(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第6865701号
(P6865701)
(45)発行日　令和3年4月28日(2021.4.28)　(24)登録日　令和3年4月8日(2021.4.8)

(51)Int.Cl.　F I
G10L 15/22 (2006.01)　G10L 15/22 470Z
G10L 19/00 (2013.01)　G10L 19/00 312F
G10L 19/00 312E
G10L 15/22 460Z

請求項の数 9　(全 17 頁)

(21)出願番号　特願2018-23711(P2018-23711)
(22)出願日　平成30年2月14日(2018.2.14)
(65)公開番号　特開2018-180519(P2018-180519A)
(43)公開日　平成30年11月15日(2018.11.15)
審査請求日　令和2年10月5日(2020.10.5)
(31)優先権主張番号　特願2017-82196(P2017-82196)
(32)優先日　平成29年4月18日(2017.4.18)
(33)優先権主張国・地域又は機関
日本国(JP)

早期審査対象出願

(73)特許権者　000004352
日本放送協会
東京都渋谷区神南２丁目２番１号
(73)特許権者　591053926
一般財団法人ＮＨＫエンジニアリングシステム
東京都世田谷区砧一丁目１０番１１号
(74)代理人　110001807
特許業務法人磯野国際特許商標事務所
(72)発明者　三島　剛
東京都世田谷区砧一丁目１０番１１号　日本放送協会放送技術研究所内
(72)発明者　佐藤　圧衛
東京都世田谷区砧一丁目１０番１１号　日本放送協会放送技術研究所内
最終頁に続く

(54)【発明の名称】音声認識誤り修正支援装置およびそのプログラム

(57)【特許請求の範囲】
【請求項１】
コンテンツに含まれる音声に対する音声認識の誤りを修正する音声認識誤り修正支援装置であって、
テキストデータである前記音声の認識結果と当該認識結果を構成する単語ごとの時間情報とにより、前記認識結果を予め定めた基準でセグメントに分割する認識結果分割手段と、
項目情報とともに前記セグメントに含まれる単語列を表示するか否かを指定するボタンを表示し、前記ボタンの選択により、編集領域を表示して前記セグメントの単語列を展開するか、前記編集領域を非表示とするかの制御を行う認識結果表示制御手段と、
前記編集領域で前記セグメントの誤りを修正する誤り修正手段と、
前記編集領域の前記セグメントに対応する音声を再生する音声再生手段と、を備え、
前記認識結果分割手段は、前記コンテンツに含まれる位置情報または時間情報の変化点で、前記認識結果を分割し、
前記誤り修正手段は、前記編集領域で指定された単語位置からの前記時間情報に対応する前記コンテンツの音声を前記音声再生手段により再生させることを特徴とする音声認識誤り修正支援装置。
【請求項２】
コンテンツに含まれる音声に対する音声認識の誤りを修正する音声認識誤り修正支援装置であって、

テキストデータである前記音声の認識結果と当該認識結果を構成する単語ごとの時間情報とにより、前記認識結果を予め定めた基準でセグメントに分割する認識結果分割手段と、
項目情報とともに前記セグメントに含まれる単語列を表示するか否かを指定するボタンを表示し、前記ボタンの選択により、編集領域を表示して前記セグメントの単語列を展開するか、前記編集領域を非表示とするかの制御を行う認識結果表示制御手段と、
前記編集領域で前記セグメントの誤りを修正する誤り修正手段と、
前記編集領域の前記セグメントに対応する音声を再生する音声再生手段と、を備え、
前記コンテンツは映像を含み、前記認識結果分割手段は、前記映像のカット点で、前記認識結果を分割し、
前記誤り修正手段は、前記編集領域で指定された単語位置からの前記時間情報に対応する前記コンテンツの音声を前記音声再生手段により再生させることを特徴とする音声認識誤り修正支援装置。
【請求項３】
コンテンツに含まれる音声に対する音声認識の誤りを修正する音声認識誤り修正支援装置であって、
テキストデータである前記音声の認識結果と当該認識結果を構成する単語ごとの時間情報とにより、前記認識結果を予め定めた基準でセグメントに分割する認識結果分割手段と、
項目情報とともに前記セグメントに含まれる単語列を表示するか否かを指定するボタンを表示し、前記ボタンの選択により、編集領域を表示して前記セグメントの単語列を展開するか、前記編集領域を非表示とするかの制御を行う認識結果表示制御手段と、
前記編集領域で前記セグメントの誤りを修正する誤り修正手段と、
前記編集領域の前記セグメントに対応する音声を再生する音声再生手段と、
複数の前記セグメントに含まれる単語から、前記セグメントごとに、ＴＦ－ＩＤＦ法により特徴単語を前記項目情報として抽出する項目情報抽出手段と、を備え、
前記認識結果表示制御手段は、前記セグメントに含まれる単語列を表示するか否かを指定するボタンを含んだ前記項目情報の一覧を表示し、
前記誤り修正手段は、前記編集領域で指定された単語位置からの前記時間情報に対応する前記コンテンツの音声を前記音声再生手段により再生させることを特徴とする音声認識誤り修正支援装置。
【請求項４】
コンテンツに含まれる音声に対する音声認識の誤りを修正する音声認識誤り修正支援装置であって、
テキストデータである前記音声の認識結果と当該認識結果を構成する単語ごとの時間情報とにより、前記認識結果を予め定めた基準でセグメントに分割する認識結果分割手段と、
項目情報とともに前記セグメントに含まれる単語列を表示するか否かを指定するボタンを表示し、前記ボタンの選択により、編集領域を表示して前記セグメントの単語列を展開するか、前記編集領域を非表示とするかの制御を行う認識結果表示制御手段と、
前記編集領域で前記セグメントの誤りを修正する誤り修正手段と、
前記編集領域の前記セグメントに対応する音声を再生する音声再生手段と、を備え、
前記誤り修正手段は、前記編集領域で指定された単語位置からの前記時間情報に対応する前記コンテンツの音声を前記音声再生手段により再生させ、前記コンテンツの音声再生中に前記編集領域の任意の単語位置を指定されることで、前記音声再生手段における音声の再生を停止させ、前記編集領域で指定された単語位置に音声の再生開始を示すポップアップメッセージを表示し、音声が停止した単語位置に音声の再生終了を示すポップアップメッセージを表示することを特徴とする音声認識誤り修正支援装置。
【請求項５】
前記誤り修正手段は、前記コンテンツの音声再生中に前記編集領域の任意の単語位置を

指定されることで、前記音声再生手段における音声の再生を停止することを特徴とする請求項１から請求項３のいずれか一項に記載の音声認識誤り修正支援装置。
【請求項６】
前記誤り修正手段は、前記コンテンツの音声再生に連動して、再生される音声に対応する前記編集領域の単語の表示属性を変更することを特徴とする請求項１から請求項５のいずれか一項に記載の音声認識誤り修正支援装置。
【請求項７】
前記誤り修正手段は、前記編集領域で指定された単語または指定区間の単語列の前記時間情報に対応する前記コンテンツの音声を、前記音声再生手段により、繰り返して再生することを特徴とする請求項１から請求項６のいずれか一項に記載の音声認識誤り修正支援装置。
【請求項８】
コンテンツに含まれる音声に対する音声認識の誤りを修正する音声認識誤り修正支援装置であって、
テキストデータである前記音声の認識結果と当該認識結果を構成する単語ごとの時間情報とにより、前記認識結果を発話内容の切り替わりごとのセグメントに分割する認識結果分割手段と、
項目情報とともに前記セグメントに含まれる単語列を表示するか否かを指定するボタンを表示し、前記ボタンの選択により、編集領域を表示して前記セグメントの単語列を展開するか、前記編集領域を非表示とするかの制御を行う認識結果表示制御手段と、
前記編集領域で前記セグメントの誤りを修正する誤り修正手段と、
前記編集領域の前記セグメントに対応する音声を再生する音声再生手段と、を備え、
前記誤り修正手段は、前記編集領域で指定された単語位置からの前記時間情報に対応する前記コンテンツの音声を前記音声再生手段により再生させることを特徴とする音声認識誤り修正支援装置。
【請求項９】
コンピュータを、請求項１から請求項８のいずれか一項に記載の音声認識誤り修正支援装置として機能させるための音声認識誤り修正支援プログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、音声認識の誤り修正を支援する音声認識誤り修正支援装置およびそのプログラムに関する。
【背景技術】
【０００２】
番組取材等で収録した音声素材（映像・音声素材を含む）の音声を文字として利用する場合、音声を文字に書き起こす作業が必須の作業となっている。通常、この作業は、作業者が収録した素材の音声を聴取し、端末のキーボード等で文字を入力することにより行っている。このとき、作業者は、音声の再生と停止を頻繁に繰り返したり、何度も同一箇所の音声を聞き直したりすることになるが、この作業は熟練者であっても素材の収録時間に対して約６倍の作業時間がかかるとも言われている。
【０００３】
従来、音声の書き起こし作業を支援する技術として、入力された音声を任意の単位に区切った文（セル）ごとに音声認識処理を施し、音声認識処理された認識結果と、これに対応する音声とを比較し、音声認識処理の誤りを修正する技術が開示されている（特許文献１参照）。
この技術は、音声認識処理においてセル単位で音声を再生し、操作者がセル単位で認識結果を修正し、セルの修正を一般的なテキストエディタの操作で行う。また、この技術では、操作者は、特殊な操作を覚える必要はなく、セルの修正後、セルの先頭から音声を再生して、操作者が認識結果を正しく修正したか否かを確認していた。

【０００４】
また、従来の音声の書き起こし作業を支援する技術として、音声の認識結果を、単語ごとに対応付けて、単語単位で修正する技術が開示されている（特許文献２，３参照）。
この技術は、字幕放送等のリアルタイム性が要求される誤り修正や、誤りの少ない認識結果を修正する場合には有効である。
【先行技術文献】
【特許文献】
【０００５】
【特許文献１】特開２０１５－１８４５６４号公報
【特許文献２】特開２００４－２２６９１０号公報
【特許文献３】特開２００５－２２８１７８号公報
【発明の概要】
【発明が解決しようとする課題】
【０００６】
特許文献１で開示されている技術は、セル単位で音声の再生および認識結果の修正を行うため、修正箇所が少なくても、修正箇所の音声と修正結果が合致するか否かを確認するために、セルの先頭から音声を再生する必要がある。
そのため、この技術は、セルの途中にある修正対象箇所の音声が再生されるまで、待ち時間が発生してしまうという問題があった。また、この技術は、セル内で、認識結果に対応する音声を操作者が聞き分ける必要があるため、認識結果が悪くなると、音声と修正対象とを対応付けることが困難になってしまうという問題があった。
【０００７】
また、特許文献２，３で開示されている技術のように、音声の認識結果を単語単位で修正する技術では、認識結果の修正と音声の確認とを素早く行うことは可能である。しかし、複数の単語に渡って認識誤りがある場合、順番に単語を指定して修正を行わなければならず、手順が複雑となり、その操作に慣れるまでに時間がかかってしまうという問題があった。
【０００８】
そこで、本発明は、音声認識の誤りを修正する際に、修正対象箇所の音声を素早く再生し、簡易な操作で音声認識の誤り修正を行うことが可能な音声認識誤り修正支援装置およびそのプログラムを提供することを課題とする。
【課題を解決するための手段】
【０００９】
前記課題を解決するため、本発明に係る音声認識誤り修正支援装置は、コンテンツの音声に対する音声認識の誤りを修正する音声認識誤り修正支援装置であって、認識結果分割手段と、認識結果表示制御手段と、誤り修正手段と、音声再生手段と、を備える構成とした。
【００１１】
かかる構成において、音声認識誤り修正支援装置は、認識結果分割手段によって、テキストデータである音声の認識結果と当該認識結果を構成する単語ごとの時間情報とにより、認識結果を予め定めた基準でセグメントに分割する。
【００１３】
そして、音声認識誤り修正支援装置は、認識結果表示制御手段によって、項目情報とともにセグメントに含まれる単語列を表示するか否かを指定するボタンを表示する。また、音声認識誤り修正支援装置は、認識結果表示制御手段によって、ボタンの選択により、編集領域を表示してセグメントの単語列を展開するか、編集領域を非表示とするかの制御を行う。これによって、認識結果表示制御手段は、音声の認識結果をすべて表示するのではなく、項目一覧によって操作者に編集対象のセグメントを指定させ、対象となったセグメントの単語列を編集領域に展開して操作者に提示する。
【００１４】

そして、音声認識誤り修正支援装置は、誤り修正手段によって、編集領域でセグメントの誤りを修正する。このとき、誤り修正手段は、編集領域で指定された単語位置からの時間情報に対応するコンテンツの音声を音声再生手段により再生させる。これによって、誤り修正手段は、認識結果またはその修正結果に対応する音声を操作者が素早く確認可能なように、指定された位置の単語から音声を再生する。

なお、音声認識誤り修正支援装置は、コンピュータを、前記した各手段として機能させるための音声認識誤り修正支援プログラムで動作させることができる。

【発明の効果】

【００１５】

本発明は、以下に示す優れた効果を奏するものである。

本発明によれば、素材コンテンツの音声認識結果を分割して、項目の一覧を表示するため、簡易な操作で音声認識の誤りを確認したい認識結果を素早く選択することができる。

また、本発明によれば、編集領域で単語の位置を指定するという簡易な操作で、対応する音声を再生するため、音声認識結果の誤りの発見や、修正確認を素早く行うことができる。

これによって、本発明は、特別なスキルを必要とせずに、音声認識結果の誤りを修正することができる。

【図面の簡単な説明】

【００１６】

【図１】本発明の実施形態に係る音声認識誤り修正支援装置の構成を示すブロック構成図である。

【図２】素材情報記憶手段が記憶する記憶内容を説明するための説明図である。

【図３】素材コンテンツを選択する素材コンテンツ選択画面の一例を示す画面構成図である。

【図４】素材コンテンツの音声認識結果を分割した項目の一覧を示す項目一覧画面の一例を示す画面構成図である。

【図５】項目一覧画面で編集領域に音声認識結果を展開した例を示す画面構成図である。

【図６】編集領域における編集作業の一例を説明するための説明図である。

【図７】音声再生に連動して編集領域の単語の表示属性を変更する例を説明するための説明図である。

【図８】編集領域における編集作業の操作内容を提示する例を説明するための説明図である。

【図９】編集領域における音声の繰り返し再生を指定する例を説明するための説明図である。

【図１０】本発明の実施形態に係る音声認識誤り修正支援装置の音声認識結果をセグメント単位で生成するセグメント情報生成動作を示すフローチャートである。

【図１１】本発明の実施形態に係る音声認識誤り修正支援装置の音声認識結果をセグメント単位で表示装置に提示するセグメント情報提示動作を示すフローチャートである。

【図１２】本発明の実施形態に係る音声認識誤り修正支援装置の音声再生を行いながら認識結果を修正するセグメント修正動作を示すフローチャートである。

【図１３】本発明の変形例の実施形態に係る音声認識誤り修正支援装置の構成を示すブロック構成図である。

【発明を実施するための形態】

【００１７】

以下、本発明の実施形態について図面を参照して説明する。

［音声認識誤り修正支援装置の構成］

最初に、図１を参照して、本発明の実施形態に係る音声認識誤り修正支援装置１の構成について説明する。

音声認識誤り修正支援装置１は、少なくとも音声を含んだ素材コンテンツにおける音声の認識誤りの修正を支援するものである。なお、本実施形態では、素材コンテンツは、映

像と音声とからなるコンテンツ、例えば、放送用素材とする。
【００１８】
音声認識誤り修正支援装置１は、図１に示すように、素材コンテンツ入力手段１０と、音声認識手段１１と、認識結果分割手段１２と、項目情報抽出手段１３と、素材情報記憶手段１４と、編集手段１５と、書き起こし結果出力手段１６と、を備える。
【００１９】
素材コンテンツ入力手段１０は、素材コンテンツを入力するものである。
素材コンテンツ入力手段１０は、例えば、外部の記憶媒体から素材コンテンツを入力するものであってもよいし、通信回線を介して入力するものであってもよい。
この素材コンテンツ入力手段１０は、入力した素材コンテンツのうち、音声については、音声認識手段１１に出力する。また、素材コンテンツ入力手段１０は、入力した素材コンテンツ（映像・音声）を、後記する編集手段１５における修正作業に使用するため、素材情報記憶手段１４に書き込み記憶する。
【００２０】
なお、素材コンテンツ入力手段１０は、素材情報記憶手段１４に素材コンテンツを書き込んだ後、音声認識手段１１に対して、素材コンテンツの書き込み完了を通知し、音声認識手段１１が素材情報記憶手段１４から音声を読み出すこととしてもよい。
【００２１】
音声認識手段１１は、素材コンテンツ入力手段１０が入力した素材コンテンツの音声を認識し、テキストデータである認識結果と当該認識結果を構成する単語ごとの時間情報とを生成するものである。
この音声認識手段１１は、図示を省略した言語モデル、音響モデル、発音辞書により、音声認識を行い、認識した単語と、その単語の音声の先頭からの経過時間を示す時間情報とを生成する。音声認識手段１１は、生成した認識結果の単語と時間情報とを認識結果分割手段１２に出力する。なお、音声認識手段１１における音声認識の手法は、例えば、特開２０１０－１７５７６５等に開示された音声から単語列を認識し、その結果を出力する手法を用いてもよい。
【００２２】
認識結果分割手段１２は、音声認識手段１１で認識された認識結果（単語列）を、予め定めた基準で分割するものである。以下、認識結果分割手段１２で生成された分割認識結果のそれぞれのかたまりをセグメントとよぶ。
認識結果分割手段１２が用いる分割の基準は、任意の基準を予め定めることができる。
例えば、分割の基準として、音声の無音区間を用いることができる。この場合、認識結果分割手段１２は、素材情報記憶手段１４に記憶されている音声から音響特徴量であるパワー等によって無音区間を検出し、音声認識手段１１による認識結果を、無音区間の前後で分割する。
【００２３】
また、例えば、分割の基準として、映像のカット点を用いることができる。この場合、認識結果分割手段１２は、素材情報記憶手段１４に記憶されている映像から、隣接するフレームの画像特徴が予め定めた基準よりも大きく異なるフレームをカット点として検出し、カット点の時間の前後で認識結果を分割する。
【００２４】
また、例えば、分割の基準として、素材コンテンツに予め付加されているメタ情報を用いてもよい。メタ情報としては、ＧＰＳ（Global Positioning System）の位置情報（ジオタグ）等がある。この場合、認識結果分割手段１２は、位置情報によって、素材コンテンツを撮影または集音した場所が異なっている時点で、認識結果を分割する。
【００２５】
認識結果分割手段１２は、音声認識結果を分割したセグメントを、項目情報抽出手段１３に出力する。また、認識結果分割手段１２は、セグメントごとに、単語とその時間情報とを素材情報記憶手段１４に書き込み記憶する。

【００２６】
項目情報抽出手段１３は、認識結果分割手段１２で分割されたセグメントごとに、当該セグメントに含まれる特徴単語を項目として抽出するものである。
この特徴単語は、セグメント内に含まれる特徴的な単語である。例えば、項目情報抽出手段１３は、ＴＦ－ＩＤＦ法（ＴＦ：Term Frequency、単語の出現頻度、ＩＤＦ:Inverse Document Frequency、逆文書頻度）によりセグメントを特徴付ける単語を抽出する。ＴＦ－ＩＤＦは、文書（本実施形態では、セグメント）中の単語に関する重みの一種であり、主に情報検索や文章要約などの分野で利用される。
具体的には、項目情報抽出手段１３は、セグメントｓ内の単語ｗの出現頻度ｔｆ（ｗ，ｓ）を、以下の式（１）で算出する。
【００２７】
【数１】

$$\mathbf{tf}(w,s)=\frac{n_{w,s}}{\sum_{t\in s} n_{t,s}} \qquad \cdots式（１）$$

【００２８】
この式（１）で、$n_{w,s}$は、ある単語ｗのセグメントｓ内での出現回数、$\Sigma_{t\in s} n_{t,s}$は、セグメントｓ内のすべての単語の出現回数の和を示す。
また、項目情報抽出手段１３は、ある単語ｗの逆文書頻度ｉｄｆ（ｗ）を、以下の式（２）で算出する。
【００２９】
【数２】

$$\mathbf{idf}(w)=\log\frac{N}{\mathbf{df}(w)}+1 \qquad \cdots式（２）$$

【００３０】
この式（２）で、Ｎは、素材コンテンツ内の全セグメント数、ｄｆ（ｗ）は、ある単語ｗが出現する素材コンテンツのセグメントの数（総セグメント数〔総文書数〕）を示す。
そして、項目情報抽出手段１３は、セグメント内の各単語について、以下の式（３）に示すように、式（１）のｔｆ値と式（２）のｉｄｆ値との積が最も大きい単語、あるいは、予め定めた基準値よりも大きい単語を、当該セグメントの特徴単語とする。
【００３１】
【数３】

$$\mathbf{tf}(w,s)\times\mathbf{idf}(w) \qquad \cdots式（３）$$

【００３２】
項目情報抽出手段１３は、抽出した項目を、セグメントに対応付けて素材情報記憶手段１４に書き込み記憶する。
なお、項目情報抽出手段１３は、ＴＦ－ＩＤＦ法を用いずに、セグメントを形態素解析し、名詞や固有名詞を特徴単語として抽出することとしてもよい。
【００３３】
また、項目情報抽出手段１３は、素材コンテンツが映像を含んでいる場合、特徴単語以外に、セグメントに対応する時間区間の映像からサムネイル画像を抽出してもよい。例えば、項目情報抽出手段１３は、セグメントに対応する時間区間の映像の先頭フレームをサムネイル画像として抽出する。項目情報抽出手段１３は、抽出したサムネイル画像を、セグメントに対応付けて素材情報記憶手段１４に書き込み記憶する。
【００３４】
素材情報記憶手段（記憶手段）１４は、音声認識の誤りを修正する対象となる素材コンテンツと、素材コンテンツをセグメントに分割した各種情報とを記憶するものである。こ

の素材情報記憶手段１４は、ハードディスク、半導体メモリ等の一般的な記憶媒体で構成することができる。

【００３５】

ここで、図２を参照（適宜図１参照）して、素材情報記憶手段１４が記憶する素材情報について具体的に説明する。

図２に示すように、素材情報記憶手段１４は、音声認識誤りを修正する対象となる素材コンテンツ（映像・音声）Ａ，Ｂ…を記憶する。この素材コンテンツ（映像・音声）Ａ，Ｂ…は、素材コンテンツ入力手段１０によって、記憶されたものである。

【００３６】

また、図２に示すように、素材情報記憶手段１４は、素材コンテンツごとに、音声認識結果をセグメントに分割した情報を記憶する。

図２の例では、素材コンテンツの識別情報（ここでは、ファイル名Ａ，Ｂ，…）ごとに、セグメント（識別情報ａ１，ａ２，…，ｂ１，…）を対応付けている。

各セグメントは、単語ｗと時間情報ｔとを複数含み、それぞれは対応付けられている。

このセグメントごとの単語ｗおよび時間情報ｔは、音声認識手段１１で対応付けられた単語および時間情報を、認識結果分割手段１２が分割した情報である。

【００３７】

また、各セグメントは、項目ｋとサムネイル画像ｇとを含む。項目ｋは、項目情報抽出手段１３が抽出した特徴単語である。サムネイル画像ｇは、項目情報抽出手段１３が当該セグメントの先頭の時間情報に対応した、素材コンテンツの映像から抽出したフレーム画像である。

なお、ここでは、素材コンテンツと、素材コンテンツの音声認識結果を分割したセグメントとを、同一の記憶手段に記憶しているが、別々の記憶手段に記憶することとしてもよい。

図１に戻って、音声認識誤り修正支援装置１の構成について説明を続ける。

【００３８】

編集手段１５は、外部に接続された修正端末（入力装置２、表示装置３、スピーカ４）を用いて、操作者が、素材情報記憶手段１４に記憶されている音声認識結果を修正するものである。なお、修正端末の表示装置３は、タッチパネルを備える構成としてもよい。

編集手段１５は、図１に示すように、素材コンテンツ選択手段１５０と、認識結果表示制御手段１５１と、誤り修正手段１５２と、映像／音声再生手段１５３と、を備える。

【００３９】

素材コンテンツ選択手段１５０は、修正対象となる素材コンテンツを選択するものである。例えば、素材コンテンツ選択手段１５０は、図３に示すように、素材情報記憶手段１４に記憶されている素材コンテンツＡ，Ｂ，Ｃのいずれかを選択するための選択ボタン３０１を含んだ素材コンテンツ選択画面３０を表示装置３に表示する。そして、素材コンテンツ選択手段１５０は、素材コンテンツ選択画面３０上の選択ボタン３０１の押下により、修正対象となる素材コンテンツを選択する。素材コンテンツ選択手段１５０は、選択された素材コンテンツのファイル名等の識別情報を、認識結果表示制御手段１５１に出力する。

【００４０】

認識結果表示制御手段１５１は、セグメントごとに、項目と当該セグメントに含まれる単語列を表示するか否かを指定する選択ボタンとを表示し、選択ボタンの押下により、セグメントの単語列を表示するか否かを制御するものである。

【００４１】

ここで、図４および図５を参照（適宜図１参照）して、認識結果表示制御手段１５１が表示する画面例について、その制御内容とともに説明する。

図４に示すように、認識結果表示制御手段１５１は、項目一覧画面３１を表示装置３の画面上に表示する。

項目一覧画面３１は、選択ボタン３１１と、項目表示欄３１２と、サムネイル画像表示

領域３１３と、タイムテーブル表示欄３１４と、スクロールバー表示欄３１５と、で構成される。

【００４２】

選択ボタン３１１は、セグメントごとに単語列を表示するか否かの選択を行うボタンである。

項目表示欄３１２は、セグメント内で抽出された項目を表示する領域である。認識結果表示制御手段１５１は、素材情報記憶手段１４から、当該セグメントに対応する項目（図２の項目ｋ）を読み出して、項目表示欄３１２に表示する。

サムネイル画像表示領域３１３は、セグメント内で抽出されたサムネイル画像を表示する領域である。認識結果表示制御手段１５１は、素材情報記憶手段１４から、当該セグメントに対応するサムネイル画像（図２のサムネイル画像ｇ）を読み出して、サムネイル画像表示領域３１３に表示する。

【００４３】

タイムテーブル表示欄３１４は、素材コンテンツの時間軸上におけるセグメント位置を示すタイムテーブルを表示する欄である。認識結果表示制御手段１５１は、素材情報記憶手段１４のセグメントの時間情報（図２の時間情報ｔ）を参照して、タイムテーブルを生成し表示する。

スクロールバー表示欄３１５は、項目一覧が画面に収まらない場合に、どの部分のセグメントを表示しているのかを示すスクロールバーを表示する欄である。認識結果表示制御手段１５１は、スクロールバーの上下によって、画面上の項目一覧を更新する。

このように、項目一覧画面３１を表示することで、操作者は、項目を確認することができ、一度に音声認識結果を表示する場合に比べて、音声認識結果を確認したいセグメントを容易に選択することができる。

【００４４】

この項目一覧画面３１において、操作者が行う入力装置２のマウスのクリック、あるいは、表示装置３のタッチパネルへのタッチによる選択ボタン（図４中、「ｏｐｅｎ」）３１１の押下により、認識結果表示制御手段１５１は、項目一覧画面３１において、セグメントの単語列の修正を行う編集領域３１６（図５参照）を表示する。

【００４５】

図５は、編集領域３１６を表示した項目一覧画面３１Ｂを示す画面例である。

この項目一覧画面３１Ｂは、図４で説明した項目一覧画面３１に対して、選択されたセグメントにおいて、動画表示領域３１３Ｂと、編集領域３１６とが表示される。

【００４６】

動画表示領域３１３Ｂは、セグメントに対応する素材コンテンツを再生する領域である。認識結果表示制御手段１５１は、当該セグメントが選択されたタイミングで、素材情報記憶手段１４のセグメントの時間情報（図２の時間情報ｔ）を参照して、対応する素材コンテンツの映像の先頭フレームを動画表示領域３１３Ｂに表示する。この動画表示領域３１３Ｂの画像領域をマウス等でクリック、あるいは再生開始ボタンｓｔを押下されることで、認識結果表示制御手段１５１は、映像／音声再生手段１５３に当該素材コンテンツの再生を指示する。

【００４７】

編集領域３１６は、セグメントに対応する単語列を表示し、編集対象となる領域である。認識結果表示制御手段１５１は、編集領域３１６に、素材情報記憶手段１４に記憶されている当該セグメントに対応する単語列（図２の単語ｗの列）を展開する。

なお、このとき、認識結果表示制御手段１５１は、選択ボタン３１１を、編集領域３１６を非表示とするボタン（図４中、「ｃｌｏｓｅ」）とする。そして、選択ボタン（図４中、「ｃｌｏｓｅ」）３１１の押下により、認識結果表示制御手段１５１は、編集領域３１６を非表示とし、動画表示領域３１３Ｂをサムネイル画像表示領域３１３として、図４の項目一覧画面３１に表示を戻す。

図１に戻って、音声認識誤り修正支援装置１の構成について説明を続ける。

【００４８】
誤り修正手段１５２は、操作者の編集操作により、編集領域３１６（図５）において、セグメントの単語列の誤りを修正するものである。この誤り修正手段１５２は、単語列を修正する編集動作においては、一般的なテキストエディタ（スクリーンエディタ）として機能する。ただし、誤り修正手段１５２は、単語列を修正する際に、音声を再生する機能を有する。
【００４９】
具体的には、誤り修正手段１５２は、編集領域３１６（図５）において、マウスのクリック、あるいは、タッチパネルへのタッチにより、選択された単語から音声を再生する。また、音声再生中、再度、任意の位置を選択されることで、誤り修正手段１５２は、音声の再生を停止する。
【００５０】
図６は、編集領域における編集作業の一例を説明するための説明図である。
例えば、図６の編集領域３１６において、「３月」が選択された場合、誤り修正手段１５２は、素材情報記憶手段１４のセグメントの時間情報（図２の時間情報ｔ）を参照して、対応する素材セグメントの位置から音声を再生するように、映像／音声再生手段１５３に指示する。なお、このとき、音声に連動して、動画表示領域３１３Ｂ（図５）において、音声再生の時間に対応する映像を再生することとしてもよい。
ここで、操作者が、誤り（ここでは、「ハタ寒い」）を発見して修正箇所をマウスでクリック等することで、誤り修正手段１５２は、音声再生を停止してカーソルＣを表示する。そして、誤り修正手段１５２は、操作者の編集操作により、誤りである「ハタ寒い」を「肌寒い」と修正する。そして、誤り修正手段１５２は、素材情報記憶手段１４に記憶されている誤りのあった単語を、修正後の単語に置き換える。これによって、音声認識誤り修正支援装置１は、操作者による修正後の保存操作を省略することができる。
【００５１】
また、誤り修正手段１５２は、マウスクリック等で指定された単語位置から音声を再生する。
図７は、音声再生に連動して編集領域の単語の表示属性を変更する例を説明するための説明図である。例えば、図７に示すように、編集領域３１６において、音声の再生を開始したい箇所をマウス等で選択された場合、誤り修正手段１５２は、素材情報記憶手段１４のセグメントの時間情報（図２の時間情報ｔ）を参照し、選択した単語から再生停止の指示があるまで音声を再生するように、映像／音声再生手段１５３に指示する。
そして、誤り修正手段１５２は、図７に示すように、音声の再生位置とセグメント中の再生有無とを明示するように、音声の再生に連動して、再生される音声に対応する各単語の表示部分の表示属性を変更する。例えば、誤り修正手段１５２は、音声に対応する単語を、白黒反転または予め定めた色でカラー表示する。
【００５２】
このとき、誤り修正手段１５２は、操作者が行った操作のフィードバック情報を画面上に提示する。例えば、図８に示すように、誤り修正手段１５２は、選択された単語位置に音声の再生開始を示すポップアップメッセージｐｏｐ１を表示し、音声が停止した単語位置に音声の再生終了を示すポップアップメッセージｐｏｐ２を表示する。これによって、操作者が不慣れであっても、自身の操作内容を把握することができ、安心して操作を行うことができる。
【００５３】
また、誤り修正手段１５２は、指定された単語または単語列に対応する音声を繰り返し再生することもできる。
例えば、図９に示すように、編集領域３１６において、音声を再生したい単語または単語列をマウス等で選択（図中、白黒反転領域）することで、誤り修正手段１５２は、ポップアップメニューｐｍを表示し、「繰り返し再生」を選択されることで、対応する単語または単語列の音声を繰り返し再生する。

図１に戻って、音声認識誤り修正支援装置１の構成について説明を続ける。

【００５４】

映像／音声再生手段１５３は、素材コンテンツの映像および音声を再生するものである。この映像／音声再生手段１５３は、認識結果表示制御手段１５１または誤り修正手段１５２から指定された位置から、素材コンテンツ（映像・音声）を再生する。

【００５５】

書き起こし結果出力手段１６は、編集手段１５で修正された音声認識結果（書き起こし結果）を、外部に出力するものである。

この書き起こし結果出力手段１６は、素材コンテンツのファイル名、または、素材コンテンツ内のセグメントの識別番号を指定されることで、素材情報記憶手段１４に記憶されている該当する素材コンテンツまたはセグメントの単語列を読み出して出力する。

【００５６】

以上説明したように音声認識誤り修正支援装置１を構成することで、音声認識誤り修正支援装置１は、簡易なテキスト編集操作で、認識結果の単語とその元となった音声とを確認しながら、音声認識の誤りを修正することができる。また、音声認識誤り修正支援装置１は、素材コンテンツに対して、セグメント単位で部分的に誤り修正を行うことができる。

なお、音声認識誤り修正支援装置１は、コンピュータを、前記した各手段として機能させるための音声認識誤り修正支援プログラムで動作させることができる。

【００５７】

［音声認識誤り修正支援装置の動作］

次に、図１０～図１２を参照して、本発明の実施形態に係る音声認識誤り修正支援装置１の動作について説明する。なお、ここでは、音声認識誤り修正支援装置１の動作として、素材コンテンツに対して音声認識による認識結果をセグメント単位で生成するセグメント情報生成動作と、認識結果をセグメント単位で表示装置３に提示するセグメント情報提示動作と、音声再生を行いながら認識結果を修正するセグメント修正動作と、について説明する。

【００５８】

（セグメント情報生成動作）

まず、図１０を参照（適宜図１参照）して、音声認識誤り修正支援装置１のセグメント情報生成動作について説明する。

ステップＳ１において、素材コンテンツ入力手段１０は、音声認識を行う素材コンテンツを入力する。このとき、素材コンテンツ入力手段１０は、入力した素材コンテンツを素材情報記憶手段１４に書き込み記憶する。

【００５９】

ステップＳ２において、音声認識手段１１は、ステップＳ１で入力した素材コンテンツの音声を認識し、テキストデータである認識結果と当該認識結果を構成する単語ごとの時間情報とを対応付けて生成する。

【００６０】

ステップＳ３において、認識結果分割手段１２は、ステップＳ２で認識された認識結果を、予め定めた基準、例えば、映像のカット点、音声の無音区間等によりセグメントに分割する。このとき、認識結果分割手段１２は、セグメント単位で、認識結果の単語と時間情報とを対応付けて、素材コンテンツを素材情報記憶手段１４に書き込み記憶する。

【００６１】

ステップＳ４において、項目情報抽出手段１３は、ステップＳ３で分割されたセグメントごとに、セグメントに含まれる特徴単語を項目として抽出するとともに、セグメントに対応する映像からサムネイル画像を抽出する。このとき、項目情報抽出手段１３は、抽出した項目およびサムネイル画像を、セグメントに対応付けて素材情報記憶手段１４に書き込み記憶する。

以上の動作によって、音声認識誤り修正支援装置１は、図２に示すように、素材情報記

憶手段１４に、素材コンテンツと、素材コンテンツをセグメントに分割した各種情報とを記憶する。

【００６２】

（セグメント情報提示動作）

次に、図１１を参照（適宜図１参照）して、音声認識誤り修正支援装置１のセグメント情報提示動作について説明する。

ステップＳ１０において、素材コンテンツ選択手段１５０は、素材情報記憶手段１４に記憶されている素材コンテンツのいずれかを選択するための選択ボタンを含んだ素材コンテンツ選択画面３０(図３参照)を表示装置３に表示する。

【００６３】

ステップＳ１１において、素材コンテンツ選択手段１５０は、画面上で選択ボタンが押下されるまで待機し（ステップＳ１１でＮｏ）、選択ボタンが押下された場合（ステップＳ１１でＹｅｓ）、ステップＳ１２以降の制御を行う認識結果表示制御手段１５１に制御を移す。

【００６４】

ステップＳ１２において、認識結果表示制御手段１５１は、素材情報記憶手段１４に記憶されている各種の情報に基づいて、セグメントごとに、項目と当該セグメントに含まれる単語列を表示するか否かを指定する選択ボタンとを含んだ項目一覧画面３１（図４参照）を表示装置３に表示する。

【００６５】

ステップＳ１３において、認識結果表示制御手段１５１は、項目一覧画面で選択ボタン（ｏｐｅｎ）が押下されるまで待機する（ステップＳ１３でＮｏ）。

一方、選択ボタン（ｏｐｅｎ）が押下された場合（ステップＳ１３でＹｅｓ）、ステップＳ１４において、認識結果表示制御手段１５１は、図５に示すように、選択されたセグメントに対応して編集領域３１６を表示し、素材情報記憶手段１４に記憶されている当該セグメントに対応する認識結果である単語列を編集領域３１６に展開する。

【００６６】

この動作以降、音声認識誤り修正支援装置１は、操作者が画面上で編集結果を修正可能な状態に移行する。なお、選択ボタン（ｏｐｅｎ）の押下により編集領域３１６を表示した場合、認識結果表示制御手段１５１は、任意のタイミングで、選択ボタン（ｃｌｏｓｅ）の押下により編集領域３１６を非表示とすることができるが、この非表示の動作については図示を省略した。また、項目一覧画面３１Ｂ（図５参照）の動画表示領域３１３Ｂにおける素材コンテンツの再生動作についてもここでは説明を省略する。

以上の動作によって、音声認識誤り修正支援装置１は、素材コンテンツをセグメント単位で、音声認識の誤りを修正することが可能になる。

【００６７】

（セグメント修正動作）

次に、図１２を参照（適宜図１参照）して、音声認識誤り修正支援装置１のセグメント修正動作について説明する。なお、セグメント修正動作は、操作者が行う任意の手順であるため、ここでは、音声再生と修正動作とを併せて行う動作の一例で説明する。

【００６８】

ステップＳ２０において、誤り修正手段１５２は、操作者のマウスのクリック、あるいは、タッチパネルへのタッチにより、編集領域３１６（図５）内の音声を再生したい単語または単語列を選択する。このとき、誤り修正手段１５２は、映像／音声再生手段１５３を介して、素材情報記憶手段１４のセグメントの時間情報を参照して、単語または単語列に対応する時間の音声を再生する。これによって、操作者は、音声と音声認識された単語列とを対比して確認することができる。

【００６９】

ステップＳ２１において、誤り修正手段１５２は、操作者のマウスのクリック、あるいは、タッチパネルへのタッチにより、修正箇所の位置の指定を受け付ける。このとき、誤

り修正手段１５２は、音声が単語列の末尾まで再生されていない、あるいは、繰り返し再生中で、音声が再生中であれば、音声の再生を停止する。

【００７０】

ステップＳ２２において、誤り修正手段１５２は、編集領域の指定された位置にカーソルを表示して、文字削除、文字挿入等の操作者の編集作業により、認識誤りを修正する。ここで、誤り修正手段１５２は、素材情報記憶手段１４の単語を修正結果で更新する。

【００７１】

ステップＳ２３において、誤り修正手段１５２は、操作者のマウスのクリック、あるいは、タッチパネルへのタッチにより、修正を行った箇所の位置の指定を受け付ける。このとき、誤り修正手段１５２は、映像／音声再生手段１５３を介して、素材情報記憶手段１４のセグメントの時間情報を参照して、単語または単語列に対応する時間の音声を再生する。これによって、操作者は、修正結果が正しいか否かを確認することができる。

【００７２】

なお、図示を省略しているが、ステップＳ２３における操作者の確認で、修正箇所がまだ正しく修正されていない場合、ステップＳ２１に戻って、動作を繰り返す。

以上の動作によって、音声認識誤り修正支援装置１は、音声認識の誤りを修正する際に、修正対象箇所の音声を素早く再生し、簡易な操作で音声認識の誤り修正することができる。

【００７３】

以上、本発明の実施形態について説明したが、本発明は、この実施形態に限定されるものではない。

ここでは、素材コンテンツを、映像および音声を含んだものとして説明したが、音声のみの素材コンテンツであっても構わない。

その場合、項目情報抽出手段１３は、項目のみを抽出し、サムネイル画像を抽出しないこととすればよい。また、映像／音声再生手段１５３は、音声のみを再生する音声再生手段とすればよい。

【００７４】

また、ここでは、音声認識誤り修正支援装置１に、直接、修正端末（入力装置２、表示装置３、スピーカ４）を接続する構成としたが、これらは、ネットワークを介して接続する形態であっても構わない。

【００７５】

また、音声認識誤り修正支援装置１は、修正端末を複数備える構成であっても構わない。その場合、認識結果表示制御手段１５１は、ある修正端末が修正を行っているセグメントについて、他の修正端末が修正対象として選択しないように排他制御し、例えば、他の修正端末において、選択ボタンを表示しないようにする。

【００７６】

また、音声認識誤り修正支援装置１の編集手段１５は、認識結果を修正するサーバとして、画面制御を行うユーザインタフェースを提供し、ネットワークを介して接続された複数の修正端末が、当該ユーザインタフェースを介して動作するクライアントとして機能させることとしてもよい。これによって、ネットワークを介して、複数の地点で、音声認識の誤りを修正することができる。

【００７７】

また、音声認識誤り修正支援装置１は、音声認識手段１１を外部に備えてもよい。

例えば、図１３に示す音声認識誤り修正支援装置１Ｂの構成としてもよい。音声認識誤り修正支援装置１Ｂは、音声認識誤り修正支援装置１（図１）の音声認識手段１１を音声認識装置として外部に備える。この場合、認識結果分割手段１２は、音声認識手段１１から出力される音声の認識結果と当該認識結果を構成する単語ごとの時間情報とを、入力インタフェースである認識結果入力手段１７を介して入力すればよい。

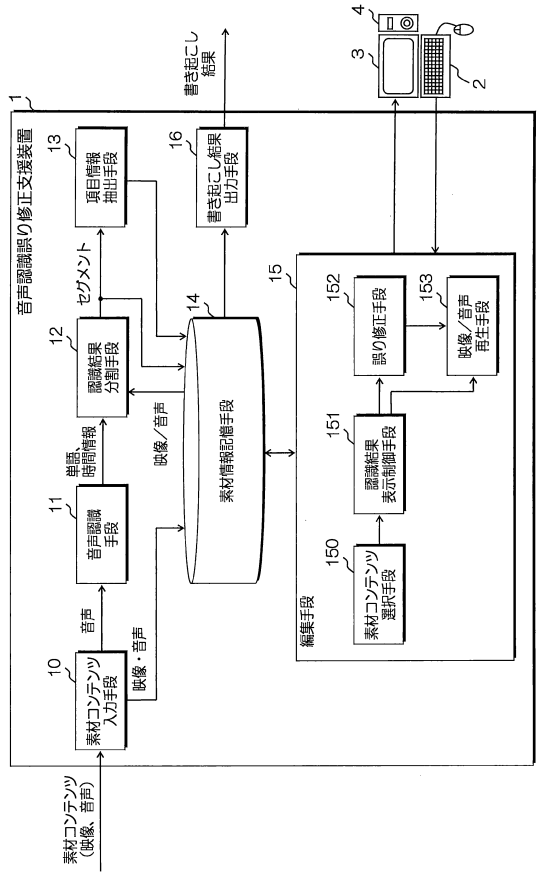
なお、音声認識誤り修正支援装置１Ｂも、コンピュータを、前記した各手段として機能させるための音声認識誤り修正支援プログラムで動作させることができる。

【符号の説明】

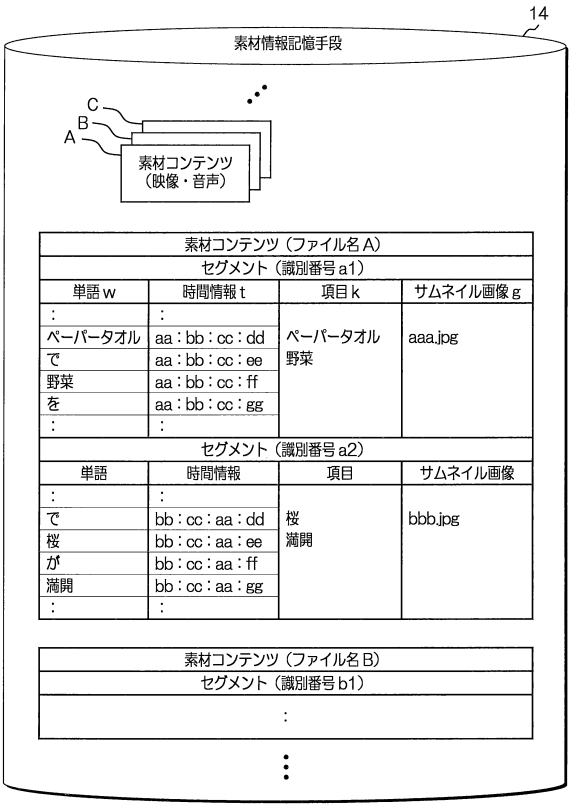
【００７８】

１，１Ｂ　音声認識誤り修正支援装置
１０　素材コンテンツ入力手段
１１　音声認識手段
１２　認識結果分割手段
１３　項目情報抽出手段
１４　素材情報記憶手段（記憶手段）
１５　編集手段
１５０　素材コンテンツ選択手段
１５１　認識結果表示制御手段
１５２　誤り修正手段
１５３　映像／音声再生手段（音声再生手段）
１６　書き起こし結果出力手段
１７　認識結果入力手段

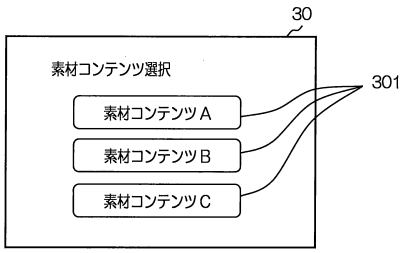
【図１】



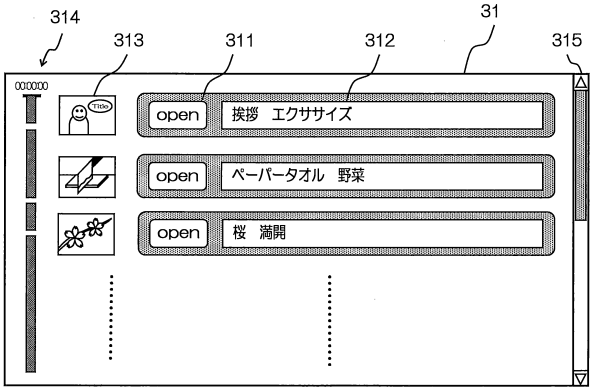
【図２】



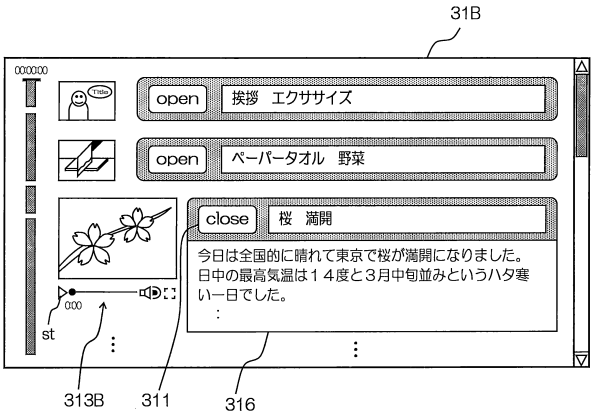
【図３】



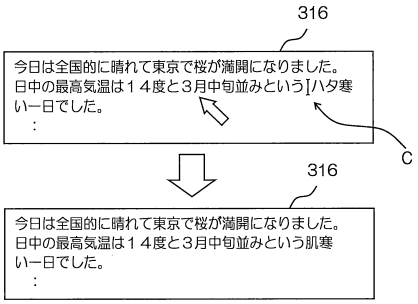
【図４】



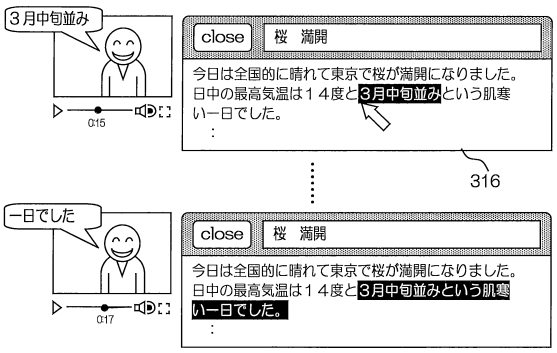
【図５】



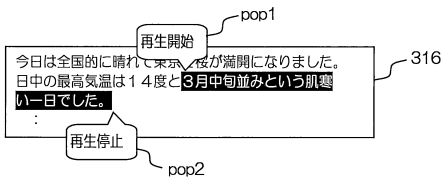
【図６】



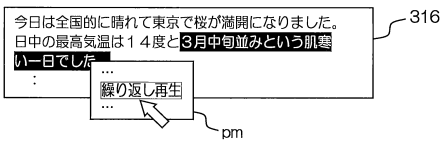
【図７】



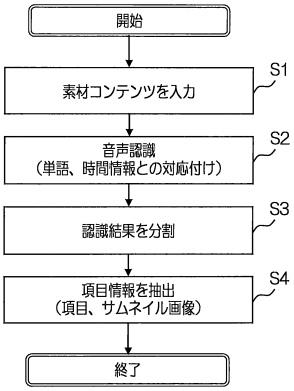
【図８】



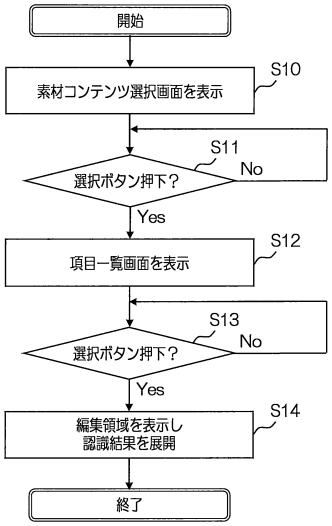
【図９】



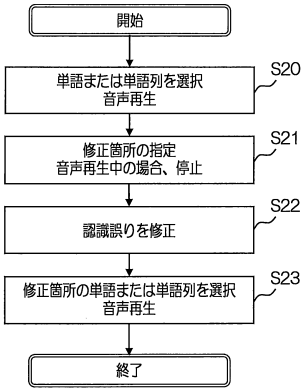
【図１０】



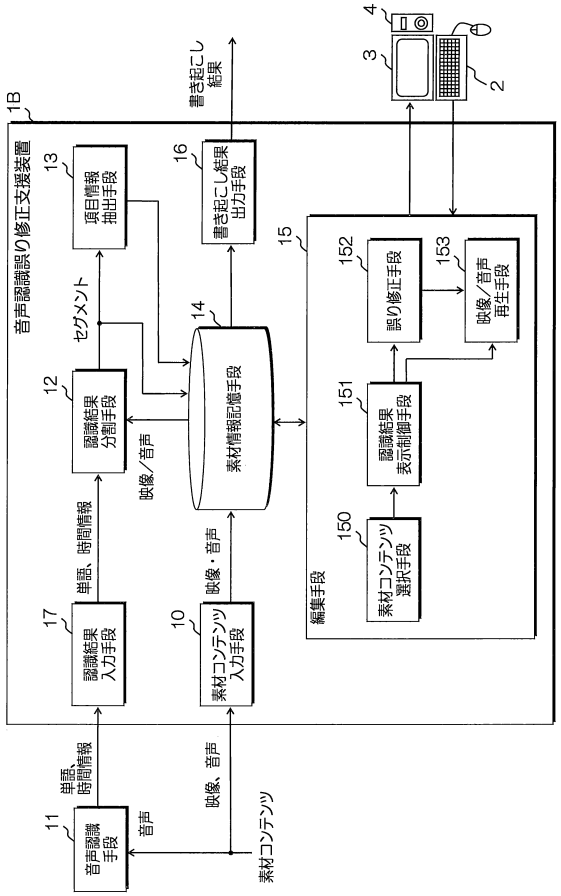
【図１１】


開始
素材コンテンツ選択画面を表示
S10
S11
No
選択ボタン押下？
Yes
項目一覧画面を表示
S12
S13
No
選択ボタン押下？
Yes
編集領域を表示し
認識結果を展開
S14
終了

【図１２】


開始
単語または単語列を選択
音声再生
S20
修正箇所の指定
音声再生中の場合、停止
S21
認識誤りを修正
S22
修正箇所の単語または単語列を選択
音声再生
S23
終了

【図１３】


音声認識誤り修正支援装置
1B
4
3
2
書き起こし結果
項目情報抽出手段
13
書き起こし結果出力手段
16
セグメント
素材情報記憶手段
14
15
誤り修正手段
152
映像／音声再生手段
153
認識結果分割手段
12
認識結果表示制御手段
151
映像／音声
単語、時間情報
映像・音声
素材コンテンツ選択手段
150
編集手段
認識結果入力手段
17
素材コンテンツ入力手段
10
単語、時間情報
映像、音声
音声認識手段
11
音声
素材コンテンツ

フロントページの続き

(72)発明者　一木　麻乃
東京都世田谷区砧一丁目１０番１１号　日本放送協会放送技術研究所内
(72)発明者　伊藤　均
東京都世田谷区砧一丁目１０番１１号　日本放送協会放送技術研究所内
(72)発明者　所澤　愛子
東京都世田谷区砧一丁目１０番１１号　日本放送協会放送技術研究所内
(72)発明者　小林　彰夫
東京都世田谷区砧一丁目１０番１１号　日本放送協会放送技術研究所内

審査官　渡部　幸和

(56)参考文献　特開２００６－３３０１７０（ＪＰ，Ａ）
特開２０１３－０２０４１１（ＪＰ，Ａ）
特開２０１４－０４４３６３（ＪＰ，Ａ）
米国特許出願公開第２０１５／００８８５０５（ＵＳ，Ａ１）
特開２００６－２０２３２１（ＪＰ，Ａ）
特開２００１－２８２２９１（ＪＰ，Ａ）
荒井　孝，”字幕放送　　ニュース字幕用音声認識システムの整備”，放送技術　　第６５巻第１２号，西村　瓊江　兼六館出版株式会社，２０１２年１１月２９日，第65巻，139～142

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ１０Ｌ　　１５／００－１９／２６