

明 細 書

発明の名称： 音声に基づいて画像を変更するためのシステム

技術分野

- [0001] この発明は、音声に基づいて画像を変更するためのシステムに関する。より詳しく説明すると、この発明は、会話に基づいてリアルタイムに共有画像が変更されるシステムに関する。

背景技術

- [0002] 一般的なウェブ会議システムは、話者が用意したプレゼンテーション資料を各端末に表示させる。そして、参加者が、その資料に関する発言をしても資料に変化は生じない。
- [0003] 特開2020-089641号公報には、音声に基づいてコマンド操作を行うシステムが記載されている。このように音声を認識し、各種操作を行わせるシステムは公知である。しかしながら、双方向の会話に基づいて、表示される画像をリアルタイムで変更されるシステムは知られていない。

先行技術文献

特許文献

- [0004] 特許文献1：特開2020-089641号公報

発明の概要

発明が解決しようとする課題

- [0005] この発明は、例えば会話に基づいてリアルタイムに共有画像が変更されるシステムを提供することを目的とする。

課題を解決するための手段

- [0006] この発明は、基本的には、音声情報に含まれるコンテンツに関する情報と、コンテンツの変化に関する情報を用いて、コンテンツを表す画像におけるコンテンツの位置を変化させることで、会話に基づいてリアルタイムに共有画像を変更できるという知見に基づく。
- [0007] このシステム1は、音声情報入力部3と、音声解析部5と、画像変化部7

とを有する。

音声情報入力部 3 は、音声情報を入力するための要素である。

音声解析部 5 は、音声情報入力部 3 により入力された音声情報を解析するための要素である。画像変化部 7 は、音声解析部 5 が解析した音声情報に含まれるコンテンツに関する情報と、音声解析部 5 が解析した音声情報に含まれるコンテンツの変化に関する情報とを用いて、コンテンツを表す画像の位置を変化させるための要素である。

[0008] このシステム 1 の画像変化部 7 は、位置を変化させたコンテンツの形状及び色のいずれか又は両方を変化させるものであってもよい。

[0009] このシステム 1 の好ましい態様は、第 1 の参加者の操作情報である第 1 の操作情報を入力するための第 1 操作入力部 2 1 と、第 2 の参加者の操作情報である第 2 の操作情報を入力するための第 2 操作入力部 2 3 と、をさらに有するものである。

そして、画像変化部 7 は、音声解析部 5 が解析した音声情報に含まれるコンテンツに関する情報、音声解析部が解析したコンテンツの変動に関する情報、第 1 の操作情報及び第 2 の操作情報に基づいて画像におけるコンテンツの位置を変化させる。画像変化部 7 は、位置を変化させたコンテンツの形状及び色のいずれか又は両方を変化させるものであってもよい。

[0010] このシステム 1 は、コンピュータを有するシステムを用いた音声に基づいて画像を変更するための方法を提供する。

この方法は、システム 1 に、音声情報が入力する音声情報入力工程と、システム 1 が、音声情報入力工程により入力された音声情報を解析するための音声解析工程と、

システム 1 が、音声解析工程において解析された音声情報に含まれるコンテンツに関する情報と、コンテンツの変化に関する情報を用いて、コンテンツを表す画像におけるコンテンツの位置を変化させる画像変換工程とを含む。

[0011] この明細書は、コンピュータを上記のシステムとして機能させるためのプ

ログラムや、そのプログラムを格納した情報記録媒体を提供する。

発明の効果

[0012] この発明は、例えば会話に基づいてリアルタイムに共有画像を変更できるシステムなどを提供できる。

図面の簡単な説明

[0013] [図1]図1は、音声に基づいて画像を変更するためのシステムを説明するためのブロック図である。

[図2]図2は、コンピュータの基本構成を示すブロック図である。

[図3]図3は、本発明のシステム例を示す概念図である。

[図4]図4は、プレゼンテーション資料の例を示す図である。

[図5]図5は、画像変化部によりコンテンツを表す画像におけるコンテンツの位置が変化した後の画像を示す概念図である。

[図6]図6は、画像変化部によりコンテンツを表す画像におけるコンテンツの位置、形状及び色が変化した後の画像を示す概念図である。

[図7]図7は、操作情報に基づいて、コンテンツを表す画像におけるコンテンツの位置が変化した後の画像を示す概念図である。

発明を実施するための形態

[0014] 以下、図面を用いて本発明を実施するための形態について説明する。本発明は、以下に説明する形態に限定されるものではなく、以下の形態から当業者が自明な範囲で適宜修正したものも含む。

[0015] 以下、音声に基づいて画像を変更するためのシステム1について説明する。このシステムは、コンピュータに基づくシステムである。

図1は、音声に基づいて画像を変更するためのシステムを説明するためのブロック図である。図1に示される通り、このシステム1は、音声情報入力部3と、音声解析部5と、画像変化部7とを有する。このシステム1は、プレゼンテーション資料などを記憶したプレゼンテーション資料記憶部11、プレゼンテーション資料を出力するプレゼンテーション出力部13、各種操作情報が入力される操作情報入力部15、音声情報を出力する音声出力部1

7をさらに有してもよい。

[0016] 図2は、コンピュータの基本構成を示すブロック図である。この図に示されるように、コンピュータは、入力部31、出力部33、制御部35、演算部37及び記憶部39を有しており、各要素は、バス41などによって接続され、情報の授受を行うことができるようにされている。例えば、記憶部には、制御プログラムが記憶されていてもよいし、各種情報が記憶されていてもよい。入力部から所定の情報が入力された場合、制御部は、記憶部に記憶される制御プログラムを読み出す。そして、制御部は、適宜記憶部に記憶された情報を読み出し、演算部へ伝える。また、制御部は、適宜入力された情報を演算部へ伝える。演算部は、受け取った各種情報を用いて演算処理を行い、記憶部に記憶する。制御部は、記憶部に記憶された演算結果を読み出して、出力部から出力する。このようにして、各種処理が実行される。以下説明する各要素は、コンピュータのいずれかの要素に対応していてもよい。

[0017] 図3は、本発明のシステム例を示す概念図である。図3に示されるように、本発明のシステム（本発明の装置を含むシステム）は、インターネット又はイントラネット43と接続された端末45と、インターネット又はイントラネット43に接続されたサーバ47とを含むものであってもよい。もちろん、単体のコンピュータや携帯端末が、本発明の装置として機能してもよいし、複数のサーバが存在してもよい。

[0018] 音声情報入力部3は、音声情報を入力するための要素（例えばマイク）である。システムは、複数の端末と、複数の端末と接続されたサーバにより構成される場合を例に説明する。この場合において、音声情報入力部3は、例えば、各端末の入力部（例えばマイク）から入力された音声情報を、サーバにおける入力部が入力する。この際のサーバ内の入力部が音声情報入力部3として機能してもよいし、各端末の入力部が音声情報入力部3として機能してもよい。いずれにせよ、サーバ内に、音声情報が入力される。

[0019] 以下、医薬情報担当者（MR）と、医療従事者とがそれぞれのコンピュータ端末（端末）を介してウェブ会議による会話をする例に基づいて説明する

。MRの端末と医療従事者の端末とは、事前に会議アクセス情報が入力される。会議アクセス情報の例は、会議を特定する情報（URL）や、会議のIDと会議のパスワードである。

[0020] MRの端末45aと医療従事者の端末45bは、会議アクセス情報を用いて、ウェブ会議用のサーバへアクセスする。MRの端末45aと医療従事者の端末45bに会議アクセス情報が入力されると、これらの端末は、会議用サーバへアクセスする。サーバは、入力部及び出力部を有しているので、それぞれの画像や音声情報がサーバを介して交換可能となる。この例では、サーバは、プレゼンテーション資料記憶部11を有している。この例では、ウェブ会議の参加者の各端末が、サーバへアクセスし、サーバに格納された画像データに基づいて、会話をを行い、画像データがその会話を反映して更新される。

[0021] 例えば、MRが入力装置（マウス）を用いて、プレゼンテーション資料記憶部21から議論の対象となるプレゼンテーション資料を読み出す。読み出したプレゼンテーション資料は、サーバの出力部からMRの端末45aと医療従事者の端末45bへと出力される。MRの端末45aと医療従事者の端末45bは、プレゼンテーション資料が入力される。MRの端末45aと医療従事者の端末45bの出力部であるモニタは、プレゼンテーション資料を表示する。なお、MR又は医療従事者の会話情報がサーバに入力され、会話情報を解析して、解析した会話情報に基づいて、プレゼンテーション資料記憶部11から議論の対象となるプレゼンテーション資料が読み出されてもよい。この場合、プレゼンテーション資料記憶部11には、各プレゼンテーション資料と関連して読み出し語が記憶され、解析した会話情報に読み出し語が含まれている場合に、その読み出し語と関連したプレゼンテーション資料が読み出されるようにすればよい。

[0022] 図4は、プレゼンテーション資料の例を示す図である。図4の例では、「下記降圧薬の効果と安全性について、先生のイメージを教えてください。」という表記がなされている。そして、この図では、縦軸が薬の効果（薬効）

であり、横軸が安全性を示すグラフが示されている。

[0023] MRの端末45aと医療従事者の端末45bを介して、両者は様々な会話をする。例えば、MRが「図に薬剤A～Dの薬効と安全性が示されていますが、先生の感覚はいかがでしょう？」と話をする。すると、MRの端末45aのマイクから、MRの端末45aに音声情報が入力される。入力された音声情報は、MRの端末45aの記憶部に一時的に記憶される。そして、MRの端末45aの出力部から音声情報が、サーバ47へと出力される。サーバ47が音声情報を受け取り、サーバの入力部から音声情報がサーバ47に入力され、サーバ47内の記憶部に記憶される。このようにして、システム1に、音声情報を入力する（音声情報入力工程）。

[0024] 医療従事者が、例えば「薬剤Cは、薬剤Bより薬効が高く、もう少し安全性が高いと思いますよ。」と会話したとする。すると、上記と同様にして、システム1に、音声情報が入力される（音声情報入力工程）。

[0025] 音声解析部5は、音声情報入力部3により入力された音声情報を解析するための要素（例えば公知の音声認識部）である。

例えば、サーバの記憶部は、プレゼンテーション資料に関連して、コンテンツ（例えば、薬剤A、薬剤B、薬剤C及び薬剤D）と、それぞれのコンテンツの画像における位置（薬剤A、薬剤B、薬剤C及び薬剤Dの座標）を記憶している。そして、音声解析部5は、「薬剤C」「薬剤B」というコンテンツに関する情報を解析結果として得る。また音声解析部5は、「薬剤Cは、薬剤Bより薬効が高く」というコンテンツの変化に関する情報を解析結果として得る。さらに音声解析部5は、「薬剤Cは、もう少し安全性が高い」というコンテンツの変化に関する情報を解析結果として得る。

このようにして、システム1が、音声情報入力工程により入力された音声情報を解析し、解析結果を得る（音声解析工程）。得られた解析結果は、適宜サーバの記憶部に記憶される。解析結果は、サーバの出力部から、MRの端末45aや医療従事者の端末45bに出力され、これらの端末の記憶部に記憶されてもよい。この処理は以下同様である。

[0026] 画像変化部 7 は、音声解析部 5 が解析した音声情報に含まれるコンテンツに関する情報と、音声解析部 5 が解析した音声情報に含まれるコンテンツの変化に関する情報とを用いて、コンテンツを表す画像におけるコンテンツの位置を変化させる。例えば、画像変化部 7 は、コンテンツに関する情報により変化させるコンテンツを特定し、コンテンツの変化に関する情報に基づいてそのコンテンツを変化させる処理を行えばよい。変化した画像は、表示部に表示される。例えば、サーバは、サーバの記憶部から、解析結果である、「薬剤 C」「薬剤 B」というコンテンツに関する情報や、「薬剤 C は、薬剤 B より薬効が高く」というコンテンツの変化に関する情報、及び「薬剤 C は、もう少し安全性が高い」というコンテンツの変化に関する情報を読み出す。そして、読み出した解析結果を用いて、コンテンツの画像における薬剤 C の位置を変化させる。そして、コンテンツの位置が変化した画像は、サーバの記憶部に記憶される。そして、MR の端末 4 5 a と医療従事者の端末 4 5 b へ出力されてもよい。すると、MR の端末 4 5 a と医療従事者の端末 4 5 b の出力部（モニタ）に変化後の画像が表示される。画像変化部 7 は、音声情報について音声を発した者（発言者）を特定し、特定した発言者に応じて変化量を調整するようにしてもよい。発言者の特定は、端末の ID や、ウェブ会議の入力の際に行えばよい。また、変化量の調整は、発言者の ID と関連して、調整に用いる係数を記憶させておき、発言者の ID に応じて読み出せばよい。このようにして発言者の個性に応じた変化を自動的に（発言者に悟られずに）行うことができることとなる。コンテンツの位置を変化させず、コンテンツの形や色を変化させるものは、この明細書に記載された上記とは別の態様である。

[0027] 図 5 は、画像変化部によりコンテンツを表す画像におけるコンテンツの位置が変化した後の画像を示す概念図である。このようにして、システム 1 が、音声解析工程において解析された音声情報に含まれるコンテンツに関する情報と、コンテンツの変化に関する情報に基づいて、コンテンツを表す画像におけるコンテンツの位置を変化させる（画像変化工程）。

[0028] 図5を見た医師は、医療従事者の端末45bの入力装置（例えばマウスやタッチパネル）を用いて、画像における薬剤Cをドラックする。そして、医師が、薬剤Cは、「このあたりかなあ」と発言する。この発言は、先の説明と同様に、システム内に入力される。また、医療従事者の端末45bによる操作情報は、医療従事者の端末45bの出力部からサーバ37へ出力される。操作情報を受け取ったサーバ37は、サーバ37の入力部から操作情報をサーバ37内へ入力する。そして、サーバ37は、操作情報に基づいて、コンテンツの画像における薬剤Cの位置を変化させる。

[0029] 図6は、画像変化部によりコンテンツを表す画像におけるコンテンツの位置、形状及び色が変わった後の画像を示す概念図である。図6に示されるように、画像変化部7は、位置を変化させたコンテンツが分かるように、その形状及び色のいずれか又は両方を変化させるものであってもよい。形状を変化させるには、大きさを変化させるものや、消滅させるもの、別のコンテンツ画像とするもの、及び枠の形を変化させるものであってもよい。後述するように、操作情報に基づいてコンテンツの位置を変化させた場合と、音声に基づいてコンテンツの位置を変化させた場合とで、変化後のコンテンツの形状及び色のいずれか又は両方を異ならせるようにしてもよい。

[0030] 図7は、操作情報に基づいて、コンテンツを表す画像におけるコンテンツの位置が変わった後の画像を示す概念図である。このように、このシステムは、操作情報に基づいて、コンテンツの位置が変わってもよい。そして、コンテンツの位置が変わった画像は、サーバの記憶部に記憶される。そして、MRの端末45aと医療従事者の端末45bへ出力されてもよい。すると、MRの端末45aと医療従事者の端末45bの出力部（例えばモニタ）に変化後の画像が表示される。このシステム1の好ましい態様は、第1の参加者の操作情報である第1の操作情報を入力するための第1操作入力部21と、第2の参加者の操作情報である第2の操作情報を入力するための第2操作入力部23と、をさらに有するものである。上記の例では、例えば、MRの端末45aの入力装置が第1操作入力部21として機能する。また、医療従事

者の端末45bの入力装置が第2操作入力部23として機能する。例えば、MRの端末45aは、マウスにより薬剤Cを右方向にドラックされたという入力情報を受け取る。受け取った入力情報は、操作情報として、サーバへ送信される。操作情報を受け取ったサーバは、画像変化部7に、操作情報に基づいた画像の変化を行わせる。変化後の画像をサーバの記憶部に記憶する。記憶された画像は、MRの端末45aと医療従事者の端末45bへと出力される。すると、MRの端末45aと医療従事者の端末45bは、受け取った画像をそれぞれの表示部に表示する。医療従事者の端末45bによる操作情報も同様である。このようにして、画像変化部7は、音声解析部5が解析した音声情報に含まれるコンテンツに関する情報、音声解析部が解析したコンテンツの変動に関する情報、第1の操作情報及び第2の操作情報に基づいて画像におけるコンテンツの位置を変化させる。上記は位置を変化させる操作情報の例であった。操作情報が、コンテンツの形状や色を変化させるものである場合、画像変化部7は、位置を変化させたコンテンツの形状及び色のいずれか又は両方を変化させるものであってもよい。

[0031] この明細書は、コンピュータを上記のシステムとして機能させるためのプログラムや、そのプログラムを格納した情報記録媒体を提供する。このプログラムは、コンピュータに実装され、コンピュータを、音声情報を入力するための音声情報入力手段と、音声情報入力手段により入力された音声情報を解析するための音声解析手段と、音声解析手段が解析した音声情報に含まれるコンテンツに関する情報と、コンテンツの変化に関する情報を用いて、コンテンツを表す画像におけるコンテンツの位置を変化させる画像変化手段とを有する、音声に基づいて画像を変更するためのシステムとして機能させる。

産業上の利用可能性

[0032] この発明は、音声認識システムに用いられるので、情報産業において利用され得る。

符号の説明

- [0033] 1 システム
- 3 音声情報入力部
- 5 音声解析部
- 7 画像変化部

請求の範囲

- [請求項1] 音声情報を入力するための音声情報入力部（３）と、
前記音声情報入力部（３）により入力された音声情報を解析するための音声解析部（５）と、
前記音声解析部（５）が解析した音声情報に含まれるコンテンツに関する情報と、前記コンテンツの変化に関する情報を用いて、前記コンテンツを特定するとともに、前記コンテンツを表す画像における前記コンテンツの位置を変化させる画像変化部（７）とを有し、
前記画像変化部（７）は、前記音声情報を発した者である発言者を特定するとともに、前記発言者に応じて前記コンテンツの位置の変化量を調整する、
音声に基づいて画像を変更するためのシステム（１）。
- [請求項2] 請求項１に記載のシステムであって、
前記画像変化部（７）は、位置を変化させたコンテンツの形状及び色のいずれか又は両方を変化させるものである、システム。
- [請求項3] 請求項１に記載のシステムであって、
第１の参加者の操作情報である第１の操作情報を入力するための第１操作入力部（２１）と、第２の参加者の操作情報である第２の操作情報を入力するための第２操作入力部（２３）と、をさらに有し、
前記画像変化部（７）は、
前記音声解析部（５）が解析した音声情報に含まれるコンテンツに関する情報、前記音声解析部が解析した前記コンテンツの変動に関する情報、第１の操作情報及び第２の操作情報に基づいて前記画像における前記コンテンツの位置を変化させるシステム。
- [請求項4] コンピュータを有するシステムを用いた音声に基づいて画像を変更するための方法であって、
前記システムに、音声情報が入力する音声情報入力工程と、
前記システムが、前記音声情報入力工程により入力された音声情報

を解析するための音声解析工程と、

前記システムが、前記音声解析工程において解析された音声情報に含まれるコンテンツに関する情報と、前記コンテンツの変化に関する情報を用いて、前記コンテンツを特定するとともに、前記コンテンツを表す画像における前記コンテンツの位置を変化させる画像変化工程とを含み、

前記画像変化工程は、前記音声情報を発した者である発言者を特定するとともに、前記発言者に応じて前記コンテンツの位置の変化量を調整する、方法。

[請求項5]

コンピュータを、

音声情報を入力するための音声情報入力手段と、

前記音声情報入力手段により入力された音声情報を解析するための音声解析手段と、

前記音声解析手段が解析した音声情報に含まれるコンテンツに関する情報と、前記コンテンツの変化に関する情報を用いて、前記コンテンツを特定するとともに、前記コンテンツを表す画像における前記コンテンツの位置を変化させる画像変化手段とを有し、

前記画像変化手段は、前記音声情報を発した者である発言者を特定するとともに、前記発言者に応じて前記コンテンツの位置の変化量を調整する、

音声に基づいて画像を変更するためのシステムとして機能させるためのプログラム。

[請求項6]

請求項5に記載のプログラムを格納した情報記録媒体。