(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第4379616号
(P4379616)

(45)発行日　平成21年12月9日(2009.12.9)　(24)登録日　平成21年10月2日(2009.10.2)

(51)Int.Cl.		F I		
G06T 17/40	(2006.01)	G06T	17/40	A
G06T 1/00	(2006.01)	G06T	1/00	340A

請求項の数 8　(全 32 頁)

(21)出願番号　特願2005-56592(P2005-56592)
(22)出願日　平成17年3月1日(2005.3.1)
(65)公開番号　特開2006-243975(P2006-243975A)
(43)公開日　平成18年9月14日(2006.9.14)
審査請求日　平成18年12月21日(2006.12.21)

特許法第３０条第１項適用　２００４年９月１０日　社団法人情報処理学会発行の「情報処理学会研究報告　情処研報Ｖｏｌ．２００４　Ｎｏ．９０」に発表

（出願人による申告）平成１６年度独立行政法人情報通信研究機構、研究テーマ「大規模コーパスベース音声対話翻訳技術の研究開発」に関する委託研究、産業活力再生特別措置法第３０条の適用を受ける特許出願

特許権者において、実施許諾の用意がある。

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

最終頁に続く

(54)【発明の名称】モーションキャプチャデータ補正装置、マルチモーダルコーパス作成システム、画像合成装置、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】

発話時の発話者の動画像から得られたモーションキャプチャデータを補正するためのモーションキャプチャデータ補正装置であって、前記モーションキャプチャデータは、複数フレームを含み、前記複数フレームの各々は、当該フレーム撮影時における前記発話者の頭部の複数個の特徴点の位置データを含み、前記複数個の特徴点は、前記発話者の首部より上であってかつ前記発話者の表情変化の影響を受けない所定箇所に配置された第１の種類の特徴点と、その他の特徴点とを含み、

前記複数フレームの各々に対する前記複数個の特徴点の位置データから、前記第１の種類の特徴点の位置データを選択するための選択手段と、

前記複数フレームの各々に対し、前記選択手段により選択された位置データを基準として、前記複数個の特徴点の各々の位置データを補正するための補正手段とを含み、

前記第１の種類の特徴点は、前記発話者の頭部であってかつ前記発話者の表情変化の影響を受けない所定箇所に配置された第１の基準特徴点と、前記発話者の首部であってかつ前記発話者の表情変化の影響を受けない所定箇所に配置された第２の基準特徴点とを含み、

前記選択手段は、

前記複数フレームの各々に対し、前記複数個の特徴点の位置データを、前記発話者の頭部の特徴点の位置データと、前記発話者の首部の特徴点の位置データとに分類するための分類手段と、

前記分類手段により分類された前記発話者の頭部の特徴点の位置データから、前記第１の基準特徴点のデータを選択するための頭部基準特徴点選択手段と、
前記頭部基準特徴点選択手段により選択された前記第１の基準特徴点のデータを基準に、同じフレームの前記頭部の特徴点の位置データを補正するための頭部補正式を算出するための頭部補正式算出手段と、
前記複数フレームの各々に対し、前記頭部の特徴点の位置データに前記頭部補正式算出手段により算出された頭部補正式を適用して補正するための頭部補正手段と、
前記分類手段により分類された前記発話者の首部の特徴点の位置データから、前記第２の基準特徴点のデータを選択するための首部基準特徴点選択手段と、
前記首部基準特徴点選択手段により選択された前記第２の基準特徴点のデータを基準に、同じフレームの前記首部の特徴点の位置データを補正するための首部補正式を算出するための首部補正式算出手段と、
前記複数フレームの各々に対し、前記首部の特徴点の位置データに前記首部補正式算出手段により算出された首部補正式を適用して補正するための首部補正手段とを含む、モーションキャプチャデータ補正装置。

【請求項２】
前記第１の種類の特徴点は、前記発話者の額領域、こめかみ領域、及び鼻の先端領域のいずれかに配置される、請求項１に記載のモーションキャプチャデータ補正装置。

【請求項３】
前記補正手段は、
前記複数フレームの各々に対し、前記選択手段により選択された前記第１の種類の特徴点の位置データを基準に、同じフレームの前記複数個の特徴点の位置データを補正するための補正式を算出するための補正式算出手段と、
前記複数フレームの各々に対し、前記複数個の特徴点の位置データに前記補正式算出手段により算出された補正式を適用して補正するための補正式適用手段とを含む、請求項１又は請求項２のいずれかに記載のモーションキャプチャデータ補正装置。

【請求項４】
発話時の発話者の顔画像を含む動画像データと、当該発話時の音声の録音データと、発話時における前記発話者の顔の予め定める複数の特徴点に関するモーションキャプチャデータとを発話ごとにそれぞれ分離し、互いに対応付けて保存するための発話分離手段と、
前記発話分離手段により分離された各発話の前記モーションキャプチャデータを補正するための、請求項１～請求項３のいずれかに記載のモーションキャプチャデータ補正装置とを含む、マルチモーダルコーパス作成システム。

【請求項５】
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３のいずれかに記載のモーションキャプチャデータ補正装置として動作させる、コンピュータプログラム。

【請求項６】
顔オブジェクトの形状を第１の座標空間における複数のノードの座標値を用いて定義した形状モデルと、所定の発話を行なっている発話者の顔画像から得られた、前記発話者の頭部の複数の特徴点の所定の第２の座標系における位置情報とを基に、前記所定の発話を行なう前記顔オブジェクトの表情を表す画像を合成するための画像合成装置であって、
前記発話者の頭部の複数の特徴点は、請求項１に記載のモーションキャプチャデータ補正装置により、各々の位置データが補正され、
前記複数の特徴点と、前記形状モデル内の任意の点との対応関係を定義することにより、前記形状モデル内に前記複数の特徴点にそれぞれ対応する複数の仮想特徴点を設定するための仮想特徴点設定手段と、
前記複数のノードの各々に対し、前記複数の仮想特徴点のうちで、当該ノードからの距離が小さいものから順番に、かつ当該ノードと仮想特徴点とを結ぶ線分が前記形状モデルに対し所定の制約条件を充足するものを所定個数だけ選定するための仮想特徴点選定手段

と、
前記複数のノードの各々に対し、前記仮想特徴点選定手段により選定された所定個数の仮想特徴点の位置情報の間の内挿により算出される座標値を割当てることにより前記形状モデルを変形させるための形状モデル変形手段と、
前記形状モデル変形手段により得られた形状モデルに基づいて前記顔オブジェクトの画像を生成するための画像生成手段とを含む、画像合成装置。
【請求項７】
前記仮想特徴点選定手段は、前記複数のノードの各々に対し、前記複数の仮想特徴点のうちで、当該ノードからの距離が小さいものから順番に、かつ当該ノードと仮想特徴点とを結ぶ線分が前記形状モデルの境界エッジを横切らないものを所定個数だけ選定するための手段を含む、請求項６に記載の画像合成装置。
【請求項８】
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６又は請求項７に記載の画像合成装置として動作させる、コンピュータプログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、音声言語処理技術に関し、特に発話時における音声及び表情変化に関する情報を含むマルチモーダルコーパスを作成するためのマルチモーダルコーパス作成装置及びシステム、並びに発話時の表情変化を表現するアニメーションを作成するための画像合成装置に関する。
【背景技術】
【０００２】
人間にとって容易かつ自然なヒューマンマシンインタフェースを実現するための技術が研究されている。例えば、音声認識及び音声合成等の音声言語処理技術においては、大規模な音声コーパスと統計的な手法とにより、高性能の処理が実現されるようになっている。さらには、音声だけでなく視聴覚情報を用いるヒューマンマシンインタフェースを実現するための技術が盛んに研究されている。そのひとつに、音声合成技術を発展させて、発話時の顔画像を合成する技術がある。
【０００３】
後掲の非特許文献１には、仮想空間上のメッシュで構成された顔の形状モデルを用いて、発話時の顔の表情変化を表現する技術が開示されている。この技術では、メッシュの各点の動きを推定し、推定した動きを基に顔のメッシュ形状を変形させる。この手法では、顔の形状モデル及びその表情変化のバリエーションに制限がなく、表情変化の豊かな顔画像を自在に合成することができる。
【０００４】
後掲の非特許文献２には、統計確率的な手法によって発話中における顔の動画像を合成する技術が開示されている。この手法では、予め発話時の顔の画像をデータベース化しておく。そして、発話内容に適した特徴を備える画像をデータベース中の顔の画像から選び再構成する。この手法で合成される画像は、撮影された画像を再構成したものであるため、大規模かつ適切なデータベースを用意すれば、自然な顔画像合成を行なうことができる。
【０００５】
また、視聴覚情報を用いるヒューマンマシンインタフェースを実現するために、音声言語処理技術における音声コーパスに相当するデータベースの整備が進められている。音声と顔の画像情報とを、言語情報に対応付けたマルチモーダルコーパスの整備が進められている。後掲の非特許文献３には、マルチモーダルコーパスを構築する種々の試みが紹介されている。
【０００６】
マルチモーダルコーパスに収められた画像の特徴量を抽出し分析することにより、当該

画像情報に基づく顔画像の合成等が可能となる。非特許文献１に記載の技術では、発話時の顔を撮影した動画像におけるピクセル情報からオプティカルフローを求めることにより、発話時における顔の各部位の動きを推定し、画像の特徴量として用いている。また、後掲の非特許文献４には、唇領域の画像におけるピクセル情報をもとに、唇の変化量を求める技術が開示されている。

【０００７】

【非特許文献１】モリシマ，Ｓ．、イワサワ，Ｓ．、サカグチ，Ｔ．、カワカミ，Ｆ．、アンドウ，Ｍ．、「より良い顔のコミュニケーション」、ＡＣＭシーグラフ'９５、インタラクティブ　コミュニティ　ビジュアル予稿集、１１７頁、１９９５年（Morishima, S., Iwasawa, S., Sakaguchi, T., Kawakami, F., and Ando, M., “Better Face Communication”, Visual Proceedings of ACM SIGGRAPH’95, Interactive Communities, p.117, 1995）

【非特許文献２】エザット，Ｔ．、ガイガー，Ｇ．、ポッジョ，Ｔ．「学習可能なビデオリアリスティック発話アニメーション」　ＡＣＭ　シーグラフ２００２予稿集、２００２年（Ezzat, T., Geiger,G. and Poggio, T. “Trainable Videorealistic Speech Animation”, Proceedings of ACM SIGGRAPH 2002）.

【非特許文献３】ナカムラ，Ｓ．、「最近のマルチモーダルコーパス活動の概要」、ＣＯＣＯＳＤＡワークショップ　２０００（Nakamura, S., “Overview on Recent Activities in Multi-Modal Corpora”, COCOSDA Workshop, 2000）

【非特許文献４】タムラ，Ｔ．、コンドウ，Ｓ．、マスコ，Ｔ．、コバヤシ，Ｔ．、「ＨＭＭからのパラメータ生成に基づくテキスト－発話音声画像合成」　ＥＵＲＯＳＰＥＥＣＨ'９９予稿集、９５９－９６２頁、１９９９年（Tamura, T., Kondo, S., Masuko, T., and Kobayashi, T., “Text-to-Audio-Visual Speech Synthesis Based on Parameter Generation from HMM”, Proceeding of EUROSPEECH, pp.959-962, 1999）

【発明の開示】

【発明が解決しようとする課題】

【０００８】

非特許文献２及び４のいずれに記載の技術においても、発話時の表情の特徴量を画像情報から得ている。しかし、この場合、次のような問題点が発生する。すなわち、顔及びその表情は立体的であるのに対し、動画像は２次元の情報である。そのため、３次元での形状変化に関する特徴量を得るのは困難である。例えば、発話中には表情を形成する顔の器官だけではなく、頭部及び首部も自由に移動回転する。顔の各器官の動画像上での位置及び形状は、頭部の動きに応じて表情とは無関係に変化する。よって、動画像から顔の器官の変化についての情報を得るのは困難である。また、画像情報はその画像を撮影するためのカメラの性能に依存する。したがって、画像情報から求める特徴量に誤差が生じる恐れがあるという問題も発生する。

【０００９】

また、非特許文献１に記載の手法をはじめとする、モデルベースの顔画像の合成方法で発話時の顔の画像を作成するには、アニメーションの各フレームにおいて、モデルのメッシュの位置を定義する必要がある。現在のアニメーションに用いられる形状モデルは、膨大な数のメッシュから構成されている。形状モデルの変形によってアニメーションを生成するには、膨大な量のメッシュについて位置をいちいち定義しなければならず、膨大な作業を要する。非特許文献２に記載の手法をはじめとする動画像の再生成手法では、作成可能な顔の表情は、顔画像のコーパスに格納されている画像により限定されてしまう。多様な容貌の顔及び多彩な表情で発話時の表情変化を表現するには、その分膨大な量の顔の画像をコーパス化する必要がある。

【００１０】

それゆえに、本発明の目的は、顔の表情を形成する各器官の動きについての正確な位置データを得ることができるモーションキャプチャデータ補正装置及びマルチモーダルコー

パス作成システムを提供することである。

【００１１】

本発明の別の目的は、多様な表情を持つ顔画像、または実際の発話者の表情を再現する顔画像のアニメーションを高精度かつ容易に合成することを可能にする画像作成装置を提供することである。

【課題を解決するための手段】

【００１２】

本発明の第１の局面に係るモーションキャプチャデータ補正装置は、発話時の発話者の動画像から得られたモーションキャプチャデータを補正するためのモーションキャプチャデータ補正装置である。モーションキャプチャデータは、複数フレームを含む。複数フレームの各々は、当該フレーム撮影時における発話者の頭部の複数個の特徴点の位置データを含む。複数個の特徴点は、発話者の首部より上であってかつ発話者の表情変化の影響を受けない所定箇所に配置された第１の種類の特徴点と、その他の特徴点とを含む。このモーションキャプチャデータ補正装置は、複数フレームの各々に対する複数個の特徴点の位置データから、第１の種類の特徴点の位置データを選択するための選択手段と、複数フレームの各々に対し、選択手段により選択された位置データを基準として、複数個の特徴点の各々の位置データを補正するための補正手段とを含む。

【００１３】

発話者の表情変化の影響を受けない所定箇所に配置された第１の種類の特徴点の位置データを基準として、発話者の顔の特徴点の位置データが補正される。一般的に発話者の頭部のモーションキャプチャデータには、頭部全体の動き、又は首部の動きによる影響が含まれる。第１の種類の特徴点の位置データは、頭部全体の動き、又は首部の動きのみによる影響を受けると考えられ、これらを基準として特徴点の位置データを補正することにより、表情変化のみに起因する特徴点の位置の変化が得られる。

【００１４】

好ましくは、第１の種類の特徴点は、発話者の額領域、こめかみ領域、及び鼻の先端領域のいずれかに配置される。

【００１５】

これら領域は、いずれも表情変化による影響を受けないか、きわめて少ない。したがってこれらの領域に配置された特徴点を基準に特徴点の位置データを補正することで、それら特徴点の、表情変化に起因する位置変化のみが正確に得られる。

【００１６】

より好ましくは、補正手段は、複数フレームの各々に対し、選択手段により選択された第１の種類の特徴点の位置データを基準に、同じフレームの複数個の特徴点の位置データを補正するための補正式を算出するための補正式算出手段と、複数フレームの各々に対し、複数個の特徴点の位置データに補正式算出手段により算出された補正式を適用して補正するための補正式適用手段とを含む。

【００１７】

第１の種類の特徴点の位置データを基準に補正式が算出され、この補正式を各特徴点の位置データに適用する。定型化した処理により、対象が別の発話者になっても新たに補正式を算出でき、安定して位置データの補正を行なうことができる。

【００１８】

さらに好ましくは、補正式算出手段は、複数フレームの各々に対して、第１の種類の特徴点の位置データに対する特異値分解により、同一フレーム内の複数個の特徴データを変換するためのアフィン変換行列を算出するための手段を含む。

【００１９】

特異値分解により座標変換のためのアフィン変換行列が得られる。その結果、簡単な行列演算で位置データの補正を行なうことができる。

【００２０】

好ましくは、第１の種類の特徴点は、発話者の頭部であってかつ発話者の表情変化の影

響を受けない所定箇所に配置された第１の基準特徴点と、発話者の首部であってかつ発話者の表情変化の影響を受けない所定箇所に配置された第２の基準特徴点とを含み、選択手段は、複数フレームの各々に対し、複数個の特徴点の位置データを、発話者の頭部の特徴点の位置データと、発話者の首部の特徴点の位置データとに分類するための分類手段と、分類手段により分類された発話者の頭部の特徴点の位置データから、第１の基準特徴点のデータを選択するための頭部基準特徴点選択手段と、頭部基準特徴点選択手段により選択された第１の基準特徴点のデータを基準に、同じフレームの頭部の特徴点の位置データを補正するための頭部補正式を算出するための頭部補正式算出手段と、複数フレームの各々に対し、頭部の特徴点の位置データに頭部補正式算出手段により算出された頭部補正式を適用して補正するための頭部補正手段と、分類手段により分類された発話者の首部の特徴点の位置データから、第２の基準特徴点のデータを選択するための首部基準特徴点選択手段と、首部基準特徴点選択手段により選択された第２の基準特徴点のデータを基準に、同じフレームの首部の特徴点の位置データを補正するための首部補正式を算出するための首部補正式算出手段と、複数フレームの各々に対し、首部の特徴点の位置データに首部補正式算出手段により算出された首部補正式を適用して補正するための首部補正手段とを含む。

【００２１】

首部の特徴点は、頭部とは別に首部の動きによる影響を受ける。したがって、頭部とは別に首部に対しても基準となる特徴点を定め、それらに基づいて首部の特徴点の位置データを補正する。こうして、顔面を含む頭部の特徴点と、首部の特徴点との各々について、表情の変化のみに起因する位置変化を算出することができる。

【００２２】

本発明の第２の局面に係るマルチモーダルコーパス作成システムは、発話時の発話者の顔画像を含む動画像データと、当該発話時の音声の録音データと、発話時における発話者の顔の予め定める複数の特徴点に関するモーションキャプチャデータとを発話ごとにそれぞれ分離し、互いに対応付けて保存するための発話分離手段と、発話分離手段により分離された各発話のモーションキャプチャデータを補正するための、上記したいずれかのモーションキャプチャデータ補正装置とを含む。

【００２３】

このマルチモーダルコーパス作成システムによれば、発話ごとに、発話者の顔画像の動画像データと、音声の録音データと、発話者の顔の特徴点のモーションキャプチャデータが得られる。そのモーションキャプチャデータをモーションキャプチャデータ補正装置を用いて補正することにより、発話者の顔の特徴点の、発話による表情変化のみに起因する位置変化が得られる。その結果、発話に起因するこの発話者の顔の特徴点の位置変化が正確に表され、発話と表情との間の関係を研究するための正確な基礎データが得られる。

【００２４】

本発明の第３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モーションキャプチャデータ補正装置として動作させる。したがってこのコンピュータプログラムにより、第１の局面に係るモーションキャプチャデータ補正装置と同様の効果を得ることができる。

【００２５】

本発明の第４の局面に係る画像合成装置は、顔オブジェクトの形状を第１の座標空間における複数のノードの座標値を用いて定義した形状モデルと、所定の発話を行なっている発話者の顔画像から得られた、発話者の頭部の複数の特徴点の所定の第２の座標系における位置情報とを基に、所定の発話を行なう顔オブジェクトの表情を表す画像を合成するための画像合成装置であって、複数の特徴点と、形状モデル内の任意の点との対応関係を定義することにより、形状モデル内に複数の特徴点にそれぞれ対応する複数の仮想特徴点を設定するための仮想特徴点設定手段と、複数のノードの各々に対し、複数の仮想特徴点のうちで、当該ノードからの距離が小さいものから順番に、かつ当該ノードと仮想特徴点とを結ぶ線分が形状モデルに対し所定の制約条件を充足するものを所定個数だけ選定するた

めの仮想特徴点選定手段と、複数のノードの各々に対し、仮想特徴点選定手段により選定された所定個数の仮想特徴点の位置情報の間の内挿により算出される座標値を割当てることにより形状モデルを変形させるための形状モデル変形手段と、形状モデル変形手段により得られた形状モデルに基づいて顔オブジェクトの画像を生成するための画像生成手段とを含む。

【００２６】

顔オブジェクトの形状モデルに、顔オブジェクトの特徴点と対応する仮想特徴点が設定され、さらに形状モデルを構成する各ノードと、当該ノードとの距離が近く、かつ所定の制約条件を充足する所定個数の仮想特徴点とが対応付けられる。各ノードに、それらに対応付けられた仮想特徴点の位置情報の間の内挿により得られた座標値を割当てることにより、各ノードに割当てられた座標値はもとの発話者の顔においてそのノードに対応する点の位置とほぼ正確に一致する。その結果、こうして得られた座標値を用いて顔オブジェクトの形状を変化させることで、元の発話者の表情変化を顔オブジェクトにより再現できる。

【００２７】

好ましくは、仮想特徴点選定手段は、複数のノードの各々に対し、複数の仮想特徴点のうちで、当該ノードからの距離が小さいものから順番に、かつ当該ノードと仮想特徴点とを結ぶ線分が形状モデルの境界エッジを横切らないものを所定個数だけ選定するための手段を含む。

【００２８】

一般に顔には、目、口、鼻の穴等、顔面を構成しない切れ目があり、形状モデルでは、それらと顔面との間は境界エッジで仕切られている。こうした切れ目を挟んだ両側のノードは互いに別の動きをするため、それらの座標位置を互いに関連付けて計算するのは不適当である。そこで、このように計算対象のノードと仮想特徴点とを結ぶ線分が境界エッジを横切るような仮想特徴点はノードの座標値の計算からは除外する。こうすることで、各ノードの座標値をより正確に、かつ実際の顔と同様に適切な表情が得られるように算出できる。

【００２９】

より好ましくは、複数フレームの位置情報を元に、仮想特徴点設定手段、仮想特徴点選定手段、形状モデル変形手段、及び画像生成手段により生成された顔オブジェクトの画像を各フレームとして時系列的に保存することにより、所定の発話を行なう顔オブジェクトの表情を表す動画を生成するための手段をさらに含む。

【００３０】

フレームごとに顔画像を作成し、それらを時系列的に保存することにより、発話時の発話者の顔の表情と同様の表情変化を持つ動画を生成できる。

【００３１】

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画像合成装置として動作させる。

【００３２】

このコンピュータプログラムによれば、上記した第４の局面に係る画像合成装置と同様の効果を得ることができる。

【発明を実施するための最良の形態】

【００３３】

以下、図面を参照しつつ、本発明の一実施の形態について説明する。なお、以下の説明に用いる図面では、同一の部品に同一の符号を付してある。それらの名称及び機能も同一である。したがって、それらについての説明は繰返さない。

【００３４】

［概要］

本実施の形態では、音声及び顔の動画像に加えて、発話時の表情に関するデータを含むマルチモーダルコーパスを作成する。本実施の形態では、音声及び動画像の収録時に、顔

の多数の部位について位置計測を併せて行なう。さらに当該位置の計測データから顔の各器官の変化を表すデータを取得し、表情に関する特徴量データとする。そして、当該顔器官の変化を表すデータ（以下、「顔器官変化量データ」と呼ぶ）と音声及び動画像のデータとを対応付けてデータベース化することにより、マルチモーダルコーパスを作成する。本実施の形態ではさらに、発話時の表情変化を表現するアニメーションを、マルチモーダルコーパスをもとに作成する。この際、顔の形状モデルに顔器官の変化を順次割り当てる。

【００３５】

［図１　システム全体の構成］

図１に、本実施の形態に係るマルチモーダルコーパス作成システム１００全体の構成を示す。図１を参照して、このマルチモーダルコーパス作成システム１００は、発話者１０２の音声及び顔の動画像を収録すると同時に、発話者１０２の顔の各部位について位置計測を行なうための収録システム１０４と、収録システム１０４による位置の計測結果を基に顔器官変化量データを生成し、収録システム１０４による収録で得られる発話時の音声のデータ及び動画像のデータ、並びに当該顔器官変化量データを発話内容と対応付けることによりマルチモーダルコーパス１０６を作成するためのマルチモーダルコーパス作成装置１０８とを含む。

【００３６】

このマルチモーダルコーパス作成システム１００はさらに、静止状態における所定の顔の形状を表す初期顔モデル１１０を記憶するための記憶装置と、入力テキストを受け、マルチモーダルコーパス１０６内の顔器官変化量データを基に、入力テキストを発話中の各時刻における顔の形状モデルを作成し動画像化することにより、入力テキスト発話時の顔の表情変化を表現するアニメーション１１２を作成するためのアニメーション作成装置１１４とを含む。

【００３７】

マルチモーダルコーパス作成システム１００はさらに、マルチモーダルコーパス作成時のユーザの操作を受け、対応する操作信号をマルチモーダルコーパス作成装置１０８に与えるための入力装置１１６Ａと、マルチモーダルコーパス作成装置１０８から、操作に用いる情報を受けて出力するための出力装置１１８Ａと、アニメーション作成時にユーザの操作を受け、対応する操作信号をアニメーション作成装置１１４に与えるための入力装置１１６Ｂと、アニメーション作成装置１１４からの出力される情報を画像及び音声等に変換して出力するための出力装置１１８Ｂとを含む。

【００３８】

初期顔モデル１１０は、静止状態における所定の顔の形状を多数の多角形（ポリゴン）によって表現した形状モデルである。図９に、初期顔モデル１１０の一例を示す。図９を参照して、この初期顔モデル１１０は、発話者１０２の顔の静止画像と所定のワイヤフレームモデルとを整合させることにより準備された形状モデルである。この顔モデルは、約７５０のポリゴンで構成されている。アニメーション作成装置１１４は、顔器官変化量データを基に、発話中における顔の各器官の変化を、初期顔モデル１１０におけるポリゴンの頂点（ノード）の各々に割当てて発話中の所定の顔の形状モデルを形成する機能を持つ。

【００３９】

［収録システム１０４の構成］

収録システム１０４は、発話時における発話者１０２の顔の各部位の位置及びその軌跡を計測しキャプチャデータとして出力するためのモーションキャプチャシステム１２０と、発話者１０２の音声を収録するための録音システム１２２と、発話時における発話者１０２の動画像を撮影するための撮影システム１２４と、発話者に発話すべき内容として提示される所定の文章、単語、文字、及び音節の記号等で構成された発話内容を格納する発話リスト１２６と、発話リスト１２６の発話内容のいずれかを発話者１０２に提示するためのテレプロンプタ１２８と、モーションキャプチャシステム１２０及び撮影システム１

２４に対してタイムコードを供給するためのタイムコードジェネレータ１３０とを含む。

【００４０】

本実施の形態に係るモーションキャプチャシステム１２０は、高再帰性光学反射マーカ（以下、単に「マーカ」と呼ぶ。）の反射光を利用して計測対象の位置を計測する光学式のシステムを含む。モーションキャプチャシステム１２０は、発話者１０２の顔面及び首部の予め定める多数の箇所にそれぞれ装着されるマーカからの赤外線反射光の映像を、所定の時間間隔のフレームごとに撮影するための複数の赤外線カメラ１３２Ａ，…，１３２Ｆ（以下これらをまとめて「赤外線カメラ１３２」と呼ぶことがある。）と、赤外線カメラ１３２からの映像信号を基にフレームごとに各マーカの位置を計測し、タイムコードジェネレータ１３０からのタイムコードを付与して出力するためのデータ処理装置１３４とを含む。

【００４１】

［図２　マーカの配置例］

図２（Ａ）及び図２（Ｂ）に、発話者１０２の首部より上へのマーカの装着例を示す。図２（Ａ）は、発話者１０２の顔面及び首部の右半分の所定位置にマーカを装着した状態での、発話者１０２の頭部及び首部の外観を示す右側面図であり、図２（Ｂ）は、同状態での発話者１０２の頭部及び首部の外観を示す正面図である。

【００４２】

図２（Ａ）及び図２（Ｂ）を参照して、発話者１０２の顔面及び首部の皮膚上には、多数のマーカ１７０Ａ，…，１７０Ｍ（以下これらをまとめて「マーカ１７０」と呼ぶことがある。）が、図示しない装着材（接着剤）により装着される。マーカ１７０は、直径３～４ｍｍの半球状又は球状の形状であり、照射光を再帰反射するよう加工されている。

【００４３】

図２（Ａ）及び図２（Ｂ）に示す例では、マーカ１７０は、眉部の９箇所、目の輪郭部の９箇所、鼻部の５箇所、口唇部の１１箇所、頬部の１８箇所、顔の輪郭部の８箇所、顎部の６箇所、首部の８箇所、及び額部の４箇所に装着されている。マルチモーダルコーパス作成においては、発話時の顔部位の詳細な変化量を計測すること、及び複数日にわたり又は複数の発話者１０２について計測を行なうことが想定される。そのため、マーカ１７０はそれぞれ、顔器官の特徴的な位置、又は装着済みのマーカとの相対的な関係によって定められる位置に、予め定めるルールにしたがい装着される。例えば、口唇部のマーカはそれぞれ次の表に示すルールにより定められた装着位置に、定められた装着順序で装着される。なお、こうして定められた装着位置を、本明細書では「特徴点」と呼ぶ。

【００４４】

【表１】

装着順	識別番号	装着位置
1	MTH1C	顔中央線と上唇上部輪郭との交点
2	MTH2C	顔中央線と上唇下部輪郭との交点から2mm～3mmほど［MTH1C］寄りの位置
3	MTH4C	顔中央線と下唇下部輪郭との交点
4	MTH3C	顔中央線と下唇上部輪郭との交点から2mm～3mmほど［MTH4C］寄りの位置
5	MTH1R(L)	右（左）の口角
6	MTH2R(L)	唇輪郭線[MTH1R(L)]-[MTH1C]間を3等分した時の[MTH1R(L)]寄りの点
7	MTH4R(L)	唇輪郭線[MTH1R(L)]-[MTH1C]間を3等分した時の[MTH1C]寄りの点
8	MTH3R(L)	唇輪郭線[MTH1R(L)]-[MTH4C]間を3等分した時の[MTH1R(L)]寄りの点
9	MTH7R(L)	唇輪郭線[MTH1R(L)]-[MTH4C]間を3等分した時の[MTH4C]寄りの点
10	MTH5R(L)	線分[MTH1R(L)]-[MTH2C]の中点から2mm～3mmほど［MTH4R(L)］寄りの位置
11	MTH6R(L)	線分[MTH1R(L)]-[MTH3C]の中点から2mm～3mmほど［MTH7R(L)］寄りの位置

再び図２を参照して、マーカ１７０のうち、額部に装着されるマーカ１７２Ａ，…，１７２Ｄは、各マーカ１７０の位置のデータを頭部の動きに応じて補正するための補正用のデータの計測に用いられるマーカである。図２（Ａ）及び（Ｂ）に示す例では、額部の皮膚の動きを抑制する拘束部材１７４を額部に貼付し、マーカ１７２Ａ，…，１７２Ｄを、

拘束部材１７４を介して間接的に額部に装着している。なお、本実施の形態では、顔全体にマーカを装着する場合、マーカは、合計１３７箇所に装着される。

【００４５】

データ処理装置１３４は、各マーカの位置の計測データ（以下、「マーカデータ」と呼ぶ。）をフレームごとにまとめてモーションキャプチャデータ１６０を生成し、マルチモーダルコーパス作成装置１０８に出力する。なお、モーションキャプチャシステム１２０には、市販の光学式モーションキャプチャシステムを利用できる。市販の光学式モーションキャプチャシステムにおける赤外線カメラ１３２及びデータ処理装置１３４の機能及び動作については周知であるので、これらについての詳細な説明はここでは繰返さない。

【００４６】

再び図１を参照して、録音システム１２２は、発話者１０２の発する音声を受音して音響信号を発生するためのマイクロホン１４０Ａ及び１４０Ｂと、マイクロホン１４０Ａ及び１４０Ｂが発生した音響信号を増幅するためのアンプ１４２と、アンプ１４２により増幅された音響信号を所定の形式でデジタル化して図示しない記録媒体に記録するための録音装置１４４とを含む。記録されたデータ１６２はマルチモーダルコーパス作成装置１０８に与えられる。本明細書では、録音装置１４４が記録し出力するデータ１６２を「音声収録データ」と呼ぶ。

【００４７】

撮影システム１２４は、マイクロホン１４０Ａ及び１４０Ｂと同様の機能を持つマイクロホン１４０Ｃと、テレプロンプタ１２８の後方にマイクロホン１４０Ｃからの出力を受けるように配置され、テレプロンプタ１２８を通して発話者１０２の顔面及び首部の動画像を撮影し、マイクロホン１４０Ｃが発生する音響信号と撮影した動画像とを、タイムコードジェネレータ１３０からのタイムコードを付与して所定の形式でデータ化し、図示しない記録媒体に記録するためのカムコーダ１５０と、動画像の撮影時の光源となる複数の照明装置１５２Ａ、１５２Ｂ、及び１５２Ｃ（以下これらをまとめて「照明装置１５２」と呼ぶことがある。）とを含む。カムコーダ１５０により記録されたデータは、マルチモーダルコーパス作成装置１０８に与えられる。本明細書では、撮影システム１２４が記録し出力するデータ１６４を「カムコーダ収録データ」と呼ぶ。

【００４８】

図１に示す収録システム１０４はさらに、動画像の背景となるクロマキスクリーン１５４と、カムコーダ１５０により撮影される動画像を発話者１０２が確認できるように表示するためのモニタ１５６とを含む。

【００４９】

［図３　マルチモーダルコーパス作成装置の構成］

図３に、マルチモーダルコーパス作成装置１０８（図１参照）の機能的構成をブロック図で示す。図３を参照して、マルチモーダルコーパス作成装置１０８は、モーションキャプチャデータ１６０をデータ処理装置１３４から取込むためのモーションキャプチャデータ取込部１８０と、音声収録データ１６２を録音装置１４４から取込むための音声収録データ取込部１８２と、カムコーダ収録データ１６４をカムコーダ１５０から取込むためのカムコーダ収録データ取込部１８４と、取込まれたモーションキャプチャデータ１６０、音声収録データ１６２、及びカムコーダ収録データ１６４（以下、これらのデータをまとめて「収録データ」と呼ぶことがある。）を発話リスト１２６を構成する発話内容ごとに切出して、発話内容ごとの収録データのセット（以下、「発話別収録データセット」と呼ぶ。）２００Ａ，…，２００Ｌ（以下これらをまとめて「発話別収録データセット２００」と呼ぶことがある。）を生成するための切出処理部１８６と、発話別収録データセット２００を記憶するための発話別収録データセット記憶部１８８とを含む。なお、カムコーダ収録データ１６４のうちの音声データよりも音声収録データ１６２の方が高音質であるため、本実施の形態では音声収録データ１６２を用いる。

【００５０】

発話別収録データセット２００は、発話内容別に収録データをまとめたものである。発

話別収録データセット２００Ａ，…，２００Ｌは各々、発話内容を表す言語データ２１０と、発話者１０２による当該発話内容の発話時に計測されたフレームのマーカデータ（マーカの測定位置データ）からなる発話別モーションキャプチャデータ２１２と、当該発話内容の発話時に収録された部分の音声収録データからなる発話別音声データ２１４と、当該発話内容が発話された区間に収録された動画像のデータからなる発話別動画像データ２１６とを含む。

【００５１】

マルチモーダルコーパス作成装置１０８はさらに、モーションキャプチャデータの入力を受けて、これを頭部全体の動きをキャンセルするように正規化し、顔の器官の変化を表す顔器官変化量データ２２０を出力するため正規化処理部１９０と、発話別収録データセット記憶部１８８内の発話別収録データセット２００Ａ，…，２００Ｌのいずれかを読出し、その中の発話別モーションキャプチャデータ２１２を正規化処理部１９０に入力し、これに応答して正規化処理部１９０により出力される顔器官変化量データ２２０で、発話別モーションキャプチャデータ２１２を置換して発話別データセット２０２Ａ，…，２０２Ｌ（以下これらをまとめて「発話別データセット２０２」と呼ぶことがある。）を生成し、マルチモーダルコーパス１０６（図１参照）に格納するための発話別データセット生成部１９２を含む。

【００５２】

発話別データセット２０２は、マルチモーダルコーパス１０６を構成するデータを発話内容別にまとめたものである。発話別データセット２０２Ａ，…，２０２Ｌはそれぞれ、同様のデータ構成を有する。例えば、発話別データセット２０２Ａは、言語データ２１０と、発話別モーションキャプチャデータ２１２を正規化することにより得られる顔器官変化量データ２２０と、発話別音声データ２１４と、発話別動画像データ２１６とを含む。

【００５３】

図４に、切出処理部１８６の構成をブロック図で示す。図４を参照して、切出処理部１８６は、取込まれたモーションキャプチャデータ１６０、音声収録データ１６２、及びカムコーダ収録データ１６４をそれぞれ一時的に記憶しておくための、モーションキャプチャデータ記憶部２３０、音声収録データ記憶部２３２、及びカムコーダ収録データ記憶部２３４と、入力装置１１６Ａ及び出力装置１１８Ａを用いて行なわれるユーザの操作、並びに発話リスト１２６に基づき、言語データ２１０の生成、及びカムコーダ収録データ１６４からの発話別動画像データ２１６の切出を行なうための動画像データ切出部２４０と、モーションキャプチャデータ１６０のタイムコード及び発話別動画像データ２１６のタイムコードに基づいて、モーションキャプチャデータ１６０から発話別モーションキャプチャデータ２１２を切出すためのモーションキャプチャデータ切出部２４２と、音声収録データ１６２を、カムコーダ収録データ１６４の音声データと同期させることにより音声収録データにタイムコードを付与するための同期処理部２４４と、この音声収録データ１６２のタイムコードと発話別動画像データ２１６のタイムコードとに基づいて、音声収録データ１６２からの発話別動画像に同期した発話別音声データ２１４を切出すための音声データ切出部２４６と、動画像データ切出部２４０により生成される言語データ２１０及び発話別動画像データ２１６、並びに当該データに対応する発話別モーションキャプチャデータ２１２及び発話別音声データ２１４をそれぞれ受けて一時的に保持し、発話内容ごとに発話別収録データセット２００（Ａ，…，Ｌ）を形成して出力するためのデータセット形成部２４８とを含む。

【００５４】

図１に示す録音装置１４４は、音声収録データにタイムコードを付与する機能を持たない。しかし音質はカムコーダ１５０により録音されたものよりも録音装置１４４により得られた音声収録データの方がよい。そこで、上記したように同期処理部２４４により音声収録データをカムコーダ収録データ１６４内の音声データに付与されたタイムコードと同期させる。より具体的には、同期処理部２４４は、カムコーダ収録データ１６４における音声のデータと、音声収録データ１６２との相互相関を計算し、相互相関が最大となるよ

うに音声収録データとカムコーダ収録データ１６４の音声データとのずれを計算し、その結果に基づいて音声収録データにタイムコードを付与する。

【００５５】

正規化処理部１９０は、発話別モーションキャプチャデータ２１２を構成する各マーカデータに対しアフィン変換を行なうことにより、顔の各器官の変化に起因するマーカ位置の変化量のみからなる（頭部の動きに起因する変化量を除いた）顔器官変化量データを生成する機能を持つ。ここに、発話別モーションキャプチャデータ２１２におけるマーカデータを同次座標系でＰ＝〈P_x，P_y，P_z，１〉と表現し、当該マーカデータを基に生成される顔器官変化量データをＰ'＝〈P'_x，P'_y，P'_z，１〉と表現すると、アフィン行列Ｍは、次の式のように表現される。

【００５６】

【数１】

$$\mathbf{P'} = \mathbf{MP} \qquad \mathbf{M} = \begin{bmatrix} M_{11} & M_{12} & M_{13} & M_{14} \\ M_{21} & M_{22} & M_{23} & M_{24} \\ M_{31} & M_{32} & M_{33} & M_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

上記式において、アフィン行列Ｍは、頭部の動きのみが含まれていると考えられる４箇所以上のマーカに対応するマーカデータから、特異値分解によって算出される。本実施の形態では、正規化用のマーカとして額部に４点、こめかみ部に２点、及び鼻部に２点のマーカを設け、それらを基準として各マーカの変化量の正規化を行なう。

【００５７】

なお、首部のマーカの変化量は頭部の動きには影響を受けず、首自身の動きに影響される。そのため、上記の頭部に対するものと同様の考え方にしたがい、別途、首部の動きの補正用マーカ４点を用意し、頭部の動きに対する正規化と同様の処理を首部のマーカに対し行なう。

【００５８】

図５に、正規化処理部１９０の構成をブロック図で示す。図５を参照して、正規化処理部１９０は、発話別モーションキャプチャデータ２１２の入力を受け、発話別モーションキャプチャデータ２１２の各フレームにおいて、マーカデータから、首部以外の、顔を含む頭部に装着されたマーカの位置を表す頭部マーカデータと、首部に装着されたマーカの位置を表す首部マーカデータと分類して出力するためのデータ分類部２６０とを含む。

【００５９】

正規化処理部１９０はさらに、データ分類部２６０から頭部マーカデータを受け、当該マーカデータの中から補正用のマーカデータを選択するための頭部補正用マーカデータ選択部２６２と、頭部補正用マーカデータ選択部２６２により選択されたマーカデータをもとに特異値分解を行ない、頭部正規化のためのアフィン行列を算出するための頭部アフィン行列算出部２６４と、頭部アフィン行列算出部２６４により算出されたアフィン行列を用いて、データ分類部２６０により出力された頭部マーカデータに対しアフィン変換を行なうことにより、頭部に装着された各マーカの変化量を算出するための頭部マーカデータ変換部２６６とを含む。

【００６０】

正規化処理部１９０はさらに、データ分類部２６０から首部マーカデータを受け、当該マーカデータの中から補正用のマーカデータを選択するための首部補正用マーカデータ選択部２７２と、首部補正用マーカデータ選択部２７２により選択されたマーカデータを基に特異値分解を行ない、首部正規化のためのアフィン行列を算出するための首部アフィン

行列算出部２７４と、データ分類部２６０により出力された首部マーカデータに対して、首部アフィン行列算出部２７４により算出されたアフィン行列を用いてアフィン変換を行なうことにより、首部に装着された各マーカの変化量を算出するための首部マーカデータ変換部２７６とを含む。

【００６１】

正規化処理部１９０はさらに、頭部マーカデータ変換部２６６から頭部に装着された各マーカの変化量を表すデータを、首部マーカデータ変換部２７６から首部に装着された各マーカの変化量を表すデータを、それぞれ受け、フレームごとに当該データを統合することにより、正規化された顔器官変化量データ２２０を作成し、発話別データセット生成部１９２に出力するためのデータ統合部２７８を含む。

【００６２】

図６に、アニメーション作成装置１１４（図１参照）の構成をブロック図で示す。図６を参照して、アニメーション作成装置１１４は、入力装置１１６Ｂ及び出力装置１１８Ｂに接続され、ユーザの操作にしたがい、図２に示すマーカ１７０に対応する仮想のマーカ（以下、単に「仮想マーカ」と呼ぶ。）を初期顔モデル１１０上に配置することにより、当該各仮想マーカの、初期顔モデル１１０を規定する座標系上での座標を設定するための仮想マーカ設定部３００と、初期顔モデル１１０内の各ノードに対して、各ノードに近接する所定数（本実施の形態では３個）の仮想マーカを当該ノードに対応するマーカに選び、その対応関係を付与した顔モデル（以下、「マーカ対応顔モデル」と呼ぶ。）３１０を作成するためのマーカ対応顔モデル作成部３０２とを含む。図２に示すマーカ１７０と、仮想マーカとの対応関係がこのようにして定義されることにより、発話時の発話者に装着された各マーカの位置を、顔モデル上の各仮想マーカの位置に割当てることができる。なおこの際、モーションキャプチャデータの座標系と顔モデルの座標系との間の変換も行なわれる。

【００６３】

アニメーション作成装置１１４はさらに、入力装置１１６Ｂ及び出力装置１１８Ｂに接続され、ユーザの操作にしたがい、マルチモーダルコーパス１０６内の発話別データセット２０２の中のいずれかを、作成予定のアニメーション１１２における発話内容に応じて選択し取得するための発話別データセット取得部３０４と、取得された発話別データセットにおける顔器官変化量データ２２０に基づき、初期顔モデル１１０が表現する顔の形状から、変形した顔モデルを順次作成するための顔モデル変形部３０６と、顔モデル変形部３０６により順次作成される変形した顔モデルに対し、テクスチャ等を付与して画像化することにより、アニメーション１１２を生成するための画像化部３０８とを含む。

【００６４】

マーカ対応顔モデル作成部３０２は、初期顔モデル１１０のノードの中から、処理の対象となるノードを選択するためのノード選択部３１２と、ノード選択部３１２により選択されたノード（以下、「選択ノード」と呼ぶ。）からの距離が最も近い仮想マーカを、仮想マーカの座標の設定値に基づき選択するための仮想マーカ選択部３１４と、仮想マーカ選択部３１４により、各ノードに対し適切な仮想マーカが所定数選択されるように仮想マーカ選択部３１４を制御し、選択された所定数の仮想マーカ（以下これらの仮想マーカを選択ノードに対する「対応マーカ」と呼ぶ。）を特定する情報を処理対象のノードに付与するための選択マーカ検査部３１６とを含む。

【００６５】

具体的には、選択マーカ検査部３１６は、仮想マーカ選択部３１４により選択された仮想マーカ（以下、「選択マーカ」と呼ぶ）が、この選択ノードに対応付ける仮想マーカとして適切であるために必要な条件を充足するかを検査する。条件が充足されなければ仮想マーカ選択部３１４に対し次にこのノードに近い仮想マーカを選択するように要求する。条件が充足されていればこの仮想マーカを当該ノードの対応マーカに指定する。さらに、対応マーカが１個指定されるたびに、対応マーカが３個選択されたかを検査し、３個に満たない場合には新たな仮想マーカを選択するように仮想マーカ選択部３１４に対し要求す

る。３個となれば、選択マーカ検査部３１６は、ノード選択部３１２に対する次の処理対象のノードの選択要求を発生する。

【００６６】

図７に、マーカ対応顔モデル作成部３０２により実行される、対応マーカの指定処理を実現するコンピュータプログラムの制御構造をフローチャートで示す。図７を参照して、対応マーカの指定処理が開始されると、ステップ３４０Ａとステップ３４０Ｂとで囲まれた、ステップ３４２からステップ３５４までの処理を、初期顔モデル１１０における全ノードに対して処理が完了するまで実行する。

【００６７】

ステップ３４２では、初期顔モデル１１０を構成するノードのうち、未処理のノードを１つ選択する。これを選択ノードとする。ステップ３４４では、選択ノードから仮想マーカまでの距離をそれぞれ算出する。さらに仮想マーカをこの距離の昇順でソートしたものをリストする。ステップ３４５では、以下の繰返しを制御するための変数ｉ及び選択されたマーカの数を表す変数ｊに０を代入する。ステップ３４６では、変数ｉに１を加算する。

【００６８】

ステップ３４７では、変数ｉの値が仮想マーカの数Mmaxを超えているか否かを判定する。変数ｉの値が数Mmaxを超えていればエラーとし、処理を終了する。このようなことは普通はないが、念のためにこのようなエラー処理を設けておく。変数ｉの値が数Mmax以下であれば制御はステップ３４８に進む。

【００６９】

ステップ３４８では、リストの先頭から変数ｉで示される位置に存在する仮想マーカ（以下これを「マーカ（ｉ）」と呼ぶ。）と選択ノードとを結ぶ線分が、初期顔モデル１１０におけるいずれの境界エッジも横切らない、という制約条件を充足しているか否かを判定する。当該線分が境界エッジのいずれかを横切るものであれば、ステップ３４５に戻る。さもなければステップ３５０に進む。

【００７０】

ステップ３５０では、この時点でのマーカ（ｉ）を選択ノードの対応マーカのひとつに指定する。すなわちマーカ（ｉ）を示す情報を、選択ノードのマーカ・ノード対応情報として保存する。この後制御はステップ３５２に進む。ステップ３５２では、変数ｊに１を加算する。ステップ３５４では、変数ｊの値が３となっているか否かを判定する。変数ｊの値が３であればステップ３４０Ｂに進む。さもなければステップ３４５に進む。

【００７１】

上記したように、選択ノードと仮想マーカとを結ぶ線分が顔モデルの境界エッジを横切るものは、ノードに対応する仮想マーカから除外される。これは以下の理由による。例えば目の上まぶたと下まぶたとのように、間に境界エッジ（例えば目のふち）が存在する場合がある。この場合、上まぶたに位置するノードと、下まぶたに位置するノードとは互いに異なる動きをする。したがって、例えば上まぶたのノードの変化量を算出する際に、下まぶたに存在するマーカの変化量を用いることは適当ではない。なお、線分がある境界エッジを横切っているか否かは、その境界エッジが、顔モデルを構成するポリゴンのうち二つによって共有されているか、一つのみに属しているかに基づいて判定する。

【００７２】

図１０に、初期顔モデル１１０における目輪郭部周辺のポリゴンと、仮想マーカとを示す。図１０を参照して、初期顔モデル１１０の目輪郭部の周囲には、多数の三角形ポリゴンが存在する。このうち例えばポリゴン４０２は、３つのエッジ４０４Ａ、４０４Ｂ、及び４０４Ｃにより囲まれている。エッジ４０４Ａ及び４０４Ｂは、他のポリゴンと共有されている。しかし、エッジ４０４Ｃは、他のポリゴンと共有されていない。エッジ４０４Ｃ等２つのポリゴンにより共有されていないエッジは、初期顔モデル１１０の切れ目との接線又は外縁にあたる。このようなエッジが境界エッジとなる。

【００７３】

再び図６を参照して、顔モデル変形部３０６は、あるフレームにおいて測定された、座標変換済みのマーカ変化量を各仮想マーカに付与する。さらに顔モデル変形部３０６は、マーカ対応顔モデル３１０のマーカ・ノード対応情報に基づき、各ノードに、対応する仮想マーカの変化量から所定の内挿式により算出される変化量ベクトルｖを割当てることにより、顔モデルの変形を行なう。マーカ対応顔モデル３１０のノードの座標をＮ、当該ノードと対応関係にある仮想マーカの座標をM_i、変形後の顔モデルにおけるマーカの座標をM'_iとすると、顔モデル変形部３０６は、ノードの座標の変化量ベクトルｖを次の内挿式によって算出する。

【００７４】

【数２】

$$\mathbf{v}=\sum_{i}^{n}(\mathbf{M'_i}-\mathbf{M_i})\cdot\left(\frac{1.0}{\mathbf{N}-\mathbf{M_i}}\right)\cdot\left(\sum_{i}^{n}\frac{1.0}{\mathbf{N}-\mathbf{M_i}}\right)^{-1}$$

なお、本実施の形態においては、ｎ＝３である。すなわち、１つのノードに対応付ける仮想マーカの数は３である。

【００７５】

［動作］

本実施の形態に係るマルチモーダルコーパス作成システム１００は以下のように動作する。まず、図１に示すマルチモーダルコーパス作成システム１００の収録システム１０４を用いた、音声、動画像、及びモーションキャプチャデータの収録プロセスについて説明する。

【００７６】

発話者１０２の顔面及び首部には、事前に、表１に示すようなルールにしたがい図２に示すように多数のマーカを予め装着しておく。図１を参照して、赤外線カメラ１３２はそれぞれ、各マーカからの反射光を受光可能な所定の位置に、受光部を発話者１０２の顔面及び首部に向けて設置される。マイクロホン１４０Ａ及び１４０Ｂはそれぞれ、発話者１０２の上部及び胸部等、発話者１０２の発する音声を受音可能な所定の位置に設置される。カムコーダ１５０は、発話者１０２の正面等、顔面及び首部の撮影に好適な位置に、受光部を発話者１０２に向けて設置される。なお、マイクロホン１４０Ｃは、発話者１０２の発する音声を受音可能で、かつカムコーダ１５０に接続可能な位置に設置される。照明装置１５２はそれぞれ、発話者１０２の顔にセルフシャドーが起こることを防止できる位置に設置される。例えば、照明装置１５２Ａ，１５２Ｂ，及び１５２Ｃはそれぞれ、発話者１０２の左右、及び正面ローアングルから、発話者に向けて光が照射されるように設置される。クロマキスクリーン１５４は、カムコーダ１５０から見て発話者１０２の背後に設置される。

【００７７】

テレプロンプタ１２８は、発話者１０２とカムコーダ１５０との間に、発話者１０２側からの光がカムコーダ１５０側に透過するよう設置される。モニタ１５６は、テレプロンプタ１２８の上部に画面を発話者１０２に向けて設置される。カムコーダ１５０は、テレプロンプタ１２８越しに発話者１０２を撮影することになる。そのため発話者１０２がテレプロンプタ１２８及びモニタの表示を見ると、発話者１０２の視線はテレプロンプタ１２８越しにカムコーダ１５０に向けられることになる。

【００７８】

収録時には、発話リスト１２６を構成する文章等を発話リスト１２６にしたがいテレプロンプタ１２８が表示する。発話者１０２は、テレプロンプタ１２８及びモニタ１５６の表示を確認しながら、発話リスト１２６により指定された内容の文章等を順次発話する。

【００７９】

発話時における顔の各部位の位置は、モーションキャプチャシステム１２０により次の

ようにして計測される。マーカはそれぞれ、発話時における顔の各器官の変化並びに頭部及び首部の動きに追従して移動する。赤外線カメラ１３２はそれぞれ、マーカによる赤外線反射光を、所定のフレームレート（例えば毎秒１２０フレーム）で撮影しその映像信号をデータ処理装置１３４に出力する。データ処理装置１３４は、赤外線カメラ１３２からの映像信号の各フレームにタイムコードジェネレータ１３０からのタイムコードを付与し、当該映像信号を基に各マーカの位置をフレームごとに算出する。データ処理装置１３４は、各マーカの位置のデータをフレームごとにまとめてモーションキャプチャデータ１６０として蓄積する。

【００８０】

発話時における発話者１０２の音声は、録音システム１２２により、次のようにして収録される。すなわち、マイクロホン１４０Ａ及び１４０Ｂは、発話者１０２の音声を受音して、音響信号を発生する。アンプ１４２は、発生した音響信号の入力を受け、当該音響信号の各々を増幅して録音装置１４４に出力する。録音装置１４４は、増幅された音響信号をアンプ１４２から受け音声収録データ１６２として記録する。

【００８１】

発話時における発話者１０２の顔の動画像は、撮影システム１２４により、次のようにして収録される。すなわち、マイクロホン１４０Ｃは、１４０Ａ及び１４０Ｂと同様に発話者１０２の音声を受音して音響信号を発生する。この音響信号は、カムコーダ１５０に与えられる。同時にカムコーダ１５０は、テレプロンプタ１２８越しに、発話中の発話者１０２のバストアップの動画像を正面から撮影する。カムコーダ１５０は、動画像とマイクロホン１４０Ｃからの音響信号とから所定の形式のカムコーダ収録データを形成し記録する。この際カムコーダ１５０は、タイムコードジェネレータ１３０のタイムコードをカムコーダ収録データ１６４の各フレームに付与する。

【００８２】

以上の収録プロセスにより、タイムコードジェネレータ１３０のタイムコードが付与されたモーションキャプチャデータ１６０と、同じタイムコードが付与された音声及び動画像のデータからなるカムコーダ収録データ１６４と、音声収録データ１６２とが同時に収録される。これらのデータは、マルチモーダルコーパス作成装置１０８に与えられる。

【００８３】

［マルチモーダルコーパス作成装置１０８の動作］

図３を参照して、マルチモーダルコーパス作成装置１０８のモーションキャプチャデータ取込部１８０は、図１に示す収録システム１０４のデータ処理装置１３４よりモーションキャプチャデータ１６０を取込む。この際モーションキャプチャデータ取込部１８０は、モーションキャプチャデータ１６０を、３次元コンピュータグラフィックスを扱うソフトウェアで利用可能な形式で取込む。

【００８４】

図８は、１フレーム分のモーションキャプチャデータ１６０に含まれるマーカデータを基に、各マーカの位置をコンピュータグラフィックスで表現した図である。図８を参照して、円形の目印はそれぞれ、当該フレームにおけるマーカの位置を表す。１フレーム分のモーションキャプチャデータは、マーカと同数のマーカデータを含む。

【００８５】

再び図３を参照して、音声収録データ取込部１８２は、録音装置１４４より音声収録データ１６２を取込む。カムコーダ収録データ取込部１８４は、カムコーダ１５０よりカムコーダ収録データ１６４を取込む。取込まれたモーションキャプチャデータ１６０と、音声収録データ１６２と、カムコーダ収録データ１６４とはそれぞれ、切出処理部１８６に与えられる。

【００８６】

切出処理部１８６は、発話リスト１２６を構成する文章、単語等の発話内容ごとに以下の動作により、発話別収録データセットを作成する。

【００８７】

図４を参照して、モーションキャプチャデータ１６０と、音声収録データ１６２と、カムコーダ収録データ１６４はそれぞれ、モーションキャプチャデータ記憶部２３０、音声収録データ記憶部２３２、及びカムコーダ収録データ記憶部２３４に格納される。音声収録データ１６２とカムコーダ収録データ１６４が格納されると、同期処理部２４４は、カムコーダ収録データ１６４における音声のデータと音声収録データ１６２との相互相関を計算し、最大の相関が得られるように音声収録データをフレームに分割し、各フレームに対応するカムコーダ収録データ１６４の音声データのフレームに付与されていたものと同じタイムコードを付与する。同期処理部２４４は、処理後の音声収録データ１６２を音声収録データ記憶部２３２に格納する。

【００８８】

ユーザが発話別のデータセット形成を指示するために入力装置１１６Ａを用いて所定の操作を行なうと、動画像データ切出部２４０は、カムコーダ収録データ記憶部２３４からカムコーダ収録データ１６４を読出す。動画像データ切出部２４０はさらに、発話リスト１２６を取得する。動画像データ切出部２４０は、ユーザの操作に応じて、カムコーダ収録データ１６４の動画像及び音声、並びに発話リストを出力装置１１８Ａを介して出力する。出力装置１１８Ａによる出力を参考にユーザが入力装置１１６Ａを用いて、１つの発話内容に対応する動画像の収録された区間の開始位置及び終了位置を指定すると、動画像データ切出部２４０は、この入力にしたがい、指定された区間の動画像のデータをカムコーダ収録データ１６４から抽出し、発話別動画像データ２１６を生成してデータセット形成部２４８に与える。発話別動画像データ２１６のうち、その開始と終了とを表すタイムコードが、モーションキャプチャデータ切出部２４２、音声データ切出部２４６、及びデータセット形成部２４８に与えられる。動画像データ切出部２４０はさらに、抽出した部分の動画像に対応する言語データ２１０を、ユーザによる入力及び発話リスト１２６を基に生成する。生成された言語データ２１０は、データセット形成部２４８に与えられる。

【００８９】

発話別動画像データ２１６の開始と終了とを表すタイムコードに応答して、モーションキャプチャデータ切出部２４２は、指定された区間を特定する。モーションキャプチャデータ切出部２４２は、モーションキャプチャデータ記憶部２３０内のモーションキャプチャデータ１６０から、当該区間に対応するデータを抽出して発話別モーションキャプチャデータ２１２を生成し、データセット形成部２４８に与える。

【００９０】

音声データ切出部２４６は、発話別動画像データ２１６の開始と終了とを表すタイムコードに応答して、音声収録データからタイムコードにより指定された区間に対応するデータを抽出して発話別音声データ２１４を生成する。生成された発話別音声データ２１４は、データセット形成部２４８に与えられる。

【００９１】

データセット形成部２４８は、言語データ２１０、発話別動画像データ２１６、発話別モーションキャプチャデータ２１２、及び発話別音声データ２１４が与えられたことに応答して、これら与えられたデータをまとめて発話別収録データセット２００を生成し、図３に示す発話別収録データセット記憶部１８８に格納する。以上の動作により、発話内容ごとの発話別収録データセット２００が形成され、発話別データセット記憶部１８８に格納される。

【００９２】

発話別データセット生成部１９２及び正規化処理部１９０は、発話別収録データセット２００の各々について以下の処理を行ない、発話別データセット２０２を生成する。すなわち、発話別データセット生成部１９２は発話別収録データセット記憶部１８８から発話別収録データセットを１セット分読出す。発話別データセット生成部１９２はさらに、発話別モーションキャプチャデータ２１２から、１フレーム分のデータを正規化処理部１９０に与える。

【００９３】

図５を参照して、正規化処理部１９０のデータ分類部２６０は、１フレーム分のデータが与えられたことに応答して、当該フレームにおけるマーカデータを、頭部マーカデータと、首部マーカデータとに分類する。データ分類部２６０は、頭部マーカデータを頭部補正用マーカデータ選択部２６２及び頭部マーカデータ変換部２６６に与え、首部マーカデータを首部補正用マーカデータ選択部２７２及び首部マーカデータ変換部２７６に与える。

【００９４】

頭部補正用マーカデータ選択部２６２は、与えられたマーカデータの中から、予め定められた８箇所のマーカデータを補正用のマーカとして選択し、それぞれ頭部アフィン行列算出部２６４に与える。頭部アフィン行列算出部２６４は、与えられたマーカデータからの特異値分解によってアフィン行列Ｍを算出し、頭部マーカデータ変換部２６６に与える。頭部マーカデータ変換部２６６は、与えられた頭部マーカデータを、このアフィン行列Ｍによって変換する。この変換により、マーカデータはそれぞれ、頭部の動きを除いた正規化した変化量に変換される。頭部マーカデータ変換部２６６は、各マーカの正規化後の変化量をデータ統合部２７８に与える。

【００９５】

首部補正用マーカデータ選択部２７２は、与えられたマーカデータの中から予め定められた４箇所の首部補正用のマーカデータを選択し、それぞれ首部アフィン行列算出部２７４に与える。首部アフィン行列算出部２７４は与えられたマーカデータを用いて首部補正用のアフィン行列を算出し首部マーカデータ変換部２７６に与える。首部マーカデータ変換部２７６は、データ分類部２６０から与えられた首部マーカデータを首部アフィン行列算出部２７４から与えられたアフィン行列で変換する。この変換により、マーカデータはそれぞれ、首部の動きを除いた正規化した変化量に変換される。首部マーカデータ変換部２７６は、各マーカの変化量をデータ統合部２７８に与える。

【００９６】

データ統合部２７８は、頭部マーカデータ変換部２６６と首部マーカデータ変換部２７６とからそれぞれ与えられるマーカの変化量のデータを統合して、１フレーム分の顔器官変化量データ２２０を生成する。データ統合部２７８は、生成した顔器官位置変化量データ２２０を発話別データセット生成部１９２（図３参照）に返す。

【００９７】

図３を参照して、発話別データセット生成部１９２は、正規化処理部１９０から１フレーム分の顔器官変化量データ２２０が返されると、発話別モーションキャプチャデータ２１２における当該フレームのデータを、そのフレームの顔器官変化量データ２２０で置換し、言語データ２１０、発話別動画像データ２１６、及び発話別音声データ２１４とともに発話別データセット２０２に出力する。発話別データセット生成部１９２はこの後、新たに１フレーム分のマーカデータを正規化処理部１９０に与え、上記と同様の処理を繰返す。

【００９８】

正規化処理部１９０及び発話別データセット生成部１９２は、以上の動作を発話別収録データセット２００の各々の全フレームについて繰返すことにより、発話別データセット２０２を形成する。形成された発話別データセット２０２は、図１に示すマルチモーダルコーパス１０６に格納される。

【００９９】

［アニメーションの作成］

次に、アニメーション作成装置１１４がアニメーション１１２を作成する動作について説明する。図６を参照して、アニメーション作成装置１１４に初期顔モデル１１０が与えられると、アニメーション作成装置１１４は、動作を開始する。図９に、初期顔モデル１１０の一例を示す。図９を参照して、この初期顔モデル１１０は、発話者１０２の顔の静止画像と所定のワイヤフレームモデルとを整合させることにより準備された形状モデルである。この顔モデルは、約７５０のポリゴンで構成されている。初期顔モデル１１０は、

仮想マーカ設定部３００と、マーカ対応顔モデル作成部３０２のノード選択部３１２及び選択マーカ検査部３１６とに与えられる。

【０１００】

仮想マーカ設定部３００は、初期顔モデル１１０を画像化して出力装置１１８Ｂに出力する等して、さらにユーザから当該初期顔モデル上における仮想マーカの位置の指定を入力装置１１６Ｂを介して受ける。初期顔モデル１１０上での仮想マーカの位置は、既に述べた表１と同様のルールにしたがって指定される。そのため、初期顔モデル１１０における顔器官と仮想マーカとの位置関係は、発話者１０２の顔器官と当該発話者１０２に装着されたマーカとの位置関係に対応する。

【０１０１】

仮想マーカ設定部３００は、ユーザによる指定を基に、各マーカのマーカデータに対しモーションキャプチャデータの座標系から顔モデルの座標系に対する座標変換を行ない、初期顔モデルの座標系における各仮想マーカの座標を特定する。仮想マーカ設定部３００は、当該各仮想マーカの識別子と当該仮想マーカの座標とを、マーカ対応顔モデル作成部３０２の仮想マーカ選択部３１４に与える。

【０１０２】

マーカ対応顔モデル作成部３０２は、初期顔モデル１１０と仮想マーカの識別子及び座標とが与えられたことに応答して、初期顔モデル１１０の各ノードに対して、当該ノードの対応マーカを次のようにして特定する。まず、ノード選択部３１２が、初期顔モデル１１０を構成するノードの中からノードを１つ選択する。このノードが選択ノードである。選択ノードと全ての仮想マーカとの距離を算出し、仮想マーカを距離の昇順にソートしてリスト化する。このリストの先頭の一つを選び、その仮想マーカと選択ノードとを結ぶ線が顔モデルの境界エッジを横切るか否かを判定する。横切らなければこの仮想マーカを選択ノードの対応ノードの一つに選択する。横切っていればリストの次の仮想マーカを選択し、同じ処理を繰返す。

【０１０３】

こうして、選択ノードに対し３つの仮想マーカが当該選択ノードの対応ノードとして特定される。対応ノードと選択ノードとを結ぶ線分のいずれも、顔モデルの境界エッジを横切らない。

【０１０４】

例えば、図１０を参照して、ノード４１２が選択ノードであるときを考える。なお、初期顔モデル１１０の目輪郭部周囲において、仮想マーカ４１０Ａ，…，４１０Ｌが設定されているものとする。仮想マーカ選択部３１４は、ノード４１２の座標と、仮想マーカ４１０Ａ，…，４１０Ｌの座標データとを基に、選択ノード４１２と仮想マーカ４１０Ａ，…，４１０Ｌとの間の距離をそれぞれ算出する。仮想マーカ選択部３１４は、仮想マーカ４１０Ａ，…，４１０Ｌの中から、ノード４１２に最も近い位置にある仮想マーカ４１０Ｋを選択する。

【０１０５】

選択マーカ検査部３１６は、ノード４１２と選択された仮想マーカ４１０Ｋとを結ぶ線分が境界エッジを横切るか否かを検査する。ノード４１２と選択された仮想マーカ４１０Ｋとを結ぶ線分は、いずれの境界エッジも横切らない。そのため、選択マーカ検査部３１６は、当該仮想マーカ４１０Ｋをノード４１２の対応マーカに指定する。選択マーカ検査部３１６はさらに、新たな仮想マーカの選択要求を仮想マーカ選択部３１４に与える。

【０１０６】

仮想マーカ選択部３１４は、選択マーカ検査部３１６からの通知及び要求に応答して、仮想マーカ４１０Ｋの次にノード４１２に近い位置にある仮想マーカを選択する。図１０に示す例では、この選択により、ノード４１２に２番目に近接する仮想マーカ４１０Ｂが選択される。

【０１０７】

選択マーカ検査部３１６は、選択された仮想マーカ４１０Ｂについての検査を上記の動

作と同様の動作で行なう。この場合、ノード４１２と仮想マーカ４１０Ｂとを結ぶ線分は、境界エッジを横切る。そのため、選択マーカ検査部３１６は、当該仮想マーカ４１０Ｂをノード４１２の対応マーカに指定せず対象から除外する。選択マーカ検査部３１６はさらに、新たな仮想マーカの選択を仮想マーカ選択部３１４に要求する。

【０１０８】

仮想マーカ選択部３１４及び選択マーカ検査部３１６が以上の動作を繰返し、ノード４１２の対応マーカとして３個の仮想マーカ（図１０に示す例では仮想マーカ４１０Ｊ，４１０Ｋ，及び４１０Ｌ）が指定されると、ノード４１２に対する仮想マーカの対応付けが完了する。選択マーカ検査部３１６はノード４１２とその対応マーカに関するマーカ・ノード対応情報をマーカ対応顔モデルの一部として出力し、ノード選択部３１２に対し新たなノードの選択要求を与える。

【０１０９】

ノード選択部３１２は、選択マーカ検査部３１６からの要求に応答して、初期顔モデル１１０を構成するノードのうち、対応付けが未完了のノードから１つのノードを選択する。以下、上記したノード選択部３１２、仮想マーカ選択部３１４、及び選択マーカ検査部３１６の動作が、全てのノードに対して対応マーカが決定されるまで繰返される。

【０１１０】

こうして、マーカ対応顔モデル作成部３０２により、各ノードに対し３個の仮想マーカを対応付けるマーカ対応顔モデル３１０（図６参照）が生成される。マーカ対応顔モデル３１０は、図６に示す顔モデル変形部３０６に与えられる。

【０１１１】

次に、アニメーション作成装置１１４が、マーカ対応顔モデル３１０を用いてアニメーションを作成する動作について説明する。図６を参照して、ユーザが入力装置１１６Ｂを用いて、発話内容等を入力すると、当該入力は、発話別データセット取得部３０４に与えられる。発話別データセット取得部３０４は、マルチモーダルコーパス１０６から、入力された発話内容等に対応する発話別データセット２０２Ａ，…，２０２Ｌ（図３参照）を読出し、当該発話別データセット内の顔器官変化量データ２２０（図３参照）を顔モデル変形部３０６に与える。

【０１１２】

この時点で顔モデル変形部３０６には、顔器官変化量データ２２０と、マーカ対応顔モデル３１０とが与えられている。マーカ対応顔モデル３１０の各ノードには、当該顔モデル上の仮想マーカが３個指定されている。顔モデル変形部３０６は、顔器官変化量データ２２０をもとに、各マーカの位置の変化量に基づき、マーカ対応顔モデル３１０中の各ノードの変化量を次のようにして算出する。

【０１１３】

すなわち、顔モデル変形部３０６はまず、マーカ対応顔モデル３１０上における仮想マーカの座標を取得する。仮想マーカはそれぞれ、顔器官変化量データ２２０におけるマーカと対応関係にある。そこで、顔モデル変形部３０６は、顔器官変化量データ２２０における１フレーム分のデータを基に、仮想マーカの各々に、当該仮想マーカに対応するマーカの変化量を付与し、当該１フレーム分の変化後の各仮想マーカの座標を算出する。

【０１１４】

さらに顔モデル変形部３０６は、１つのノードの変化量を、ノードに対し指定された３個の対応マーカの座標を基に決定する。ここに、あるノード座標をＮとする。また当該ノードの対応マーカの変化前の座標をそれぞれM_i（$1 \leqq i \leqq n = 3$）とする。さらに、当該対応マーカについて、１フレーム分の変化量が付与された後の座標をM'_iとする。顔モデル変形部３０６は、ノードの変化量ベクトルｖを次の式により算出する。

【０１１５】

【数３】

$$\mathbf{v}=\sum_{i}^{n}(\mathbf{M'_i}-\mathbf{M_i})\cdot(\frac{1.0}{\mathbf{N}-\mathbf{M_i}})\cdot(\sum_{i}^{n}\frac{1.0}{\mathbf{N}-\mathbf{M_i}})^{-1}$$

上記の式でノードの変化量ベクトルｖを、変形前の当該ノードに対し付与することにより、変化後のノードの座標が算出される。顔モデル変形部３０６は、フレームごとに、マーカ対応顔モデルの各ノードに対しこの処理を実行する。これにより、各ノードの座標は変更され、変形した顔モデルがフレームごとに生成される。顔モデル変形部３０６は、変形した顔モデルの各々を、画像化部３０８に与える。

【０１１６】

画像化部３０８は、フレームごとの変形した顔モデルを受けると、それらにテクスチャなどを付与してそれらを画像化することにより、アニメーション１１２における各コマの画像を生成する。さらに、必要に応じて、コマの間引き等の処理を行ない、一連の動画像を形成する。形成した動画像が、アニメーション１１２となる。

【０１１７】

図１１に、発話中における発話者１０２の顔画像と、顔器官変化量データ２２０及び図９に示す初期顔モデル１１０をもとに作成したアニメーション１１２における顔画像とを対比して示す。図１１を参照して、１段目には、マルチモーダルコーパス１０６に格納された動画像データのうち、異なる５つの発話内容の発話中にそれぞれ撮影された動画像中のフレームの画像を示す。発話内容に応じて、発話者１０２の口及び目等顔の各器官の形状が変化している。これらの画像のいずれにおいても、発話者の頭部の向き、大きさ、及び傾きは他の画像におけるそれらとは僅かながら異なる。この相違は、各画像における額部のマーカ及び拘束部材の位置に顕著に現れている。

【０１１８】

２段目の画像は、１段目の各画像の収録と同じ時点での顔器官変化量データ２２０に基づき変形した顔モデルにおけるポリゴンの形状を表す画像である。３段目は、アニメーション１１２において、２段目に示すポリゴン形状の顔モデルをもとに画像化されたフレームの画像である。対応する１段目の画像と比較すると、２段目及び３段目の画像における口及び目等、顔の各器官の形状は、１段目の動画像と同様に変化している。また、２段目及び３段目の画像においては、顔の各器官の変化量に応じて顔モデルを変形させているため、頭部の向き、大きさは一定に保たれている。３段目の各画像における額のマーカの位置は、一定している。

【０１１９】

以上のように、本実施の形態では、マルチモーダルコーパスを、発話者による発話中の音声、動画像、及び顔部位の位置の計測データを基に作成する。顔部位の位置の計測には光学式モーションキャプチャシステムを用いるため、顔部位の位置を動画像から推定しなくてもよく、高速度で３次元の位置計測が行なえる。その結果、顔部位の特徴量の算出が容易になる。また、発話の収録時には、顔部位の同定に用いる多数のマーカを、事前に定めたルールにしたがい発話者１０２に装着する。したがって、高精度かつ詳細に顔部位の変化量を得ることができる。また、複数の発話者から、又は同一の発話者から複数回にわたってそれぞれ収録を行なう場合であっても、計測条件を安定させることが容易で再現性の高い計測をすることが可能となる。その結果、大規模なマルチモーダルコーパスを作成することが可能になる。

【０１２０】

本実施の形態のマルチモーダルコーパス作成装置は、モーションキャプチャデータを基に、顔器官の変化量の算出を行なう。そのため、動画像の光学的な誤差に影響を受けることなく発話中の各器官の変化を正確にコーパス化できる。マルチモーダルコーパス作成装置は、モーションキャプチャデータを正規化して、発話中の顔器官の変化量を算出するた

め、発話者の頭部全体の回転及び移動等に影響されることなく、顔器官の変化量を得ることができる。よって、顔器官の変化量をより高精度にコーパス化できる。

【０１２１】

また、音声、動画、及び顔器官の変化量のデータを同期させてコーパス化するため、音声と顔器官の変化量との対応関係を詳細に得ることができる。そのため、音声言語処理技術において確立している種々の手法を、発話中の顔器官の変化量に関する処理に適用することができる。

【０１２２】

さらに、コーパスを構成するデータに対応する顔器官の位置は一定に保たれる。よって、当該コーパスの利用が容易になる。

【０１２３】

本実施の形態のアニメーション作成装置は、発話者の顔部位の計測データを基に構築されたマルチモーダルコーパスに基づき、発話中の顔器官の変化量をモデルに割当てることにより、アニメーションを作成する。よって、動画像を用いた手法と同様に自然なアニメーションを作成することができる。

【０１２４】

また、本実施の形態のアニメーション作成装置は、モデルベースでアニメーションを作成するため、バリエーションに富んだアニメーションの作成が可能となる。

【０１２５】

さらに、本実施の形態のアニメーション作成装置は、マルチモーダルコーパスを基に、発話中の顔器官の変化量をモデルに割当てることによりアニメーションを作成する。顔器官の特徴点は事前にルールとして定められている。したがって、どのようなモデルに対しても、当該モデルにおける特徴点をルールにしたがい指定するだけで、モデルを発話時の音声及び各器官の動きに適切に同期した自然なアニメーションを作成できる。よって、手軽に高度なアニメーションを作成することができる。

［コンピュータによる実現及び動作］

なお、本実施の形態のマルチモーダルコーパス作成装置１０８及びアニメーション作成装置１１４は、コンピュータハードウェアと、そのコンピュータハードウェアにより実行されるプログラムと、コンピュータハードウェアに格納されるデータとにより実現される。図１２はこのコンピュータシステム５００の外観を示し、図１３はコンピュータシステム５００の内部構成を示す。

【０１２６】

図１２を参照して、このコンピュータシステム５００は、ＦＤ（フレキシブルディスク）ドライブ５２２及びＣＤ－ＲＯＭ（コンパクトディスク読出専用メモリ）ドライブ５２０を有するコンピュータ５１０と、キーボード５１６と、マウス５１８と、モニタ５１２とを含む。

【０１２７】

図１３を参照して、コンピュータ５１０は、ＦＤドライブ５２２及びＣＤ－ＲＯＭドライブ５２０に加えて、ハードディスク５２４と、ＣＰＵ（中央処理装置）５２６と、ＣＰＵ５２６、ハードディスク５２４、ＦＤドライブ５２２、及びＣＤ－ＲＯＭドライブ５２０に接続されたバス５３６と、ブートアッププログラム等を記憶する読出専用メモリ（ＲＯＭ）５２８と、バス５３６に接続され、プログラム命令、システムプログラム、及び作業データ等を記憶するランダムアクセスメモリ（ＲＡＭ）５３０とを含む。コンピュータシステム５００はさらに、プリンタ５１４を含んでいる。コンピュータ５１０はさらに、データ処理装置１３４（図１参照）及びバス５３６に接続されたデータインタフェース５４０と、録音装置１４４（図１参照）及びバス５３６に接続されたメディアコンバータ５４２と、カムコーダ１５０（図１参照）及びバス５３６に接続されたキャプチャカード５４４とを含む。

【０１２８】

ここでは示さないが、コンピュータ５１０はさらにローカルエリアネットワーク（ＬＡ

N）への接続を提供するネットワークアダプタボードを含んでもよい。

【０１２９】

コンピュータシステム５００にマルチモーダルコーパス作成装置１０８又はアニメーション作成装置１１４の機能を実現させるためのコンピュータプログラムは、ＣＤ－ＲＯＭドライブ５２０又はＦＤドライブ５２２に挿入されるＣＤ－ＲＯＭ５３２又はＦＤ５３４に記憶され、さらにハードディスク５２４に転送される。又は、プログラムは図示しないネットワークを通じてコンピュータ５１０に送信されハードディスク５２４に記憶されてもよい。プログラムは実行の際にＲＡＭ５３０にロードされる。ＣＤ－ＲＯＭ５３２から、ＦＤ５３４から、又はネットワークを介して、直接にＲＡＭ５３０にプログラムをロードしてもよい。

【０１３０】

このプログラムは、コンピュータ５１０にこの実施の形態のマルチモーダルコーパス作成装置１０８又はアニメーション作成装置１１４の機能を実現させるための複数の命令を含む。この機能を実現させるのに必要な基本的機能のいくつかはコンピュータ５１０上で動作するオペレーティングシステム（ＯＳ）又はサードパーティのプログラム、若しくはコンピュータ５１０にインストールされる各種ツールキットのモジュールにより提供される。したがって、このプログラムはこの実施の形態のシステム及び方法を実現するのに必要な機能全てを必ずしも含まなくてよい。このプログラムは、命令のうち、所望の結果が得られるように制御されたやり方で適切な機能又は「ツール」を呼出すことにより、上記したマルチモーダルコーパス作成装置１０８又はアニメーション作成装置１１４が行なう処理を実行する命令のみを含んでいればよい。コンピュータシステム５００の動作は周知であるので、ここでは繰返さない。

【０１３１】

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

【図面の簡単な説明】

【０１３２】

【図１】マルチモーダルコーパス作成システム１００全体の構成を示す図である。

【図２】マーカが設置された状態での、発話者１０２の顔面及び首部の外観の一例を示す正面図及び側面図である。

【図３】マルチモーダルコーパス作成装置１０８の構成を示すブロック図である。

【図４】切出処理部１８６の構成を示すブロック図である。

【図５】正規化処理部１９０の構成を示すブロック図である。

【図６】アニメーション作成装置１１４の構成を示すブロック図である。

【図７】対応マーカの指定処理の制御構造を示すフローチャートである。

【図８】１フレーム分のモーションキャプチャデータ１６０により表現されるマーカの位置を模式的に示す図である。

【図９】初期顔モデル１１０の一例を示す図である。

【図１０】初期顔モデル１１０の目輪郭部周辺におけるポリゴン、仮想マーカの概要を示す図である。

【図１１】動画像における発話者１０２の顔の画像と、アニメーション１１２における顔の画像との変化を示す図である。

【図１２】本発明の実施の形態に係るマルチモーダルコーパス作成装置１０８及びアニメーション作成装置１１４の機能を実現するコンピュータシステムの外観の一例を示す図である。

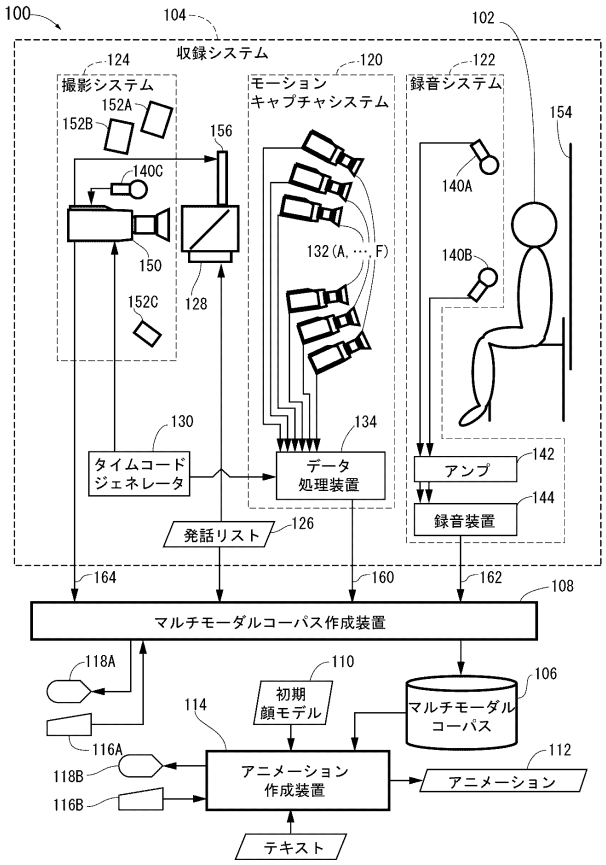
【図１３】図１２に示すコンピュータシステムのブロック図である。

【符号の説明】

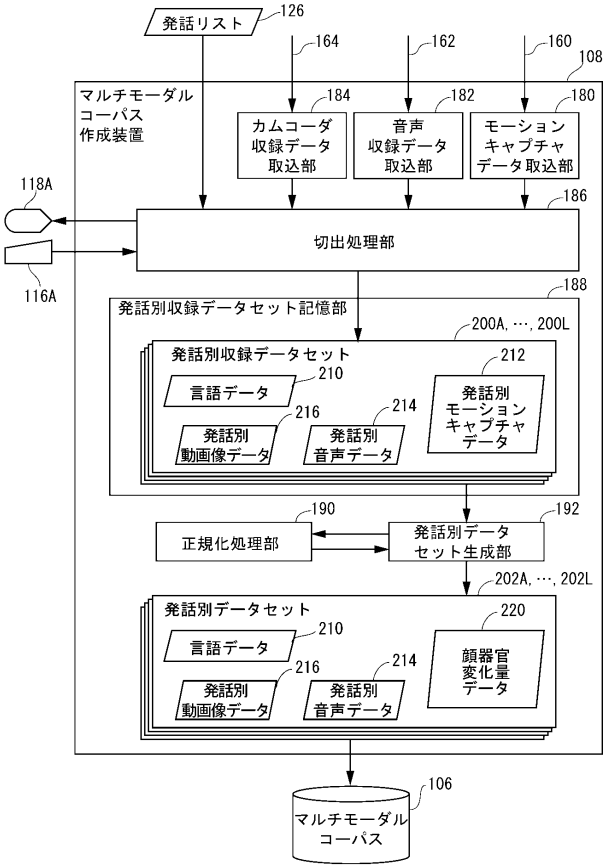
【０１３３】

１００　システム
１０４　収録システム
１０６　マルチモーダルコーパス
１０８　マルチモーダルコーパス作成装置
１１０　初期顔モデル
１１２　アニメーション
１１４　アニメーション作成装置
１２２　録音システム
１２４　撮影システム
１３４　データ処理装置
１７０Ａ，…，１７０Ｍ　マーカ
１８０　モーションキャプチャデータ取込部
１８２　音声収録データ取込部
１８４　カムコーダ収録データ取込部
１８６　切出処理部
１８８　発話別収録データセット記憶部
１９０　正規化処理部
１９２　発話別データセット生成部
２００Ａ，…，２００Ｌ　発話別収録データセット
２０２Ａ，…，２０２Ｌ　発話別データセット
２１２　発話別モーションキャプチャデータ
２１４　発話別音声データ
２１６　発話別動画像データ
２２０　顔器官変化量データ
２３０　モーションキャプチャデータ記憶部
２３２　音声収録データ記憶部
２３４　カムコーダ収録データ記憶部
２４０　動画像データ切出部
２４２　モーションキャプチャデータ切出部
２４４　同期処理部
２４６　音声データ切出部
２４８　データセット形成部
２６０　データ分類部
２６２　頭部補正用マーカデータ選択部
２６４　頭部アフィン行列算出部
２６６　頭部マーカデータ変換部
２７２　首部補正用マーカデータ選択部
２７４　首部アフィン行列算出部
２７６　首部マーカデータ変換部
２７８　データ統合部
３００　仮想マーカ設定部
３０２　マーカ対応顔モデル作成部
３０４　発話別データセット取得部
３０６　顔モデル変形部
３０８　画像化部

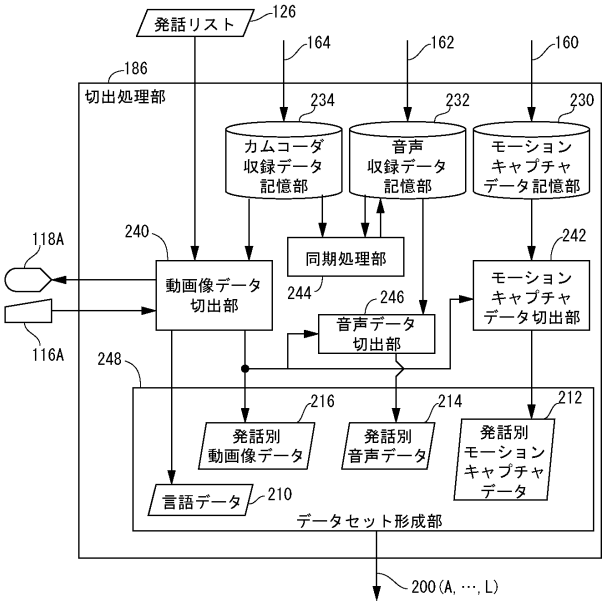
【図１】


100
104
収録システム
102
124
撮影システム
120
モーション
キャプチャシステム
122
録音システム
152A
152B
156
154
140C
140A
132(A, …, F)
140B
150
128
152C
130
134
タイムコード
ジェネレータ
データ
処理装置
アンプ
142
144
録音装置
発話リスト
126
164
160
162
108
マルチモーダルコーパス作成装置
118A
110
初期
顔モデル
106
マルチモーダル
コーパス
114
116A
118B
アニメーション
作成装置
112
アニメーション
116B
テキスト

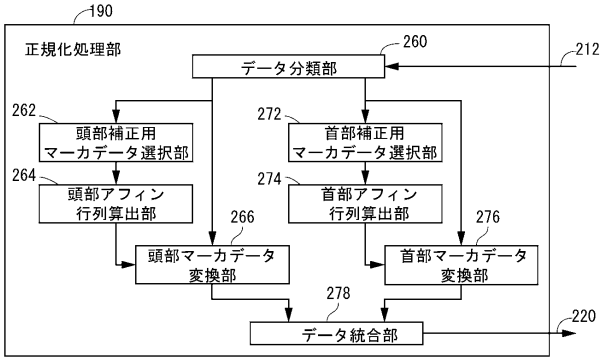
【図３】


発話リスト
126
164
162
160
108
マルチモーダル
コーパス
作成装置
184
カムコーダ
収録データ
取込部
182
音声
収録データ
取込部
180
モーション
キャプチャ
データ取込部
118A
186
切出処理部
116A
188
発話別収録データセット記憶部
200A, …, 200L
発話別収録データセット
212
言語データ
210
発話別
モーション
キャプチャ
データ
216
214
発話別
動画像データ
発話別
音声データ
190
192
正規化処理部
発話別データ
セット生成部
202A, …, 202L
発話別データセット
220
言語データ
210
顔器官
変化量
データ
216
214
発話別
動画像データ
発話別
音声データ
106
マルチモーダル
コーパス

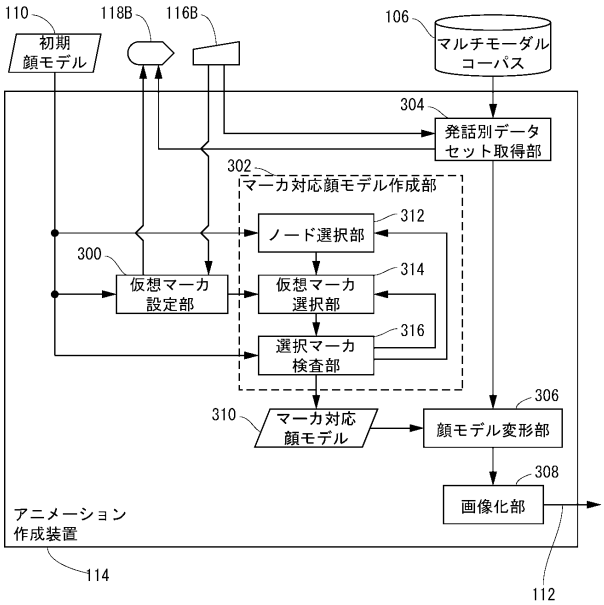
【図４】


発話リスト
126
164
162
160
186
切出処理部
234
232
230
カムコーダ
収録データ
記憶部
音声
収録データ
記憶部
モーション
キャプチャ
データ記憶部
118A
240
242
同期処理部
動画像データ
切出部
244
モーション
キャプチャ
データ切出部
246
音声データ
切出部
116A
248
216
214
212
発話別
動画像データ
発話別
音声データ
発話別
モーション
キャプチャ
データ
言語データ
210
データセット形成部
200(A, …, L)

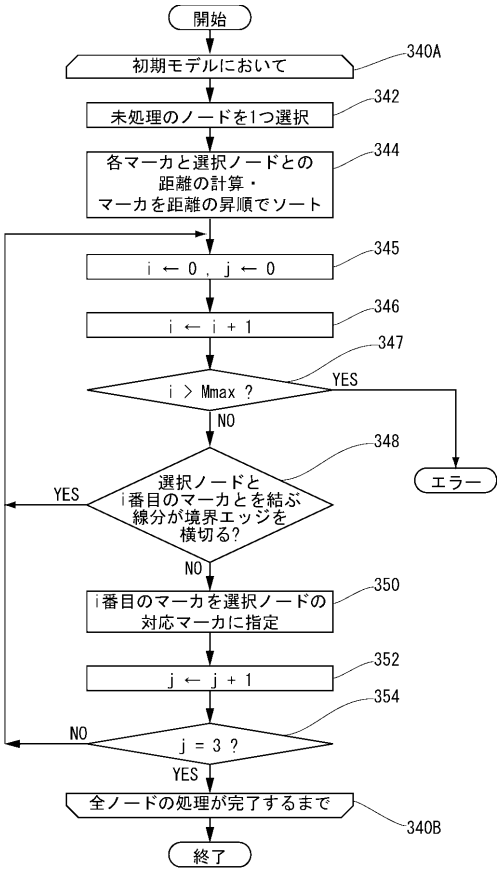
【図５】


190
正規化処理部
260
データ分類部
212
262
頭部補正用
マーカデータ選択部
272
首部補正用
マーカデータ選択部
264
頭部アフィン
行列算出部
274
首部アフィン
行列算出部
266
276
頭部マーカデータ
変換部
首部マーカデータ
変換部
278
220
データ統合部

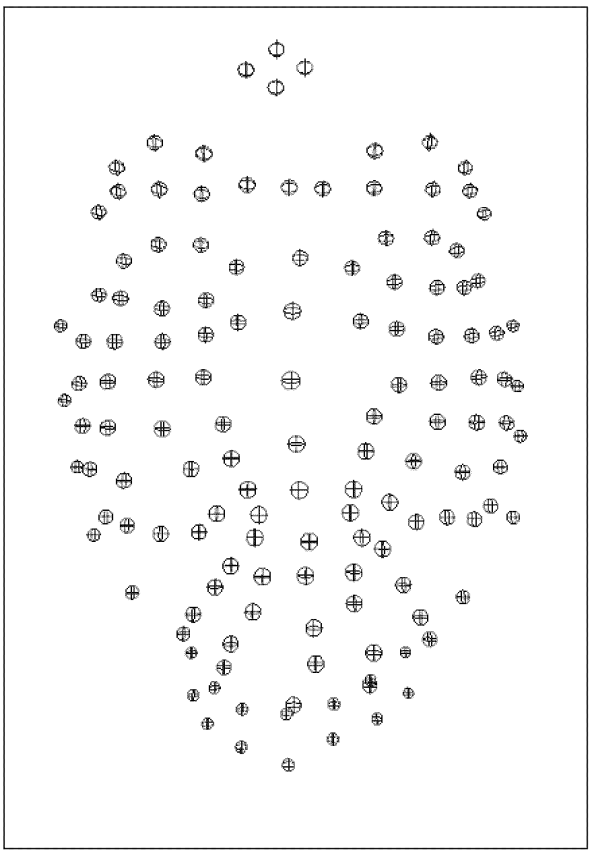
【図６】


110
初期
顔モデル
118B
116B
106
マルチモーダル
コーパス
304
発話別データ
セット取得部
302
マーカ対応顔モデル作成部
312
ノード選択部
300
仮想マーカ
設定部
314
仮想マーカ
選択部
316
選択マーカ
検査部
306
310
マーカ対応
顔モデル
顔モデル変形部
308
画像化部
アニメーション
作成装置
114
112

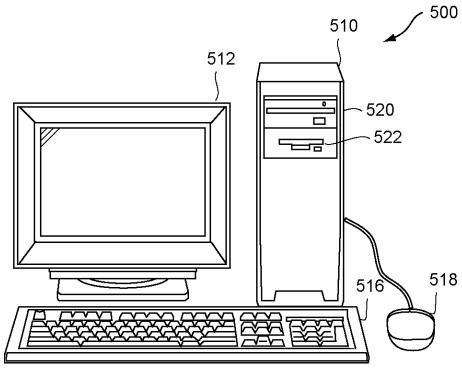
【図７】


開始
初期モデルにおいて
340A
未処理のノードを1つ選択
342
各マーカと選択ノードとの
距離の計算・
マーカを距離の昇順でソート
344
i ← 0 , j ← 0
345
i ← i + 1
346
i > Mmax ?
347
YES
NO
エラー
選択ノードと
i番目のマーカとを結ぶ
線分が境界エッジを
横切る?
348
YES
NO
i番目のマーカを選択ノードの
対応マーカに指定
350
j ← j + 1
352
j = 3 ?
354
NO
YES
全ノードの処理が完了するまで
340B
終了

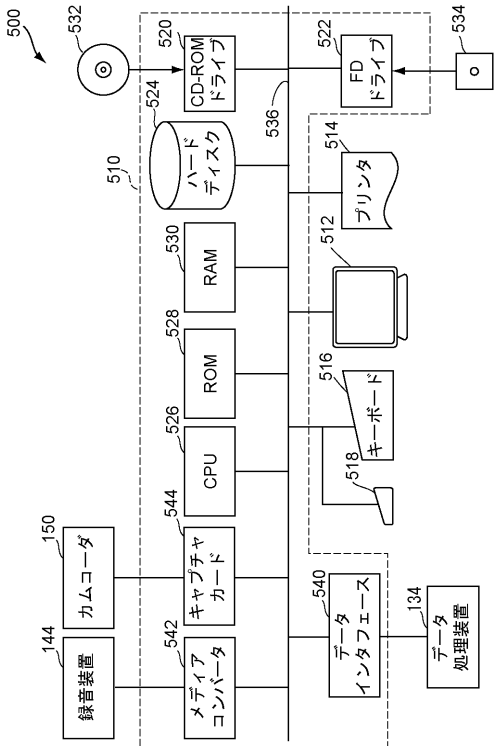
【図８】



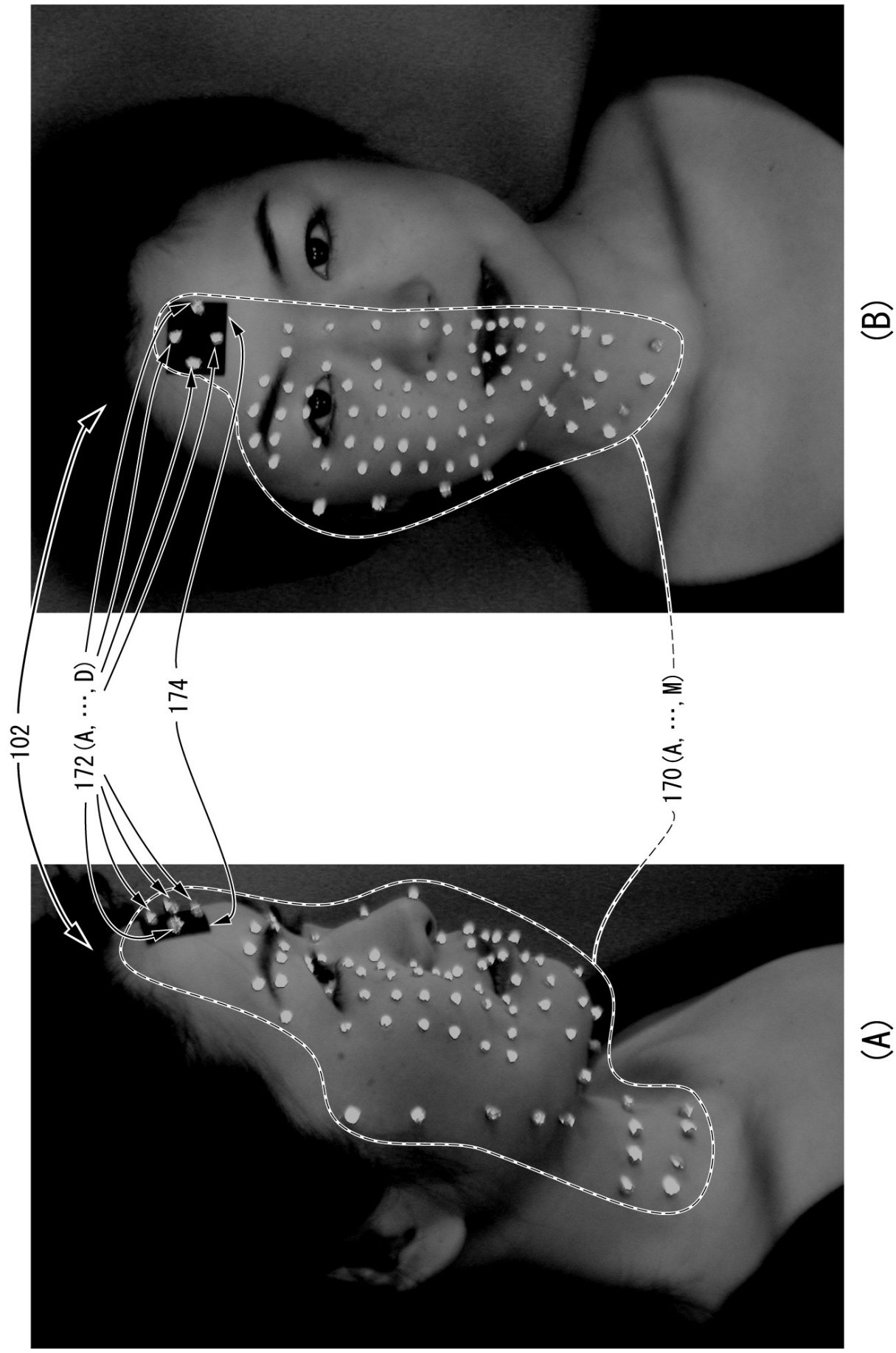
【図１２】


500
510
512
520
522
516
518

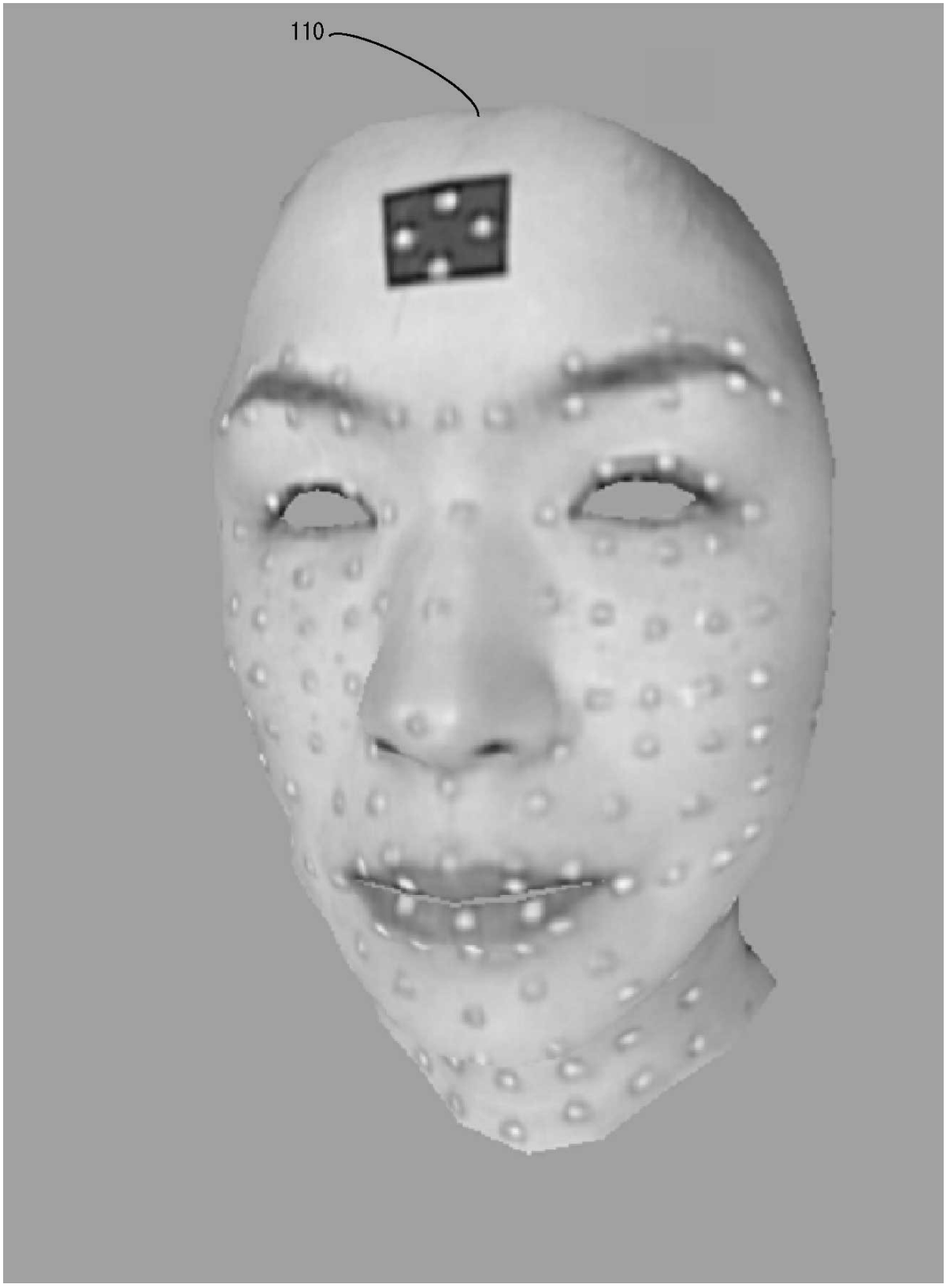
【図１３】


500
532
520
CD-ROM
ドライブ
522
FD
ドライブ
534
524
536
ハード
ディスク
510
514
プリンタ
530
RAM
512
528
ROM
516
キーボード
526
CPU
518
544
キャプチャ
カード
150
カムコーダ
542
メディア
コンバータ
144
録音装置
540
データ
インタフェース
134
データ
処理装置

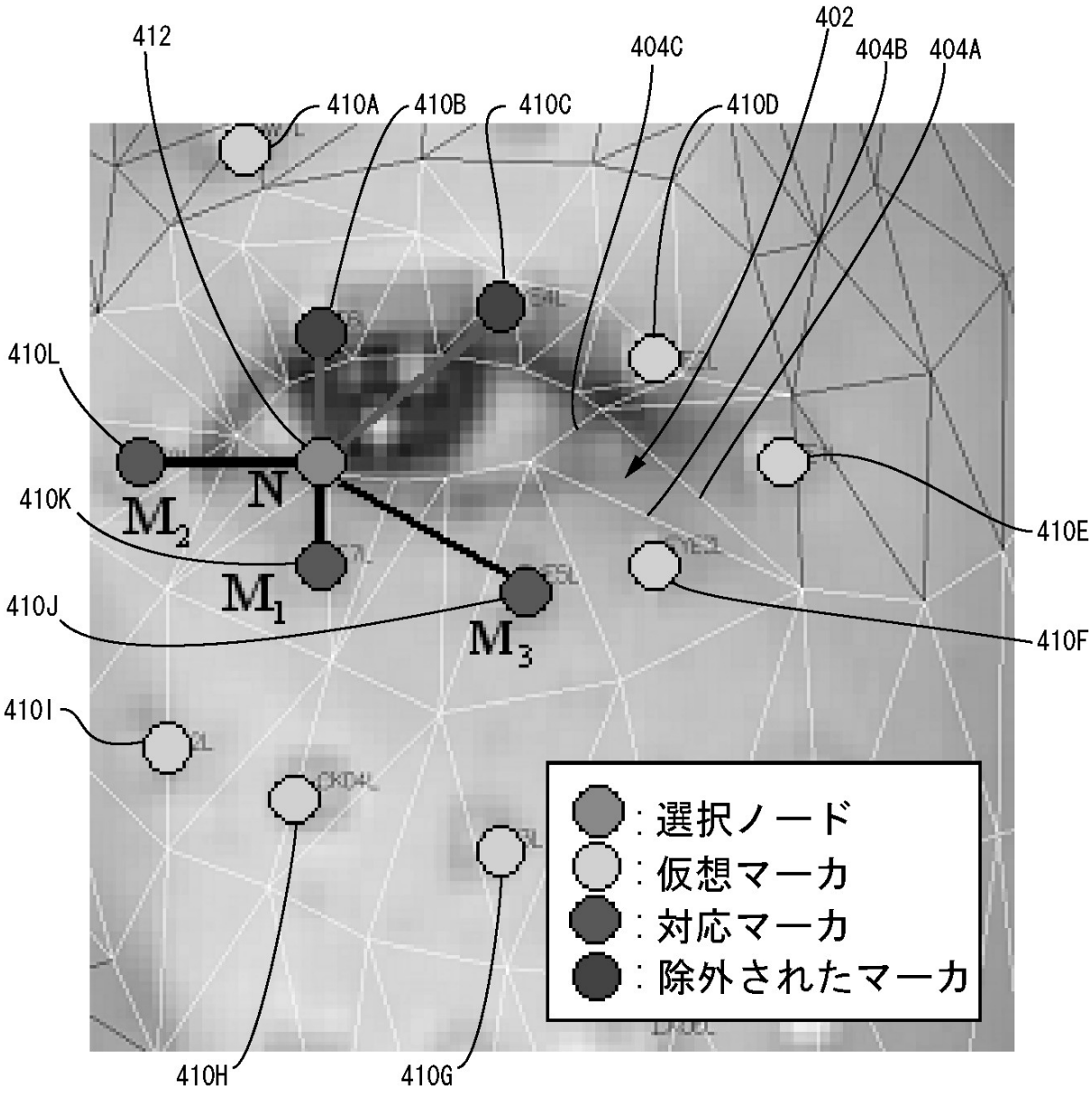
【図２】



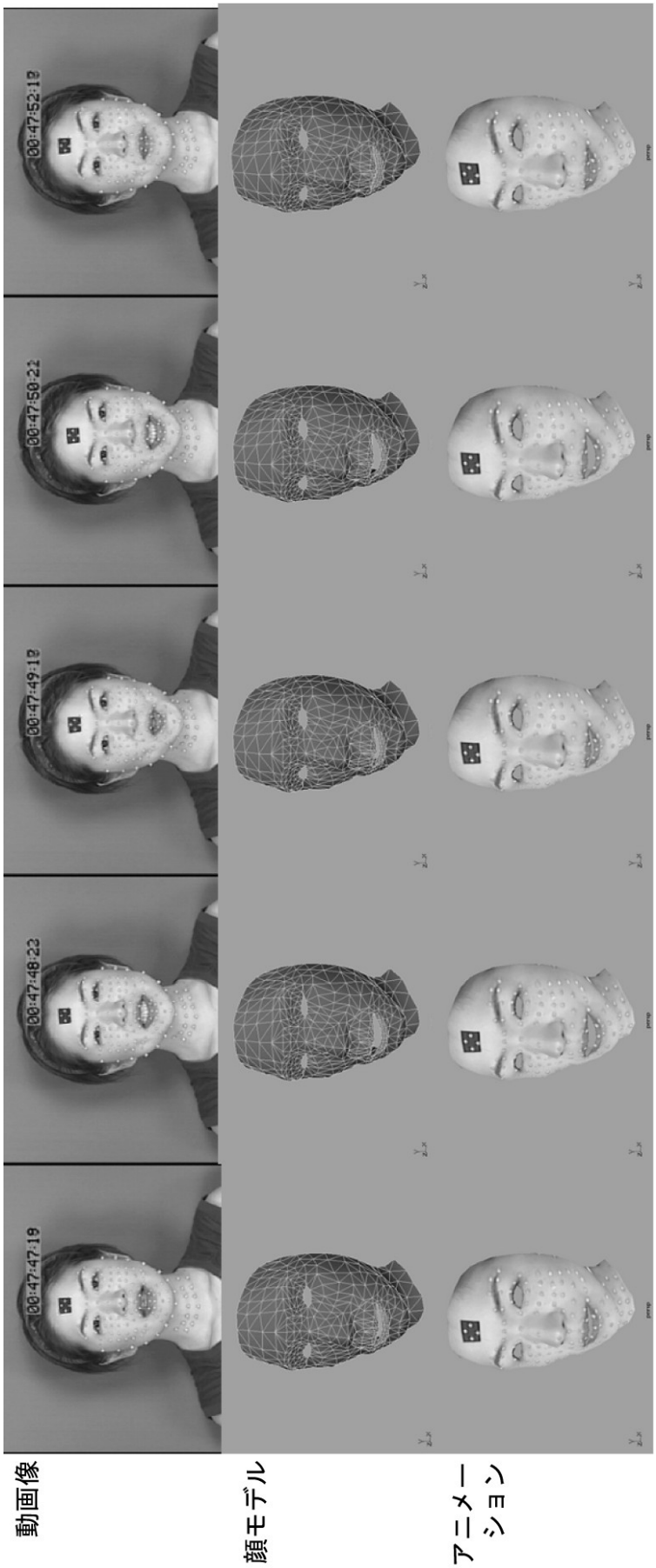
【図９】



【図１０】


412
402
404C
404B
404A
410A
410B
410C
410D
410L
410K
410J
410I
410E
410F
410H
410G
N
M2
M1
M3
：選択ノード
：仮想マーカ
：対応マーカ
：除外されたマーカ

【図１１】


動画像
顔モデル
アニメーション
00:47:19
00:47:48:23
00:47:49:18
00:47:50:21
00:47:52:19

フロントページの続き

審査官　松永　隆志

(56)参考文献　特開２０００－３１５２５９（ＪＰ，Ａ）
楊大昭、外５名，人体の行動計測による自然な人体モーションと顔表情の合成手法の検討，映像情報メディア学会技術報告　ヒューマンインフォメーション　ネットワーク映像メディア，日本，社団法人映像情報メディア学会，１９９８年　６月　１日，Ｖｏｌ．２２　Ｎｏ．２８，Ｐ65～Ｐ72

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｔ　１７／４０
Ｇ０６Ｔ　１／００