

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5109038号
(P5109038)

(45) 発行日 平成24年12月26日(2012.12.26)

(24) 登録日 平成24年10月19日(2012.10.19)

(51) Int. Cl.		F I	
G06T 13/40	(2011.01)	G06T 13/40	
G10L 15/00	(2006.01)	G10L 15/00	200G
G10L 13/00	(2006.01)	G10L 13/00	100S

請求項の数 16 (全 62 頁)

(21) 出願番号	特願2008-210639 (P2008-210639)	(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
(22) 出願日	平成20年8月19日(2008.8.19)	(74) 代理人	100099933 弁理士 清水 敏
(65) 公開番号	特開2009-87328 (P2009-87328A)	(72) 発明者	川本 真一 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
(43) 公開日	平成21年4月23日(2009.4.23)	(72) 発明者	四倉 達夫 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
審査請求日	平成23年8月11日(2011.8.11)	(72) 発明者	中村 哲 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
(31) 優先権主張番号	特願2007-234239 (P2007-234239)		
(32) 優先日	平成19年9月10日(2007.9.10)		
(33) 優先権主張国	日本国(JP)		

最終頁に続く

(54) 【発明の名称】 リップシンクアニメーション作成装置及びコンピュータプログラム

(57) 【特許請求の範囲】

【請求項1】

予め準備された統計的音響モデルと、予め準備された音素及び視覚素の間のマッピング定義と、前記視覚素に対応する、予め準備された複数個の顔画像の顔モデルとを用い、入力される発話データからリップシンクアニメーションを作成するためのリップシンクアニメーション作成装置であって、

前記統計的音響モデル、前記マッピング定義、及び前記発話データに対するトランスクリプションを使用して、前記発話データに含まれる音素及び対応する視覚素を求め、デフォルトのブレンド率が付与された継続長付きの視覚素シーケンスを作成するための視覚素シーケンス作成手段を含み、前記視覚素シーケンスの継続長内の所定位置にはキーフレームが定義され、前記視覚素シーケンスの各視覚素の継続長内に定義されるキーフレームによりキーフレームシーケンスが定義され、

前記リップシンクアニメーション作成装置はさらに、前記キーフレームシーケンス内のキーフレームのうち、視覚素に対応する顔モデルの形状の変化の速さが最も大きいものから順番に、所定の割合又は所定の数のキーフレームを削除するためのキーフレーム削除手段と、

前記キーフレーム削除手段により一部のキーフレームが削除されたキーフレームシーケンスに基づき、キーフレーム間のブレンドにより顔画像のアニメーションを作成するためのブレンド処理手段とを含み、

前記視覚素シーケンス作成手段は、前記キーフレームシーケンスの作成時には、第1の

フレームレートのフレームの任意のものをキーフレームとして選択し、

前記リップシンクアニメーション作成装置はさらに、前記第1のフレームレートよりも小さな第2のフレームレートを指定する入力と、前記キーフレーム削除手段により出力されるキーフレームシーケンスとを受けると接続され、前記キーフレーム削除手段により出力されるキーフレームシーケンスを、前記第2のフレームレートのキーフレームシーケンスに変換するためのフレームレート変換手段を含み、

前記フレームレート変換手段は、前記第2のフレームレートのキーフレームシーケンスの各キーフレームに、前記キーフレーム削除手段の出力するキーフレームシーケンス内で、当該キーフレームの継続長内に始端を有するキーフレームに割当てられた視覚素のいずれかを割当て、

10

前記ブレンド処理手段は、前記フレームレート変換手段によりフレームレートが変換された前記キーフレームシーケンスに基づき、キーフレーム間のブレンドにより顔画像のアニメーションを作成するための手段を含む、リップシンクアニメーション作成装置。

【請求項2】

前記キーフレーム削除手段は、前記キーフレームシーケンス内のキーフレームのうち、当該キーフレームの視覚素に対応する顔モデルを構成する各特徴点と、隣接するキーフレームの視覚素に対応する顔モデルを構成する、対応する各特徴点との間の変化の速さが最も大きいものから順番に、所定の割合又は所定の数のキーフレームを削除するための手段を含む、請求項1に記載のリップシンクアニメーション作成装置。

【請求項3】

20

前記複数個の顔モデルの内から選ばれる2個の顔モデルの組合せの全てに対し、顔モデルを構成する特徴点の間の動きベクトルを算出するための動きベクトル算出手段と、

前記2個の顔モデルの特徴点を、前記動きベクトル算出手段により算出された動きベクトルに対する所定のクラスタリング方法によってクラスタ化し、各クラスタの代表ベクトルを算出することにより、クラスタ化された顔モデルを作成するための手段と、

前記クラスタ化された顔モデルを記憶するためのクラスタ化顔モデル記憶手段とをさらに含み、

前記キーフレーム削除手段は、

前記キーフレームシーケンス内のキーフレームの各々に対し、当該キーフレームの視覚素と、隣接するキーフレームの視覚素との組合せに対応するクラスタ化された顔モデルを前記クラスタ化顔モデル記憶手段から読み出し、各クラスタに属する特徴点のキーフレーム間の変化の速さを当該クラスタの代表ベクトルを用いて算出するための移動量算出手段と

30

、
前記移動量算出手段により算出された変化の速さが最も大きいものから順番に、所定の割合又は所定の数のキーフレームを前記キーフレームシーケンスから削除するための手段とを含む、請求項1に記載のリップシンクアニメーション作成装置。

【請求項4】

前記キーフレーム削除手段により一部のキーフレームが削除されたキーフレームシーケンスを受け、当該キーフレームシーケンス内のキーフレームの視覚素に対応する音素の発話パワーを前記発話データから算出するための発話パワー算出手段と、

40

前記キーフレームシーケンス内の各キーフレームに対し、前記発話パワー算出手段により、当該キーフレームを含む視覚素の継続長について算出された平均発話パワーが小さければ小さいほどブレンド率が小さくなるような、前記平均発話パワーに対する所定の関数により、ブレンド率を調整するための、発話パワーによるブレンド率調整手段とをさらに含み、

前記ブレンド処理手段は、前記発話パワーによるブレンド率調整手段によってブレンド率が調整されたキーフレームシーケンスに基づき、キーフレーム間のブレンドにより顔画像のアニメーションを作成する、請求項1に記載のリップシンクアニメーション作成装置。

【請求項5】

50

前記キーフレーム削除手段により一部のキーフレームが削除されたキーフレームシーケンスを受け、前記キーフレームの視覚素に対応する顔モデルを構成する頂点と、隣接するキーフレームの視覚素に対応する顔モデルを構成する頂点との間の変化の速さを算出するための変化の速さ算出手段と、

前記キーフレーム削除手段により一部のキーフレームが削除されたキーフレームシーケンスに含まれる各キーフレームのうち、前記変化の速さ算出手段により算出された変化の速さが所定のしきい値よりも大きなキーフレームについて、そのブレンド率がより小さな値となるような、前記変化の速さに対する所定の関数を用いてブレンド率を更新するための頂点速度によるブレンド率調整手段とをさらに含み、

前記ブレンド処理手段は、前記頂点速度によるブレンド率調整手段によってブレンド率が調整されたキーフレームシーケンスに基づき、キーフレーム間のブレンドにより顔画像のアニメーションを作成する、請求項1に記載のリップシンクアニメーション作成装置。

【請求項6】

予め準備された統計的音響モデルと、予め準備された音素及び視覚素の間のマッピング定義と、予め準備された複数個の顔画像の顔モデルとを用い、入力される発話データからリップシンクアニメーションを作成するためのリップシンクアニメーション作成装置であって、前記発話データに対するトランスクリプションが利用可能であり、

前記統計的音響モデル、前記マッピング定義、及び前記トランスクリプションを使用して、前記発話データに含まれる音素及び対応する視覚素を求め、デフォルトのブレンド率が付与された継続長付きの視覚素シーケンスを作成するための視覚素シーケンス作成手段を含み、

前記視覚素シーケンスの継続長内の所定位置にはキーフレームが定義され、前記視覚素シーケンスの各視覚素の継続長内に定義されるキーフレームによりキーフレームシーケンスが定義され、

前記キーフレームシーケンス内のキーフレームの視覚素に対応する音素の発話パワーを前記発話データから算出するための発話パワー算出手段と、

前記キーフレームシーケンス内の各キーフレームに対し、前記発話パワー算出手段により、当該キーフレームを含む視覚素の継続長について算出された平均発話パワーが小さければ小さいほどブレンド率が小さくなるような、前記平均発話パワーに対する所定の関数により、ブレンド率を調整するための、発話パワーによるブレンド率調整手段と、

前記ブレンド率調整手段によりブレンド率が調整された視覚素シーケンスに基づき、キーフレーム間のブレンドにより顔画像のアニメーションを作成するためのブレンド処理手段とを含み、

前記視覚素シーケンス作成手段は、前記キーフレームシーケンスの作成時には、第1のフレームレートのフレームの任意のものをキーフレームとして選択し、

前記リップシンクアニメーション作成装置はさらに、前記第1のフレームレートよりも小さな第2のフレームレートを指定する入力と、前記ブレンド率調整手段により出力されるキーフレームシーケンスとを受けるとに接続され、前記ブレンド率調整手段により出力されるキーフレームシーケンスを、前記第2のフレームレートのキーフレームシーケンスに変換するためのフレームレート変換手段を含み、

前記フレームレート変換手段は、前記第2のフレームレートのキーフレームシーケンスの各キーフレームに、前記ブレンド率調整手段の出力するキーフレームシーケンス内で、当該キーフレームの継続長内に始端を有するキーフレームに割当てられた視覚素のいずれかを割当て、

前記ブレンド処理手段は、前記フレームレート変換手段によりフレームレートが変換された前記キーフレームシーケンスに基づき、キーフレーム間のブレンドにより顔画像のアニメーションを作成するための手段を含む、リップシンクアニメーション作成装置。

【請求項7】

前記視覚素シーケンス作成手段の出力するキーフレームシーケンスに含まれるキーフレームのうち、空白音素に対応する視覚素が割当てられたキーフレームの直前のキーフレーム

10

20

30

40

50

の継続長の終端位置を、当該キーフレーム内の前記発話データの発話パワー系列の最大点以後で、かつ当該キーフレームの継続長内の位置に移動させることにより、発話終端位置を補正するための発話終端補正手段をさらに含み、

前記キーフレーム削除手段は、前記発話終端補正手段により発話終端が補正されたキーフレームシーケンスを入力として受ける請求項 1 ~ 請求項 6 のいずれかに記載のリップシンクアニメーション作成装置。

【請求項 8】

前記フレームレート変換手段は、前記第 2 のフレームレートのキーフレームシーケンスの各キーフレームに割当てた視覚素が、直前のキーフレームに割当てた視覚素と異なるものとなるように視覚素を割当て、請求項 1 ~ 請求項 7 のいずれかに記載のリップシンクアニメーション作成装置。

10

【請求項 9】

前記リップシンクアニメーション作成装置は、前記複数個の顔画像の顔モデルを記憶するための顔モデル記憶手段をさらに含む、請求項 1 ~ 請求項 8 のいずれかに記載のリップシンクアニメーション作成装置。

【請求項 10】

前記予め準備された音素は、予め定められた標準音素と、前記標準音素以外の一般音素とを含み、

前記複数個の顔画像の顔モデルは、前記標準音素に対応する顔モデルから成る標準視覚素モデルと、前記一般音素に対応する顔モデルから成る一般視覚素モデルとを含み、

20

前記リップシンクアニメーション作成装置はさらに、前記予め準備された音素に対応して予め分類された、対応する音素を発話しているときの発話者の顔画像の特徴点の 3 次元位置の実測値から成るキャプチャデータと前記標準視覚素モデルとを用い、前記一般視覚素モデルを生成するための一般視覚素生成手段を含む、請求項 9 に記載のリップシンクアニメーション作成装置。

【請求項 11】

前記一般視覚素生成手段は、前記標準音素に対応する前記キャプチャデータの線形和で、前記一般音素に対応する前記キャプチャデータを近似するための、前記標準音素の数と同数の係数を、所定の近似誤差を最小とするように算出するための係数算出手段と、

前記一般視覚素モデルを、当該一般視覚素モデルに対応する一般音素について前記係数算出手段により算出された係数を用いた前記標準視覚素モデルの線形和により計算し、前記標準視覚素モデルとともに対応する一般音素と関連付けて前記顔モデル記憶手段に記憶させるための線形和計算手段とを含む、請求項 10 に記載のリップシンクアニメーション作成装置。

30

【請求項 12】

コンピュータにより実行されると、当該コンピュータを、請求項 1 ~ 請求項 11 のいずれかに記載のリップシンクアニメーション作成装置として機能させる、コンピュータプログラム。

【請求項 13】

前記所定の関数は、前記平均発話パワーの線形関数である、請求項 4 又は請求項 6 に記載のリップシンクアニメーション作成装置。

40

【請求項 14】

前記所定の関数は、前記平均発話パワーの非線形関数である、請求項 4 又は請求項 6 に記載のリップシンクアニメーション作成装置。

【請求項 15】

前記所定の関数は、前記変化の速さの線形関数である、請求項 5 に記載のリップシンクアニメーション装置。

【請求項 16】

前記所定の関数は、前記変化の速さの非線形関数である、請求項 5 に記載のリップシンクアニメーション作成装置。

50

【発明の詳細な説明】

【技術分野】

【0001】

この発明は音声からアニメーションを作成するアニメーション作成装置に関し、特に、発話音声にあわせて口等の形が変わる顔画像等のアニメーションを自動的に生成する装置に関する。

【背景技術】

【0002】

コンピュータ技術の発達により、以前は大部分が手作業で行なわれていた仕事がコンピュータによる作業に置き換えられるケースが多くなっている。その代表的なものに、アニメーションの作成がある。

【0003】

以前は、アニメーションといえば次のような手法で作成されることが一般的であった。登場するキャラクタをアニメーションの演出家が決め、絵コンテと呼ばれる、主要なシーンのラフな原画を作成する。これら絵コンテに基づき、アニメーションの各フレームの絵をアニメーターと呼ばれる作業者が作成する。それら絵を仕上げ担当者がセル画に仕上げる。セル画を順にフィルムに写し、所定のフレームレートで再生すればアニメーションの画像の部分が出来上がる。

【0004】

このアニメーションの画像を再生しながら、声優がアニメーションの台本に基づいて台詞をつけていく。いわゆる「アフレコ」である。

【0005】

このような作業で最も人手がかかるのはセル画の作成である。一方、原画をCG（コンピュータ・グラフィックス）で作成する場合、原画を加工してセル画を作成するのは比較的単純な作業である。一枚一枚撮影する必要もない。そのため、この部分については原画のCG化とあわせてかなりコンピュータ化されている。

【0006】

一方、残りの作業のうちで比較的むずかしいのは、アフレコの作業である。アニメーションの動きにあわせて、なおかつ状況にあわせた声で台詞をしゃべる必要があるため、アフレコの作業にはそれなりの時間がかかり、習熟も必要である。

【0007】

そこで、アフレコの逆に、先に音声を収録し、その音声にあわせてアニメーションを作成する手法が考えられた。これは「プレスコ」又は「プレレコ」（以下「プレスコ等」と呼ぶ。）と呼ばれる。これはもともと米国等で手作業でアニメーションを作成する際に採用されていた手法である。この手法でアニメーションを作成する場合には、次のような作業手順となる。

【0008】

まず、アニメーションに登場するキャラクタを決める。絵コンテも従来と同様に作成する。声優が、絵コンテと台本に基づいて発話し、それを音声として収録する。この音声にあわせて、アニメーションを作成する。

【0009】

このプレスコ等の手法によるアニメーション作成をコンピュータで実現する場合には、音声からアニメーションをいかにして自動的に作成するか、という点が問題となる。特に、人物等のアニメーションの口の動きを、予め録音した声優の音声にあわせて自然な形で生成するのは難しく、これを自動的にこなす手法が求められている。

【0010】

このための一手法として提案されたものに、特許文献1に記載された手法がある。特許文献1に記載された手法では、口形状の基本パターンを予め複数個用意しておく。そして、任意の音声に対応する口形状を、これら基本パターンの加重和により求める。そのため

10

20

30

40

50

に、声優の音声の所定の特徴量から、各基本パターンの加重パラメータに変換するための変換関数を、重回帰分析によって予め求めておく。台本に沿って録音された声優の音声の所定の特徴量をこの変換関数で加重パラメータに変換し、その加重パラメータを用いて口形状の基本パターンの加重和を算出することで、声優の音声に対応する口形状及び顔画像を作成する。こうした処理をアニメーションの各フレームに相当する時刻に行なうことで、アニメーションのフレームシーケンスを作成する。

【 0 0 1 1 】

図 1 に、このような従来のアニメーション作成装置の前提となるアニメーション作成過程 3 0 の概略を示す。図 1 を参照して、アニメーション作成過程 3 0 においては、話者 4 0 が台本 4 4 に基づき台詞を発話すると、その音声信号 4 2 に対し、音声認識装置による音素セグメンテーション（発話から、発話を構成する音素列を生成すること）が行なわれる。

10

【 0 0 1 2 】

予め、主要な音素については、その音素を発音するときの口の形状を含む顔画像 6 0 ~ 6 8 が準備されており、音声認識の結果得られる各音素 5 0 ~ 5 8 に対し、これら顔画像を割当ててアニメーション化する。

【 0 0 1 3 】

なお、個々の音素に対して発話画像を一つずつ割当てても滑らかな画像が得られないため、特許文献 1 にも記載のように、主要な画像の間の加重和により、中間の画像を作成する。例えば、主要な顔画像として「あ（ / a / ）」「い（ / i / ）」「う（ / u / ）」「え（ / e / ）」「お（ / o / ）」という 5 つの音素に対する 5 つの顔画像、及び音素「ん / N / 」に対する顔画像の、合計 6 つの顔画像を準備する。「ん / N / 」に対する顔画像は後述するように他の顔画像の基本となる画像であり、本明細書では「無表情の顔画像」とも呼ぶ。「あ」～「お」の 5 つの音素はそれぞれ対応の顔画像に割当て、残りの音素についてはそれぞれ上記した 6 つの顔画像のいずれかに割当てる。これを以下、音素から顔画像へのマッピングと呼ぶ。

20

【 0 0 1 4 】

図 2 に、使用される顔画像の例を示す。顔画像は、他の全ての顔画像の基本となる無表情の顔画像 8 0 と、前述した「あ」～「お」までの顔画像 6 0 ~ 6 8 とを含む。顔画像 6 0 ~ 6 8 は、ワイアフレーム画像に予め準備した顔のテクスチャを貼り付けることで生成する。顔画像 6 0 ~ 6 8 及び 8 0 のワイアフレーム画像は、いずれもワイアフレームを構成する各頂点の 3 次元座標により定義される。ただし、基本となる無表情の顔画像 8 0 については各頂点の座標が予め定義されるが、顔画像 6 0 ~ 6 8 の各頂点の座標は、無表情の顔画像 8 0 に対する相対座標により定義される。顔画像 6 0 ~ 6 8 及び 8 0 を構成する各頂点の座標の集合からなる顔モデルを以下「視覚素」と呼ぶ。

30

【 0 0 1 5 】

このように準備した顔画像に基づいてアニメーションを作成する場合、従来は以下のような手作業による手順を採っている。すなわち、音声を聞きながら、ある時点での「あ」の音声の発話時に「あ」の顔画像を割当て、「お」の音声の発話時に「お」の顔画像を割当てる、という作業を、そのような割当てが必要と思われるフレームの全てに対して手作業で行なう。このように特定の音声の発話時の顔画像が割当てられたフレームを「キーフレーム」と呼ぶ。

40

【 0 0 1 6 】

次に、このようにして割当てられたキーフレームに基づき、キーフレームの間の任意の時点の顔画像を、その時点をはさむ二つのキーフレームに割当てられた顔画像の間のブレンドによって合成する。

【 0 0 1 7 】

図 3 に、キーフレームの割当て例を示す。図 3 に示す例では、「あ」を表す顔画像 6 0 については、縦棒 1 0 0 及び 1 0 2 で示されるように、二つのキーフレームに割当てられている。同様に、顔画像 6 2 については縦棒 1 1 0 により、顔画像 6 4 については縦棒 1 2

50

0により、顔画像66については縦棒130により、そして顔画像68については縦棒140により、それぞれ示されるように、一つのフレームに割当てられている。

【0018】

これらフレーム(キーフレーム)での顔画像は、指定された顔画像と一致するように作成されるが、それ以外のフレームでは、そのフレームをはさむ二つのキーフレームの顔画像の間のブレンドにより作成される。特許文献1でいう「加重和」がこれに相当する概念である。図3のグラフ104、112、122、132、及び142は、それぞれ顔画像60~68のブレンド率を表したものである。ブレンド率=0の区間ではその顔画像はアニメーション作成に使用されない。ブレンド率0の区間では、その顔画像とブレンド率とを掛け合わせたものを、他の顔画像とそのブレンド率とを掛け合わせたものと加算して顔画像を作成する。

10

【0019】

ブレンド率とは、特定の顔画像を100%、顔画像/N/を0%として、顔画像/N/から特定の顔画像に至るまでの特徴点の移動量の割合で中間の顔画像を表すものである。従って、顔画像/A/、/I/、/U/、/E/、/O/をそのまま音素に割当てた場合、そのブレンド率はいずれも100%となる。ブレンド率50%の顔画像/A/とは、顔画像/N/からの特徴点の移動量の割合が、顔画像/A/の特徴点の移動量の50%となっているような顔画像のことをいう。顔画像/N/での位置を始点とするベクトルで顔画像の特徴点の移動量を表せば、ブレンド率B%の顔画像とは、各特徴点を表すベクトルが、方向はブレンド率100%の顔画像のベクトルと等しく、長さがブレンド率B%に相当するだけ縮小されたものとなっている顔画像に相当する。

20

【0020】

図4に、このようにしてブレンドにより作成された顔画像の例を示す。図4(A)には、/a/の顔画像に対するブレンド率が100%のときの顔画像を示す。図4(D)には、/i/の顔画像に対するブレンド率が100%のときの顔画像を示す。図4(B)には、/a/のブレンド率65%、/i/のブレンド率35%のときの顔画像を、図4(C)には、/a/のブレンド率35%、/i/のブレンド率65%のときの顔画像を、それぞれ示す。

【0021】

図4(A)~(D)から分かるように、ブレンド率を変化させて二つの顔画像をモデル上でブレンドして新たな顔画像を作成することにより、二つの顔画像の中間的な顔画像を作成できる。

30

【0022】

【特許文献1】特開平7-44727号公報

【非特許文献1】Linde Y., Buzo A., Gray R., "An algorithm for vector quantizeer design," IEEE Transactions on Communications. COM-28 (1980), 84-95.

【発明の開示】

【発明が解決しようとする課題】

【0023】

上記した従来技術によって自動的に顔画像のアニメーションを作成する場合、どこにキーフレームを設定するか、及びそのブレンド率をどのように設定するかが問題となる。従来はいずれも人間が手作業で行なっており、その結果得られるアニメーションはかなり高い品質となっている。しかし、キーフレームとそのブレンド率とを自動的に設定することができ、かつ人間の手作業による結果と同様に滑らかなアニメーションを作成できる技術については、従来は知られていない。

40

【0024】

キーフレームの設定及びブレンド率の設定は、上記したブレンドによるアニメーションの作成において最も重要で、かつ熟練を要する作業であり、この作業を自動化する技術が望まれている。

50

【 0 0 2 5 】

また、アニメーションは、映画とは異なり、単に滑らかな映像が得られれば良い、というものではない。例えば、従来の手作業によるアニメーションでは、単位時間あたりのフレーム数が少ないため、動きがぎこちない、という問題があったが、こうした弱点を逆にアニメーションの魅力であると感じる人もいる。リップシンクアニメーションでも、必要であればこのように手作業によるアニメーションのような動きを実現できることが望ましい。

【 0 0 2 6 】

さらに、文化のグローバル化に伴い、外国で日本語のアニメーションが作成されることも多くなってきたが、今後は日本語で作成したアニメーションを外国での放送用に変更することも考えられる。従来は、映画と同じようにいわゆる吹替えによってこれを実現しているが、吹替えの場合にはどうしても口の動きと音声とが一致しない。リップシンクアニメーションを使用すると、先に音声を収録してからその音声にあわせてアニメーションを作成するので、こうした問題にはうまく対処することができる。しかしその場合には、それぞれの言語で使用される音声にあわせてアニメーション作成に必要な資源を準備する必要がある。そのような準備作業は、できるだけ少なくすることが望ましい。

10

【 0 0 2 7 】

したがって本発明の目的は、人間の発話の音声データから顔画像のアニメーションを作成する際に、滑らかで自然なアニメーションが得られるようにキーフレーム及びそのブレンド率を自動的に設定できるリップシンクアニメーション作成装置を提供することである。

20

【 0 0 2 8 】

本発明の他の目的は、人間の発話の音声データから顔画像のアニメーションを作成する際に、滑らかで自然なアニメーションも、ぎこちない動きのアニメーションも、必要に応じて得られるようにキーフレーム及びそのブレンド率を自動的に設定できるリップシンクアニメーション作成装置を提供することである。

【 0 0 2 9 】

本発明のさらに他の目的は、多言語の人間の発話の音声データから、それぞれの言語の音声に合致した顔画像のアニメーションを作成する際に、できるだけ作業量を少なくしながら、滑らかで自然なアニメーションが得られるようにキーフレーム及びそのブレンド率を自動的に設定できるリップシンクアニメーション作成装置を提供することである。

30

【課題を解決するための手段】

【 0 0 3 0 】

本発明の第1の局面に係るリップシンクアニメーション作成装置は、予め準備された統計的音響モデルと、予め準備された音素及び視覚素の間のマッピング定義と、視覚素に対応する、予め準備された複数個の顔画像の顔モデルとを用い、入力される発話データからリップシンクアニメーションを作成するためのリップシンクアニメーション作成装置であって、発話データに対するトランスクリプションが利用可能である。このリップシンクアニメーション作成装置は、統計的音響モデル、マッピング定義、及びトランスクリプションを使用して、発話データに含まれる音素及び対応する視覚素を求め、デフォルトのブレンド率が付与された継続長付きの視覚素シーケンスを作成するための視覚素シーケンス作成手段を含む。視覚素シーケンスの継続長内の所定位置にはキーフレームが定義され、視覚素シーケンスの各視覚素の継続長内に定義されるキーフレームによりキーフレームシーケンスが定義される。リップシンクアニメーション作成装置はさらに、キーフレームシーケンス内のキーフレームのうち、視覚素に対応する顔モデルの形状の変化の速さが最も大きいものから順番に、所定の割合又は所定の数のキーフレームを削除するためのキーフレーム削除手段と、キーフレーム削除手段により一部のキーフレームが削除されたキーフレームシーケンスに基づき、キーフレーム間のブレンドにより顔画像のアニメーションを作成するためのブレンド処理手段とを含む。

40

【 0 0 3 1 】

50

視覚素シーケンス作成手段は、統計的音響モデル、マッピング定義、及びトランスクリプションを使用して、発話データから視覚素シーケンスを作成する。この視覚素シーケンスには継続長が付されている。視覚素シーケンスの継続長内の所定位置にはキーフレームが定義され、視覚素シーケンスの各視覚素の継続長内に定義されるキーフレームによりキーフレームシーケンスが定義される。これらのキーフレームからブレンドによりアニメーションを作成することもできるが、そうすると作成されるアニメーションの動きは不自然になる。そこで、キーフレーム削除手段によって、キーフレームシーケンス内のキーフレームのうち、顔モデルの形状の変化の速さが最も大きいものから順番に、所定の割合又は所定の数のキーフレームを削除する。動きが速くなる部分のキーフレームを削除することにより、デフォルトのブレンド率を使用しても、作成されるアニメーションの動きは自然なものとなる。その結果、滑らかで自然なアニメーションが得られるようにキーフレーム及びそのブレンド率を自動的に設定できるリップシンクアニメーション作成装置を提供できる。この割合は、調整可能としてもよい。

10

【 0 0 3 2 】

好ましくは、キーフレーム削除手段は、キーフレームシーケンス内のキーフレームのうち、当該キーフレームの視覚素に対応する顔モデルを構成する各特徴点と、隣接するキーフレームの視覚素に対応する顔モデルを構成する、対応する各特徴点との間の変化の速さが最も大きいものから順番に、所定の割合又は所定の数のキーフレームを削除するための手段を含む。

【 0 0 3 3 】

顔モデルを構成する各特徴点について、隣接するキーフレームとの間での変化の速さを算出することにより、計算量は大きくなるが計算結果に含まれる誤差が少なくなり、自然なアニメーションを作成できる。

20

【 0 0 3 4 】

より好ましくは、リップシンクアニメーション作成装置は、複数個の顔モデルの内から選ばれる2個の顔モデルの組合せの全てに対し、顔モデルを構成する特徴点の間の動きベクトルを算出するための動きベクトル算出手段と、2個の顔モデルの特徴点を、動きベクトル算出手段により算出された動きベクトルに対する所定のクラスタリング方法によってクラスタ化し、各クラスタの代表ベクトルを算出することにより、クラスタ化された顔モデルを作成するための手段と、クラスタ化された顔モデルを記憶するためのクラスタ化顔モデル記憶手段とをさらに含む。キーフレーム削除手段は、キーフレームシーケンス内のキーフレームの各々に対し、当該キーフレームの視覚素と、隣接するキーフレームの視覚素との組合せに対応するクラスタ化された顔モデルをクラスタ化顔モデル記憶手段から読み出し、各クラスタに属する特徴点のキーフレーム間の変化の速さを当該クラスタの代表ベクトルを用いて算出するための移動量算出手段と、移動量算出手段により算出された変化の速さが最も大きいものから順番に、所定の割合又は所定の数のキーフレームをキーフレームシーケンスから削除するための手段とを含む。

30

【 0 0 3 5 】

予め、顔モデルの組合せの全てについて、動きベクトルを求め、それら動きベクトルに対する所定のクラスタリング、例えばベクトル量子化クラスタリングによって各特徴点をクラスタに分類する。クラスタ化された顔モデルを作成するための手段は、各クラスタについて、代表ベクトルを算出する。移動量算出手段は、キーフレームシーケンス内のキーフレームの各々に対し、当該キーフレームの視覚素と、隣接するキーフレームの視覚素との組合せに対応するクラスタ化された顔モデルをクラスタ化顔モデル記憶手段から読み出し、各クラスタに属する特徴点のキーフレーム間の変化の速さを当該クラスタの代表ベクトルを用いて算出する。算出された変化の速さが最も大きいものから順番に、所定の割合又は所定の数のキーフレームがキーフレームシーケンスから削除される。各特徴点の変化の速さを算出する代わりに、一つのクラスタに属する特徴点を一つの代表点で代表させてそれらの変化の速さを算出するので、演算に要する時間が短縮できる。

40

【 0 0 3 6 】

50

さらに好ましくは、リップシンクアニメーション作成装置は、キーフレーム削除手段により一部のキーフレームが削除されたキーフレームシーケンスを受け、当該キーフレームシーケンス内のキーフレームの視覚素に対応する音素の発話パワーを発話データから算出するための発話パワー算出手段と、キーフレームシーケンス内の各キーフレームに対し、発話パワー算出手段により、当該キーフレームを含む視覚素の継続長について算出された平均発話パワーが小さければ小さいほどブレンド率が小さくなるような、平均発話パワーに対する所定の関数により、ブレンド率を調整するための、発話パワーによるブレンド率調整手段とをさらに含む。ブレンド処理手段は、発話パワーによるブレンド率調整手段によってブレンド率が調整されたキーフレームシーケンスに基づき、キーフレーム間のブレンドにより顔画像のアニメーションを作成する。

10

【0037】

発話パワーが小さいところでは、ブレンド率が小さくなる。一般に、発話パワーが小さいときには、人間はあまりはっきりと口を開いていない。したがって、このようにすることにより、実際の発話時の発話者の口に近い動きをする顔画像のアニメーションを実現できる。その結果、滑らかで自然なアニメーションが得られるようにキーフレーム及びそのブレンド率を自動的に設定できるリップシンクアニメーション作成装置を提供できる。

【0038】

リップシンクアニメーション作成装置は、キーフレーム削除手段により一部のキーフレームが削除されたキーフレームシーケンスを受け、キーフレームの視覚素に対応する顔モデルを構成する頂点と、隣接するキーフレームの視覚素に対応する顔モデルを構成する頂点との間の変化の速さを算出するための変化の速さ算出手段と、キーフレーム削除手段により一部のキーフレームが削除されたキーフレームシーケンスに含まれる各キーフレームのうち、変化の速さ算出手段により算出された変化の速さが所定のしきい値よりも大きなキーフレームについて、そのブレンド率がより小さな値となるような、変化の速さに対する所定の関数を用いてブレンド率を更新するための頂点速度によるブレンド率調整手段とをさらに含んでもよい。ブレンド処理手段は、頂点速度によるブレンド率調整手段によってブレンド率が調整されたキーフレームシーケンスに基づき、キーフレーム間のブレンドにより顔画像のアニメーションを作成する。

20

【0039】

好ましくは、リップシンクアニメーション作成装置は、複数個の顔モデルの内から選ばれ2個の顔モデルの組合せの全てに対し、顔モデルを構成する特徴点の間の動きベクトルを算出するための動きベクトル算出手段と、2個の顔モデルの特徴点を、動きベクトル算出手段により算出された動きベクトルに対する所定のクラスタリング方法によってクラスタ化し、各クラスタの代表ベクトルを算出することにより、クラスタ化された顔モデルを作成するための手段と、クラスタ化された顔モデルを記憶するためのクラスタ化顔モデル記憶手段とをさらに含む。リップシンクアニメーション作成装置はさらに、キーフレーム削除手段により一部のキーフレームが削除されたキーフレームシーケンスを受け、各キーフレームのうち、当該キーフレームの視覚素と、隣接するキーフレームの視覚素との組合せに対応するクラスタ化された顔モデルの組合せをクラスタ化顔モデル記憶手段から読み出し、各クラスタに属する特徴点のキーフレーム間の変化の速さを当該クラスタの代表ベクトルを用いて算出するための変化の速さ算出手段と、キーフレーム削除手段により一部のキーフレームが削除されたキーフレームシーケンスに含まれる各キーフレームのうち、変化の速さ算出手段により算出された変化の速さが所定のしきい値よりも大きなキーフレームについて、そのブレンド率がより小さな値となるような、変化の速さに対する所定の関数を用いてブレンド率を更新するための頂点速度によるブレンド率調整手段とをさらに含む。ブレンド処理手段は、頂点速度によるブレンド率調整手段によってブレンド率が調整されたキーフレームシーケンスに基づき、キーフレーム間のブレンドにより顔画像のアニメーションを作成する。

30

40

【0040】

本発明の第2の局面に係るリップシンクアニメーション作成装置は、予め準備された統

50

計的音響モデルと、予め準備された音素及び視覚素の間のマッピング定義と、予め準備された複数個の顔画像の顔モデルとを用い、入力される発話データからリップシンクアニメーションを作成するためのリップシンクアニメーション作成装置であって、発話データに対するトランスクリプションが利用可能であり、統計的音響モデル、マッピング定義、及びトランスクリプションを使用して、発話データに含まれる音素及び対応する視覚素を求め、デフォルトのブレンド率が付与された継続長付きの視覚素シーケンスを作成するための視覚素シーケンス作成手段を含む。視覚素シーケンスの継続長内の所定位置にはキーフレームが定義され、視覚素シーケンスの各視覚素の継続長内に定義されるキーフレームによりキーフレームシーケンスが定義される。リップシンクアニメーション作成装置はさらに、キーフレームシーケンス内のキーフレームの視覚素に対応する音素の発話パワーを発話データから算出するための発話パワー算出手段と、キーフレームシーケンス内の各キーフレームに対し、発話パワー算出手段により、当該キーフレームを含む視覚素の継続長について算出された平均発話パワーが小さければ小さいほどブレンド率が小さくなるような、平均発話パワーに対する所定の関数により、ブレンド率を調整するための、発話パワーによるブレンド率調整手段と、ブレンド率調整手段によりブレンド率が調整された視覚素シーケンスに基づき、キーフレーム間のブレンドにより顔画像のアニメーションを作成するためのブレンド処理手段とを含む。

10

【0041】

好ましくは、リップシンクアニメーション作成装置は、発話パワーによるブレンド率調整手段によりブレンド率が調整されたキーフレームシーケンスを受け、当該キーフレームシーケンスに含まれる各キーフレームの視覚素に対応する顔モデルを構成する頂点と、隣接するキーフレームの視覚素に対応する顔モデルを構成する頂点との間の変化の速さを算出するための変化の速さ算出手段と、発話パワーによるブレンド率調整手段によりブレンド率が調整されたキーフレームシーケンスに含まれる各キーフレームのうち、変化の速さ算出手段により算出された変化の速さが所定のしきい値よりも大きなキーフレームについて、そのブレンド率が、より小さな値となるような、変化の早さに対する所定の関数を用いてブレンド率を更新するための頂点速度によるブレンド率調整手段とをさらに含む。ブレンド処理手段は、頂点速度によるブレンド率調整手段によってブレンド率が調整されたキーフレームシーケンスに基づき、キーフレーム間のブレンドにより顔画像のアニメーションを作成する。

20

30

【0042】

より好ましくは、リップシンクアニメーション作成装置は、複数個の顔モデルの内から選ばれる2個の顔モデルの組合せの全てに対し、顔モデルを構成する特徴点の間の動きベクトルを算出するための動きベクトル算出手段と、2個の顔モデルの特徴点を、動きベクトル算出手段により算出された動きベクトルに対する所定のクラスタリング方法によってクラスタ化し、各クラスタの代表ベクトルを算出することにより、クラスタ化された顔モデルを作成するための手段と、クラスタ化された顔モデルを記憶するためのクラスタ化顔モデル記憶手段とをさらに含む。リップシンクアニメーション作成装置はさらに、発話パワーによるブレンド率調整手段によりブレンド率が調整されたキーフレームシーケンスを受け、各キーフレームのうち、当該キーフレームの視覚素と、隣接するキーフレームの視覚素との組合せに対応するクラスタ化された顔モデルの組合せをクラスタ化顔モデル記憶手段から読み出し、各クラスタに属する特徴点のキーフレーム間の変化の速さを当該クラスタの代表ベクトルを用いて算出するための変化の速さ算出手段と、キーフレームシーケンスに含まれる各キーフレームのうち、変化の速さ算出手段により算出された変化の速さが所定のしきい値よりも大きなキーフレームについて、そのブレンド率がより小さな値となるような、変化の速さに対する所定の関数を用いてブレンド率を更新するための頂点速度によるブレンド率調整手段とを含む。ブレンド処理手段は、頂点速度によるブレンド率調整手段によってブレンド率が調整されたキーフレームシーケンスに基づき、キーフレーム間のブレンドにより顔画像のアニメーションを作成する。

40

【0043】

50

本発明の第3の局面に係るリップシンクアニメーション作成装置は、予め準備された統計的音響モデルと、予め準備された音素及び視覚素の間のマッピング定義と、予め準備された複数の顔画像の顔モデルとを用い、入力される発話データからリップシンクアニメーションを作成するためのリップシンクアニメーション作成装置であって、発話データに対するトランスクリプションが利用可能である。リップシンクアニメーション作成装置は、統計的音響モデル、マッピング定義、及びトランスクリプションを使用して、発話データに含まれる音素及び対応する視覚素を求め、デフォルトのブレンド率が付与された継続長付きの視覚素シーケンスを作成するための視覚素シーケンス作成手段を含む。視覚素シーケンス中の各視覚素の継続長中にはキーフレームが定義され、これらキーフレームによりキーフレームシーケンスが定義される。リップシンクアニメーション作成装置はさらに、当該キーフレームシーケンスに含まれる各キーフレームの視覚素に対応する顔モデルを構成する頂点と、隣接するキーフレームの視覚素に対応する顔モデルを構成する頂点との間の変化の速さを算出するための変化の速さ算出手段と、キーフレームシーケンスに含まれる各キーフレームのうち、変化の速さ算出手段により算出された変化の速さが所定のしきい値よりも大きなキーフレームについて、そのブレンド率が、より小さな値となるような、変化の速さに対する所定の関数を用いてブレンド率を更新するための頂点速度によるブレンド率調整手段と、頂点速度によるブレンド率調整手段によりブレンド率が調整されたキーフレームシーケンスに基づき、キーフレーム間のブレンドにより顔画像のアニメーションを作成するためのブレンド処理手段とを含む。

10

【0044】

20

本発明の第4の局面に係るリップシンクアニメーション作成装置は、予め準備された統計的音響モデルと、予め準備された音素及び視覚素の間のマッピング定義と、予め準備された複数の顔画像の顔モデルとを用い、入力される発話データからリップシンクアニメーションを作成するためのリップシンクアニメーション作成装置であって、発話データに対するトランスクリプションが利用可能である。リップシンクアニメーション作成装置は、複数の顔モデルの内から選ばれる2個の顔モデルの組合せの全てに対し、顔モデルを構成する特徴点の間の動きベクトルを算出するための動きベクトル算出手段と、2個の顔モデルの特徴点を、動きベクトル算出手段により算出された動きベクトルに対する所定のクラスタリング方法によってクラスタ化し、各クラスタの代表ベクトルを算出することにより、クラスタ化された顔モデルを作成するための手段と、クラスタ化された顔モデルを記憶するためのクラスタ化顔モデル記憶手段と、統計的音響モデル、マッピング定義、及びトランスクリプションを使用して、発話データに含まれる音素及び対応する視覚素を求め、デフォルトのブレンド率が付与された継続長付きのキーフレームシーケンスを作成するためのキーフレームシーケンス作成手段とを含む。視覚素シーケンス中の各視覚素の継続長中にはキーフレームが定義され、これらキーフレームによりキーフレームシーケンスが定義される。リップシンクアニメーション作成装置はさらに、キーフレームシーケンスを受け、各キーフレームのうち、当該キーフレームの視覚素と、隣接するキーフレームの視覚素との組合せに対応するクラスタ化された顔モデルの組合せをクラスタ化顔モデル記憶手段から読出し、各クラスタに属する特徴点のキーフレーム間の変化の速さを当該クラスタの代表ベクトルを用いて算出するための変化の速さ算出手段と、キーフレームシーケンスに含まれる各キーフレームのうち、変化の速さ算出手段により算出された変化の速さが所定のしきい値よりも大きなキーフレームについて、そのブレンド率が、より小さな値となるような、変化の速さに対する所定の関数を用いてブレンド率を更新するための頂点速度によるブレンド率調整手段と、頂点速度によるブレンド率調整手段によりブレンド率が調整されたキーフレームシーケンスに基づき、キーフレーム間のブレンドにより顔画像のアニメーションを作成するためのブレンド処理手段とを含む。

30

40

【0045】

好ましくは、リップシンクアニメーション作成装置は、視覚素シーケンス作成手段の出力するキーフレームシーケンスに含まれるキーフレームのうち、空白音素に対応する視覚素が割当てられたキーフレームの直前のキーフレームの継続長の終端位置を、当該キーフ

50

フレーム内の発話データの発話パワー系列の最大点以後で、かつ当該キーフレームの継続長内の位置に移動させることにより、発話終端位置を補正するための発話終端補正手段をさらに含む。キーフレーム削除手段は、発話終端補正手段により発話終端が補正されたキーフレームシーケンスを入力として受ける。

【0046】

空白音素に対応する視覚素が割当てられたキーフレームの直前のキーフレームについて、その終端位置が補正される。補正後の終端は、そのキーフレーム内の発話パワー系列の最大点以後の位置とする。補正後の終端をこのようにもとの終端位置より前に移動させることにより、発話の最後における視覚素が早めに空白音素に対応する視覚素となり、発話アニメーションが自然なものとなる。

10

【0047】

より好ましくは、発話終端補正手段は、視覚素シーケンス作成手段の出力するキーフレームシーケンスに含まれるキーフレームのうち、空白音素に対応する視覚素が割当てられたキーフレームの直前のキーフレームの、発話パワーの最大値を与える第1の時刻を検出するための手段と、第1の時刻以後で、かつ処理対象のキーフレームの終端時刻以前に、発話パワーの最大値より所定の割合、又は所定の量だけ発話パワーが減少する第2の時刻を検出するための手段と、処理対象のキーフレームの終端位置を、第2の時刻まで移動させるようにキーフレームを補正するための手段とを含む。

【0048】

発話パワーの最大値を与える第1の時刻以後で、最大値より所定の割合又は所定の量だけ発話パワーが減少する第2の時刻に、キーフレームの終端位置を移動させる。各キーフレームにおける発話パワーの絶対値の大きさは無関係に、最大値からの減衰率で終端位置の移動位置を決定するので、発話パワーの大きさの変動にかかわらず、発話の最後に安定したタイミングで口を閉じる画像が得られる。

20

【0049】

さらに好ましくは、キーフレーム作成手段は、キーフレームシーケンスの作成時には、第1のフレームレートのフレームの任意のものをキーフレームとして選択する。リップシンクアニメーション作成装置はさらに、第1のフレームレートよりも小さな第2のフレームレートを指定する入力と、キーフレーム削除手段により出力されるキーフレームシーケンスとを受けるとして接続され、キーフレーム削除手段により出力されるキーフレームシーケンスを、第2のフレームレートのキーフレームシーケンスに変換するためのフレームレート変換手段を含む。フレームレート変換手段は、第2のフレームレートのキーフレームシーケンスの各キーフレームに、キーフレーム削除手段の出力するキーフレームシーケンス内で、当該キーフレームの継続長内に始端を有するキーフレームに割当てられた視覚素のいずれかを割当てる。ブレンド処理手段は、フレームレート変換手段によりフレームレートが変換されたキーフレームシーケンスに基づき、キーフレーム間のブレンドにより顔画像のアニメーションを作成するための手段を含む。

30

【0050】

キーフレーム作成手段は第1のフレームレートのフレームのうちの任意のフレームを用いてキーフレームシーケンスを作成する。第1のフレームレートよりも小さな第2のフレームレートが指定されると、キーフレームレート変換手段が第1のフレームレートのキーフレームシーケンスを第2のフレームレートのキーフレームシーケンスに変換する。このとき、第1のフレームレートのキーフレームシーケンスのうち複数のキーフレームが、第2のフレームレートのキーフレームシーケンス中のキーフレームに対応する可能性が有る。フレームレート変換手段は、そうした場合には、第2のキーフレームレートのキーフレームシーケンス中のキーフレームの継続長内に始端を有する、第1のキーフレームレートのキーフレームシーケンスのキーフレームの視覚素のいずれかを、変換後のキーフレームに割当てる。第2のキーフレームレートのキーフレームシーケンス中のキーフレームに、必ずそのキーフレームの継続長内に始端を有するキーフレームの視覚素が割当てられるため、実際の音声の発声の前に視覚素にしたがって口形状の変化が始まることになる。こ

40

50

の順序は実際の人間の発声時に観測される順序と一致するので、自然な発話をする顔画像アニメーションが得られる。

【0051】

フレームレート変換手段は、第2のフレームレートのキーフレームシーケンスの各キーフレームに割当てた視覚素が、直前のキーフレームに割当てた視覚素と異なるものとなるように視覚素を割当てるようにしてもよい。

【0052】

同一の視覚素が割当てられたキーフレームが連続すると、同じ口形状が長く続くことになり、発話中の顔画像としては不自然になる。直前のキーフレームに割当てられた視覚素と異なる視覚素を各キーフレームに割当てるようにすることにより、そのような不自然さを回避することができ、より自然な顔画像アニメーションを作成できる。

10

【0053】

より好ましくは、ブレンド処理手段は、第2のフレームレートのキーフレームシーケンスからアニメーションを作成するときには、第2のキーフレームレートよりも高い第3のフレームレートでフレームごとの画像を作成する機能を有し、かつ隣接するキーフレームの間の補間により、当該隣接するキーフレームの間のフレームの画像を生成する機能を有する。リップシンクアニメーション作成装置はさらに、フレームレート変換手段の出力する第2のフレームレートのキーフレームシーケンス内のキーフレームの各々について、当該キーフレームと、当該キーフレームの直後のキーフレームとの間のフレーム位置に、当該キーフレームと同じキーフレームをコピーするためのキーフレームコピー手段を含む。

20

【0054】

さらに好ましくは、キーフレームコピー手段は、フレームレート変換手段の出力する第2のフレームレートのキーフレームシーケンス内のキーフレームの各々について、当該キーフレームの直後のキーフレームの直前のフレーム位置に、当該キーフレームと同じキーフレームをコピーするための手段を含む。

【0055】

ブレンド処理手段が、第2のフレームレートの隣接する二つのキーフレーム間に、第3のフレームレートにしたがったフレームを作成するようになっており、しかもそれらのフレームにおける画像を、それら二つのキーフレームの間の補間により作成する場合、二つのキーフレーム間に、滑らかに変化する第3のフレームレートにしたがったフレームが挿入される。そのような補間処理をすると、画像の変化は滑らかになるが、時にアニメーションに求められる「リミット感」を持つ映像（「カクカク」と変化する映像）が得られない。その場合、隣接する二つのキーフレームのうち、後者の直前のフレーム位置に、前者のキーフレームをそのままコピーする。その結果、前者のキーフレーム位置から、コピーされたフレーム位置まではブレンド処理手段による補間を行なっても画像は安定し、変化せず、その直後の次のキーフレームではじめて画像が変化することになる。その結果、第2のフレームレートより大きな第3のフレームレートにしたがってフレームシーケンスを作成する場合で、しかも隣接するキーフレーム間のフレームの画像を補間によって作成する機能を持つブレンド処理手段をそのまま使用する場合にも、リミット感を持つアニメーションを作成できる。

30

40

【0056】

さらに好ましくは、リップシンクアニメーション作成装置は、複数個の顔画像の顔モデルを記憶するための顔モデル記憶手段をさらに含む。

【0057】

複数の顔画像の顔モデルを、顔モデル記憶手段によって記憶することができる。アニメーションを繰返し作成する場合であっても、顔モデルを外部から繰返し受信することなく、同じ顔モデルを何度でも用いて、アニメーションを作成することができる。

【0058】

さらに好ましくは、予め準備された音素は、予め定められた標準音素と、標準音素以外の一般音素とを含み、複数個の顔画像の顔モデルは、標準音素に対応する顔モデルから成

50

る標準視覚素モデルと、一般音素に対応する顔モデルから成る一般視覚素モデルとを含む。リップシンクアニメーション作成装置はさらに、予め準備された音素に対応して予め分類された、対応する音素を発話しているときの発話者の顔画像の特徴点の3次元位置の実測値から成るキャプチャデータと標準視覚素モデルとを用い、一般視覚素モデルを生成するための一般視覚素生成手段を含む。

【0059】

標準視覚素モデルのみを手作業で予め作成しておき、発話時の実際の発話者の顔のキャプチャデータを準備しておけば、装置が一般視覚素作成手段によって標準視覚素モデル以外の一般視覚素モデルを自動的に生成する。したがって、手作業による顔モデル作成のための作業量を少なくし、口の動きと音声とが一致したさらに滑らかで自然な顔画像アニメーションが得られる。

10

【0060】

さらに好ましくは、一般視覚素生成手段は、標準音素に対応するキャプチャデータの線形和で、一般音素に対応するキャプチャデータを近似するための、標準音素の数と同数の係数を、所定の近似誤差を最小とするように算出するための係数算出手段と、一般視覚素モデルを、当該一般視覚素モデルに対応する一般音素について係数算出手段により算出された係数を用いた標準視覚素モデルの線形和により計算し、標準視覚素モデルとともに対応する一般音素と関連付けて顔モデル記憶手段に記憶させるための線形和計算手段とを含む。

【0061】

20

装置が、近似誤差が最小となるような標準視覚素モデルの線形和で一般視覚素モデルを生成する。標準視覚素モデルだけでなく、一般視覚素モデルも用いて各音素に対する顔画像を生成できるので、滑らかで自然な顔画像アニメーションが得られる。

【0062】

本発明の第5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、当該コンピュータを、上記したいずれかのリップシンクアニメーション作成装置として機能させる。

【0063】

本発明の第6の局面に係る顔モデル生成装置は、予め準備された音素及び視覚素の間のマッピング定義を用い、視覚素に対応する顔画像の顔モデルを生成するための顔モデル生成装置であって、予め準備された音素は、予め定められた標準音素と、標準音素以外の一般音素とを含み、複数個の顔画像の顔モデルは、標準音素に対応する顔モデルから成る標準視覚素モデルと、一般音素に対応する顔モデルから成る一般視覚素モデルとを含み、顔モデル生成装置は、視覚素に対応する複数個の顔画像の顔モデルを記憶するための顔モデル記憶手段と、予め準備された音素に対応して予め分類された、対応する音素を発話しているときの発話者の顔画像の特徴点の3次元位置の実測値から成るキャプチャデータ及び標準視覚素モデルを用い、一般視覚素モデルを生成するための一般視覚素生成手段とを含む。

30

【0064】

好ましくは、一般視覚素生成手段は、標準音素に対応するキャプチャデータの線形和で、一般音素に対応するキャプチャデータを近似するための、標準音素の数と同数の係数を、所定の近似誤差を最小とするように算出するための係数算出手段と、一般視覚素モデルを、当該一般視覚素モデルに対応する一般音素について係数算出手段により算出された係数を用いた標準視覚素モデルの線形和により計算し、標準視覚素モデルとともに対応する一般音素と関連付けて顔モデル記憶手段に記憶させるための線形和計算手段とを含む。

40

【0065】

さらに好ましくは、平均発話パワーに対する所定の関数は、平均発話パワーの線形関数でもよく、非線形関数でもよい。

【0066】

変化の速さに対する所定の関数も、変化の速さの線形関数でも非線形関数でもよい。

50

【 0 0 6 7 】

リップシンクアニメーション作成装置はさらに、線形関数又は非線形関数の振舞を制御するパラメータの入力を受けるための手段をさらに含んでもよい。

【 発明を実施するための最良の形態 】

【 0 0 6 8 】

以下、本発明について、実施の形態に基づいて説明する。以下の説明では、基本となる顔画像を6種類使用しているが、顔画像の数はこれには限定されない。6種類よりも少なくてもよいし、6種類よりも多くてもよい。

【 0 0 6 9 】

[第1の実施の形態]

< 構成 >

【 0 0 7 0 】

図5に、本発明に係るアニメーション作成装置の一例として、本発明の第1の実施の形態に係るリップシンクアニメーション作成装置200の概略ブロック図を示す。図5を参照して、リップシンクアニメーション作成装置200は、発話記憶部152に記憶された発話の音声データと、トランスクリプション記憶部154に記憶された、発話記憶部152に記憶された発話の書き起こしテキスト(トランスクリプション)とを入力として受け、キャラクタモデル記憶部156に記憶された、/a/~/o/及び/N/からなる6つの基本となる顔画像に相当する3Dキャラクターモデルを用いて顔画像のアニメーション260を作成するためのものである。

【 0 0 7 1 】

キャラクタモデル記憶部156に記憶される顔画像の例を図7に示す。図7(A)~(F)は、それぞれ音素/a/、/i/、/u/、/n/、/e/、/o/に対応する顔画像である。本明細書では、これら画像をそれぞれ顔画像/A/、/I/、/U/、/N/、/E/、及び/O/と表記することにする。

【 0 0 7 2 】

なお、本実施の形態では、顔画像/A/、/I/、/U/、/E/、/O/は、いずれも顔画像/N/を基準とし、各特徴点が、顔画像の定義されている3次元空間において、顔画像/N/の対応する特徴点からどの程度移動しているかを示す3次元ベクトル情報によって定義されている。従って、例えば顔画像/A/と顔画像/N/との間で、その中間の顔画像を定義することもできる。本実施の形態では、特定の顔画像と顔画像/N/との間の中間の顔画像を定義するために、上記した「ブレンド率」という概念を使用する。

【 0 0 7 3 】

二つの顔画像の間のブレンドについては前述したとおりである。

【 0 0 7 4 】

リップシンクアニメーション作成装置200は、発話者の音声から予め作成された音響モデルを記憶するための音響モデル記憶部170と、予め準備された、音素と視覚素との間のマッピングテーブルを記憶するための音素-視覚素マッピングテーブル記憶部176と、音響モデル記憶部170に記憶された音響モデル及び音素-視覚素マッピングテーブル記憶部176に記憶された音素-視覚素マッピングテーブルを用い、発話データに対し、トランスクリプション記憶部154に記憶されたトランスクリプションに基づいた音素セグメンテーションを行なって音素シーケンスを作成し、さらに、得られた音素シーケンス内の各音素を音素-視覚素マッピングテーブル記憶部176に記憶された音素-視覚素マッピングテーブルを用いて対応の視覚素に変換することにより、継続長付き視覚素シーケンスを作成するための視覚素シーケンス作成部230と、視覚素シーケンス作成部230により出力される視覚素シーケンスを記憶するための視覚素シーケンス記憶部180とを含む。視覚素の継続期間は、対応する音素継続期間の先頭から開始する。したがって視覚素シーケンス記憶部180に記憶された視覚素シーケンスのうち、各視覚素の先頭フレームがキーフレームとなる。視覚素シーケンス内のキーフレームにより、キーフレームシーケンスが構成される。なお、視覚素シーケンス作成部230は、各視覚素に対し、置換

10

20

30

40

50

前の音素と、デフォルトのブレンド率（例えば100%）を付して視覚素シーケンスを作成するものとする。

【0075】

リップシンクアニメーション作成装置200はさらに、キャラクタモデル記憶部156に記憶された3Dキャラクタモデルの各顔画像を構成する頂点に対し、任意の二つの顔画像の間での動きベクトルを用いたVQ（ベクトル量子化）クラスタリングを行ない、任意の二つの顔画像の間での各頂点の動きを、各頂点が属するクラスタの代表ベクトルによって表した動きベクトルデータと、そのときのクラスタリング後の顔画像のモデルとを出力するためのクラスタリング処理部232と、クラスタリング処理部232の出力する、任意の顔画像モデルの組合せに対するクラスタリング後の顔画像モデルと動きベクトルデータとを記憶するためのクラスタ化顔モデル記憶部234と、キャラクタモデル記憶部156に記憶された顔画像モデルと、クラスタ化顔モデル記憶部234に記憶されたクラスタリング後の顔モデル及び動きベクトルデータとのいずれか一方を使用して、キーフレームの中で頂点の動きが速いものを検出し、そのようなキーフレームを所定の割合又は所定の数だけ削除するためのキーフレーム削除部236とを含む。なお、本実施の形態では、あるキーフレームを削除した場合、そのキーフレームの継続長だった部分は、削除されたキーフレームの直前のキーフレームの継続長に統合される。

10

【0076】

リップシンクアニメーション作成装置200はさらに、キーフレーム削除部236によるキーフレームの削除の際の、全体のキーフレーム数のうち、削除されるキーフレームの数が示す割合を指定するための削除率入力部201と、キーフレーム削除部236によるキーフレーム削除の際の速度計算に、キャラクタモデル記憶部156に記憶されたモデルをそのまま使用するか、クラスタ化顔モデル記憶部234に記憶されたクラスタリング後の動きベクトルによるモデルを使用するかを指定するためのクラスタ処理指定部202とを含む。キーフレーム削除部236の詳細については後述する。

20

【0077】

リップシンクアニメーション作成装置200はさらに、発話記憶部152に記憶された発話データから、各フレームにおける発話パワーを算出するための発話パワー算出部238と、発話パワー算出部238により算出された発話パワーを記憶するための発話パワー記憶部240と、キーフレーム削除部236により出力された視覚素シーケンスに対し、発話パワー記憶部240に記憶された各フレームにおける発話パワーに基づいて、後述するように、キーフレームのブレンド率を調整するための発話パワーによるブレンド率調整部244とを含む。

30

【0078】

リップシンクアニメーション作成装置200はさらに、発話パワーによるブレンド率調整部244において、あるキーフレームのブレンド率を減衰させる関数の振舞を決めるパラメータ（以下「減衰率」と呼ぶ。0 < < 1である。）をユーザが入力するための減衰率入力部206と、発話パワーによるブレンド率調整部244によるブレンド率の調整を行なうか否かをユーザが指示する際に使用する発話パワー使用指示入力部204と、発話パワー使用指示入力部204により発話パワーが指示されたときにはキーフレーム削除部236の出力を発話パワーによるブレンド率調整部244に与え、それ以外ときにはキーフレーム削除部236の出力を発話パワーによるブレンド率調整部244をバイパスして後続する処理部に与えるために、一对の選択部242及び246とを含む。

40

【0079】

リップシンクアニメーション作成装置200はさらに、クラスタ処理指定部202により指定された値にしたがい、キャラクタモデル記憶部156に記憶された顔画像モデルのデータ及びクラスタ化顔モデル記憶部234に記憶された動きベクトルのいずれかを用い、各キーフレームにおける頂点の動きの速さを算出して、動きの速さが所定の基準より大きなキーフレームについて、ブレンド率を小さくなるように調整するための頂点速度によるブレンド率調整部250と、ブレンド率調整部250によるブレンド率の調整の際の、

50

ブレンド率に対する関数の振舞を決めるパラメータである減衰率（ただし $0 < \text{減衰率} < 1$ ）を入力するためにユーザが使用する減衰率入力部 210 と、ブレンド率調整部 250 によるブレンド率調整を行なうか否かをユーザが指定するための頂点速度使用指示入力部 208 と、使用指示入力部 208 により入力された指示にしたがい、選択部 246 の出力をブレンド率調整部 250 に与えるか、発話パワーによるブレンド率調整部 244 をバイパスして後続する処理部に与えるかを選択する一対の選択部 248 及び 252 とを含む。

【0080】

リップシンクアニメーション作成装置 200 はさらに、選択部 252 の出力する、ブレンド率の調整が完了した継続長付き視覚素シーケンスを記憶するための視覚素シーケンス記憶部 254 と、視覚素シーケンス記憶部 254 に記憶された継続長付き視覚素シーケンスに基づき、キャラクタモデル記憶部 156 に記憶された各顔画像モデルを用いたブレンド処理を行なうことによつて、顔画像のアニメーション 260 を作成するためのブレンド処理部 256 を含む。

10

【0081】

図 6 に、図 5 の視覚素シーケンス作成部 230 の詳細な構成を示す。図 6 を参照して、視覚素シーケンス作成部 230 は、音響モデル記憶部 170 に記憶された音響モデルを用い、発話記憶部 152 に記憶された発話データに対して、トランスクリプション記憶部 154 に記憶されたトランスクリプションに基づいた音素セグメンテーションを行ない、音素シーケンスをその継続長を示す情報とともに出力するための音素セグメンテーション部 172 と、音素セグメンテーション部 172 から出力された継続長付き音素シーケンスを記憶するための音素シーケンス記憶部 174 とを含む。

20

【0082】

視覚素シーケンス作成部 230 はさらに、音素と視覚素との間のマッピングテーブルを記憶するための音素 - 視覚素マッピングテーブル記憶部 176 と、音素 - 視覚素マッピングテーブル記憶部 176 に記憶された音素 - 視覚素マッピングテーブルを参照しながら、音素シーケンス記憶部 174 に記憶された音素シーケンス内の各音素を対応する視覚素に変換することにより、継続長付き視覚素シーケンスを出力するための音素 - 視覚素変換処理部 178 とを含む。なお、前述したとおり、音素 - 視覚素変換処理部 178 の出力する継続長付き視覚素シーケンスの各視覚素には、対応の音素と、デフォルトのブレンド率とが付されている。

30

【0083】

音素セグメンテーション部 172 は、発話記憶部 152 に含まれる発話データに対する音素セグメンテーションをし、音素列と、それぞれの継続時間長が分かる時間データとを出力できるものであればどのようなものでもよい。発話内容がトランスクリプション記憶部 154 に記憶されたトランスクリプションにより予め分かっているため、音素セグメンテーション部 172 は音声データを精度高く音素列に変換できる。

【0084】

テーブル 1 に、マッピングテーブル記憶部 176 に記憶されたマッピングテーブルの例の一部を示す。

【0085】

40

【表 1】
テーブル1

視覚素	音素
/A/	/a/
/I/	/i/
/U/	/u/
/E/	/e/
/O/	/o/
/N/	/N/, /p/, /b/, /m/
NOP	/h/, /j/, /q/, /r/
直前の80%	その他

10

テーブル1を参照して、本実施の形態では、マッピングテーブルは、音素/a/を視覚素/A/に、音素/i/を視覚素/I/に、音素/u/を視覚素/U/に、音素/e/を視覚素/E/に、音素/o/を視覚素/O/にそれぞれ対応付けている。マッピングテーブルでは、図3に示す顔画像/A/, /I/, /U/, /E/, /O/のように、予めある音素に対して準備された視覚素には、その音素を必ず対応付けるようにする。さもないと得られる顔の動画像が発話内容とちぐはぐになってしまう。また音素/N/, /p/, /b/, /m/等、唇を閉じるような音素は無表情の顔画像/N/に対応付ける。音素/h/, /j/, /q/, /r/については無視し、視覚素に変換しない。第1のテーブルに記載された音素以外の音素については、直前の音素のブレンド率の80%のブレンド率を割当てる。

20

【0086】

図8～図10を用いて、クラスタリング処理部232による処理について説明する。クラスタリング処理部232による処理は、簡略に言えば、以下のようなものとなる。

【0087】

キャラクタモデル記憶部156に含まれる顔モデルのうちの任意の二つの組合せの全てについて、以下の処理を行なう。

【0088】

まず、一方の顔画像の全ての頂点の座標ベクトルを、他方の対応する頂点の座標ベクトルから減算する。この減算により、一方の顔画像から他方の顔画像に変化する際の各頂点の動きベクトルが求められる。図8は、一方の顔画像として視覚素/N/の各頂点からなる顔画像280を、他方の顔画像として視覚素/O/の各頂点からなる顔画像282を例とし、視覚素/N/から視覚素/O/への動きベクトルの集合からなる画像284を示してある。なお、図8において、横軸はX軸、縦軸はZ軸であり、Y軸については図示していない。

30

【0089】

こうして求めた動きベクトルの集合に対し、クラスタリング処理部232は、概略、以下のアルゴリズムによってクラスタリングを行なう。

【0090】

(1) クラスタ数Nを決定する。

40

【0091】

(2) 動きベクトルの集合の中からN個のベクトルを任意に選択し、初期コードブックとする。

【0092】

(3) 動きベクトルの集合の中の全ベクトルを、初期コードブックとの間のユークリッド距離に基づいてN個のクラスタに分類する。この場合、各動きベクトルは、ユークリッド距離が最も小さくなるコードブックにより代表されるクラスタに分類される。

【0093】

(4) 各クラスタに属するベクトルの平均を算出することにより、新たなN個のコード

50

ブックを作成する。

【0094】

(5) コードブックが変化しなくなるか、その間の差がしきい値より小さくなるまでステップ(3)及び(4)を繰り返す。

【0095】

なお、本実施の形態においては、各クラスタの代表頂点は、そのクラスタについて求められたセントロイド(重心)に最も近い頂点とする。

【0096】

以上のようにして得られたクラスタリングの結果、各画像の組合せについて各頂点が複数個のクラスタのいずれかに属することになる。図9にそうしたクラスタリングの結果を顔画像にマッピングした例を示す。図9を参照して、画像300と他の画像(図示せず)との間の動きベクトルのクラスタリングにより、画像300を構成する顔モデルを構成する各頂点は、画像302に示すように、クラスタ310, 312, 314, 316, 318, 320, 322及び324に分類される。この例の場合、クラスタの個数は8、頂点数は1483個である。

10

【0097】

図9から分かるように、口付近の頂点はその位置により明確にクラスタ化されるが、それ以外の領域の頂点の動きにはあまり差がない。

【0098】

図10には、同様の処理でクラスタ数 = 128、頂点数1483個の場合のクラスタリングにより得られたクラスタを顔画像にマッピングした結果340を示す。このようにクラスタ数を多くすると、口付近以外の各頂点もクラスタ化されてくることが分かる。

20

【0099】

このようにクラスタ化するのは以下の理由による。例えば図5に示すキーフレーム削除部236及びブレンド率調整部250における処理において、全ての頂点について移動量又は速度を算出すると、頂点の数だけ計算する必要があり処理に長時間を要する。これに対し、頂点をクラスタ化した場合、各頂点の移動量又は速度を、その頂点が属するクラスタの代表頂点の移動量又は速度で近似することができる。したがって、実質的な計算量はクラスタの数まで削減され、計算時間を大幅に短縮することができる。

【0100】

30

例えば口付近の画像だけを短時間で処理する必要があるればクラスタ数を小さくし、計算時間が多少長くても、口だけでなく頭部全体の画像もある程度の精密さで求める必要があるればクラスタ数を大きくすればよい。さらに、計算に要する時間に制限がないのであれば、こうしたクラスタリングを行わず、全ての頂点について個別にその移動量又は速度を計算すればよい。

【0101】

図11は、キーフレーム削除部236の機能をコンピュータプログラムで実現する際の、プログラムの制御構造を示すフローチャートである。図11を参照して、ステップ360において、削除率を所定の記憶領域から読出す。この削除率は、図5に示す削除率入力部201を用いてユーザにより予め入力され、所定の記憶領域に記憶されていたものである。

40

【0102】

ステップ362において、この削除率に基づき、削除すべきキーフレーム数Kを算出する処理が行なわれる。視覚素シーケンス記憶部180に記憶された視覚素シーケンス中のキーフレーム数をa、削除率を%とすると、本実施の形態では、削除すべきキーフレーム数Kを $a \times \text{削除率} \times 100$ により求める。ここで、計算結果を四捨五入するか、切り上げるか、切り捨てるかは設計事項である。

【0103】

ステップ364では、以下の繰り返し処理のための繰り返し変数iに0を代入する。ステップ366で変数iに1を加算し、ステップ368で変数iの値が削除すべきキーフレーム

50

数Kより大きくなったか否かを判定する。判定結果がYESであればステップ382に進み、それ以外の場合にはステップ370に進む。

【0104】

ステップ370では、以下の計算において、クラスタ化顔モデル記憶部234に記憶されたクラスタリング後の顔画像のモデルを使用するか、又はキャラクタモデル記憶部156に記憶された元の顔画像のモデルを使用するかを判定する。この判定は、クラスタ処理指定部202を用いてユーザにより予め入力されており、所定の記憶領域に記憶されていた情報に基づいて行なわれる。クラスタ化後のモデルを使用する場合にはステップ376に進み、使用しない場合にはステップ372に進む。

【0105】

ステップ372では、視覚素シーケンス中で隣接するキーフレームの組合せの全てにおいて、全ての頂点を用いてキーフレーム間の距離Dを以下の式により算出する。

【0106】

【数1】

$$D(k) = \sum_{v \in \text{頂点集合}} \sqrt{(x_{v,k} - x_{v,k+1})^2 + (y_{v,k} - y_{v,k+1})^2 + (z_{v,k} - z_{v,k+1})^2}$$

ここで、D(k)はk番目のキーフレームと、k+1番目のキーフレームとの間の全頂点のユークリッド距離の合計を表す。この距離D(k)を、以後k番目のキーフレームとk+1番目のキーフレームとの間のキーフレーム間の距離と呼ぶ。

【0107】

続いてステップ374において、ステップ372で算出されたキーフレーム間の距離に基づいて、以下の式によって削除すべきキーフレームを決定する。

【0108】

【数2】

$$k_{target} = \arg \max_k \left(\frac{D(k-1) + D(k)}{Dur_k} \right)$$

ただしDur_kはk番目のキーフレームの継続長を示す。

【0109】

要するに、ステップ372及びステップ374の処理により、一つ前のキーフレームからの全ての頂点の移動速度と、一つ後のキーフレームまでの全ての頂点の移動速度との合計が最も大きなキーフレームが削除対象のキーフレームとして決定される。ステップ380でこのキーフレームを削除し、ステップ366に戻る。

【0110】

一方、ステップ370においてクラスタリング後のモデルを使用すると判定された場合には、ステップ376において、以下の式により、視覚素シーケンス中で隣接するキーフレームの組合せの全てにおいて、各クラスタの代表頂点を用いてキーフレーム間の距離D'を以下の式により算出する。

【0111】

【数3】

$$D'(k) = \sum_{r \in \text{代表点の集合}} m_r \sqrt{(x_{r,k} - x_{r,k+1})^2 + (y_{r,k} - y_{r,k+1})^2 + (z_{r,k} - z_{r,k+1})^2}$$

ただしm_rは代表頂点rにより代表されるクラスタに属する頂点の数を示す。

【0112】

ステップ378では、ステップ376で算出されたキーフレーム間の距離D'に基づいて、以下の式によって削除すべきキーフレームを決定する。

【0113】

10

20

30

40

【数4】

$$k_{target} = \arg \max_k \left(\frac{D'(k-1) + D'(k)}{Dur_k} \right)$$

要するに、ステップ376及び378の処理により、キーフレーム間の全ての頂点の移動速度を、代表頂点の移動速度で近似し、それらを用いて一つ前及び一つ後のキーフレームの間の頂点の移動速度の合計が最も大きなキーフレームが削除対象のキーフレームとして決定される。ステップ380でこのキーフレームを削除し、ステップ366に戻る。

【0114】

ステップ372での処理は、顔画像のモデルを構成する全ての頂点について行なう必要がある。一方、ステップ376での処理は、各クラスタの代表頂点のみに対して行なえばよい。したがって、ステップ376での処理に要する時間はステップ372での処理に要する時間と比較してはるかに少なくなる。ただし、ステップ376で得られる距離D'は、ステップ372の処理で得られる距離Dと比較すると概算値となり、誤差を含み、場合によっては削除されるキーフレームが両者で異なってくる。

10

【0115】

なお、ステップ368で変数iの値が削除フレーム数Kより大きいと判定された場合、ステップ382において、K個のキーフレームが削除された後の視覚素シーケンスが出力され、処理を終了する。

【0116】

図12に、キーフレーム削除部236によって行なわれるキーフレームの削除の概念を示す。図12(A)を参照して、視覚素シーケンス中に、4つのキーフレーム400、402、404及び406があるものとする。これらの全ての組合せについて、前記した距離D又はD'を算出する。そして、これらの中で前後のキーフレームとの間の頂点の移動速度の合計値として最小値を与えるキーフレームを削除する。図12(A)で示す例では、キーフレーム402がそうしたキーフレームであるものとする。すると、図12(B)に示すようにキーフレーム402を視覚素シーケンスから削除し、新たに3つの視覚素を含む視覚素シーケンスに対し、前記した処理が行なわれることになる。

20

【0117】

図5に示す発話パワーによるブレンド率調整部244によって行なわれる処理について、図13を参照して説明する。発話パワーによるブレンド率調整部244は、各キーフレームに対応する音素の継続長にわたる発話パワーを、発話記憶部152に記憶された発話データ及び視覚素シーケンス記憶部180に記憶された視覚素シーケンスに含まれる音素シーケンスの継続長から算出する。ある音素の発話パワーは、各音素の継続長の中央における音声信号の振幅の二乗和により求める。

30

【0118】

例えば、図13に示すように、実際の音声信号の波形がグラフ420で示されるものであり、グラフ420により示される音声信号中に、音素/a/、/i/、/o/、/e/、及び/u/からなる音素シーケンスがあったものとする。音素/a/については、その継続長の先頭から次のキーフレーム/i/に代わるまでの期間にわたる平均の発話パワーを算出する。他の音素/i/、/o/、/e/、及び/u/についても同様であり、それぞれの継続長の先頭から、次のキーフレームに代わるまでの期間にわたる平均の発話パワーを、線分430、432、434、436及び438により示すようにそれらの継続長の全体にわたり算出する。発話パワーによるブレンド率調整部244は、こうして算出された発話パワーの平均値に基づき、各音素に対応する視覚素のブレンド率を調整する。

40

【0119】

図14に、発話パワーによるブレンド率調整部244が行なう処理をコンピュータプログラムにより実現する際の、プログラムの制御構造をフローチャート形式で示す。

【0120】

図14を参照して、ステップ450において、減衰率を所定の記憶領域から読出す。

50

この減衰率は、図5に示す減衰率入力部206を用いてユーザにより入力され、所定の記憶領域に格納されていたものである。

【0121】

ステップ452では、音素シーケンス中の全ての音素について、その継続長にわたる発話パワーの平均を算出する。以下、N番目のキーフレームの音素の、その継続長全体にわたる発話パワーの平均をSP(N)と書く。

【0122】

ステップ454では、ステップ452で算出された全ての発話パワーの平均値の中で、最大のものMAX(SP)と、最小のものMIN(SP)とを決定する。

【0123】

ステップ456では、平均発話パワーの最大値を与えるキーフレームを除く全てのキーフレームについて、次の式(1)にしたがい、ブレンド率を更新する。なお、以下、N番目のキーフレームのブレンド率をBR(N)と書く。

【0124】

【数5】

$$BR(N) = BR(N) \times \left(1 - \alpha \times \frac{(MAX(SP) - SP(N))}{(MAX(SP) - MIN(SP))} \right) \quad (1)$$

平均発話パワーの最大値を与えるキーフレームを除く全てのキーフレームに対してこの式(1)によるブレンド率の調整を行なうと、発話パワーによるブレンド率調整部244による処理は終了する。なお、減衰率は、最小値を与えるキーフレームのブレンド率をどの程度減衰させるかを表していることが上の式から分かる。式(1)は、平均発話パワーSP(N)に対する線形関数であり、算出された平均発話パワーが小さければ小さいほどブレンド率が小さくなるような関数である。

【0125】

この処理による結果の一例を次のテーブルにより示す。調整前のブレンド率及び平均発話パワーを全てのキーフレームの音素に対して示したのがテーブル2であり、発話パワーによるブレンド率調整部244による調整後のブレンド率を示したのがテーブル3である。

【0126】

【表2】

テーブル2

音素	ブレンド率BR	平均発話パワーSP
/a/	90	387
/i/	90	600
/o/	90	387
/e/	90	100
/u/	90	171

【0127】

【表3】

テーブル3

音素	ブレンド率BR	平均発話パワーSP
/a/	60	387
/i/	90	600
/o/	60	387
/e/	20	100
/u/	30	171

ブレンド率に対しこのような調整を行なうことにより、平均発話パワーが最大となるキ

10

20

30

40

50

ーフレームのブレンド率は変化しないが、平均の発話パワーが小さくなればなる程、ブレンド率が小さくなる。その結果、話し声が小さい場合には口の動きも小さくなるアニメーションが作成でき、アニメーションの動きがより自然に近くなる。

【 0 1 2 8 】

図 1 5 に、図 5 のブレンド率調整部 2 5 0 が行なう処理をコンピュータプログラムで実現する際の、プログラムの制御構造をフローチャート形式で示す。

【 0 1 2 9 】

図 1 5 を参照して、ステップ 4 7 0 において、減衰率 を所定の記憶領域から読出す。減衰率 は、図 5 に示す減衰率入力部 2 1 0 を用いてユーザにより入力され、所定の記憶領域に記憶されていたものである。減衰率 の意味は以下から明らかとなるが、本実施の形態では、キーフレームの間で頂点の動きに基づいてブレンド率を調整しないキーフレーム（以下「不変フレーム」と呼ぶ。）の割合を示す値が用いられる。

【 0 1 3 0 】

ステップ 4 7 2 では、ステップ 4 7 0 で読出された減衰率 を、全体のキーフレーム数に乘算することにより、不変フレームの数 L を算出する。不変フレームの数 L について、切り上げにより求めるか、四捨五入により求めるか、切り捨てにより求めるかは設計事項である。

【 0 1 3 1 】

ステップ 4 7 4 では、クラスタリング後のモデルを使用するか否かを判定する。この判定は、クラスタ処理指定部 2 0 2 を用いてユーザにより入力され、所定の記憶領域に格納されていた値を用いて行なわれる。クラスタリング後のモデルを使用する場合はステップ 4 8 0 に進み、使用しない場合にはステップ 4 7 6 に進む。

【 0 1 3 2 】

ステップ 4 7 6 では、全てのキーフレームに対し、その前後のキーフレームとの間での、全頂点の平均速度を算出する。この算出方法は図 1 1 のステップ 3 7 2 及び 3 7 4 で行なうのと同様である。

【 0 1 3 3 】

ステップ 4 7 8 では、全キーフレームを、ステップ 4 7 6 で算出された平均速度の降順にソートする。

【 0 1 3 4 】

ステップ 4 8 4 では、このようにソートされたキーフレームのデータのうち、下位から L 個のキーフレームの中の、平均速度の最大値 $\langle VS \rangle$ を決定する。

【 0 1 3 5 】

ステップ 4 8 6 では、ステップ 4 8 4 で決定された値 $\langle VS \rangle$ より大きな平均速度を持つキーフレームにおいて、ブレンド率 $BR(N)$ を以下の式 (2) にしたがって調整する。

【 0 1 3 6 】

【 数 6 】

$$BR(N) = BR(N) \times \frac{\langle VS \rangle}{VS(N)} \quad (2)$$

ただし $VS(N)$ は N 番目のキーフレームの平均速度である。式 (2) は平均速度 $VS(N)$ の非線形関数である。式 (2) によれば、平均速度 $VS(N)$ は平均値 $\langle VS \rangle$ より大きいので、調整後のブレンド率は調整前のブレンド率より小さくなる。ステップ 4 8 6 の後、処理を終了する。

【 0 1 3 7 】

一方、クラスタリング後のモデルを使用する場合、ステップ 4 8 0 において、全てのキーフレームに対し、その前後のキーフレームとの間での頂点の平均速度を、各クラスタの代表頂点を用いて算出する。ここでの処理は、図 1 1 のステップ 3 7 6 及び 3 7 8 で行なったのと同様の考え方により行なう。

10

20

30

40

50

【 0 1 3 8 】

ステップ 4 8 2 では、全キーフレームをステップ 4 8 0 で算出された平均速度の降順でソートする。以下、ステップ 4 8 4 の処理に進む。

【 0 1 3 9 】

ここでの処理は、要するに、各頂点の動く速度が速いキーフレームについては、他のキーフレームの速さを基準として、口の動きが小さくなるようにブレンド率を調整する、というものである。頂点の動きがキーフレーム間であまりに速い場合、キーフレームでの口の形を元のままに維持すると、口の動きが不自然に見える。そこで、そうした場合にはブレンド率を小さく調整することにより、口の動きが小さくなるようにする。

【 0 1 4 0 】

次の表に、ブレンド率調整部 2 5 0 によるブレンド率の調整前後におけるブレンド率の変化の例を示す。テーブル 4 は平均速度の調整後でキーフレームのソート前、テーブル 5 はソート後でかつブレンド率の調整前を示す。

【 0 1 4 1 】

【表 4】

テーブル4

音素	ブレンド率BR	平均速度VS
/a/	60	100
/i/	90	200
/o/	60	150
/e/	20	24
/u/	30	50

【 0 1 4 2 】

【表 5】

テーブル5

音素	ブレンド率BR	平均速度VS
/i/	90	200
/o/	60	150
/a/	60	100
/u/	30	50
/e/	20	24

ここで、減衰率 = 6 0 % とすると、不変フレーム数 L は $5 \times 0.6 = 3$ となる。したがって表 x における下 3 行についてはブレンド率の調整は行なわず、上 2 行のみのブレンド率の調整を行なう。ステップ 4 8 4 で決定する平均速度の最大値 $\langle VS \rangle$ は、音素 / a / の平均速度「100」となる。

【 0 1 4 3 】

$\langle VS \rangle = 100$ を用いてステップ 4 8 6 の処理を行なうと、上位の二つの音素 / i / 及び / o / のブレンド率がそれぞれ以下のように訂正される。すなわち、音素 / i / については $BR(N) = 90 \times 100 / 200 = 45$ となり、音素 / o / については $BR(N) = 60 \times 100 / 150 = 40$ となる。その結果、ブレンド率調整部 2 5 0 によるブレンド率調整後の各キーフレームのブレンド率は以下ようになる。

【 0 1 4 4 】

10

20

30

40

【表 6】
テーブル6

音素	ブレンド率BR	平均速度VS
/i/	45	200
/o/	40	150
/a/	60	100
/u/	30	50
/e/	20	24

すなわち、不変フレームの中の最大の平均速度より大きな平均速度を持つキーフレームのブレンド率が当初より小さな値に調整される。しかも、そのキーフレームの平均速度が大きいほど、ブレンド率は小さくなるため、キーフレームの頂点の移動速度が速いほど、そのキーフレームにおける口の位置の変化が小さくなり、一連のアニメーションはより滑らかで自然なものとなる。

【 0 1 4 5 】

< 動作 >

以上構成を説明したリップシンクアニメーション作成装置 2 0 0 は以下のように動作する。図 5 を参照して、最初に発話記憶部 1 5 2 に、所定の発話者の発話を記録した発話データが準備され、その発話の書き起こしデータであるトランスクリプションがトランスクリプション記憶部 1 5 4 に準備される。また、前述した 6 つの視覚素に対応した 6 つの顔画像のキャラクタモデルがワイアフレーム画像としてキャラクタモデル記憶部 1 5 6 に準備される。

【 0 1 4 6 】

顔画像のアニメーション 2 6 0 の作成のためには、種々の準備作業が必要である。以下それらの準備作業を順番に述べる。

【 0 1 4 7 】

- 視覚素シーケンスの作成 -

まず、視覚素シーケンス作成部 2 3 0 が音響モデル記憶部 1 7 0 に記憶された音響モデル、及び音素 - 視覚素マッピングテーブル記憶部 1 7 6 に記憶された音素 - 視覚素マッピングテーブル記憶部 1 7 6 を用い、以下のようにして視覚素シーケンスを作成し視覚素シーケンス記憶部 1 8 0 に記憶させる。

【 0 1 4 8 】

図 6 を参照して、視覚素シーケンス作成部 2 3 0 の音素セグメンテーション部 1 7 2 が、発話記憶部 1 5 2 中の発話データを読み、トランスクリプション記憶部 1 5 4 と音響モデル記憶部 1 7 0 とを用いて発話データに対する音素セグメンテーションを行なう。この処理の結果、音素セグメンテーション部 1 7 2 からは音素シーケンスが、各音素の継続長を表すデータとともに出力される。この継続長付き音素シーケンスは音素シーケンス記憶部 1 7 4 に記憶される。

【 0 1 4 9 】

音素 - 視覚素変換処理部 1 7 8 が、音素シーケンス記憶部 1 7 4 から音素シーケンスを読み出し、音素 - 視覚素マッピングテーブル記憶部 1 7 6 に記憶された音素 - 視覚素マッピングテーブルを用いて、音素シーケンス中の音素に対応する視覚素に置き換えることにより、継続長付き視覚素シーケンスを生成する。ただしここでは、置換前の音素も各視覚素に付してあるものとする。この継続長付き視覚素シーケンスは視覚素シーケンス記憶部 1 8 0 に記憶される。

【 0 1 5 0 】

- 顔画像の頂点のクラスタリング -

クラスタリング処理部 2 3 2 は、キャラクタモデル記憶部 1 5 6 に格納された 6 つの顔画像に対し、二つの顔画像の全ての組合せに対し、以下の処理を実行する。

【 0 1 5 1 】

10

20

30

40

50

まず、一方の顔画像から他方の顔画像に変化する際の頂点の動きベクトルを算出する。この動きベクトルの集合に対し、前述したとおりのVQクラスタリングを行なうことにより、一方の顔画像を所定個数のクラスタに分類する。逆方向の動きについては、動きベクトルの向きが逆になるだけであるから、クラスタリングは正逆で同じになる。

【0152】

このようにしてクラスタリングを行なった結果、二つの顔画像の全ての組合せに対し、クラスタリング後の顔モデルと、各クラスタの代表頂点とが算出される。この顔モデルが、各クラスタの代表頂点とともにクラスタ化顔モデル記憶部234に記憶される。

【0153】

- 発話パワーの算出 -

発話パワー算出部238は、視覚素シーケンス記憶部180に記憶された各視覚素に付された音素の情報に基づき、発話記憶部152中の各音素の平均発話パワーを算出し、発話パワーとして発話パワー記憶部240に記憶させる。

【0154】

- アニメーションの作成 -

アニメーションの作成においては、様々な選択肢がある。第1の選択肢は、キーフレームの削除率である。キーフレームの削除は常に行なわれるので、この指定は必須である。ただし、指定がない場合には所定のデフォルトの値を使用するようにしてもよい。第2の選択肢は、キーフレーム削除部236での処理及びブレンド率調整部250での処理において、クラスタリングの結果を使用するか否かの指定である。第3の選択肢は、発話パワーによるブレンド率調整部244の処理を行なうか否かである。さらに、発話パワーによるブレンド率調整部244の処理を実行する場合には減衰率を指定する必要がある。第4の選択肢は、ブレンド率調整部250の処理を行なうか否かである。ブレンド率調整部250の処理を行なう場合にはさらに、減衰率を指定する必要がある。

【0155】

発話パワーによるブレンド率調整部244による処理を行なうことが指定された場合には、選択部242及び246は、キーフレーム削除部236の出力を発話パワーによるブレンド率調整部244に与え、さらに発話パワーによるブレンド率調整部244の出力を選択部248に与えるように、接続を切替える。それ以外の場合には、選択部242及び246は、キーフレーム削除部236の出力を直接に選択部248に与えるように接続を切替える。

【0156】

一方、ブレンド率調整部250による処理を行なうことが指定された場合には、選択部248及び252は、選択部246の出力をブレンド率調整部250に与え、ブレンド率調整部250の出力を視覚素シーケンス記憶部254に与えるように接続を切替える。それ以外の場合には、選択部248及び252は、選択部246の出力を直接に視覚素シーケンス記憶部254に与えるように接続を切替える。

【0157】

以下、一般性を失わずに、発話パワーによるブレンド率調整部244による処理及びブレンド率調整部250による処理がともに選択されることを前提とし、クラスタリング後のモデルを使用しない場合と使用する場合とについて、それぞれキーフレーム削除部236、発話パワーによるブレンド率調整部244、及びブレンド率調整部250の動作を説明する。

【0158】

(1) クラスタリング後のモデルを使用しない場合

- キーフレーム削除部236の動作 -

キーフレーム削除部236は、削除率入力部201により入力された削除率を読み出し(図11、ステップ360)、視覚素シーケンス記憶部180に記憶された視覚素シーケンス中の視覚素の数の削除率を乗ずることにより、削除フレーム数Kを算出する(ステップ362)。

10

20

30

40

50

【 0 1 5 9 】

キーフレーム削除部 2 3 6 はさらに、ステップ 3 6 8 で削除フレーム数 K だけのキーフレームを削除したか否かを判定する。通常は最初の判定では削除フレーム数 K だけのキーフレームの削除は行われていない。したがってステップ 3 7 0 に進む。ステップ 3 7 0 では、クラスタリング後のモデルを使用することが指定されていないので、ステップ 3 7 2 に進む。

【 0 1 6 0 】

ステップ 3 7 2 では、視覚素シーケンス内の隣り合う全てのキーフレーム間で、全ての頂点を用いてキーフレーム間の距離 D を算出し、ステップ 3 7 4 でこの距離に基づいて各点の移動速度の合計が最も早いキーフレームを削除ターゲットに定める。そしてステップ 3 8 0 でこのキーフレームを削除する。この後ステップ 3 6 6 に戻る。

10

【 0 1 6 1 】

以後、削除したキーフレームの数が削除フレーム数 K より大きくなると処理を終了する。

【 0 1 6 2 】

キーフレーム削除部 2 3 6 によりこのようにして K 個のキーフレームが削除された視覚素シーケンスは選択部 2 4 2 を介して発話パワーによるブレンド率調整部 2 4 4 に与えられる。

【 0 1 6 3 】

- 発話パワーによるブレンド率調整部 2 4 4 の動作 -

20

発話パワーによるブレンド率調整部 2 4 4 は、最初に減衰率 を読出す (図 1 4 のステップ 4 5 0)。ステップ 4 5 2 で、キーフレーム削除部 2 3 6 の出力する視覚素シーケンス中の音素に関する情報に基づいて、発話記憶部 1 5 2 に記憶された発話データから、各音素の継続期間にわたる平均発話パワーを算出する。

【 0 1 6 4 】

ステップ 4 5 4 では、こうして算出された平均発話パワーのうち、最大パワー MAX (SP) と最小パワー MIN (SP) とを算出し、ステップ 4 5 6 において、減衰率 を用いた式により、各キーフレームについてブレンド率 BR (N) を調整する。全てのキーフレームについてブレンド率を調整された視覚素シーケンスは、選択部 2 4 6 及び選択部 2 4 8 を介してブレンド率調整部 2 5 0 に与えられる。

30

【 0 1 6 5 】

- 頂点速度によるブレンド率調整部 2 5 0 の動作 -

頂点速度によるブレンド率調整部 2 5 0 は、最初に減衰率 を読出し (図 1 5、ステップ 4 7 0)、選択部 2 4 8 から与えられた視覚素シーケンス中に含まれるキーフレームにこの減衰率 を乗算して不変フレーム数 L を算出する (ステップ 4 7 2)。続くステップ 4 7 4 では、ステップ 4 7 6 が選択される。

【 0 1 6 6 】

ステップ 4 7 6 では、選択部 2 4 8 から与えられた視覚素シーケンス中の全てのキーフレームに対し、その前後のキーフレームとの間での、全頂点の平均速度を算出する。ステップ 4 7 8 では、このようにして算出された平均速度をソートキーに、平均速度の降順にキーフレームをソートする。

40

【 0 1 6 7 】

ステップ 4 8 4 では、ステップ 4 7 8 でソートされたキーフレームの下位から L 個のキーフレームのうちの平均速度の最大値を < VS > の値に設定する。ステップ 4 8 6 で、ステップ 4 8 4 において設定された速度 < VS > の値を用い、前述した式によって、不変フレーム以外のキーフレームの各々について、そのブレンド率を調整する。不変フレーム以外の全てのキーフレームについてブレンド率の調整が終了すると、ブレンド率の調整が完了した視覚素シーケンスを図 5 に示す視覚素シーケンス記憶部 2 5 4 に出力する。

【 0 1 6 8 】

ブレンド処理部 2 5 6 は、視覚素シーケンス記憶部 2 5 4 に記憶された視覚素シーケン

50

スを読み出し、各キーフレームに対応する時刻にはそのキーフレームで指定された視覚素を用い、キーフレーム間のフレームの時刻では、そのフレームの両隣のキーフレームの間で、キーフレームに付されたブレンド率を用いた内挿によって中間の画像を作成する。このようにして、一定時間間隔のフレームの各々で、キーフレームの画像とそのブレンド率とを用いた内挿によって画像を作成することにより、アニメーションが作成される。

【 0 1 6 9 】

(2) クラスタリング後のモデルを使用する場合

クラスタリング後のモデルを使用する場合には、リップシンクアニメーション作成装置 2 0 0 の各部は以下のように動作する。

【 0 1 7 0 】

- キーフレーム削除部 2 3 6 の動作 -

図 1 1 を参照して、キーフレーム削除部 2 3 6 は、ステップ 3 6 0 ~ 3 6 8 までの処理についてはクラスタリング後のモデルを使用しない場合と同様に動作する。しかし、ステップ 3 7 0 の判定ではステップ 3 7 6 を選択する。ステップ 3 7 6 では、隣り合う全てのキーフレームの間で、代表頂点を用いて距離 D' を算出する。代表頂点を用いた距離 D' の算出については前述したとおりであるが、代表頂点の移動距離に、その代表頂点により代表されるクラスタ内の頂点の数を乗算し、その値を全てのクラスタにわたり合計することにより距離 D' が得られる。

【 0 1 7 1 】

ステップ 3 7 8 では、こうして算出された距離 D' を用い、頂点の動きが最も早いキーフレームを削除対象のキーフレームに決定する。ステップ 3 8 0 以下の処理は、クラスタリング後のモデルを使用しない場合と同様である。

【 0 1 7 2 】

- 発話パワーによるブレンド率調整部 2 4 4 の動作 -

発話パワーによるブレンド率調整部 2 4 4 は、クラスタリング後のモデルを使用しない場合と全く同様である。したがってここではその詳細は繰返さない。

【 0 1 7 3 】

- ブレンド率調整部 2 5 0 の動作 -

この場合、ブレンド率調整部 2 5 0 は以下のように動作する。図 1 5 を参照して、ステップ 4 7 0 及び 4 7 2 の処理はクラスタリング後のモデルを使用しない場合と同様である。ただし、ステップ 4 7 4 の判定ではステップ 4 8 0 が選択される。

【 0 1 7 4 】

ステップ 4 8 0 では、全キーフレームに対し、その前後のキーフレームとの間の頂点の平均速度を、各頂点が属するクラスタの代表頂点の動きベクトルを用いて算出する。ここでの算出方法はキーフレーム削除部 2 3 6 での算出方法と同様である。そしてステップ 4 8 2 において、このようにして算出された平均速度をソートキーに、全てのキーフレームを降順にソートする。この後は、ステップ 4 8 4 及び 4 8 6 をクラスタリング後のモデルを使用しない場合と同様に実行する。

【 0 1 7 5 】

図 1 6 に、キーフレーム削除部 2 3 6 によるキーフレーム削除の結果の一例を示す。図 1 6 (A) はキーフレーム削除部 2 3 6 によるキーフレームの削除なし (視覚素シーケンス作成部 2 3 0 による出力のまま。ただしブレンド率については発話パワーによって初期値を付与してある。) を示し、図 1 6 (B) 及び図 1 6 (C) はそれぞれ削除率 = 2 0 % 及び 3 0 % に設定したときの結果を示す。図 1 6 (D) は従来の方法にしたがい、人間のアニメーターが音声を聞きながら手作業によってキーフレームを設定した結果を示す。自動的な処理で図 1 6 (D) に近い結果が得られると好ましい。

【 0 1 7 6 】

図 1 6 (A) と図 1 6 (B) とを比較すると、キーフレーム 5 0 0 及び 5 0 2 が削除されていることが分かる。この結果、図 1 6 (B) と図 1 6 (D) とはかなり近い結果となっている。さらに図 1 6 (B) と図 1 6 (C) とを比較すると、キーフレーム 5 1 0 が削

10

20

30

40

50

除されている。この結果を図16(D)と比較すると、両者が非常に類似していることが分かる。特に図16(C)の結果から合成したアニメーションと、図16(D)の手作業による結果から合成したアニメーションとは、前半部分において非常によく一致しており、主観的な評価ではほとんど差がなかった。

【0177】

図17の上段(A)(B)は、従来の方法によって得られた顔画像の口付近のアニメーション結果(A)と、上記実施の形態によって得られたアニメーション結果(B)とを対比して示す。図17の下段(C)(D)は、対応する各キーフレームのブレンド率を示す。従来の方法によるブレンド率を図17(D)に、本発明の実施の形態によるブレンド率を図17(C)に、それぞれ示す。図17(C)における枠530、図17(D)における枠532に相当する部分の顔アニメーションが図17(B)及び(A)に該当する。

10

【0178】

図17(C)及び(D)を参照して、従来の方法によるブレンド率のグラフ522と、本実施の形態によるブレンド率のグラフ520とを比較すると、本実施の形態では全体にブレンド率が低くなり、その結果口画像の動きが滑らかになっていることが分かる。

【0179】

以上のように本実施の形態に係る視覚素シーケンス作成部230によれば、発話音声及びそのトランスクリプションと、視覚素に相当する基本的な顔画像のモデルとから、自動的に音声に対応して滑らかに変化する顔画像を作成することができる。発話パワーが小さい部分、又は隣接するキーフレームとの間のモデルの各頂点の動きが速すぎるキーフレームなどにおいては、ブレンド率は低くなるように調整される。その結果、得られる顔画像のアニメーションはいわゆる「うるさい」アニメーションではなく、滑らかで、手作業によってキーフレーム及びそのブレンド率を調整した場合に近いアニメーションを作成することができる。

20

【0180】

なお、上記実施の形態では、削除率入力部201によって入力された割合にしたがってキーフレームを削除する。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。例えば、削除率入力部201に代えて、削除されるべきキーフレームの数を入力するための機能部を設け、入力された数だけのキーフレームを削除するようにしてもよい。

【0181】

[コンピュータによる実現]

上述の実施の形態は、コンピュータシステム及びコンピュータシステム上で実行されるプログラムによって実現され得る。図18はこの実施の形態で用いられるコンピュータシステム550の外観を示し、図19はコンピュータシステム550のブロック図である。ここで示すコンピュータシステム550は単なる例であって、他の構成も利用可能である。

30

【0182】

図18を参照して、コンピュータシステム550はコンピュータ560と、全てコンピュータ560に接続された、モニタ562と、キーボード566と、マウス568と、スピーカ558と、マイクロフォン590と、を含む。さらに、コンピュータ560はDVD-ROM(Digital Versatile Disk Read-Only-Memory: デジタル多用途ディスク読出専用メモリ)ドライブ570と、半導体メモリドライブ572とを含む。

40

【0183】

図19を参照して、コンピュータ560はさらに、DVD-ROMドライブ570と半導体メモリドライブ572とに接続されたバス586と、全てバス586に接続された、CPU576と、コンピュータ560のブートアッププログラムを記憶するROM578と、CPU576によって使用される作業領域を提供するとともにCPU576によって実行されるプログラムのための記憶領域となるRAM580と、音声データ、音響モデル、言語モデル、レキシコン、及びマッピングテーブルを記憶するためのハードディスクド

50

ライブ574と、ネットワーク552への接続を提供するネットワークインターフェイス596を含む。

【0184】

図5に示す発話記憶部152、トランスクリプション記憶部154、キャラクタモデル記憶部156、音響モデル記憶部170、音素-視覚素マッピングテーブル記憶部176、視覚素シーケンス記憶部180、クラスタ化顔モデル記憶部234、発話パワー記憶部240、視覚素シーケンス記憶部254などは、いずれも図19に示すハードディスクドライブ574又はRAM580により実現される。また、削除率入力部201、クラスタ処理指定部202、発話パワー使用指示入力部204、減衰率入力部206、使用指示入力部208及び減衰率入力部210等は、いずれも図18及び図19に示すモニタ562並びにキーボード566及びマウス568を用いるグラフィカルユーザインタフェースを実現するプログラムによって実現される。そのような入力のプログラムの構成は周知であるので、ここではその詳細については説明しない。

10

【0185】

顔画像のアニメーション260の再生は、図示しないアニメーション再生プログラムによって実現される。アニメーション再生プログラム自体は、所定のタイムテーブルにしたがい、一定のフレーム間隔でフレームシーケンスを順次表示する、という機能を提供するものであればよい。

【0186】

上述の実施の形態のシステムを実現するソフトウェアは、DVD-ROM582又は半導体メモリ584等の媒体に記録されたオブジェクトコードの形で流通し、DVD-ROMドライブ570又は半導体メモリドライブ572等の読出装置を介してコンピュータ560に提供され、ハードディスクドライブ574に記憶される。CPU576がプログラムを実行する際には、プログラムはハードディスクドライブ574から読出されてRAM580に記憶される。図示しないプログラムカウンタによって指定されたアドレスから命令がフェッチされ、その命令が実行される。CPU576はハードディスクドライブ574から処理すべきデータを読み出し、処理の結果をこれもまたハードディスクドライブ574に記憶する。スピーカ558とマイクロフォン590とは、直接に本発明とは関係ないが、スピーカ558は、作成されたアニメーションの再生時の音声の発生に必要である。発話データの収録にコンピュータシステム550を使用するときには、マイクロフォン590が必要となる。

20

30

【0187】

コンピュータシステム550の一般的動作は周知であるので、詳細な説明は省略する。

【0188】

ソフトウェアの流通の方法に関して、ソフトウェアは必ずしも記憶媒体上に固定されたものでなくてもよい。例えば、ソフトウェアはネットワークに接続された別のコンピュータから分配されてもよい。ソフトウェアの一部がハードディスクドライブ574に記憶され、ソフトウェアの残りの部分をネットワークを介してハードディスクドライブ574に取込み、実行の際に統合する様にしてもよい。

【0189】

40

典型的には、現代のコンピュータはコンピュータのオペレーティングシステム(OS)によって提供される一般的な機能を利用し、所望の目的に従って制御された態様で機能を達成する。従って、OS又はサードパーティから提供されうる一般的な機能を含まず、一般的な機能の実行順序の組合せのみを指定したプログラムであっても、そのプログラムが全体として所望の目的を達成する制御構造を有する限り、そのプログラムがこの発明の範囲に包含されることは明らかである。

【0190】

[第2の実施の形態]

<概略>

上記した第1の実施の形態により、音声を基にして滑らかな顔画像のアニメーションを

50

作成することができる。しかし、商品としてのアニメーションでは、単に画像が滑らかなことに留まらず、様々な制約が与えられることがある。例えば、通常のアニメーションは、テレビ(30fps(frame per second))又は、映画(24fps)と同様のフレームレートで作成される。しかし、商業的なアニメーションでは、これよりも小さな(遅い)フレームレートでアニメーションを作成することが要請される場合がある。例えば、12fps、8fpsなどでアニメーションを作成することが要請される場合があり得る。こうした場合には、次のような問題が生じる。

【0191】

第1の実施の形態に係る装置では、アニメーション作成時のフレームレートは高く設定されており、従って滑らかな映像を得ることができる。しかし、敢えて低いフレームレートでアニメーションを作成する場合には、一つのキーフレームの継続長内に複数の音素が含まれる場合が多くなる。すると、本来は複数の視覚素を含む期間内に、口の画像が1種類しか含まれないこととなる。そのため、口画像にどの視覚素を割当てればよいか問題となる。この場合、一つのキーフレームの継続長に含まれる複数の視覚素のうちのいずれかを、そのキーフレームの視覚素に割当てることが妥当である。しかし、そうすると、場合によっては連続するキーフレームに同じ視覚素が割当てられてしまう場合があり得る。一般的に、8fpsという遅いフレームレートでアニメーションを作成する場合にも、最終的にはテレビ、映画などのフレームレートと同じフレームレートの画像を作成することになるため、連続するキーフレームに同じ視覚素が割当てられると、かなり長い期間にわたり同じ視覚素が続いてしまうということになり、アニメーションが不自然になってしまう恐れがある。

【0192】

これと関連した問題であるが、現在使用されているアニメーション作成プログラムでは、あるキーフレームと、その次のキーフレームとにそれぞれの形状を割当てると、その間に存在するフレームの映像については、これら二つのキーフレームの映像を自動的に補間して各フレームの画像を作成するという機能が標準的に備わっている。そうした場合、キーフレーム間の画像の変化は劣せずして滑らかなものとなるが、遅いフレームレートを前提として作成するアニメーションの場合には、意図したものと異なった動きが生成されることになる。遅いフレームレートの場合には、結果として作成されるアニメーションは「カクカク」とした動きをするものとなる。これは「リミット感」と呼ばれてアニメーション作成上の一つの技法とされている。そのようなリミット感を生成することが意図されたアニメーションでは、このような自動的な補間機能があるために、かえって意図したりリミット感を達成することができないという問題点が生ずる。

【0193】

更に、人間の発話の場合、発話終端で口を開いたままにするということはよくあるが、アニメーションでは、そのような形で発話を終わらせると不自然に感じられることがある。そこで、発話の終端では必ず口を閉じるように補正することが考えられる。しかし、この場合、どのように補正すれば自然に見えるかが問題となる。

【0194】

以後に説明する第2の実施の形態に係るリップシンクアニメーション作成装置は、こうした問題を解決するためのものである。

【0195】

- 発話終端補正 -

最初に、発話の終端で口を閉じるように補正するためのアニメーションの補正方法(以後この補正を「発話終端補正」と呼ぶ。)について説明する。図20を参照して、発話者の音声から得られたキーフレーム列610が、4つの連続するキーフレーム620, 622, 624, 626を含むものとする。これらのうち、キーフレーム626は発話後の空白期間を表している。

【0196】

本実施の形態では、発話の終端に相当するキーフレーム624について、以下のように

10

20

30

40

50

してその終端位置を調整する。

【0197】

図20を参照して、キーフレーム列610を作成するもととなった発話者の音声信号の発話パワー系列630を考える。本実施の形態では、キーフレーム624の終端位置(キーフレーム626の開始位置)からこの発話パワー系列630を時間軸上でさかのぼるようにして、キーフレーム624に相当する期間内で発話パワーが最大となる点640を探索する。次にこの点640における発話パワーの値から、所定の減衰量642(dB)だけ減衰した発話パワーを算出し、同じくキーフレーム624の終端から時間軸をさかのぼって、その発話パワーが減衰後の発話パワーと等しくなる点644を探索する。この点644に相当する時間軸上の位置をキーフレーム624の終端位置とする。

10

【0198】

その結果、図20に示されるように、キーフレーム626の位置が点644の位置まで進み、新たなキーフレーム652となり、その継続長はキーフレーム624の継続長が短縮された分だけ長くなる。こうして得られたキーフレーム列650を用いてアニメーションを作成すると、発話の最後において口が閉じる時期が早くなり、アニメーションとして自然なものになる。

【0199】

なお、この実施の形態では、最大の発話パワーを与える点640の後で所定の割合だけ発話パワーが減衰する時間軸上の位置をキーフレームの終端位置としている。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。例えば、発話パワーの大きさがある値だけ減少した時点をキーフレームの終端位置としてもよい。

20

【0200】

- フレームレート変換及び視覚素の割当処理 -

次に、低いフレームレートの時に、各キーフレームにどの視覚素を割当てるか、についての本実施の形態における決定方法について説明する。図21を参照して、キーフレーム列670が、6つのキーフレーム680, 682, 684, 686, 688及び690を含むものとする。フレームレートが8fps程度に遅くなると、キーフレームの時刻はフレーム位置に固定されてしまう。すなわち、キーフレームと所定のフレームレートの画像のフレーム位置とが、図21に示されるように一致する。

【0201】

一方、第1の実施の形態に係るリップシンクアニメーション装置によって得られたキーフレーム列672から、図21の上段に示すキーフレーム列670を生成する場合を考える。なお、キーフレーム列672は、キーフレーム700, 702, 704, 706, 708, 710, 712, 714及び716を含むものとする。

30

【0202】

この場合、キーフレーム列670の各キーフレームの継続長は、キーフレーム列672の各キーフレームの継続長と比較して長くなるため、キーフレーム列670の一つのキーフレームの継続長に対し、キーフレーム列672の複数のキーフレームの視覚素が対応する。例えば、キーフレーム682に対しては、時間的に隣接する三つのキーフレーム702, 704及び706の視覚素が割当てられる可能性がある。同様にキーフレーム688に対しては、キーフレーム714及び716の視覚素が割当てられる可能性がある。このように一つのキーフレームに複数の視覚素が割当てられる可能性があるときに、どの視覚素を選択すればよいか問題となる。

40

【0203】

ところで、実際の発話では、音声の発生を行なうに先立って口の動きが生ずるのが観察される。しがたって、音声より先にその音声に対応するように口を動かせるのが自然である。本実施の形態では、そのような考え方にしたいが、図21に示すキーフレーム列670の各キーフレームに視覚素を割当てると、キーフレーム列672の中で、そのキーフレームの継続長内に視覚素の始端を有するキーフレームの視覚素を割当てることとする。

【0204】

50

例えば、図 2 2 を参照して、楕円 7 3 0 で示したキーフレーム 6 8 2 について考える。前述したように、このキーフレーム 6 8 2 に対しては、キーフレーム列 6 7 2 の三つのキーフレーム 7 0 2 , 7 0 4 及び 7 0 6 が対応する可能性がある。しかしこれらのうち、キーフレーム 7 0 2 についてはその始端がキーフレーム 6 8 2 の継続長内にないため、候補からは外れる。キーフレーム 6 8 2 の継続長内に始端を有するという条件を充足するのは、キーフレーム 7 0 4 及び 7 0 6 である。このように二つ以上の視覚素がキーフレーム 6 8 2 内に存在する場合、先に生ずる視覚素をこのキーフレーム 6 8 2 に割当てるのが自然である。したがって本実施の形態では、矢印 7 3 4 で示されるように、キーフレーム 7 0 4 の視覚素 N (/ n /) をキーフレーム 6 8 2 に割当てることとする。点線の矢印 7 3 2 及び 7 3 6 で示されるように、二つのキーフレーム 7 0 2 及び 7 0 6 の視覚素は、キーフレーム 6 8 2 には割当てられない。

10

【 0 2 0 5 】

ところでこうした場合、得られる映像に問題が生ずる可能性がある。例えば図 2 2 において楕円 7 4 0 で示すように、キーフレーム 6 8 8 に対し、その継続長内に始端を有するキーフレーム 7 1 4 及び 7 1 6 がある。これらのいずれもキーフレーム 6 8 8 の視覚素に割当てするための条件は充足している。しかし、例えば図 2 2 に示すように、その直前のキーフレーム 6 8 6 に対し、視覚素 A (/ a /) が割当てられている場合、キーフレーム 6 8 8 に対しキーフレーム 7 1 4 の視覚素 A (/ a /) を割当てると、二つのキーフレーム 6 8 6 及び 6 8 8 の視覚素が全く同一となってしまふ。前述したようにこの場合、かなり長い時間にわたって同じ視覚素が連続してしまうため、アニメーションが不自然になると

20

いう問題点がある。

【 0 2 0 6 】

そこでこうした場合には、キーフレーム 7 1 4 ではなく、2 番目のキーフレーム 7 1 6 の視覚素 I (/ i /) をキーフレーム 6 8 8 に割当てることとする。

【 0 2 0 7 】

このようにすることにより、元々高速なフレームレートを想定して作成されたキーフレーム列 6 7 2 から、かなり低いフレームレートのキーフレーム列 6 7 0 を作成し、しかもそこから得られるアニメーションの顔画像に不自然さがそれほどないものを作成することができる。

【 0 2 0 8 】

以上のようにして、図 2 2 において実線の矢印 7 5 0 , 7 3 4 , 7 5 2 , 7 5 4 , 及び 7 4 4 で示される視覚素がキーフレーム列 6 7 0 の各キーフレームに割当てられる。なお、図 2 2 においてキーフレーム列 6 7 0 の最後尾に示されているキーフレーム 6 9 0 には、キーフレーム列 6 7 2 の、図示されない次のキーフレームの視覚素が矢印 7 5 6 によって示される様に割当てられる。

30

【 0 2 0 9 】

- 形状安定化処理 -

ところで、先ほど述べたリミット感について、図 2 2 に示すようにキーフレーム 6 8 6 及び 6 8 8 に異なる視覚素の口形状を割当てたとする。通常使用されているアニメーション作成プログラムでは、この二つのキーフレームの間のフレームの画像については、この二つのキーフレームの間の補間を行なうことによって生成するのが一般的である。その結果、意図したりミット感が得られなくなるという問題がある。この問題を図 2 3 (A) を参照して説明する。

40

【 0 2 1 0 】

図 2 3 (A) を参照して、キーフレーム 6 8 6 に相当する時刻を時刻 t 、キーフレーム 6 8 8 に相当する時刻を時刻 $t + 6$ とする。すなわち、この二つのキーフレームの間に、5 つのフレームが存在している。時刻 t では、このキーフレーム 7 9 0 における視覚素 / a / のブレンド率は、印 7 7 0 によって示されるように 1 0 0 % であり、視覚素 / i / のブレンド率は 印 7 7 4 で示されるように 0 % である。一方、時刻 $t + 6$ では、逆に視覚素 / i / のブレンド率は 印 7 7 6 で示されるように 1 0 0 % であり、視覚素 / a / の

50

ブレンド率は 印 7 7 2 で示されるように 0 % となる。そしてこの間の両者のブレンド率は、ブレンド率曲線 7 8 0 及び 7 8 2 によって示されるように計算される。時刻 t 及び時刻 $t + 6$ の間の各フレームでは、このブレンド率によってこの二つの視覚素の顔画像をブレンドした顔画像が作成される。このようなブレンドを行なうと画像は滑らかに変化するが、それによってリミット感が失われ、小さなフレームレートでアニメーションを作成するという要請を充足することができなくなるという問題点がある。

【 0 2 1 1 】

そこで本実施の形態では、図 2 3 (B) に示されるように、時刻 $t + 6$ の直前のフレームに相当する時刻 $t + 5$ に、時刻 t における視覚素 / a / 及び / i / のブレンド率をそのままにして、キーフレーム 7 9 0 をキーフレーム 7 9 2 としてコピーする。その結果、アニメーション作成プログラムによって自動的なブレンドが行なわれる場合であっても、時刻 $t \sim t + 5$ の間では、直線 8 0 0 及び 8 0 2 によって示されるように視覚素 / a / のブレンド率は 1 0 0 %、視覚素 / i / のブレンド率は 0 % に維持される。顔画像の変化は時刻 $t + 5 \sim t + 6$ の間で行なわれることになり、上記したリミット感を達成することができる。

【 0 2 1 2 】

< 構成 >

図 2 4 に、この第 2 の実施の形態に係るリップシンクアニメーション作成装置 8 1 0 のブロック図を示す。このリップシンクアニメーション作成装置 8 1 0 の構成は、図 5 に示す第 1 の実施の形態に係るリップシンクアニメーション作成装置 2 0 0 の構成とほぼ同様であるが、図 5 に示すキーフレーム削除部 2 3 6 と選択部 2 4 2 との間に、前述した発話終端の補正を行なうための発話終端補正部 8 2 2、及びこの発話終端補正部 8 2 2 の機能を利用するか否かを選択するための選択部 8 2 0 及び 8 2 4 を更に含む点と、継続長付き視覚素シーケンス記憶部 2 5 4 の出力を受けるように接続され、継続長付き視覚素シーケンスのフレームレートを、フレームレート入力 8 3 2 によって指定されたフレームレートに変換するためのフレームレート変換部 8 4 0 と、フレームレート変換部 8 4 0 の出力する視覚素シーケンスについて、アニメーション作成プログラムによるブレンドによってリミット感が失われるのを防ぐための形状安定化処理を実行するための形状安定化処理部 8 4 2 と、形状安定化処理部 8 4 2 の出力するフレームレート変換後の継続長付き視覚素シーケンスを記憶するための継続長付き視覚素シーケンス記憶部 8 4 6 と、継続長付き視覚素シーケンス記憶部 2 5 4 及び 8 4 6 の出力にそれぞれ接続された第 1 及び第 2 の入力を有し、フレームレート変換を使用するか否かを指定する使用指示入力 8 3 0 の指示にしたがい、継続長付き視覚素シーケンス記憶部 2 5 4 の出力又は継続長付き視覚素シーケンス記憶部 8 4 6 の出力のいずれかを選択してブレンド処理部 2 5 6 に与えるための選択部 8 4 8 とを含む点において、図 5 に示すリップシンクアニメーション作成装置 2 0 0 と異なっている。

【 0 2 1 3 】

なお、図 2 4 に示す選択部 8 2 0 及び 8 2 4 は、発話終端補正を行なうか否かを指定する使用指示入力 8 2 6 にしたがって、キーフレーム削除部 2 3 6 の出力を発話終端補正部 8 2 2 を経由して選択部 2 4 2 に与える処理と、発話終端補正部 8 2 2 を経由せず直接に選択部 2 4 2 に与える処理とを選択的に行なう。また発話終端補正部 8 2 2 には、図 2 0 を参照して説明した減衰率 (dB) の入力 8 2 8 が与えられる。使用指示入力 8 2 6 と使用指示入力 8 3 0 とは、互いに同一の指示を用いるようにしてもよい。

【 0 2 1 4 】

既に述べたように、このリップシンクアニメーション作成装置 8 1 0 の発話終端補正部 8 2 2、フレームレート変換部 8 4 0、及び形状安定化処理部 8 4 2 は、コンピュータハードウェアと、そのハードウェア上で実行されるコンピュータプログラムとにより実現され得る。以下、それらプログラムの制御構造について説明する。

【 0 2 1 5 】

図 2 5 は、発話終端補正部 8 2 2 を実現するためのコンピュータプログラムの制御構造

10

20

30

40

50

を示すフローチャートである。

【0216】

図25を参照して、このプログラムは、キーフレーム削除部236から出力されるキーフレーム列のうち、未処理の発話終端を探すステップ870と、未処理の発話終端があったか否かを判定し、発話終端がない場合には処理を終了し、発話終端があった場合には次のステップに制御を移す判定ステップ872と、未処理の発話終端があると判定ステップ872で判定された場合に、その発話終端の直前のキーフレームの視覚素継続長内の音声パワーの最大値 P_{max} を求めるステップ874とを含む。

【0217】

ステップ870における未処理の発話終端を探す処理は、空白の視覚素が割当てられたキーフレームの直前の、空白以外の視覚素の割当てられたキーフレームを探すことにより行なわれる。ステップ874で行なわれる最大値 P_{max} を求める処理については、図20を参照して説明した通りである。ここでいう最大値 P_{max} を与える点は、図20における点640に相当する。

【0218】

このプログラムは更に、ステップ874の後、処理中の視覚素継続長の終端からさかのぼり、音声パワーが $P_{max} - (dB)$ となる最初の時間 t を求めるステップ876と、そのような条件を充足する点があるか否かを判定し、条件を充足する点がない場合にはステップ870に分岐し、条件を充足する点がある場合には次のステップに処理を分岐させるステップ878と、ステップ878において条件を充足する点があると判定されたことに応答して実行され、その視覚素継続長の終端を、ステップ876で発見された時間 t に変更し、あわせてその直後のキーフレームの始端を同じく時間 t に変更する処理を行なうステップ880とを含む。ステップ880の後、制御はステップ870に戻る。ステップ876で求める時間 t の点は、図20で説明した点644に相当する。

【0219】

図26は、図24に示すフレームレート変換部840を実現するためのコンピュータプログラムの制御構造を示すフローチャートである。図26を参照して、このプログラムは、以後の繰返し処理において処理対象のキーフレーム数を表す変数 i に値0を設定するステップ900と、変数 i に1を加算するステップ902と、ステップ902での加算処理の結果、変数 i が全てのキーフレームの数より大きくなったか否かを判定し、大きくなった場合にはこの処理を終了し、それ以外の場合には次のステップに制御を分岐させるステップ904とを含む。

【0220】

このプログラムは更に、ステップ904において、変数 i がキーフレーム数より大きくないと判定されたことに応答して実行され、 i 番目のキーフレーム(以後このキーフレームを「キーフレーム(i)」と書く。)の継続長内に始端が含まれる視覚素を探すステップ906と、ステップ906で見つけられた視覚素の数 N が0か否かを判定し、その結果によって処理を分岐させるステップ908と、ステップ908で、視覚素の数 $N=0$ と判定されたことに応答して実行され、キーフレーム(i)を破棄する処理を行ない、更にステップ902に制御を戻すステップ910と、ステップ908によって視覚素の数 N が0でない判定されたことに応答して実行され、視覚素の数 N が1か否かを判定し、その判定結果にしたがって制御を分岐させる処理を行なうステップ912と、ステップ912において視覚素の数 N が1であると判定されたことに応答して実行され、キーフレーム(i)に、ステップ906で見つけられた視覚素(この視覚素はこの場合一つしかないのをこれを視覚素(1)と書く。)を割当て、制御をステップ902に戻すステップ914と、ステップ912において視覚素の数 N が1でない判定されたことに応答して実行され、以後の処理でキーフレーム(i)の継続長内に始端が含まれる視覚素の、先頭からの数を表す変数 j に0を設定するステップ916とを含む。

【0221】

このプログラムは更に、ステップ916に引き続いて、変数 j に1を加算するステップ

10

20

30

40

50

918と、ステップ918での加算の結果、変数jの値が、キーフレーム(i)内の視覚素の数Nより大きくなったか否かを判定し、その判定結果にしたがって制御を分岐するステップ920と、ステップ920において変数jの値が視覚素の数Nより大きいと判定されたことに応答して実行され、キーフレーム(i)に、キーフレーム(i)内に始端を有する先頭の視覚素(視覚素(1))を割当て、制御をステップ902に戻すステップ922と、ステップ920において変数jの値が視覚素の数Nより大きくはないと判定されたことに応答して実行され、キーフレーム(i)内のj番目の視覚素(これを「視覚素(j)」と書く。)が、一つ前のキーフレーム(キーフレーム(i-1))の視覚素と同一か否かを判定し、その判定結果にしたがって制御を分岐させるステップ924とを含む。

【0222】

ステップ924において、視覚素(j)がキーフレーム(i-1)の視覚素と一致すると判定された場合には、制御はステップ918に戻り、それ以外の場合には制御は次に進む。

【0223】

このプログラムは更に、ステップ924において視覚素(j)がキーフレーム(i-1)の視覚素ではないと判定されたことに応答して実行され、キーフレーム(i)に視覚素(j)を割当て、更に制御をステップ902に戻す処理を行なうステップ926を含む。

【0224】

図27に、図24に示す形状安定化処理部842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。図27を参照して、このプログラムは、以後の処理において処理対象となるキーフレームの番号を表す変数iに1を設定するステップ950と、変数iに1を加算するステップ952と、ステップ952での加算処理の結果、変数iの値が処理対象のキーフレーム数より大きくなったか否かを判定し、変数iの値がキーフレーム数を上回った場合に処理を終了させるステップ954と、ステップ954において変数iの値がキーフレーム数を上回ってはいないと判定されたことに応答して実行され、キーフレーム(i)の直前のフレームに、キーフレーム(i-1)をコピーして新たなキーフレームとする処理を行ない、その後ステップ952に制御を戻す処理を行なうステップ956等を含む。

【0225】

<動作>

図24に示すリップシンクアニメーション作成装置810は以下のように動作する。以下の説明では、使用指示入力826と830とは、同一の値をリップシンクアニメーション作成装置810に指示するものとする。使用指示入力826及び830が、発話終端補正部822による処理、フレームレート変換部840による処理、及び形状安定化処理部842による処理を使用しないことを指定する値である場合、選択部820及び824はキーフレーム削除部236の出力を選択部242の入力に直接に与える。選択部848は、継続長付き視覚素シーケンス記憶部254の出力をブレンド処理部256に与える。したがってこの場合リップシンクアニメーション作成装置810の構成は事実上図5に示すリップシンクアニメーション作成装置200と同一となり、リップシンクアニメーション作成装置200と同様の動作を行なう。

【0226】

使用指示入力826及び830が、発話終端補正部822、フレームレート変換部840、及び形状安定化処理部842を使用することを指定する値である場合、選択部820はキーフレーム削除部236の出力を発話終端補正部822に与える。発話終端補正部822の出力は選択部824を介して選択部242の入力に与えられる。

【0227】

一方、選択部848は、継続長付き視覚素シーケンス記憶部254の出力ではなく、継続長付き視覚素シーケンス記憶部846の出力を選択し、ブレンド処理部256に与える。フレームレート変換部840は、フレームレート入力832に応答し、継続長付き視覚素シーケンス記憶部254に記憶された視覚素シーケンスを順に読出し、図21及び図2

10

20

30

40

50

2 に示した手法を用いてフレームレートを変換し、さらに各フレームに視覚素を割当てて、フレームレート変換後の視覚素シーケンスを形状安定化処理部 8 4 2 に与える。形状安定化処理部 8 4 2 は、フレームレート変換部 8 4 0 から出力される視覚素シーケンスの中で、各キーフレームを、次のキーフレームの直前のフレームにコピーする処理を行なう。この処理は図 2 3 に示した通りである。この処理を全てのキーフレームに対して行なった後、その結果を継続長付き視覚素シーケンス記憶部 8 4 6 に出力する。

【 0 2 2 8 】

既に述べたように選択部 8 4 8 は継続長付き視覚素シーケンス記憶部 8 4 6 の出力を選択してブレンド処理部 2 5 6 に与える。ブレンド処理部 2 5 6 は、継続長付き視覚素シーケンス記憶部 8 4 6 に記憶されたキーフレーム列を読み込み、隣接するキーフレームの間で、それぞれ指定されたブレンド率をその間のフレームに内挿することにより、アニメーションを作成して出力する。こうして作成されるアニメーション 2 6 0 のフレームレートは、テレビ又は映画のフレームレートと同じフレームレートであるが、フレームレート変換部 8 4 0 によってキーフレームが削除され、更に形状安定化処理部 8 4 2 によって、隣接するキーフレーム間でのアニメーションの内挿を防止するように形状安定化処理が行なわれているため、実質的にフレームレート入力 8 3 2 で指定されたフレームレートの値にしたがった低いフレームレートのアニメーションと同様のリミット感を得ることができる。

【 0 2 2 9 】

[第 3 の実施の形態]

< 概略 >

上記した第 1 及び第 2 の実施の形態により、視覚素 / A /、/ I /、/ U /、/ E /、/ O /、及び / N / (以下「標準視覚素」と呼び、これらに対応する音素を「標準音素」と呼ぶ。)に基づいた顔画像のアニメーションを作成することができる。しかし、日本語の場合、視覚素は標準視覚素を含めて十数種類あるので(/ K /、/ S /、/ T / 等)、標準視覚素のみでは、日本語の滑らかなアニメーションを作成するには十分ではない可能性がある。また、上記実施の形態において、標準視覚素のための顔画像は予め用意されていたが、他の視覚素も用いて日本語のアニメーションを作成するのであれば、準備しなければならない顔画像の数が増加する。こうした顔画像のための顔モデルは、アニメーション作成に使用する基準となる標準顔モデルに対して手作業で編集を加えて作成するため、多くの視覚素のための顔画像を用意するのは困難である。英語、中国語等のような外国語のアニメーションを作成するときには、さらに異なる視覚素について顔画像を作成しなくてはならず、したがってさらに困難になる。

【 0 2 3 0 】

以後に説明する第 3 の実施の形態に係るリップシンクアニメーション作成装置は、標準視覚素と、標準視覚素以外の視覚素(以下、これらを「一般視覚素」と呼ぶ。)とを含む視覚素群を用いた日本語のリップシンクアニメーションの作成、及びその多言語への拡張のためのものである。

【 0 2 3 1 】

< 構成 >

図 2 8 に、この第 3 の実施の形態に係るリップシンクアニメーション作成装置 1 0 0 0 のブロック図を示す。図 2 8 に示すこのリップシンクアニメーション作成装置 1 0 0 0 の構成は、図 2 4 に示す第 2 の実施の形態に係るリップシンクアニメーション作成装置 8 1 0 の構成とほぼ同様であるが、標準視覚素のみではなく、一般視覚素も用いて日本語の顔画像のアニメーション 2 6 0 を作成するためのものである点において、図 2 4 に示すリップシンクアニメーション作成装置 8 1 0 と異なっている。

【 0 2 3 2 】

具体的には、リップシンクアニメーション作成装置 1 0 0 0 は、図 2 4 に示す音素 - 視覚素マッピングテーブル記憶部 1 7 6 に代え、それと同様の構成ではあるが、日本語の音素の各々に対し、標準視覚素と、それ以外の視覚素とを含む視覚素群の中から、一つの視覚素を関連付ける点で図 2 4 に示す音素 - 視覚素マッピングテーブル記憶部 1 7 6 と異な

10

20

30

40

50

る音素 - 視覚素マッピングテーブルを記憶するための音素 - 視覚素マッピングテーブル記憶部 1002 を含む点と、図 24 に示す、標準視覚素に対応した顔モデル（以下「標準視覚素モデル」と呼ぶ。）を格納した 3D キャラクタモデル記憶部 156 に代えて、標準視覚素だけでなく、それ以外の日本語の視覚素のための、標準顔モデルを基準とした顔モデル（以下「一般視覚素モデル」と呼ぶ。）からなる 3D キャラクタモデルを記憶する 3D キャラクタモデル記憶部 1004 を含む点とにおいて図 24 に示すリップシンクアニメーション作成装置 810 と異なっている。

【0233】

リップシンクアニメーション作成装置 1000 はさらに、ある発話者が日本語の文を発音しているときにキャプチャした、顔の特徴点の 3 次元データ（以下「キャプチャデータ」と呼ぶ。）を、そのとき発音していた音素と関連付けて記憶するキャプチャデータ記憶部 1006 と、標準視覚素モデルを記憶した標準視覚素モデル記憶部 1008 と、キャプチャデータ記憶部 1006 に記憶されたキャプチャデータ及び標準視覚素モデル記憶部 1008 に記憶された標準視覚素モデルを使用して、標準音素以外の音素（/k/、/s/、/t/ 等）に対応するキャプチャデータの各々を、標準音素に対応するキャプチャデータの線形和で近似するための係数を算出するための係数算出部 1010 と、係数算出部 1010 により算出された係数を用いて、標準視覚素モデル記憶部 1008 に記憶された標準視覚素モデルの線形和で一般視覚素モデルを表し、標準視覚素モデルと一般視覚素モデルとを使用して 3D キャラクタモデルを作成してキャラクタモデル記憶部 1004 に格納するためのキャラクタモデル合成部 1012 を含む点において、図 24 に示すリップシンクアニメーション作成装置 810 と異なっている。

【0234】

一般視覚素の数をいくつにするか、一般視覚素として、どのようなものを選択するか、及び日本語の各音素を標準視覚素及び一般視覚素のうちどの視覚素と対応付けるかは設計事項に属する。ただし、標準音素は常に標準視覚素に対応付ける必要がある。

【0235】

図 29 を参照して、図 28 のキャプチャデータ記憶部 1006 に記憶されたキャプチャデータ、及び標準視覚素モデル記憶部 1008 に記憶された標準視覚素モデルを使用して、標準視覚素モデルによる線形和で一般視覚素モデルを近似するための係数を求める処理について説明する。

【0236】

図 29 を参照して、キャプチャデータ記憶部 1006 に、日本語の音素 /a/、/i/、/u/、/e/、/o/、/n/、/k/、/s/、/t/、/h/、及び /b/ 等を発話しているときの発話者の顔のキャプチャデータである、

【0237】

【数 7】

$/\tilde{A}/, /i/, /\tilde{U}/, /E/, /O/, /N/, /K/, /S/, /T/, /H/,$ 及び $/B/$

等がそれぞれ記憶されているものとする。/ $\sim N$ /（記号「 \sim 」は式中文字の上に付されている。）は、音素 /n/ を発話中の発話者の顔の特徴点のキャプチャデータである。/ $\sim N$ / 以外のキャプチャデータはいずれも、/ $\sim N$ / を基準とし、顔画像の各特徴点が、顔画像の定義されている 3 次元空間において、キャプチャデータ / $\sim N$ / の対応する特徴点からどの程度移動しているかを示す 3 次元ベクトル情報によって表されたものである。

【0238】

図 29 を参照して、標準視覚素モデル記憶部 1008 は、標準視覚素モデルである /A/、/I/、/U/、/E/、及び /O/ を、基準となる視覚素モデル /N/ からの、各特徴点の移動ベクトルの集合という形で記憶している。これら視覚素モデルはいずれも、アニメーションのキャラクタとして使用される標準視覚素モデルについて作成されたものである。

【0239】

10

20

30

40

50

係数算出部 1010 の機能は以下のとおりである。ここでは、例として、キャプチャデータ記憶部 1006 に記憶されているキャプチャデータから、音素 / k / に対応付けられた、アニメーション作成のための一般視覚素モデル / K / を求める方法について説明する。

【0240】

一般視覚素モデル / ~ K / を以下のように定式化する。

【0241】

【数 8】

$$|\tilde{K}| = \tilde{\alpha}_{KA} |\tilde{A}| + \tilde{\alpha}_{KI} |\tilde{I}| + \tilde{\alpha}_{KU} |\tilde{U}| + \tilde{\alpha}_{KE} |\tilde{E}| + \tilde{\alpha}_{KO} |\tilde{O}| + \varepsilon_K.$$

10

ただし、 $\tilde{\alpha}_{KA}$ 、 $\tilde{\alpha}_{KI}$ 、 $\tilde{\alpha}_{KU}$ 、 $\tilde{\alpha}_{KE}$ 、及び $\tilde{\alpha}_{KO}$ (記号「 $\tilde{\alpha}$ 」は式中文字の上に付されている。) は実数の値をとる変数であり、 ε_K は誤差変数である。この式は、一般視覚素モデル / ~ K / を構成する各特徴点の位置を表すベクトルの全てについてたてることができる。すなわち、キャプチャデータを構成する特徴点の数が M 個であれば、M 個のベクトルの線形和の等式が得られる。

【0242】

これら M 個のベクトルの線形和の等式の全てに関して算出した $\tilde{\alpha}_K$ の自乗和が最小となるような、 $\tilde{\alpha}_{KA}$ 、 $\tilde{\alpha}_{KI}$ 、 $\tilde{\alpha}_{KU}$ 、 $\tilde{\alpha}_{KE}$ 、及び $\tilde{\alpha}_{KO}$ を算出する。算出された $\tilde{\alpha}_{KA}$ 、 $\tilde{\alpha}_{KI}$ 、 $\tilde{\alpha}_{KU}$ 、 $\tilde{\alpha}_{KE}$ 、及び $\tilde{\alpha}_{KO}$ の値をそれぞれ α_{KA} 、 α_{KI} 、 α_{KU} 、 α_{KE} 、及び α_{KO} とする。係数算出部 1010 が行なう処理は、この係数を算出することである。

20

【0243】

キャラクタモデル合成部 1012 の機能は、係数算出部 1010 により算出されたこれら係数 α_{KA} 、 α_{KI} 、 α_{KU} 、 α_{KE} 、及び α_{KO} を用いて、一般視覚素モデルを構成する特徴点の各々の位置を表す 3 次元ベクトルの値を、標準視覚素モデルの線形和として算出し、キャラクタモデル記憶部 1004 に格納することである。

【0244】

以下では、音素 / k / に対応付ける、アニメーション作成のための一般視覚素モデル / K / を算出する場合を例としてキャラクタモデル合成部 1012 の機能を説明する。キャラクタモデル合成部 1012 は、一般視覚素モデル / K / を次の式にしたがって算出する。

30

【0245】

【数 9】

$$|K| \equiv \alpha_{KA} |A| + \alpha_{KI} |I| + \alpha_{KU} |U| + \alpha_{KE} |E| + \alpha_{KO} |O|.$$

この式は、一般視覚素モデル / K / を構成する特徴点の位置を表す 3 次元ベクトルの全てを、標準視覚素モデル / A /、/ I /、/ U /、/ E / 及び / O / を構成する特徴点の位置を表す 3 次元ベクトルの線形和で表すことを意味する。

【0246】

キャラクタモデル合成部 1012 は、同様にして、一般視覚素モデル / S /、/ T /、/ H /、及び / B / 等を、標準視覚素モデル / A /、/ I /、/ U /、/ E / 及び / O / の線形和として求める。

40

【0247】

そのようにして求められた一般視覚素モデルを、標準視覚素モデルとともにキャラクタモデル記憶部 1004 に記憶させる。

【0248】

テーブル 7 に、音素 - 視覚素マッピングテーブル記憶部 1002 に記憶されたマッピングテーブルの例を示す。

【0249】

【表 7】
テーブル7

視覚素	音素
/A/	/a/
/I/	/i/
/U/	/u/
/E/	/e/
/O/	/o/
/N/	/n/
/K/	/k/
/S/	/s/
/T/	/t/
/H/	/h/
/B/	/b/
:	:

10

テーブル7を参照して、本実施の形態では、上から1行目の音素/a/から5行目の/o/までは、第1の実施の形態で用いられたテーブル1と同様である。ただし、テーブル1と異なり、音素/n/は視覚素/N/にのみ対応付けられている。7行目では、音素/k/が、一般視覚素/K/に対応付けられている。8行目以下の音素/s/等についても7行目の音素/k/と同様である。このようなマッピングテーブルを用いると、音素が与えられるとそれに対応する視覚素が分かり、その視覚素のラベルと一致する視覚素ラベルを持つ視覚素モデルをキャラクタモデル記憶部1004から読出すことができる。

20

【0250】

<動作>

以上、構成を説明したリップシンクアニメーション作成装置1000は以下のように動作する。図28に示すリップシンクアニメーション作成装置1000の動作は、図24に示すリップシンクアニメーション作成装置810とほぼ同様であり、使用する日本語用3Dキャラクタモデルのみが異なっている。したがって、以下においては、本実施の形態において追加された、一般視覚素モデルを含む3Dキャラクタモデルを作成する際のリップシンクアニメーション作成装置1000の動作についてのみ詳細を述べ、それ以外の動作に関する説明は概略にとどめて、その詳細な説明は繰返さない。

30

【0251】

本実施の形態に係るリップシンクアニメーション作成装置1000では、顔画像のアニメーション260の作成のためには、音素-視覚素マッピングテーブルの作成と、一般視覚素モデルを含む3Dキャラクタモデルの作成という準備作業が必要である。以下それらの準備作業について述べる。

【0252】

- 音素-視覚素マッピングテーブルの作成 -

日本語の音素と、視覚素とを手作業で対応付け、機械可読な形式の音素-視覚素マッピングテーブルを作成し、音素-視覚素マッピングテーブル記憶部1002に記憶させる。このとき、第2の実施の形態と異なり、標準音素以外の音素を標準視覚素に対応付けなければならないわけではない。任意の音素を標準視覚素以外の視覚素(一般視覚素)に対応付けてもよい。こうして作成された音素-視覚素マッピングテーブルの一例が上記したテーブル7である。

40

【0253】

- 日本語用3Dキャラクタモデル記憶部1004の作成 -

係数算出部1010及びキャラクタモデル合成部1012は、以下のようにして標準視覚素モデルとともに一般視覚素モデルも含む3Dキャラクタモデルを作成する。ここで作成の対象となる一般視覚素モデルは、上記した音素-視覚素マッピングテーブルで音素と

50

対応付けられた視覚素の全てである。

【 0 2 5 4 】

図 29 を参照して、係数算出部 1 0 1 0 は、音素 - 視覚素マッピングテーブルで音素に対応付けられている任意の音素 - 視覚素のペアを選択し、キャプチャデータ記憶部 1 0 0 6 に記憶されているキャプチャデータのうち、選択されたペアの音素のラベルが付されたキャプチャデータ（これを便宜上「合成対象キャプチャデータ」と呼ぶ。）を読み出す。係数算出部 1 0 1 0 はさらに、キャプチャデータ記憶部 1 0 0 6 に記憶されているキャプチャデータのうち、標準音素に対応するキャプチャデータを全て読み出す。そして、既に述べたように、合成対象キャプチャデータを、標準音素に対応するキャプチャデータの線形和で近似するための係数を算出する。そして、この係数群に、合成対象キャプチャデータの音素と対応付けられている視覚素のラベルを付してキャラクタモデル合成部 1 0 1 2 に与える。

10

【 0 2 5 5 】

係数算出部 1 0 1 0 は、これと同様の処理を、音素 - 視覚素マッピングテーブル記憶部 1 0 0 2 に記憶されている音素 - 視覚素マッピングのうち、一般視覚素を含むもの全てについて繰り返す。

【 0 2 5 6 】

キャラクタモデル合成部 1 0 1 2 は、係数算出部 1 0 1 0 から与えられる係数群及び視覚素ラベルに基づき、次のような処理を行なう。すなわち、キャラクタモデル合成部 1 0 1 2 は、与えられた視覚素ラベルに対応する一般視覚素モデルを、標準視覚素モデル記憶部 1 0 0 8 に記憶された標準視覚素の線形和で表し、このとき、その係数として係数算出部 1 0 1 0 から与えられた係数を使用する。この結果、与えられた視覚素ラベルに対応する一般視覚素モデルが、標準視覚素モデルの線形和として表される。

20

【 0 2 5 7 】

キャラクタモデル合成部 1 0 1 2 は、係数算出部 1 0 1 0 から与えられる係数群及び視覚素ラベルからなる全ての組に対して上記した処理を繰り返す。結果をキャラクタモデル記憶部 1 0 0 4 に記憶させる。キャラクタモデル記憶部 1 0 0 4 に記憶される一般視覚素モデルには、該当する視覚素ラベルが付されている。

【 0 2 5 8 】

キャラクタモデル合成部 1 0 1 2 はまた、標準視覚素モデル記憶部 1 0 0 8 に記憶されている標準視覚素モデルも、対応する視覚素ラベルを付してキャラクタモデル記憶部 1 0 0 4 に記憶させる。

30

【 0 2 5 9 】

以上の処理により、日本語用の 3 D キャラクタモデルが完成する。

【 0 2 6 0 】

3 D キャラクタモデルが完成すると、後のリップシンクアニメーション作成装置 1 0 0 0 の動作は、第 2 の実施の形態に係るリップシンクアニメーション作成装置 8 1 0 と異なるところがない。ただし、アニメーションのキーフレームに使用される顔画像として、標準視覚素モデルから得られたものだけでなく、一般視覚素モデルから得られたものも使用できる。このため、作成されるリップシンクアニメーションは、第 2 の実施の形態において得られたものよりもさらに滑らかなものとなる。

40

【 0 2 6 1 】

[多言語への拡張]

上述の第 3 の実施の形態の説明においては、リップシンクアニメーション作成装置 1 0 0 0 が日本語のアニメーションを作成するための装置であることを前提としていた。しかし、実は上記第 3 の実施の形態における日本語用 3 D キャラクタモデルの作成方法は、英語、中国語等、日本語と異なる言語のアニメーションの作成にも、日本語の標準音素及び標準視覚素モデルを用いて拡張することができる。そして、そのような 3 D キャラクタモデルを使用する限り、リップシンクアニメーション作成装置 1 0 0 0 においてリップシンクアニメーションを作成する部分の構成の基本的部分はそのまま使用することができる。

50

【 0 2 6 2 】

例えば、英語のアニメーションを作成する場合における考え方を説明する。使用される言語が英語であるため、図 28 に示すリップシンクアニメーション作成装置 1000 において、次のような変更が必要となる。発話者が異なることを前提とすると、音響モデル記憶部 170 に記憶される音響モデルを英語の話者に対応したものに変更する必要がある。当然、アニメーション作成のための発話記憶部 152 及びトランスクリプション記憶部 154 も変わってくる。音素 - 視覚素マッピングテーブル記憶部 1002 についても、英語の音素とその音素の発音時の視覚素とに基づいて新たに作成する必要がある。話者が異なることが前提となっているため、キャプチャデータ記憶部 1006 に記憶されるキャプチャデータも英語の発話者から収録したものとする必要があるのである。

10

【 0 2 6 3 】

そしてこの場合、キャラクタモデル記憶部 1004 に記憶される 3D キャラクタモデルは以下のようにして作成する。図 30 に、英語のアニメーションを作成するための 3D キャラクタモデルを準備するための方法について説明する。

【 0 2 6 4 】

図 30 を参照して、この場合には、図 29 に示すキャプチャデータ記憶部 1006 には、英語の発話時の発話者の顔の特徴点の位置を表すキャプチャデータを準備する。このキャプチャデータは、頭部の揺動によるグローバルな座標変動を補正により除去した後、無音時のキャプチャデータを基準として、各特徴点が無音時の位置からどの程度移動したかによって表される。このキャプチャデータの中には、日本語の標準音素に相当する音素の発話時のキャプチャデータも含まれるものとする。

20

【 0 2 6 5 】

係数算出部 1010 は、音素 - 視覚素マッピングテーブル記憶部 1002 に記憶されている英語の音素 - 視覚素マッピングを参照し、そこに出現している音素 - 視覚素の組合わせごとに、その音素のラベルが付されているキャプチャデータを、日本語の標準音素に相当する音素の発話時のキャプチャデータの線形和で近似するよう、その係数群を最小自乗基準で決定する。音素 - 視覚素マッピングテーブルに出現する全ての音素について、この係数群を用いた線形和で一般視覚素モデルを作成し、標準視覚素モデルとともにキャラクタモデル記憶部 1004 に記憶し、対応する視覚素ラベルを付しておく。

【 0 2 6 6 】

以上のように、英語用の音素 - 視覚素マッピングテーブルを準備し、英語用 3D キャラクタモデルを準備し、英語用の発話者用の音響モデル記憶部 170 を準備し、英語の発話記憶部 152 とそのトランスクリプション記憶部 154 とを準備すると、後は第 3 の実施の形態において日本語のリップシンクアニメーションを作成した場合と全く同様に、英語のリップシンクアニメーションを作成することができる。キャラクタモデル記憶部 1004 に記憶された一般視覚素は全て日本語の標準視覚素の線形和で表されたものであるが、その線形和は英語のキャプチャデータに基づいて求められたものであるため、英語の発話時の顔画像をよく再現することができる。

30

【 0 2 6 7 】

以上の説明は日本語の標準顔モデルを用いて英語のリップシンクアニメーションを作成する場合に関するものであった。しかし、以上の説明から明らかなように、第 3 の実施の形態に係るリップシンクアニメーション作成装置 1000 は、そのような言語の組合せのみに限定的に適用可能なわけではない。任意の言語の組合せに対し、それらの発話時の発話者の顔画像の 3次元の位置を表すキャプチャデータが得られれば、全く同様にこのリップシンクアニメーション作成装置 1000 を適用してリップシンクアニメーションを作成できる。

40

【 0 2 6 8 】

[第 4 の実施の形態]

上記第 1 ~ 第 3 の実施の形態では、キーフレームの削除、発話終端の補正、発話パワーによるブレンド率の調整、及び頂点速度によるブレンド率の調整を行なっている。これら

50

の処理のうち、発話パワーによるブレンド率の調整と、頂点速度によるブレンド率の調整とは、それぞれ減衰率 α 及び β を指定することによってアニメーションの口形状の変化の度合いを調整できるという点で画期的な技術である。

【0269】

このように、減衰率 α によるブレンド率の調整においても、減衰率 β によるブレンド率の調整においても、各頂点の平均速度の調整が行なわれる場合には、各頂点の形状の変化の割合が一様になるため、安定したアニメーションが得られる。

【0270】

しかし、アニメーションを作成する場合には、口形状の変化により大きな変化の幅を持たせたり、頂点によってその変化の度合いを異ならせたりする効果を実現することが望ましいことがある。そのような技術が実現できれば、アニメーションの作成者にとっては、少なくともアニメーションの表現の幅を広げることができる可能性が与えられることになる。

10

【0271】

以下に、図31を参照して説明する本発明の第4の実施の形態に係るリップシンクアニメーション作成装置1040は、そうした技術を提供するものである。

【0272】

図31を参照して、リップシンクアニメーション作成装置1040が第3の実施の形態に係るリップシンクアニメーション作成装置1000(図28参照)と異なるのは、図28の減衰率入力部206に代えて、発話パワーによるブレンド率を調整するための減衰率 α と、減衰率 β に対するブレンド率の変化の感度を調整する感度率 ρ とを入力するための減衰率入力部1050を含むこと、及び図28の減衰率入力部210に代えて、頂点速度によるブレンド率の調整に用いられる減衰率 α と、この減衰率 α の値に対するブレンド率の変化の感度を調整するための感度率 ρ とを入力するための感度率入力部1060を含むことである。感度率 ρ 及び ρ は、発話パワーに対する関数及び頂点速度に対する関数の振舞を制御するためのパラメータである。

20

【0273】

リップシンクアニメーション作成装置1040がリップシンクアニメーション作成装置1000と異なるのはさらに、図28に示す発話パワーによるブレンド率調整部244に代えて、各キーフレームにおける発話パワーと、減衰率 α 及びその感度率 ρ とに基づいて、後述する式によってブレンド率を調整するための、発話パワーによるブレンド率調整部244を含むこと、及び図28に示す頂点速度によるブレンド率調整部250に代えて、減衰率 α 及びその感度率 ρ を用い、各キーフレームの頂点速度に基づき後述する式によってブレンド率を調整するための、頂点速度によるブレンド率調整部1062を含むことである。その他の点では、リップシンクアニメーション作成装置1040の各部は図28において同じ符号で示された各部と同一である。

30

【0274】

図31に示すリップシンクアニメーション作成装置1040において、頂点速度によるブレンド率調整部1062は、図28の頂点速度によるブレンド率調整部250と異なり、クラスタ化された顔モデルを用いて頂点速度を算出する機能は持たず、キャラクターモデルの各頂点の速度を実際に計算する。しかし、頂点速度によるブレンド率調整部1062を拡張して、クラスタ化された顔モデルを用いるようにすることは、第3の実施の形態の記述から容易に行なえる。

40

【0275】

図32は、図31の発話パワーによるブレンド率調整部1052を実現するためのコンピュータプログラムの制御構造をフローチャート形式で示す。図32を参照して、このプログラムは、減衰率 α 及び減衰感度 ρ をメモリから読出すステップ1080と、音素シークス中の全ての音素について、その継続長にわたる発話パワーの平均を算出するステップ452と、ステップ452で算出された全ての発話パワーの平均値の中で、最大のものMAX(SP)と、最小のものMIN(SP)とを決定するステップ454と、平均発

50

さくなり、発話パワーが大きいところでは逆に口形状の変化率が大きくなることわかる。

【0283】

この結果、式(3)におけるブレンド率の減衰の係数の値は、減衰感度 ρ の変化に対して図35のような傾向で変化することがわかる。なお図35は、減衰率 $\rho = 1$ として計算した。

【0284】

図35を参照して、減衰感度 ρ の値が1であれば、式(3)は式(1)と等しい。減衰感度 ρ の値が1より小さい場合、減衰の係数は発話パワーの小さいところでは減衰感度 $\rho = 1$ の場合より小さく、発話パワーの大きなところでは逆に減衰感度 $\rho = 1$ の場合より大きくなる。この場合、発話パワーの小さな発話シーンでは口形状の変化は非常に小さくなり、発話パワーの大きな発話シーンでは口形状の動きが通常よりも大きく感じられることになる。

10

【0285】

式(3)において減衰感度 $\rho = 0$ とすると、ブレンド率の変化に対する発話パワーの寄与はなくなる。すなわち、式(3)において減衰率 ρ はブレンド率に対する単なる定数係数となる。

【0286】

減衰率 ρ の値が大きくなれば、発話パワーの小さなところではブレンド率の減衰の係数は減衰感度 $\rho = 1$ の場合より大きく、その変化率も大きくなる。逆に発話パワーの大きなところでは1に近くなってそれほど変化しない。したがってこの場合、発話パワーの小さな発話シーンでは発話者の口形状は比較的是っきりと変化するが、発話パワーの大きな発話シーンでは口形状はほとんど全開となり、発話パワーが多少変化してもそれほど形状は変化しないことになる。

20

【0287】

図32のステップ1082において、全てのキーフレームに対して式(3)によるブレンド率の調整を行なうと、発話パワーによるブレンド率調整部1052による処理は終了する。

【0288】

図36は、図31に示す頂点速度によるブレンド率調整部1062を実現するプログラムのフローチャートである。この処理では、図を明りょうにするために、頂点速度の計算はクラスタの平均速度によるのではなく、各頂点の実速度の平均によって行なうこととしている。

30

【0289】

図36を参照して、このプログラムは、減衰率 ρ 及び減衰感度 ρ をメモリから読出すステップ1120と、ステップ1120に続き、処理対象の全キーフレームに対して、その前後のキーフレームとの間での全頂点の平均速度 V を算出するステップ1122と、ステップ1122の処理を受けて、全てのキーフレームについて算出された平均速度 V のうちの最大速度 $MAX(V)$ 及び最小速度 $MIN(V)$ を決定するステップ1124とを含む。

40

【0290】

このプログラムはさらに、ステップ1124の後、以下に述べる各処理を各キーフレームに対して行なうプロセス1126を含む。以後の処理は、口形状の変化を、同時調音を考慮して第1～第3の実施の形態でのスムージングよりもさらに滑らかにするためのものである。第1～第3の実施の形態での削除率 ρ を調整することにより、口形状の動きを滑らかにすることはできるが、同時に発話に対する口形状の正確さは失われてしまう。以後の処理は、1つのキーフレームだけでなく、前後のキーフレームでの口形状を考慮に入れて口形状の変化をスムーズでかつ発話に対してより正確なものにするための処理である。

【0291】

そのために、プロセス1126では、最初に処理対象のキーフレームを削除したものと

50

仮定して、その前後のキーフレームから処理対象のキーフレームでの口形状を内挿により求める。この内挿により求めた口形状と、処理対象のキーフレームの口形状とをブレンドすることにより、処理対象のキーフレームにおける新たな口形状を求める。このブレンドのときの、処理対象のキーフレームに割当てられるブレンド重みを r とする。

【0292】

プロセス1126は、次の式(4)により、ブレンド重み r を算出するステップ1130を含む。

【0293】

【数12】

$$r = 1 - \beta \times \left(\frac{V(n) - \text{MIN}(V)}{\text{MAX}(V) - \text{MIN}(V)} \right)^{\beta_p} \quad (4)$$

10

ただし、 β は第3の実施の形態で使用した頂点速度によるブレンド率調整における減衰率と同じ意味合いの減衰率、 β_p は減衰率 β の減衰感度、 $V(n)$ は n 番目のキーフレームの頂点の平均速度。

【0294】

式(4)から、ブレンド重み r は減衰感度 $\beta_p = 1$ のときには頂点速度 $V(n)$ に対する線形関数であるが、 $\beta_p > 1$ のときには非線形関数である。

【0295】

プロセス1126はさらに、ステップ1130に続き、次の式(5)にしたがって処理対象のキーフレーム(n 番目のキーフレームとする。)の前後のキーフレーム($n-1$ 番目と $n+1$ 番目のキーフレーム)から内挿した、処理対象のキーフレームにおける仮想的な口形状 S'_n を求めるステップ1132を含む。

【0296】

【数13】

$$S'_n = \frac{\{(t_{n+1} - t_n)S_{n-1} + (t_n - t_{n-1})S_{n+1}\}}{t_{n+1} - t_{n-1}} \quad (5)$$

20

ただし、 t_n 、 t_{n-1} 及び t_{n+1} はそれぞれ、 n 番目のキーフレーム、 $n-1$ 番目のキーフレーム、及び $n+1$ 番目のキーフレームの時刻であり、 S_{n-1} 及び S_{n+1} はそれぞれ、 $n-1$ 番目及び $n+1$ 番目のキーフレームにおける口形状。

30

【0297】

プロセス1126はさらに、ステップ1132に続き、ステップ1132で仮想的に内挿によって求められた口形状 S'_n 、ステップ1130で求められたブレンド重み r 、処理対象のキーフレームについて算出されている口形状 S_n 、及び処理対象のキーフレームに対して割当てられているブレンド率 $BR(n)$ とを用い、次の式(6)によって処理対象のキーフレームにおける口形状 \hat{S}_n を求めるためのブレンド率を求めるステップ1134とを含む。

【0298】

【数14】

$$\hat{S}_n = rBR(n)S'_n + (1-r)S_n \quad (6)$$

40

式(6)によれば、処理対象のキーフレームに対して予め求められている口形状 S_n にブレンド率 $BR(n)$ を乗じたものと、前後のキーフレームから内挿により仮想的に求められた口形状 S'_n との間で、式(4)により求められたブレンド重み r でブレンドを行なう。その結果、新たなキーフレーム \hat{S}_n が得られる。式(6)の S'_n に、式(5)を代入することで、新たなキーフレーム \hat{S}_n をキーフレーム S_{n-1} 、 S_n 及び S_{n+1} の加重和で算出するためのブレンド率が計算できる。この新たなブレンド率が減衰感度 $\beta_p = 1$ のときには頂点速度 $V(n)$ の頂点速度 $V(n)$ の線形関数であり、 $\beta_p > 1$ のと

50

きには非線形関数であることはいうまでもない。

【0299】

式(5)によって、時刻 t_{n-1} 及び t_{n+1} におけるキーフレーム S_{n-1} と S_{n+1} との間で、キーフレーム S_n の時刻 t_n における仮想的な口形状 S_n' が求められることは、図37から容易に理解可能である。

【0300】

この第4の実施の形態に係るリップシンクアニメーション作成装置1040の動作は、第3の実施の形態のリップシンクアニメーション作成装置1000において頂点速度によるブレンド率の算出でクラスタ化した顔モデルを用いない点、発話パワーによるブレンド率調整部1052及び頂点速度によるブレンド率調整部1062において、リップシンクアニメーション作成装置1000で用いた式ではなく、上記した式(3)及び式(4)～式(6)を用いる点において異なる。その他の点ではリップシンクアニメーション作成装置1040はリップシンクアニメーション作成装置1000と同様に動作する。したがってその詳細な説明は繰返さない。

10

【0301】

この第4の実施の形態では減衰感度 p 及び p を導入した。これら減衰感度を使用する式(3)及び式(4)～式(6)によって、キーフレームの口形状を算出するためのブレンド率を算出する。この結果、口形状の変化は単なる線形的な変化ではなくなる。例えば発話パワーの小さな部分では口形状がそれほど変わらず、発話パワーの大きな部分で口形状が大きく変化するようなアニメーションが得られる。それとは逆に、発話パワーの小さな部分では比較的大きく口形状が変化し、発話パワーの大きな部分では口形状は飽和してそれほど大きく変化しないようなアニメーションも得られる。したがって、アニメーションの作成者が、自分の好みと、アニメーションの要求仕様とにあわせて微妙にアニメーションの登場人物の口形状を変化させることが容易に行なえる。しかもそのためには、パラメータを変化させるだけでよい。様々な要求仕様と作成者の好みとにあわせたアニメーションを容易に作成することが可能になる。

20

【0302】

[第5の実施の形態]

上記した第1～第4の実施の形態では、各処理を実行するか否かの指定は可能ではあるものの、キーフレームの削除、発話終端補正、発話パワーによるブレンド率の調整、及び頂点速度によるブレンド率の調整を、この順序で行なっている。しかし本発明はそのような実施の形態に限定されるわけではない。特にコンピュータを用いたソフトウェアによって本発明のリップシンクアニメーション作成装置を実現する場合には、上記した処理はそれぞれ独立したプログラムとして作成されるのが通常である。また、これら処理の実行順序を入替えても特に問題は生じない。第5の実施の形態に係るリップシンクアニメーション作成装置は、操作者が最初に処理の順序を指定することにより、指定された順番で上記したキーフレームの削除、発話終端補正、発話パワーによるブレンド率の調整、及び頂点速度によるブレンド率の調整を行なう。処理の一部を実行しない指定も可能である。

30

【0303】

図38に、この実施の形態に係るリップシンクアニメーション装置のうち、上記した処理を実現するためのメインプログラムの制御構造をフローチャート形式で示す。図38を参照して、このプログラムは、上記した4つの処理を画面に表示し、その処理順序の指定と各処理で必要とされるパラメータの入力とを受けて、メモリ中の配列に各処理の識別子とパラメータとを記憶するステップ1160を含む。この処理では、予め4つのボックスを表示し、各ボックスについてどの処理を実行するかを選択させればよい。Graphical User Interfaceを用い、例えば各処理を示すアイコンを各ボックスにドロップすることで処理を指定するようにしてもよい。

40

【0304】

このプログラムはさらに、ステップ1160の後、以下の繰返し処理を制御するための変数 i に0を代入するステップ1162と、ステップ1162の後、変数 i に1を加算す

50

るステップ1164と、変数*i*の値が最大値MAXを超えたか否かを判定し、変数*i*の値が最大値MAXを超えたときには処理を終了させるステップ1166とを含む。最大値MAXは、本実施の形態の場合には4であるが、2又は3であってもよい。また、上記した処理以外の処理をアニメーションに対して実行する場合には、最大値MAXの値は5以上でもよい。

【0305】

このプログラムはさらに、ステップ1166で変数*i*の値が最大値MAXより大きくないと判定されたことに応答して実行され、*i*番目の処理の識別子が配列に記憶されているか否かを判定し判定結果に応じて制御の流れを分岐させるステップ1168と、ステップ1168において*i*番目の処理の識別子が記憶されていると判定されたことに応答して実行され、*i*番目の処理の識別子により特定されるプログラムの入力ファイル及び出力ファイルを、*i*番目の入力ファイル及び出力ファイルとして予め準備されているファイル名に指定しなおすステップ1170と、ステップ1170で指定しなおされた入力ファイルから継続長付視覚素データを読み出して*i*番目の処理のプログラムで処理し、ステップ1170で指定しなおされた出力ファイルに出力して制御をステップ1164に戻すステップと、ステップ1168で*i*番目の処理の識別子が記憶されていないと判定されたことに応答して実行され、*i*番目の処理で実際に使用される予定だった入力ファイルの内容を実際に使用される予定だった出力ファイルにコピーして制御をステップ1164に戻すステップ1174とを含む。このプログラムでは、*i*番目の処理の出力ファイルが、*i*+1番目の処理の入力ファイルとして使用される。

【0306】

この第5の実施の形態に係るリップシンクアニメーション作成装置は以下のように動作する。最初に、ステップ1160でどの処理をどのような順序で実行するかに関する指示を受け、メモリ内の配列に各プログラムの識別子として記憶する。同時に、キーフレーム削除処理における削除率、ブレンド率調整のための減衰率及び、並びに減衰感度 p 及び p のうち、指定された処理で必要とされるものの入力を受け、各プログラムに関連付けてメモリに格納する。

【0307】

続いて、ステップ1162で変数*i*に0を代入し、ステップ1164で変数*i*に1を加算する($i = 1$)。ステップ1166の判定結果はNOとなり、最初に指定された処理に対応するプログラムが実行される。この際、メインプログラムは、プログラム内で指定されている入力ファイルを、継続長付視覚素シーケンスが格納されているファイルに指定し直し、プログラム内で指定されている出力ファイルを、1番目の処理に対応するプログラムの出力ファイルとして実際に準備されているものに指定し直す。1番目の処理に対応するプログラムは、メインプログラムにより指定し直された入力ファイルから継続長付視覚素シーケンスを読み出して処理する。さらにこのプログラムは、処理後の継続長付視覚素シーケンスをメインプログラムにより指定し直された出力ファイルに出力する。

【0308】

次に変数*i*の値に1を加算する($i = 2$)。ステップ1166での判定結果はNOであり、ステップ1168の処理が実行される。仮に2番目の処理が指定されていない場合、ステップ1174で、2番目の処理の入力ファイルとして準備されていたファイル(すなわち1番目の処理の実際出力ファイル)の内容を、2番目の処理の出力ファイルとして準備されていたファイル(すなわち3番目の処理の実際入力ファイル)にコピーする。この後、制御はステップ1164に進む。

【0309】

このようにして、変数*i*の値が定数MAX=4より大きくなるまでプログラムの実行又はファイルのコピーが繰返され、最終的に必要な処理を全て受けた継続長付シーケンスファイルが、4番目の処理の出力ファイルとして得られる。この出力ファイルからアニメーションを作成する処理は、第1~第4の実施の形態における処理と同様であるので、ここではその詳細は繰返さない。

【0310】

以上のようにこの実施の形態に係るリップシンクアニメーション作成装置1040では、キーフレーム削除処理、発話終端補正処理、発話パワーによるブレンド率調整処理、及び頂点速度によるブレンド率調整処理のうち、所望の処理に対応するプログラムを、所望の順番で実行することができる。このようにして種々の処理をパラメータ、実行される処理の選択、及び処理の実行順序を変えながら行なうことで、口形状の変化態様が微妙に異なる多くの継続長付視覚素シーケンスを作成できる。それらの中で、要求使用に最も合致したアニメーションを生成することができたときのパラメータ、実行される処理の選択、及び処理の実行順序を採用することで、要求使用に合致し、しかも全体として一貫した印象のリップシンクアニメーションを作成することが可能になる。

10

【0311】

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及び範囲内でのすべての変更を含む。

【図面の簡単な説明】

【0312】

【図1】本発明の第1の実施の形態に係るアニメーション作成装置によるアニメーション作成過程30の概略を示す図である。

【図2】本発明の第1の実施の形態で使用される視覚素に対応する顔画像を示す図である。

20

【図3】ブレンド率の概念を説明するための図である。

【図4】ブレンドによる顔画像の変化例を示す図である。

【図5】本発明の第1の実施の形態に係るリップシンクアニメーション作成装置200の概略の機能的構成を示すブロック図である。

【図6】図5の視覚素シーケンス作成部230のより詳細なブロック図である。

【図7】各音素に対応する視覚素のうち、口周辺の画像を示す図である。

【図8】二つの視覚素の間の動きベクトルを説明するための図である。

【図9】クラスタリング後の顔画像の例を示す図である。

【図10】クラスタリング後の顔画像の他の例を示す図である。

30

【図11】図5のキーフレーム削除部236を実現するコンピュータプログラムの制御構造を示すフローチャートである。

【図12】キーフレームの削除を説明するための図である。

【図13】平均発話パワーの算出方法を説明するための図である。

【図14】図5の発話パワーによるブレンド率調整部244を実現するコンピュータプログラムの制御構造を示すフローチャートである。

【図15】図5の頂点速度によるブレンド率調整部250を実現するコンピュータプログラムの制御構造を示すフローチャートである。

【図16】本発明の実施の形態における種々の条件でのキーフレームの生成結果と、手作業によるキーフレームの指定結果とを対比して示す図である。

40

【図17】本発明の一実施の形態によって得られるアニメーションの結果を、従来の方法によるものと比較して示す図である。

【図18】コンピュータシステム550のハードウェア外観を示す図である。

【図19】コンピュータシステム550のブロック図である。

【図20】本発明の第2の実施の形態における発話終端補正の概略を説明するための模式図である。

【図21】本発明の第2の実施の形態における、フレームレート変換の概念を示す模式図である。

【図22】第2の実施の形態における、フレームレート変換後の各キーフレームに対し割当てる視覚素の決定方法を説明するための模式図である。

50

【図 2 3】第 2 の実施の形態における形状安定化処理を説明するための模式図である。

【図 2 4】第 2 の実施の形態に係るリップシンクアニメーション作成装置 8 1 0 の概略ブロック図である。

【図 2 5】図 2 4 に示す発話終端補正部 8 2 2 を実現するためのコンピュータプログラムのフローチャートである。

【図 2 6】図 2 4 に示すフレームレート変換部 8 4 0 を実現するためのコンピュータプログラムのフローチャートである。

【図 2 7】図 2 4 に示す形状安定化処理部 8 4 2 を実現するためのコンピュータプログラムのフローチャートである。

【図 2 8】第 3 の実施の形態に係るリップシンクアニメーション作成装置 1 0 0 0 の概略ブロック図である。

【図 2 9】図 2 8 のキャラクタモデル記憶部 1 0 0 4 に記憶される 3 Dキャラクタモデルを準備するためのより詳細な図である。

【図 3 0】英語のアニメーションを作成するための詳細な図である。

【図 3 1】第 4 の実施の形態に係るリップシンクアニメーション作成装置 1 0 4 0 の概略ブロック図である。

【図 3 2】リップシンクアニメーション作成装置 1 0 4 0 の発話パワーによるブレンド率調整部 1 0 5 2 を実現するプログラムの制御構造を示すフローチャートである。

【図 3 3】減衰感度 ρ の値を変化させたときの、発話パワーと正規化後の発話パワーとの関係を示す図である。

【図 3 4】減衰感度 ρ の変化による正規化後の発話パワーの変化を比較して示すグラフである。

【図 3 5】減衰感度 ρ の変化による、キーフレームの口形状に対する減衰係数の変化を比較して示す図である。

【図 3 6】リップシンクアニメーション作成装置 1 0 4 0 の頂点速度によるブレンド率調整部 1 0 6 2 を実現するプログラムの制御構造を示すフローチャートである。

【図 3 7】頂点速度によるブレンド率調整部 1 0 6 2 によるブレンド率の計算時の、内挿による仮想的なアニメーションの形状変化を算出する方法を模式的に示す図である。

【図 3 8】本発明の第 5 の実施の形態に係るリップシンクアニメーション作成装置において、キーフレーム削除処理、発話パワーによるブレンド率調整処理、及び頂点速度によるブレンド率調整処理を任意の順序で実行するためのプログラムの制御構造を示すフローチャートである。

【符号の説明】

【 0 3 1 3 】

4 0 話者

4 2 音声信号

4 4 台本

5 0 ~ 5 8 音素

6 0 ~ 6 8 , 8 0 顔画像

1 5 2 発話記憶部

1 5 4 トランスクリプション記憶部

1 5 6 , 1 0 0 4 キャラクタモデル記憶部

1 7 0 音響モデル記憶部

1 7 2 音素セグメンテーション部

1 7 4 音素シーケンス記憶部

1 7 6 , 1 0 0 2 音素 - 視覚素マッピングテーブル記憶部

1 7 8 音素 - 視覚素変換処理部

1 8 0 , 2 5 4 視覚素シーケンス記憶部

1 8 2 アニメーション作成部

2 0 0 , 8 1 0 , 1 0 0 0 リップシンクアニメーション作成装置

10

20

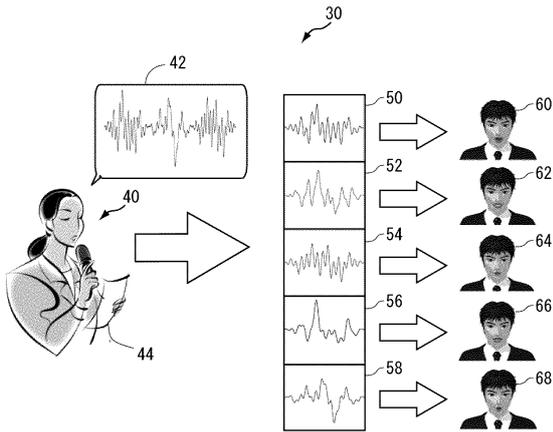
30

40

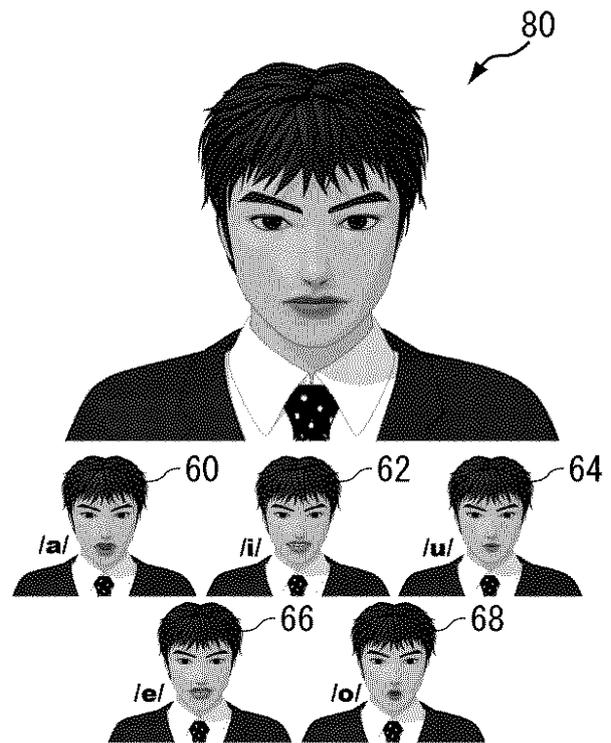
50

- 2 0 2 クラスタ処理指定部
- 2 0 4 発話パワー使用指示入力部
- 2 3 0 視覚素シーケンス作成部
- 2 3 2 クラスタリング処理部
- 2 3 4 クラスタ化顔モデル記憶部
- 2 3 6 キーフレーム削除部
- 2 3 8 発話パワー算出部
- 2 4 0 発話パワー記憶部
- 2 4 4 発話パワーによるブレンド率調整部
- 2 5 0 頂点速度によるブレンド率調整部 10
- 2 5 6 ブレンド処理部
- 2 6 0 顔画像のアニメーション
- 6 1 0 , 6 5 0 , 6 7 0 , 6 7 2 キーフレーム列
- 6 2 0 , 6 2 2 , 6 2 4 , 6 2 6 , 6 8 0 , 6 8 2 , 6 8 4 , 6 8 6 , 6 8 8 , 6 9 0
- , 7 0 0 , 7 0 2 , 7 0 4 , 7 0 6 , 7 0 8 , 7 1 0 , 7 1 2 , 7 1 4 , 7 1 6 , 7 9 0
- , 7 9 2 キーフレーム
- 8 2 2 発話終端補正部
- 8 4 0 フレームレート変換部
- 8 4 2 形状安定化処理部
- 1 0 0 6 キャプチャデータ記憶部 20
- 1 0 0 8 標準視覚素モデル記憶部
- 1 0 1 0 係数算出部
- 1 0 1 2 キャラクタモデル合成部

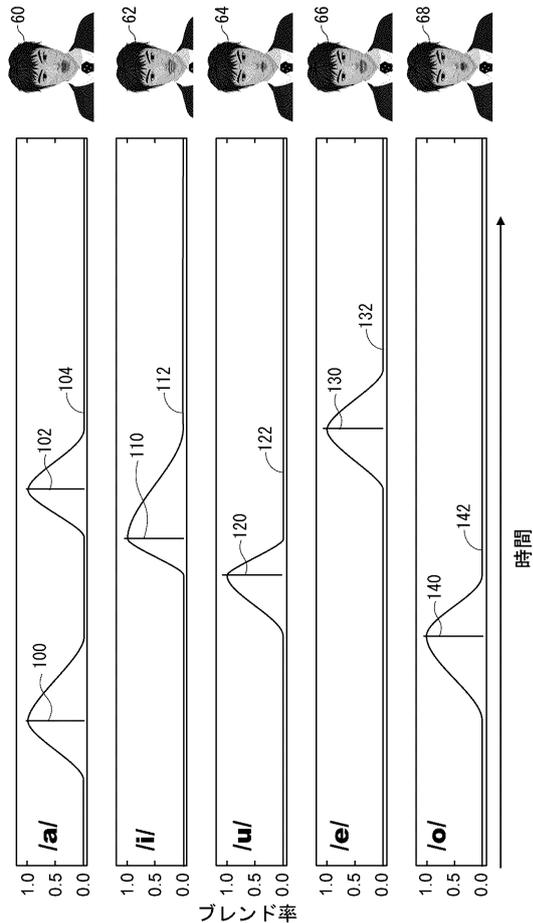
【図 1】



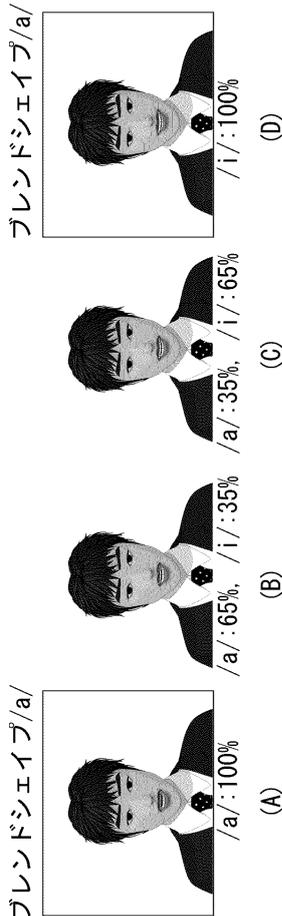
【図 2】



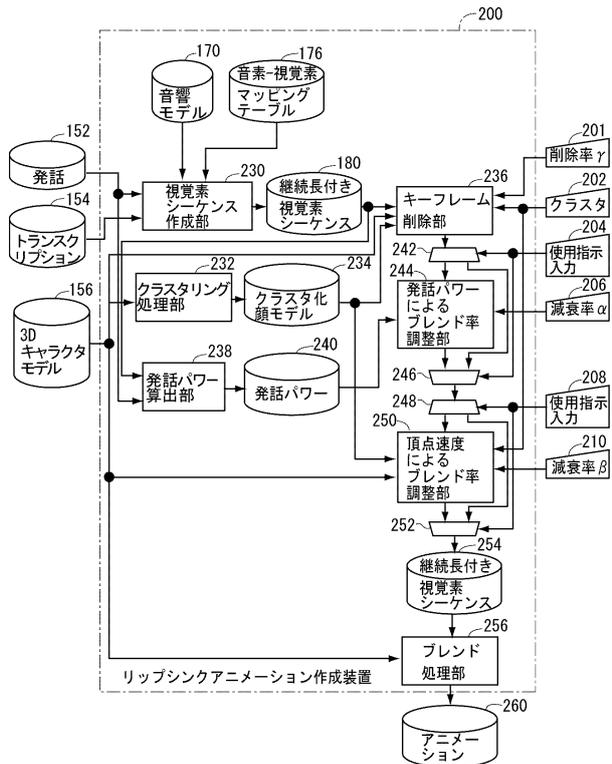
【図3】



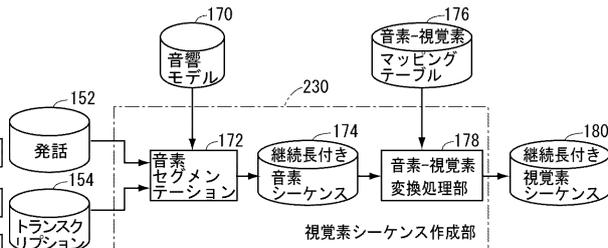
【図4】



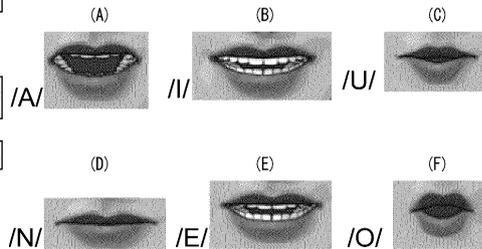
【図5】



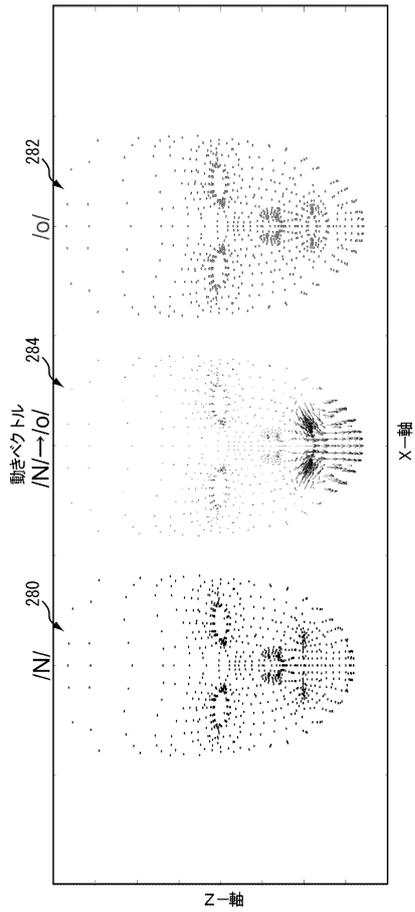
【図6】



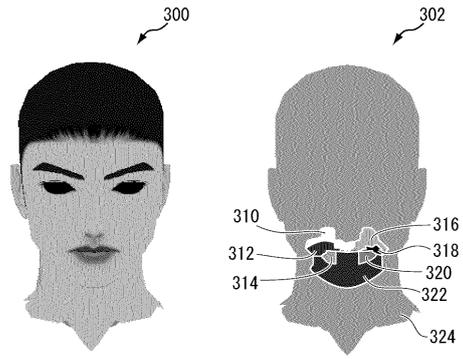
【図7】



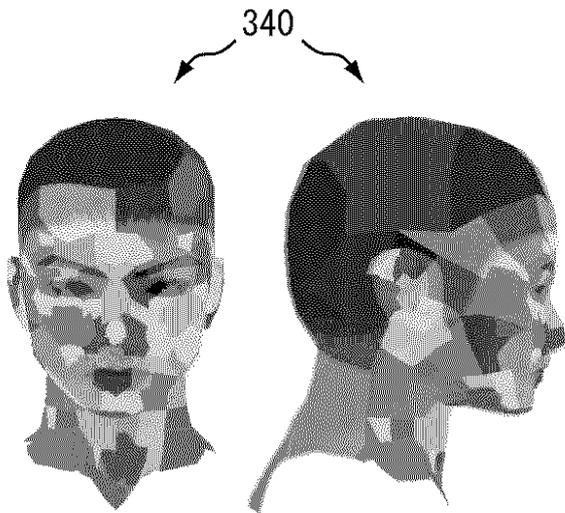
【図 8】



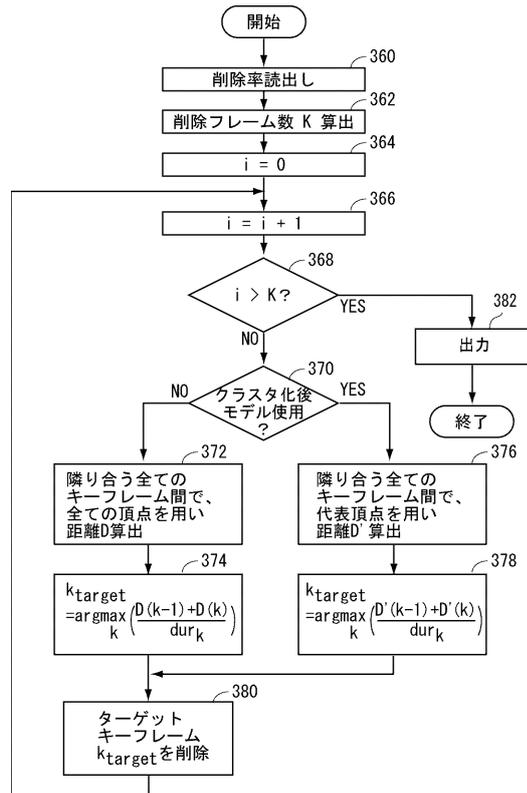
【図 9】



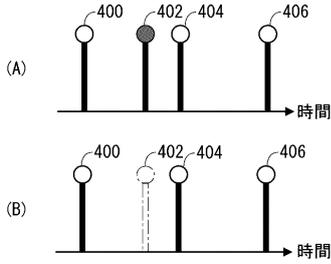
【図 10】



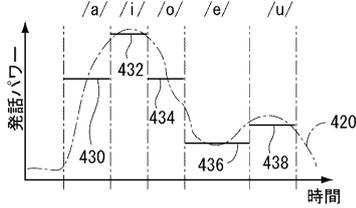
【図 11】



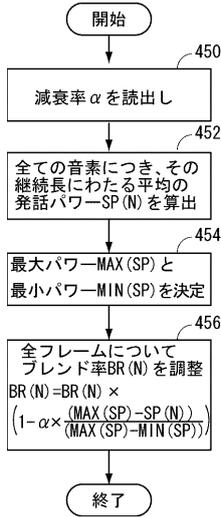
【図12】



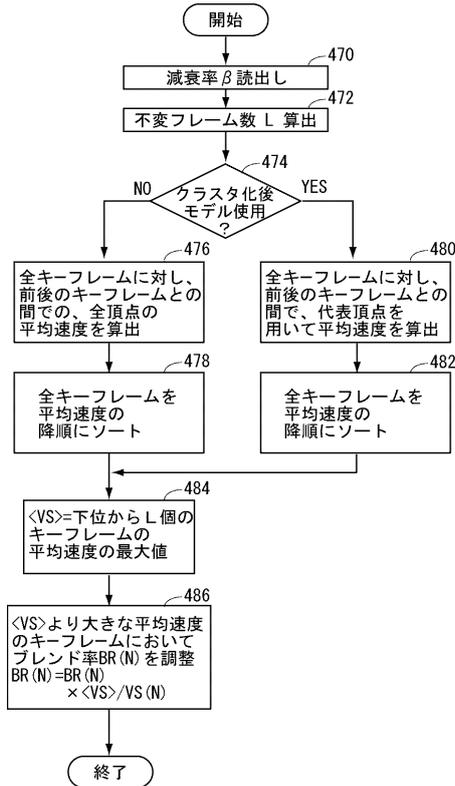
【図13】



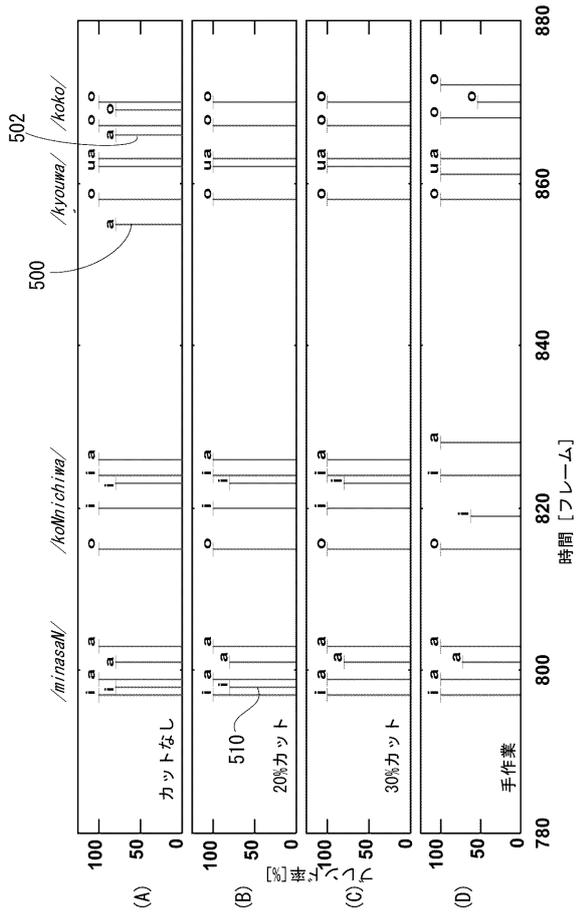
【図14】



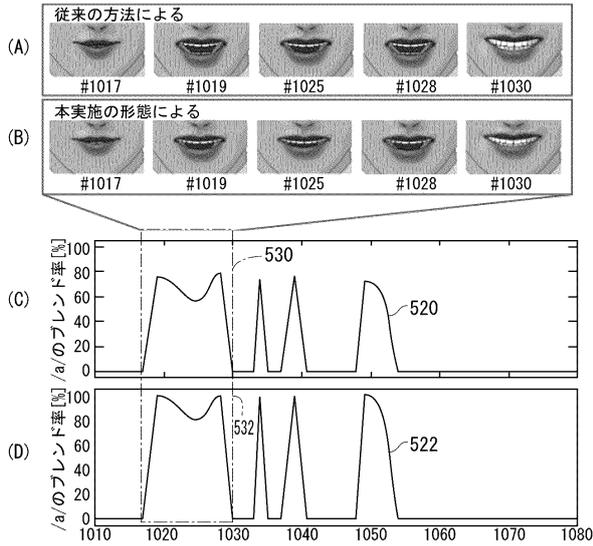
【図15】



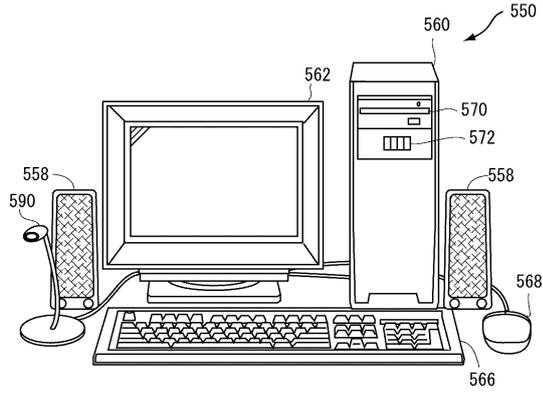
【図16】



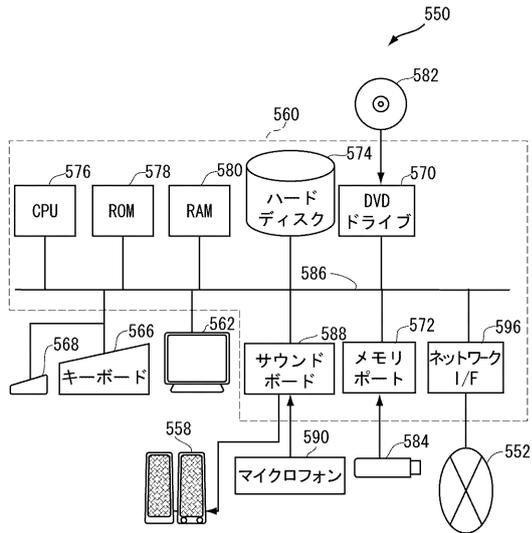
【図17】



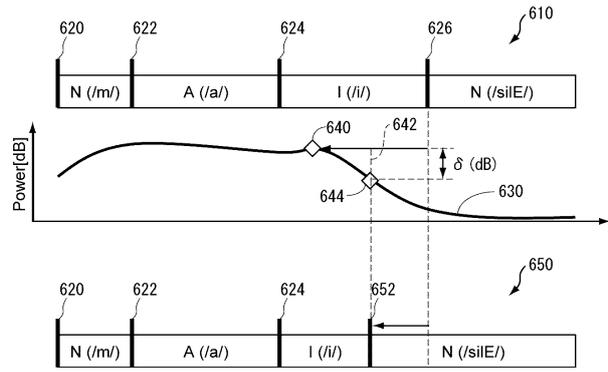
【図18】



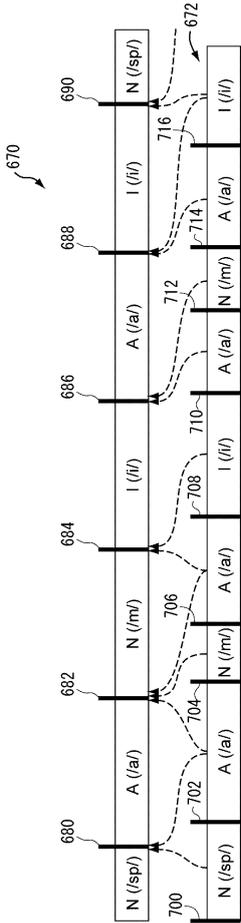
【図19】



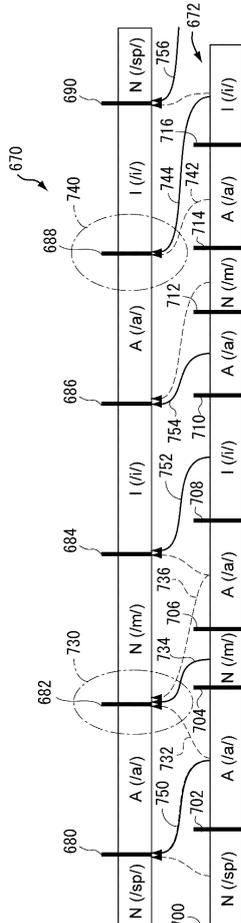
【図20】



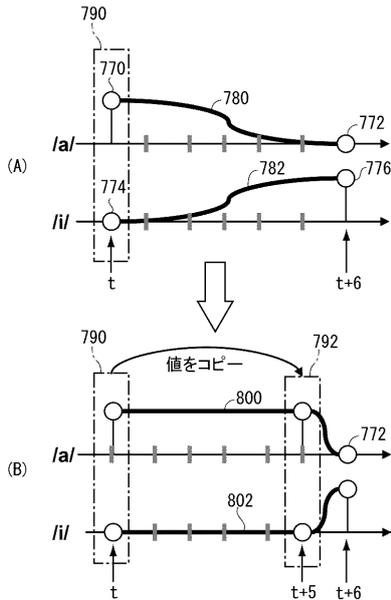
【図 2 1】



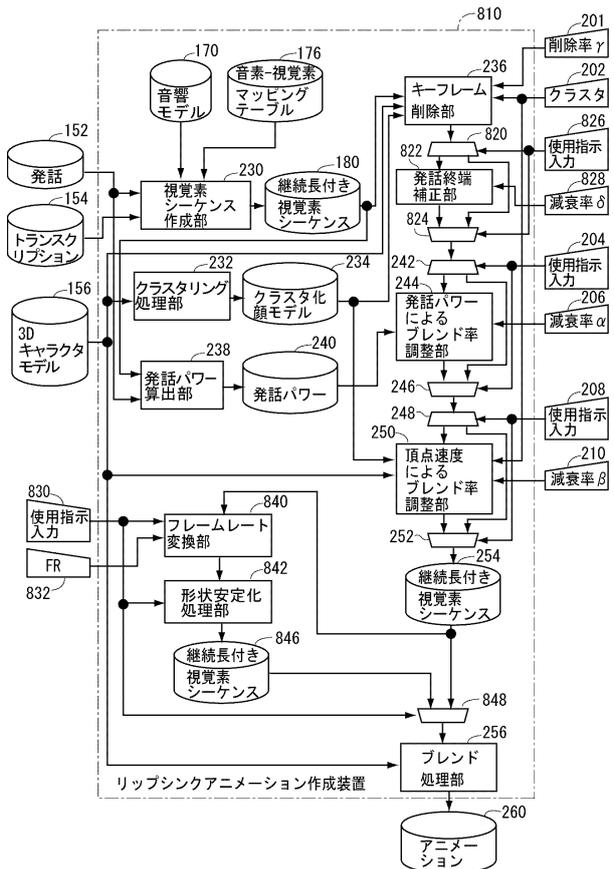
【図 2 2】



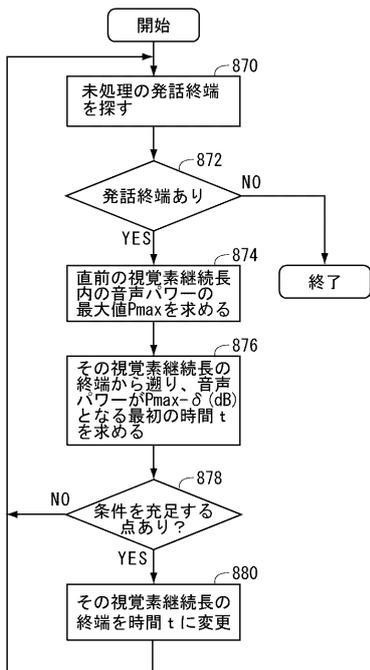
【図 2 3】



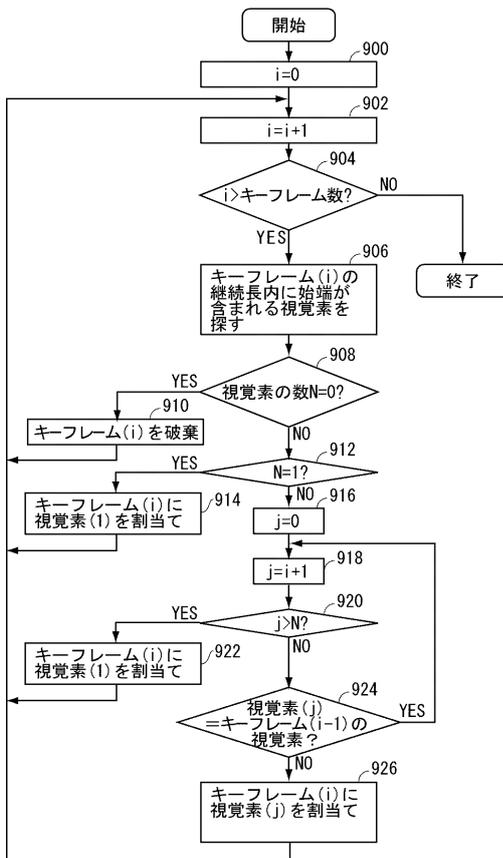
【図 2 4】



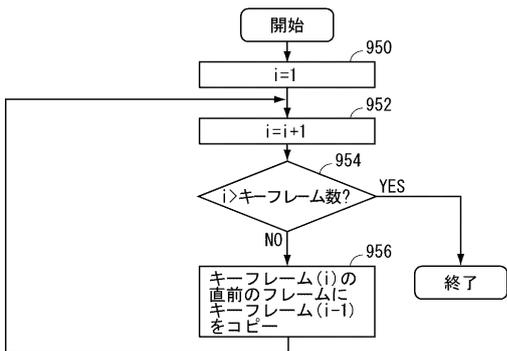
【図25】



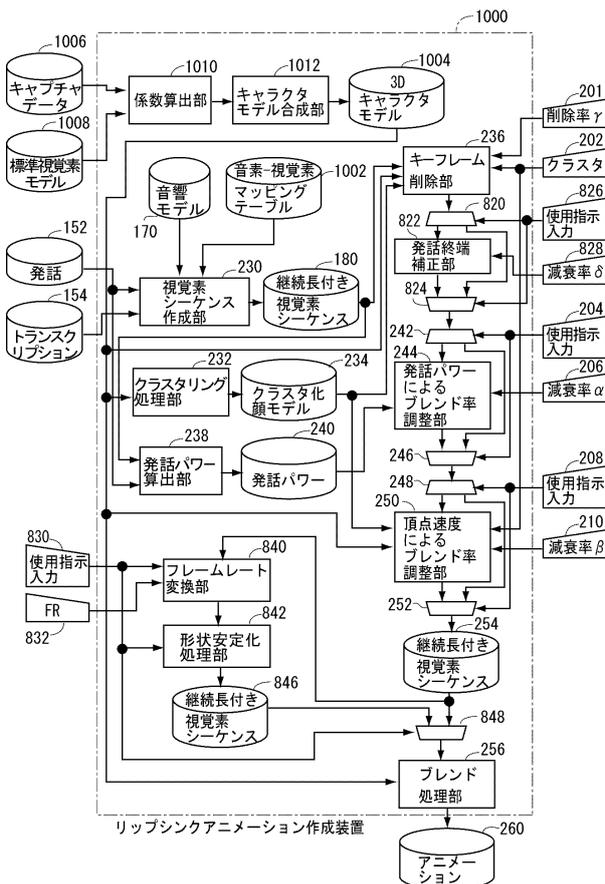
【図26】



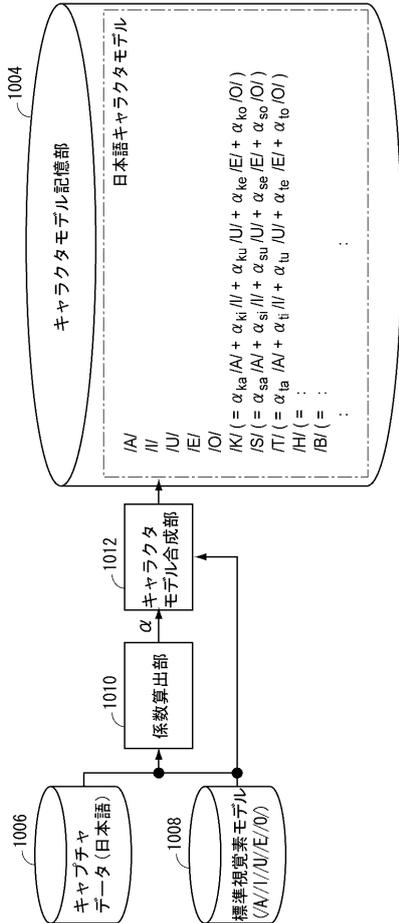
【図27】



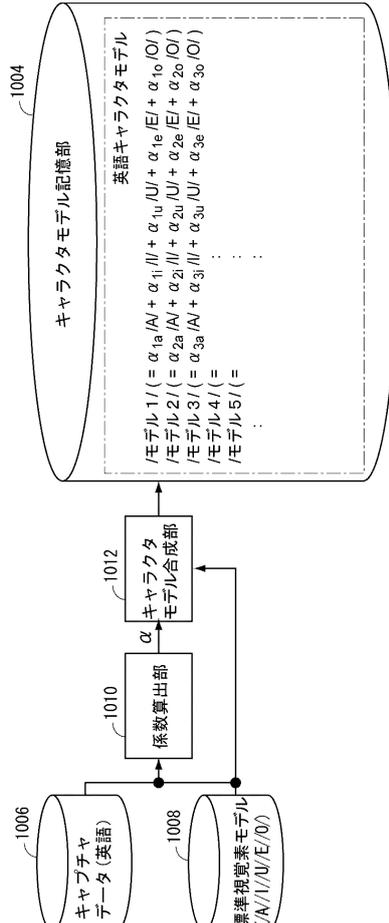
【図28】



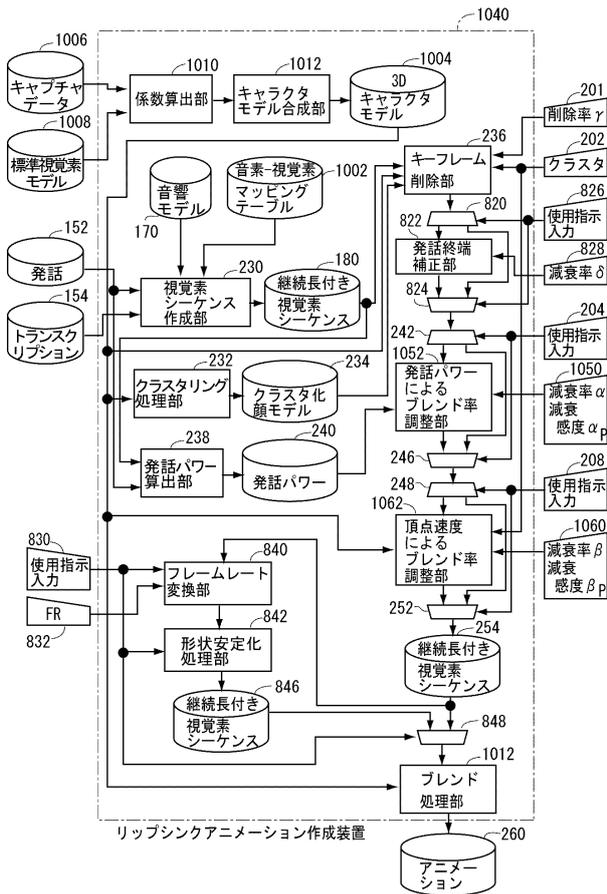
【図29】



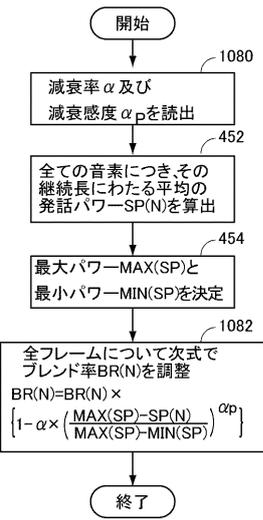
【図30】



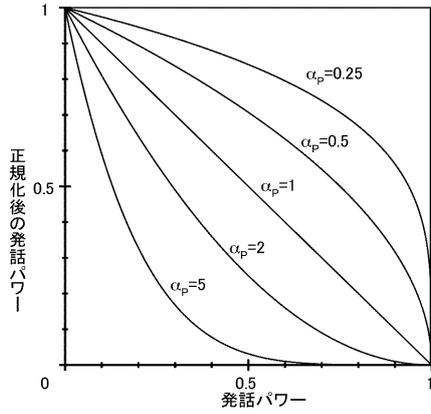
【図31】



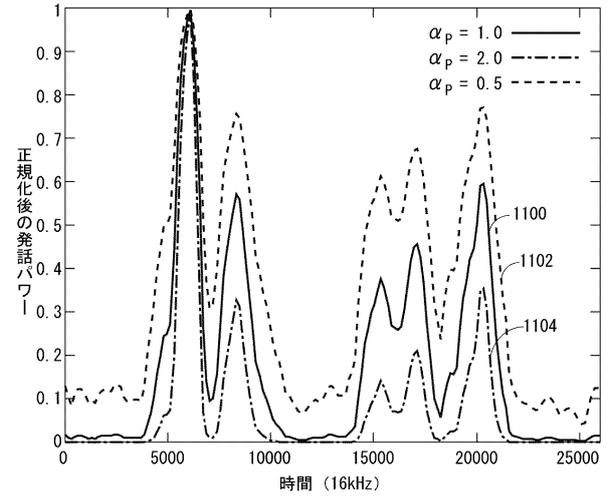
【図32】



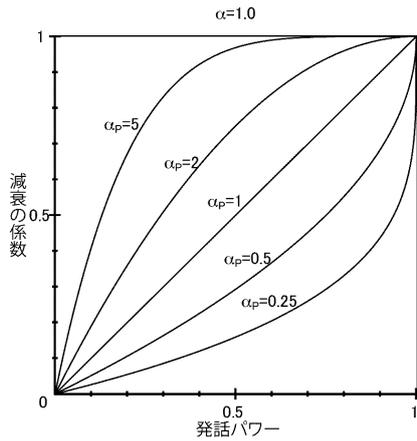
【図 3 3】



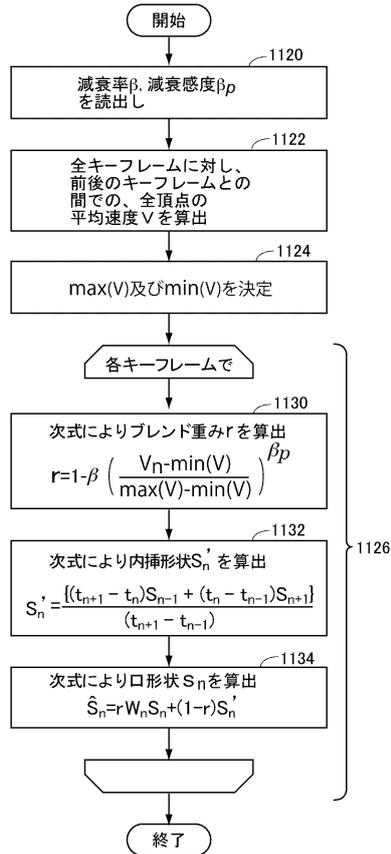
【図 3 4】



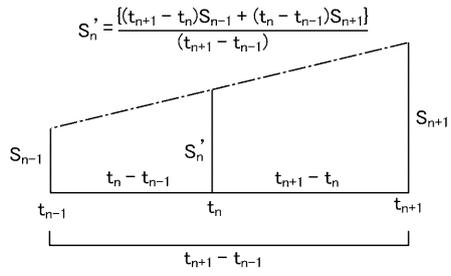
【図 3 5】



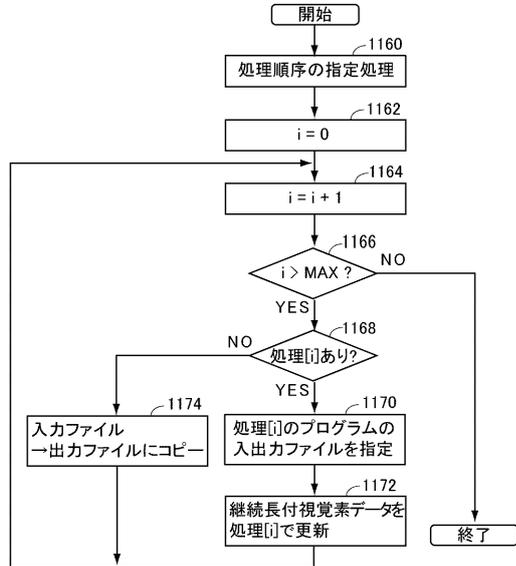
【図 3 6】



【図37】



【図38】



フロントページの続き

審査官 真木 健彦

(56)参考文献 特開2007-058846(JP,A)

特開平07-044727(JP,A)

特開2003-281567(JP,A)

川本真一 四倉達夫 中村哲, 電子情報通信学会2007年総合大会論文集 基礎・境界, 電子情報通信学会2007年総合大会, 日本, 社団法人電子情報通信学会, 2007年 3月 7日, A-16-18

前島謙宜 四倉達夫 森島繁生 中村哲, 雑環境下での音声の聞き取り実験による合成発話顔アニメーションの評価, 電子情報通信学会論誌, 日本, 社団法人電子情報通信学会, 2005年 1月 1日, Vol.J88-A No.1, P.71-82

(58)調査した分野(Int.Cl., DB名)

G06T 13/40