１の実施の形態と同様に行なうことができる。これらの中

間表情付けによって発生される音声は、原点３８０に対応する平坦な音声から、表情付けの種類の強度を音声Ａ、Ｂ、Ｃ軸に沿って変化させて発生したものとなる。

【０１０８】

一方、音声Ａ軸と音声Ｂ軸との間の中間表情付け４００、音声Ｂ軸と音声Ｃ軸との間の中間表情付け４０４、および音声Ｃ軸と音声Ａ軸との間の中間表情付け４０２も考えられる。これらは、ある種類の表情付けがされた軸上の音声を、他軸上の別種類の表情付けがされた音声に変えることであるから、音声に対する表情付けの「種類」の変化に相当すると考えることができる。

【０１０９】

図１６および図１７を参照して説明したような方向の中間音声の発生だけでなく、それらを所定の割合で混合した音声も発生可能である。その方法は図１１を参照して説明した方法と同様である。

【０１１０】

例えば、図１８に示すように、音声Ａ軸、音声Ｂ軸、および音声Ｃ軸の各々において、各音声の割合が１００％となる点４２０、４２２、および４２４を決める。次にこれら３点４２０、４２２、および４２４を互いに結ぶことによって得られる空間上の３つの半直線４１２、４１４および４１６を考える。これら３つの半直線４１２、４１４および４１６は空間上で一つの三角形を規定する。この三角形を図１１に示す方法と同様に分割することでメッシュ４１０が得られる。

【０１１１】

さらに、音声Ａ軸、音声Ｂ軸および音声Ｃ軸のそれぞれにおいて、原点３８０と前述した点４２０、４２２および４２４との間を１０分割する。この分割により、各軸上には０％から１００％まで、１１個の点が規定される。それら点のうち、各軸上でそれぞれ１０％に相当する３点を互いに結んで３角形を考えることで、メッシュ４１０と同様のメッシュが形成できる。同様にして、２０％から９０％までの各々の率について、メッシュ４１０と同様のメッシュが形成できる。

【０１１２】

中間音声として、平坦な音声と、音声Ａ、ＢおよびＣとを準備し、これら音声を用いて予め上記のように得られたメッシュの各交点に相当する混合割合の中間音声を作成しておく。音声の発生時には、ユーザが３次元空間上のある点（図１８において原点３８０と３点４２０、４２２および４２４により形成される三角錐内のある点）を指定すると、上記したメッシュの交点のうち、指定された点に最も近い交点を定めることができる。その交点に対応する中間音声で音声を発生させる。

【０１１３】

このようにすることにより、ｎｏｒｍａｌ音声を中心として、３種類の表情付けのうち、任意のものを選択し、任意の強度でｎｏｒｍａｌ音声に対しそれらの表情付けを行なうことができる。さらにそれだけでなく、３種類の音声の種類を互いに入れ替えたり、それら音声の持つ表情を任意の割合で混合した音声を作成したりすることができる。

【０１１４】

もちろん、中心におく音声はｎｏｒｍａｌ音声に限定されるわけではなく、利用者の意図に応じて任意の表情付けを持つ音声を中心としてもよい。もっとも、中心に置く音声をｎｏｒｍａｌ音声とすると、中間音声の表情がどのようなものになるか直感的に分りやすいと思われる。したがって中心にｎｏｒｍａｌ音声を置くことが実用的である。

【０１１５】

なお、複数（例えば二つ）の表情付音声の間でモーフィング率＝０．５としてモーフィングを行なった場合でも、得られる音声はｎｏｒｍａｌなものとは異なるものとして知覚されることが実験によって確認されている。３つ以上の音声についても同様で、全ての表情付音声について互いにモーフィング率が等しくなるような条件でモーフィングを行なったとしても、得られる音声はｎｏｒｍａｌとは異なって知覚されると思われる。したがって、図１６～図１８に示すようにｎｏｒｍａｌ音声を中心におき、この音声を基準として

様々な表情を付ける形でモーフィングを行なうようにすることが好ましい。

【０１１６】

このようにｎｏｒｍａｌ音声を基準として、他に例えば３種類の表情付音声をモーフィングする場合、すなわち４種類の音声の間でのモーフィングを行なう場合でも、図１５に示す３個のスライダ３５０、３５２および３５４を用いてモーフィング率を指定できる。各スライダには、ｎｏｒｍａｌ音声を基準とした３個の表情付音声のモーフィング率を指定する。したがって、３個のスライダにより指定された値が全て０の場合にはｎｏｒｍａｌ音声が発生されることになる。

【０１１７】

さらに、上記したように表情付けの強度を任意に変化させたり、種類の変化を任意に行なわせたりする場合、音声の種類は３種類または４種類に限定されるわけではない。理論的には、５種類以上の音声の間でも中間音声を同様にして定め、利用することができる。

【０１１８】

なお、上記した実施の形態では同一話者による複数声質の音声を用いて音声モーフィングを行なっている。しかし本発明はそのような実施の形態には限定されない。別の話者による音声の間での音声モーフィングを行なってもよい。発声される内容はテキスト朗読でもよいし、ある歌の歌声でもよい。

【０１１９】

さらに、上記した実施の形態では、２種類の音声の間で音声モーフィングを行なって得る中間音声の数を９個としたが、中間音声の数が９個に限定されるわけではないことはもちろんである。一般的には、基準となる音声の数が多くなると、音声の変化を滑らかにするためには中間音声の数を多くする必要がある。

【０１２０】

今回開示された実施の形態は単に例示であって、本発明が上記した実施の形態のみに制限されるわけではない。本発明の範囲は、発明の詳細な説明の記載を参酌した上で、特許請求の範囲の各請求項によって示され、そこに記載された文言と均等の意味および範囲内でのすべての変更を含む。

【図面の簡単な説明】

【０１２１】

【図１】本発明の一実施の形態に係る音声モーフィングシステム２０の全体の概略ブロック図である。

【図２】音声モーフィングシステム２０を実現するコンピュータシステム５０の外観図である。

【図３】コンピュータシステム５０のブロック図である。

【図４】図１に示す音声モーフィング装置２２の概略ブロック図である。

【図５】モーフィング率を決定するためのシグモイド関数を示すグラフである。

【図６】モーフィング後音声作成処理の制御の流れを示すフローチャートである。

【図７】図１に示すモーフィング音声選択・再生装置３２のブロック図である。

【図８】モーフィング音声選択・再生装置３２における入出力画面１６４を示す図である。

【図９】スライド目盛の軌跡および音声シーケンスの記録例を示す図である。

【図１０】入出力画面１６４のラベルテキスト領域２１０において音声ラベルを変更した状態を示す図である。

【図１１】３種類の音声から中間音声を作成する方法を説明するための図である。

【図１２】３種類の音声から作成された中間音声のうち、いずれを選択するかに関する方法を説明するための図である。

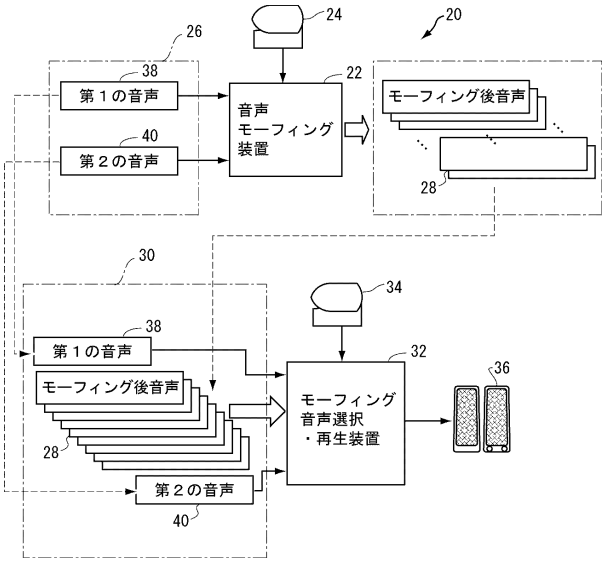
【図１３】３種類の音声から作成した二つの中間音声の間でさらに中間音声を作成する方法を説明するための図である。

【図１４】３次元センサによる中間音声の選択方法を説明するための図である。

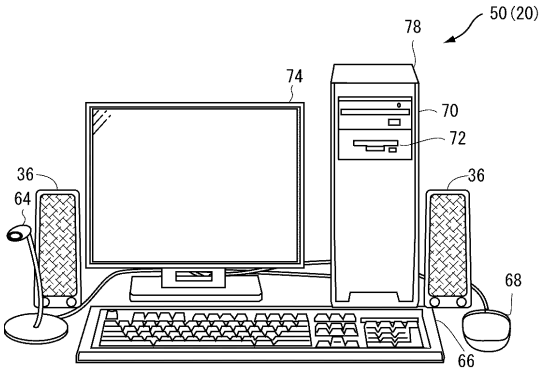
【図１５】３種類の音声から得られた中間音声を選択する際のスライド表示を説明するた

めの図である。
【図１６】声質の種類の選択および各音声種類における声質の強さの指定のために使用する座標系を説明するための図である。
【図１７】平坦な音声と３種類の音声との間の声質の種類と強度との指定方法を説明するための図である。
【図１８】平坦な音声と３種類の音声との間での中間音声の指定方法を説明するための図である。
【符号の説明】
【０１２２】
２０　音声モーフィングシステム
２２　音声モーフィング装置
２４　入出力インタフェース
２６　基準音声記憶装置
２８　モーフィング後音声
３０　音声データ記憶装置
３２　モーフィング音声選択・再生装置
３４　ユーザインタフェース
３８　第１の音声
４０　第２の音声
５０　コンピュータシステム
１００　パラメータ入力部
１０２　シグモイド関数決定部
１０４　モーフィング率決定部
１０６　繰返し制御部
１０８　モーフィング実行部
１１０　保存処理部
１２０　音声記憶装置

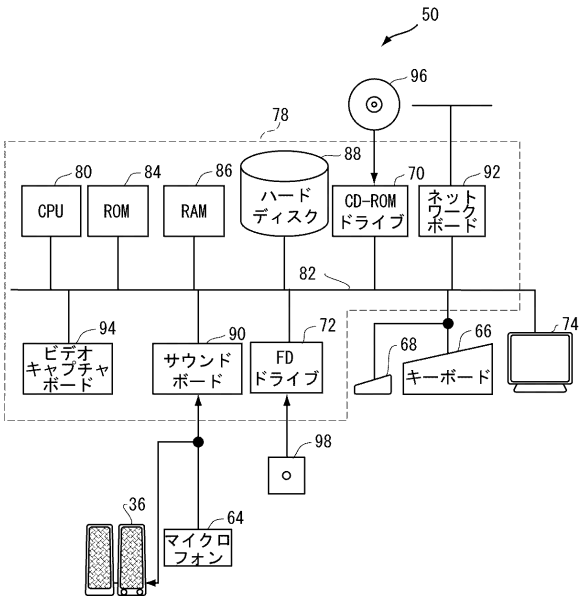
【図１】


24
20
26
38
第１の音声
40
第２の音声
22
音声
モーフィング
装置
モーフィング後音声
28
30
34
38
第１の音声
モーフィング後音声
28
32
モーフィング
音声選択
・再生装置
36
第２の音声
40

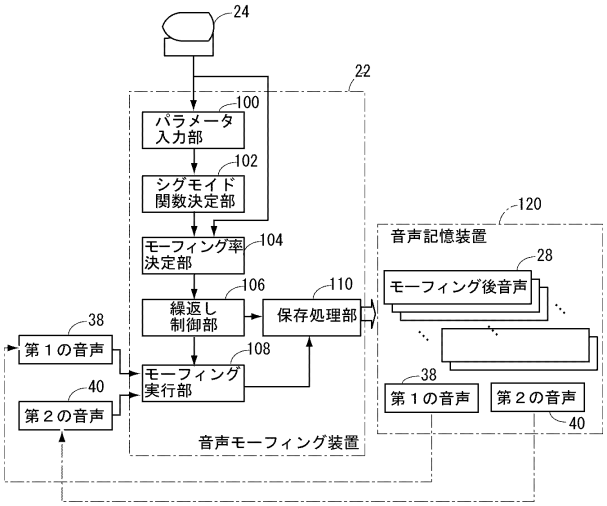
【図２】


50(20)
78
74
70
72
36
64
36
68
66

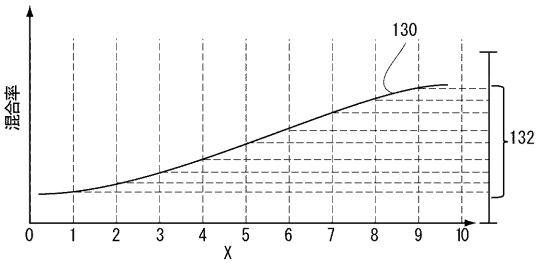
【図３】


50
96
78
88
80
84
86
70
92
CPU
ROM
RAM
ハード
ディスク
CD-ROM
ドライブ
ネット
ワーク
ボード
82
94
90
72
68
66
74
ビデオ
キャプチャ
ボード
サウンド
ボード
FD
ドライブ
キーボード
98
36
64
マイクロ
フォン

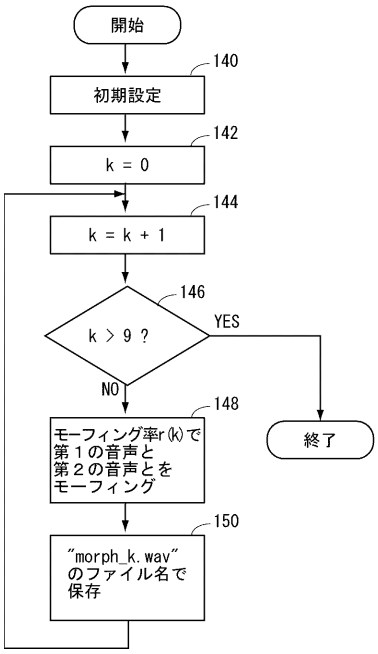
【図４】


24
22
100
パラメータ
入力部
102
シグモイド
関数決定部
104
モーフィング率
決定部
106
繰返し
制御部
110
保存処理部
108
モーフィング
実行部
38
第１の音声
40
第２の音声
音声モーフィング装置
120
音声記憶装置
28
モーフィング後音声
38
第１の音声
第２の音声
40

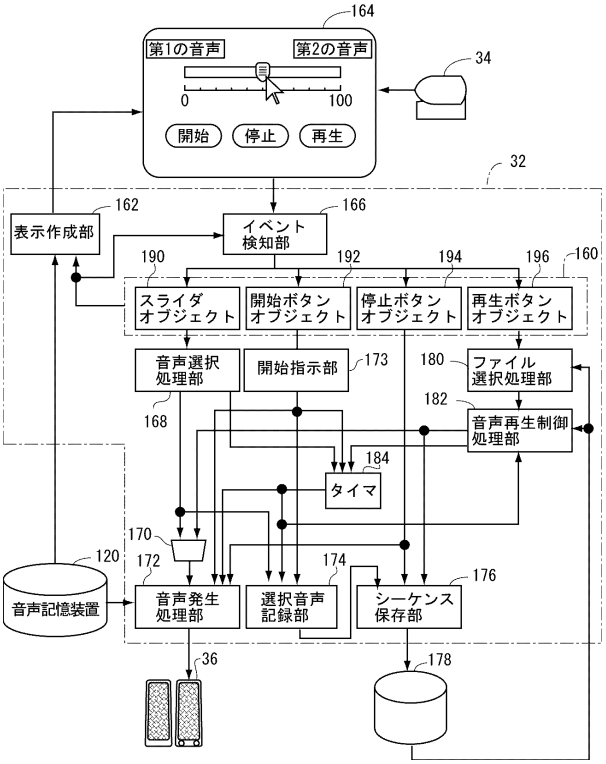
【図５】


130
混合率
132
0
1
2
3
4
5
6
7
8
9
10
X

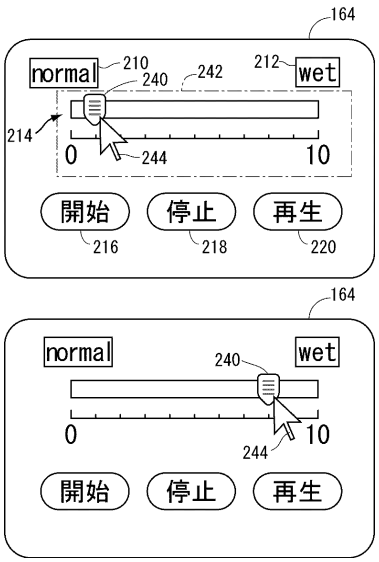
【図６】



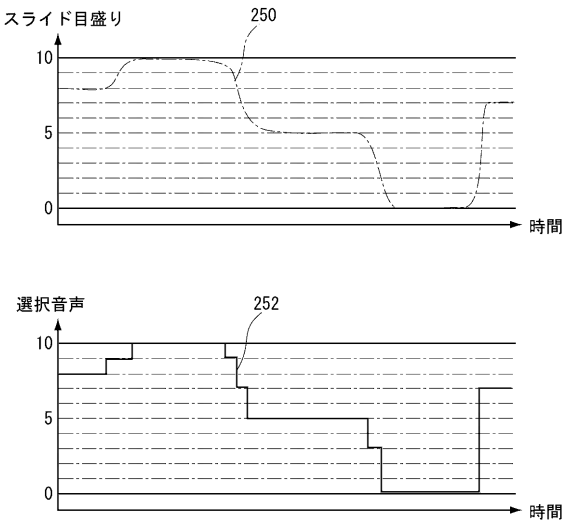
【図７】



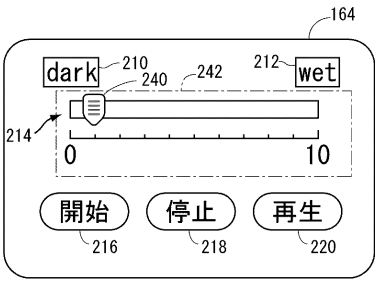
【図８】



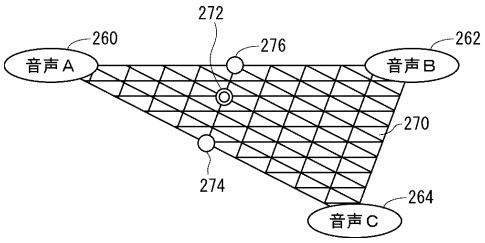
【図９】



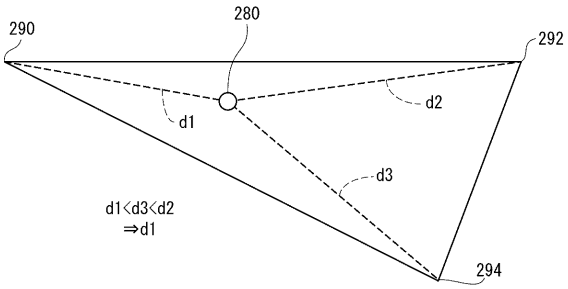
【図１０】



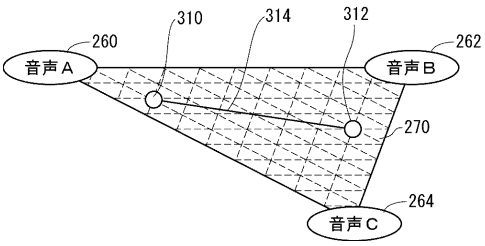
【図１１】



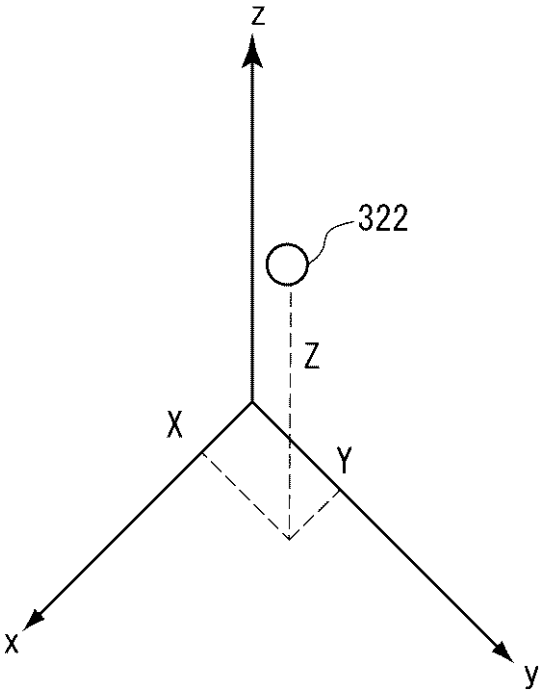
【図１２】



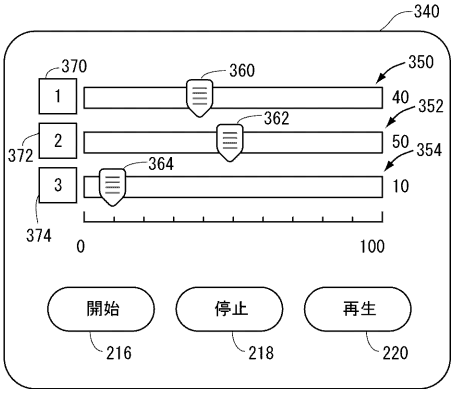
【図１３】



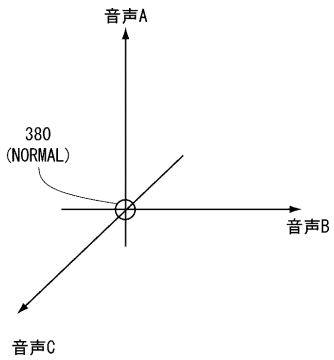
【図１４】



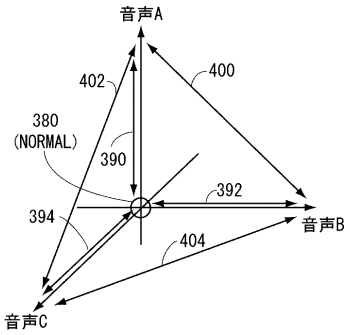
【図１５】



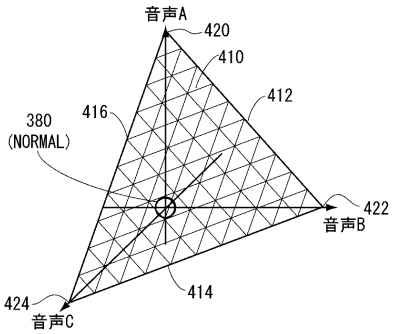
【図１６】



【図１７】


音声A
400
402
380
(NORMAL)
390
392
394
音声B
404
音声C

【図１８】


音声A
420
410
412
416
380
(NORMAL)
422
音声B
414
424
音声C

フロントページの続き

審査官　清水　正一

(56)参考文献　特開２００１－１０９９０１（ＪＰ，Ａ）
特開２００２－０９４８８１（ＪＰ，Ａ）
特開２００４－００５２６５（ＪＰ，Ａ）
特開２００３－２９５８８２（ＪＰ，Ａ）
特開２００３－２１９２６２（ＪＰ，Ａ）
特開平０９－０５０２９５（ＪＰ，Ａ）
特開平０９－１５２８９２（ＪＰ，Ａ）
特開２００３－２８３６１３（ＪＰ，Ａ）
特開２００１－１１７５６４（ＪＰ，Ａ）
特開平１０－２５４５００（ＪＰ，Ａ）
特開平０９－２４４６９３（ＪＰ，Ａ）
特開平０９－１４６５９７（ＪＰ，Ａ）
特開２００４－１０２１１８（ＪＰ，Ａ）
特開２００２－３３３８９７（ＪＰ，Ａ）
坂野　秀樹，武田　一哉，鹿野　清宏，板倉　文忠，包絡と音源の独立操作による音声モーフィング，電子情報通信学会論文誌，日本，社団法人電子情報通信学会，１９９８年　2月，第J81-A巻
鈴木　紀子、米澤　朋子、片桐　恭弘，男性－女性間モーフィング音声の印象評定，日本音響学会２００４年秋季研究発表会講演論文集－Ｉ－，日本，社団法人日本音響学会，２００４年　９月２１日，p.409-410

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ１０Ｌ　２１／００－２１／０６
Ｇ１０Ｌ　１３／００－１３／０８