

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関  
国際事務局

(43) 国際公開日  
2020年9月24日(24.09.2020)



(10) 国際公開番号

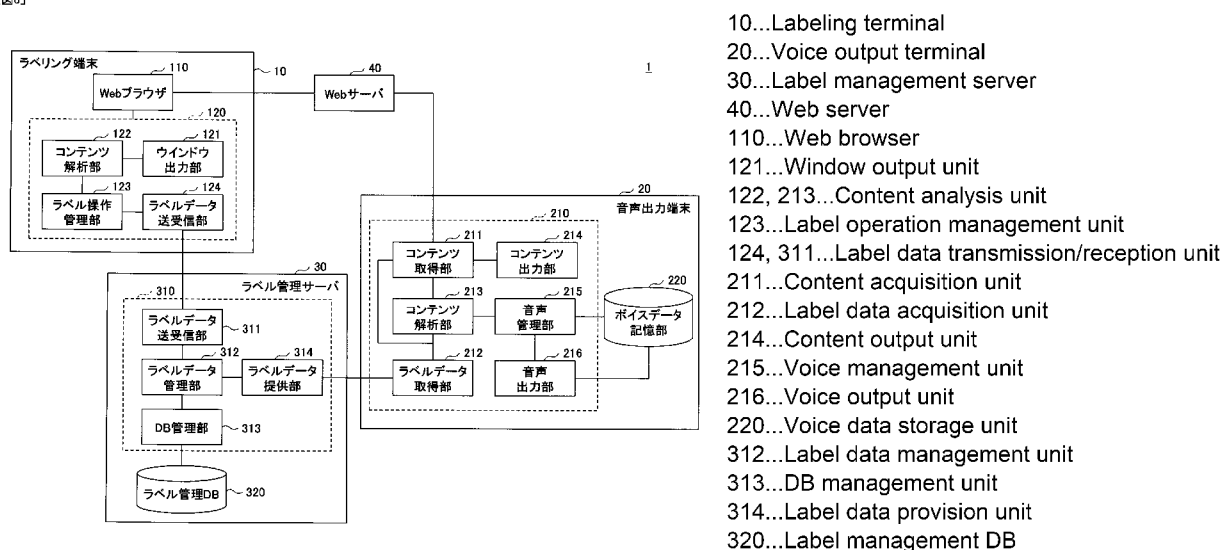
WO 2020/189376 A1

- (51) 国際特許分類:  
G10L 13/00 (2006.01) G10L 13/10 (2013.01)
- (21) 国際出願番号: PCT/JP2020/010032
- (22) 国際出願日: 2020年3月9日(09.03.2020)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:  
特願 2019-050337 2019年3月18日(18.03.2019) JP
- (71) 出願人: 日本電信電話株式会社 (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) [JP/JP]; 〒1008116 東京都千代田区大手町一丁目5番1号 Tokyo (JP).
- (72) 発明者: 白井 良成 (SHIRAI, Yoshinari); 〒1808585 東京都武蔵野市緑町3丁目9-11 NTT 知的財産センタ内 Tokyo (JP).  
藤田 早苗 (FUJITA, Sanae); 〒1808585 東京都武蔵野市緑町3丁目9-11 NTT 知的財産センタ内 Tokyo (JP).
- (74) 代理人: 伊東 忠重, 外 (ITO, Tadashige et al.); 〒1000005 東京都千代田区丸の内二丁目1番1号 丸の内 M Y P L A Z A (明治安田生命ビル) 16階 Tokyo (JP).
- (81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO,

(54) Title: VOICE OUTPUT METHOD, VOICE OUTPUT SYSTEM, AND PROGRAM

(54) 発明の名称: 音声出力方法、音声出力システム及びプログラム

[図6]



(57) Abstract: A voice output method that is executed by a voice output system that includes a first terminal, a server, and a second terminal. The method is characterized in that the first terminal executes a first label assignment step in which label data that indicate the attributes of speakers for the reading of character strings included in content in synthetic voices are assigned to the character strings and a transmission step in which the label data are transmitted to the server, the server executes a storage step in which the label data transmitted from the first terminal are associated with content identification information that identifies the content and stored in a database, and the second terminal executes an acquisition step in which the label data that correspond to the content identification information for the content are acquired from the server, a second label assignment step in which the acquired label data are assigned to the character strings included in the content, a specification step in which the label data assigned to each character string included in the content is used to specify,

DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

- (84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類:

- 国際調査報告 (条約第21条(3))

---

from a plurality of voice data, voice data for a synthetic voice that is to be used for the reading of the character string, and a voice output step in which voice is outputted such that each character string included in the content is read in a synthetic voice that uses the specified voice data.

(57) 要約: 第1の端末と、サーバと、第2の端末とが含まれる音声出力システムが実行する音声出力方法であって、前記第1の端末は、コンテンツに含まれる文字列に対して、該文字列を合成音声で読み上げる場合における話者の属性を表すラベルデータを付与する第1のラベル付与手順と、前記ラベルデータを前記サーバに送信する送信手順と、を実行し、前記サーバは、前記第1の端末から送信されたラベルデータを、前記コンテンツを識別するコンテンツ識別情報と対応付けてデータベースに保存する保存手順を実行し、前記第2の端末は、前記コンテンツのコンテンツ識別情報に対応するラベルデータを前記サーバから取得する取得手順と、取得したラベルデータを、前記コンテンツに含まれる各文字列に付与する第2のラベル付与手順と、前記コンテンツに含まれる各文字列にそれぞれ付与されているラベルデータを用いて、複数の音声データの中から、前記文字列の読み上げに用いられる合成音声の音声データをそれぞれ特定する特定手順と、前記コンテンツに含まれる各文字列のそれぞれを、特定された音声データを用いた合成音声により読み上げて音声出力する音声出力手順と、を実行する、ことを特徴とする。

## 明 細 書

### 発明の名称：音声出力方法、音声出力システム及びプログラム 技術分野

[0001] 本発明は、音声出力方法、音声出力システム及びプログラムに関する。

#### 背景技術

[0002] 音声合成（speech synthesis）と呼ばれる技術が従来から知られている。音声合成は、例えば、視覚障がい者への情報伝達や、ディスプレイを十分にすることができない状況下での情報伝達（例えば、車の運転中におけるカーナビによる情報伝達）等に利用されてきた。近年では、少し聞いただけでは人間の声と区別がつかないほど合成音声の性能が向上しており、スマートフォンやスマートスピーカ等の普及とも相まって音声合成が広く普及しつつある。

[0003] 音声合成は、典型的にはテキストを合成音声に変換することに用いられる。この場合、音声合成は、しばしばテキスト音声合成（TTS:Text-to-speech）とも呼ばれる。テキスト音声合成の有効な利用方法として、スマートフォン等を用いた電子書籍の読み上げやWebページの読み上げ等がある。例えば、青空文庫等の小説サイトのテキストを合成音声で読み上げるスマートフォン用アプリケーションが知られている（非特許文献1）。

[0004] 音声合成を利用することで、視覚障がい者だけでなく一般の人も、例えば、混雑した電車の中や運転中等のスマートフォンが操作困難な状況下でも電子書籍やWebページを合成音声で読み上げさせて聞くこと等ができるようになる。また、これ以外にも、例えば、能動的に文字を読むのが面倒だと感じた場合等に、合成音声で読み上げさせることで、受動的に情報を得ることができるようになる。

[0005] 他方で、小説の読者の理解を助けるために、小説中の発話文の発話者を推定する研究が行われている（非特許文献2）。

#### 先行技術文献

## 非特許文献

[0006] 非特許文献1：“青空司書”，[online]，〈URL: <https://sites.google.com/site/aozorashisho/>〉

非特許文献2：He, et.al, “Identification of Speakers in Novels”, Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pages 1312-1320.

## 発明の概要

### 発明が解決しようとする課題

[0007] ところで、音声合成を利用してテキストを読み上げる場合、合成音声の声（以下、「ボイス」とも表す。）は、例えば、スマートフォンに搭載されているOS（Operating System）やアプリケーション上でユーザが予め設定したボイスに固定される。このため、例えば、ユーザがイメージしていた声とは異なる声でテキストが読み上げられてしまう場合がある。

[0008] 例えば、年配の男性のような声質のボイスが設定された状態で、音声合成により小説を読み上げた場合、若い女性がイメージされる登場人物の発話文も、年配の男性のような声質のボイスで読み上げられてしまう。

[0009] これに対して、例えば、読み上げ対象のコンテンツ（電子書籍やWebページ等）中の部分文字列をどのような年齢や性別のボイスで読み上げればよいかを識別し、この識別結果に応じてボイスを切り替えながらテキストを読み上げればよいと考えられる。しかしながら、テキスト中の部分文字列の主体（例えば、会話文であれば発話者の属性等）を識別することは容易ではない。また、仮に識別できたとしても、その識別結果に応じて音声合成のボイスを変更して出力するようなアプリケーションは現状では存在しない。

[0010] 本発明は、上記の点に鑑みてなされたもので、コンテンツに付与された属性情報に応じた音声出力することを目的とする。

### 課題を解決するための手段

[0011] 上記目的を達成するため、本発明の実施の形態では、第1の端末と、サー

バと、第2の端末とが含まれる音声出力システムが実行する音声出力方法であって、前記第1の端末は、コンテンツに含まれる文字列に対して、該文字列を合成音声で読み上げる場合における話者の属性を表すラベルデータを付与する第1のラベル付与手順と、前記ラベルデータを前記サーバに送信する送信手順と、を実行し、前記サーバは、前記第1の端末から送信されたラベルデータを、前記コンテンツを識別するコンテンツ識別情報と対応付けてデータベースに保存する保存手順を実行し、前記第2の端末は、前記コンテンツのコンテンツ識別情報に対応するラベルデータを前記サーバから取得する取得手順と、取得したラベルデータを、前記コンテンツに含まれる各文字列に付与する第2のラベル付与手順と、前記コンテンツに含まれる各文字列にそれぞれ付与されているラベルデータを用いて、複数の音声データの中から、前記文字列の読み上げに用いられる合成音声の音声データをそれぞれ特定する特定手順と、前記コンテンツに含まれる各文字列のそれぞれを、特定された音声データを用いた合成音声により読み上げて音声出力する音声出力手順と、を実行する、ことを特徴とする。

### 発明の効果

[0012] コンテンツに付与された属性情報に応じた音声を出力することができる。

### 図面の簡単な説明

[0013] [図1]読み上げ対象のコンテンツの一例を説明するための図である。

[図2]ボイスの割り当ての一例を説明するための図である。

[図3]XML形式のタグでラベルの付与を実現する場合の一例を説明するための図である。

[図4]本発明の実施の形態における音声出力システムの全体構成の一例を示す図である。

[図5]ラベリング画面の一例を示す図である。

[図6]本発明の実施の形態における音声出力システムの機能構成の一例を示す図である。

[図7]ラベル管理DBに格納されているラベルデータの構成の一例を示す図であ

る。

[図8]本発明の実施の形態におけるラベル付与処理の一例を示すフローチャートである。

[図9]本発明の実施の形態におけるラベルデータ保存処理の一例を示すフローチャートである。

[図10]本発明の実施の形態における音声出力処理の一例を示すフローチャートである。

[図11]コンピュータのハードウェア構成の一例を示す図である。

### 発明を実施するための形態

[0014] 以下、本発明の実施の形態について説明する。本発明の実施の形態では、ヒューマンコンピュータ技術によりコンテンツ中の部分文字列にラベルを付与した上で、部分文字列に付与されたラベルに応じてボイスを切り替えながら合成音声を出力する音声出力システム1について説明する。これにより、本発明の実施の形態における音声出力システム1では、イメージした声に近いボイスで、コンテンツ中の各部分文字列を音声出力することが可能となる。

[0015] ここで、ラベルとは、コンテンツ中の部分文字列を音声合成により読み上げる際に、当該部分文字列を読み上げる話者の識別情報（例えば、話者の名前等）や属性（例えば、年齢や性別等）を表す情報のことである。また、コンテンツとは、テキスト（つまり、文字列）で表現される電子データのことであり、例えば、Webページや電子書籍等が挙げられる。本発明の実施の形態では、コンテンツは、一例として、Webページ上の文章（例えばWebページ上で公開されている小説等）であるものとする。

[0016] 更に、ヒューマンコンピュータ技術とは、一般に、コンピュータにとって解くことが困難な課題を、人間の処理能力を利用して解決する技術のことである。本発明の実施の形態では、コンテンツ中の部分文字列に対するラベルの付与をヒューマンコンピュータ技術により実現する（つまり、部分文字列に対するラベルの付与は、後述するラベリング画面等のUI（ユ

ーザインタフェース）を用いて人手で行われる。）。

[0017] なお、本発明の実施の形態では、異なるボイスで読み上げられる複数の部分文字列がコンテンツ中に存在することを想定しているが、これに限られない。本発明の実施の形態は、例えば、1つのコンテンツ中の全文字列が1つのボイスで読み上げられる場合であっても同様に適用することができる（ただし、この場合、コンテンツ中の部分文字列とは全文字列を意味する。）。

[0018] <コンテンツとボイスの割り当て>

まず、音声合成による読み上げ対象のコンテンツ中の部分文字列に対するボイスの割り当てについて説明する。

[0019] 図1は、読み上げ対象のコンテンツの一例である。図1では、コンテンツの一例として、夏目漱石の小説「こころ」の一部を取り出したものを表している。小説等のコンテンツには、一人称視点で記載されている文書や三人称視点で記載されている文章、或る登場人物の発話を表す文章等が存在する。

[0020] 例えば、図1に示す例では、『これからどこへ行くという目的のない私は、ただ先生の歩く方へ歩いて行った。先生はいつもより口数を利かなかった。それでも私はさほどの窮屈を感じなかったので、ぶらぶらいっしょに歩いて行った。』との文章は一人称視点で記載されている文章であり、『「すぐお宅へお帰りですか」』との文章は登場人物「私」の発話を表す文章である。同様に、『「ええ別に寄る所ありませんから」』との文章は登場人物「先生」の発話を表す文章であり、『二人はまた黙って南の方へ坂を下りた。』との文章は三人称視点で記載されている文章である。また、『「先生のお宅の墓地はあすこにあるんですか」と私がまた口を利き出した。』との文章は、括弧（「」）内の文章が登場人物「私」の発話を表す文章、それ以降の文章が一人称視点で記載されている文章である。

[0021] 図1に示すコンテンツを音声合成で読み上げる場合、登場人物「私」の発話文を読み上げるボイスと、登場人物「先生」の発話文を読み上げるボイスとは異なるボイスで、かつ、それぞれが常に同じボイスで読み上げられることが好ましいと考えられる。

[0022] また、発話文（つまり、括弧内の文章）以外の文章では、三人称視点であれば登場人物の発話に用いられているボイスとは異なるボイスで読み上げられることが好ましいと考えられる。一方で、一人称視点であれば該当の登場人物（図1に示す例では「私」）と同一のボイスで読み上げられることが好ましいと考えられる。

[0023] 以上により、図1に示すコンテンツを音声合成で読み上げる場合には、例えば、図2に示すように、登場人物「私」を表すボイス1と、登場人物「先生」を表すボイス2と、第三者視点の文章を読み上げるためのナレーションを表すボイス3とを用いて、コンテンツ中の各部分文字列に対して該当のボイスを割り当てて読み上げることが好ましい。

[0024] つまり、例えば小説等のコンテンツでは、一般に、同一登場人物の発話文は同一ボイスを割り当てた上で一貫してそのボイスで読み上げることが好ましく、地の文章（発話文でない文章）は三人称視点や一人称視点等に応じたボイスを割り当てた上で一貫してそのボイスで読み上げることが好ましい。

[0025] なお、図1に示す例では、コンテンツの一例として小説を挙げたが、限られないことはいうまでもない。コンテンツは、電子書籍等の小説以外にも、論説や論文、漫画等でもよいし、例えばニュースサイト等のWebページであってもよい。

[0026] 特に、例えば、ニュースサイトのWebページでは、ユーザによっては男性キャスター風に読み上げて欲しい人もいれば女性キャスター風に読み上げて欲しい人もいると思われる。また、例えば、ニュースサイトの記事中に登場する政治家のコメント等は、その政治家の性別や年齢に応じたボイスで読み上げて欲しいこともある。更に、論文等も、地の文は第一著者の性別や年齢に応じたボイスで読み上げ、引用箇所等では別のボイスで読み上げると、論文内容の利用が進む可能性がある。本発明の実施の形態では、これらの場合についても適用可能である。

[0027] <部分文字列に対するラベルの付与>

上記の読み上げを実現するために、コンテンツ中の部分文字列にラベルを

付与する方法について説明する。

[0028] 例えば、Webページ上のコンテンツ中の各部分文字列に対して、図3に示すようなラベル（つまり、XML形式のタグ）が付与されていれば、図2に示すようなボイスの割り当てが可能となると考えられる。このようなラベルが付与されていれば、合成音声により読み上げを行うアプリケーションプログラムは、タグで囲まれた文章（部分文字列）の年齢（age）及び性別（gender）の属性値に応じて、その年齢及び性別に最も近いボイスを選択して読み上げることが可能となるためである。また、同一登場人物の発話であるか否かをid（識別情報）で管理し、同一idが割り振られた発話は同一ボイスで一貫して読み上げることが可能となる。

[0029] なお、図3に示す例では、SSML（Speech Synthesis Markup Language）に類似するラベルを用いたが、例えば、以下の参考文献1に記載されているように、発話文への発話者情報付与に関する既存ラベルを用いることも考えられる。

[0030] [参考文献1]

宮寄由美，柏野和佳子，山崎誠，”発話文への発話者情報付与の基本設計：『現代日本語書き言葉均衡コーパス』収録の小説を対象に”，言語資源活用ワークショップ発表論文集，2017.

[0031] しかしながら、上記のように、ラベルをコンテンツに埋め込む場合には、コンテンツに対する更新権限を持っている者（例えば、コンテンツの作成者等）しか、ラベルの付与や更新をすることができない。例えば、Webページ上に小説等のコンテンツを作成及び公開しているコンテンツ作成者がラベルを付与したり、更新したりするのはコンテンツ作成者にとっては負担が大きいと考えられる。また、Webページのコンテンツを複数のボイスで読み上げてもらいたいという強いモチベーションがコンテンツ作成者に存在するとも限らない。

[0032] そこで、本発明の実施の形態では、ヒューマンコンピューテーション技術を用いて、コンテンツ作成者とは異なる第三者（例えば、コンテンツの利用者

等)がWebページ上のコンテンツにラベルを付与するものとする。本発明の実施の形態では、ラベルを付与する第三者(この第三者を「ラベラー」とも表す。)が、コンテンツ中の部分文字列に対して、この部分文字列を読み上げる話者の識別情報と、性別と、年齢とを設定することで、ラベルを付与する。これにより、コンテンツ中の各部分文字列を、この部分文字列に付与されているラベルに応じたボイスで読み上げることを可能とする。なお、ラベルを付与の具体的な方法については後述する。

[0033] <音声出力システム1の全体構成>

次に、本発明の実施の形態における音声出力システム1の全体構成について、図4を参照しながら説明する。図4は、本発明の実施の形態における音声出力システム1の全体構成の一例を示す図である。

[0034] 図4に示すように、本発明の実施の形態における音声出力システム1には、1以上のラベリング端末10と、1以上の音声出力端末20と、ラベル管理サーバ30と、Webサーバ40とが含まれる。これらの各端末及び各サーバは、例えばインターネット等の通信ネットワークNを介して通信可能に接続される。

[0035] ラベリング端末10は、コンテンツ中の部分文字列に対するラベル付与に用いられるコンピュータである。ラベリング端末10としては、例えば、PC(パーソナルコンピュータ)、スマートフォン、タブレット端末等が利用可能である。

[0036] ラベリング端末10には、Webブラウザ110と、このWebブラウザ110に対するアドオン120とが搭載されている。なお、アドオン120とは、Webブラウザ110に対する拡張機能を提供するプログラムのことである。アドオンはアドインと称されることもある。

[0037] ラベリング端末10は、Webブラウザ110によりコンテンツを表示させることができる。また、ラベリング端末10は、アドオン120により、Webブラウザ110上に表示されたコンテンツ中の部分文字列に対してラベルを付与することができる。このとき、ラベリング端末10には、アドオン120

により、コンテンツ中の部分文字列に対してラベルを付与するためのラベリング画面が表示される。ラベラーは、このラベリング画面上で、コンテンツ中の部分文字列に対するラベルを付与することができる。なお、ラベリング画面については後述する。

[0038] そして、ラベリング端末10は、アドオン120により、当該部分文字列に付与されたラベルを表すデータ（以降、「ラベルデータ」とも表す。）をラベル管理サーバ30に送信する。

[0039] 音声出力端末20は、音声合成によりコンテンツを読み上げさせたいユーザが利用するコンピュータである。音声出力端末20としては、例えば、PC、スマートフォン、タブレット端末等が利用可能である。これら以外にも、例えば、ゲーム機器、デジタル家電、カーナビゲーション端末等の車載機、ウェアラブルデバイス、スマートスピーカ等が用いられてもよい。

[0040] 音声出力端末20は、音声出力アプリケーション210と、ボイスデータ記憶部220とを有する。音声出力端末20は、音声出力アプリケーション210により、コンテンツ中の部分文字列に付与されているラベルのラベルデータをラベル管理サーバ30から取得する。そして、音声出力端末20は、ボイスデータ記憶部220に記憶されているボイスデータを用いて、コンテンツ中の各部分文字列に付与されているラベルに応じたボイスで読み上げた音声を出力する。

[0041] ラベル管理サーバ30は、ラベルデータを管理するコンピュータである。ラベル管理サーバ30は、ラベル管理プログラム310と、ラベル管理DB320とを有する。ラベル管理サーバ30は、ラベル管理プログラム310により、ラベリング端末10から送信されたラベルデータをラベル管理DB320に格納する。また、ラベル管理サーバ30は、ラベル管理プログラム310により、音声出力端末20からの要求に応じて、ラベル管理DB320に格納されているラベルデータを当該音声出力端末20に送信する。

[0042] Webサーバ40は、コンテンツを管理するコンピュータである。Webサーバ40は、コンテンツ作成者により作成されたコンテンツを管理している。Web

サーバ40は、ラベリング端末10や音声出力端末20からの要求に応じて、この要求に係るコンテンツをラベリング端末10や音声出力端末20に送信する。

[0043] なお、図1に示す音声出力システム1の構成は一例であって、他の構成であってもよい。例えば、ラベリング端末10と音声出力端末20とが異なる端末でなくてもよい（つまり、1台の端末が、ラベリング端末10の機能と音声出力端末20の機能とを有していてもよい。）。

[0044] <ラベリング画面>

ここで、ラベリング端末10上に表示されるラベリング画面1000を図5に示す。図5は、ラベリング画面1000の一例を示す図である。図5に示すラベリング画面1000は、ラベリング端末10に搭載されているWebブラウザ110若しくはアドオン120（又はその両方）により表示される。

[0045] ラベリング画面1000には、コンテンツ表示欄1100と、ラベリングウィンドウ1200とが含まれる。コンテンツ表示欄1100は、コンテンツ及びラベリング結果が表示される表示領域である。ラベリングウィンドウ1200は、コンテンツ表示欄1100に表示されているコンテンツ中の部分文字列に対するラベル付与を行うためのダイアログウィンドウである。

[0046] ラベリングウィンドウ1200には、名前（NAME）と性別（SEX）と年齢（AGE）とが設定された話者がラジオボタンで選択可能なリストとして表示される。ここで、リスト中の各話者はラベルを表し、名前が識別情報、性別及び年齢が属性に相当する。

[0047] 図5に示す例では、名前「default」、性別「F」及び年齢「20」の話者と、名前「老爺」、性別「M」及び年齢「70」の話者と、名前「メロス」、性別「M」及び年齢「23」の話者と、名前「王様」、性別「M」及び年齢「43」の話者とがリストとして表示されている。

[0048] また、ラベリングウィンドウ1200には、ADDボタン、DELボタン、SAVEボタン及びLOADボタンが含まれる。ラベラーによりADDボタンが押下された場合、リストに話者が1人追加される。また、DELボタンが押下された場合、ラ

ジオボタンで選択されている話者がリストから削除される。SAVEボタンが押下された場合、コンテンツ中の部分文字列に付与されたラベルのラベルデータがラベル管理サーバ30に送信される。一方で、LOADボタンが押下された場合は、ラベル管理サーバ30で管理されているラベルデータが取得され、当該コンテンツの現在のラベリング状況が表示される。

[0049] コンテンツ表示欄1100に表示されているコンテンツ中の部分文字列に対してラベルを付与する場合、ラベラーは、ラベリングウィンドウ1200から所望の話者をラジオボタンで選択した上で、所望の部分文字列をマウス等で選択する。これにより、選択された部分文字列に対して、選択された話者及びその属性（年齢及び性別）を表すラベルが付与される。このとき、ラベルが付与された部分文字列は、どの話者及びその属性を表すラベルが付与されたかがわかるように、付与されたラベルが表す話者固有の色でマーキングされたり、当該話者固有の表示態様で表示されたりして、ラベリング状況が可視化される。

[0050] 図5に示す例では、コンテンツ表示欄1100に表示されているコンテンツ中の部分文字列『「王様は、人を殺します。」』に対して、話者「老爺」及びその属性（性別「M」、年齢「70」）を表すラベルが付与されている。同様に、図5に示す例では、部分文字列『「なぜ殺すのだ。」』に対して、話者「メロス」及びその属性（性別「M」、年齢「23」）を表すラベルが付与されている。

[0051] なお、名前「default」の話者は、ラベラーにより明示的にラベルが付与された部分文字列以外の部分文字列に付与されるラベルを表す。図5に示す例では、名前「老爺」、名前「メロス」及び名前「王様」を表すラベルが付与されていない部分文字列に対して、名前「default」の話者を表すラベルが付与されている。

[0052] 以上のように、ラベラーは、ラベリング画面1000上で、コンテンツ中の各部分文字列に対してラベルを付与することができる。これにより、後述するように、音声出力端末20の音声出力アプリケーション210では、部

分文字列に付与されているラベルに応じたボイスで、当該部分文字列を読み上げて音声出力をすることができる（言い換えれば、部分文字列に対してラベルが付与されることで、当該部分文字列にはラベルに応じたボイスが割り当てられることになる。）。

[0053] <音声出力システム１の機能構成>

次に、本発明の実施の形態における音声出力システム１の機能構成について、図６を参照しながら説明する。図６は、本発明の実施の形態における音声出力システム１の機能構成の一例を示す図である。

[0054] 《ラベリング端末１０》

図６に示すように、本発明の実施の形態におけるラベリング端末１０は、機能部として、ウインドウ出力部１２１と、コンテンツ解析部１２２と、ラベル操作管理部１２３と、ラベルデータ送受信部１２４とを有する。これら各機能部は、アドオン１２０がプロセッサ等に実行させる処理により実現される。

[0055] ウインドウ出力部１２１は、Webブラウザ１１０上に、上述したラベリングウインドウを表示する。

[0056] コンテンツ解析部１２２は、Webブラウザ１１０により表示されているコンテンツ（例えばWebページ等）の構造を解析する。ここで、コンテンツの構造としては、例えば、DOM（Document Object Model）等が挙げられる。

[0057] ラベル操作管理部１２３は、コンテンツ中の部分文字列に対するラベル付与に関する操作を管理する。例えば、ラベル操作管理部１２３は、ラベリングウインドウ中のリストから話者をラジオボタンで選択する操作やコンテンツ中の部分文字列をマウスで選択する操作等を受け付ける。

[0058] また、ラベル操作管理部１２３は、コンテンツ解析部１２２による解析結果に基づいて、例えば、マウスで選択された部分文字列が属するHTML（Hyper Text Markup Language）エレメントを取得し、ラベリング状況を可視化する処理（つまり、当該HTMLエレメントをラベル固有の色でマーキングする等の処理）を行う。

[0059] ラベルデータ送受信部 124 は、ラベリングウィンドウでSAVEボタンが押下された場合に、現在のコンテンツ中の部分文字列に付与されたラベルのラベルデータをラベル管理サーバ30に送信する。このとき、ラベルデータ送受信部 124 は、ラベリングしたコンテンツのURL (Uniform Resource Locator) もラベル管理サーバ30に送信する。なお、このとき、ラベルデータ送受信部 124 は、必要に応じて、ラベリングを行ったラベラーの情報（例えば、ラベラーのユーザID等）をラベル管理サーバ30に送信してもよい。

[0060] また、ラベルデータ送受信部 124 は、ラベリングウィンドウでLOADボタンが押下された場合に、ラベル管理サーバ30で管理されているラベルデータを受信する。これにより、ラベラーは、例えば、或るコンテンツを途中でラベリングしてラベルデータをラベル管理サーバ30に送信した場合に、その続きからラベリングを行うことが可能となる。

[0061] 《音声出力端末 20》

図6に示すように、本発明の実施の形態における音声出力端末20は、機能部として、コンテンツ取得部211と、ラベルデータ取得部212と、コンテンツ解析部213と、コンテンツ出力部214と、音声管理部215と、音声出力部216とを有する。これら各機能部は、音声出力アプリケーション210がプロセッサ等に実行させる処理により実現される。

[0062] また、本発明の実施の形態における音声出力端末20は、記憶部として、ボイスデータ記憶部220を有する。当該記憶部は、例えば、音声出力端末20が備える記憶装置等を用いて実現可能である。

[0063] コンテンツ取得部211は、Webサーバ40からコンテンツ（例えば、小説等の文章が公開されたWebページ）を取得する。

[0064] ラベルデータ取得部212は、コンテンツ取得部211により取得されたコンテンツのURL（つまり、コンテンツの識別情報）に対応するラベルデータをラベル管理サーバ30から取得する。ラベルデータ取得部212は、例えば、当該コンテンツのURLを含む取得要求をラベル管理サーバ30に送信することで、この取得要求の応答としてラベルデータを取得することができる。

- [0065] コンテンツ解析部 213 は、コンテンツ取得部 211 により取得されたコンテンツを解析して、当該コンテンツに含まれるテキストのどの部分文字列に対してどのラベルデータが付与されているかを特定する。
- [0066] コンテンツ出力部 214 は、コンテンツ取得部 211 により取得されたコンテンツを表示する。ただし、コンテンツ出力部 214 は、必ずしもコンテンツを表示する必要はない。コンテンツを表示しない場合には、音声出力端末 20 は、コンテンツ出力部 214 を有していなくてもよい。
- [0067] 音声管理部 215 は、コンテンツ解析部 213 による解析結果に基づいて、コンテンツ中の各部分文字列を、ボイスデータ記憶部 220 に記憶されているどのボイスデータで読み上げるかを特定する。すなわち、音声管理部 215 は、各部分文字列にそれぞれ付与されているラベルが表す属性を用いて、ボイスデータ記憶部 220 に記憶されているボイスデータの中から属性が最も近いボイスデータを検索し、検索されたボイスデータを、当該部分文字列を読み上げるボイスデータとして特定する。これにより、コンテンツ中の各部分文字列に対してボイスが割り当てられる。
- [0068] 音声出力部 216 は、コンテンツ中の各部分文字列を該当のボイスデータを用いた合成音声により読み上げて音声を出力する。このとき、音声出力部 216 は、音声管理部 215 で特定されたボイスデータを用いて、各部分文字列を読み上げて音声を出力する。なお、音声出力端末 20 のユーザは、合成音声の出力開始（つまり、再生）、一時停止、早送り（又は、次の部分文字列を再生）、巻き戻し（又は、1 つ前の部分文字列を再生）等の操作を行うことができてもよい。この場合、音声出力部 216 は、これらの操作に応じて、ボイスデータによる音声出力を制御する。
- [0069] ボイスデータ記憶部 220 は、コンテンツ中の部分文字列の読み上げに用いられるボイスデータを記憶する。ここで、ボイスデータ記憶部 220 では、各ボイスデータに対して属性（例えば、性別及び年齢）が対応付けて記憶されている。なお、これらのボイスデータは任意のボイスデータを利用可能であり、例えば、任意のサーバ等から予めダウンロードされる。ただし、ダ

ウンロードされたボイスデータに対して属性が対応付けされていない場合には、音声出力端末20のユーザは、当該ボイスデータに対して属性を対応付ける必要がある。

[0070]      《ラベル管理サーバ30》

図6に示すように、本発明の実施の形態におけるラベル管理サーバ30は、機能部として、ラベルデータ送受信部311と、ラベルデータ管理部312と、DB管理部313と、ラベルデータ提供部314とを有する。これら各機能部は、ラベル管理プログラム310がプロセッサ等を実行させる処理により実現される。

[0071]      また、本発明の実施の形態におけるラベル管理サーバ30は、記憶部として、ラベル管理DB320を有する。当該記憶部は、例えば、ラベル管理サーバ30が備える記憶装置、又はラベル管理サーバ30と通信ネットワークNを介して接続される記憶装置等を用いて実現可能である。

[0072]      ラベルデータ送受信部311は、ラベリング端末10からのラベルデータを受信する。また、ラベルデータ送受信部311は、ラベリング端末10にラベルデータを送信する。

[0073]      ラベルデータ管理部312は、ラベルデータ送受信部311によりラベルデータが受信された場合に、当該ラベルデータを検証する。ラベルデータの検証とは、例えば、ラベルデータのフォーマット（データ形式）が正しいか否かの検証等である。

[0074]      DB管理部313は、ラベルデータ管理部312により検証されたラベルデータをラベル管理DB320に格納する。

[0075]      なお、例えば、同一の部分文字列に対して異なるラベルを表すラベルデータがラベル管理DB320に既に格納されている場合、DB管理部313は、古いラベルデータを新しいラベルデータで更新してもよいし、古いラベルデータと新しいラベルデータとを共存させてもよい。又は、同一の部分文字列に対するラベルデータであっても、ラベラーのユーザIDが異なる場合には異なるラベルデータと扱ってもよい。

- [0076] ラベルデータ提供部314は、音声出力端末20からの取得要求に応じて、該当のラベルデータ（つまり、当該取得要求に含まれるURLに対応するラベルデータ）をラベル管理DB320から取得して、当該取得要求に対する応答として、取得したラベルデータを当該音声出力端末20に送信する。
- [0077] ラベル管理DB320は、ラベルデータが格納されている。ここで、ラベルデータは、上述したように、コンテンツ中の部分文字列に対して付与されるラベルを表すデータであり、ラベルは当該部分文字列を読み上げる話者の識別情報及び属性を表す。したがって、ラベルデータでは、コンテンツと、当該コンテンツ中で該当の部分文字列を特定可能な情報と、当該部分文字列を読み上げる話者の識別情報と、当該話者の属性とが少なくとも対応付けられていればよい。
- [0078] このようなラベルデータをラベル管理DB320に格納する際に、どのようなデータ構成で格納するかは任意のデータ構成を採用することが可能であるが、一例として、話者テーブルと部分文字列テーブルとを用いてラベル管理DB320に格納した場合を図7に示す。図7は、ラベル管理DB320に格納されているラベルデータの構成の一例を示す図である。
- [0079] 図7に示すように、話者テーブルには1つ以上の話者データが格納されており、各話者データには、データ項目として、「SPEAKER\_ID」と、「SEX」と、「AGE」と、「NAME」と、「COLOR」と、「URL」とが含まれる。
- [0080] データ項目「SPEAKER\_ID」には、話者データを識別するIDが設定される。データ項目「SEX」には、話者の属性として性別が設定される。データ項目「AGE」には、話者の属性として年齢が設定される。データ項目「NAME」には、話者の名前が設定される。データ項目「COLOR」には、ラベリング状況を可視化する際の話者固有の色が設定される。データ項目「URL」には、コンテンツのURLが設定される。
- [0081] なお、図7に示す例では、データ項目「NAME」に同一の名前が設定される場合も考慮して、データ項目「SPEAKER\_ID」に設定されるIDを話者の識別情報としている。ただし、例えば、データ項目「NAME」に同一の名前が設定され

ない場合には、話者の名前を識別情報としてもよい。

[0082] また、図7に示すように、部分文字列テーブルには1つ以上の部分文字列データが格納されており、各部分文字列データには、データ項目として、「TEXT」と、「POSITION」と、「SPEAKER\_ID」と、「URL」とが含まれる。

[0083] データ項目「TEXT」には、ラベラーにより選択された部分文字列が設定される。データ項目「POSITION」には、コンテンツ中における先頭からの当該部分文字列の出現回数が設定される。データ項目「SPEAKER\_ID」には、ラベラーにより選択された話者（つまり、ラベリングウィンドウで選択された話者）が設定される。データ項目「URL」には、コンテンツのURLが設定される。

[0084] 例えば、図7に示す部分文字列テーブルの3行目の部分文字列データには、データ項目「TEXT」に『「先生のお宅の墓地はあすこにあるんですか」と私がまた口を利き出した。』、データ項目「POSITION」に「0」、データ項目「SPEAKER\_ID」に「1」がそれぞれ設定されている。これは、部分文字列『「先生のお宅の墓地はあすこにあるんですか」と私がまた口を利き出した。』は、コンテンツ中の先頭から当該部分文字列までに同様の部分文字列が存在を表し、かつ、当該部分文字列はSPEAKER\_IDが「1」の話者データ（つまり、名前（NAME）が「私」の話者）のボイスにより読み上げられることを表す。

[0085] 同様に、図7に示す部分文字列テーブルの6行目の部分文字列データには、データ項目「TEXT」に『「いいえ」』、データ項目「POSITION」に「1」、データ項目「SPEAKER\_ID」に「2」がそれぞれ設定されている。これは、部分文字列『「いいえ」』は、コンテンツ中の先頭から当該部分文字列までに同様の部分文字列が1回出現することを表し、かつ、当該部分文字列はSPEAKER\_IDが「2」の話者データ（つまり、名前（NAME）が「先生」の話者）のボイスにより読み上げられることを表す。

[0086] 部分文字列データに対してデータ項目「POSITION」を設けることで、音声出力アプリケーション210がコンテンツ中の部分文字列を読み上げる際に、コンテンツの先頭からの出現回数も用いて、ラベルが付与された部分文字

列を検索することが可能となる。また、例えば、Webページ（コンテンツ）が更新された場合であっても、当該部分文字列の先頭からの出現位置が変わらなければ、Webページの更新前に付与されたラベルを利用可能である。

[0087] ここで、コンテンツ中の部分文字列で、部分文字列テーブルに格納されていない部分文字列は、SPEAKER\_IDが「0」の話者データ（つまり、データ項目「NAME」に「default」が設定されている話者データ）のボイスにより読み上げられる。

[0088] 以上により、図7に示す構成により、ラベルデータは、話者データと部分文字列テーブルとの組、又は、話者データのみで表される。例えば、コンテンツ中の発話文（つまり、括弧内の文章）を表す部分文字列や一人称視点の文章を表す部分文字列に付与されたラベルのラベルデータは、話者データと部分文字列データとの組で表される。一方で、コンテンツ中の三人称視点の文章を表す部分文字列に付与されたラベルのラベルデータは、データ項目「SPEAKER\_ID」に「0」が設定された話者データのみで表される。

[0089] なお、図7に示すラベルデータの構成は一例であって、他の構成であってもよい。例えば、Webページ（コンテンツ）のソースファイルをコピーし、コピーしたソースファイルにラベルを埋め込んでDBで保持する方法も考えられる。しかしながら、この場合、Webページが更新された場合には、Webページの更新前後におけるラベルと部分文字列との対応付けが難しいことがあるため、上記の図7に示す構成の方が好ましい。

[0090] <ラベル付与処理>

以降では、ラベリング端末10を用いて、ラベラーがコンテンツ中の部分文字列にラベルを付与する場合の処理（ラベル付与処理）の流れについて、図8を参照しながら説明する。図8は、本発明の実施の形態におけるラベル付与処理の一例を示すフローチャートである。

[0091] まず、ラベリング端末10のWebブラウザ110及びウィンドウ出力部121は、ラベリング画面を表示する（ステップS101）。すなわち、ラベリング端末10は、Webブラウザ110によりコンテンツを取得して画面上に表

示すると共に、ウインドウ出力部 121 によりラベリングウインドウを当該画面上に表示することで、ラベリング画面を表示する。

[0092] 次に、ラベリング端末 10 のコンテンツ解析部 122 は、Web ブラウザ 110 により表示されているコンテンツの構造を解析する（ステップ S102）。

[0093] 次に、ラベリング端末 10 のラベル操作管理部 123 は、ラベラーによるラベリング操作を受け付ける（ステップ S103）。ラベリング操作とは、ラベリングウインドウ中のリストから話者をラジオボタンで選択した上で、コンテンツ中の部分文字列をマウスで選択する操作のことである。これにより、当該部分文字列に対してラベルが付与され、例えば、当該部分文字列が話者固有の色でマーキングされる等のラベリング状況の可視化が行われる。

[0094] 最後に、ラベリング端末 10 のラベルデータ送受信部 124 は、例えば、ラベリングウインドウで SAVE ボタンが押下された場合、現在のコンテンツ中の部分文字列に付与されたラベルのラベルデータをラベル管理サーバ 30 に送信する（ステップ S104）。なお、上述したように、このとき、ラベルデータ送受信部 124 は、ラベリングしたコンテンツの URL もラベル管理サーバ 30 に送信する。

[0095] 以上により、ラベラーによってコンテンツ中の部分文字列にラベルが付与され、このラベルのラベルデータがラベル管理サーバ 30 に送信される。

[0096] <ラベルデータ保存処理>

以降では、ラベリング端末 10 から送信されたラベリングデータをラベル管理サーバ 30 が保存する処理（ラベルデータ保存処理）の流れについて、図 9 を参照しながら説明する。図 9 は、本発明の実施の形態におけるラベルデータ保存処理の一例を示すフローチャートである。

[0097] まず、ラベル管理サーバ 30 のラベルデータ送受信部 311 は、ラベリング端末 10 からのラベルデータを受信する（ステップ S201）。

[0098] 次に、ラベル管理サーバ 30 のラベルデータ管理部 312 は、上記のステップ S201 で受信したラベルデータを検証する（ステップ S202）。

[0099] 次に、ラベル管理サーバ30のDB管理部313は、上記のステップS202で検証が成功した場合、当該ラベルデータをラベル管理DB320に保存する（ステップS203）。

[0100] 以上により、ラベラーによってコンテンツ中の部分文字列に付与されたラベルのラベルデータがラベル管理サーバ30に保存される。

[0101] <音声出力処理>

以降では、音声出力端末20でコンテンツ中の部分文字列を、当該部分文字列に付与されたラベルに応じたボイスで読み上げる場合の処理（音声出力処理）の流れについて、図10を参照しながら説明する。図10は、本発明の実施の形態における音声出力処理の一例を示すフローチャートである。

[0102] まず、音声出力端末20のコンテンツ取得部211は、Webサーバ40からコンテンツを取得する（ステップS301）。

[0103] 次に、音声出力端末20のコンテンツ出力部214は、上記のステップS301で取得したコンテンツを表示する（ステップS302）。

[0104] 次に、音声出力端末20のラベルデータ取得部212は、上記のステップS301で取得したコンテンツのURLに対応するラベルデータをラベル管理サーバ30から取得する（ステップS303）。

[0105] 次に、音声出力端末20のコンテンツ解析部213は、上記のステップS301で取得したコンテンツを解析する（ステップS304）。上述したように、この解析では、コンテンツに含まれるテキストのどの部分文字列に対してどのラベルデータが付与されているかが特定される。

[0106] 次に、音声出力端末20の音声管理部215は、上記のステップS304での解析結果に基づいて、コンテンツ中の各部分文字列に対して、当該部分文字列の読み上げに用いるボイスデータをボイスデータ記憶部220の中から特定する（ステップS305）。すなわち、上述したように、音声管理部215は、各部分文字列にそれぞれ付与されているラベルが表す属性を用いて、ボイスデータ記憶部220に記憶されているボイスデータの中から属性が最も近いボイスデータを検索し、検索されたボイスデータを、当該部分文

字列を読み上げるボイスデータとして特定する。このとき、話者の識別情報（例えば、SPEAKER\_ID）が同一のラベルデータが付与されている部分文字列に対しては、同一のボイスデータを特定する。これにより、コンテンツ中の各部分文字列に対して、一貫性があるボイスが割り当てられる。

[0107] 最後に、音声出力端末20の音声出力部216は、上記のステップS305で割り当てられたボイス（を用いた合成音声）で各部分文字列を読み上げて音声を出力（ステップS306）。

[0108] 以上により、コンテンツ中の各部分文字列が、当該部分文字列に付与されたラベルに応じたボイスで読み上げられる。

[0109] <音声出力システム1のハードウェア構成>

次に、本発明の実施の形態における音声出力システム1に含まれるラベリング端末10、音声出力端末20、ラベル管理サーバ30及びWebサーバ40のハードウェア構成について説明する。これらの各端末及び各サーバは、1台以上のコンピュータ500を用いて実現可能である。図11は、コンピュータ500のハードウェア構成の一例を示す図である。

[0110] 図11に示すコンピュータ500は、ハードウェアとして、入力装置501と、表示装置502と、外部I/F503と、RAM(Random Access Memory)504と、ROM(Read Only Memory)505と、プロセッサ506と、通信I/F507と、補助記憶装置508とを有する。これら各ハードウェアは、それぞれがバスBを介して通信可能に接続されている。

[0111] 入力装置501は、例えばキーボードやマウス、タッチパネル等である。表示装置502は、例えばディスプレイ等である。なお、ラベル管理サーバ30やWebサーバ40は、入力装置501及び表示装置502の少なくとも一方を有していなくてもよい。

[0112] 外部I/F503は、外部装置とのインタフェースである。外部装置には、記録媒体503a等がある。コンピュータ500は、外部I/F503を介して、記録媒体503aの読み取りや書き込み等を行うことができる。

[0113] RAM504は、プログラムやデータを一時保持する揮発性の半導体メモ

りである。ROM 505は、電源を切ってもプログラムやデータを保持することができる不揮発性の半導体メモリである。ROM 505には、例えば、OSに関する設定情報や通信ネットワークNに関する設定情報等が格納されている。

[0114] プロセッサ506は、例えばCPU (Central Processing Unit) 等である。通信I/F 507は、コンピュータ500を通信ネットワークNに接続するためのインタフェースである。

[0115] 補助記憶装置508は、例えばHDD (Hard Disk Drive) やSSD (Solid State Drive) 等であり、プログラムやデータを格納している不揮発性の記憶装置である。補助記憶装置508に格納されているプログラムやデータには、例えば、OS、当該OS上で各種機能を実現するアプリケーションプログラム等がある。

[0116] なお、本発明の実施の形態における音声出力端末20は、上記の各ハードウェアに加えて、音声を外部に出力するためのハードウェア（例えば、イヤホン等を接続するためのI/Fやスピーカ等）を有する。

[0117] 本発明の実施の形態におけるラベリング端末10、音声出力端末20、ラベル管理サーバ30及びWebサーバ40は、図11に示すコンピュータ500により実現される。なお、上述したように、本発明の実施の形態におけるラベリング端末10、音声出力端末20、ラベル管理サーバ30及びWebサーバ40は、複数台のコンピュータ500で実現されていてもよい。また、1台のコンピュータ500には、複数のプロセッサ506や複数のメモリ（RAM 504やROM 505、補助記憶装置508等）が含まれていてもよい。

[0118] <まとめ>

以上のように、本発明の実施の形態における音声出力システム1では、ヒューマンコンピュータ技術によりコンテンツ中の部分文字列にラベルを付与した上で、部分文字列に付与されたラベルに応じてボイスを切り替えながら合成音声を出力することができる。これにより、本発明の実施の形態における音声出力システム1では、イメージした声に近いボイスで、コンテ

ンツ中の各部分文字列を音声出力することが可能となる。

- [0119] なお、本発明の実施の形態では、ラベラーと音声出力端末20のユーザとは、同一人物であることを要しない。つまり、コンテンツ中の部分文字列に付与されたラベルのラベルデータの利用者はラベラーに限定されない。また、ラベル管理サーバ30で管理されているラベルデータは複数のラベラーの間で共有可能であってもよい。このとき、例えば、ラベル管理サーバ30等により、ラベリングを行ったラベラーのランキングや多く利用されたラベルデータのランキング等が提供されてもよい。これにより、ラベリングに対するラベラーのモチベーション維持に貢献することが可能となる。
- [0120] また、例えば、Webページ等のコンテンツでは、同一のコンテンツが複数のWebページに分割されて提供される場合もある。このような場合では、Webページ間でボイスの割り当てが一貫されていることが好ましい。すなわち、或る小説が複数のWebページに分割されている場合には、Webページが異なっても、同一人物の発話文は同一のボイスで読み上げられることが好ましい。したがって、このような場合には、例えば、図7に示す話者データのデータ項目「URL」に複数のWebページのURLを設定可能とすることが考えられる。また、このとき、音声出力端末20上では、話者の識別情報が同一のラベルデータが付与されている部分文字列を読み上げるボイスデータを、当該識別情報と対応付けて保持しておく必要がある。
- [0121] また、本発明の実施の形態では、年齢や性別等の属性に応じたボイスで部分文字列を読み上げる場合について説明したが、これら年齢や性別以外にも、コンテンツ中の発話文のイメージと合成音声とのギャップを生じさせる様々な属性が存在する。
- [0122] 例えば、小説中で落ち着いたイメージの人の発話文が快活な声質のボイスで再生されたり、悲しげな場面での発話文が嬉しそうな声質のボイスで再生されたり、といった場合がある。また、小説等では、作中の展開によって登場人物の子どもから大人に成長したり、回想場面では逆に大人が子ともに場面点検したりといったことが起こり得る。このため、年齢や性別以外にも、

例えば、様々な属性（例えば、場面の状況や登場人物の性格等）も表すラベルを部分文字列に付与し、このラベルデータに応じたボイスで音声出力してもよい。また、このラベルに応じて、ボイスの設定（例えば、発話速度（SpeechRate）やピッチ（Pitch）等）が変更されてもよい。

[0123] 本発明は、具体的に開示された上記の実施形態に限定されるものではなく、特許請求の範囲から逸脱することなく、種々の変形や変更が可能である。

### 符号の説明

|        |       |              |
|--------|-------|--------------|
| [0124] | 1     | 音声出力システム     |
|        | 1 0   | ラベリング端末      |
|        | 2 0   | 音声出力端末       |
|        | 3 0   | ラベル管理サーバ     |
|        | 4 0   | Webサーバ       |
|        | 1 1 0 | Webブラウザ      |
|        | 1 2 0 | アドオン         |
|        | 1 2 1 | ウインドウ出力部     |
|        | 1 2 2 | コンテンツ解析部     |
|        | 1 2 3 | ラベル操作管理部     |
|        | 1 2 4 | ラベルデータ送受信部   |
|        | 2 1 0 | 音声出力アプリケーション |
|        | 2 1 1 | コンテンツ取得部     |
|        | 2 1 2 | ラベルデータ取得部    |
|        | 2 1 3 | コンテンツ解析部     |
|        | 2 1 4 | コンテンツ出力部     |
|        | 2 1 5 | 音声管理部        |
|        | 2 1 6 | 音声出力部        |
|        | 2 2 0 | ボイスデータ記憶部    |
|        | 3 1 0 | ラベル管理プログラム   |
|        | 3 1 1 | ラベルデータ送受信部   |

- 3 1 2 ラベルデータ管理部
- 3 1 3 DB管理部
- 3 1 4 ラベルデータ提供部
- 3 2 0 ラベル管理DB

## 請求の範囲

### [請求項1]

第1の端末と、サーバと、第2の端末とが含まれる音声出力システムが実行する音声出力方法であって、

前記第1の端末は、

コンテンツに含まれる文字列に対して、該文字列を合成音声で読み上げる場合における話者の属性を表すラベルデータを付与する第1のラベル付与手順と、

前記ラベルデータを前記サーバに送信する送信手順と、を実行し、  
前記サーバは、

前記第1の端末から送信されたラベルデータを、前記コンテンツを識別するコンテンツ識別情報と対応付けてデータベースに保存する保存手順を実行し、

前記第2の端末は、

前記コンテンツのコンテンツ識別情報に対応するラベルデータを前記サーバから取得する取得手順と、

取得したラベルデータを、前記コンテンツに含まれる各文字列に付与する第2のラベル付与手順と、

前記コンテンツに含まれる各文字列にそれぞれ付与されているラベルデータを用いて、複数の音声データの中から、前記文字列の読み上げに用いられる合成音声の音声データをそれぞれ特定する特定手順と、

前記コンテンツに含まれる各文字列のそれぞれを、特定された音声データを用いた合成音声により読み上げて音声出力する音声出力手順と、を実行する、

ことを特徴とする音声出力方法。

### [請求項2]

前記ラベルデータには、前記話者を識別する話者識別情報が含まれ、

前記特定手順は、

同一の話者識別情報が含まれるラベルデータが付与されている文字列に対しては、同一の音声データを特定する、ことを特徴とする請求項 1 に記載の音声出力方法。

[請求項3] 前記保存手順は、

前記ラベルデータを、前記話者及び該話者の属性を表す話者データと、前記文字列を表す文字列データとで表して前記データベースに保存する、ことを特徴とする請求項 1 又は 2 に記載の音声出力方法。

[請求項4] 前記文字列データには、

前記コンテンツ中の先頭から前記文字列までにおける該文字列と同一の文字列の出現回数が含まれる、ことを特徴とする請求項 3 に記載の音声出力方法。

[請求項5] 前記第 1 のラベル付与手順は、

前記コンテンツに含まれる文字列のうちのユーザにより選択された文字列に対して、前記ユーザにより選択された話者の属性を表すラベルデータを付与する、ことを特徴とする請求項 1 乃至 4 の何れか一項に記載の音声出力方法。

[請求項6] 前記属性には、前記話者の性別及び年齢が少なくとも含まれる、ことを特徴とする請求項 1 乃至 5 の何れか一項に記載の音声出力方法。

[請求項7] 第 1 の端末と、サーバと、第 2 の端末とが含まれる音声出力システムであって、

前記第 1 の端末は、

コンテンツに含まれる文字列に対して、該文字列を合成音声で読み上げる場合における話者の属性を表すラベルデータを付与する第 1 のラベル付与手段と、

前記ラベルデータを前記サーバに送信する送信手段と、を有し、

前記サーバは、

前記第 1 の端末から送信されたラベルデータを、前記コンテンツを識別するコンテンツ識別情報と対応付けてデータベースに保存する保

存手段を有し、

前記第2の端末は、

前記コンテンツのコンテンツ識別情報に対応するラベルデータを前記サーバから取得する取得手段と、

取得したラベルデータを、前記コンテンツに含まれる各文字列に付与する第2のラベル付与手段と、

前記コンテンツに含まれる各文字列にそれぞれ付与されているラベルデータを用いて、複数の音声データの中から、前記文字列の読み上げに用いられる合成音声の音声データをそれぞれ特定する特定手段と、

前記コンテンツに含まれる各文字列のそれぞれを、特定された音声データを用いた合成音声により読み上げて音声出力する音声出力手段と、を有する、

ことを特徴とする音声出力システム。

[請求項8]      請求項1乃至6の何れか一項に記載の音声出力方法における各手順をコンピュータに実行させることを特徴とするプログラム。

[図1]

これからどこへ行くという目的のない私は、ただ先生の歩く方へ歩いて行った。先生はいつもより口数を利かなかった。それでも私はさほどの窮屈を感じなかったので、ぶらぶらいつしよに歩いて行った。

「すぐお宅へお帰りですか」―― 発話(私)

「ええ別に寄る所ありませんから」―― 発話(先生)

二人はまた黙って南の方へ坂を下りた。―― 第三者視点

「先生のお宅の墓地はあすこにあるんですか」と私がまた口を利き出した。―― 発話(私)―― 人称視点

「いいえ」―― 発話(先生)

「どなたのお墓があるんですか。―― ご親類のお墓ですか」―― 発話(私)

「いいえ」―― 発話(先生)

先生はこれ以外に何も答えなかった。私もその話はそれぎりにして切り上げた。すると一町ほど歩いた後で、先生が不意にそこへ戻って来た。

「あすこには私の友達の墓があるんです」―― 発話(先生)―― 人称視点

[図2]

これからどこへ行くという目的のない私は、ただ先生の歩く方へ歩いて行っただ。先生はいつもより口数を利かなかった。それでも私はさほどの窮屈を感じなかったので、ぶら

ぶらいつしよに歩いて行っただ。

「すぐお宅へお帰りですか」

「ええ別に寄る所ありませんから」

二人はまた黙って南の方へ坂を下りた。

「先生のお宅の墓地はあすこにあるんですか」と私がまた口を利き出した。

「いいえ」

「どなたのお墓があるんですか。——ご親類のお墓ですか」

「いいえ」

先生はこれ以外に何も答えなかった。私もその話はそれぎりにして切り上げた。すると一町ほど歩いた後で、先生が不意にそこへ戻って来た。

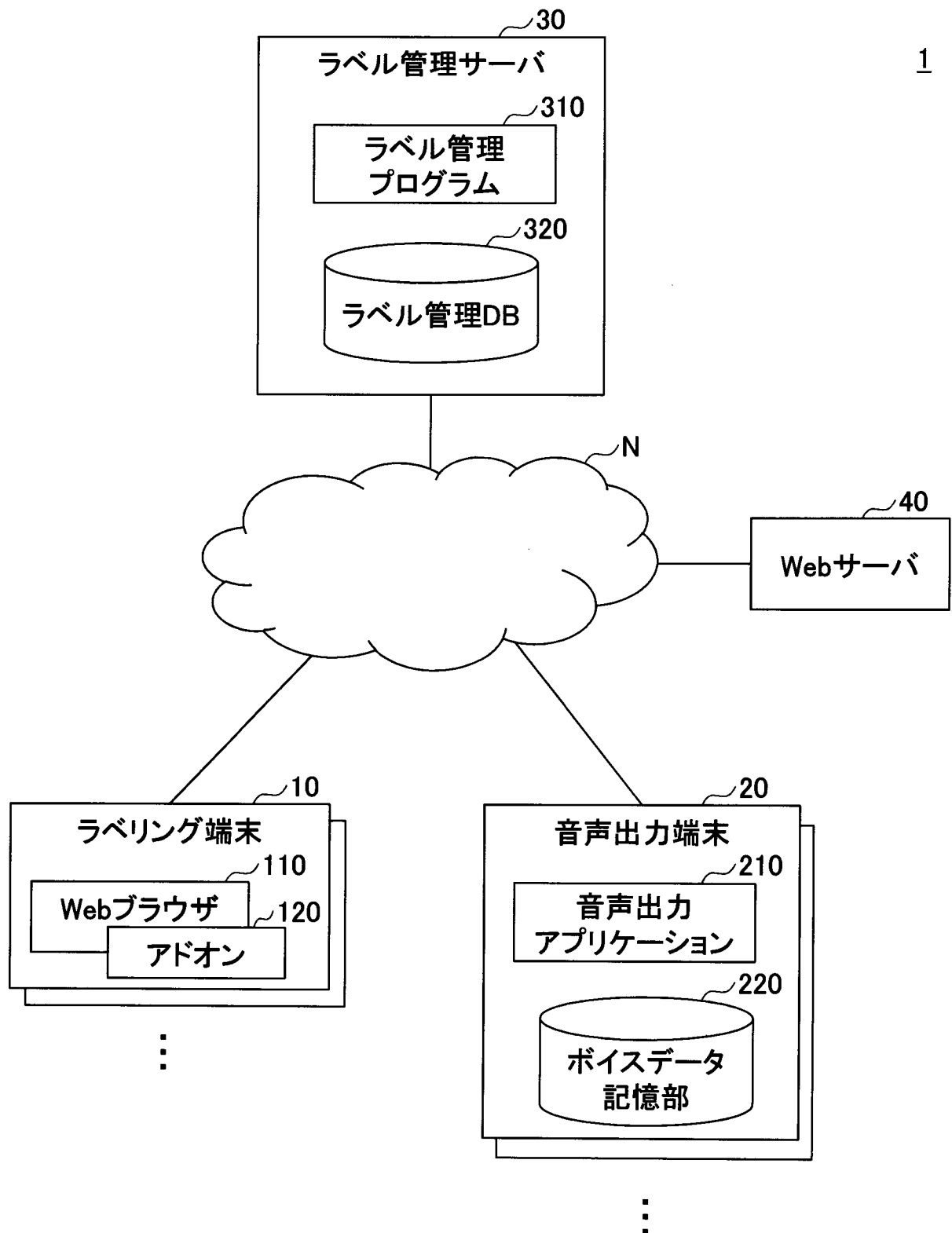
「あすこには私の友達の墓があるんです」

▨ : ボイス1(私)  
▤ : ボイス2(先生)  
▧ : ボイス3(ナレーション)

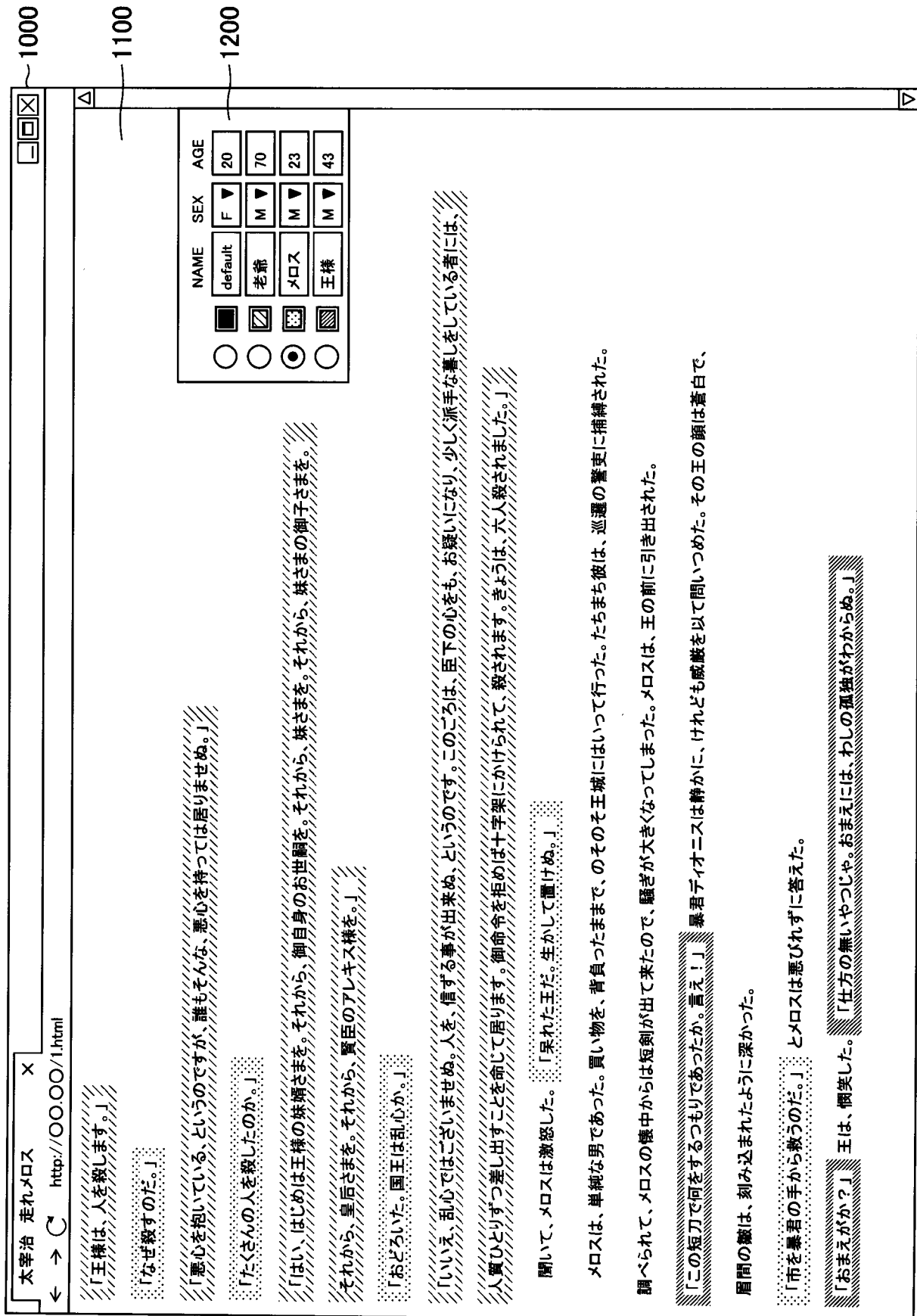
[図3]

<voice gender="female" age="30" id="0">  
<voice gender="male" age="20" id="1">これからどこへ行くという目的のない私は、ただ先生の歩く方へ歩いて行った。先生はいつもより口数を利かなかった。それでも私はさほどの窮屈を感じなかったので、ぶらぶらいっしょに歩いて行った。  
「すぐお宅へお帰りですか」</voice>  
<voice gender="male" age="50" id="2">「ええ別に寄る所ありませんから」</voice>  
二人はまた黙って南の方へ坂を下りた。  
<voice id="1">「先生のお宅の墓地はあすこにあるんですか」と私がまた口を利き出した。  
</voice>  
<voice id="2">「いいえ」</voice>  
<voice id="1">「どなたのお墓があるんですか。――ご親類のお墓ですか」</voice>  
<voice id="2">「いいえ」</voice>  
<voice id="1">先生はこれ以外に何も答えなかった。私もその話はそれぎりにして切り上げた。すると一町ほど歩いた後で、先生が不意にそこへ戻って来た。</voice>  
<voice id="2">「あすこには私の友達の墓があるんです」</voice>  
</voice>

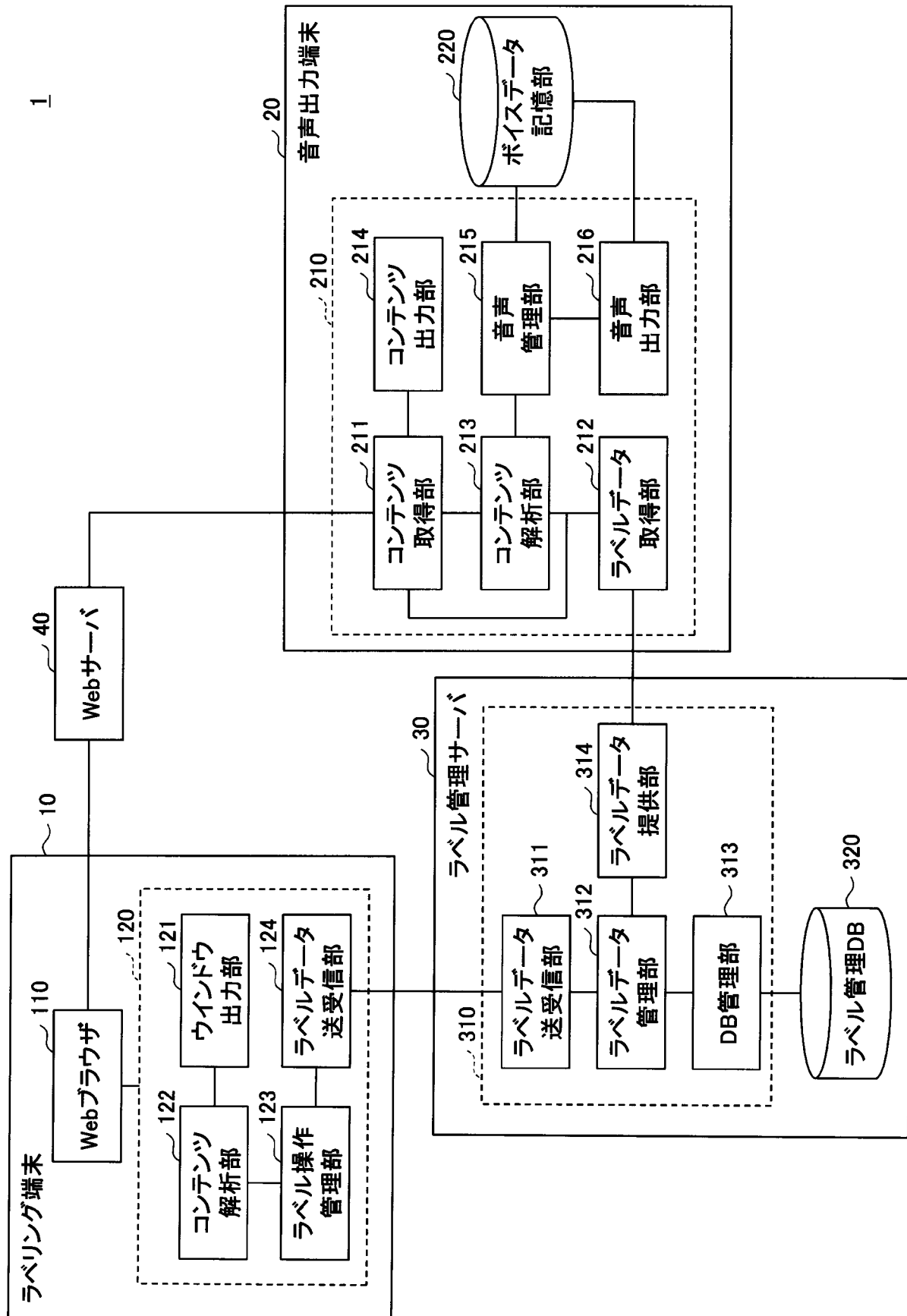
[図4]



[図5]



[図6]



[図7]

話者テーブル

| SPEAKER_ID | SEX | AGE | NAME    | COLOR | URL                 |
|------------|-----|-----|---------|-------|---------------------|
| 0          | F   | 30  | default | blue  | http://〇〇.〇〇/1.html |
| 1          | M   | 20  | 私       | green | http://〇〇.〇〇/1.html |
| 2          | M   | 50  | 先生      | red   | http://〇〇.〇〇/1.html |

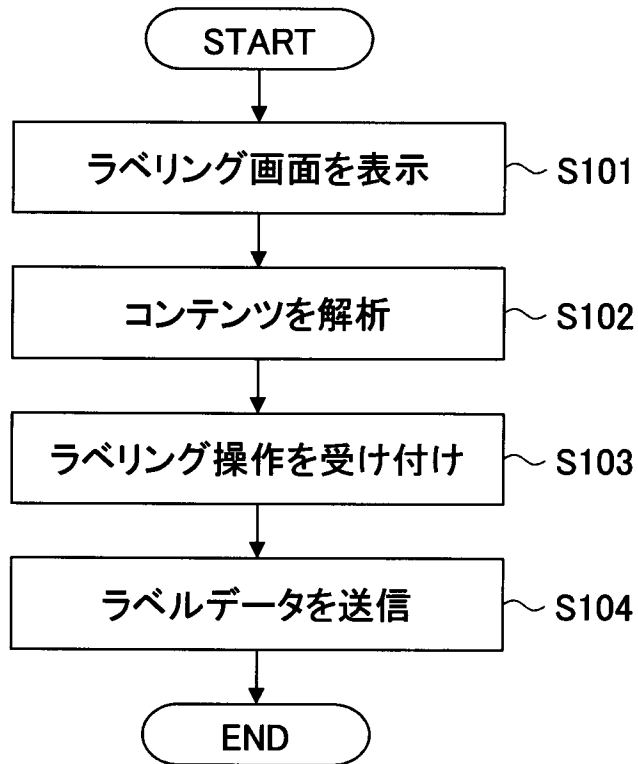
話者データ

部分文字列テーブル

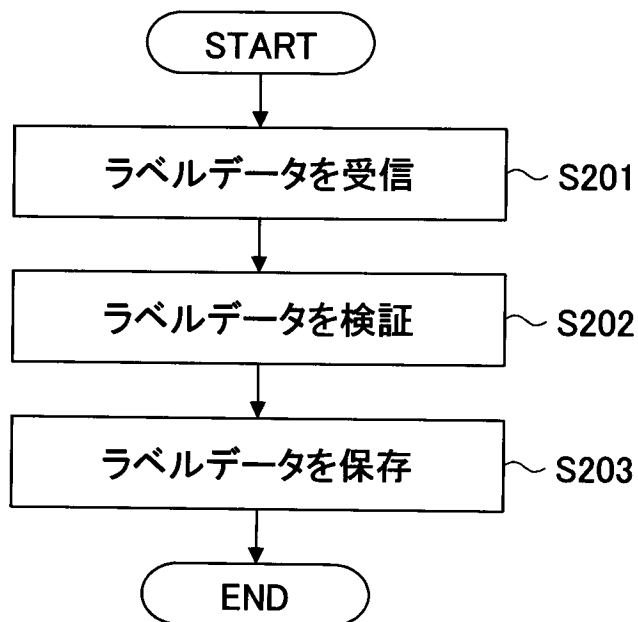
| TEXT                                                                                                                  | POSITION | SPEAKER_ID | URL                 |
|-----------------------------------------------------------------------------------------------------------------------|----------|------------|---------------------|
| これからどこへ行くという目的のない私は、ただ先生の歩く方へ歩いて行った。<br>先生はいつもより口数を利かなかった。それでも私はさほどの窮屈を感じな<br>かったので、ぶらぶらいいしよに歩いて行った。<br>「すぐお宅へお帰りですか」 | 0        | 1          | http://〇〇.〇〇/1.html |
| 「ええ別に寄る所ありませんから」                                                                                                      | 0        | 2          | http://〇〇.〇〇/1.html |
| 「先生のお宅の墓地はあすこにあるんですか」と私がまた口を利き出した。                                                                                    | 0        | 1          | http://〇〇.〇〇/1.html |
| 「いいえ」                                                                                                                 | 0        | 2          | http://〇〇.〇〇/1.html |
| 「どなたのお墓があるんですか。——ご親類のお墓ですか」                                                                                           | 0        | 1          | http://〇〇.〇〇/1.html |
| 「いいえ」                                                                                                                 | 1        | 2          | http://〇〇.〇〇/1.html |
| 先生はこれ以外に何も答えなかった。私もその話はそれぎりにして切り上げた。<br>すると一町ほど歩いた後で、先生が不意にそこへ戻って来た。                                                  | 0        | 1          | http://〇〇.〇〇/1.html |
| 「あすこには私の友達のお墓があるんです」                                                                                                  | 0        | 2          | http://〇〇.〇〇/1.html |

部分  
文字列  
データ

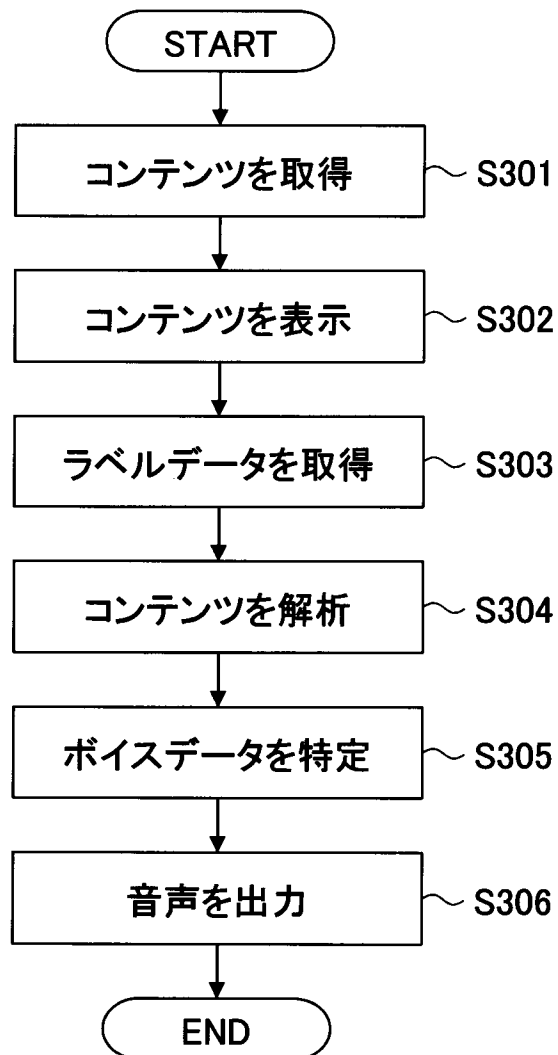
[図8]



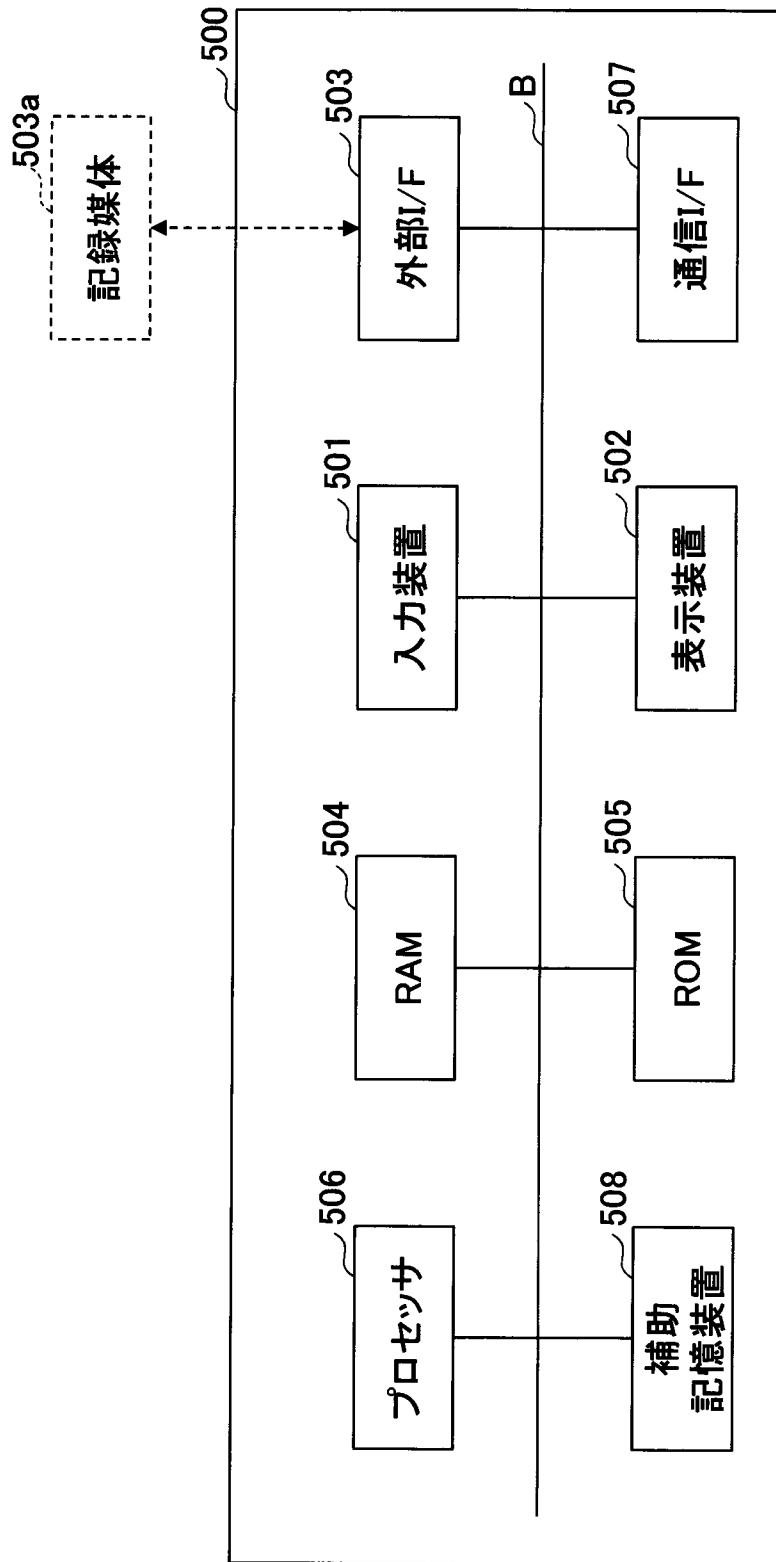
[図9]



[図10]



[図11]



# INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2020/010032

## A. CLASSIFICATION OF SUBJECT MATTER

G10L 13/00 (2006.01) i; G10L 13/10 (2013.01) i

Fl: G10L13/10 112B; G10L13/00 100B; G10L13/00 100K; G10L13/00 100R

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L13/00; G10L13/10

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

|                                                          |           |
|----------------------------------------------------------|-----------|
| Published examined utility model applications of Japan   | 1922-1996 |
| Published unexamined utility model applications of Japan | 1971-2020 |
| Registered utility model specifications of Japan         | 1996-2020 |
| Published registered utility model applications of Japan | 1994-2020 |

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|-----------|------------------------------------------------------------------------------------|-----------------------|
| A         | JP 8-272388 A (CANON INC.) 18.10.1996 (1996-10-18)<br>entire text, all drawings    | 1-8                   |

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

\* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search  
22 May 2020 (22.05.2020)

Date of mailing of the international search report  
02 June 2020 (02.06.2020)

Name and mailing address of the ISA/  
Japan Patent Office  
3-4-3, Kasumigaseki, Chiyoda-ku,  
Tokyo 100-8915, Japan

Authorized officer

Telephone No.

**INTERNATIONAL SEARCH REPORT**  
Information on patent family members

International application No.

PCT/JP2020/010032

Patent Documents  
referred in the  
Report

Publication  
Date

Patent Family

Publication  
Date

JP 8-272388 A

18 Oct. 1996

(Family: none)

## A. 発明の属する分野の分類（国際特許分類（IPC））

G10L 13/00(2006.01)i; G10L 13/10(2013.01)i

FI: G10L13/10 112B; G10L13/00 100B; G10L13/00 100K; G10L13/00 100R

## B. 調査を行った分野

調査を行った最小限資料（国際特許分類（IPC））

G10L13/00; G10L13/10

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報 1922-1996年

日本国公開実用新案公報 1971-2020年

日本国実用新案登録公報 1996-2020年

日本国登録実用新案公報 1994-2020年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

## C. 関連すると認められる文献

| 引用文献の<br>カテゴリー* | 引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示                      | 関連する<br>請求項の番号 |
|-----------------|--------------------------------------------------------|----------------|
| A               | JP 8-272388 A（キヤノン株式会社）18.10.1996（1996-10-18）<br>全文、全図 | 1-8            |

☐ C欄の続きにも文献が列挙されている。☒ パテントファミリーに関する別紙を参照。

\* 引用文献のカテゴリー

“A” 特に関連のある文献ではなく、一般的技術水準を示すもの

“E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの

“L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）

“O” 口頭による開示、使用、展示等に言及する文献

“P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献

“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの

“X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの

“Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの

“&amp;” 同一パテントファミリー文献

国際調査を完了した日

22.05.2020

国際調査報告の発送日

02.06.2020

名称及びあて先

日本国特許庁(ISA/JP)

〒100-8915

日本国

東京都千代田区霞が関三丁目4番3号

権限のある職員（特許庁審査官）

大野 弘 5Z 9175

電話番号 03-3581-1101 内線 3591

国際調査報告  
パテントファミリーに関する情報

国際出願番号  
PCT/JP2020/010032

| 引用文献          | 公表日        | パテントファミリー文献 | 公表日 |
|---------------|------------|-------------|-----|
| JP 8-272388 A | 18.10.1996 | (ファミリーなし)   |     |