

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5137031号
(P5137031)

(45) 発行日 平成25年2月6日(2013.2.6)

(24) 登録日 平成24年11月22日(2012.11.22)

(51) Int.Cl.	F I	
G 1 0 L 13/10	(2013.01)	G 1 0 L 13/08 1 3 0 B
G 1 0 L 13/02	(2013.01)	G 1 0 L 13/02 1 2 3
G 1 0 L 13/00	(2006.01)	G 1 0 L 13/00 1 0 0 V
G 1 0 L 21/055	(2013.01)	G 1 0 L 21/04 2 0 0 A

請求項の数 8 (全 44 頁)

(21) 出願番号	特願2008-273704 (P2008-273704)	(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10月24日(2008.10.24)	(74) 代理人	100099933 弁理士 清水 敏
(65) 公開番号	特開2009-122667 (P2009-122667A)	(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6月4日(2009.6.4)	(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25日(2011.8.25)	(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2007-278619 (P2007-278619)		
(32) 優先日	平成19年10月26日(2007.10.26)		
(33) 優先権主張国	日本国(JP)		
(出願人による申告)平成20年度、支出負担行為担当官文部科学省科学技術・学術政策局長、研究テーマ「重要課題解決型研究等の推進 新映像技術ダイブイントゥザムービーの研究」に関する委託研究、産業技術力強化法第19条の適用を受ける特許出願			
			最終頁に続く

(54) 【発明の名称】 台詞音声作成装置、発話音声収録装置、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】

各々、発話のタイミング、その長さ、及び発話者が予め定められている複数の台詞の音声を作成するための台詞音声作成装置であって、

前記複数の台詞の各々の発話タイミング、発話内容の台詞、及び発話時間を特定する台詞情報を記憶するための台詞情報記憶手段と、

ある発話者に割当てられている台詞の少なくとも一部を当該ある発話者に発話させ、その発話音声を収録するための発話音声収録手段と、

前記発話音声収録手段により収録された発話音声に基づき、前記ある発話者に割当てられている台詞の全てについて、発話のタイミング及びその長さが前記台詞情報記憶手段に記憶された前記台詞情報に合致するように台詞音声を作成するための台詞音声作成手段とを含み、

前記台詞音声作成手段は、

各々、前記発話音声収録手段により収録された前記発話音声に基づいて、発話のタイミング及びその長さが前記台詞情報による指定と合致するように前記ある発話者の台詞音声を生成するための複数の台詞音声生成手段と、

前記複数の台詞の各々について、前記複数の台詞音声生成手段について台詞ごとに予め定められた優先順位と、前記発話音声収録手段により発話音声が収録されたか否かにしたがって、前記複数の台詞音声生成手段のいずれかを選択して台詞音声を生成するための選択手段とを含み、

前記複数の台詞音声生成手段は、

前記発話音声収録手段により収録された前記発話音声を、発話のタイミング及びその長さが前記台詞情報記憶手段に記憶された前記台詞情報による指定と合致するように加工して、対応する台詞の台詞音声を生成するための台詞音声加工手段と、

前記台詞の少なくとも一部の発話音声に基づき、任意の台詞について、前記ある発話者の音声ではなく、前記ある発話者の音声と類似した声質の台詞音声を生成するための手段とを含む、台詞音声作成装置。

【請求項 2】

前記複数の台詞の各々について、対応する台詞音声に加えるべき音響効果を指定する音響効果情報を記憶するための音響効果情報記憶手段と、

前記台詞音声作成手段により作成された前記台詞音声の各々に対し、前記音響効果情報記憶手段に記憶された音響効果情報により指定される音響効果を実現するための音響処理を行なうための音響処理手段とをさらに含む、請求項 1 に記載の台詞音声作成装置。

【請求項 3】

前記台詞音声作成手段により、複数の発話者の各々に対して作成された台詞音声を、前記台詞情報記憶手段に記憶された台詞情報にしたがったタイミングで順番に再生するための音声再生手段をさらに含む、請求項 1 又は請求項 2 に記載の台詞音声作成装置。

【請求項 4】

前記発話音声収録手段は、

表示装置一体型のタッチパネルと、

各台詞が発話される場面の映像を、台詞ごとに、台詞とともに前記タッチパネル上に繰返し再生して表示するための表示手段と、

発話者の直前の所定時間分の発話音声を録音するための発話録音手段と、

発話者による録音停止の指示にตอบสนองして、前記発話録音手段を停止させ、前記発話録音手段により録音された発話音声のうち、前記指示の直前に録音された、前記映像の持続時間分の録音音声を、当該映像に対応する台詞に対する前記発話者の発話音声として保存するための保存手段と、

前記表示手段、前記発話録音手段、及び前記保存手段を、対象となる台詞を順次変えながら繰返して動作するように制御するための繰返し制御手段とを含む、請求項 1 ~ 請求項 3 のいずれかに記載の台詞音声作成装置。

【請求項 5】

前記発話音声収録手段はさらに、前記表示手段による映像の表示時に、発話者が発話音声のタイミングを定めるための補助となるタイミング補助情報を前記映像に重畳して表示するためのタイミング補助情報重畳手段を含む、請求項 4 に記載の発話音声収録装置。

【請求項 6】

前記発話音声収録手段はさらに、

前記複数の台詞ごとに、予め録音された音声を記憶するための音声記憶手段と、

前記表示手段による映像の表示ごとに、表示される映像に対応する台詞の音声を前記音声記憶手段から読出し、前記映像の表示に同期して再生するための音声再生手段とを含む、請求項 4 又は請求項 5 に記載の発話音声収録装置。

【請求項 7】

前記発話音声収録手段はさらに、

発話者の発話時の音声レベルを測定するための音声レベル測定手段と、

前記音声レベル測定手段により測定された音声レベルを示す補助情報を前記表示手段による映像の表示に重畳して表示するための音声レベル補助情報重畳手段とを含む、請求項 4 ~ 請求項 6 のいずれかに記載の発話音声収録装置。

【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、

複数の台詞の各々の発話タイミング、発話内容の台詞、及び発話時間を特定する台詞情報を記憶するための台詞情報記憶手段と、

10

20

30

40

50

ある発話者に割当てられている台詞の少なくとも一部を当該ある発話者に発話させ、その発話音声を受録するための発話音声収録手段と、

前記発話音声収録手段により収録された発話音声に基づき、前記ある発話者に割当てられている台詞の全てについて、発話のタイミング及びその長さが前記台詞情報に合致するように台詞音声を作成するための台詞音声作成手段として機能させる、コンピュータプログラムであって、

前記台詞音声作成手段は、

各々、前記発話音声収録手段により収録された前記発話音声に基づいて、発話のタイミング及びその長さが前記台詞情報による指定と合致するように前記ある発話者の台詞音声を生成するための複数の台詞音声生成手段と、

10

前記複数の台詞の各々について、前記複数の台詞音声生成手段について台詞ごとに予め定められた優先順位と、前記発話音声収録手段により発話音声を受録されたか否かにしたがって、前記複数の台詞音声生成手段のいずれかを選択して台詞音声を生成するための選択手段とを含み、

前記複数の台詞音声生成手段は、

前記発話音声収録手段により収録された前記発話音声を、発話のタイミング及びその長さが前記台詞情報記憶手段に記憶された前記台詞情報による指定と合致するように加工して、対応する台詞の台詞音声を生成するための台詞音声加工手段と、

前記台詞の少なくとも一部の発話音声に基づき、任意の台詞について、前記ある発話者の音声ではなく、前記ある発話者の音声と類似した声質の台詞音声を生成するための手段とを含む、コンピュータプログラム。

20

【発明の詳細な説明】

【技術分野】

【0001】

この発明は映画、アニメーション等、出演者がシナリオにしたがって発話するマルチメディア製作物等における台詞音声作成装置に関し、特に、所定のシナリオにあわせて台詞を効率的に収録し、再生することが可能な台詞音声作成装置およびそのための発話音声収録装置に関する。

【背景技術】

【0002】

30

コンピュータ技術、特に映像及び音響処理技術が発達することにより、利用者を参加させたマルチメディア製作物をごく短時間に製作することが可能なシステムが実用化されつつある。たとえば、博覧会のアトラクションとして、そのようなシステムを導入し、博覧会を訪れた参加者を登場人物とする映画をその場で製作し上映するようなアトラクションが実現できると、より多くの訪問者をひきつけるような効果が期待できる。

【0003】

そうしたアトラクションシステムが、特許文献1に開示されている。特許文献1に開示されたシステムは、参加者の立体的顔画像を撮影する複数の3次元スキャナ及び画像処理パーソナルコンピュータ(以下単に「PC」と呼ぶ。)と、予め準備された映画のシナリオを登場人物の画像、背景画像等とともに記憶するためのシナリオ保存サーバと、これら3次元スキャナによって撮影された参加者の立体的顔画像に基づき、シナリオ保存サーバに保存されている映画の登場人物の顔を参加者の顔画像に入替えることにより、参加者が登場人物として登場する映画を生成するためのアトラクション映像生成装置と、生成された映画を映写するための映像送出装置とを含む。

40

【0004】

複数の参加者はそれぞれ、映画の中の所望の登場人物を指定することでその登場人物として映画の中に登場することができる。

【特許文献1】特開2005-115740号公報

【発明の開示】

【発明が解決しようとする課題】

50

【 0 0 0 5 】

しかし、上記したシステムでは、登場人物の顔画像を参加者の顔画像に入替えることはできても、声までを入替えることはできない。これは、顔画像の場合には3次元スキャナで参加者の顔画像を撮影することにより、どのようなシーンでもその顔画像を利用することができるのに比べ、音声の場合にはそのようなことができないためである。

【 0 0 0 6 】

音声の場合には、シナリオにあわせ、かつ映像にあわせて台詞を読上げる作業をする必要がある。こうした作業は単に難しいだけでなく、長時間を有する作業であるという問題点がある。限られた時間しかないアトラクションの参加者にそのような作業を強いることはできず、結果として上記したアトラクションシステムにおいて参加者の声を使用することができなかつた。

10

【 0 0 0 7 】

こうしたことは、映画に限らず他の場面でも起こりうる。たとえば音声のみを使用したラジオドラマのようなものを作成する場合にも、参加者がさくことのできる時間が短ければ、参加者の声を利用して長時間のドラマを作成することは困難である。また、アニメーションを吹替える場合、又は実写の動物に人間の声をあてはめるための吹替えをする場合にも同様の問題が生じる。

【 0 0 0 8 】

また、こうした問題はアトラクションの参加者のように一時的に来訪している人の声を使用する場合だけではなく、いわゆる声優のように吹替えを職業としている人の場合にも生じえる。ある長さのシナリオの全てを声優の声で吹替える場合には、最低限必要な時間が決まってしまう、利用可能な時間が非常に限定されているときには吹替えを完全に行なうことが不可能な場合さえ生じ得る。

20

【 0 0 0 9 】

仮に、既に他人の声で台詞が多数録音されている場合には、もしかしたら参加者の声とよく似た人物の声を使用して吹替えを行なうことができるかもしれない。しかしそのためには、男女、年齢、声の質等を考慮して、できるだけ多数の人物の声を収録しておかなければならず、それは非常に困難である。

【 0 0 1 0 】

それ故に本発明の目的は、登場人物の台詞が分かっているマルチメディア製作物を製作するために、登場人物の音声をユーザの声で容易に短時間で置換することが可能な台詞音声作成装置を提供することである。

30

【 0 0 1 1 】

本発明の他の目的は、登場人物の台詞が分かっているマルチメディア製作物を製作するために、登場人物の音声をユーザの声に近い声質で容易に短時間で置換することが可能な台詞音声作成装置を提供することである。

【 0 0 1 2 】

本発明のさらに他の目的は、登場人物の台詞が分かっているマルチメディア製作物を製作するために、登場人物の音声をユーザの声に近い声質で容易に短時間で置換することが可能なように、多数のユーザの声を収集し、それらを音声の置換において利用できる台詞音声作成装置を提供することである。

40

【 0 0 1 3 】

本発明の別の目的は、発話者のみで容易に品質の高い台詞の発話を収録できる発話音声収録装置を提供することである。

【 課題を解決するための手段 】

【 0 0 1 4 】

本発明の第1の局面に係る台詞音声作成装置は、各々、発話のタイミング、その長さ、及び発話者が予め定められている複数の台詞の音声を作成するための台詞音声作成装置であって、複数の台詞の各々の発話タイミング、発話内容の台詞、及び発話時間を特定する台詞情報を記憶するための台詞情報記憶手段と、ある発話者に割当てられている台詞の少

50

なくとも一部を当該ある発話者に発話させ、その発話音声を収録するための発話音声収録手段と、発話音声収録手段により収録された発話音声に基づき、ある発話者に割当てられている台詞の全てについて、発話のタイミング及びその長さが台詞情報記憶手段に記憶された台詞情報に合致するように台詞音声を作成するための台詞音声作成手段とを含む。

【0015】

ある発話者の台詞に対し、発話音声収録手段によってその一部しか収録されなくても、台詞音声作成手段によって、その発話者に割当てられている台詞の全てについて、台詞情報に合致した台詞音声を作成することができる。その結果、台詞の吹替え等に経験の浅い発話者で、録音時間が十分でなく、台詞の一部しか録音しなかった場合にも台詞音声の全てを作成できる。その結果、登場人物の台詞が分かっているマルチメディア製作物において、登場人物の音声をユーザの声で容易に短時間で置換することが可能な台詞音声作成装置を提供できる。また、発話者の音声を効率よく収集することができる。

10

【0016】

好ましくは、台詞音声作成手段は、各々、発話音声収録手段により収録された発話音声に基づいて、発話のタイミング及びその長さが台詞情報による指定と合致するようにある発話者の台詞音声を生成するための複数の台詞音声生成手段と、複数の台詞の各々について、複数の台詞音声生成手段について台詞ごとに予め定められた優先順位と、発話音声収録手段により発話音声収録されたか否かにしたがって、複数の台詞音声生成手段のいずれかを選択して台詞音声を生成するための選択手段とを含む。複数の台詞音声生成手段は、発話音声収録手段により収録された発話音声を、発話のタイミング及びその長さが台詞情報記憶手段に記憶された台詞情報による指定と合致するように加工して、対応する台詞の台詞音声を生成するための台詞音声加工手段と、台詞の少なくとも一部の発話音声に基づき、任意の台詞について、ある発話者の音声ではなく、ある発話者の音声と類似した声質の台詞音声を生成するための手段とを含む。

20

【0017】

台詞音声加工手段は、発話音声収録手段により収録された音声を加工して台詞音声を生成する。この場合、発話者の音声の特徴はほとんどそのまま維持されるので、音質の劣化は少ないが、一部しか発話音声収録されなかった場合には、生成できない台詞音声が発生する。一方、台詞音声を生成するための手段は、その発話者ではなく、その発話者の音声と類似した声質の台詞音声を生成する。一般に、この場合には台詞音声の品質は落ちるが、全ての台詞音声を生成することができる。選択手段が、優先順位と、発話音声収録手段により発話音声収録されたか否かによって台詞音声生成手段を選択するので、全ての台詞音声を、発話者の音声に近い品質で簡単に作成することができる。

30

【0018】

さらに好ましくは、台詞音声作成装置は、複数の台詞の各々について、対応する台詞音声に加えるべき音響効果を指定する音響効果情報を記憶するための音響効果情報記憶手段と、台詞音声作成手段により作成された台詞音声の各々に対し、音響効果情報記憶手段に記憶された音響効果情報により指定される音響効果を実現するための音響処理を行なうための音響処理手段とをさらに含む。

【0019】

映画、ラジオドラマ、テレビドラマ等では、シーンにより発話者の環境が変わる。そのため、シーンにあわせて音声についても環境による音響的な変化を反映する必要がある。この台詞音声作成装置では、台詞の各々について、音響効果を指定する音響効果情報を記憶しておき、音響処理手段が、台詞の各々について、音響効果情報にしたがった音響効果を台詞音声に加える。そのため、シーンの変化にあわせた臨場的な台詞音声を簡単に短時間で作成できる。

40

【0020】

台詞音声作成装置は、台詞音声作成手段により、複数の発話者の各々に対して作成された台詞音声を、台詞情報記憶手段に記憶された台詞情報にしたがったタイミングで順番に再生するための音声再生手段をさらに含んでもよい。

50

【 0 0 2 1 】

音声再生手段が、一連の台詞音声を、それぞれ所定のタイミングで順番に再生する。そのため、たとえば予め登場人物が発言する時間が決まっている映画、テレビドラマ、ラジオドラマ等の製品の音声を、この台詞音声作成装置により容易に実現することができる。

【 0 0 2 2 】

発話音声収録手段は、表示装置一体型のタッチパネルと、各台詞が発話される場面の映像を、台詞ごとに、台詞とともにタッチパネル上に繰り返し再生して表示するための表示手段と、発話者の直前の所定時間分の発話音声を録音するための発話録音手段と、発話者による録音停止の指示に回答して、発話録音手段を停止させ、発話録音手段により録音された発話音声のうち、指示の直前に録音された、映像の持続時間分の録音音声を、当該映像 10
に対応する台詞に対する発話者の発話音声として保存するための保存手段と、表示手段、発話録音手段、及び保存手段を、対象となる台詞を順次変えながら繰り返し動作するように制御するための繰り返し制御手段とを含んでもよい。

【 0 0 2 3 】

表示手段は、各台詞が発話される場面の映像を、台詞ごとに、台詞と共にタッチパネル上に繰り返し再生して表示する。発話者は、この繰り返しのリズムにあわせて台詞を発話することができる。発話録音手段は、直前の所定時間分の発話音声を録音している。録音停止の指示があると、保存手段が、映像の持続時間分の直前の録音音声を、その映像に対応する台詞に対する発話者の発話音声として保存する。繰り返し制御手段により、上記処理が 20
各台詞に対して行なわれることにより、複数の台詞の全てに対して発話者の発話音声を収録できる。

【 0 0 2 4 】

映像を繰り返し表示することによって、発話者は発話のタイミングを捉えやすくなり、台詞の発声に失敗する危険性が小さくなる。さらに、発話を繰り返し、発話者がよく発話できたと思う発話を簡単な操作で収録音声として保存できるため、映像によく合致した発話を、アシスタントなしでも簡単に収録することが可能になる。

【 0 0 2 5 】

好ましくは、発話音声収録手段はさらに、表示手段による映像の表示時に、発話者が発話のタイミングを定めるための補助となるタイミング補助情報を映像に重畳して表示するためのタイミング補助情報重畳手段を含む。 30

【 0 0 2 6 】

発話のタイミング補助情報が映像に重畳して表示されるため、発話者は発話のタイミングをさらに捉えやすくなり、台詞の発話タイミングを逃すおそれが小さくなる。

【 0 0 2 7 】

より好ましくは、発話音声収録手段はさらに、複数の台詞ごとに、予め録音された音声を記憶するための音声記憶手段と、表示手段による映像の表示ごとに、表示される映像に対応する台詞の音声を音声記憶手段から読み出し、映像の表示に同期して再生するための音声再生手段とを含む。

【 0 0 2 8 】

映像とともに、その映像に対応する台詞の音声再生される。発話者はこの音声を参考にして発話することができる。その結果、発話者のみにより、映像中の登場キャラクターの発話画像とタイミングが一致した、品質の高い発話音声容易に収録することが可能になる。 40

【 0 0 2 9 】

さらに好ましくは、発話音声収録手段はさらに、発話者の発話時の音声レベルを測定するための音声レベル測定手段と、音声レベル測定手段により測定された音声レベルを示す補助情報を表示手段による映像の表示に重畳して表示するための音声レベル補助情報重畳手段とを含む。

【 0 0 3 0 】

発話者の音声レベルを示す補助情報が映像に重畳して表示される。発話レベルが大きす 50

ぎたり、小さすぎたりするという失敗の危険性を小さくでき、アシスタントなしで、発話者のみによって品質の高い発話音声を容易に収録することができる。

【0031】

本願発明の第2の局面に係るコンピュータプログラムは、コンピュータにより実行されると、当該コンピュータを、複数の台詞の各々の発話タイミング、発話内容の台詞、及び発話時間を特定する台詞情報を記憶するための台詞情報記憶手段と、ある発話者に割り当てられている台詞の少なくとも一部を当該ある発話者に発話させ、その発話音声を収録するための発話音声収録手段と、発話音声収録手段により収録された発話音声に基づき、ある発話者に割り当てられている台詞の全てについて、発話のタイミング及びその長さが台詞情報に合致するように台詞音声を作成するための台詞音声作成手段として機能させる。

10

【0032】

この発明の第3の局面に係る発話音声収録装置は、各々、発話のタイミング、その長さ、及び発話者が予め定められている複数の台詞の音声を発話者に発話させ、その発話音声を収録するための発話音声収録装置であって、表示装置一体型のタッチパネルと、各台詞が発話される場面の映像を、台詞ごとに、台詞とともにタッチパネル上に繰返し再生して表示するための表示手段と、発話者の直前の所定時間分の発話音声を録音するための発話録音手段と、発話者による録音停止の指示にตอบสนองして、発話録音手段を停止させ、発話録音手段により録音された発話音声のうち、指示の直前に録音された、映像の持続時間分の録音音声を、当該映像に対応する台詞に対する発話者の発話音声として保存するための保存手段と、表示手段、発話録音手段、及び保存手段を、対象となる台詞を順次変えながら繰返して動作するように制御するための繰返し制御手段とを含む。

20

【0033】

好ましくは、発話音声収録装置はさらに、表示手段による映像の表示時に、発話者が発話音声のタイミングを定めるための補助となるタイミング補助情報を映像に重畳して表示するためのタイミング補助情報重畳手段を含む。

【0034】

さらに好ましくは、発話音声収録装置はさらに、複数の台詞ごとに、予め録音された音声を記憶するための音声記憶手段と、表示手段による映像の表示ごとに、表示される映像に対応する台詞の音声を音声記憶手段から読み出し、映像の表示に同期して再生するための音声再生手段とを含む。

30

【0035】

発話音声収録装置はさらに、発話者の発話時の音声レベルを測定するための音声レベル測定手段と、音声レベル測定手段により測定された音声レベルを示す補助情報を表示手段による映像の表示に重畳して表示するための音声レベル補助情報重畳手段とを含んでもよい。

【0036】

この発明の第4の局面に係るコンピュータプログラムは、表示装置一体型のタッチパネルが接続されるコンピュータにより実行されると、当該コンピュータを、各々、発話のタイミング、その長さ、及び発話者が予め定められている複数の台詞の音声を発話者に発話させ、その発話音声を収録するための発話音声収録装置として機能させるためのコンピュータプログラムである。このコンピュータプログラムは、コンピュータを、各台詞が発話される場面の映像を、台詞ごとに、台詞とともにタッチパネル上に繰返し再生して表示するための表示手段と、発話者の直前の所定時間分の発話音声を録音するための発話録音手段と、発話者による録音停止の指示にตอบสนองして、発話録音手段を停止させ、発話録音手段により録音された発話音声のうち、指示の直前に録音された、映像の持続時間分の録音音声を、当該映像に対応する台詞に対する発話者の発話音声として保存するための保存手段と、表示手段、発話録音手段、及び保存手段を、対象となる台詞を順次変えながら繰返して動作するように制御するための繰返し制御手段として機能させる。

40

【発明を実施するための最良の形態】

50

【 0 0 3 7 】

以下、本発明の台詞音声作成装置の実施の形態に係るマルチメディア制作システムについて説明する。以下の説明及び図面において、同一の部品には同一の名称及び参照番号を付してある。それらの機能も同一である。したがってそれらについての詳細な説明は繰返さない。

【 0 0 3 8 】

[第 1 の実施の形態]

図 1 に、本発明の第 1 の実施の形態に係るマルチメディア制作システム 5 0 のブロック図を示す。図 1 を参照して、マルチメディア制作システム 5 0 は、特許文献 1 に記載されたものと同様の複数の 3 次元スキャナからなる三次元スキャナ群 6 0 と、三次元スキャナ群 6 0 により撮影された参加者の顔画像の三次元モデルを作成するための画像処理 P C 6 2 と、映画のシナリオを登場人物（以下「キャラクタ」と呼ぶ。）の顔画像及びその他の画像とともに記憶するための図示しないシナリオ保存サーバと、画像処理 P C 6 2 により生成された参加者の顔画像を用い、シナリオ保存サーバに保存されているキャラクタの顔画像を置換して、参加者の顔を持つ人物が登場する映像を生成し映像データ 6 6 として出力するための映像生成装置 6 4 と、この映像データ 6 6 を保存するための映像データ記憶装置とを含む。

10

【 0 0 3 9 】

マルチメディア制作システム 5 0 はさらに、最終的な映像データ 6 6 を作成するための映像素材を記憶するための映像素材 D B （データベース） 7 0 と、この映画のキャラクタの中で、参加者による吹替えの対象となるキャラクタの台詞に関する台詞情報を記憶するための台詞情報記憶部 7 2 と、映画の中のキャラクタの台詞を標準的な音声で発話した標準音声を記憶するための標準音声記憶部 7 4 と、映画の中の各台詞がどのようなシーンで発話されているか、それにより台詞の音声にどのような音響効果を加えるべきかを示すカット情報を記憶するためのカット情報記憶部 7 6 とを含む。

20

【 0 0 4 0 】

マルチメディア制作システム 5 0 はさらに、映像素材 D B 7 0 に記憶された映像、台詞情報記憶部 7 2 に記憶された台詞情報、標準音声記憶部 7 4 に記憶された標準音声による台詞の発話データ、及びカット情報記憶部 7 6 に記憶されたカット情報を用い、参加者（ユーザ）の音声を収録してその音声に基づき、映画の特定のキャラクタの台詞の音声をユーザの音声に入替える処理（いわゆる「吹替え」と同様の処理）を実行し、ユーザの音声により台詞を発話した音声からなる台詞音声データ 8 6 と、台詞音声データ 8 6 の中の台詞の発話開始時間、発話時間、対応する音声ファイル名等をテーブルとして記憶するための台詞音声テーブル 8 8 とを出力するための台詞音声データ作成部 9 0 とを含む。

30

【 0 0 4 1 】

台詞音声データ作成部 9 0 は、三次元スキャナ群 6 0 と同様、複数のユーザの音声を処理することが可能なように構成されている。各ユーザは後述するように識別子（ I D ）によって区別され、三次元スキャナ群 6 0、画像処理 P C 6 2、及び映像生成装置 6 4 からなる映像処理系と、台詞音声データ作成部 9 0 とで同じユーザについては同じ I D を割当てて管理する。こうすることにより、映画の複数のキャラクタの顔と音声とを同時に特定のユーザの顔と音声とにより入替えることが可能になる。

40

【 0 0 4 2 】

マルチメディア制作システム 5 0 はさらに、台詞音声データ作成部 9 0 がユーザの音声からキャラクタの台詞音声を作成する際に、どのような手法を用いるかを示す情報を台詞ごとに記憶した手法リストテーブル 7 8 と、ユーザによる発話の収録ができなかった台詞について、台詞音声データ作成部 9 0 がユーザの発話に代えて台詞音声データの作成に使用する、予めこの映画の各台詞を種々の声優の音声により発話した音声データを記憶した声優音声 D B 8 0 と、台詞音声データ作成部 9 0 がユーザの発話に代えて音声合成によりユーザの声に似た性質の台詞音声データを生成する際に使用する音声素片を、それらの特徴量データとともに記憶した素片 D B 8 2 とを含む。

50

【 0 0 4 3 】

マルチメディア製作システム 5 0 はさらに、映像生成装置 6 4 から出力された映像データ 6 6 と、台詞音声データ作成部 9 0 から出力された台詞音声データ 8 6 とを台詞音声テーブル 8 8 を使用して互いに同期させて再生することにより、キャラクタの一部の顔画像及び音声ユーザの顔画像及び音声に入替えられたマルチキャラクタ製作物を上演するための映像・音声再生装置 9 2 を含む。

【 0 0 4 4 】

前述したとおり、台詞音声データ作成部 9 0 は、複数のユーザの音声を収録し、それらに基づき、別々のキャラクタの台詞音声を生成する機能を持つ。そのために台詞音声データ作成部 9 0 は、各々が処理対象のユーザに関する識別情報、性別、氏名、年齢、吹替え対象となるキャラクタを特定する情報等を含むユーザ情報の入力を受けるための複数のユーザ情報入力部 1 0 0 , 1 0 0 A , ... , 1 0 0 N と、これらユーザ情報入力部 1 0 0 , 1 0 0 A , ... , 1 0 0 N が受けたユーザ情報に基づいて、各々が対応するユーザの音声を収録し、収録した音声に基づいて種々の手法によりユーザの音声の声質で対応するキャラクタの台詞音声を生成し出力するための複数のキャラクタ音声作成部 1 0 2 , 1 0 2 A , ... , 1 0 2 N と、複数のキャラクタ音声作成部 1 0 2 , 1 0 2 A , ... , 1 0 2 N の出力する、ユーザ音声の声質に置換えられた種々のキャラクタの台詞音声を、台詞情報記憶部 7 2 に記載された台詞情報に基づいて 1 つのマルチメディア製作物の音声を構成するように台詞の番号順にしたがって統合し、台詞音声データ 8 6 及び台詞音声テーブル 8 8 として出力するための音声統合部 1 0 4 とを含む。

【 0 0 4 5 】

なお、ユーザ情報入力部 1 0 0 , 1 0 0 A , ... , 1 0 0 N により入力されたユーザ情報は、画像処理 P C 6 2 にも与えられ、ユーザの顔画像の管理にも用いられる。

【 0 0 4 6 】

複数のキャラクタ音声作成部 1 0 2 , 1 0 2 A , ... , 1 0 2 N の構成はいずれも同じである。したがって以下では、キャラクタ音声作成部 1 0 2 の構成を代表として説明する。

【 0 0 4 7 】

図 2 は、キャラクタ音声作成部 1 0 2 の機能的ブロック図である。図 2 を参照して、キャラクタ音声作成部 1 0 2 は、ユーザ情報を受けて、映像素材 D B 7 0 に格納されている映像素材、台詞情報記憶部 7 2 に記憶されている台詞情報、及び標準音声記憶部 7 4 に記憶されている標準音声による台詞音声を利用して、ユーザによる吹替え対象となるキャラクタの台詞音声をユーザに発話させ、その発話音声をユーザ音声 D B 1 2 0 に収録するための音声収録部 1 1 4 と、音声収録部 1 1 4 における発話の収録を制御するためにアテンダントが音声収録部 1 1 4 を操作するため、及びユーザによる発話の補助を行なうために使用する入出力装置 1 1 2 とを含む。

【 0 0 4 8 】

ところで、一般的に、1 つの映画を構成する台詞は多数あり、あるキャラクタの台詞のみに限ってもユーザによるその台詞の発話音声の収録にはかなりの時間を要することが予測される。映画の音声の発話となると、キャラクタの動きにあわせて発話を行なう必要があり、この収録にはさらに時間がかかる可能性が高い。特に、アトラクション等では、時間的制限もあって、全ての発話音声の収録を行なうことが難しい場合が多い。また、収録できたとしても発話時間が短すぎたり長すぎたりすることも多く、そのまま収録音声を用いることができない場合が多い。そこで本実施の形態に係るキャラクタ音声作成部 1 0 2 では、あるキャラクタの台詞のうち、ユーザによる発話を収録できた分、及びできなかった分の双方に対して、それぞれ所定の音声生成手法を用いて、台詞音声をできるだけユーザの声質に近い声質で生成することを目標としている。手法リストテーブル 7 8 には、台詞ごとにどのような手法をどのような優先順位で使用するかを示す台詞ごとの手法リストが記憶されており、キャラクタ音声作成部 1 0 2 は音声生成にこの手法リストテーブル 7 8 を使用する。

【 0 0 4 9 】

キャラクタ音声作成部 102 はさらに、音声収録部 114 によってユーザ音声 DB 120 に記憶されたユーザ音声に対し、手法リストテーブル 78 を参照して、ユーザの吹替え対象のキャラクタの台詞ごとに条件に合致した手法を決定するための合成手法決定部 116 と、合成手法決定部 116 によって決定された手法を用いてキャラクタの台詞音声をユーザの声質にあわせて作成し、台詞ごとに音声ファイル 110 として出力するための音声作成部 118 とを含む。音声作成部 118 はこの合成の際に、手法によって声優音声 DB 80、ユーザ音声 DB 120、素片 DB 82、標準音声記憶部 74 等に記憶された音声を適宜利用する。また音声作成部 118 は、生成された台詞の発話音声に対し、カット情報記憶部 76 に記憶されたカット情報に基づいて決定される音響効果を加えて最終的な音声ファイル 110 を出力する。

10

【0050】

キャラクタ音声作成部 102 はさらに、ユーザ音声 DB 120 に格納されたユーザの音声を声優音声 DB 80 に新たな声優音声として登録する処理を行なう音声 DB 更新部 122 と、ユーザ音声 DB 120 に記憶されたユーザ音声を音素片（素片）に分解し、それらの所定の音響特徴量データ、音素ラベル、及びユーザ ID とともに素片 DB 82 に追加するための素片 DB 更新部 124 とを含む。素片 DB 更新部 124 による音声の素片への分解においては、音声認識技術を利用し、台詞情報記憶部 72 に記憶された台詞情報にあわせてユーザ音声 DB 120 に記憶されたユーザの音声を細分化するセグメンテーションを行なう。

【0051】

20

図 3 は、台詞情報記憶部 72 に記憶される台詞情報テーブルの構成を示す。図 3 を参照して、台詞情報記憶部 72 は、作成対象となる映画の台詞の全てを通し番号（No）で管理するためのものである。各台詞情報は、その台詞の通し番号（以下「台詞番号」と呼ぶ。）と、その台詞を発話する映画のキャラクタを識別するキャラクタ ID と、台詞の内容であるテキストデータと、その台詞を標準音声で発話したものを記録した、標準音声記憶部 74 内の音声ファイルのファイル名と、映画の進行経過の中でその台詞の発話が始まる時点を示す開始時刻と、その発話の継続時間を示す発話時間とを含む。台詞情報記憶部 72 の台詞情報テーブルがこのような構成を有しているため、同じキャラクタ ID の台詞を抽出することにより、あるキャラクタの台詞を全てリスト化することができる。また、ある台詞について、ユーザによる音声を利用できないときに、対応する標準音声を音声ファイル名により示される音声ファイルから得ることができる。

30

【0052】

図 4 に、本実施の形態に係るマルチメディア制作システム 50 における、ユーザによる録音状況としてあり得るいくつかの場合を示す。なお、たとえば図 4（A）を参照して、あるユーザについて録音すべき発話全体が発話集合 140 を形成するものとする。この発話集合 140 は、音声合成、声質変換等のために必要で、必ず収録すべき発話からなる必須発話部分 142 と、対応するキャラクタの台詞全体からなる台詞部分 144 とからなる。ユーザの収録にかかる時間、ユーザの発話の巧拙等により、必須発話部分 142 はともかく、台詞部分 144 については、全て収録できる場合、一部のみしか収録できないとき、全く収録できないとき、の 3 通りがあり得る。図 4 には、それらの場合を分けて、収録できた部分に斜線を付し、収録できなかった部分は白抜きのままで例示してある。

40

【0053】

たとえば図 4（A）には、発話集合 140 の全てを収録できた場合を示す。図 4（B）には、必須発話部分 142 と、一部の台詞部分 146 のみが収録でき、残りの部分 148 が収録できなかった場合を示す。図 4（C）には、必須発話部分 142 の部分のみが収録でき、他の台詞部分 150 が全く収録できなかった場合を示す。

【0054】

図 4（A）に示す場合には、基本的にはユーザの音声のみを用いて台詞音声を作成することができる。ただしこの場合にも、ユーザの巧拙によって話速を変換したり、発話レベルを調整したりする加工が必要なときがある。それらは台詞ごとに異なる。

50

【 0 0 5 5 】

図 4 (B) に示す場合には、収録できた台詞部分 1 4 6 については、図 4 (A) に示す場合と同様に処理できるが、収録できなかった台詞部分 1 4 8 については何らかの手法を用いてユーザの音声以外からユーザの音声に似た台詞音声を生成する必要がある。

【 0 0 5 6 】

図 4 (C) に示す場合には、台詞部分 1 5 0 の全てについて台詞音声を生成する必要がある。その場合、たとえば必須発話部分 1 4 2 からユーザの声質を表す特徴量を抽出し、声優音声 DB から類似の声質の声優の台詞音声を抽出したり、標準音声の声質をユーザの声質に近くなるように変換したりする処理 (声質変換) を行なったりする必要がある。

【 0 0 5 7 】

図 2 に示す手法リストテーブル 7 8 には、台詞ごとに、どのような優先順位でそうした手法を使用するかが示されている。本実施の形態では、9 種類の手法を用いて台詞音声を生成する。それら手法の詳細については後述する。

【 0 0 5 8 】

図 5 及び図 6 は、図 2 に示す音声収録部 1 1 4 で行なわれる音声収録処理を、コンピュータハードウェア上で実現するコンピュータプログラムのフローチャートである。既に述べたように、映画のキャラクタの台詞の吹替えを行なうことは難しい。たとえばある台詞について、決められた時間で明瞭に発話する必要がある。通常、発話時間が長すぎても短すぎても吹替えとして不適當になる場合がある。ましてや、声優ではないユーザに台詞の吹替えを間違いなく行なわせるのは困難である。そこで、本実施の形態では、様々な方策を講じてできるだけ正確に所望の台詞音声を収録することができるようにしている。たとえば、図 7 に示されるように、台詞音声収録時にユーザに提示される入出力装置 1 1 2 の画面に、台詞の発話時の映像 2 4 6 と、発話すべき台詞のテキスト 2 4 0 とを表示し、発話の進行にあわせて伸びるプログレスバー 2 4 2 を表示したり、台詞のテキスト 2 4 0 のうち、発話が終了しているべき部分 2 4 4 の色を、これから発話すべき部分の色と違う色で表示したりする、という方法を採用する。

【 0 0 5 9 】

図 5 を参照して、このプログラムは、ユーザ情報をユーザ情報入力部 1 0 0 から受信し所定の記憶領域に保存するステップ 1 7 0 と、ステップ 1 7 0 に続き、受信したユーザ情報にしたがって、処理対象のユーザに対し、指定されたキャラクタを割当てるステップ 1 7 2 と、ステップ 1 7 2 に続き、共通の練習用台詞及び対応する標準音声、ステップ 1 7 2 で割当てられたキャラクタの台詞及び対応する標準音声を図 2 に示す台詞情報記憶部 7 2 及び標準音声記憶部 7 4 から抽出するステップ 1 7 4 と、ステップ 1 7 4 に続き、ユーザ音声テーブルと呼ばれる、ユーザの台詞音声を管理するためのテーブルを生成し、全ての台詞について未収録状態に初期化するステップ 1 7 6 とを含む。

【 0 0 6 0 】

ユーザ音声テーブルは、図 2 に示すユーザ音声 DB 1 2 0 の一部を構成する。図 8 を参照して、ユーザ音声 DB 1 2 0 は、ユーザの発話を台詞ごとに収録した音声ファイルを記憶するユーザ音声記憶部 2 6 2 と、ユーザ音声記憶部 2 6 2 に記憶された音声ファイルの管理を行なうためのユーザ音声テーブル 2 6 0 とを含む。

【 0 0 6 1 】

ユーザ音声テーブル 2 6 0 は、ユーザが吹替えを行なうキャラクタの台詞と、対応するユーザ音声とを管理するためのものであって、先頭にはユーザ ID が付され、さらに、このキャラクタの台詞の各々について、抽出された台詞の台詞番号と、ユーザによるその台詞の発話の収録が完了したか否かを示す録音フラグと、収録した発話音声データを格納した音声ファイルの名称と、その発話時間とを記憶するためのものである。録音フラグは、1 のときに発話音声収録済であることを示し、0 のときには未収録であることを示す。なお、実際には発話開始時間、発話時間等は 1 秒よりも細かい単位で管理する必要があるが、以下の説明及び図面では、理解を容易にするため、これら時間は秒単位で管理するものとする。

10

20

30

40

50

【 0 0 6 2 】

再び図5を参照して、ステップ176では、上記したユーザ音声テーブル260が新たに作成され、台詞番号には抽出された台詞に付されている通し番号が、録音フラグには全て0が、音声ファイル名には全て空白が、発話時間には全て0が、それぞれ代入される。

【 0 0 6 3 】

このプログラムはさらに、ステップ176に続き、収録に要した時間を測定するためのタイマを起動するステップ178と、ユーザ音声テーブル260内の先頭の台詞を選択するステップ180と、直前のステップで選択された台詞のテキストをユーザの前に置かれたモニタに表示するステップ182と、この台詞に対応する標準音声を標準音声記憶部74から取出し、再生するステップ184とを含む。ステップ182及び184においても、図7に示したような表示が行なわれる。

10

【 0 0 6 4 】

このプログラムはさらに、ステップ184に続いて、ユーザの発話練習の時間として設けられたステップ186と、ステップ182に戻って再度練習を行なうか、次のステップに進んでもよいかをアテンダントが判定して入力する判定結果にしたがい、制御の流れを分岐させるステップ188とを含む。ステップ188での判定結果が再度練習を行なうべきことを示すときには、制御はステップ182に戻る。

【 0 0 6 5 】

このプログラムはさらに、ステップ188で練習を終了しても良いことを示す入力があったことに応答して実行され、選択中の台詞を再度表示するステップ190と、選択中の台詞の通常の発話速度にしたがって変化するプログレスバーの表示を開始するステップ192とを含む。

20

【 0 0 6 6 】

続いて図6を参照して、このプログラムは、ステップ192の次に配置され、ユーザの発話する台詞音声を録音するステップ194と、ステップ194で録音された台詞音声を再生するステップ196と、ステップ196で再生された台詞音声の発話時間、発話の明瞭さ及び自然さ等に基づいて、この台詞の収録を完了するか否かについてアテンダントが判定して入力した結果にしたがい、制御の流れを分岐させるステップ198と、ステップ198においてこの台詞の収録を完了することを示す入力が行なわれたことに応答して実行され、ステップ194で録音された音声を音声ファイルとしてユーザ音声記憶部262内に保存し、ユーザ音声テーブル260内の当該台詞の音声ファイル名欄にその音声ファイル名を、発話時間欄に録音音声の持続時間を、それぞれ代入するステップ200と、収録フラグに「1」を代入するステップ201とを含む。

30

【 0 0 6 7 】

このプログラムはさらに、ステップ200の後、対象のキャラクタの次の台詞の選択を試みるステップ202と、ステップ202で選択を試みた次の台詞が存在しているか否か、すなわち対象のキャラクタの台詞を全て処理したか否かを判定し、その判定結果にしたがって制御の流れを分岐させるステップ204と、ステップ204においてまだ台詞が残っていると判定されたことに応答して、タイマを参照し、録音開始から所定時間が経過したか否かを判定し、判定結果にしたがって制御の流れを分岐させるステップ212とを含む。ステップ212においてまだ所定時間が経過していないと判定された場合には、制御は図5のステップ182に戻る。

40

【 0 0 6 8 】

このプログラムはさらに、ステップ204で対象のキャラクタの全台詞について収録が完了したと判定された場合、及びステップ212において所定時間が経過したと判定されたことに応答して実行され、録音した全音声を、対応する台詞のテキストに基づいてセグメンテーションし、音声素片に分解するステップ206と、ステップ206で生成された素片の各々について、F0、スペクトル分布等、所定の音響特徴量を算出するステップ208と、ステップ206で作成された素片を、ステップ208で算出された音響特徴量、対応する音素のラベル、及び話者のIDとともに素片DB82に追加して処理を終了する

50

ステップ 210 とを含む。

【0069】

このプログラムはさらに、ステップ 198 において、録音をやり直すことを示す入力が入力されたことにより行なわれたことに応答して実行され、ステップ 194 で録音された音声データを破棄するステップ 214 と、ステップ 214 の後に配置され、タイマの時間を参照して所定時間が経過したか否かを判定し、判定結果にしたがって制御の流れを分岐させるステップ 216 と、ステップ 216 においてまだ所定時間が経過していないと判定されたときに実行され、どこから処理を再開するかを決めるアテンダントの入力にしたがって、台詞音声の収録から再開するときにはステップ 190 に、発話の練習から再開するときにはステップ 182 に、それぞれ制御の流れを分岐させるステップ 220 と、ステップ 216 で既に所定時間が経過していると判定されたことに応答して実行され、現在収録中の台詞が必須部分であればステップ 220 に、それ以外であればステップ 206 に、それぞれ制御を分岐させるステップ 218 とを含む。

10

【0070】

図 9 は、図 2 に示す音声作成部 118 のより詳細なブロック図を示す。図 9 を参照して、音声作成部 118 は、それぞれ第 1 ~ 第 9 の手法によって台詞音声を生成するための第 1 ~ 第 9 の音声生成部 300, 302, 304, 306, 308, 310, 312, 314, 及び 316 と、合成手法決定部 116 によって決定された手法にしたがって、第 1 ~ 第 9 の音声生成部 300, 302, 304, 306, 308, 310, 312, 314, 及び 316 のいずれかを選択的に能動化し、ユーザ音声を与えて指定した手法で音声を生

20

【0071】

第 1 の音声生成部 300 は、ある台詞についてユーザの台詞音声を収録することができたときの手法である。この場合には、原則として収録した音声をそのまま使用する。

【0072】

第 2 の音声生成部 302 も、ある台詞についてユーザの台詞音声を収録することができたときの手法である。ただし、この手法では、収録した台詞音声の発話速度を調整して台詞音声を生成する。

30

【0073】

第 3 の音声生成部 304 は、台詞のうち、一部についてユーザの台詞音声を収録することができなかつたときにも有効な手法である。この手法では、収録することができた台詞についてはユーザの台詞音声の話速変換をして台詞音声を生成する。収録することができなかつた台詞については、ユーザの音声を使用せず、標準音声記憶部 74 に記憶された標準音声のうち、ユーザ情報に合致した台詞音声（性、年齢等）を用いる。

【0074】

第 4 の音声生成部 306 も、台詞のうち、一部についてユーザの台詞音声を収録することができなかつたときにも有効な手法である。この手法では、収録することができた台詞についてはユーザの台詞音声の話速変換をして台詞音声を生成する。収録することができなかつた台詞については、声優音声 DB 80 に記憶されている声優による台詞音声のうち、ユーザの音声にもっとも近い声質を持つ声優の台詞音声を採用される。このときの声優音声の決定には、練習用台詞が得られたユーザ音声の所定の特徴量（基本周波数、スペクトル分布等）を用いた声質間の距離比較が用いられる。

40

【0075】

第 5 の音声生成部 308 も、台詞のうち、一部についてユーザの台詞音声を収録することができなかつたときにも有効な手法である。この手法では、収録することができた台詞についてはユーザの台詞音声の話速変換をして台詞音声を生成する。収録することができ

50

なかった台詞については、声優音声DB80に記憶されている声優による台詞音声のうち、ユーザ音声と声質がもっとも類似のものを特定し、その台詞音声にさらにユーザの性質を反映させた声質変換を行なって台詞音声とする。

【0076】

第6の音声生成部310も、台詞のうち、一部についてユーザの台詞音声を収録することができなかつたときにも有効な手法である。この手法では、収録することができた台詞についてはユーザの台詞音声の話速変換をして台詞音声を生成する。収録することができなかつた台詞については、収録できたユーザ音声から生成した音声素片のうち、母音の音声素片と、素片DB82に記憶されている全子音の音声素片のうち、ユーザの音声に類似した特徴量を持つ音声素片とを用いて音声合成をする。発話の個人的特徴は主として母音に現れるので、このような音声合成をすることによって、かなりユーザの音声に似た声質の合成音声を生成することができる。

10

【0077】

第7の音声生成部312は、必須発話部分以外の台詞音声が全く収録できなかつたときに採用される手法である。この手法では、声優音声DB80に記憶されている声優音声のうち、ユーザ音声と最も声質が類似した音声台詞音声として使用される。このときの声質の類似の判定には、必須発話部分の音声から抽出される特徴量が使用される。

【0078】

第8の音声生成部314も、必須発話部分以外の台詞音声が全く収録できなかつたときに有効な手法である。この手法では、声優音声DB80に記憶されている声優音声のうち、ユーザ音声と最も声質が類似した音声を、その声優音声にさらにユーザ音声の性質を用いた声質変換を行なって台詞音声を生成する。

20

【0079】

第9の音声生成部316も、必須発話部分以外の台詞音声が全く収録できなかつたときに有効な手法である。この手法では、必須発話部分について収録したユーザ音声から生成した音声素片のうち、母音の音声素片と、素片DB82に記憶されている子音の音声素片のうち、ユーザの音声に類似した特徴量を持つ音声素片とを用いて音声合成をする。前述のとおり、このような音声合成をすることによって、かなりユーザの音声に似た性質の合成音声を生成することができる。

30

【0080】

以上の各手法の説明から明らかなように、台詞情報記憶部72に記憶された台詞情報は第1～第9の音声生成部300, 302, 304, 306, 308, 310, 312, 314, 及び316の全てにより参照される。標準音声記憶部74に記憶された標準音声は、第3の音声生成部304に参照される。声優音声DB80に記憶された声優音声は、第4の音声生成部306、第5の音声生成部308、第7の音声生成部312、及び第8の音声生成部314により参照される。素片DB82は、第6の音声生成部310、及び第9の音声生成部316により参照される。

【0081】

図10は、図2に示す合成手法決定部116で行なわれる音声の生成手法の決定処理を、コンピュータハードウェア上で実現するコンピュータプログラムのフローチャートである。図10を参照して、このプログラムは、以下の繰返しを制御するための変数*i*に0を代入するステップ340と、変数*i*に1を加算するステップ342と、変数*i*の値が台詞の数MAXを超えたか否かを判定し、超えた場合には処理を終了するステップ344と、ステップ344で変数*i*の値がMAX以下であると判定されたことに応答して実行され、台詞番号が*i*の台詞(以下これを「台詞(*i*)」と書く。)に対応する手法リストを手法リストテーブル78から読み出し、作業用のリスト変数WLISTに格納するステップ346とを含む。

40

【0082】

手法リストテーブル78の詳細を図11に示す。図11を参照して、手法リストテーブル78は、台詞番号ごとに、利用可能な手法の識別子をリストした手法リストを含む。通

50

常は、この手法リストにリストされた手法のいずれかをを用いれば必ず台詞を処理できるように手法リストテーブル78は予め作成されている。ただし、手法リストの中に、利用可能なものが含まれない場合も含めて、たとえば標準音声の台詞音声を出力する、というデフォルトの手法が予め準備されている。

【0083】

再び図10を参照して、このプログラムはさらに、ステップ346の後に配置され、リスト変数WLISTの要素数を変数CMAXに代入するステップ348と、ステップ348に続き、以下の繰返しを制御する変数jに0を代入するステップ350とを含む。なお、一般的にリスト変数の要素のインデックスは0から始まる。

【0084】

このプログラムはさらに、ステップ350に続き、j+1がCMAXの値を超えたか否かを判定し、判定結果にしたがって制御の流れを分岐させるステップ352と、ステップ352においてj+1がCMAX以下であると判定されたことに応答して実行され、リスト変数WLISTのうち、リスト要素WLIST[j]によって示される手法が、与えられたユーザ音声によって実現可能か否かを判定し、判定結果にしたがって制御を分岐させるステップ354とを含む。これら手法の各々について採用できるか否かは、処理対象の台詞音声の収録状況により異なる。基本的には、第1の手法と第2の手法とについては対応の台詞音声の収録がされていなければ利用できないが、それ以外の手法は対応の台詞音声の収録がされていなくても利用可能である。その理由については各手法の説明から明らかとなるであろう。

【0085】

このプログラムはさらに、ステップ354においてリスト要素WLIST[j]によって示される手法が利用可能でないと判定されたことに応答して実行され、変数jの値に1を加算してステップ352に制御を戻すステップ356と、ステップ354においてリスト要素WLIST[j]によって示される手法が利用可能であると判定されたことに応答して実行され、台詞(i)をWLIST[j]により示される手法を用いて処理し、制御をステップ342に戻すステップ358と、ステップ352においてj+1の値がCMAXより大きいと判定されたことに応答して実行され、台詞(i)をデフォルトの手法で処理し、制御をステップ342に戻すステップ360とを含む。

【0086】

図12は、図2に示すカット情報記憶部76の構成を示す。図12を参照して、カット情報記憶部76は、台詞ごとに、台詞番号と、その台詞に対して適用すべき音響効果を列挙した音響効果リストとを記憶している。ある台詞について音響効果を加えようとする場合、音声信号処理部320は、このカット情報記憶部76の、処理対象の台詞番号に対応する音響効果リストを調べ、それを順に先頭から実行する。

【0087】

図13は、図9に示す第1の音声生成部300を実現するプログラムのフローチャートである。図13を参照して、このプログラムは、ユーザ音声DBから台詞音声(i)を讀出すステップ380を含む。ステップ380によってこの処理は終了する。讀出された台詞音声(i)は、音声信号処理部320に与えられ、処理される。音声信号処理部320の処理の詳細は図23を参照して後述する。

【0088】

この第1の手法は、対象となる台詞についてユーザの音声を収録することができたときの手法であり、台詞音声としてユーザの音声をそのまま使用する。

【0089】

図14は図9に示す第2の音声生成部302を実現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。図14を参照して、このプログラムは、ユーザ音声DB120からユーザの台詞音声(i)とその発話時間とを讀出すステップ410と、ステップ410に続き、台詞情報テーブルから台詞(i)の発話時間Tiを讀出すステップ412と、ステップ410で讀出した発話時間tiとステップ412で讀出した発話時間Tiとを

10

20

30

40

50

用い、ユーザの台詞音声 (i) の発話時間が t_i から T_i となるように話速変換を行なって処理を終了するステップ 4 1 4 とを含む。

【 0 0 9 0 】

図 1 5 は、図 9 に示す第 3 の音声生成部 3 0 4 を実現するためのプログラムのフローチャートである。図 1 5 を参照して、このプログラムは、ユーザ音声 DB 1 2 0 のユーザ音声テーブル 2 6 0 から録音フラグ (i) を読出すステップ 4 4 0 と、ステップ 4 4 0 において読出された録音フラグの値が 1 か否かを判定し、その結果に応じて制御の流れを分岐させるステップ 4 4 2 とを含む。

【 0 0 9 1 】

このプログラムはさらに、ステップ 4 4 2 において録音フラグが 1 でない (すなわちこの台詞についてユーザ音声の収録ができなかった) と判定されたことに応答して実行され、標準音声記憶部 7 4 から台詞 (i) の標準音声を読出し、台詞音声 (i) として出力し、処理を終了するステップ 4 4 4 と、ステップ 4 4 2 において録音フラグが 1 であると判定されたことに応答して実行され、ユーザ音声 DB 1 2 0 から台詞音声 (i) と発話時間 t_i とを読出すステップ 4 4 6 と、台詞情報記憶部 7 2 に記憶された台詞情報テーブルから台詞 (i) の発話時間 T_i を読出すステップ 4 4 8 と、ステップ 4 4 6 及び 4 4 8 でそれぞれ読出された発話時間 t_i 及び T_i を用いて、ユーザの台詞音声 (i) の発話時間が T_i となるように、話速変換を行なって出力し、処理を終了するステップ 4 5 0 とを含む。

10

【 0 0 9 2 】

図 1 6 は、図 9 に示す第 4 の音声生成部 3 0 6 を実現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。図 1 6 を参照して、このプログラムは、ユーザ音声 DB 1 2 0 のユーザ音声テーブル 2 6 0 から i 番目の台詞音声に対する録音フラグ (i) を読出すステップ 4 7 0 と、ステップ 4 7 0 で読出された録音フラグ (i) の値が 1 か否かによって制御の流れを分岐させるステップ 4 7 2 と、ステップ 4 7 2 において録音フラグ (i) の値が 1 でない (すなわち 0 である。) と判定されたことに応答して実行され、声優音声 DB 8 0 中に記憶されている台詞 (i) の声優音声のうち、ユーザ音声と最も声質が類似したものを読出して台詞音声 (i) として出力し、処理を終了するステップ 4 7 4 とを含む。

20

【 0 0 9 3 】

このプログラムはさらに、ステップ 4 7 2 において録音フラグが 1 であると判定されたことに応答して実行され、ユーザ音声 DB 1 2 0 から台詞音声 (i) と発話時間 t_i とを読出すステップ 4 7 6 と、ステップ 4 7 6 に続き、台詞情報記憶部 7 2 の台詞情報テーブルから台詞 (i) の発話時間 T_i を読出すステップ 4 7 8 と、発話時間 t_i 及び T_i を用い、ステップ 4 7 6 で読出されたユーザの台詞音声 (i) の発話時間が t_i から T_i となるように、話速変換を行なって、台詞音声 (i) として出力し、処理を終了するステップ 4 8 0 とを含む。

30

【 0 0 9 4 】

図 1 7 は、図 9 に示す第 5 の音声生成部 3 0 8 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 5 を参照して、このプログラムは、ユーザ音声 DB 1 2 0 から録音フラグ (i) を読出すステップ 5 0 0 と、読出された録音フラグの値が 1 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 5 0 2 と、ステップ 5 0 2 において録音フラグの値が 1 ではないと判定されたことに応答して実行され、声優音声 DB 8 0 に格納されている台詞 (i) の声優音声のうち、ユーザの声質に最も類似したものを特定するステップ 5 0 4 と、ステップ 5 0 4 で特定された台詞 (i) の声優音声を、ユーザ音声の特徴を用いて声質変換し、台詞音声 (i) として出力し処理を終了するステップ 5 0 6 とを含む。

40

【 0 0 9 5 】

このプログラムはまた、ステップ 5 0 2 において録音フラグが 1 であると判定されたことに応答して実行され、ユーザ音声 DB 1 2 0 から台詞音声 (i) と発話時間 t_i とを読

50

出すステップ508と、台詞情報記憶部72の台詞情報テーブルから台詞(i)の発話時間 T_i を読み出すステップ510と、ユーザの台詞音声(i)の発話時間が t_i から T_i になるように話速変換を行なって、台詞音声(i)として出力し処理を終了するステップ512とを含む。

【0096】

図18は、図9に示す第6の音声生成部310を実現するためのプログラムのフローチャートである。図18を参照して、このプログラムは、ユーザ音声DB120から録音フラグ(i)を読み出すステップ530と、この録音フラグの値が1か否かを判定し、判定結果にしたがって制御の流れを分岐させるステップ532と、ステップ532において録音フラグの値が1でないと判定されたことに応答して実行され、台詞(i)、ユーザ音声の特徴量、ユーザの母音の音声素片、素片DB82の子音の音声素片を使用して音声合成を行なって台詞音声(i)を生成し出力するステップ534とを含む。

10

【0097】

このプログラムはさらに、ステップ532において録音フラグ=1であると判定されたことに応答して実行され、ユーザ音声DB120から台詞音声(i)と発話時間 t_i とを読み出すステップ536と、台詞情報記憶部72の台詞情報テーブルから台詞(i)の発話時間 T_i を読み出すステップ538と、ユーザの台詞音声(i)の発話時間が t_i から T_i となるようにユーザの台詞音声(i)の話速変換を行なって台詞音声(i)として出力するステップ540とを含む。

【0098】

20

図19は、図9に示す第7の音声生成部312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19を参照して、このプログラムは、声優音声DB80の台詞(i)の音声の中で、ユーザ音声の声質と最も類似した音声を読み出し、台詞音声(i)として出力し、処理を終了するステップ560を含む。

【0099】

図20は、図9に示す第8の音声生成部314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20を参照して、このプログラムは、声優音声DB80に記憶されている、台詞(i)の声優音声のうち、ユーザ音声の声質と最もよく類似した音声を特定し読み出すステップ580と、ステップ580で読み出された音声(i)を、ユーザの発話必須部分の音声の特徴を用いて、ユーザの声質に近い声質に声質変換することにより、ユーザの台詞音声(i)を生成し出力するステップ582とを含む。

30

【0100】

図21は、図9に示す第9の音声生成部316を実現するためのプログラムのフローチャートである。図21を参照して、このプログラムは、台詞(i)、ユーザ音声の特徴量、ユーザの母音の音声素片、及び素片DB82に記憶された全子音の素片を用いて、台詞(i)の音声合成を行ない、台詞音声(i)として出力し処理を終了するステップ600を含む。

【0101】

図22は、図9に示す音声信号処理部320を実現するプログラムのフローチャートである。音声信号処理部320は、合流部292の出力する台詞音声(i)に対し、以下のような処理を行なう。すなわち、このプログラムは、カット情報記憶部76から台詞(i)の音響効果リストE L I S Tを読み出すステップ382と、ステップ382の後、音響効果リストE L I S Tの要素数を変数E M A Xに代入するステップ384と、ステップ384の後、以後の繰返しを制御するための変数kに0を代入するステップ386と、ステップ386に続いて配置され、ステップ388に続き、 $k+1$ の値がE M A Xより大きいか否かを判定し、判定結果にしたがって制御を分岐させるステップ390と、ステップ390において $k+1$ の値がE M A X以下であると判定されたことに応答して実行され、台詞音声(i)にE L I S T[k]の音響効果を付与するステップ392と、ステップ392の後、変数kの値に1を加算してステップ390に制御を戻すステップ388とを含む。

40

【0102】

50

このプログラムはさらに、ステップ390において $k + 1$ の値がE M A Xより大きいと判定されたことに応答して実行され、台詞音声(i)を音声ファイルに書き出すステップ394と、ステップ394の後、台詞音声テーブル88の台詞(i)の音声ファイル名を新たなファイル名で更新して処理を終了するステップ396とを含む。

【0103】

ステップ396で更新される台詞音声テーブル88の構成を図24に示す。図24を参照して、台詞音声テーブル88は、台詞番号と、台詞の再生開始時刻と、台詞の再生(発話)時間と、その台詞の音声(台詞音声)が格納された台詞音声データ86中のファイル名と、再生フラグとを含む。台詞の再生開始時刻は、作成される映画の先頭を所定の時刻とし、その時刻を基準として台詞の再生を開始するように定められる時刻である。再生時間は台詞の再生の継続時間をさす。再生ファイル名は、すでに述べたように台詞音声データ86中で、台詞音声を格納したファイルのファイル名である。再生フラグは、0であれば映画の再生時に音声を再生することを示し、1であれば再生しないことを示す。この再生フラグは、後述するように音声の重なり(二人以上の登場人物が同時に発話すること)を実現するために用いられる。その手法については後述する。

10

【0104】

図23は、本実施の形態に係るマルチメディア制作システム50によって作成された映画を再生する再生システムのブロック図である。図23を参照して、この再生システムは、映像データ66から映像信号と映像・同期信号と効果音の音声信号とを出力するための映像信号再生部620と、映像信号再生部620により再生された映像信号を再生して映像を表示するための表示装置622と、映像信号再生部620が出力する効果音の音声信号を音声に変換して出力するための効果音出力装置624と、映像の再生に先立って、台詞音声データ86及び台詞音声テーブル88を入力として受け、台詞音声テーブル88に記憶された各台詞の発話開始時刻及び発話時間に基づいて互いに同じ時間に重なって発話されるべき台詞の組合せを検出し、それらの音声ファイルの音声を合成して新たな音声ファイルを作成して、重なりが検出された台詞の内の1つの音声ファイルと入替え、さらにそれ以外の台詞音声の発話フラグを「1」に更新することによって、同時に発話する台詞音声統合されるように台詞音声データ86及び台詞音声テーブル88を更新するための同時音声統合処理部632とを含む。

20

【0105】

再生システムはさらに、再生時に映像信号再生部620からの同期信号を受け、台詞音声テーブル88を参照して、同期信号により示される時刻と一致する発話開始時刻の台詞音声であって、かつ対応する再生フラグが「0」であるものを検出して台詞音声データ86から読み出し、再生して音声信号を出力するための同期再生部638と、同期再生部638の出力する音声信号を音声に変換して出力するための台詞音声出力装置640とを含む。

30

【0106】

すなわち、この再生システムは、効果音と、台詞音声とを完全に分離して生成し、台詞音声をその発話開始時間の順番にしたがって、順に再生する。そのため、効果音を活かしながら、登場人物の音声と顔画像とをユーザのものに置換した映画を再生できる。

40

【0107】

図24は、前述したとおり、台詞音声テーブル88の構成を示す。図25は、図24に示す台詞音声テーブル88のうち、発話時間が重なっている台詞(台詞1, 2, 3)を同時音声統合処理部632によって統合した後の台詞音声テーブル88の構成を示す。

【0108】

図25を参照して、台詞音声テーブル88の構成自体は更新前と同様である。異なっているのは、台詞1の再生時間が7秒から11秒に増加していること、台詞1の再生ファイル名が「wave0001.wav」から「comb0001.wav」に変更されていること、及び台詞2及び3の再生フラグが「0」から「1」に変更されていることである。これは以下の理由による。

50

【 0 1 0 9 】

図 2 4 に示す台詞音声テーブル 8 8 において、台詞 1 の再生開始時刻は 0 時 0 分 3 秒、再生時間が 7 秒であるから、再生終了時刻は 0 時 0 分 1 0 秒である。一方、台詞 2 の再生開始時刻は 0 時 0 分 8 秒、再生時間は 5 秒であるから再生終了時刻は 0 時 0 分 1 3 秒となる。すると、台詞 1 の発話時間と台詞 2 の発話時間とは、一部において互いに重なっている。本実施の形態では、このように互いに発話の時間帯が重なっている台詞については、それらの音声を統合して新たな音声ファイルとし、一方の台詞の音声ファイル（通常は再生開始時間の早い方）の音声ファイルと入替え、その発話時間も新たな音声ファイルの発話時間で更新する。そして、他方の台詞の音声ファイルについては再生フラグを 1 とする。

10

【 0 1 1 0 】

図 2 4 に示す例では、台詞 1、2、及び 3 の再生時刻が重なっていたために、これらが統合され、最終的に図 2 5 に示すように台詞 1 の再生時間が 1 1 秒、台詞 2 及び台詞 3 の再生フラグが 1（すなわち再生せず）となっている。

【 0 1 1 1 】

図 2 6 は、この同時音声統合処理部 6 3 2 を実現するためのプログラムのフローチャートである。図 2 6 を参照して、このプログラムは、処理中の台詞の台詞番号を表す変数 X に初期値として 0 を代入するステップ 6 6 0 と、この変数 X に 1 を加算するステップ 6 6 2 と、ステップ 6 6 2 の処理結果を受け、X 番目の台詞（X）の台詞音声が存在するか否か（すなわち全ての台詞音声を処理し終わったか否か）を判定し、判定結果に応じて制御を分岐させるステップ 6 6 4 とを含む。ステップ 6 6 4 では、全ての台詞音声を処理し終わっていれば処理を終了する。

20

【 0 1 1 2 】

このプログラムはさらに、ステップ 6 6 4 において台詞音声（X）が存在すると判定されたことに応答して実行され、台詞音声テーブル 8 8 のその台詞音声（X）の再生フラグの値が 0 か否かを判定し、判定結果に応じて制御を分岐させるステップ 6 6 6 を含む。ステップ 6 6 6 において再生フラグが 0 でないと判定された場合、台詞音声（X）を再生する必要はない。したがってこの場合、制御はステップ 6 6 2 に戻り、次の台詞音声の処理に移る。

【 0 1 1 3 】

このプログラムはさらに、ステップ 6 6 6 で台詞音声（X）の発話フラグの値が 0 であると判定されたことに応答して実行され、台詞音声（X）と音声重なっているか否かを判定する台詞音声の台詞番号を示す変数 Y に X の値を代入するステップ 6 6 8 と、ステップ 6 6 8 の後、この変数 Y の値に 1 を加算するステップ 6 7 0 と、ステップ 6 7 0 の処理結果を受け、台詞音声（Y）が存在するか否か、すなわち全ての台詞音声について台詞音声（X）との重なりを調べる処理が完了したか否かを判定し、判定結果にしたがって制御の流れを分岐させるステップ 6 7 2 とを含む。ステップ 6 7 2 において Y 番目の台詞が存在していないと判定された場合、制御はステップ 6 6 2 に戻る。

30

【 0 1 1 4 】

このプログラムはさらに、ステップ 6 7 2 において Y 番目の台詞音声が存在すると判定されたことに応答して実行され、台詞音声（Y）の再生フラグの値が 0 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 6 7 4 を含む。ステップ 6 7 4 で台詞音声（Y）の再生フラグの値が 0 でないと判定されたときには、制御はステップ 6 7 0 に戻り、次の台詞音声に対して台詞音声（X）との重なりを調べる処理に移る。

40

【 0 1 1 5 】

このプログラムはさらに、ステップ 6 7 4 において台詞音声（Y）の再生フラグの値が 0 であると判定されたことに応答して実行され、台詞音声テーブル 8 8 に記憶された双方の台詞音声の発話開始時間及び発話時間の値に基づき、台詞（X）と台詞（Y）との発話時間の少なくとも一部が重なっているか否かを判定し、判定結果に応じて制御を分岐させるステップ 6 7 6 を含む。ステップ 6 7 6 で発話時間が重なっていないと判定された場合

50

には、制御はステップ 670 に戻る。

【0116】

このプログラムはさらに、ステップ 676 において台詞 (X) と台詞 (Y) との発話時間の少なくとも一部が重なっていると判定されたことに応答して実行され、台詞音声 (X) と台詞音声 (Y) とを混合して新たな台詞音声を作成し、台詞音声 (X) として台詞音声データ 86 を更新するステップ 678 と、この新たな台詞音声 (X) の発話時間 t を、重複修正前の台詞音声 (X) の発話時間 t_x と台詞音声 (Y) の発話時間 t_y との間で、以下のようにして計算して求め、これを新たな台詞音声 (X) の発話時間 t_x として台詞音声テーブル 88 を更新するステップ 680 と、ステップ 680 に続き、台詞音声テーブル 88 の、台詞音声 (Y) の再生フラグの値を「1」に更新し、制御をステップ 670 に戻すステップ 682 とを含む。

10

【0117】

図 27 は、図 23 に示す同期再生部 638 を実現するためのプログラムのフローチャートである。図 27 を参照して、このプログラムは、図 23 に示す映像信号再生部 620 から与えられる同期信号を読み込むステップ 700 と、ステップ 700 で読み込んだ同期信号により示される時刻が、台詞音声テーブル 88 に格納されたいずれかの台詞であって、かつその再生フラグが 0 であるものの発話開始時刻に達したか否かを判定し、判定結果にしたがって制御の流れを分岐させるステップ 702 とを含む。ステップ 702 において、同期信号により示される時刻がいずれの台詞音声の再生開始時間にもなっていないと判定されたときには、制御はステップ 700 に戻り、再度、同期信号を読み込む。

20

【0118】

このプログラムはさらに、ステップ 702 において、同期信号により示される時刻がいずれかの台詞音声の発話開始時刻になったと判定されたことに応答して実行され、その台詞音声の再生を開始し、制御をステップ 700 に戻すステップ 704 とを含む。

【0119】

図 28 は、音声信号処理部 320 が実行する音響効果処理のうち、話速変換と音量正規化処理の内容を説明するための図である。

【0120】

図 28 (A) を参照して、話速変換処理とは、台詞の発話時間の基準となる参照音声 720 での発話時間と比較して、収録音声 722 の収録時間が図 28 (A) に示されるように短すぎたり、逆に長すぎたりした場合に、この収録音声 722 の話速を変換して参照音声 720 の発話時間と等しい発話時間の補正音声 724 を生成する処理のことである。話速変換には、既存の話速変換技術を使用することができる。

30

【0121】

図 28 (B) は、音量正規化を示す。参照音声 740 の平均レベル L_0 と比較して、収録音声 742 の平均レベル L_1 が図 28 (B) に示すように低すぎたり、逆に高すぎたりしたときに、収録音声のレベルを補正して、ほぼ平均レベル L_0 と等しい平均レベル L_3 とするのが音量正規化処理である。このような音量正規化処理は、複数のユーザによって収録される音声の大きさにばらつきがあってはいけなかったり、逆に場面によってはユーザにより音声の大きさに差をつけたりする必要があるために行なわれる。この音量正規化についても、既存の技術を使用することができる。

40

【0122】

[コンピュータによる実現]

図 29 は、このマルチメディア制作システム 50 においてユーザの音声を収録するための台詞音声データ作成部 90 のハードウェアの外観図を示す。図 29 を参照して、台詞音声データ作成部 90 は、実質的にはコンピュータシステム 830 からなる。図 30 は、このコンピュータシステム 830 の内部構成を示す。

【0123】

図 29 を参照して、コンピュータシステム 830 は、リムーバブルメモリ用のメモリポート 852 及び DVD (Digital Versatile Memory) ドライブ

50

850を有するコンピュータ840と、文字情報及びコマンド操作の入力を行なうためのキーボード846と、ポインティングデバイスであるマウス848と、2台のモニター842及び844と、2台のマイクロフォン868及び870と、2組のスピーカセット872及び874とを含む。これらのうち、モニター844、スピーカセット874、及びマイクロフォン868は、コンピュータシステム830の本体部分と分離して図29に示すようにユーザの録音用ブースに設置されており、ユーザの台詞音声の録音時にユーザとの入出力インタフェースとして用いられる。

【0124】

図30を参照して、コンピュータ840は、メモリポート852と、DVDドライブ850と、マイクロフォン868及び870と、スピーカセット872及び874とに加えて、CPU(中央処理装置)856と、CPU856、メモリポート852及びDVDドライブ850に接続されたバス866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)358と、バス866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)860と、バス866、マイクロフォン868及び870、並びにスピーカセット872及び874に接続されるサウンドボード884とを含む。

10

【0125】

コンピュータ840はさらに、他のコンピュータと通信を行なうためのローカルエリアネットワーク(LAN)876への接続を提供するネットワークインターフェイスカード(NIC)878を含む。

20

【0126】

コンピュータシステム830にマルチメディア製作システム50としての動作を行なわせるための、上記した各種のコンピュータプログラムは、DVDドライブ850又はメモリポート852に挿入されるDVD862又はリムーバブルメモリ864に記憶され、さらにハードディスク854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840に送信されハードディスク854に記憶されてもよい。プログラムは実行の際にRAM860にロードされる。DVD862から、リムーバブルメモリ864から、又はネットワーク876を介して、直接にRAM860にプログラムをロードしてもよい。

【0127】

これらのプログラムは、コンピュータ840にこの実施の形態のマルチメディア製作システム50としての動作を行なわせる複数の命令を含む。この動作を行なわせるのに必要な基本的機能のいくつかはコンピュータ840上で動作するオペレーティングシステム(OS)若しくはサードパーティのプログラム、又はコンピュータ840にインストールされる音声処理及び統計モデル処理用の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様に制御されたやり方で適切な機能又は「ツール」を呼出す事により、上記した台詞音声作成装置としての動作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830の動作は周知であるので、ここでは繰返さない。

30

40

【0128】

なお、図1に示されるシステムのうち、台詞音声データ作成部90はユーザごとの収録をするための、いずれもコンピュータシステム830と同様の構成の複数のコンピュータシステムと、音声統合部104を実現するための1つのコンピュータシステムとを含む。音声統合部104を実現するコンピュータシステムも、ハードウェア構成はコンピュータシステム830と同様であるが、マイクロフォン及びスピーカ等は必要ない。

【0129】

また、本実施の形態では、図23に示す映像・音声再生装置92のうち、映像信号再生部620は1つのコンピュータシステムにより実現され、同時音声統合処理部632及び同期再生部638はそれとは別の1つのコンピュータシステムにより実現される。

50

【 0 1 3 0 】

本システムで使用されるコンピュータシステムは、いずれもネットワーク 8 7 6 を介して互いに通信を行ない、最終的に映像データ 6 6、台詞音声データ 8 6、及び台詞音声テーブル 8 8 を映像・音声再生装置 9 2 のハードディスクに作成し、そこから再生を行なう。

【 0 1 3 1 】

〔動作〕

以上に構成を説明したマルチメディア製作システム 5 0 は以下のように動作する。

【 0 1 3 2 】

図 1 を参照して、複数のユーザがマルチメディア製作システム 5 0 を利用するものとして、予め各ユーザには識別情報が割当てられているものとする。また各ユーザには、映画の登場人物の誰と入替わることが決定されているものとする。

10

【 0 1 3 3 】

マルチメディア製作システム 5 0 では、予め映像素材 DB 7 0 には映像素材が、台詞情報記憶部 7 2 には台詞情報が、標準音声記憶部 7 4 には各台詞に対し、男性、女性、年齢に応じた標準音声、カット情報記憶部 7 6 には音響効果情報が、それぞれ格納されている。また、声優音声 DB 8 0 には各台詞を複数の声優がそれぞれ発話したものが台詞別、声優別に格納されているものとする。各声優の音声については、予め音響分析が行なわれており、それぞれの声質を表す音響特徴量が算出されている。また素片 DB 8 2 には、標準音声及び声優音声をセグメンテーションすることによって作成された音声素片が格納されている。各音声素片には、対応する音素の音素ラベルと、音響特徴量と、元の音声の識別子と、発話者の識別子とが付されている。

20

【 0 1 3 4 】

各ユーザのユーザ情報が、ユーザ情報入力部 1 0 0、1 0 0 A、...、1 0 0 N で入力され、画像処理 PC 6 2 及び複数のキャラクタ音声作成部 1 0 2、1 0 2 A、...、1 0 2 N のうち、対応するものに送られる。

【 0 1 3 5 】

三次元スキャナ群 6 0 は、各ユーザの顔をスキャンし、画像処理 PC 6 2 に 3 次元スキャンデータを送る。以下、画像処理 PC 6 2 はユーザの 3 次元スキャンデータを用いてユーザの三次元顔モデルを作成し、さらに任意の角度からの 3 次元顔画像を作成して映像生成装置 6 4 に与える。映像生成装置 6 4 は、登場人物の顔画像を、画像処理 PC 6 2 で作成されたユーザの顔画像で置換し、映像データ 6 6 として出力する。なお、映像データ 6 6 には、音声との同期をとるための同期信号再生用のデータが含まれている。

30

【 0 1 3 6 】

一方、複数のキャラクタ音声作成部 1 0 2、1 0 2 A、...、1 0 2 N は、いずれも、以下のようにして対応するユーザの台詞音声を収録し、この収録音声に基づいて、第 1 の音声生成部 3 0 0 ~ 第 9 の音声生成部 3 1 6 を用いてユーザの声を活かした映画用の音声データを作成し、出力する。このときの第 1 の音声生成部 3 0 0 ~ 第 9 の音声生成部 3 1 6 の処理はいずれも同様である。以下では、キャラクタ音声作成部 1 0 2 の動作について説明する。

40

【 0 1 3 7 】

図 2 を参照して、音声収録部 1 1 4 は、ユーザ情報をユーザ情報入力部 1 0 0 から受信し（図 5 のステップ 1 7 0）、以後の処理ではこのユーザ情報を用いる。続いて、ユーザに割当てられたキャラクタに関する情報が入力される（図 5 のステップ 1 7 2）。音声収録部 1 1 4 は、入力されたキャラクタの台詞に関する台詞情報を台詞情報記憶部 7 2 から読み出し、対応する標準音声を標準音声記憶部 7 4 から、対応する映像がもしあれば映像素材 DB 7 0 から、それぞれ読み出す（図 5 のステップ 1 7 4）。音声収録部 1 1 4 はさらに、ユーザ音声テーブル 2 6 0 を作成し、全ての台詞情報の録音フラグを 0 に初期化する。

【 0 1 3 8 】

音声収録部 1 1 4 は、続いてタイマをスタートさせ（ステップ 1 7 8）、台詞の収録を

50

開始する。台詞の収録では、発話対象の台詞を選択し（ステップ180）、映像と、台詞情報の表示とを行ない（ステップ182）、同時に標準音声の再生を開始する。その結果、入出力装置112の画面（モニタ844の画面）に図7に示すような表示が行なわれる。この後、ユーザが標準音声をまねて、練習としてその発話を行なう（ステップ186）。

【0139】

コンピュータシステム830を操作しながらユーザの発話を聞いているアテンダントが、その発話についての練習を終了してよいか否かを判断し（ステップ188）、もし練習をさらにする必要があるれば（ステップ188においてNO）、その発話について再度同じ処理を繰り返す操作を行なう。練習を終了してよいと判定されると（ステップ188においてYES）、再度選択した台詞と、対応する映像とを表示し（ステップ190）、プログレスバーの表示を開始し（ステップ192）、ユーザの音声を収録する（ステップ194）。

10

【0140】

もしも収録した音声正しい内容で、発声内容も明瞭で、発話時間も許容範囲内であれば、アテンダントは収録した音声を音声ファイルとしてユーザ音声記憶部262に保存し、図8に構成を示すユーザ音声テーブル260の処理中の台詞の行の音声ファイル名の欄にユーザ音声記憶部262に保存した音声ファイルの名称を代入し、発話時間の欄にユーザの台詞音声の実際の発話時間（*t_i*）を代入する（ステップ200）。さらに音声収録部114は、その行の録音フラグを1に更新し（ステップ201）、次の台詞を選択する（ステップ202）。もしも全ての台詞についてのユーザの台詞音声の収録が完了していれば（ステップ204でYES）、収録されたユーザの全発話を音素にセグメンテーションして素片化し（ステップ206）、各音素素片の音響特徴量を算出して（ステップ208）、素片DB82に追加する。

20

【0141】

もしもステップ204でまだ全ての台詞についての台詞音声の収録が終わっていないと判定されると、ステップ212でタイマを参照し、予め収録時間として定められていた時間を超過しているか否かを判定する。もしも超過していれば、ステップ206に進み、以後は全ての台詞について収録を完了した場合と同様の処理が行なわれる。もしもまだ所定時間に達していいなければ、図5のステップ182に戻り、このユーザに対応するキャラクターの次の台詞について、上述した処理を繰り返す。

30

【0142】

仮にステップ198で、収録された音声好ましくないもの（たとえば内容が本来の発話テキストと著しく異なっているもの、発話が不明瞭なもの、発話時間が許容範囲外のもの）であるとアテンダントが判定したときには、ステップ214でその収録音声破棄される。続いてタイマをチェックすることで、収録のための時間を超過しているか否かを判定する（ステップ216）。時間が超過していなければ、処理中の台詞の標準音声による発声（ステップ182）からやり直すか、単にユーザによる発話の収録（ステップ190）からやり直すかをアテンダントが判定し、判定結果にしたがって指示を入力する。音声収録部114は、その指示にしたがって制御を分岐させ（ステップ220）、その結果、ステップ182又はステップ190から処理が再開される。

40

【0143】

一方、ステップ216ですでに収録に要した時間が、所定の時間を超過していると判定された場合には、ステップ218で現在収録中の台詞が必須部分の台詞か否かを判定する。必須部分であれば、この収録は必ずする必要があるので、制御はステップ220に進み、アテンダントの判定にしたがって、収録を再開する。もしも必須部分でなければ、収録作業を終了すべきであるから、制御はステップ206に進む。以後、全ての台詞の収録が完了したときと同様の動作がステップ206、208及び210で実行される。

【0144】

こうして、音声収録部114によって、図8に示すユーザ音声記憶部262には、ある

50

キャラクタの台詞についてのユーザの台詞音声の音声ファイルが格納され、ユーザ音声テーブル260には各台詞について、録音できたか否かを示す録音フラグと、ユーザ音声記憶部262中の対応する音声ファイルの名称と、ユーザによる発話時間とが記録される。

【0145】

複数のキャラクタ音声作成部102, 102A, ..., 102Nの各々が上記した処理を実行する結果、これらからはそれぞれのキャラクタの台詞音声ユーザ音声DB120(ユーザ音声テーブル260とユーザ音声記憶部262)の形で出力される。音声統合部104は、これら種々のキャラクタのユーザの台詞音声を台詞情報記憶部72に記憶された台詞情報に基づいて所定の順番で読出せるよう統合し、台詞音声データ86及び台詞音声テーブル88を出力する。音声収録部114は、このようにして、対象のユーザについて

10

の音声の収録が完了すると合成手法決定部116に対し、台詞音声の生成を開始するよう指示を出す。

【0146】

この指示に回答して、対応の合成手法決定部116は以下のような処理を実行する。図10を参照して、ステップ340~ステップ344によって、処理すべき台詞のうち、先頭の台詞を選択する。そして、その台詞の台詞番号をキーに、手法リストテーブル78を検索し、その台詞に対する手法リストWLISTを入手する。

【0147】

続いてステップ348~354の処理により、手法リストWLISTに記載された手法について、先頭から順番に調べ、利用可能な手法で最初に発見された手法を用い、処理対象の台詞をその手法で処理することを決定し、その手法を特定する情報を音声作成部118に与え、処理させる。手法リストは必ずその中に利用可能なものがあるように作成するが、仮にない場合でもデフォルトの手法を用いて台詞音声の生成ができるようにしておく。

20

【0148】

このようにして処理対象の中の最初の台詞について、ユーザの収録音声に基づいて、音声作成部118の第1~第9の中で、選択された手法に対応するものに対し、台詞音声の生成を指示する。このとき合成手法決定部116は、分岐部280を制御して、選択された音声生成部にユーザ音声を与え、その出力する台詞音声を選択して出力するように合流部292を制御する。こうして、先頭の台詞について台詞音声の生成を開始させると、合成手法決定部116は再度ステップ342から処理を再開し、次の台詞について台詞音声生成の手法を決定し、対応する音声生成部に台詞音声を生成させる。こうして対象となるキャラクタの台詞について全て台詞音声の生成が完了すると、合成手法決定部116の処理は終了である。

30

【0149】

図9を参照して、音声作成部118は以下のように動作する。分岐部280は、合成手法決定部116からの指示にしたがい、指定された音声生成部を能動化し、ユーザ音声を与える。第1の音声生成部300~第9の音声生成部316のうち、能動化されたものは、与えられたユーザ音声に基づき、それぞれの手法を用いて台詞音声を生成する。出力される台詞音声は合流部292によって選択され、音声信号処理部320に与えられる。

40

【0150】

ここで、第1の手法が選択された場合、図9に示す第1の音声生成部300は、ユーザ音声DB120から台詞音声(i)を読出す(ステップ380)。ステップ380によってこの処理は終了する。

【0151】

第2の手法が選択された場合、図9に示す第2の音声生成部302は以下のように動作する。図14を参照して、まず、第2の音声生成部302は、ユーザ音声DB120からユーザの台詞音声(i)とその発話時間t_iとを読出す(ステップ410)。第2の音声生成部302は続いて、台詞情報テーブル台詞(i)の発話時間T_iを読出す(ステップ412)。さらに、第2の音声生成部302は、ステップ410で読出した発話時間t_i

50

とステップ412で読出した発話時間 T_i とを用い、ユーザの台詞音声(i)の発話時間が t_i から T_i となるように話速変換を行なう(ステップ414)。

【0152】

第3の手法が選択された場合、図9に示す第3の音声生成部304は以下のように動作する。図15を参照して、第3の音声生成部304は、まずユーザ音声DB120のユーザ音声テーブル260から録音フラグ(i)を読出す(ステップ440)。次に第3の音声生成部304は、読出された録音フラグの値が1か否かを判定し、録音フラグが1でないときには標準音声記憶部74から台詞(i)の標準音声を読出し、台詞音声(i)として出力し、処理を終了する(ステップ444)。ステップ442において録音フラグが1であると判定された場合には、ユーザ音声DB120から台詞音声(i)と発話時間 t_i とを読出し(ステップ446)、台詞情報記憶部72に記憶された台詞情報テーブルから台詞(i)の発話時間 T_i を読出す(ステップ448)。そして、ステップ446及び448でそれぞれ読出された発話時間 t_i 及び T_i を用いて、ユーザの台詞音声(i)の発話時間が T_i となるように、話速変換を行なって出力する(ステップ450)。

10

【0153】

第4の手法が選択された場合、図9に示す第4の音声生成部306は以下のように動作する。図16を参照して、第4の音声生成部306は、ユーザ音声DB120のユーザ音声テーブル260から i 番目の台詞音声に対する録音フラグ(i)を読出す(ステップ470)。次に、ステップ470で読出された録音フラグ(i)の値が1でない場合、声優音声DB80中に記憶されている台詞(i)の声優音声のうち、ユーザ音声と最も声質が類似したものを読出して台詞音声(i)として出力する(ステップ474)。ステップ472において録音フラグが1であると判定されれば、第3の音声生成部304は、ユーザ音声DB120から台詞音声(i)と発話時間 t_i とを読出す(ステップ476)。次に、台詞情報記憶部72の台詞情報テーブルから台詞(i)の発話時間 T_i を読出し(ステップ478)、発話時間 t_i 及び T_i を用い、読出されたユーザの台詞音声(i)の発話時間が t_i から T_i となるように、話速変換を行なって、台詞音声(i)として出力する(ステップ480)。

20

【0154】

第5の手法が選択された場合、図9に示す第5の音声生成部308は以下のように動作する。図17を参照して、第5の音声生成部308は、ユーザ音声DB120から録音フラグ(i)を読出す(ステップ500)。読出された録音フラグの値が1ではない場合、声優音声DB80に格納されている台詞(i)の声優音声のうち、ユーザの声質に最も類似したものを特定し(ステップ504)、ステップ504で特定された台詞(i)の声優音声を、ユーザ音声の特徴を用いて声質変換し、台詞音声(i)として出力し処理を終了する(ステップ506)。ステップ502において録音フラグが1である場合、第5の音声生成部308は、ユーザ音声DB120から台詞音声(i)と発話時間 t_i とを読出す(ステップ508)。次に、台詞情報記憶部72の台詞情報テーブルから台詞(i)の発話時間 T_i を読出す(ステップ510)。最後に、ユーザの台詞音声(i)の発話時間が t_i から T_i になるように話速変換を行なって、台詞音声(i)として出力し処理を終了する(ステップ512)。

30

40

【0155】

第6の手法が選択された場合、図9に示す第6の音声生成部310は以下のように動作する。図18を参照して、第6の音声生成部310は、ユーザ音声DB120から録音フラグ(i)を読出す(ステップ530)。この録音フラグの値が1でなければ、台詞(i)、ユーザ音声の特徴量、ユーザの母音の音声素片、素片DB82の子音の音声素片を使用して音声合成を行なって台詞音声(i)を生成し出力する(ステップ534)。録音フラグ=1であれば、第6の音声生成部310は、ユーザ音声DB120から台詞音声(i)と発話時間 t_i とを読出す(ステップ536)。次に、台詞情報記憶部72の台詞情報テーブルから台詞(i)の発話時間 T_i を読出す(ステップ538)。最後に、ユーザの台詞音声(i)の発話時間が t_i から T_i となるようにユーザの台詞音声(i)の話速変

50

換を行なって台詞音声 (i) として出力する (ステップ 5 4 0) 。

【 0 1 5 6 】

第 7 の手法が選択された場合、図 9 に示す第 7 の音声生成部 3 1 2 は以下のように動作する。図 1 9 を参照して、第 7 の音声生成部 3 1 2 は、声優音声 D B 8 0 の台詞 (i) の音声の中で、ユーザ音声の声質と最も類似した音声を読み出し、台詞音声 (i) として出力し、処理を終了する (ステップ 5 6 0) 。

【 0 1 5 7 】

第 8 の手法が選択された場合、第 8 の音声生成部 3 1 4 は以下のように動作する。図 2 0 を参照して、第 8 の音声生成部 3 1 4 は、声優音声 D B 8 0 に記憶されている、台詞 (i) の声優音声のうち、ユーザ音声の声質と最もよく類似した音声を特定し読み出す (ステップ 5 8 0) 。次に、ステップ 5 8 0 で読み出された音声 (i) を、ユーザの発話必須部分の音声の特徴を用いて、ユーザの声質に近い声質に声質変換することにより、ユーザの台詞音声 (i) を生成し出力する (ステップ 5 8 2) 。

【 0 1 5 8 】

第 9 の手法が選択された場合、図 9 に示す第 9 の音声生成部 3 1 6 は以下のように動作する。図 2 1 を参照して、第 9 の音声生成部 3 1 6 は、台詞 (i) 、ユーザ音声の特徴量、ユーザの母音の音声素片、及び素片 D B 8 2 に記憶された子音の素片を用いて、台詞 (i) の音声合成を行ない、台詞音声 (i) として出力し処理を終了する (ステップ 6 0 0) 。

【 0 1 5 9 】

合流部 2 9 2 から出力される台詞音声には、いずれも図 9 に示す音声信号処理部 3 2 0 により以下のようにしてカット情報記憶部 7 6 により指定される音響効果が追加される。すなわち、図 2 2 を参照して、音声信号処理部 3 2 0 は、合流部 2 9 2 の出力する台詞音声 (i) に対し、カット情報記憶部 7 6 から台詞 (i) の音響効果リスト E L I S T を読み出す (ステップ 3 8 2) 。音声信号処理部 3 2 0 はさらに、音響効果リスト E L I S T の要素を順番に調べ、それらの要素により特定される音響効果を台詞音声 (i) に全て加え、その後、音響効果が加えられた台詞音声 (i) を音声ファイルに書き出す (ステップ 3 9 4) 。このとき、音量正規化処理 (図 2 8 (B)) 等の処理も同時に実行する。音声信号処理部 3 2 0 は、この後、台詞音声テーブル 8 8 の台詞 (i) の音声ファイル名を新たなファイル名で更新して処理を終了する (ステップ 3 9 6) 。

【 0 1 6 0 】

この音声信号処理部 3 2 0 の機能により、図 2 4 に示すような台詞音声テーブル 8 8 と、音響効果が加えられた台詞音声データ 8 6 とが生成される。

【 0 1 6 1 】

このようにして、全てのキャラクタの全ての台詞について台詞音声を作成され、それらに対応する台詞音声データ 8 6 と台詞音声テーブル 8 8 とが作成されると、映像データ 6 6 とあわせて映像・音声再生装置 9 2 により映画を再生することができる。このときには、図 2 3 に示す映像信号再生部 6 2 0 、同時音声統合処理部 6 3 2 、及び同期再生部 6 3 8 は以下のように動作する。

【 0 1 6 2 】

最初に、同時音声統合処理部 6 3 2 は、図 2 6 に示すような制御構造を持つプログラムにより、互いに重なっている台詞の音声を 1 つのファイルに統合し、それにあわせて台詞音声テーブル 8 8 の音声ファイル名称を更新し、さらに統合により再生不要となった音声ファイルに対応する再生フラグを 1 とする処理を実行する。この処理により最終的に再生可能な台詞音声データ 8 6 及び台詞音声テーブル 8 8 が生成される。このときの台詞音声テーブル 8 8 の各台詞の再生開始時刻には、各台詞の再生を開始する時刻が記録されている。

【 0 1 6 3 】

映画の再生が開始されると、映像信号再生部 6 2 0 は映像信号と背景音等の効果音を示す音声信号を再生し、表示装置 6 2 2 及び効果音出力装置 6 2 4 にそれぞれ与える。表示

10

20

30

40

50

装置 6 2 2 はこの映像信号を再生し、映像を表示する。効果音出力装置 6 2 4 は効果音の音声信号を音声に変換する。この映画の登場人物の顔画像は、ユーザの顔画像と入替えられている。

【 0 1 6 4 】

一方、映像信号再生部 6 2 0 は、映像信号の再生と同期して映像データ中に記録されている同期データに基づいて、同期信号を生成し同期再生部 6 3 8 に与える。

【 0 1 6 5 】

同期再生部 6 3 8 は、この同期信号を常に監視し、同期信号により表される時刻が台詞音声テーブル 8 8 に記憶されている台詞音声の再生開始時刻と一致すると、その台詞音声を再生し台詞音声出力装置 6 4 0 に与える。台詞音声出力装置 6 4 0 はこの音声を再生する。台詞音声は、上記したいずれかの手法にしたがって再生又は合成された音声である。この音声は、基本的には各ユーザの音声そのままか、それが話速変換されたものか、又はできるだけユーザの音声の声質に似るように選択され、又は合成された音声である。もちろん、中には標準音声をそのまま再生する場合もあり得るが、台詞の全体を見ると、それぞれのキャラクタの声は、対応するユーザの声質に似たものを感じられる。

【 0 1 6 6 】

[第 2 の実施の形態]

以下、本発明の第 2 の実施の形態に係るマルチメディア製作システムの構成及び動作について説明する。この第 2 の実施の形態に係るシステムは、第 1 の実施の形態に係るマルチメディア製作システム 5 0 の構成のうち、音声収録部 1 1 4 を以下に説明する音声収録部で置き換えた構成を有する。それ以外の部分の構成は、マルチメディア製作システム 5 0 のものと同じである。したがって、以下では第 2 の実施の形態に係るシステムの音声収録部のみについてその構成及び動作を説明する。

【 0 1 6 7 】

第 1 の実施の形態では、ユーザ音声収録時には、アシスタントによる補助作業が必要であった。できればそのようなアシスタントの介在なしにユーザの音声を効率よく収録できると好ましい。第 2 の実施の形態の音声収録部は、アシスタントなしで、ユーザが効率よく映画の登場人物の台詞を効率よく収録できるようにするための種々の機能を備えたものである。そのために本実施の形態では、音声収録部にタッチパネルを使用する。そして、同じ映像を繰返し表示しながらユーザが台詞を繰返して発声し、うまく発声できたと思われる音声のみを収録音声として保存する。

【 0 1 6 8 】

図 3 1 は、第 2 の実施の形態におけるシステムの音声収録部で使用されるタッチパネル 9 0 0 の画面とその表示例とを示す模式図である。図 3 1 を参照して、このタッチパネル 9 0 0 の画面には、音声の収録対象となる映画のシーンが繰返して表示される映像表示領域 9 0 2 と、映像表示領域 9 0 2 に表示されているシーンでユーザが発話すべき台詞 9 1 4 を映像と収録時刻の進行とにあわせて同期して表示するための台詞表示領域 9 0 4 と、ユーザにより入力された音声のレベルを示すレベルメータ 9 0 6 と、映像の表示とともに、予め準備されている標準音声による台詞の再生を行なうか否かを指定するボイスオーバーボタン 9 1 0 と、繰返して台詞を発声するユーザが、うまく発声できたと思ったときに押すことにより、直前に収録された音声を保存するようにシステムに指示するために操作するストップボタン 9 0 8 とが表示される。映像表示領域 9 0 2 の右上には、シーンの進行にあわせた時間情報画像 9 1 2 がリアルタイムで表示される。

【 0 1 6 9 】

レベルメータ 9 0 6 はユーザにより入力された音声のレベルをバー形式で表示するものである。レベルメータ 9 0 6 の各ブロックのうち、下側の複数個のブロックは、適正範囲の音声レベルを緑色で表示するためのものである。上の 2 つのブロックは、音声レベルが適正範囲を超えたことを赤色で表示するためのものである。レベルメータ 9 0 6 の表示を見ることで、ユーザは、自分の発声が適正なレベルか否かをリアルタイムで判定することができる。

10

20

30

40

50

【 0 1 7 0 】

ボイスオーバーボタン 9 1 0 は、映像ファイル 9 4 6 の各シーンごとに予め準備された音声ファイルを映像の再生と同時に再生するか否かをシステムに指示するためのものである。このように映像の再生と同時に台詞の再生を行なうことを「ボイスオーバー (V O) 」と呼ぶ。ボイスオーバーボタン 9 1 0 のうち「 O N 」ボタンを押せばボイスオーバーフラグがセットされ、「 O F F 」ボタンを押せばボイスオーバーフラグがリセットされる。

【 0 1 7 1 】

図 3 2 は、この実施の形態に係るマルチメディア製作システムで使用される音声収録部 9 3 0 のブロック図である。図 3 2 を参照して、音声収録部 9 3 0 は、タッチパネル 9 0 0 に加え、映画の映像を複数個のシーンに分割したものをそれぞれ保持した複数個の映像ファイル 9 4 6 と、映画の台詞を、映像ファイル 9 4 6 の各映像ファイルのシーンに対応して予め分割してテキストファイルとして格納した複数個の台詞ファイル 9 4 8 と、映像表示領域 9 0 2 に表示される映像の、表示開始からの経過時間を計時するためのタイマ 9 5 0 と、ヘッドセット 9 4 4 と、ヘッドセット 9 4 4 が接続されたサウンドボード 8 8 4 と、サウンドボード 8 8 4 にヘッドセット 9 4 4 から入力された音声信号のレベルを測定し、レベル信号を出力するレベル測定部 9 5 2 と、映像ファイル 9 4 6 、台詞ファイル 9 4 8 、タイマ 9 5 0 、及びレベル測定部 9 5 2 の出力に接続され、図 1 に示すような表示画面をリアルタイムで作成しタッチパネル 9 0 0 に表示させるための表示作成部 9 3 2 とを含む。

【 0 1 7 2 】

音声収録部 9 3 0 はさらに、タッチパネル 9 0 0 に対する利用者のタッチ操作を検出して、利用者による命令を判定し対応する命令信号を出力したり、入力された情報を出力したりする G U I (G r a p h i c a l U s e r I n t e r f a c e) 部 9 3 6 と、サウンドボード 8 8 4 がヘッドセット 9 4 4 から受信する音声信号に対する録音のための処理をするための録音処理部 9 4 2 と、録音処理部 9 4 2 により処理される音声信号の直近の所定時間分を記憶するためのリングバッファ 9 6 2 と、映画の台詞の音声、映像ファイル 9 4 6 の各映像ファイルのシーンに対応して予め分割して保持する複数の音声ファイル 9 6 0 と、音声ファイル 9 6 0 のうち、表示作成部 9 3 2 によって表示されているシーンに対応するものを読み出し、再生してサウンドボード 8 8 4 を介してヘッドセット 9 4 4 に与え、音声として出力させるための音声再生部 9 4 0 とを含む。

【 0 1 7 3 】

音声収録部 9 3 0 はさらに、 G U I 部 9 3 6 から与えられる信号、タイマ 9 5 0 の出力にしたがって、表示作成部 9 3 2 、音声再生部 9 4 0 及び録音処理部 9 4 2 を繰返し制御して音声収録処理を実行し、ストップボタン 9 0 8 (図 3 1 参照) が操作されたときに、リングバッファ 9 6 2 に記憶されている音声のうち、収録中のシーンの長さに対応する時間分を取出して保存する処理を各台詞に対して実行するための制御部 9 3 8 と、制御部 9 3 8 により使用される、ボイスオーバーがオンかオフかを示す V O フラグを記憶するための V O フラグ記憶部 9 5 4 と、音声収録の対象となっている利用者が選択する対象となる、映画のキャラクタ情報を記憶するためのキャラクタ情報記憶部 9 5 6 と、制御部 9 3 8 により保存される音声をシーンごとに記憶する複数の録音ファイル 9 6 4 と、ネットワークに対するアクセスを提供する N I C 8 7 8 とを含む。

【 0 1 7 4 】

映像ファイル 9 4 6 、台詞ファイル 9 4 8 、音声ファイル 9 6 0 及び録音ファイル 9 6 4 は、本実施の形態ではハードディスクに記憶される。 V O フラグ記憶部 9 5 4 、キャラクタ情報記憶部 9 5 6 、リングバッファ 9 6 2 等の領域は R A M 中に設けられる。

【 0 1 7 5 】

制御部 9 3 8 、表示作成部 9 3 2 、音声再生部 9 4 0 及び録音処理部 9 4 2 は実質的には一つのマイクロプロセッサにより実行される複数のプロセスによって実現される。これらプロセスは、図 3 2 中には図示しないメモリに記憶された、後述する制御構造を有するプログラムをマイクロプロセッサが実行することにより実現される。これら機能のうち、

10

20

30

40

50

音声再生部 9 4 0 及び録音処理部 9 4 2 についてはその構成は周知であるので、ここではそれらについての詳細な説明は行なわない。

【 0 1 7 6 】

図 3 3 は、音声収録部 9 3 0 のうち、制御部 9 3 8 の機能を実現するためにマイクロプロセッサが実行するプログラムの主要部の制御構造を示すフローチャートである。図 3 3 を参照して、このプログラムは、音声収録部 9 3 0 の電源が投入されることにより実行を開始し、所定の初期画面をタッチパネル 9 0 0 上に表示するステップ 1 7 0 を含む。初期画面は図示しないが、映画の中のどのキャラクタを選択するかに関する入力欄、利用者に関する情報の入力欄、音声収録のスタートボタン、及び図 3 1 に示すボイスオーバーボタン 9 1 0 と同様のボイスオーバーボタンを含んでいる。この画面はタッチパネル 9 0 0 の画面に表示され、GUI 部 9 3 6 を通じて利用者の入力及び指示が制御部 9 3 8 に与えられる。

10

【 0 1 7 7 】

このプログラムはさらに、ステップ 1 7 0 に続き、利用者の入力に基づいて利用者に映画のキャラクタを割当て、その情報をキャラクタ情報記憶部 9 5 6 に記憶するステップ 1 7 2 と、ステップ 1 7 2 で利用者に割当てられたキャラクタに応じた台詞（キャラクタ台詞）及び音声と、他のキャラクタと共通に割当てられた台詞及び音声とをNIC 8 7 8 を介してサーバに記憶された、処理対象の映画に関する情報から抽出しキャラクタ情報記憶部 9 5 6、映像ファイル 9 4 6、台詞ファイル 9 4 8、及び音声ファイル 9 6 0 に書込むステップ 1 7 4 と、ステップ 1 7 4 の後、図 3 2 に示す録音ファイル 9 6 4 及びリングバッファ 9 6 2 を未収録状態に初期化するステップ 1 7 6 とを含む。

20

【 0 1 7 8 】

このプログラムはさらに、GUI 部 9 3 6 の出力に基づいて、収録処理の開始を指示するスタートキー（図示せず）が押されたか否かを判定し、判定結果に応じて制御の流れを分岐させるステップ 9 7 0 と、ステップ 9 7 0 においてスタートキーが押されていないと判定されたときに実行され、ボイスオーバーボタンのいずれかが押されたか否かを判定し、判定結果に応じて制御を分岐させるステップ 9 7 2 と、ステップ 9 7 2 でボイスオーバーボタンが押されたらと判定されたときに、VO フラグ記憶部 9 5 4 に記憶された VO フラグの値を利用者の指示に応じて更新し、制御をステップ 9 7 0 に戻すステップ 9 7 4 とを含む。ステップ 9 7 2 でボイスオーバーボタンが押されていないと判定されたときには制御はステップ 9 7 0 にもどる。

30

【 0 1 7 9 】

このプログラムはさらに、ステップ 9 7 0 でスタートボタンが押されたらと判定されたことに応答して実行され、最初の映像シーン、台詞、音声を選択し、それぞれ映像ファイル 9 4 6、台詞ファイル 9 4 8、及び音声ファイル 9 6 0 から対応するファイルを読み出すステップ 9 7 5 と、ステップ 9 7 5 の後、読み出された映像ファイル、台詞ファイル、及び音声ファイルを使用して図 3 1 に示すような表示を行なうことにより、収録対象のシーンを所定時間ごとに繰り返し再生する処理を開始させるステップ 9 7 6 とを含む。本実施の形態では、所定時間ごとにタッチパネル 9 0 0 の画面を更新する処理を、図 3 3 に示すものは別のプログラムスレッドで実行する。

40

【 0 1 8 0 】

なお、図 3 1 に示す例では、映像表示領域 9 0 2 には選択されたシーンの映像がムービーとしてストップボタン 9 0 8 が押されるまで繰り返し表示される。台詞 9 1 4 は、映像の表示が開始されると表示されるが、台詞の進行にあわせて台詞 9 1 4 の文字の色を左から順番に変えながら表示される。このためには、いわゆるカラオケで使用されている歌詞表示の技術を使用すればよい。

【 0 1 8 1 】

このプログラムはさらに、ステップ 9 7 6 に続いて実行され、録音処理部 9 4 2 に指示して利用者の音声をリングバッファ 9 6 2 に録音する処理を開始させるステップ 9 7 8 と、ステップ 9 7 8 に続き、図 1 に示すボイスオーバーボタン 9 1 0 のいずれかが押されたら

50

否かを判定するステップ980と、ボイスオーバーボタン910のいずれかが押されたときに実行され、押されたボタンにしたがってVOフラグの値を更新するステップ982と、ステップ980でボイスオーバーボタン910が押されていないと判定されたときに、ストップボタン908が押されたか否かを判定するステップ984とを含む。

【0182】

ステップ984でストップボタン908が押されていないと判定された場合には制御はステップ980に戻る。

【0183】

このプログラムはさらに、ステップ984でストップボタン908が押されたと判定されたときに実行され、録音処理部942に指示してリングバッファ962への追加の録音を終了させるステップ988と、ステップ988の後、リングバッファ962に格納された音声のうち、直前に再生されていたシーンの持続時間に相当する時間分だけさかのぼった部分までを取り出し、録音ファイル964としてハードディスクに追加して格納するステップ990と、ステップ990に続いて再生処理を停止させるステップ991と、ステップ991に続き、次のシーンを選択するステップ992と、ステップ992に続き、ステップ992で全てのシーンに関する音声の収録が終了したか否かを判定し、終了していない場合にはステップ976に制御を分岐させるステップ994と、ステップ994で全シーンの音声の収録が終了していると判定されたことに応答して、収録された録音ファイル964を利用者に関する情報とともにサーバに送信するステップ994とを含む。サーバでは、これら音声を素片に分解し、素片DBに組込む。

【0184】

図34は、図32の表示作成部932に相当する処理をプロセッサにより実現するための、プロセッサによって実行されるプログラムの制御構造を示すフローチャートである。このプログラムがプロセッサによって実行されるときには、音声収録処理のプロセスとは別プロセスとして実行され、音声収録処理から表示作成処理を終了する指示をプロセッサ間通信によって受けるまで、同じ映像ファイル、台詞ファイル、及び音声ファイルの再生を繰返して実行する。

【0185】

図34を参照して、このプログラムは、音声収録処理によって起動され、音声収録処理によって指定された映像ファイル、台詞ファイル、及び音声ファイルをオープンするステップ1100と、ステップ1100に続き、映像ファイル及び台詞ファイルの続き（最初の場合には先頭）のレコードを読み出すステップ1101と、ステップ1101に続き、指定された映像ファイルの終わりに達したか否かを判定し、判定結果に応じて制御の流れを分岐させるステップ1102と、ステップ1102で映像ファイルの終わりに達したと判定されたときに実行され、現在開いている映像ファイル及び台詞ファイルの読み出位置を先頭レコードに戻して読み出すステップ1104と、ステップ1102でファイルの終わりに達していないと判定された場合、及びステップ1104の処理実行後に実行され、ステップ1101又はステップ1104で読み出されたファイル内容に対してデコード等の処理をした上で、映像メモリに映像として書込むステップ1106とを含む。

【0186】

このプログラムはさらに、ステップ1106に続き、図32に示すレベル測定部952から出力される、利用者からの音声の音声レベル信号を入力ポートから読むステップ1114と、読み出した音声レベルにしたがって音声レベルメーカの画像を作成し、映像メモリに書込むステップ1116と、ステップ1116の後、タイマ950から時刻情報を読み出すステップ1118と、ステップ1118に続き、読み出した時刻情報にしたがって時間情報画像912を生成し、映像メモリに書込むステップ1120と、映像メモリの内容をタッチパネル900の表示用フレームメモリに転送するステップ1122とを含む。

【0187】

このプログラムはさらに、ステップ1122に続き、VOフラグがセットされているか否かを判定して、判定結果に応じて制御の流れを分岐させるステップ1130と、ステッ

10

20

30

40

50

ブ 1 1 3 0 で V O フラグがセットされていると判定されたときに実行され、音声収録プログラムによって指定された音声ファイルのレコードの続きを読出すステップ 1 1 3 2 と、ステップ 1 1 3 2 の結果、読出位置が音声ファイルの終りに到達したか否かを判定し、判定結果に応じて制御を分岐させるステップ 1 1 3 4 と、ステップ 1 1 3 4 で読出位置が音声ファイルの終りに到達したと判定されたときに実行され、音声ファイルの読出位置をその先頭位置に戻してその内容を読出すステップ 1 1 3 6 と、ステップ 1 1 3 0、又はステップ 1 1 3 6 で読出された音声ファイルのレコードをデコードして音声の出力ポートに順次出力する処理を実行し、レコードの終りまでデータを処理すると制御をステップ 1 1 0 1 に戻すステップ 1 1 3 8 とを含む。

【 0 1 8 8 】

10

〔動作〕

この第 2 の実施の形態に係る音声収録部 9 3 0 は以下のように動作する。音声収録部 9 3 0 の電源が投入されると、音声収録部 9 3 0 の内部は初期化され、タッチパネル 9 0 0 の画面には初期画面が表示される（図 3 3、ステップ 1 7 0）。この初期画面では、利用者へのキャラクタの割当て、ユーザ情報の入力等が行なわれる（ステップ 1 7 2）。利用者にキャラクタが割当てられると、N I C 8 7 8 を介してサーバをアクセスし、割当てられたキャラクタに関する台詞及び音声、並びに共通の台詞及び音声のファイルをサーバから抽出し、映像ファイル 9 4 6、台詞ファイル 9 4 8、及び音声ファイル 9 6 0 として音声収録部 9 3 0 内に保存する。

【 0 1 8 9 】

20

音声収録部 9 3 0 はさらに、音声収録部 9 3 0 内の録音ファイル 9 6 4 及びリングバッファ 9 6 2 の内容をクリアし、音声収録部 9 3 0 を未収録状態に初期化する（ステップ 1 7 6）。この処理により、収録対象となる音声以外の音声がリングバッファ 9 6 2 に残っていたりして、収録対象となる音声を正しく収録できないという事態を避けることができる。

【 0 1 9 0 】

続いて利用者によりスタートボタン又はボイスオーバーボタンが押されるまで待機し、ボイスオーバーボタンが押されたら押されたボタンにしたがって V O フラグ記憶部 9 5 4 の値を更新する。スタートボタンが押されると、音声の収録が開始される。

【 0 1 9 1 】

30

表示作成部 9 3 2 は、音声収録の対象となっている映画のうち、最初のシーンを選択し（ステップ 9 7 5）、そのシーンを指定して表示作成処理を開始させる（ステップ 9 7 6）。この処理により、図 3 4 に示す表示作成プログラムの実行が開始される。この表示作成プログラムは音声収録プログラムと並行して実行される。まず、表示作成プログラムにより表示作成処理がどのように行なわれるかを説明する。

【 0 1 9 2 】

図 3 4 を参照して、ステップ 1 1 0 0 で、音声収録処理により指定されたシーンの映像ファイル及び台詞ファイルの先頭のレコードを読出す。ファイルの終りではないので、それらのレコードの内容を画像メモリに書込む（ステップ 1 1 0 6）。さらにレベル測定部 9 5 2 から利用者の入力音声レベルを示す信号を読み（ステップ 1 1 1 4）、その値にしたがってレベルメータ画像を生成し、画像メモリ中の所定位置に書き込む（ステップ 1 1 1 6）。続いてタイマの出力を読み（ステップ 1 1 1 8）、タイマ画像を生成して画像メモリに書込み、さらに、タイマ出力に応じ、時間が経過した部分までの台詞の文字を、ステップ 1 1 0 6 と異なる色で画像メモリに書き込む（ステップ 1 1 2 0）。ステップ 1 1 2 2 で画像メモリの内容をタッチパネル 9 0 0 に転送して画像を表示させる。

40

【 0 1 9 3 】

続いて V O フラグがセットされているか否かを判定し（ステップ 1 1 3 0）、セットされていないならば、この繰り返し処理が開始してから所定時間が経過するのを待って制御をステップ 1 1 0 1 に戻す。V O フラグがセットされていれば、処理中のシーンに対応する音声ファイルを読出し（ステップ 1 1 3 2 ~ 1 1 3 6）、音声の再生を行なって（ステップ

50

1138) からステップ1101に制御を戻す。

【0194】

このように、表示作成処理プログラムは、所定時間ごとに各ファイルの内容を読み出し、再生する処理を繰り返す。表示作成処理プログラムは、図33のステップ991で再生終了処理が実行されるまで、上記した処理を繰り返して実行する。

【0195】

再び図33を参照して、ステップ978で録音処理部942を制御して録音処理を開始する。録音処理部942は独立した処理であり、この音声収録処理とは別プロセスで並列に実行を続ける。

【0196】

このようにして、音声収録処理では、処理対象のシーンの映像が繰り返して再生される。その映像においては、台詞がシーンの進行時刻に応じて変化する色で表示され、同時にシーンの進行時刻に関する時刻情報が映像の左上に重ねてリアルタイムで表示される。VOフラグがセットされていれば、そのシーンに対応する台詞が標準的な音声で再生される。VOフラグは、音声収録処理の実行中であれば、いつでもボイスオーバーボタン910を操作することによりオンにもオフにも設定することができる。また、利用者の音声レベルを示すレベルメータ906(図31参照)が表示されるので、利用者は自分の音声が大きすぎるのか、小さすぎるのかをリアルタイムで把握することができる。

【0197】

利用者は、このようにして、同じシーンの映像を繰り返して再生させながら、そのシーンの台詞を台詞914を見ながら行なう。このような繰り返しを行なうことにより、利用者には、どのタイミングで、どのような大きさを発声をすればよいかについての感覚をつかむことが容易になり、好ましい発声を行なうことが可能になる。この音声収録処理では、一旦スタートボタンを押すと常に最新の所定時間の利用者の発話が繰り返して録音され、リングバッファ962に格納される。

【0198】

利用者が、上記した繰り返しの中で、うまく台詞を発音できたと思ったときには、利用者は図31に示すストップボタン908を押す。これに回答して、録音が終了され(ステップ988)、リングバッファ962に録音されている音声のうち、処理中のシーンに対応する、直前の所定時間の音声はリングバッファ962から取り出され、新たな録音ファイル964としてハードディスクに追加格納される(ステップ990)。

【0199】

音声収録部930は、次のシーンを選択する(ステップ992)。全てのシーンに対して音声収録が完了したら(ステップ994でYES)、録音ファイル964の全てを利用者情報とともにサーバに転送して(ステップ998)、初期画面を再表示し(ステップ180)、次のユーザの処理開始を待つ。次のシーンが残っていれば(ステップ994でNO)、そのシーンに対して表示作成処理及び録音処理を開始し(ステップ976, 978)、利用者がストップボタンを押すまで、そのシーンを繰り返して表示しながら利用者の発声を繰り返して録音する。

【0200】

このようにして、ある利用者について、全てのシーンに対する台詞の発話を収録することができる。

【0201】

この実施の形態の音声収録部930では、収録のためのアシスタントが不要である。利用者のみで、タッチパネル900を用いた簡単な操作で全ての台詞を録音することができる。しかもこの場合、上に述べたように同じシーンを繰り返して表示しながら、同じ台詞を繰り返して発声させるようにすると、そのリズムによって、利用者が容易に適切な発話を行うことが可能となる。音声レベルも表示されるため、利用者は適切な音量で発音することができ、録音の失敗を少なくすることができる。シーンの進行に伴い台詞の色を代えて表示したり、画面に時刻情報を表示したりすることによっても発話のタイミングを知ることが

10

20

30

40

50

でき、利用者は、適切な台詞音声の発声を容易に行なうことができる。また、ボイスオーバーを行なうこともできるので、利用者は、標準的な発話音声を聞きながら、それにあわせて台詞の発話を行なうことができる。

【0202】

以上のように、この第2の実施の形態に係る音声収録部930では、利用者が適切なタイミングでアシスタントなしに適切な台詞音声の収録を行なうことができるように、様々な補助情報を利用している。それらのうち、あるものについてはオンしたりオフしたりすることができ、利用者にとってもやりやすい形で台詞の音声収録ができる。特に、複数の補助情報が表示されるので、利用者は自分が最も利用しやすいと思った補助情報を使用して台詞の発話を行なうことができ、失敗が少ないという特徴がある。また、同じシーンの台詞の収録を繰り返さない、その中で利用者がうまくいったと思う発話を選択して収録できるため、収録の失敗が極めて少ないという特徴がある。また、同じシーンを繰り返しながら発話するので、利用者はリズムをつかんで適切な速さ及びタイミングで発話することが容易になるという効果がある。

10

【0203】

なお、第2の実施の形態では、利用者が台詞を発話しやすいように複数の補助情報を利用している。この補助情報の組合せは、第2の実施の形態において説明したものに限定されず、第2の実施の形態の説明で利用したもの以外のものを利用してよいし、一部を利用しないようにしてもよいことはいうまでもない。

20

【0204】

なお、上の実施の形態の説明では、ユーザの音声を素片化し、素片DB82に追加している。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、ユーザの音声のうち、高品質に録音できた台詞音声は、声優音声DB80に登録するようにしてもよい。こうすることで、多数のユーザの音声を声優音声DB80に追加することが可能になり、さまざまな音声を効率よく収集することが可能になる。

【0205】

したがって、マルチメディア制作システム50によって、予め準備された映画の登場人物の顔画像だけでなく、その台詞まで、ユーザの音声に置換えたかのような映画を作成し、上映することができる。その結果、登場人物の台詞が分かっているマルチメディア制作物において、登場人物の音声をユーザの声で容易に短時間で置換することが可能になる。また、登場人物の音声をユーザの声に近い声質で容易に短時間で置換することが可能になる。さらに、登場人物の音声をユーザの声に近い声質で容易に短時間で置換することが可能なように、多数のユーザの声を収集し、それらを音声の置換において利用することが可能になる。

30

【0206】

上記した実施の形態は、映画を製作するためのマルチメディア制作システムに本発明を適用したものであった。本発明が適用可能なシステムはこれには限らず、たとえばテレビジョン番組、ラジオドラマ等、一般に台詞を発話するタイミング、その長さ、等が発話者ごとに決まったシナリオにしたがって進行するものであればどのようなものにも適用することができる。

40

【0207】

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の全ての変更を含む。

【図面の簡単な説明】

【0208】

【図1】本発明の第1の実施の形態に係るマルチメディア制作システム50の機能ブロック図である。

【図2】キャラクタ音声作成部102の機能ブロック図である。

50

【図 3】台詞情報記憶部 7 2 に記憶される台詞情報テーブルの構成を示す図である。

【図 4】録音終了時の台詞音声の録音状態の例を示す図である。

【図 5】図 2 に示す音声収録部 1 1 4 で行なわれる音声収録処理を、コンピュータハードウェア上で実現するコンピュータプログラムの前半部分のフローチャートである。

【図 6】図 2 に示す音声収録部 1 1 4 で行なわれる音声収録処理を、コンピュータハードウェア上で実現するコンピュータプログラムの後半部分のフローチャートである。

【図 7】ユーザ情報入力部 1 0 0 がユーザの音声収録時に入出力装置 1 1 2 の画面上に表示する表示の一例を示す図である。

【図 8】ユーザ音声 DB 1 2 0 の構成を示すブロック図である。

【図 9】音声作成部 1 1 8 の構成を示す機能ブロック図である

【図 10】図 2 に示す合成手法決定部 1 1 6 で行なわれる音声の生成手法の決定処理を、コンピュータハードウェア上で実現するコンピュータプログラムのフローチャートである。

【図 11】手法リストテーブル 7 8 の構成を示す図である。

【図 12】カット情報記憶部 7 6 に記憶された音響効果リストテーブルの構成を示す模式図である。

【図 13】図 9 に示す第 1 の音声生成部 3 0 0 を実現するプログラムのフローチャートである。

【図 14】図 9 に示す第 2 の音声生成部 3 0 2 を実現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。

【図 15】図 9 に示す第 3 の音声生成部 3 0 4 を実現するためのプログラムのフローチャートである。

【図 16】図 9 に示す第 4 の音声生成部 3 0 6 を実現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。

【図 17】図 9 に示す第 5 の音声生成部 3 0 8 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 18】図 9 に示す第 6 の音声生成部 3 1 0 を実現するためのプログラムのフローチャートである。

【図 19】図 9 に示す第 7 の音声生成部 3 1 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 20】図 9 に示す第 8 の音声生成部 3 1 4 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 21】図 9 に示す第 9 の音声生成部 3 1 6 を実現するためのプログラムのフローチャートである。

【図 22】図 9 に示す音声信号処理部 3 2 0 を実現するプログラムのフローチャートである。

【図 23】マルチメディア製作システム 5 0 によって作成された映画を再生する再生システムのブロック図である。

【図 24】台詞音声テーブル 8 8 の例示的構成を模式的に示す図である。

【図 25】図 2 3 に示す同時音声統合処理部 6 3 2 による同時音声統合処理後の台詞音声テーブル 8 8 の例示的構成を示す図である。

【図 26】同時音声統合処理部 6 3 2 を実現するためのプログラムのフローチャートである。

【図 27】図 2 3 に示す同期再生部 6 3 8 を実現するためのプログラムのフローチャートである。

【図 28】音声信号処理部 3 2 0 が実行する音響効果処理のうち、話速変換と音量正規化処理の内容を説明するための図である。

【図 29】マルチメディア製作システム 5 0 においてユーザの音声を収録するための台詞音声データ作成部 9 0 を実現するコンピュータ 8 4 0 のハードウェア構成の外観図である。

10

20

30

40

50

【図30】コンピュータ840の内部構成を示すブロック図である。

【図31】本発明の第2の実施の形態に係るマルチメディア製作システムにおける音声収録部で使用されるタッチパネル900の表示形態を模式的に示す図である。

【図32】第2の実施の形態に係るマルチメディア製作システムで使用される音声収録部930の機能ブロック図である。

【図33】第2の実施の形態に係るマルチメディア製作システムの音声収録部930において実行される、音声収録処理を実現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。

【図34】音声収録部930において実行される、表示作成処理を実現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。

10

【符号の説明】

【0209】

50 マルチメディア製作システム

60 三次元スキャナ群

62 画像処理PC

64 映像生成装置

66 映像データ

70 映像素材DB

72 台詞情報記憶部

74 標準音声記憶部

76 カット情報記憶部

78 手法リストテーブル

80 声優音声DB

82 素片DB

86 台詞音声データ

88 台詞音声テーブル

90 台詞音声データ作成部

92 映像・音声再生装置

100 ~ 100N ユーザ情報入力部

102 ~ 102N キャラクタ音声作成部

104 音声統合部

112 入出力装置

114, 930 音声収録部

116 合成手法決定部

118 音声作成部

120 ユーザ音声DB

122 音声DB更新部

124 素片DB更新部

280 分岐部

292 合流部

300 ~ 316 第1の音声生成部 ~ 第9の音声生成部

320 音声信号処理部

620 映像信号再生部

622 表示装置

624 効果音出力装置

632 同時音声統合処理部

638 同期再生部

640 台詞音声出力装置

900 タッチパネル

902 映像表示領域

20

30

40

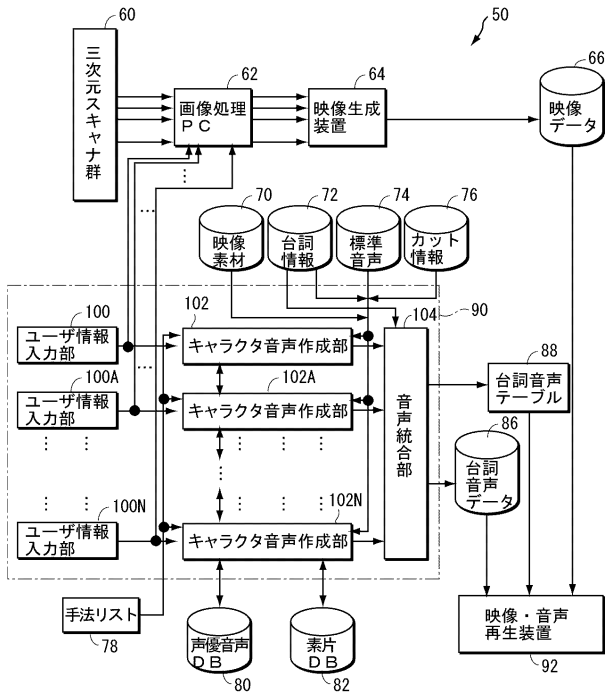
50

- 9 0 4 台詞表示領域
- 9 0 6 レベルメータ
- 9 0 8 ストップボタン
- 9 1 0 ボイスオーバボタン
- 9 1 2 時間情報画像
- 9 1 4 台詞
- 9 3 0 音声収録部
- 9 3 2 表示作成部
- 9 3 6 G U I 部
- 9 3 8 制御部
- 9 4 0 音声再生部
- 9 4 2 録音処理部
- 9 4 4 ヘッドセット
- 9 4 6 映像ファイル
- 9 4 8 台詞ファイル
- 9 5 2 レベル測定部
- 9 5 4 V O フラグ記憶部
- 9 5 6 キャラクタ情報記憶部
- 9 6 0 音声ファイル
- 9 6 2 リングバッファ
- 9 6 4 録音ファイル

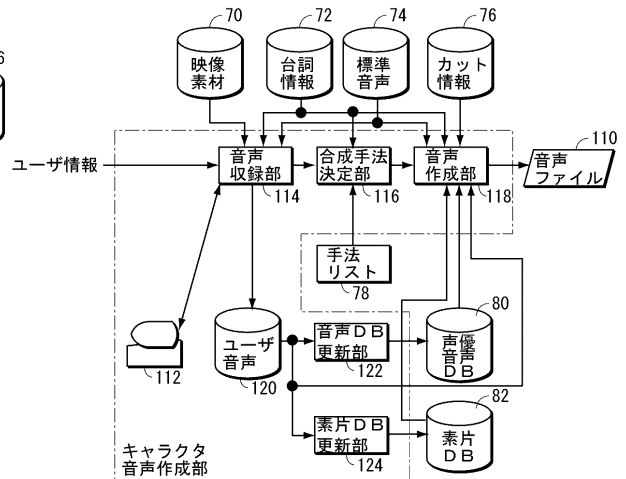
10

20

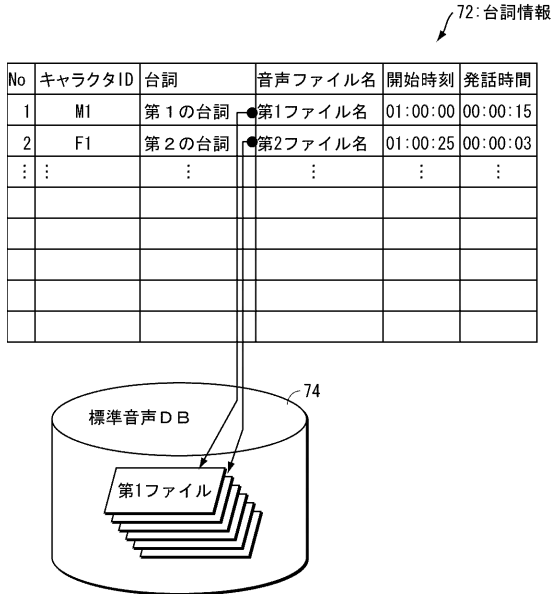
【図 1】



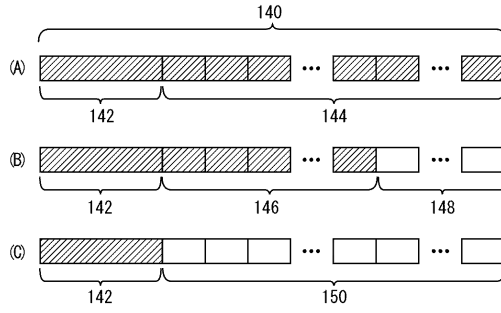
【図 2】



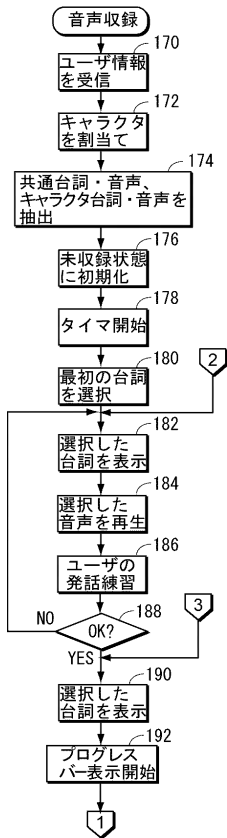
【図3】



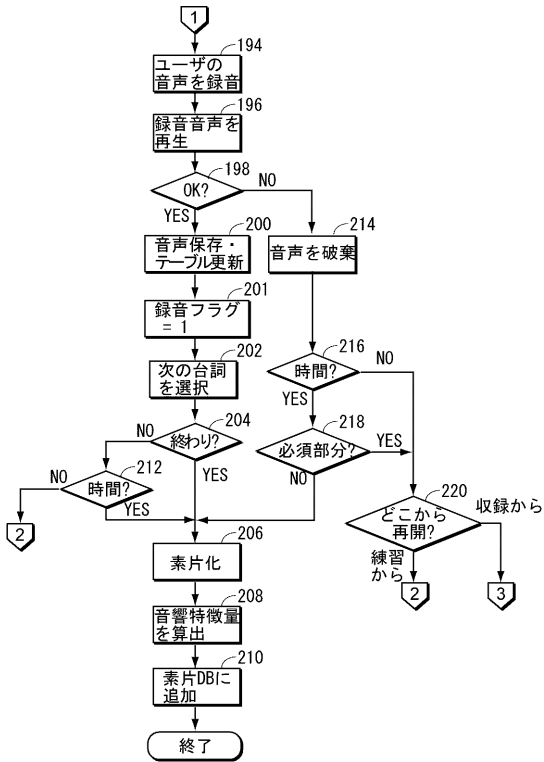
【図4】



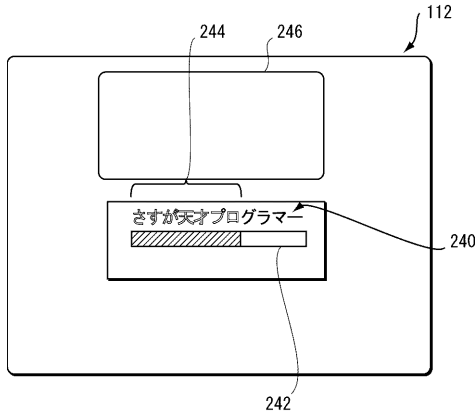
【図5】



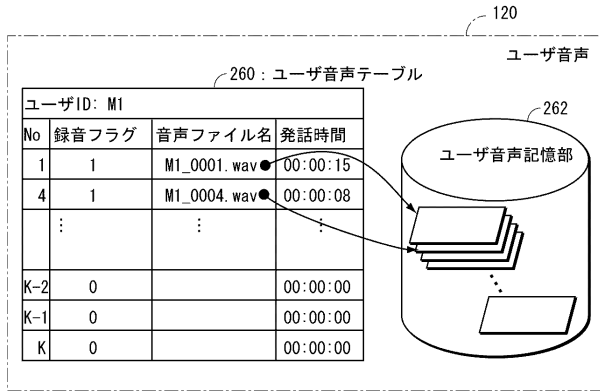
【図6】



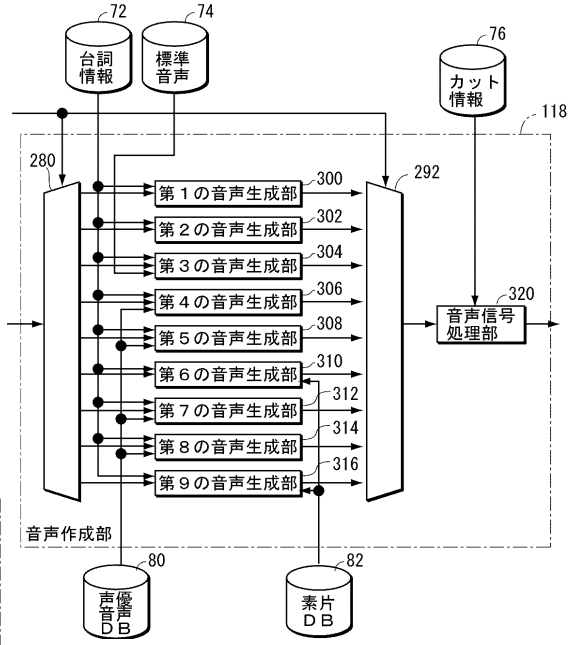
【図7】



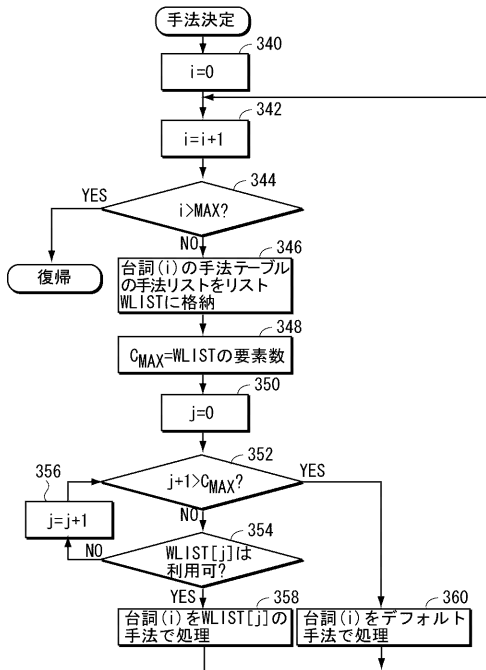
【図8】



【図9】



【図10】



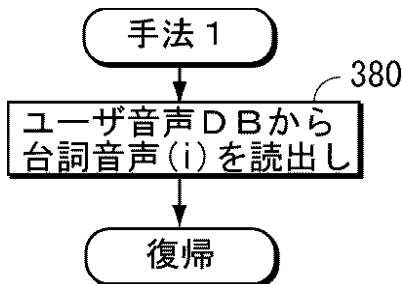
【図11】

台詞番号	手法リスト
1	手法 1, 手法 2, 手法 4, 手法 3, 手法 6
2	手法 1, 手法 2, 手法 4, 手法 3, 手法 6, 手法 5, 手法 7
⋮	⋮
⋮	⋮
MAX	手法 1, 手法 2, 手法 4, 手法 3, 手法 6, 手法 7, 手法 9

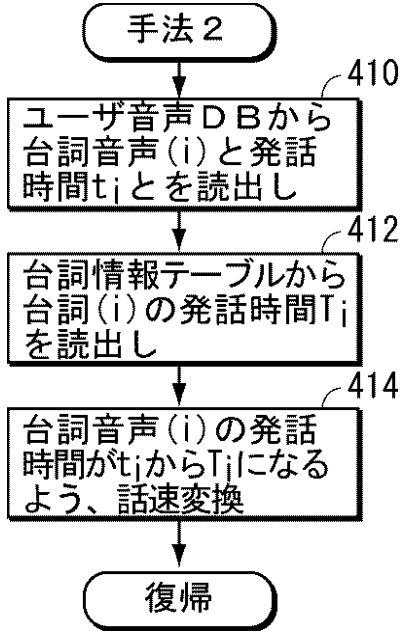
【図12】

台詞番号	音響効果リスト
1	音響効果 1, 音響効果 2
2	音響効果 1
⋮	⋮
⋮	⋮
MAX	音響効果 8

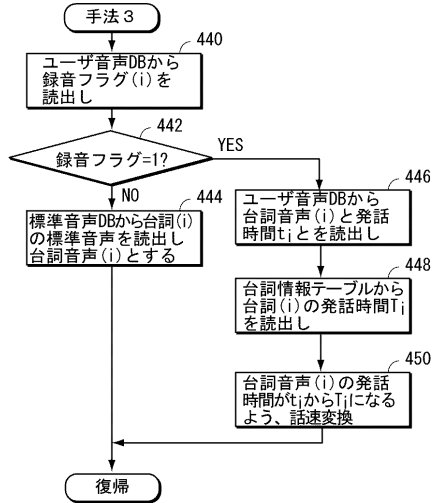
【図13】



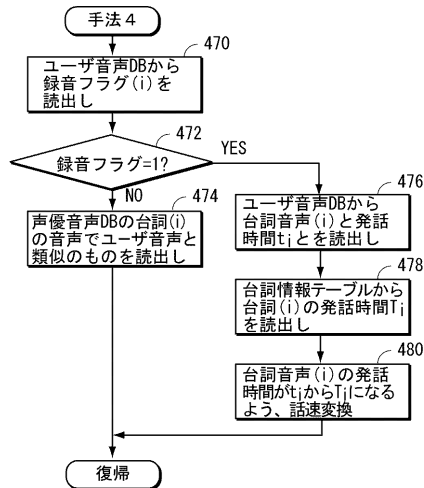
【図14】



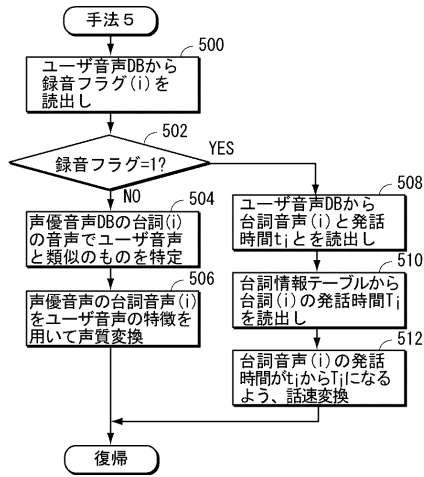
【図15】



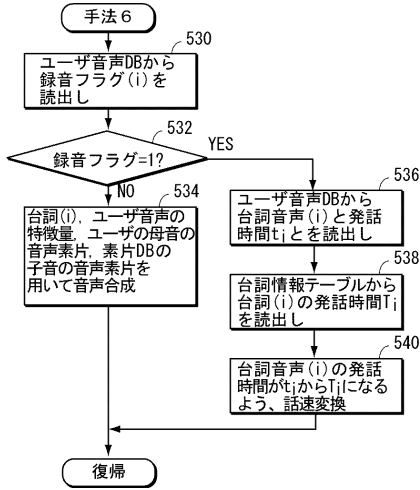
【図16】



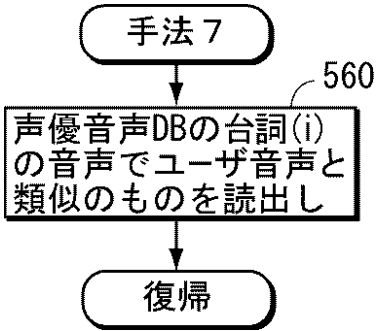
【図17】



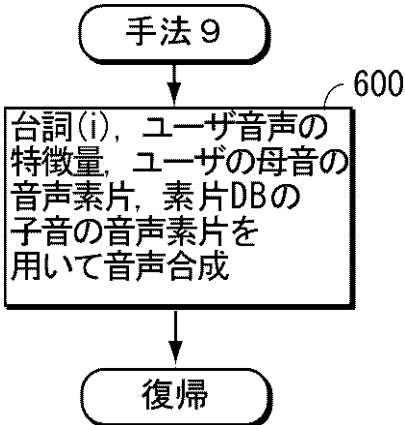
【図18】



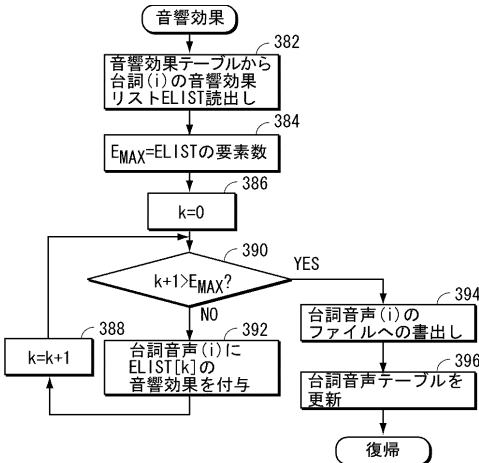
【図19】



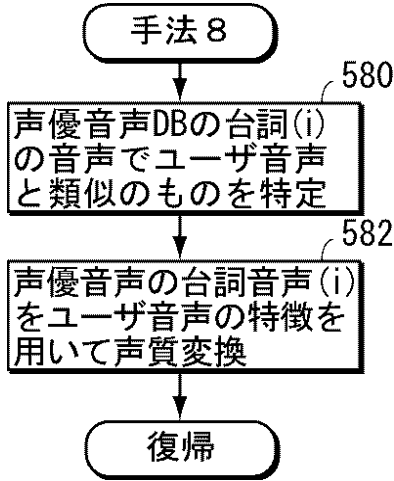
【図21】



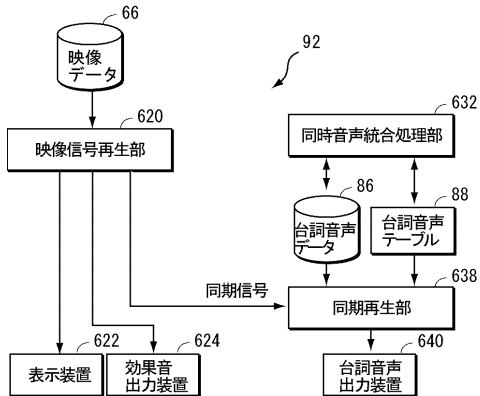
【図22】



【図20】



【図23】



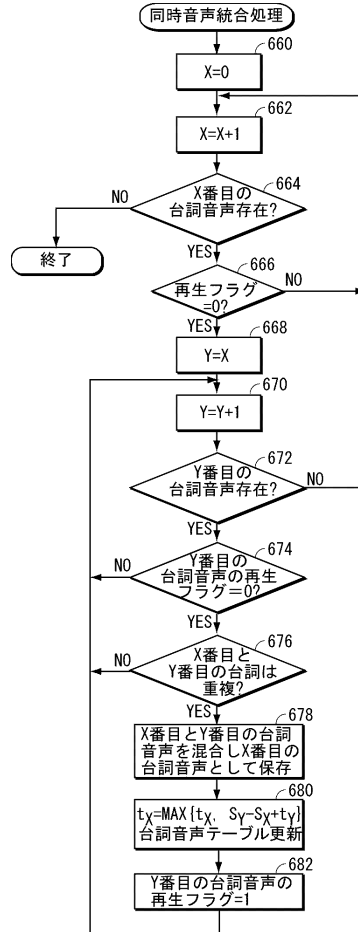
【図24】

番号	再生開始時刻	再生時間	再生ファイル名	再生フラグ
1	00:00:03	00:00:07	wave0001.wav	0
2	00:00:08	00:00:05	wave0002.wav	0
3	00:00:11	00:00:03	wave0003.wav	0
4	00:00:23	00:00:04	wave0004.wav	0
5	00:00:28	00:00:02	wave0005.wav	0
⋮	⋮	⋮	⋮	⋮

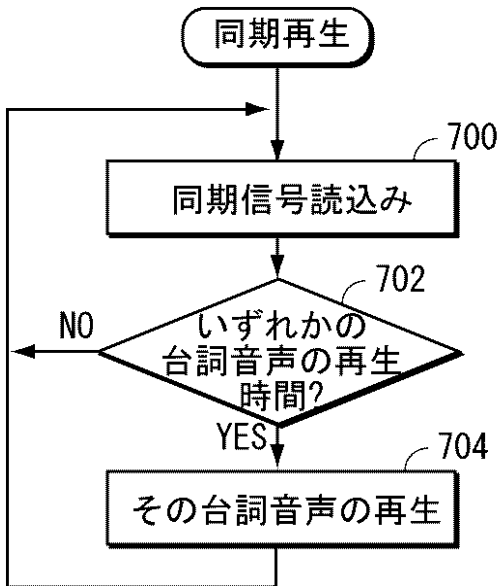
【図25】

番号	再生開始時刻	再生時間	再生ファイル名	再生フラグ
1	00:00:03	00:00:11	comb0001.wav	0
2	00:00:08	00:00:05	wave0002.wav	1
3	00:00:11	00:00:03	wave0003.wav	1
4	00:00:23	00:00:04	wave0004.wav	0
5	00:00:28	00:00:02	wave0005.wav	0
⋮	⋮	⋮	⋮	⋮

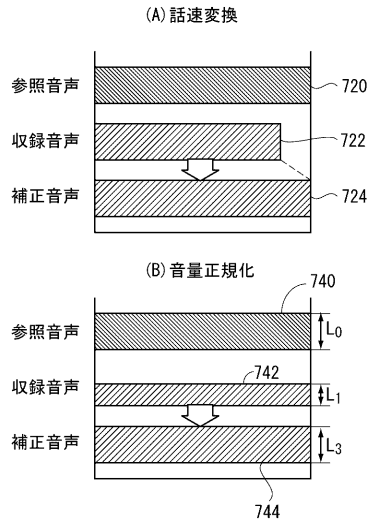
【図26】



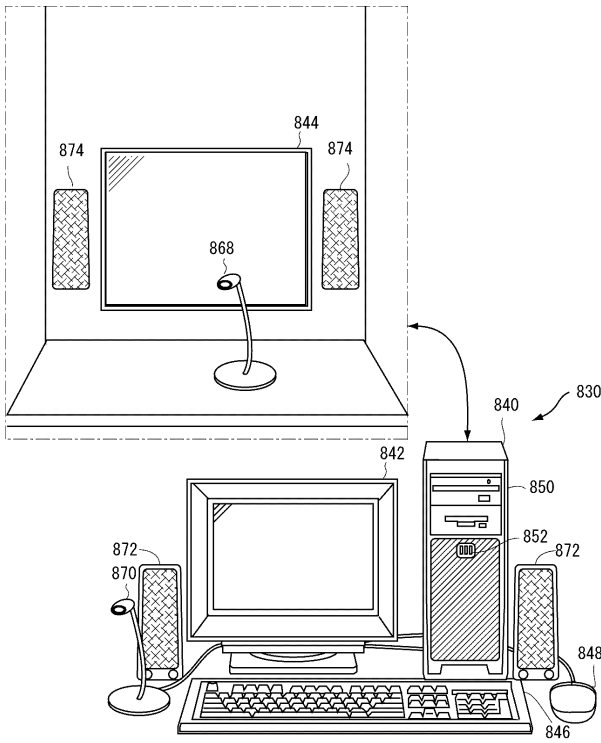
【図27】



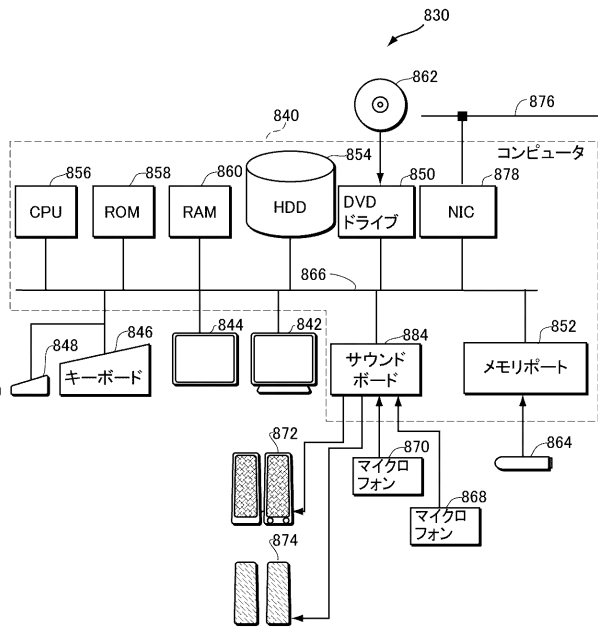
【図28】



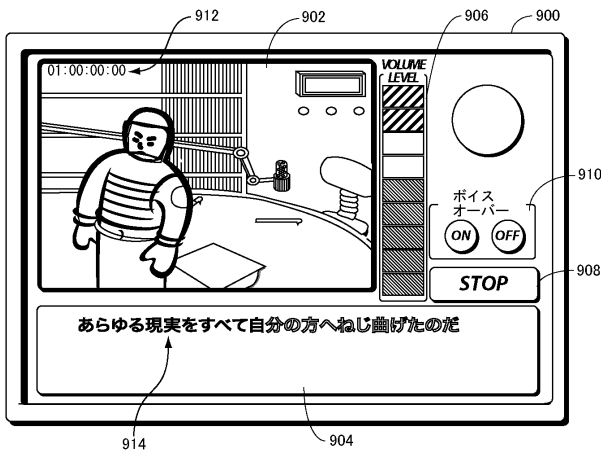
【図29】



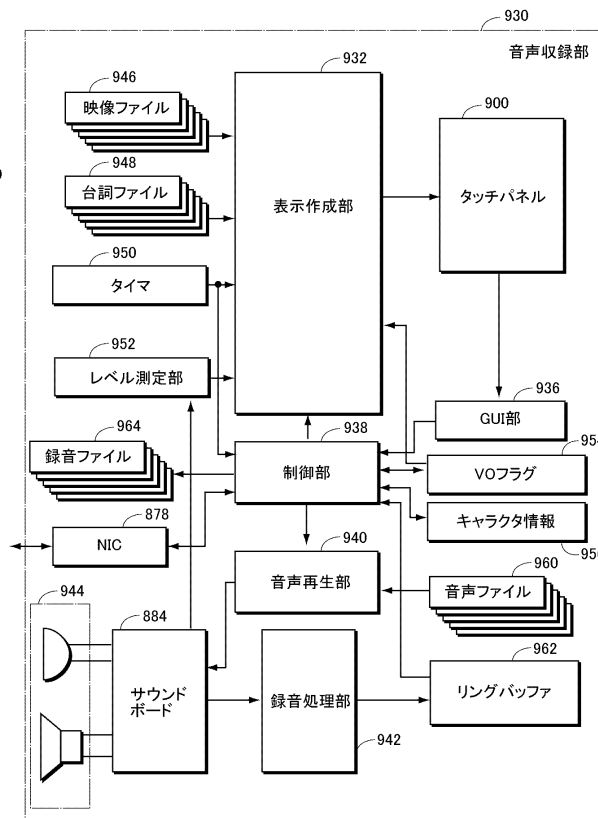
【図30】



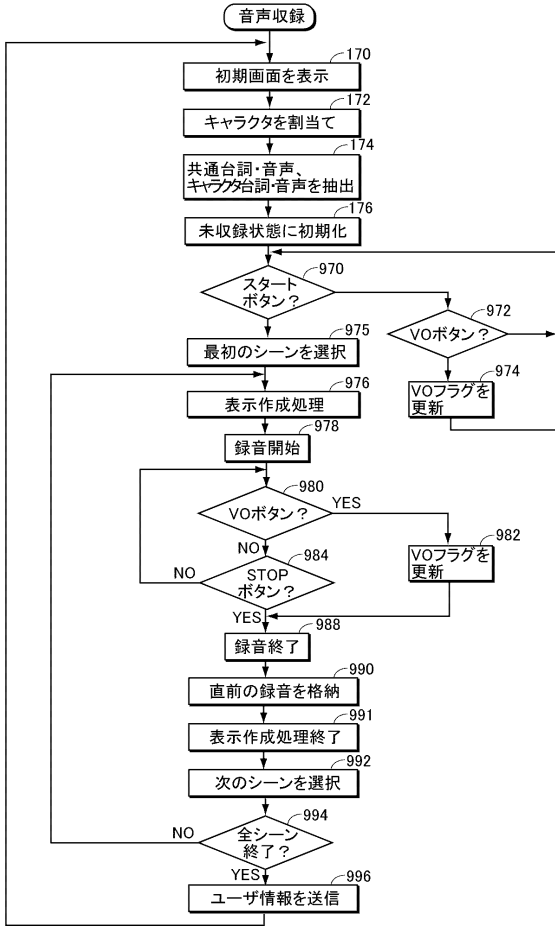
【図31】



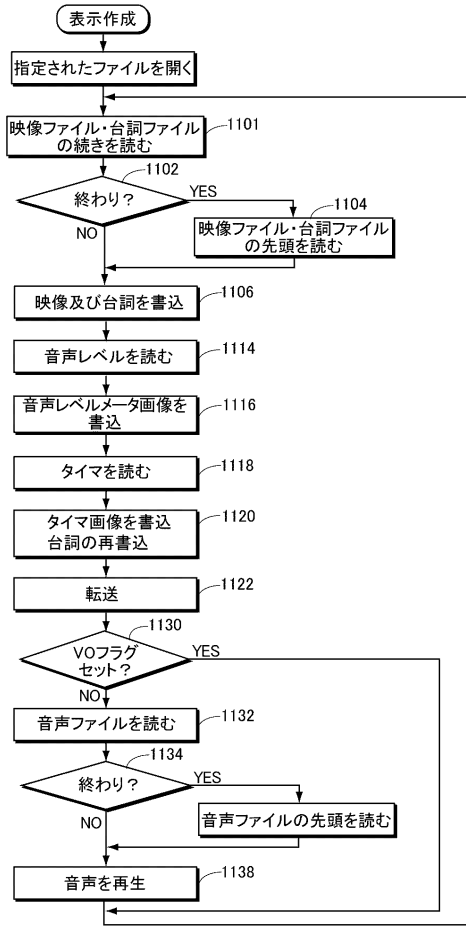
【図32】



【図33】



【図34】



フロントページの続き

(72)発明者 足立 吉広

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

審査官 山下 剛史

(56)参考文献 特開2004-229706(JP,A)
特開2002-239221(JP,A)
特開2008-22979(JP,A)
特開2004-71013(JP,A)
特開2007-192931(JP,A)
特開2004-110947(JP,A)
特開2006-346284(JP,A)
特開2002-358092(JP,A)
特開昭63-157226(JP,A)
特開2000-231396(JP,A)

(58)調査した分野(Int.Cl., DB名)

G10L 11/00 - 21/06

G11B 27/00 - 27/034