

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2005-346721  
(P2005-346721A)

(43) 公開日 平成17年12月15日(2005.12.15)

(51) Int.Cl.<sup>7</sup>  
G06T 13/00

F I  
G06T 13/00

テーマコード(参考)  
5B050

審査請求 有 請求項の数 17 O L (全 14 頁)

(21) 出願番号 特願2005-163428 (P2005-163428)  
(22) 出願日 平成17年6月3日(2005.6.3)  
(31) 優先権主張番号 093116054  
(32) 優先日 平成16年6月4日(2004.6.4)  
(33) 優先権主張国 台湾(TW)

(71) 出願人 503021065  
甲尚股▲ふん▼有限公司  
台湾台北縣新店市寶橋路235巷126号  
2樓  
(74) 代理人 100111442  
弁理士 小原 英一  
(72) 発明者 盧 澤人  
台湾 台北縣新店市寶橋路235巷126  
号2F 甲尚股▲ふん▼有限公司 内  
Fターム(参考) 5B050 BA08 EA04 EA24 FA10

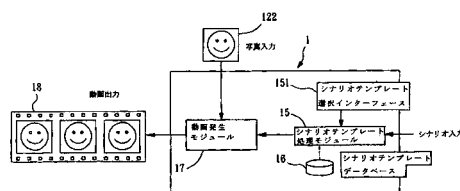
(54) 【発明の名称】 自動的に動画を生み出すシステム

(57) 【要約】

【課題】 自動的に動画を生み出すシステムと方法を提供する。

【解決手段】 主に音声変化分析のデータを通してシナリオテンプレートデータベース中の顔表情調整パラメータを自動的に対応し保存し、顔表情は時間軸に従い調整され、自動的に音声及び表情変化を具えた動画効果を生み出すシステム及び方法である。該シナリオテンプレートデータベースは複数の顔の表情の調整パラメータを含み、異なる顔表情調整パラメータの配列組合せを通して異なる表情変化シナリオを生み出すことができる。さらに、音声屈折の変化を対応させ、豊富で変化に富んだ動画効果を自動的に発生することができる。

【選択図】 図 1



## 【特許請求の範囲】

## 【請求項 1】

主に音声或いはイベント駆動を通して、使用者の選択のシナリオに基づき自動的に発生する動画であって、それはシナリオ選択インターフェース、シナリオテンプレートデータベース、シナリオテンプレート処理モジュール、動画発生モジュールから構成され、

該シナリオ選択インターフェースにより使用者はシナリオテンプレートを選択し、

該シナリオテンプレートデータベースはシナリオテンプレートデータを保存し、

該シナリオテンプレート処理モジュールは入力したシナリオに基づき選択したシナリオテンプレートデータを配置し、動画パラメータを生じ、

該動画発生モジュールは配置が完成した動画パラメータに基づき、動画データを発生することを特徴とする自動的に動画を生み出すシステム。 10

## 【請求項 2】

前記動画を生み出すシステムは、さらに特徴点識別モジュール、特徴点对応モジュール、音声分析モジュールの構成が付加され、

該特徴点識別モジュールは映像の特徴点を識別し、

該特徴点对応モジュールは該識別が完了した映像の特徴点によりメッシュデータを形成し、

該音声分析モジュールは音声データを識別及び分析することを特徴とする請求項 1 記載の自動的に動画を生み出すシステム。

## 【請求項 3】

前記動画発生モジュールは、前記配置が完成した動画パラメータに基づきメッシュデータを調整し、音声再生及び口唇データに対応し動画を生み出すことを特徴とする請求項 2 記載の自動的に動画を生み出すシステム。 20

## 【請求項 4】

前記特徴点对応モジュールは漸進的特徴点对応方式を採用し、以下のステップを含むものであって、

(a)ポートレート映像の顔の最も精密な特徴点を五官特性に基づき異なるグループに区分し、(b)前記映像により生み出した最も精密な特徴点を数個のレベルに区分し、各レベル間の対応関係を確立し、

(c)特徴点識別結果を入力し、 30

(d)該特徴点を利用しさらに精密なレベルの特徴点に調整し、

(e)ステップ(d)を繰返し、最も精密な特徴点を獲得し、

(f)最も精密な特徴点を利用し顔のメッシュデータを確立することを特徴とする請求項 2 記載の自動的に動画を生み出すシステム。

## 【請求項 5】

前記シナリオテンプレートデータは、さらに以下のデータを含むものであって、

(a)複数組の動画パートのデータは順番通りの動画を表示し、

(b)各動画パートは複数組の動画状態のデータを含み、インデックス或いは確率によって前記音声セクションに対応し、

(c)対応する各組の動画状態の動画データ、 40

(d)上記各データのデータ構造を記録し、階層に応じて配置することを特徴とする請求項 1 或いは 2 記載の自動的に動画を生み出すシステム。

## 【請求項 6】

前記シナリオテンプレート処理モジュールの処理過程は以下のステップを含むものであって、

(a)入力した音声或いはシナリオを動画パート数量と同数のセクションに分割し、

(b)シナリオテンプレートデータ中の動画パートに対応し、動画パートの手順の不変を維持し、

(c)シナリオテンプレートデータ中の動画状態に対応し、インデックス或いは確率に基づき対応し、動画パートを組成し、 50

(d)シナリオテンプレートデータ中の動画データを展開し、各動画状態が対応するキーフレームデータを展開し、結果を出力することを特徴とする請求項 1 或いは 2 記載の自動的に動画を生み出すシステム。

【請求項 7】

前記シナリオテンプレートは動態連続の顔表情或いは五官比率、位置、サイズ、回転変化の動画パラメーターであることを特徴とする請求項 1 或いは 2 記載の自動的に動画を生み出すシステム。

【請求項 8】

前記シナリオテンプレートは人の皮膚質感(テクスチャー)或いは映像色調、明暗変化の動画パラメーターであることを特徴とする請求項 1 或いは 2 記載の自動的に動画を生み出すシステム。

10

【請求項 9】

前記シナリオテンプレートは、動態連続の漫画符号効果を組合せた動画パラメーターに対応可能であることを特徴とする請求項 1 或いは 2 記載の自動的に動画を生み出すシステム。

【請求項 10】

主に以下のステップを含むものであって、

(a)動画発生モジュールを初期化し、  
(b)シナリオ選択インターフェースにより、シナリオテンプレートデータベース中より手動或いは自動でシナリオテンプレートデータを選択し、

20

(c)シナリオテンプレート処理モジュールにより、該選択したシナリオテンプレートデータに対応し、動画パラメーターを展開し、

(d)動画発生モジュールにより、該完成した動画パラメーターに基づき、メッシュデータを調整し、動画データを発生させることを特徴とする自動的に動画を生み出すシステム。

【請求項 11】

前記ステップ(a)はさらに以下のステップを含み、

(a1)あるポートレート映像を入力し、  
(a2)特徴点識別モジュールにより該ポートレート映像のポートレート特徴を識別及び定位し、

30

(a3)特徴点对応モジュールにより該識別が完了したポートレート映像の特徴点をメッシュデータに生成させることを特徴とする請求項 10 記載の自動的に動画を生み出すシステム。

【請求項 12】

前記ステップ(c)はさらに以下のステップを含むものであって、

(c1)音声データを入力、  
(c2)音声分析モジュールにより該音声データを分析し複数の音声セクションに区分し、  
(c3)シナリオテンプレート処理モジュールにより該選択したシナリオテンプレートを音声セクションに配置し、動画パラメーターを出力することを特徴とする請求項 10 記載の自動的に動画を生み出すシステム。

40

【請求項 13】

前記ステップ(c3)はさらに以下のステップを含むものであって、

(c3-1) 入力した音声セクションを動画パートの数量と同量のセクションに分割し、  
(c3-2) シナリオテンプレートデータ中の動画パートに対応し、動画パートの順序の不変を維持し、

(c3-3) シナリオテンプレートデータ中の動画状態に対応し、インデックス或いは確率により対応し、動画パートを組成し、

(c3-4) シナリオテンプレートデータ中の動画データを展開し、各動画状態がそれぞれ対応するキーフレームデータを展開し、結果を出力することを特徴とする請求項 10 記載の自動的に動画を生み出すシステム。

50

**【請求項 14】**

前記ステップの順序は(a)(b)(c)(d)を(b)(c)(a)(d)に変更可能であることを特徴とする請求項10、11、或いは12記載の自動的に動画を生み出すシステム。

**【請求項 15】**

前記シナリオテンプレートは動態連続の顔の表情の変化、五官比率、位置、回転、或いはサイズの動画パラメーターであることを特徴とする請求項10、11、或いは12記載の自動的に動画を生み出すシステム。

**【請求項 16】**

前記シナリオテンプレートは人の皮膚質感(テクスチャー)或いは映像色調、明暗変化の動画パラメーターであることを特徴とする請求項10、11、或いは12記載の自動的に動画を生み出すシステム。

10

**【請求項 17】**

前記シナリオテンプレートは動態連続の漫画符号効果を組合わせた動画パラメーターであることを特徴とする請求項10、11、或いは12記載の自動的に動画を生み出すシステム。

**【発明の詳細な説明】****【技術分野】****【0001】**

本発明は一種の自動的に動画を生み出すシステムに関する。特に、音声分析に顔の表情変化を組合せ自動的に動画を生み出すシステムと方法で、その係音声分析のデータに自動的に顔の表情を対応させパラメーターを調整し、音声及び表情の変化を具えた動画効果を生み出すシステム及び方法に係る。

20

**【背景技術】****【0002】**

従来公知の動画編集技術において、例えば、特許文献1等に開示されており、音声分析技術を利用し、対応する音声時間軸の口唇形状データを導き、該データにより映像を駆動し、擬似で話しをするキャラクターを生み出す。このような処理は自動化されてはいるが、生み出される動画は口唇形状のみであるため、表情の変化に乏しく、生命力とリアリティに欠ける。

現行の方法において表情の変化をプラスするためには、使用者は対応する音声の時間軸上において、時間毎編集(Timeline Editor)などの適当な制作ツールを利用し動画編集を行い(これはKey Frame Animation方法と呼ばれる)、表情変化の効果を生み出す必要がある。このような制作ツールでは通常は音声波形を時間軸画像表示の制作インターフェースとし、画面上である時間点を選択し、該時間点上にキーフレーム(Key Frame(動画フレーム))を加え、このキーフレーム(Key Frame(動画フレーム))の内容を編集し、変更(Transition)指定をするなどのステップを含み、前記ステップを数回繰り返した後、豊富な表情を具えた動画編集を完成することができる。通常は、制作の便のため、削除、コピーなどの基本の編集機能も該制作ツールに含んでいる。

30

**【特許文献1】**特開平11-219446号公報

**【発明の開示】**

40

**【発明が解決しようとする課題】****【0003】**

上述した従来公知の編集技術には、以下の欠点があった。

すなわち、公知方式の時間軸上で行う表情変化の編集は非常に複雑で、通常使用者は高度な動画制作の専門知識が必要である。

さらに、時間軸上で行う動画の編集には煩雑な編集ツール及び入力装置が必要で、結果が出るまでには非常に長い時間がかかる。しかも限られた入力装置(携帯電話端末など)上でのこのような機能の実現は容易ではない。

加えて、編集結果は特定の音声時間軸で行われる編集に対応するため、音声データが変化すると再度編集し直さなければならず、データを繰り返し利用することができない。

50

本発明は上記構造の問題点を解決した自動的に動画を生み出すシステムを提供するものである。

【課題を解決するための手段】

【0004】

上記課題を解決するため、本発明は、下記の自動的に動画を生み出すシステムを提供する。

それは、主に自動的に動画を生み出すシステムと方法で、特に音声分析に顔の表情変化を組合せ自動的に動画を生み出すシステムと方法で、音声分析のデータに対して自動的に顔の表情を対応させパラメータを調整し、音声及び表情の変化を具えた動画効果を生み出すシステム及び方法で、

さらに、音声或いはイベント駆動を通して、シナリオテンプレートを当てはめるシステム及び方法で、音声入力後、使用者は好みの「シナリオ」(Scenario)を選択するだけで、豊富な表情を具えた動画を自動的に生じることができる。ここで、イベント駆動とは、特徴的な出来事やイベントを分割点と見なし、シナリオテンプレート処理モジュールのパート対応を行うオペレーションの方式である。

また、シナリオテンプレートデータベースを提供し、シナリオテンプレートは元々のキーフレーム(Key Frame(動画フレーム))中の顔の表情調整パラメータデータを異なるシナリオに基づき分類し、分類後のデータはシナリオテンプレートを形成し、シナリオテンプレートデータベース中に保存し、使用者がシナリオを選択後、本発明システム及び方法は入力された音声に対して分析を行い、異なる特性のパートを探し、選択されたシナリオに基づき異なる動画変化を当てはめ、こうして同様のテンプレートは長さが異なる音声に運用され、加えて、それは簡単な動画発生システム及び方法を提供し、該システム及び方法を通して、使用者は写真を入力し、音声を入力し、テンプレートを選択するだけで豊富な動画出力を完成することができ、限られた入力装置で頻繁に使用する状況下(携帯電話端末でメールを送信するなど)での操作使用に非常に適していることを特徴とする自動的に動画を生み出すシステムである。

【0005】

すなわち、請求項1の発明は、主に音声或いは出来事の動きを通して、使用者の選択のシナリオに基づき自動的に発生する動画であって、それはシナリオ選択インターフェース、シナリオテンプレートデータベース、シナリオテンプレート処理モジュール、動画発生モジュールから構成され、該シナリオ選択インターフェースにより使用者はシナリオテンプレートを選択し、該シナリオテンプレートデータベースはシナリオテンプレートデータを保存し、該シナリオテンプレート処理モジュールは入力したシナリオに基づき選択したシナリオテンプレートデータを配置し、動画パラメータを生じ、該動画発生モジュールは配置が完成した動画パラメータに基づき、動画データを発生させることを特徴とする自動的に動画を生み出すシステムである。

【0006】

請求項2の発明は、前記動画を生み出すシステムは、さらに特徴点識別モジュール、特徴点对応モジュール、音声分析モジュールの構成が付加され、該特徴点識別モジュールは映像の特徴点を識別し、該特徴点对応モジュールは該識別が完了した映像の特徴点によりメッシュデータを形成し、該音声分析モジュールは音声データを識別及び分析することを特徴とする請求項1記載の自動的に動画を生み出すシステムである。

請求項3の発明は、前記動画発生モジュールは、前記配置が完成した動画パラメータに基づきメッシュデータを調整し、音声再生及び口唇データに対応し動画を生み出すことを特徴とする請求項2記載の自動的に動画を生み出すシステムである。

請求項4の発明は、前記特徴点对応モジュールは漸進的特徴点对応方式(Progressive Geometry Construction)を採用し、以下のステップを含むものであって、

(a)ポートレート映像の顔の最も精密な特徴点を五官特性に基づき異なるグループに区分し、

(b)前記映像により生み出した最も精密な特徴点を数個のレベル(水準(Level))毎に区

10

20

30

40

50

分し、各レベル間の対応関係を確立し、

(c)特徴点識別結果を入力し、

(d)該特徴点を利用しさらに精密なレベルの特徴点に調整し、

(e)ステップ(d)を繰返し、最も精密な特徴点を獲得し、

(f)最も精密な特徴点を利用し顔のメッシュデータを確立することを特徴とする請求項 2 記載の自動的に動画を生み出すシステムである。

請求項 5 の発明は、前記シナリオテンプレートデータは、さらに以下のデータを含むものであって、

(a)複数組の動画パートのデータは順番通りの動画を表示し、

(b)各動画パートは複数組の動画状態のデータを含み、インデックス或いは確率によって前記音声セクションに対応し、

(c)対応する各組の動画状態の動画データ、

(d)上記各データのデータ構造を記録し、階層に応じて配置することを特徴とする請求項 1 或いは 2 記載の自動的に動画を生み出すシステムである。

#### 【0007】

請求項 6 の発明は、前記シナリオテンプレート処理モジュールの処理過程は以下のステップを含むものであって、

(a)入力した音声或いはシナリオを動画パート数量と同数のセクションに分割し、

(b)シナリオテンプレートデータ中の動画パートに対応し、動画パートの手順の不変を維持し、

(c)シナリオテンプレートデータ中の動画状態に対応し、インデックス或いは確率に基づき対応し、動画パートを組成し、

(d)シナリオテンプレートデータ中の動画データを展開し、各動画状態が対応するキーフレームデータを展開し、結果を出力することを特徴とする請求項 1 或いは 2 記載の自動的に動画を生み出すシステムである。

請求項 7 の発明は、前記シナリオテンプレートは動態連続の顔表情或いは五官比率、位置、サイズ、回転変化の動画パラメーターであることを特徴とする請求項 1 或いは 2 記載の自動的に動画を生み出すシステムである。

請求項 8 の発明は、前記シナリオテンプレートは人の皮膚質感(テクスチャー)或いは影像色調、明暗変化の動画パラメーターであることを特徴とする請求項 1 或いは 2 記載の自動的に動画を生み出すシステムである。

#### 【0008】

請求項 9 の発明は、前記シナリオテンプレートは動態連続の漫画符号効果を組合せた動画パラメーターを対応可能であることを特徴とする請求項 1 或いは 2 記載の自動的に動画を生み出すシステムである。

請求項 10 の発明は、主に以下のステップを含むものであって、

(a)動画発生モジュールを初期化し、

(b)シナリオ選択インターフェースにより、シナリオテンプレートデータベース中より手動或いは自動でシナリオテンプレートデータを選択し、

(c)シナリオテンプレート処理モジュールにより、該選択したシナリオテンプレートデータに対応し、動画パラメーターを展開し、

(d)動画発生モジュールにより、該完成した動画パラメーターに基づき、メッシュデータ調整し、動画データを発生させることを特徴とする自動的に動画を生み出すシステムである。

#### 【0009】

請求項 11 の発明は、前記ステップ(a)はさらに以下のステップを含み、

(a1)あるポートレート影像を入力し、

(a2)特徴点識別モジュールにより該ポートレート影像のポートレート特徴を識別及び定位し、

(a3)特徴点对応モジュールにより該識別が完了したポートレート影像の特徴点をメッシュ

データに生成させることを特徴とする請求項 10 記載の自動的に動画を生み出すシステム。

請求項 12 の発明は、前記ステップ(c)はさらに以下のステップを含むものであって、  
 (c1)音声データを入力、  
 (c2)音声分析モジュールにより該音声データを分析し複数の音声セクションに区分し、  
 (c3)シナリオテンプレート処理モジュールにより該選択したシナリオテンプレートを音声セクションに配置し、動画パラメータを出力することを特徴とする請求項 10 記載の自動的に動画を生み出すシステムである。

【0010】

請求項 13 の発明は、前記ステップ(c3)はさらに以下のステップを含むものであって、  
 (c3-1) 入力した音声セクションを動画パートの数量と同量のセクションに分割し、  
 (c3-2) シナリオテンプレートデータ中の動画パートに対応し、動画パートの順序の不变を維持し、  
 (c3-3) シナリオテンプレートデータ中の動画状態に対応し、インデックス或いは確率により対応し、動画パートを組成し、  
 (c3-4) シナリオテンプレートデータ中の動画データを展開し、各動画状態がそれぞれ対応するキーフレームデータを展開し、結果を出力することを特徴とする請求項 10 記載の自動的に動画を生み出すシステムである。

【0011】

請求項 14 の発明は、前記ステップの順序は(a)(b)(c)(d)を(b)(c)(a)(d)に変更可能であることを特徴とする請求項 10、11、或いは 12 記載の自動的に動画を生み出すシステムである。

請求項 15 の発明は、前記シナリオテンプレートは動態連続の顔の表情の変化、五官比率、位置、回転、或いはサイズの動画パラメータであることを特徴とする請求項 10、11、或いは 12 記載の自動的に動画を生み出すシステムである。

請求項 16 の発明は、前記シナリオテンプレートは人の皮膚質感(テクスチャー)或いは影像色調、明暗変化の動画パラメータであることを特徴とする請求項 10、11、或いは 12 記載の自動的に動画を生み出すシステムである。

請求項 17 の発明は、前記シナリオテンプレートは動態連続の漫画符号効果を組合わせた動画パラメータであることを特徴とする請求項 10、11、或いは 12 記載の自動的に動画を生み出すシステムである。

【発明の効果】

【0012】

上記のように、本発明は音声分析のデータに自動的に顔の表情を対応させパラメータを調整し、音声及び表情の変化を具えた動画効果を生み出すことができる。さらに本発明では音声入力後、使用者は好みの「シナリオ」(Scenario)を選択するだけで、豊富な表情を具えた動画を自動的に生じることができる。また本発明は簡単なシステム及び方法であるため、使用者は写真入力、音声入力、テンプレート選択のステップを踏むだけで豊富な動画出力を完成することができ、限られた入力装置で頻繁に使用する状況下での操作に適している。

【発明を実施するための最良の形態】

【0013】

図 1 に示すように、本発明の自動的に動画を生み出すシステム 1 は、シナリオ選択インターフェース 151、シナリオテンプレートデータベース 16、シナリオテンプレート処理モジュール 15、動画発生モジュール 17 からなる。該シナリオテンプレート選択インターフェース 151 により使用者はシナリオテンプレートを選択する。該シナリオテンプレートデータベース 16 により使用者はシナリオテンプレートデータを保存する。該シナリオテンプレート処理モジュール 15 は使用者が選択したシナリオテンプレートデータを処理する。該動画発生モジュール 17 により、使用者はシナリオテンプレート及びポートレートデータを配置しキーフレーム(Key Frame)データを完成し、動画データを生じる。

10

20

30

40

50

先ず、本発明の動画発生モジュール初期化により動画状態を生じ、動画発生モジュール初期化後、使用者は該シナリオ選択インターフェース151を用い、該シナリオテンプレートデータベース16中からシナリオテンプレートを選択する。

次に、該オリジナルポートレイト映像122及び該選択したシナリオテンプレートを、該シナリオテンプレート処理モジュール15により処理し、最後に本発明の動画発生モジュール17は該シナリオテンプレートの配置データにより、動画データ18を生じる。

#### 【0014】

次に、図2に、本発明の別の実施例のシステム構造を示すが、本発明の自動的に動画を生み出すシステム1はさらに特徴点検出モジュール(Feature Detection Module)12、特徴点对応モジュール(Geometry Construction Module)13、音声分析モジュール(Audio Analysis Module)14からなる。 10

本発明の自動的に動画を生み出すシステム外部の該映像読み取りユニットは、先ずオリジナルポートレイト映像121を読み取る。該オリジナルポートレイト映像121の読み取り後、該特徴点検出モジュール12中に入力し、特徴点の識別を行う。識別完成後、関連するポートレイト特徴は定位を完了される。

続いて、該特徴点对応モジュール(Geometry Construction Module)13は、特徴点検出モジュールを利用し発生する特徴点に対して予め内蔵する1組の一般的メッシュ(Generic Mesh)131により比較調整を行い、動画処理を行うことができるメッシュデータ131とする。ここで、メッシュデータとは特徴点を線で繋いだ(図4に示すような)網の目(mesh)のデータである。 20

#### 【0015】

図3に示すように、本システムは漸進的特徴点对応方式(Progressive Geometry Construction)を採用する。この方式は、特徴点を五官特性に基づきグループに区分し、さらに画質に応じて数個のレベル(Level)に区分し、各レベル間の対応関係を確立するものである。また、一般的メッシュ(Generic Mesh)も特徴点(Feature Point)と対応し区分され、処理時には該特徴点(Feature Point)は対応する一般的メッシュ(Generic Mesh)の調整を担当し、絶え間ない調整演算により正確なメッシュ出力を得ることができる。上記調整演算は、演算リソースが十分なシステム(ディスクトップPCなど)において執行するなら、特徴点は完全に検出され、精密な結果を得ることができる。

しかし、演算リソースが限られたハンドヘルドデバイス(携帯電話端末及びPDAなど)で実行し、低いレベルの検出しかなされなくとも、やはり近似した結果を得ることができる。 30

実際の応用状況においては、前者はコンテンツプロバイダーが提供するデータで、後者は使用者がハンドヘルドデバイス上でリアルタイムに操作し得られるものである。該オリジナルポートレイト映像121は該特徴点検出モジュール12及び該特徴点对応モジュール13により処理され、図4に示すような結果を生じる。

#### 【0016】

本発明の音声分析モジュール14(図2参照)は公知の技術により制作される音声識別ユニット、及び音声特性を分析する特性分析ユニットを含む。

使用者は、一定時間の音声データを録音することができ、該音声分析モジュール14により音声の識別及び分析を行う。該音声分析ユニットは入力された音声を音声表記に識別し、各音声表記はそれが発生した時間を含む。該特性分析ユニット音声の特性に基づき、音声を異なる特性のセクション区分し、該セクションの特性データ(音声強度など)及び時間情報(開始時間、音声の長さなど)を記録する。音声識別及び分析を経た結果は図5に示す。 40

図5に示すように、音声データは該音声分析モジュール14(図2参照)により識別を完了後、5個の音声屈折点41、42、43、44、45に区分され、ある人のある状況下(怒り、喜びなど)での話し声の音声変化の状態を表す。

該音声データが該音声分析モジュールにより処理され、数個の特性データを含む音声セクションに区分された後(図6参照)、本発明のシナリオ処理モジュールは音声セクション 50



とシナリオテンプレート中のデータの対応(match)を行う。

【0017】

図7に示すように、シナリオテンプレートデータは動画パート(Animation Part) 061、動画状態(Animation State) 062、動画データ(Animation Parameter) 063の3個の主要な階層に区分される。

該動画パート061では動画の手順性を表示し、1個の該動画パート061は1個或いは1個以上の音声セクションに対応可能である。

該動画状態062ではそれが所属する動画パート061を構成し、該動画パート061中の1個の動画状態062は1個の音声セクションにのみ対応可能であるが、繰返し出現可能で、該動画状態062中にはインデックスを含む。

該動画データ063ではそれが所属する動画状態062が位置する相対時間軸上のキーフレームデータ(Key Frame Data)を表示し、動画発生モジュールを駆動可能な動画データを生じる。

図8は「喜びが極まって泣く」シナリオテンプレートの構造を示す。シナリオテンプレート処理モジュールは4項の主要ステップを通して、シナリオテンプレートと音声セクションの対応を行う。該3項とは分割音声セクション、動画パート対応、動画状態対応、動画データ展開で、その過程は図9に示す。

動画パート対応はシナリオテンプレート中の動画パートの数量に基づき、先ず音声セクションに等量分割を行い、音声セクションのエネルギーの差異を計算する。次に分割点を移動させ改めて音声セクションのエネルギーの差異を計算し、エネルギーの最大差異を得るまで繰返し演算する。この時の分割点は最良の分割点と見なされ、対応処理の結果動画パート手順は不変で、かつ分割点は最良の位置に位置する。

【0018】

さらに「喜びが極まって泣く」シナリオテンプレートの動画パート対応の状況を表す図10に示すように、それは「喜」と「泣」の2組の動画パート、091等量分割による対応結果表示、092最良の分割後の対応結果取得表示からなる。

動画状態対応は各1組の動画パート中の動画状態に対して処理を行う。その目的は動画パート中の各1個の音声セクションをすべて1個の動画状態に対応させ、かつ動画状態を繰返し出現可能とすることである。処理方式はインデックス音声特性により分析された確率などの方法に基づく。

【0019】

続いて「喜びが極まって泣く」動画状態対応結果を示す図11で説明する101は対応が完成した動画パート、102はインデックスに基づき対応した動画状態、103は音声特性により対応した確率対応の動画状態である。

動画データの展開は対応後の動画状態を時間軸上の動画キーフレームに転換する。シナリオテンプレート中の各1個の動画状態はすべて相対時間軸上に位置する、ある段階の動画トラック(Animation Track)を含む。該段階の動画が繰返し標記されていないかどうかは、動画状態対応後に、それが表示する動画トラックを対応する音声セクション開始時間まで移動させ、該段動画データを完成することができる。さらに、該動画データが繰返し標記されていないかどうかに基づき、動画データを音声セクションに繰返しコピーし終了する。

【0020】

前記のように、本発明シナリオテンプレート処理モジュール(Scenario Template...)の機能はポートレイト映像と音声データを適当に対応(match)させ、動画を生じるものである。

該シナリオテンプレート(Scenario Template)は通常用いられる一種のテンプレート(Template)で、ある特定の顔の表情動画シナリオの表示に用いる。それは動画パート(Animation Part)、動画状態(Animation State)、動画データ(Animation Parameter)を含む。該シナリオテンプレート(Scenario Template)はまたツールが予め制作したデータを利用し、シナリオテンプレートデータベース(Scenario Template Database)中或いは一般の保存

10

20

30

40

50

装置中に保存することができ、テンプレートによりインターフェース151を選択後、本発明のシステム中において使用することができる。実際の状況では、異なる応用ニーズに応じて異なるシナリオテンプレートを設計し、その数量は応用状況に応じて決定する。また、シナリオテンプレート(Scenario Template)はネットワーク(インターネットなど)或いは他の伝送方式(携帯電話端末など)を利用し応用の設備中にダウンロードしデータを拡充可能なシステムを達成する。

#### 【0021】

ポートレート映像データと音声データが上記手順処理を経て本発明の動画発生モジュールに入力されると、最終の動画映像を生じる。

本発明の動画発生モジュールは2D或いは3Dのモジュールとすることができ、音声再生及びキーフレームデータ(Key frame Data)に応じて動画出力を行う。 10

#### 【0022】

本発明の音声駆動の自動表情動画発生システム中の各実行ユニット相互間のシステム関係について説明するために、本発明の音声駆動の自動表情動画発生システムの操作過程について以下に記載する。

図12の本発明システム操作フローチャートに示すように、本発明の音声駆動の自動表情動画発生システムは外部の映像読み取りユニットによりオリジナルポートレート映像(ステップ111)を読み取り、該オリジナルポートレート映像は読み取り後に本発明システム中の特徴点検出モジュール(Feature Detection Module)中に入力され、特徴点の識別(ステップ112)を行い、識別完成後は、関連するポートレート特徴の定位を完了される。 20

続いて、本発明中の特徴点对応モジュール(Geometry Construction Module)は特徴点検出モジュールが発生する特徴点を利用し、1組の既に予め内蔵する一般的メッシュ(Generic Mesh)に対してデータ比較調整を行い、動画処理を行うことができるメッシュ(mesh)データ(ステップ113)とする。

#### 【0023】

上記オリジナルポートレート映像識別手順処理の前、後、或いは同時に、使用者は一定時間の音声データを録音することができ、本発明の音声分析モジュールにより音声の識別及び分析(ステップ114)を行う。音声分析ユニットは入力された音声を音声表記に識別し、各音声表記はそれが発生した時間を含む。特性分析ユニットは音声の特性に基づき、音声を異なる特性のセクションに区分する。該セクションの時間情報を含む。 30

ポートレート映像の特徴点検出及び特徴点对応の処理手順処理が完了し、かつ音声データの音声分析モジュールによる識別及び分析が完了すると、処理が完了したポートレート映像データ及び音声データは本発明シナリオテンプレート処理モジュール(Scenario Template Module)に入力される。

本発明シナリオテンプレート処理モジュールのシナリオテンプレート(Scenario Template)は一種のテンプレート(Template)で、特定の動画シナリオの表示に用いる。この手順においては、使用者は手動或いは自動的方式で、シナリオテンプレートデータベース中(Senario Template Database)より特定のシナリオを選択し、選択されたシナリオは自動的に識別が完了した音声データに基づき対応(Distribution)の処理(ステップ115)を行う。

例えば、使用者が「喜びが極まって泣く」のシナリオを選択すると、本発明のシナリオテンプレート処理モジュールは自動的に音声データ中の挫折抑揚の音声変化に「喜」及び「泣」シナリオ中の顔映像を対応させパラメーターを調整し、音声再生を形成する時に同時に「喜びが極まって泣く」顔の映像変化を表示する。 40

ポートレート映像データ、顔の映像調整パラメーター、音声データが上記手順により処理後は本発明の動画発生モジュール(ステップ116)に入力され次の処理が行われ、最終の動画映像(ステップ117)が生じる。

#### 【0024】

上記システム中において、もし音声分析モジュールの音声特性データをうっかり忘れてしまったなら、3個の動画パートのみを具えたシステムに単純化することができる。該3個の動画パートはそれぞれイントロパート(Intro Part)、再生パート(Play Part)、終了 50

パート(Ending Part)で、音声開始及び音声終了を分割点とし、シナリオテンプレート処理モジュールのパート対応を行う。

この種の簡易システム中において、イントロパート及び終了パートは1個の動画状態のみを含み、かつ繰り返さない。再生パートは1個或いは数個の動画状態で、インデックス或いは再配置が可能である。このようなシステムはハンドヘルドデバイス、携帯電話などの演算リソースが限られたシステムにおいて、音声と比較的短い音声データへの応用に非常に適している。

#### 【0025】

前記システムで明らかなように、もし音声分析を行わなくても音声再生に従い豊富な顔動画の効果が発生することができる。そのオペレーションの方式はイベント駆動(Event Driven)により、すなわち、特徴的な出来事やイベントを分割点と見なし、シナリオテンプレート処理モジュールのパート対応を行うものである。

本システムのシナリオテンプレート処理モジュールは音声分析モジュールが分析し導き出す音声特性データを直接駆動動画パートのイベント(Event)とすることができる。その方法は異なる音声特性を異なる動画パートに対応させ、シナリオテンプレート過程中に当てはめることができ、音声特性を利用し直接動画を制御する目的を達成することができる。

図12、13に示すように、例えば音程が高い音声121は驚喜を表現する動画パート123に対応し、音程が低い音声122は悲しみを表現する動画パート124に対応すると、この2組の動画パートは音程の高低により制御され、対応する音声特性を分析し導き出す時、直接出現する。

#### 【0026】

音声分析モジュールは音声を分割する時にも音声の全体特性を考慮することができる。例えば音楽のリズムの分析を分割点とし、シナリオテンプレート処理モジュールを利用し、各音楽の小節に動画を当てはめる。こうしてもし該動画発生モジュールが人体全身のキャラクター動画システムを含むなら、自動的に音楽リズムに従いダンスのステップを生じる。

#### 【図面の簡単な説明】

#### 【0027】

【図1】本発明の実施例のシステム構造図である。

【図2】本発明の別の実施例のシステム構造図である。

【図3】本発明のポートレイト特徴識別実施例の指示図である。

【図4】本発明のポートレイト特徴識別実施例の別種の指示図である。

【図5】本発明の音声識別実施例の指示図である。

【図6】本発明のシナリオテンプレートと音声配置実施例の指示図である。

【図7】本発明のシナリオテンプレートの指示図である。

【図8】本発明のシナリオテンプレートの実施例指図である。

【図9】本発明のシナリオテンプレート処理モジュールのフローチャートである。

【図10】本発明のシナリオテンプレートの動画パート対応の指示図である。

【図11】本発明の別種のシナリオテンプレートの動画状態対応の指示図である。

【図12】本発明のシステムフローチャートである。

【図13】本発明のさらに別種のシナリオテンプレートの動画状態対応の指示図である。

#### 【符号の説明】

#### 【0028】

- 1 本発明
- 12 特徴点検出モジュール
- 13 特徴点对応モジュール
- 14 音声分析モジュール
- 15 シナリオテンプレート処理モジュール
- 16 シナリオテンプレートデータベース

10

20

30

40

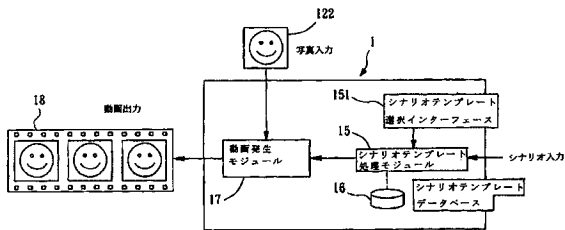
50

- 17 動画発生モジュール
- 121 オリジナルポートレート映像
- 122 オリジナルポートレート映像
- 131 一般的な顔のメッシュデータ
- 141 音声入力
- 151 シナリオテンプレート選択インターフェース
- 18 動画出力
- 41、42、43、44、45 音声屈折点
- 50、51、52、5N、5N+1 シナリオテンプレート状態
- 061 動画パート
- 062 動画状態
- 063 動画データ
- 091 対応ステップ
- 092 対応ステップ
- 101 対応ステップ
- 102 対応ステップ
- 103 対応ステップ
- 111、112、113、114、115、116、117 ステップ
- 21 音程が高い音声
- 122 音程が低い音声
- 23 驚喜の動画パート
- 124 悲しみの動画パート

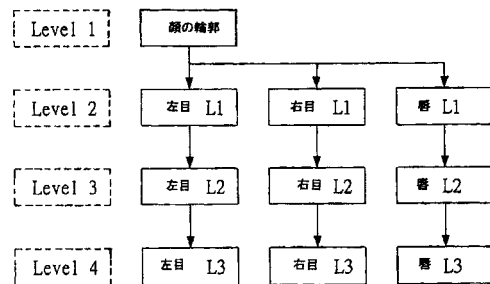
10

20

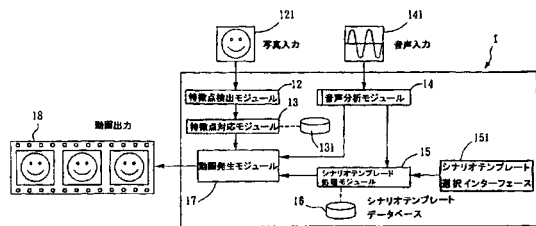
【図1】



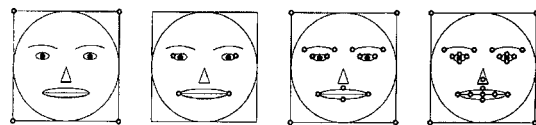
【図3】



【図2】

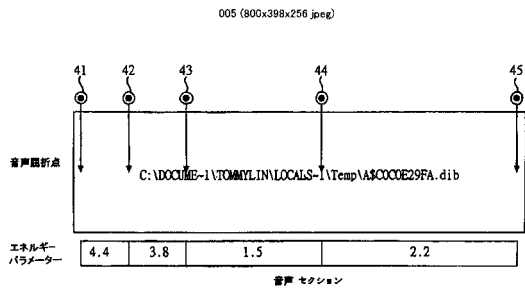


【図4】

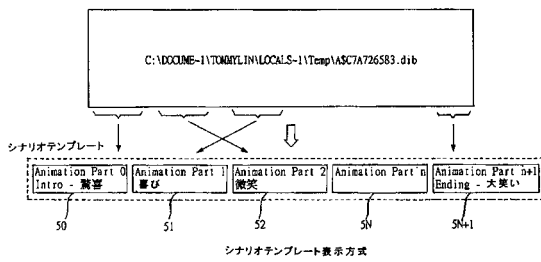


異なるグレードの特徴点定義 (Feature Point)

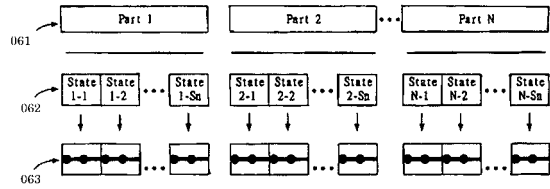
【 図 5 】



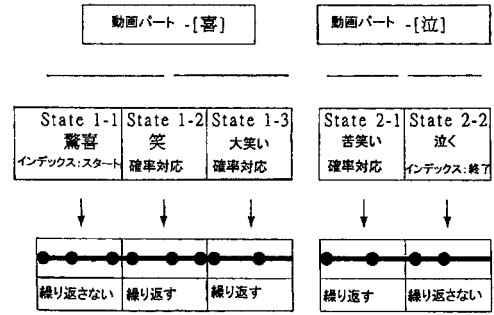
【 図 6 】



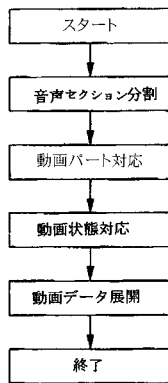
【 図 7 】



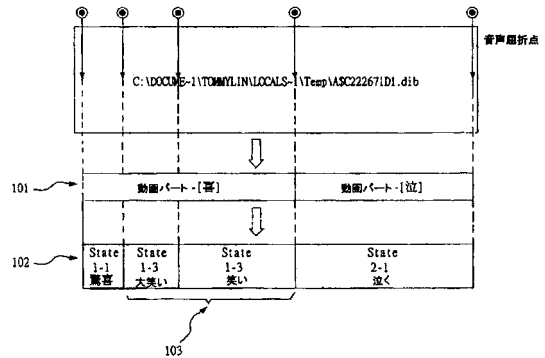
【 図 8 】



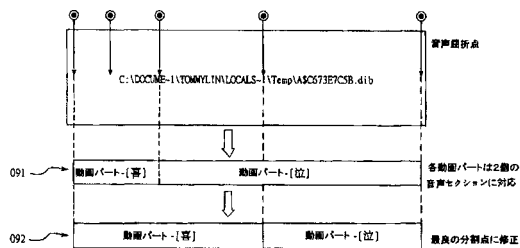
【 図 9 】



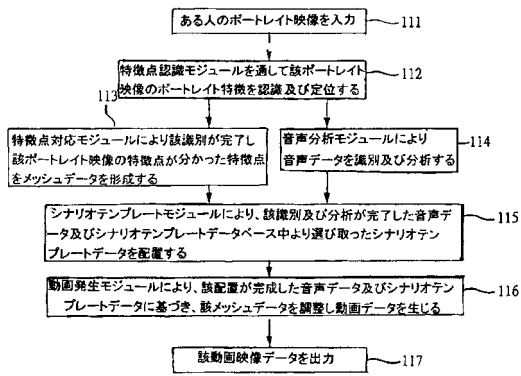
【 図 11 】



【 図 10 】



【 図 1 2 】



【 図 1 3 】

