

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4780704号  
(P4780704)

(45) 発行日 平成23年9月28日(2011.9.28)

(24) 登録日 平成23年7月15日(2011.7.15)

(51) Int.Cl. F I  
H04N 7/14 (2006.01) H04N 7/14

請求項の数 3 (全 16 頁)

<p>(21) 出願番号 特願2005-345192 (P2005-345192)</p> <p>(22) 出願日 平成17年11月30日(2005.11.30)</p> <p>(65) 公開番号 特開2007-150955 (P2007-150955A)</p> <p>(43) 公開日 平成19年6月14日(2007.6.14)</p> <p>審査請求日 平成20年3月31日(2008.3.31)</p> <p>特許法第30条第1項適用 平成17年9月14日 電子情報通信学会人工知能と知識処理研究専門委員会主催の「社会情報学フェア2005」において文書をもって発表</p> <p>(出願人による申告)平成17年度独立行政法人情報通信研究機構、研究テーマ「軽度脳障害者のための情報セラピーインタフェースの研究開発」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</p>	<p>(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</p> <p>(74) 代理人 100090181 弁理士 山田 義人</p> <p>(72) 発明者 桑原 和宏 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</p> <p>(72) 発明者 桑原 教彰 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</p> <p>(72) 発明者 安部 伸治 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</p> <p style="text-align: right;">最終頁に続く</p>
---	--

(54) 【発明の名称】 対話型アノテーションシステム

(57) 【特許請求の範囲】

【請求項1】

ネットワークを介して接続される介護者端末と被介護者端末とを備え、前記ネットワークにはさらに写真データを予め記憶しておくデータベースが結合され、前記介護者端末および前記被介護者端末はそれぞれ、音声入力手段および音声出力手段を有し、介護者と被介護者とは前記音声入力手段および前記音声出力手段によって会話できる、対話型アノテーションシステムであって、

前記データベースから写真のデータを読み出して、前記2つの端末に同じ写真を同期的に表示させる写真表示手段、

前記介護者端末において前記写真中の場所を指定する場所指定手段、

前記場所指定手段が前記写真中の場所を指定したとき、前記被介護者端末に表示されている写真上でその場所に関連した領域に前記被介護者の注目を誘うための注目映像処理を施す注目映像処理手段、

前記介護者端末に設けられ、前記領域に前記注目映像処理が施されているとき、前記介護者端末に付属する前記音声入力手段から入力した介護者の音声を録音する録音手段、および

前記写真の写真データ、前記写真中の場所を含む領域を示す座標データを示す座標データおよび前記録音手段で録音した介護者の音声データを紐付けして前記データベースに保存する手段を備える、対話型アノテーションシステム。

【請求項2】

10

20

ネットワークを介して接続される介護者端末と被介護者端末とを備え、前記ネットワークにはさらに写真データを予め記憶しておくデータベースが結合され、前記介護者端末および前記被介護者端末はそれぞれ、音声入力手段および音声出力手段を有し、介護者と被介護者とは前記音声入力手段および前記音声出力手段によって会話できる、対話型アノテーションシステムであって、

前記データベースから写真のデータを読み出して、前記2つの端末に同じ写真を同期的に表示させる写真表示手段、

前記介護者端末において前記写真中の場所を指定する場所指定手段、

前記場所指定手段が前記写真中の場所を指定したとき、前記被介護者端末に表示されている写真上でその場所に関連した領域に前記被介護者の注目を誘うための注目映像処理を施す注目映像処理手段、

前記介護者端末に設けられ、前記領域に前記注目映像処理が施されているとき、前記介護者端末に付属する前記音声入力手段から入力した介護者の音声を録音する録音手段、および

前記写真の写真データ、前記写真中の場所を含む領域を示す座標データを示す座標データおよび前記録音手段で録音した介護者の音声データを紐付けして前記データベースに保存する手段を備え、

前記録音手段は前記音声入力手段からの音声を録音し続け、さらに、録音開始を指示する録音開始指示手段を設け、前記録音開始指示手段による指示が入力されたとき、直前の無音部分を検索し、その無音部分以後の音声を切り出して録音するようにした、対話型アノテーションシステム。

【請求項3】

前記介護者端末および前記被介護者端末はそれぞれ撮影手段、および相手側の前記撮影手段で撮影した映像を表示する相手映像表示手段をさらに含む、請求項1または2記載の対話型アノテーションシステム。

【発明の詳細な説明】

【技術分野】

【0001】

この発明は対話型アノテーションシステムに関し、特にたとえば、認知症者の過去の写真をナレーションとともにスライドショーの形式に編集して認知症者に提示する、思い出ビデオのようなビデオコンテンツを作成する際に写真にアノテーションを付与するための対話型アノテーションシステムに関する。

【背景技術】

【0002】

高齢の認知症者に視聴覚刺激を提示することで、彼（女）らを活性化することを目的とした研究が幾つかなされている。たとえば、非特許文献1に示すビデオレスパイトでは、ビデオの視聴者である認知症者に語り掛けを行うキャラクタが提示される。また、非特許文献2に示すプロジェクトCIRCAでは、昔の有名な歌や画像、映像を含んだ、よくデザインされたマルチメディアコンテンツが利用されている。

【0003】

発明者等は、認知症者の長期記憶を刺激するという観点から、視聴覚刺激の素材として、思い出ビデオを選定した。思い出ビデオは認知症者の古いアルバム中の写真を用いて作成されたスライドショービデオである。その臨床的な有効性は非特許文献3で実験的に示されている。

【0004】

しかし、思い出ビデオを作成するのは単純な作業ではない。まず、古いアルバムを用意し、そこから適当な写真を選択する。使用する写真は、視聴者である認知症者の遠い昔の記憶を呼び覚ますものでなくてはならない。次に、ビデオカメラで写真をムービー仕立てに撮影するが、必要に応じてパン、ズームの映像効果を付与する。またナレーションは、認知症者をビデオにより一層引き付けるために付与される。

10

20

30

40

50

## 【 0 0 0 5 】

発明者等は、非特許文献 4 で、思い出ビデオの作成作業を支援することを目指して、写真のどの部分に誰が写っているなどのアノテーションを写真に付与し、そのアノテーションを基にパン、ズームなどの映像効果を自動的に付与するオーサリングツールを提案している。アノテーションを活用することにより、同じ写真群から異なった内容の思い出ビデオを容易に作成できるという利点がある。

【非特許文献 1】Lund, D.A., Hill, R.D., Caserta, M.S., and Wright, S.D.: Video Re-  
espice: an innovative resource for family, professional caregivers, and persons  
with dementia, The Gerontologist, Vol. 35, Issue 5 (1995) 683-687.

【非特許文献 2】Gowans, G., Campbell, J., Alm, N., Dye, R., Astell, A., and Elli- 10  
s, M.: Designing a multimedia conversation aid for reminiscence therapy in demen-  
tia care environments, Extended abstracts of the 2004 conference on Human Factor  
s and Computing Systems (2004) 825 - 836.

【非特許文献 3】安田ほか：認知症者への思い出写真ビデオの作成と集中度の評価。第 2  
8 回高次脳機能障害学会総会（2004）

【非特許文献 4】Kawahara, N., Kuwabara, K., Tetsutani, N., and Yasuda, K., "Usi-  
ng Photo Annotations to Produce a Reminiscence Video for Dementia Patients," 3rd  
International semantic Web Conference (ISWC2004) Demo Papers, 2004

## 【発明の開示】

## 【発明が解決しようとする課題】 20

## 【 0 0 0 6 】

しかし、適切なアノテーションを写真に付与するのは容易ではない。特に思い出ビデオへの応用を想定すると、写真に写っている認知症者(被介護者)と関係の深い人々がアノテーションの主な対象となり、正確なアノテーションを付けることができる人は、被介護者または、被介護者をよく知っている人(たとえば家族)になる。

## 【 0 0 0 7 】

さらに、写真の写っている対象によっては、嫌な思い出になる場合も想定される。そのような場合は、思い出ビデオの対象外とした方が適当である。嫌な思い出かどうかは、被介護者本人にしかわからないことも多い。したがって、アノテーションをつける際には、介護者が、写真をもとに、被介護者と話をしながら、被介護者の反応を踏まえたうえで、写真にアノテーションをつけていくのが適当である。 30

## 【 0 0 0 8 】

それゆえに、この発明の主たる目的は、新規な、対話型アノテーションシステムを提供することである。

## 【 0 0 0 9 】

この発明の他の目的は、被介護者と話をしながら写真にアノテーションをつけられる、対話型アノテーションシステム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 0 】

請求項 1 の発明は、ネットワークを介して接続される介護者端末と被介護者端末とを備え、ネットワークにはさらに写真データを予め記憶しておくデータベースが結合され、介護者端末および被介護者端末はそれぞれ、音声入力手段および音声出力手段を有し、介護者と被介護者とは音声入力手段および音声出力手段によって会話できる、対話型アノテーションシステムであって、データベースから写真のデータを読み出して、2つの端末に同じ写真を同期的に表示させる写真表示手段、介護者端末において写真中の場所を指定する場所指定手段、場所指定手段が写真中の場所を指定したとき、被介護者端末に表示されている写真上でその場所に関連した領域に被介護者の注目を誘うための注目映像処理を施す注目映像処理手段、介護者端末に設けられ、領域に注目映像処理が施されているとき、介護者端末に付属する音声入力手段から入力した介護者の音声を録音する録音手段、および写真の写真データ、写真中の場所を含む領域を示す座標データを示す座標データおよび録 40 50

音手段で録音した介護者の音声データを紐付けしてデータベースに保存する手段を備える、対話型アノテーションシステムである。

【0011】

請求項1の発明では、対話型アノテーションシステム(10：実施例で相当する部分または要素を例示する参照符号。以下、同じ。)は、ネットワーク(12)およびサーバ(14)を介して繋がる介護者端末(介護者用コンピュータ16)および被介護者端末(被介護者用テレビ電話装置18)を含み、写真表示手段の機能はサーバ(14)が持つ。つまり、実施例では、サーバ(14)が、写真データを予め記憶しているデータベースから写真データを読み出して、2つの端末(16, 18)のGUI画面(40, 50)の写真表示領域(42, 52)に同じ写真を表示させる。そして、場所指定手段(519)が介護者端末で位置指定したとき、被介護者端末に映像処理コマンドが、たとえば介護者端末からサーバを介して、送られる。したがって、被介護者端末では指定位置を含む領域をたとえばズームアップしたりハイライトまたは点滅表示するなどの注目映像処理を実行する。したがって、被介護者が確実にその指定位置に注目することが期待できる。

10

しかも、音声入力手段(マイク22)および音声出力手段(スピーカ24)を用いて、介護者と被介護者とで会話できるので、この会話を通して、介護者は容易にアノテーションを収集できる。つまり、介護者端末、たとえばコンピュータ(16)が自分側のマイク(22)からの音声を録音することによって、写真に対するナレーションデータをアノテーションとして採取することができる。したがって、別にナレーションデータを作成する必要がない。

20

そして、実施例ではサーバがデータベース(28)に、写真の写真データ、写真中の場所を含む領域を示す座標データを示す座標データおよび録音手段で録音した介護者の音声データを紐付けして保存することによって、たとえばナレーション付の思い出ビデオのようなビデオコンテンツを作成できる。

【0016】

請求項2の発明は、ネットワークを介して接続される介護者端末と被介護者端末とを備え、ネットワークにはさらに写真データを予め記憶しておくデータベースが結合され、介護者端末および被介護者端末はそれぞれ、音声入力手段および音声出力手段を有し、介護者と被介護者とは音声入力手段および音声出力手段によって会話できる、対話型アノテーションシステムであって、データベースから写真のデータを読み出して、2つの端末に同じ写真を同期的に表示させる写真表示手段、介護者端末において写真中の場所を指定する場所指定手段、場所指定手段が写真中の場所を指定したとき、被介護者端末に表示されている写真上でその場所に関連した領域に被介護者の注目を誘うための注目映像処理を施す注目映像処理手段、介護者端末に設けられ、領域に注目映像処理が施されているとき、介護者端末に付属する音声入力手段から入力した介護者の音声を録音する録音手段、および写真の写真データ、写真中の場所を含む領域を示す座標データを示す座標データおよび録音手段で録音した介護者の音声データを紐付けしてデータベースに保存する手段を備え、録音手段は音声入力手段からの音声を録音し続け、さらに、録音開始を指示する録音開始指示手段を設け、録音開始指示手段による指示が入力されたとき、直前の無音部分を検索し、その無音部分以後の音声を切り出して録音するようにした、対話型アノテーションシステムである。

30

40

【0017】

請求項2の発明では、たとえば介護者端末のGUIに録音ボタンを表示し、それを操作したとき、直前の無音部分までさかのぼって、そこから後の会話(介護者音声)だけを自動的に録音できる。

【0018】

請求項3の発明は、介護者端末および被介護者端末はそれぞれ撮影手段、および相手側の撮影手段で撮影した映像を表示する相手映像表示手段をさらに含む、請求項1または2記載の対話型アノテーションシステムである。

【0019】

50

請求項3の発明では、2つの端末にそれぞれ撮影手段(20)を設け、互いに相手のリアル動画映像を見ることができる。したがって、介護者がアノテーションを採取する際に被介護者の状態や表情を観察できるので、それらが録音手段によるアノテーション採取のときの参考になり、あるいはその映像からアノテーションを採取したりすることもできる。

【発明の効果】

【0020】

この発明によれば、被介護者と話をしながら写真にアノテーションをつけられるので、正確なアノテーションを付与できる。

【0021】

この発明の上述の目的、その他の目的、特徴および利点は、図面を参照して行う以下の実施例の詳細な説明から一層明らかとなる。

【発明を実施するための最良の形態】

【0022】

実施例の説明に先立って、まず、アノテーションについて説明する。アノテーションとは、概略、写真にメタ情報を付与することであり、メタ情報とは、データに関する構造化した情報を意味し、実施例の場合であれば、写真画像データの詳細を構造的に記述する情報のことである。ただし、実施例では、これらのメタ情報以外にも、たとえば、その写真を使って思い出ビデオのようなビデオコンテンツを作る際に必要なナレーションの音声データなどもアノテーションとして、登録できるようにする。

【0023】

このようなアノテーションを付与するためのアノテーションツールとしては、たとえば、DublinCore (<http://dublincore.org>)、ImageRegions (<http://www.w3.org>)、FOAF (<http://www.foaf-project.org>)、Jena2(2.1) (<http://jena.sourceforge.net>)などが利用可能である。Dublin Coreはメタ情報を取り扱う代表的なツールとして知られていて、さらに、写真中の人物の顔やオブジェクトの領域(以後、「リージョン」と呼ぶ。)を複数個指定し、これをメタ情報として保持するためにImage Regionsを利用する。リージョンが人の場合には、FOAFを利用して人のメタ情報を付与する。リージョンは静止画である写真に対して、そこへのズーム、リージョン間のパンといったエフェクトを加えるために用いる。メタ情報をRDFの形式でデータベースに格納するために、Jena2(2.1)を使用する。

【0024】

たとえば、図1に示す写真PHにおいては2つのリージョンR1およびR2を設定する場合、それらのリージョンR1およびR2は、Image Regionsによって図2に示すメタ情報で表すことができる。リージョンR1についていえば、その原点(矩形枠の左上の角)の座標(x11、y11)および原点の対角の座標(x12、y12)、さらには高さh1および幅w1のメタ情報で表現できる。同様に、リージョンR2についても、原点座標x21、対角座標y21、高さh2、および幅w2で記述できる。

【0025】

図3を参照して、実際のアノテーションデータを説明する。この図3の例では、図1に示した写真を用いる。そして、その写真に2つのリージョンが設定されているものとする。図3において、楕円形の中に「dc:」とあり、それに関連する矩形の中のデータが、Dublin Coreで登録したメタ情報である。たとえば「dc:date」では日付「20040716(2004年7月16日)」が、「dc:title」では名称「at Disney Animal Kingdom(ディズニー動物王国にて)」が、「dc:description」では説明文「They are very happy.(彼等は非常に楽しそう)」がメタ情報として登録される。

【0026】

楕円形のなかに「imgReg:」とあるメタ情報は、Image Regionsでリージョンを指定したときのメタ情報である。「imgReg:hasregion」はリージョンが設定されていることを示すメタ情報であり、「imgReg:Rectangle」はリージョンが矩形であることを示し、「imgReg

10

20

30

40

50

:regionDepict」はリージョンの説明で、「imgReg:boundingBox」はリージョンの原点位置とサイズ（高さh、幅w）とを含む。「imgReg:coords」はリージョンの原点位置および対角位置の座標である。

【0027】

また、「foaf:gender」で与えられるメタ情報は、リージョンが人である場合の性別（例示では「female(女)」）であり、「foaf:name」で与えられるメタ情報は名前（例示では「Haruka(はるか)」）であり、「foaf:Person」で与えられるメタ情報はViewer（この思い出ビデオを観る人）と写真に写った人との関係を示し、例示では、「孫（grandchild）」であることがわかる。

【0028】

このようなアノテーションデータは、写真を使って思い出ビデオのようなビデオコンテンツを作る際に、使用する写真やナレーションを検索するために利用できる。たとえば、FOAFではリージョンの種類が人であるとき、その人を特定するメタ情報を付与するのであるから、このFOAFのメタ情報を利用して、「特定の人物の写っている写真」を検索することができる。複数の人物を同時に検索できるが、この場合には、検索した名前の全員が写っている写真が対象となる。

【0029】

Dublin Coreのメタ情報を利用する場合には、「撮影年月日」で検索できる。たとえば、「From（第1指定日）~To（第2指定日）」で第1指定日以降第2指定日以前に撮影した全ての写真が検索できる。「From（指定日）」でその指定日以降に撮影した全ての写真が検索できる。同様に、「To（指定日）」でその指定日以前に撮影した全ての写真が検索できる。また、「特定のプロパティに特定の値が含まれている写真」を検索できる。たとえば、「dc:title」に「Disney」の文字が含まれる写真など。

【0030】

このように、使用可能性のある写真にアノテーションを付与しておけば、ビデオコンテンツを作成する際の検索などに非常に便利である。そこで、図4に示すこの発明の一実施例の対話型アノテーションシステム10は、介護者が、たとえば認知症患者である被介護者と、たとえばIP（Internet Protocol）テレビ電話を通して会話しながら上述のアノテーションデータを取得しようとするシステムであり、図4に示すように、たとえばインターネットやLANのようなネットワーク12を介して接続される、サーバ14、介護者用コンピュータ16（以下、単に「コンピュータ16」ということがある。）および被介護者用IPテレビ電話装置18（以下、単に「テレビ電話18」ということがある。）を含む。サーバ14は、コンピュータ16およびテレビ電話18の間でコマンドの転送を行うリレーサーバの役目を果たすものである。

【0031】

コンピュータ16はキーボードやマウスさらにはモニタを含む一般的なコンピュータであるが、IPテレビ電話機能が組み込まれている。そのために、コンピュータ16には、付屬的に、カメラ20、マイク（音声入力手段）22およびスピーカ（音声出力手段）24が設けられている。カメラ20はこのコンピュータ16の前の介護者（図示せず）を撮影するもので、マイク22はその介護者の発した声や周囲の音を取り込む。テレビ電話18にも同様に、カメラ20、マイク22およびスピーカ24が付屬的に設けられている。カメラ20はテレビ電話18の前の被介護者（図示せず）を撮影するもので、マイク22はその被介護者の発した声や周囲の音を取り込む。そして、コンピュータ16とテレビ電話18とはネットワーク12およびサーバ14を介して、映像や音声をやり取りすることができる。両方のスピーカ24はそのようにして送られてくる相手の声を再生する。

【0032】

コンピュータ16には、図5に示すように、ハードディスクやRAMのような内部メモリ26が設けられていて、その内部メモリ26には、上述のアノテーションツール、たとえば、DublinCore、ImageRegions、FOAF、Jena2(2.1)などが設定されている。

【0033】

10

20

30

40

50

さらに、サーバ14には、インタフェース30を介してデータベース28が結合される。この実施例では、データベース28には、たとえばPostgreSQL7.4 (<http://www.postgresql.org>)というリレーショナルデータベースを用いる。

【0034】

コンピュータ16は図示しないが、グラフィックボードまたはプロセッサやサウンドボードまたはプロセッサを内蔵していて、それらを通して、サーバ14(図4)から送られてくる写真や相手(被介護者)の映像がモニタ32に表示されるとともに、スピーカ24からサーバ14から送られてくる被介護者の声などの音声が出力される。

【0035】

コンピュータ16にはさらに、写真データ入力装置34が接続される。この写真データ入力装置34は、イメージスキャナ、デジタルカメラ(デジカメ)、インターネット(Web)などの少なくとも1つを含む。イメージスキャナは、認知症者の過去の写真をスキャンしてカラーまたはモノクロの写真画像データを入力する。デジカメはリアルタイムで撮影した写真画像データを入力できる他、過去の写真を撮影してそれらの写真画像データを入力するという使い方もできる。インターネットは、遠隔地から送信される認知症者の過去の写真の写真画像データを入力し、さらには必要に応じて認知症者の過去にまつわる事象の写真画像データを取り込むために使用できる。さらに他の種類の写真データ入力装置が用いられてもよい。

【0036】

コンピュータ16の内部メモリ26には、さらに、たとえばインテル社が公開しているOpenCVとよばれるオープンソースのプログラムを設定している。したがって、写真データ入力装置34を用いて被介護者の過去の写真の写真画像データをコンピュータ16に入力するとき、写真中の顔の領域が特定され、図6に示すように、その写真の画像データとともに、領域候補リストの領域データ(たとえば、対角2点の座標データと高さhおよび幅w)がサーバ14にアップロードされ、データベース28に蓄えられる。同時に、このデータベース28には、図6に示すように、介護者が逐次取得するナレーションデータを含むアノテーションも、コンピュータ16からサーバ14へアップロードすることによって、その写真データや顔領域データとともに紐付けされて、記憶される。

【0037】

コンピュータ16では、相手のテレビ電話18にどれかの写真を表示させたいとき、サーバ14のURLを開いて、そこでその写真を指定することになる。そして、指定した写真のすべてのデータ(画像データ、座標データ、音声データ)は、その都度、サーバ14からコンピュータ16およびテレビ電話18に同期的にダウンロードされる。ただし、テレビ電話18には画像データだけが送られる。テレビ電話18は被介護者側であるので、アノテーションデータが不要であるからである。このようにして、コンピュータ16およびテレビ電話18のそれぞれのモニタに同じ写真が同時に(同期して)表示されるので、両方で同じ写真を見ながら、さらに会話をしながら、介護者が被介護者から必要な情報を聞き出しながら、介護の一環として、アノテーション付与作業を行える。

【0038】

コンピュータ16のモニタ32には図7に示すようなGUI(Graphical User Interface)40が表示される。このGUI40は、モニタ画面の左方の大部分を占める、写真表示領域42を含む。この写真表示領域42は、そこに入力した写真画像データの写真を表示し、さらには、リージョンを指定するなどの編集作業のために利用される。したがって、たとえばマウスなどを使ってその位置や場所を指定すると、コンピュータ16内にその位置や場所の情報が取り込まれるとともに、ネットワーク12を通してサーバ14に伝えられる。ただし、図7のこの領域42に表示されている写真の1人の顔に付着している黒丸印は、後に説明するように、介護者がマウスで位置を指定したときの、または、テレビ電話18の後述のタッチパネル51を通して被介護者が位置を指定したときの、その位置を誇張して表現したものである。

【0039】

10

20

30

40

50

GUI 40は、また、モニタ画面の右方に形成される相手映像表示領域44を含む。相手映像表示領域44には、テレビ電話18から送られるカメラ20で撮影した被介護者の動画映像がリアルタイムで表示される。

【0040】

GUI 40には、モニタ画面の下方に形成される、アノテーション入力領域46が設けられる。アノテーション入力領域46には、領域42に表示されている写真全般に関わることかつアノテーションツール(Dublin Core やFOAF)で定義されたメタ情報を入力する。そのメタ情報の入力ためには、図4に示すキーボードやマウスが利用できる。

【0041】

アノテーション入力領域46の上方には、いくつかの操作ボタンが設けられるが、この実施例では、録音ボタン48が重要である。この録音ボタン48は、介護者が被介護者と会話しているとき、自分の発言をビデオコンテンツのナレーションとして利用しようと思ったとき、その自分の発言をデータベース28にアノテーションとして登録する場合に操作される。たとえば、同じ写真を見ながら被介護者と会話していく訳であるが、たとえば「かわいい赤ちゃんですね。」のような被介護者への語りかけを録音する。

【0042】

テレビ電話18のモニタには図8に示すようなGUI 50が表示される。このGUI 50は、モニタ画面の左方の大部分を占める、写真表示領域52を含む。この写真表示領域52は、そこに入力した写真画像データの写真を表示するために利用される。GUI 50は、モニタ画面の右方に形成される相手映像表示領域54を含む。相手映像表示領域54には、コンピュータ16から送られるカメラ20で撮影した介護者の動画映像がリアルタイムで表示される。このテレビ電話18のGUI 50には、アノテーション入力領域は設けない。

【0043】

ただし、被介護者(図示せず)が写真表示領域52に表示された写真上の人や物を指し示すために、このモニタ(GUI)50の上には、タッチパネル51が少なくとも写真表示領域52を覆うように、設けられる。このタッチパネル51をタッチすることによって入力される位置や場所のデータは、ネットワーク12を介してサーバ14に伝えられ、必要に応じてさらにコンピュータ16に伝えられる。

【0044】

図4に示す実施例の対話型アノテーションシステム10を用いる場合、基本的には、介護者側が主導的役割を果たす。したがって、写真を表示する場合には、コンピュータ16を用いて介護者が写真を指定する。すなわち、図9の最初のステップS11において、介護者(図示せず)はキーボードやマウスなどの適宜の入力手段(図4)を用いて、サーバ14のURLにアクセスし、それを開く。ただし、この操作に先立って、介護者がアノテーション付与を意図しているすべての写真は、先に説明したように写真入力装置34を用いてコンピュータ16内に取り込まれ、サーバ14にアップロードされ、そのサーバ14に接続されたデータベース28に保存されているものとする。

【0045】

コンピュータ16の操作によってサーバ14の該当のURLを開き、介護者は、次のステップS13において、そのURLから、アノテーションしたい写真を、たとえばマウスで指定する。

【0046】

そうすると、サーバ14は、その指定された写真の写真データや顔領域候補の座標データ(図6)をコンピュータ16およびテレビ電話18の両方にダウンロードするので、ステップS15および図10のステップS101において、コンピュータ16もテレビ電話18もそのダウンロードデータを受ける。ただし、テレビ電話18には顔領域候補の座標データは要らないので、サーバ14はコンピュータ16が指定した写真の画像データだけをテレビ電話18に送る。したがって、ステップS17および図10のステップS103において、コンピュータ16のモニタ32のGUI 40およびテレビ電話18のGUI 5

10

20

30

40

50



0のそれぞれの写真表示領域42および52には、図7および図8に示すように、介護者がコンピュータ16で指定した写真が表示される。

【0047】

その状態で、介護者は、マウスを用いて、その写真中の顔の中の点を指定する。そうすると、図7に示す黒丸がコンピュータ16のモニタ32上の写真表示領域42に表示される。そして、コンピュータ16は、次のステップS19において、その位置データ（座標データ）を、ズームアップコマンドとともにサーバ14に送る。

【0048】

サーバ14はその位置データを受け取り、コンピュータ16からアップロードされている顔領域候補座標データを参照して、その位置が写真中の顔の領域に含まれるかどうかを判断する。介護者が指定した位置が顔領域を示すとき、サーバ14は、コンピュータ16が送ったコマンド（たとえばズームアップコマンド）と一緒に、その顔領域を表すデータ（たとえば、対角2点の座標データと高さhおよび幅w）をテレビ電話18に送る。

【0049】

応じて、テレビ電話18はステップS109で“YES”となり、ステップS111で、そのコマンドとデータとに基づいて、指定された画像処理を実行する。たとえば、指定領域が顔領域であり、コマンドがズームアップの場合、図11(A)に示すように、テレビ電話18のモニタ上のGUI50の写真表示領域52に、まず、その顔領域（図1で示すリージョンR1）を示す4角形を表示し、ついで、図11(B)に示すように、その顔領域を写真表示領域52の全面に拡大して表示する（ズームアップ処理）。このような映像処理によって、テレビ電話18のモニタを見ている被介護者がその顔領域に注目することが期待できる。

【0050】

ただし、そのような注目映像処理としては、ズームアップの他に、たとえば図11(A)に示す4角形および/またはその枠内をハイライト表示するとか、点滅表示するとか、色々考えられる。

【0051】

図9に示すステップS21で、介護者と被介護者とが、同じ写真の同じ顔を見ながら、マイク22およびスピーカ24を使って、会話することができる。コンピュータ16は、自分のマイク22から入力される音声信号を音声データに変換し、内部メモリ26（図5）に一時的に記憶する。コンピュータ16は、自分のマイク22からのすべての音声を録音している。

【0052】

つまり、被介護者と会話しながら録音するために、この実施例のシステム10では、コンピュータ16は常に録音状態としておき、介護者のマイク22を通して被介護者に対して行う呼びかけを常時録音しておく。介護者が発した呼びかけに対して、被介護者がどのように反応しているかは、GUI40（図7）の相手映像表示領域54に表示されている被介護者のリアルタイム動画映像を見れば分かる。そして、被介護者が好ましい反応を出した際に、介護者は、図7のGUI40に設定されているナレーション録音ボタン48を押す（クリックする）。ただし、このような仮想ボタンに代えて、フットスイッチなどの実際のスイッチを用いるようにしてもいい。

【0053】

ナレーション録音ボタン48が押されると、コンピュータ16は、直前に録音された音声から、無音部分の区切りを検出し、ナレーションとして使用する部分の音声をデータベース内にアノテーションデータとして格納する。たとえば、今注目している顔が赤ちゃんのものであったとして、介護者が被介護者に、たとえば「かわいい赤ちゃんですね」のような共感を呼び起こすことをねらった呼びかけを行なうことがよく行われる。そして、この呼びかけに被介護者が好ましい反応を示したとして、図12に示すように、その呼びかけの終了以後、録音ボタン48を操作すると、この「かわいい赤ちゃんですね」の呼びかけが始まる前の無音部分を検索し、その無音部分以後の「かわいい赤ちゃんですね」の音

10

20

30

40

50

声データだけを切出して、サーバ14経由でデータベース28に登録すればよい。

【0054】

このとき、このナレーションデータは図6に示すように、写真と紐付けされているので、後にすぐに検索することができる。つまり、格納された音声データと、録音時に注目していた写真中の領域へリンクをはっておけば、その写真中の領域が、音声データのアノテーション情報となるわけである。通常の録音システムのように、録音ボタンを押してから、録音するのではなく、発音が終わってから、いいと思ったもののみをデータベースに登録することで、効率よくナレーションデータの収集が行える。

【0055】

このように、介護者の発した語り掛けの音声を録音し、データベース28に格納することにより、写真のアノテーションとしてナレーションデータも同時に獲得することができる。特に、この実施例のように介護者と被介護者とが離れた場所でテレビ電話を介した場合には、コンピュータ16のマイク22は介護者側の音声のみを拾うことになるので、被介護者の発声は録音されず、雑音の少ないナレーションデータが録音できるという利点もある。しかも、介護者側の音声そのままナレーションとして録音されるので、たとえば別の人に別途を吹き込んでもらったり、音声合成でナレーションデータを作成したりする必要がない。

10

【0056】

また、この会話を通して被介護者から得た情報によって、図3に示す多くのアノテーションを取得できる。たとえば「dc:date」で定義される日付、「dc:title」で定義される写真またはその写真を撮ったときの行事などの名称、「dc:description」で定義される説明文、「foaf:gender」で定義される人の性別、「foaf:name」で定義される人の名前、「foaf:Person」で定義できる被介護者との関係など、必要なアノテーションを正確に会話から採取することができる。ただし、「imgReg:」で定義できる一連のリージョン情報は、先に説明した「顔領域候補データ」から取得できる。

20

【0057】

このように、テレビ電話を利用して介護者と被介護者とが会話しながらアノテーションデータを獲得するようになれば、介護の一環として、つまり介護しながらアノテーションも取得できるので、アノテーションをつける作業自体は、介護の中に組み込まれ、オーバーヘッドとはなくなる。しかも、当事者に確認しながらアノテーションデータを採取できるので、写真に対して正確なアノテーションを付与することができる。

30

【0058】

図9に戻って、ステップS25でコンピュータ16はその写真のアノテーション採取は終了したかどうか判断する。この判断は、そのとき表示されている写真中の顔候補すべてについてアノテーションを付与したかどうか、つまり、すべての顔領域をコンピュータ16側で指定したかどうかで判断できる。ただし、コンピュータ16を操作している介護者が判断して次の写真を表示するための操作ボタン(図示せず)を操作するようにしてもよい。

【0059】

ステップS25で“NO”なら、つまり指定していない顔候補が未だ残っているときは、コンピュータ16はステップS27において、ズームダウンコマンドをサーバ14に送る。ズームアップは上述のように、指定した顔を画面いっぱい拡大表示するような映像処理であったが、ズームダウンはその逆で、画面52(図8)に写真全体が表示できるように縮小する映像処理のことである。このズームダウンコマンドはサーバ14からテレビ電話18に送られるので、テレビ電話18では、図10のステップS111で、ズームダウン処理する。それによって、図8に示すGUI50の領域52に写真が全体表示されるので、次のステップS19での位置指定が明瞭に認識できる。つまり、ズームアップしたままであったら、次の位置指定の過程が被介護者には見えないので、位置指定を変更したと、はっきり認識できない場合があるが、ズームダウンして一旦「引いた」後再びズームアップ(ステップS21)するので、次の位置指定がはっきりと認識できる。

40

50

## 【 0 0 6 0 】

その後、次の指定位置についてステップ S 1 9 S 2 5 を繰り返し実行し、やがてステップ S 2 5 で “ Y E S ” が判断される。ステップ S 2 5 で “ Y E S ” の判断をした次に、ステップ S 2 9 において、コンピュータ 1 6 は、次の写真があるかどうか判断する。この判断は、コンピュータ 1 6 が何枚の写真のデータをサーバ 1 4 にアップロードしてデータベース 2 8 に保存したか把握できているので、処理した写真がその枚数に達したかどうか判断することで容易に行なえる。

## 【 0 0 6 1 】

ステップ S 2 9 で “ Y E S ” が判断されたときには、ステップ S 1 3 まで戻って、再びステップ S 1 3 S 2 7 を実行すればよい。ステップ S 2 9 で “ N O ” なら、つまり、すべての写真についてアノテーションデータを採取または収集できたなら、処理を終了する。ただし、それまでに取得したアノテーションデータは、サーバ 1 4 の処理によって、その都度データベース 2 8 に更新的に蓄積されているので、処理が終了したときには、データベース 2 8 内に写真およびそれに紐付けされたアノテーションデータが保存されていることになる。

## 【 0 0 6 2 】

なお、テレビ電話 1 8 の G U I 5 0 の上にタッチパネル 5 1 を装着しているが、このタッチパネルは被介護者が操作するもので、たとえば、被介護者が会話中に、「これは妹と（場所）へいったときの写真です。」という返答が帰ってきた場合を想定して、その場合、介護者は「妹さんはどの人」などと会話を進行させることがある。その問いかけに対して、被介護者は写真の上で「妹」を指し示そうとする。そのとき、写真の上にタッチパネル 5 1（図 8）があるので、被介護者が「妹」にタッチすれば即座に、その位置データが取得できる。つまり、マウスなどの被介護者にとって煩わしい操作なしに、被介護者が簡単に位置（写真中の人物や物）を指示することができる。

## 【 0 0 6 3 】

そして、この被介護者のタッチパネル 5 1 からの位置データは、サーバ 1 4 を通じてコンピュータ 1 6 に入力されるので、コンピュータ 1 6 は、図 9 のステップ S 1 3 で、自分で位置指定したときと同様に、たとえば図 7 の写真表示領域 4 2 に黒丸表示を行なう。それによって、介護者は被介護者が指し示した「妹」が写真中のどの人かを容易に理解または認識できる。

## 【 0 0 6 4 】

なお、図 4 の実施例において、被介護者端末としてのテレビ電話 1 8 において、所定の時間（たとえば 1 秒間）何も動作がないときには、その被介護者端末 1 8 から、サーバ 1 4 に実行すべきコマンドがあるかを問い合わせ、返却値として帰ってくるコマンドをステップ S 1 1 1 で実行する（ポーリング）。同じように、介護者端末としてのコンピュータ 1 6 において所定の時間（たとえば 1 秒間）何も動作がないときには、介護者端末 1 6 は、被介護者端末 1 8 でスクリーン（タッチパネル 5 1）にタッチしたという情報がないかをサーバ 1 4 に問い合わせ、その返却値として帰る情報をもとに次の動作を決めればよい。

## 【 0 0 6 5 】

サーバ 1 4 は、上で説明したように、2 つの端末に同期的に同じ写真を表示させる手段として機能するが、さらに、サーバ 1 4 はこのように、被介護者端末 1 8 と介護者端末 1 6 とのコネクションを管理し、コマンドを双方の端末間で転送するとともに、送られてきたコマンドや通知を一旦蓄え、次に端末からアクセスがあったときにその返却値として、蓄えておいた情報を該当の端末に返すという機能も持っている。

## 【 0 0 6 6 】

ただし、2 つの端末間で直接データの授受ができるようなネットワーク環境であれば、サーバ 1 4 を省略することも可能である。その場合には、コンピュータ 1 6 にデータベース 2 8 をインタフェース 3 0 を介して接続すればよい。

## 【 0 0 6 7 】

以上説明したように、実施例の対話型アノテーションシステム 1 0 によれば、介護者が

10

20

30

40

50

被介護者の相手をしながら、並行して写真のアノテーションをつける作業を容易にする。アノテーションがついた写真からは思い出ビデオのようなビデオコンテンツが容易に生成できるので、介護者が不在の場合は、思い出ビデオを再生することで、介護の一助とすることができる。さらに、介護者がネットワークを介して、被介護者の相手をする事が可能になる。介護ボランティアが被介護者の自宅に訪問することなく、相手をする事が可能となり、被介護者にとって介護を受ける機会が増えることにつながる。

【図面の簡単な説明】

【0068】

【図1】この発明の実施例の対話型アノテーションシステムで用いるアノテーションの例を説明するための図解図である。

10

【図2】図1に示すリージョンを示すアノテーションデータを例示する図解図である。

【図3】実施例のシステムでデータベースに登録するアノテーションの一例を示す図解図である。

【図4】この発明の一実施例の対話型アノテーションシステムを示すブロック図である。

【図5】図4実施例の介護者用コンピュータの構成を示すブロック図である。

【図6】写真とそれに関連する顔領域候補データおよびアノテーションの紐付けの状態を図解する図解図である。

【図7】介護者用コンピュータのモニタに形成するGUIを例示する図解図である。

【図8】被介護者用テレビ電話装置のモニタに形成するGUIを例示する図解図である。

【図9】図4実施例における介護者用コンピュータの動作例を示すフロー図である。

20

【図10】図4実施例における被介護者用テレビ電話装置の動作例を示すフロー図である。

【図11】被介護者用テレビ電話装置のモニタに形成するGUIにおける映像処理の一例（ズームアップ）を図解する図解図である。

【図12】ナレーション録音の方法の一例を示す図解図である。

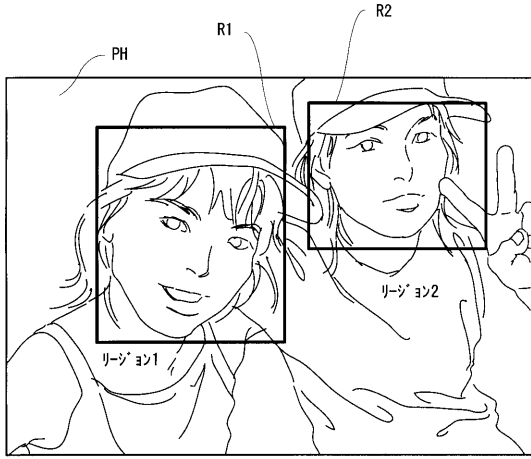
【符号の説明】

【0069】

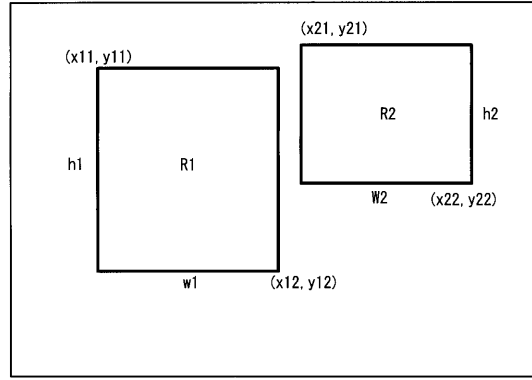
- 10 ...対話型アノテーションシステム
- 12 ...ネットワーク
- 14 ...サーバ
- 16 ...介護者用コンピュータ(コンピュータ)
- 18 ...被介護者用テレビ電話装置(テレビ電話)
- 20 ...カメラ
- 22 ...マイク
- 24 ...スピーカ
- 26 ...内部メモリ
- 28 ...データベース
- 40, 50 ...GUI

30

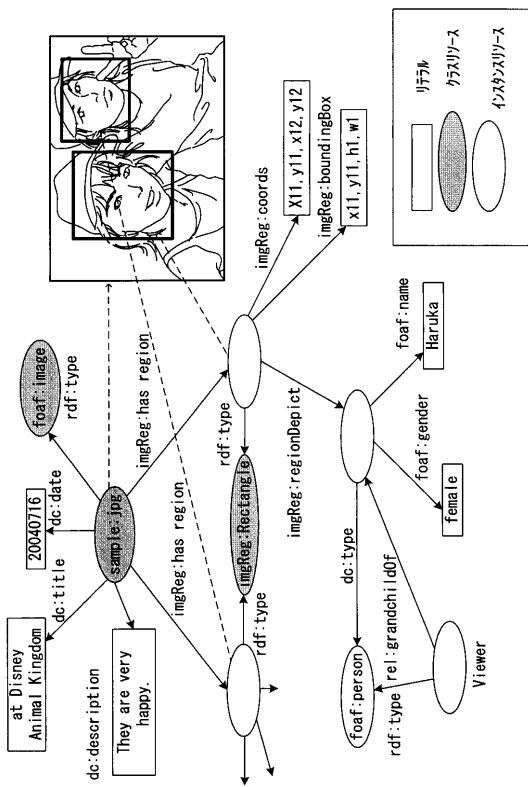
【図1】



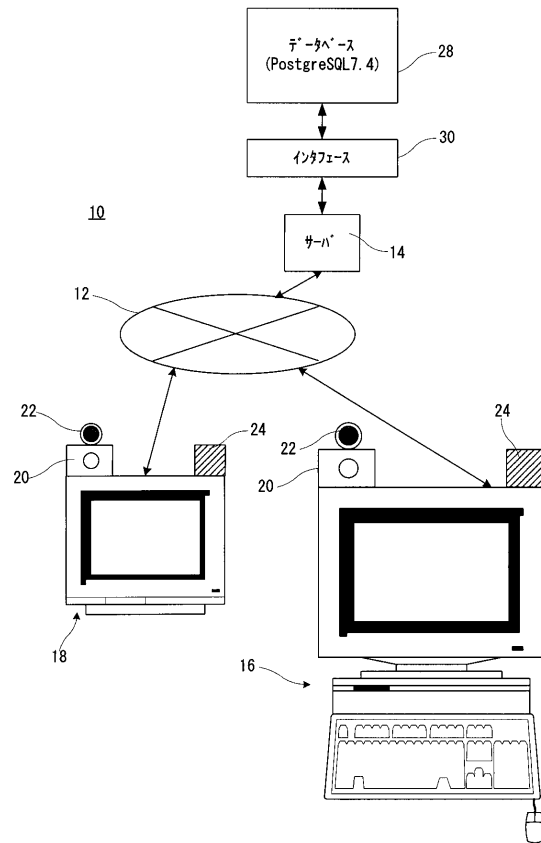
【図2】



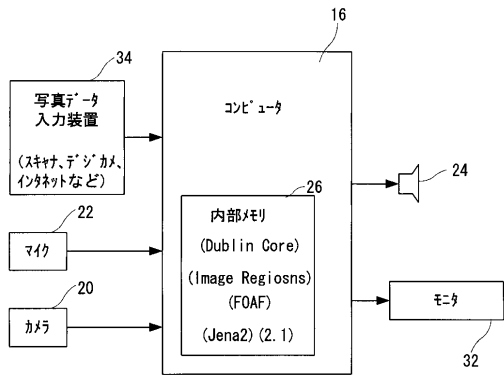
【図3】



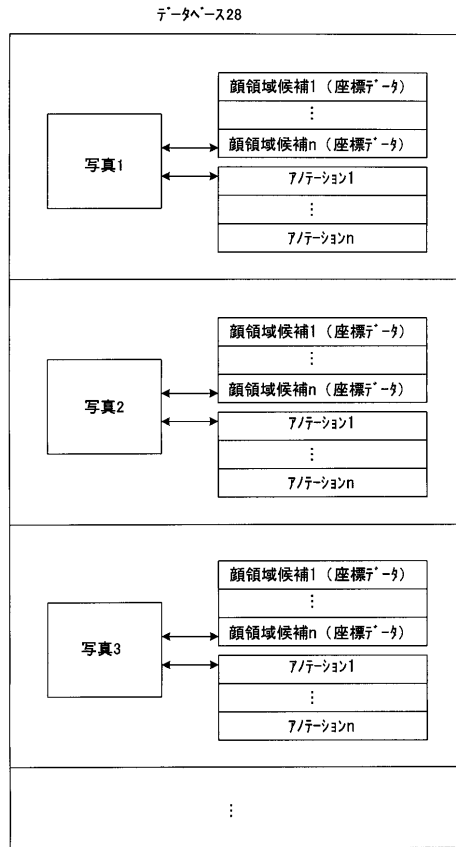
【図4】



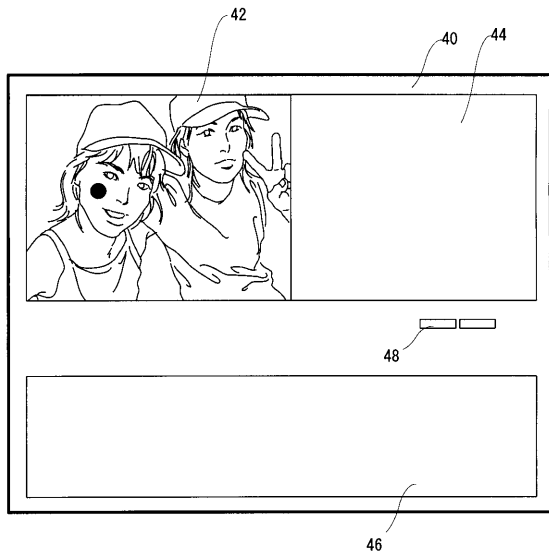
【図5】



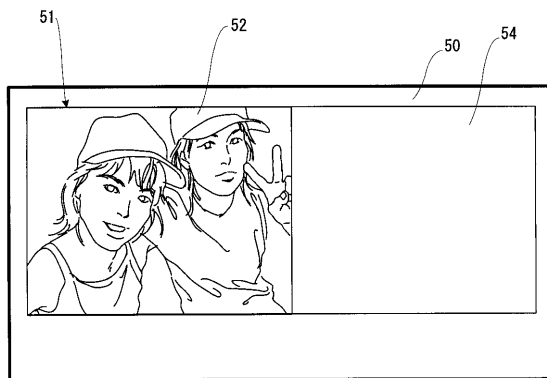
【図6】



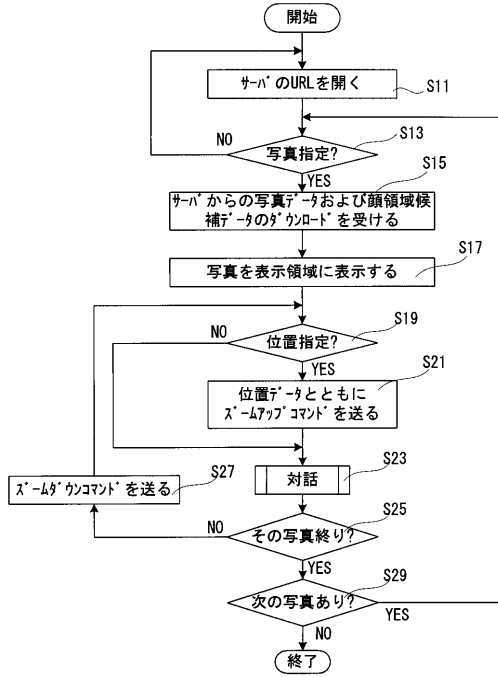
【図7】



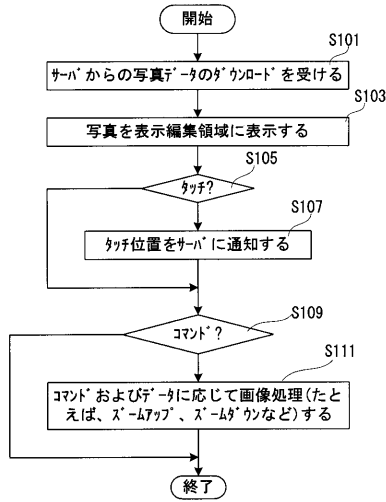
【図8】



【図9】



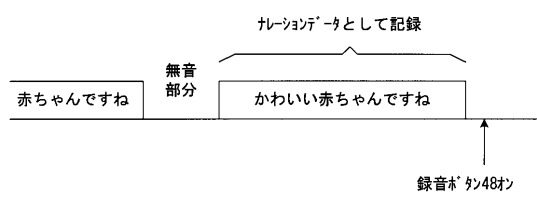
【図10】



【図11】



【図12】



---

フロントページの続き

(72)発明者 安田 清

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 三森 雄介

(56)参考文献 桑原 教彰, “セマンティクスを考慮したメディア同期の提案”, 電子情報通信学会技術研究報告 Vol. 105 No. 105, 日本, 社団法人電子情報通信学会, 2005年 5月24日, pp.15-20

(58)調査した分野(Int.Cl., DB名)

H04N 7/14 - 7/173