(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第3945778号
(P3945778)
(45)発行日　平成19年7月18日(2007.7.18)　(24)登録日　平成19年4月20日(2007.4.20)

(51)Int.Cl.		F I		
G10L 15/22	(2006.01)	G10L	15/22	460Z
G10L 15/00	(2006.01)	G10L	15/00	200G
G10L 15/10	(2006.01)	G10L	15/10	300G
G10L 15/08	(2006.01)	G10L	15/08	300Z

請求項の数 16　(全 27 頁)

(21)出願番号　特願2004-71229(P2004-71229)
(22)出願日　平成16年3月12日(2004.3.12)
(65)公開番号　特開2005-258198(P2005-258198A)
(43)公開日　平成17年9月22日(2005.9.22)
審査請求日　平成17年1月14日(2005.1.14)

(73)特許権者　390009531
インターナショナル・ビジネス・マシーンズ・コーポレーション
ＩＮＴＥＲＮＡＴＩＯＮＡＬ　ＢＵＳＩＮＥＳＳ　ＭＡＳＣＨＩＮＥＳ　ＣＯＲＰＯＲＡＴＩＯＮ
アメリカ合衆国１０５０４　ニューヨーク州　アーモンク　ニュー　オーチャード　ロード
(74)代理人　100086243
弁理士　坂口　博
(74)代理人　100091568
弁理士　市位　嘉宏
(74)代理人　100108501
弁理士　上野　剛史

最終頁に続く

(54)【発明の名称】設定装置、プログラム、記録媒体、及び設定方法

(57)【特許請求の範囲】

【請求項１】

内容が予め定められた音声の再生に同期して前記内容を表示する表示タイミングを設定する設定装置であって、

前記音声の内容を示す内容データを取得する内容データ取得部と、

再生される前記音声を音声認識した文字データを分割して複数の認識データを生成する音声認識部と、

前記複数の認識データの各々に一致する文字列を前記内容データから検出する文字列検出部と、

前記文字列検出部により一致する文字列が検出されなかった各認識データについて、当該認識データに含まれる各文字に一致する文字を前記内容データから検出することにより、当該認識データに一致する文字列を前記内容データから検出する文字検出部と、

前記認識データのうち前記文字検出部により一致する文字が検出されなかった各文字について、当該文字の読みに含まれる音素に一致する音素を、前記内容データの読みの中から検出する音素検出部とを備え、

前記文字検出部は、前記認識データのうち前記音素検出部により一致する音素が検出された文字に一致する文字として、前記内容データにおいて当該音素を含む文字を更に検出し、

前記内容データに含まれる文字列の各々を表示させる表示タイミングを、当該文字列に一致する認識データとして音声認識された音声の再生時に設定する表示設定部を更に備え

る設定装置。
【請求項２】
前記内容データの読み方の候補である読み候補を複数生成する読付部を更に備え、
前記音素検出部は、前記認識データのうち前記文字検出部により前記内容データに一致する文字が検出されなかった各文字について、当該文字の読みに含まれる音素に一致する音素を、前記読付部により生成された複数の前記読み候補の何れかの中から検出する
請求項１記載の設定装置。
【請求項３】
前記読付部は、前記内容データにおける複数の前記読み候補の各々を、当該読み候補により読まれる可能性を示す情報に対応付けて生成し、
前記音素検出部は、前記認識データに含まれる文字の読みに含まれる音素を、読まれる可能性が高い順に、前記複数の読み候補の各々と比較する
請求項２記載の設定装置。
【請求項４】
前記複数の認識データの各々が文字列に一致する確度である信頼度を算出する信頼度算出部を更に備え、
前記表示設定部は、前記内容データにおける複数の前記文字列のうち連続した２つの文字列について、先に表示すべき文字列に対応する信頼度が、後に表示すべき文字列に対応する信頼度より高い場合に、先に表示すべき前記文字列の末尾に後に表示すべき前記文字列を連結した文字列を、先に表示すべき前記文字列を表示すべき時刻に表示させる設定を行う
請求項１記載の設定装置。
【請求項５】
前記信頼度算出部は、前記文字列検出部により一致する文字列が検出された認識データに対応付けて、前記文字検出部により一致する文字列が検出された認識データと比較して、より高い信頼度を算出する
請求項４記載の設定装置。
【請求項６】
前記認識データのうち前記文字検出部により一致する文字が検出されなかった各文字について、当該文字の読みに含まれる音素に一致する音素を、前記内容データの読みの中から検出する音素検出部を更に備え、
前記文字検出部は、前記認識データのうち前記音素検出部により一致する音素が検出された文字に一致する文字として、前記内容データにおいて当該音素を含む文字を検出し、
前記信頼度算出部は、前記音素検出部により一致する音素が検出された文字を含む認識データに対応付けて、前記音素検出部により一致する音素が検出されることなく前記文字検出部により一致する文字が検出された認識データと比較して、より低い信頼度を生成する
請求項４記載の設定装置。
【請求項７】
内容が予め定められた音声の再生に同期して前記内容を表示する表示タイミングを設定する設定装置であって、
前記音声の内容を示す内容データを取得する内容データ取得部と、
再生される前記音声を音声認識した文字データを分割して複数の認識データを生成する音声認識部と、
前記複数の認識データの各々が文字列に一致する確度である信頼度を算出する信頼度算出部と、
前記複数の認識データの各々に一致する文字列を前記内容データから検出すると共に、予め定められた基準信頼度未満の信頼度の認識データである低信頼データについては、さらに、当該低信頼データに後続する認識データに一致する文字列を検出できなかった場合に、当該低信頼データに一致する文字列は検出できないと判断する文字列検出部と、

前記文字列検出部により一致する文字列が検出されなかった各認識データについて、当該認識データに含まれる各文字に一致する文字を前記内容データから検出することにより、当該認識データに一致する文字列を前記内容データから検出する文字検出部と、
前記内容データに含まれる文字列の各々を表示させる表示タイミングを、当該文字列に一致する認識データとして音声認識された音声の再生時に設定する表示設定部と
を備える設定装置。

【請求項８】
内容が予め定められた音声の再生に同期して前記内容を表示する表示タイミングを設定する設定装置であって、
前記音声の内容を示す内容データを取得する内容データ取得部と、
再生される前記音声を音声認識した文字データを分割して複数の認識データを生成すると共に、さらに、音声認識した前記複数の認識データが、再生される音声の内容と一致する可能性を示す音声認識確信度を、認識データ毎に生成する音声認識部と、
前記複数の認識データの各々について、音声認識確信度がより高い認識データに一致する文字列を、当該認識データと比較して音声認識確信度が低い認識データに先立って前記内容データから検出すると共に、第１の前記認識データに一致する第１の文字列及び第２の前記認識データに一致する第２の文字列を検出した場合に、前記第１の認識データに後続しかつ前記第２の認識データに先行する認識データに一致する文字列として、前記第１の文字列に後続し前記第２の文字列に先行する文字列を検出する文字列検出部と、
前記文字列検出部により一致する文字列が検出されなかった各認識データについて、当該認識データに含まれる各文字に一致する文字を前記内容データから検出することにより、当該認識データに一致する文字列を前記内容データから検出する文字検出部と、
前記内容データに含まれる文字列の各々を表示させる表示タイミングを、当該文字列に一致する認識データとして音声認識された音声の再生時に設定する表示設定部と
を備える設定装置。

【請求項９】
内容が予め定められた音声の再生に同期して前記内容を表示する表示タイミングを設定する設定装置であって、
再生される前記音声の内容を示す内容データに含まれる複数の文字列の各々に対応付けて、当該文字列を表示すべき時刻、及び、当該時刻に当該文字列を内容とする音声が再生される確度である信頼度を取得する信頼度取得部と、
前記複数の文字列のうち連続した２つの文字列について、先に表示すべき文字列に対応する信頼度が、後に表示すべき文字列に対応する信頼度より高い場合に、先に表示すべき前記文字列の末尾に後に表示すべき前記文字列を連結した文字列を、先に表示すべき前記文字列を表示すべき時刻に表示させる設定を行う表示設定部と
を備える設定装置。

【請求項１０】
前記表示設定部は、先に表示すべき前記文字列に対応する信頼度が、後に表示すべき前記文字列に後続する後続文字列に対応する信頼度より高い場合に、連結した前記文字列の末尾に前記後続文字列を更に連結した文字列を、先に表示すべき前記文字列を表示すべき時刻に表示させる設定を行う
請求項９記載の設定装置。

【請求項１１】
内容が予め定められた音声の再生に同期して前記内容を表示する表示タイミングを設定する設定装置として、コンピュータを機能させるプログラムであって、
前記コンピュータを、
前記音声の内容を示す内容データを取得する内容データ取得部と、
再生される前記音声を音声認識した文字データを分割して複数の認識データを生成する音声認識部と、
前記複数の認識データの各々に一致する文字列を前記内容データから検出する文字列検

出部と、
前記文字列検出部により一致する文字列が検出されなかった各認識データについて、当該認識データに含まれる各文字に一致する文字を前記内容データから検出することにより、当該認識データに一致する文字列を前記内容データから検出する文字検出部と、
前記認識データのうち前記文字検出部により一致する文字が検出されなかった各文字について、当該文字の読みに含まれる音素に一致する音素を、前記内容データの読みの中から検出する音素検出部として機能させ、
前記文字検出部は、前記認識データのうち前記音素検出部により一致する音素が検出された文字に一致する文字として、前記内容データにおいて当該音素を含む文字を更に検出し、
前記コンピュータを、更に、
前記内容データに含まれる文字列の各々を表示させる表示タイミングを、当該文字列に一致する認識データとして音声認識された音声の再生時に設定する表示設定部と
して機能させるプログラム。
【請求項１２】
内容が予め定められた音声の再生に同期して前記内容を表示する表示タイミングを設定する設定装置として、コンピュータを機能させるプログラムであって、
前記コンピュータを、
再生される前記音声の内容を示す内容データに含まれる複数の文字列の各々に対応付けて、当該文字列を表示すべき時刻、及び、当該時刻に当該文字列を内容とする音声が再生される確度である信頼度を取得する信頼度取得部と、
前記複数の文字列のうち連続した２つの文字列について、先に表示すべき文字列に対応する信頼度が、後に表示すべき文字列に対応する信頼度より高い場合に、先に表示すべき前記文字列の末尾に後に表示すべき前記文字列を連結した文字列を、先に表示すべき前記文字列を表示すべき時刻に表示させる設定を行う表示設定部と
して機能させるプログラム。
【請求項１３】
内容が予め定められた音声の再生に同期して前記内容を表示する表示タイミングを設定する設定方法であって、
コンピュータにより、
前記音声の内容を示す内容データを取得する内容データ取得段階と、
再生される前記音声を音声認識した文字データを分割して複数の認識データを生成する音声認識段階と、
前記複数の認識データの各々に一致する文字列を前記内容データから検出する文字列検出段階と、
前記文字列検出段階において一致する文字列が検出されなかった各認識データについて、当該認識データに含まれる各文字に一致する文字を前記内容データから検出することにより、当該認識データに一致する文字列を前記内容データから検出する文字検出段階と、
前記認識データのうち前記文字検出段階において一致する文字が検出されなかった各文字について、当該文字の読みに含まれる音素に一致する音素を、前記内容データの読みの中から検出する音素検出段階部と、
前記認識データのうち前記音素検出段階において一致する音素が検出された文字に一致する文字として、前記内容データにおいて当該音素を含む文字を更に検出する段階と、
前記内容データに含まれる文字列の各々を表示させる表示タイミングを、当該文字列に一致する認識データとして音声認識された音声の再生時に設定する表示設定段階と
を備える設定方法。
【請求項１４】
内容が予め定められた音声の再生に同期して前記内容を表示する表示タイミングを設定する設定方法であって、
コンピュータにより、

再生される前記音声の内容を示す内容データに含まれる複数の文字列の各々に対応付けて、当該文字列を表示すべき時刻、及び、当該時刻に当該文字列を内容とする音声が再生される確度である信頼度を取得する信頼度取得段階と、
前記複数の文字列のうち連続した２つの文字列について、先に表示すべき文字列に対応する信頼度が、後に表示すべき文字列に対応する信頼度より高い場合に、先に表示すべき前記文字列の末尾に後に表示すべき前記文字列を連結した文字列を、先に表示すべき前記文字列を表示すべき時刻に表示させる設定を行う表示設定段階と
を備える設定方法。
【請求項１５】
内容が予め定められた音声の再生に同期して前記内容を表示する表示タイミングを設定する設定装置として、コンピュータを機能させるプログラムであって、
前記コンピュータを、
前記音声の内容を示す内容データを取得する内容データ取得部と、
再生される前記音声を音声認識した文字データを分割して複数の認識データを生成する音声認識部と、
前記複数の認識データの各々が文字列に一致する確度である信頼度を算出する信頼度算出部と、
前記複数の認識データの各々に一致する文字列を前記内容データから検出すると共に、予め定められた基準信頼度未満の信頼度の認識データである低信頼データについては、さらに、当該低信頼データに後続する認識データに一致する文字列を検出できなかった場合に、当該低信頼データに一致する文字列は検出できないと判断する文字列検出部と、
前記文字列検出部により一致する文字列が検出されなかった各認識データについて、当該認識データに含まれる各文字に一致する文字を前記内容データから検出することにより、当該認識データに一致する文字列を前記内容データから検出する文字検出部と、
前記内容データに含まれる文字列の各々を表示させる表示タイミングを、当該文字列に一致する認識データとして音声認識された音声の再生時に設定する表示設定部と
して機能させるプログラム。
【請求項１６】
内容が予め定められた音声の再生に同期して前記内容を表示する表示タイミングを設定する設定装置として、コンピュータを機能させるプログラムであって、
前記コンピュータを、
前記音声の内容を示す内容データを取得する内容データ取得部と、
再生される前記音声を音声認識した文字データを分割して複数の認識データを生成すると共に、さらに、音声認識した前記複数の認識データが、再生される音声の内容と一致する可能性を示す音声認識確信度を、認識データ毎に生成する音声認識部と、
前記複数の認識データの各々について、音声認識確信度がより高い認識データに一致する文字列を、当該認識データと比較して音声認識確信度が低い認識データに先立って前記内容データから検出すると共に、第１の前記認識データに一致する第１の文字列及び第２の前記認識データに一致する第２の文字列を検出した場合に、前記第１の認識データに後続しかつ前記第２の認識データに先行する認識データに一致する文字列として、前記第１の文字列に後続し前記第２の文字列に先行する文字列を検出する文字列検出部と、
前記文字列検出部により一致する文字列が検出されなかった各認識データについて、当該認識データに含まれる各文字に一致する文字を前記内容データから検出することにより、当該認識データに一致する文字列を前記内容データから検出する文字検出部と、
前記内容データに含まれる文字列の各々を表示させる表示タイミングを、当該文字列に一致する認識データとして音声認識された音声の再生時に設定する表示設定部と
として機能させるプログラム。
【発明の詳細な説明】
【技術分野】
【０００１】

本発明は、設定装置、プログラム、記録媒体、及び設定方法に関する。特に本発明は、音声に同期して音声の内容を表示する処理を制御する設定装置、プログラム、記録媒体、及び設定方法に関する。

【背景技術】

【０００２】

近年、ＩＴ機器の性能が飛躍的に向上し、ブロードバンド等のコンピュータネットワークが整備されるのに伴って、動画などのデジタルコンテンツが盛んに配信されるようになってきている。動画は、文字などの静的なコンテンツと比較して多くの利用者にとって分かり易く、かつ説得力が高い。更に、ケーブルテレビ及びＣＳテレビ等の普及により、テレビ番組のチャンネル数が増加しており、動画コンテンツは様々な分野において更に広く用いられていくことが予想される。

【０００３】

動画により提供される情報をより多くの利用者に適切に提供するには、動画に対応付けて音声の内容を示す字幕を表示することが必要である。更に、２００７年には、放送等される全ての動画に字幕を付与することが目標として掲げられている。このため、動画に対して適切な字幕を表示する技術の進歩が社会的に要請されている。

【０００４】

従来、音声を認識して音声の内容を示す文字列を生成する音声認識技術により、字幕を生成する方法が提案されている。しかしながら、音声認識技術は、音声を誤認識して誤った文字列を生成する場合がある。また、句読点又は記号等は、音声として表現されないので、音声認識技術によってこれらの記号を適切に表示させることはできない。このため、音声認識技術をそのまま字幕生成に適用することはできず、音声認識結果を修正して字幕を作成していた（特許文献２参照。）。

【０００５】

他の方法として、動画の音声の台本を、適切な長さの文字列毎に分割して、適切なタイミングで表示する方法が提案されている。しかしながら、高機能な動画編集ソフトウェアを用いた場合であっても、手作業で適切なタイミングを決定するのは困難であった。このため、従来、再生される音声と台本とを比較して、台本中の各文字列を表示すべきタイミングを決定する技術が提案されている（特許文献１及び３参照。）。

【０００６】

非特許文献１については後述する。

【特許文献１】特開平１０－２５４４７８号公報

【特許文献２】特開２０００－８９７８６号公報

【特許文献３】特開平１０－１３６２６０号公報

【非特許文献１】「テレビドラマのシナリオと音声トラックの自動対応付け」、谷村正剛ら著、自然言語情報処理２６－４、１９９９年５月２８日発行

【発明の開示】

【発明が解決しようとする課題】

【０００７】

特許文献１及び３の技術は、まず、音声を分析することにより、音声が所定期間発せられなかった部分を文の切れ目と判断する。そして、音声を分析した結果得られた文の文頭の音素と、台本における各文の文頭に含まれる音素とを比較することにより、音声と台本との対応付けを生成する。これにより、台本中の各文を、その文に一致した音声が発せられる時に表示すべきであることが分かる。

【０００８】

しかしながら、音声が発せられない部分は、文の切れ目とは限らない。例えば、話者は、迷ったり困ったとき、息継ぎをするとき、一瞬考えたとき、又は強調したいとき等、様々な場合に間をおいて話す。従って、この技術により文の切れ目を適切に判断することは難しく、台本中の各文と一致する音声を検出するのは困難である。音声と字幕の表示タイミングが相違すると、例えば、話者が話し始めているにもかかわらず字幕が表示されない

という不都合、又は、話者が話し始めていないのにも関わらずクイズの答えなどが先に表示されてしまう不都合が発生してしまう。

【０００９】

更に、この技術は、台本中の各文をそのまま字幕として表示するため、利用者の読みやすさ又は表示装置の画面のサイズ等を考慮して文を分割又は結合することはできない。また、この技術は、音声認識の精度に関わらず同様の字幕を生成するため、音声認識の認識率が将来向上したとしても、字幕の表示精度を向上させることはできない。

【００１０】

そこで本発明は、上記の課題を解決することのできる設定装置、プログラム、記録媒体、及び設定方法を提供することを目的とする。この目的は特許請求の範囲における独立項に記載の特徴の組み合わせにより達成される。また従属項は本発明の更なる有利な具体例を規定する。

【課題を解決するための手段】

【００１１】

上記課題を解決するために、本発明の第１の形態においては、内容が予め定められた音声の再生に同期してその内容を表示する表示タイミングを設定する設定装置であって、音声の内容を示す内容データを取得する内容データ取得部と、再生される音声を音声認識した文字データを分割して複数の認識データを生成する音声認識部と、複数の認識データの各々に一致する文字列を内容データから検出する文字列検出部と、文字列検出部により一致する文字列が検出されなかった各認識データについて、当該認識データに含まれる各文字に一致する文字を内容データから検出することにより、当該認識データに一致する文字列を内容データから検出する文字検出部と、前記認識データのうち前記文字検出部により一致する文字が検出されなかった各文字について、当該文字の読みに含まれる音素に一致する音素を、前記内容データの読みの中から検出する音素検出部とを備え、前記文字検出部は、前記認識データのうち前記音素検出部により一致する音素が検出された文字に一致する文字として、前記内容データにおいて当該音素を含む文字を更に検出し、内容データに含まれる文字列の各々を表示させる表示タイミングを、当該文字列に一致する認識データとして音声認識された音声の再生時に設定する表示設定部を更に備える設定装置、当該設定装置を用いた設定方法、コンピュータを当該設定装置として機能させるプログラム、当該プログラムを記録した記録媒体を提供する。

【００１２】

なお、上記の発明の概要は、本発明の必要な特徴の全てを列挙したものではなく、これらの特徴群のサブコンビネーションもまた、発明となりうる。

【発明の効果】

【００１３】

本発明によれば、音声の再生に同期して、その音声の内容を示す文字列を表示することができる。

【発明を実施するための最良の形態】

【００１４】

以下、発明の実施の形態を通じて本発明を説明するが、以下の実施形態は特許請求の範囲にかかる発明を限定するものではなく、また実施形態の中で説明されている特徴の組み合わせの全てが発明の解決手段に必須であるとは限らない。例えば、明快な方法として全体的にまたは部分的に重み付けなどを利用して幾つかの処理を一括処理させるような手段も想定しうる事はいうまでもない。

【００１５】

図１は、字幕設定装置１０の機能ブロック図である。字幕設定装置１０は、内容が予め定められた音声を含む動画の再生に同期して、その内容を字幕として表示するタイミングを字幕行表示装置１５に設定することを目的とする。字幕設定装置１０は、内容データ取得部１００と、音声認識部１１０と、文字列検出部１２０と、文字検出部１３０と、読付部１４０と、音素検出部１５０と、信頼度算出部１６０と、信頼度取得部１６５と、表示

設定部１７０とを備える。

【００１６】

内容データ取得部１００は、音声の内容を示す内容データ２０を取得する。音声認識部１１０は、話者により話された音声をマイクなどで取得することにより、再生される音声を音声認識する。音声認識には、既存の多様な技術を適用可能である。そして、音声認識部１１０は、音声認識した文字データを分割して複数の認識データを生成する。例えば、生成した複数の認識データを、認識データ３０－１～Ｎとする。ここで、認識データ３０－１～Ｎの各々は、音声認識された時刻に対応付けられていることが望ましい。音声認識部１１０は、生成した認識データ３０－１～Ｎを文字列検出部１２０に送る。

【００１７】

文字列検出部１２０は、認識データ３０－１～Ｎの各々に一致する文字列を内容データ２０から検出し、検出結果を信頼度算出部１６０に送る。文字列検出部１２０は、文字検出部１３０から受け取った検出結果に基づいて一致する文字列を検出してもよい。文字検出部１３０は、文字列検出部１２０により一致する文字列が検出されなかった各認識データについて、その認識データに含まれる各文字に一致する文字を内容データ２０から検出することにより、その認識データに一致する文字列を内容データ２０から検出する。また、文字検出部１３０は、音素検出部１５０により一致する音素が検出された文字に一致する文字として、内容データ２０においてその音素を含む文字を検出してもよい。そして、文字検出部１３０は、検出結果を文字列検出部１２０に送る。

【００１８】

読付部１４０は、内容データ２０の読み方の候補である読み候補を複数生成する。更に、読付部１４０は、これらの読み候補の各々を、その読み候補により読まれる可能性を示す情報に対応付けて生成してもよい。音素検出部１５０は、認識データ３０－１～Ｎのうち文字検出部１３０により一致する文字が検出されなかった各文字を、その文字の読みを含む音素に展開する。例えば、音素検出部１５０は、漢字をその漢字の読みを示す平仮名に変換してもよい。

【００１９】

そして、音素検出部１５０は、認識データ３０－１～Ｎのうち文字検出部１３０により一致する文字が検出されなかった各文字について、その文字の読みに含まれる音素に一致する音素を、読付部１４０により生成された複数の読み候補の何れかの中から検出する。更に、音素検出部１５０は、一致する音素を検出できなかった文字については、その文字に含まれる音素が発せられる時間の長さ等に基づいて、一致する文字を検出してもよい。そして、音素検出部１５０は、検出結果を文字検出部１３０に送る。

【００２０】

信頼度算出部１６０は、認識データ３０－１～Ｎの各々が文字列に一致する確度である信頼度を、認識データ毎に算出する。ここで、信頼度とは、各認識データが音声認識された時刻に、その認識データに一致する文字列を内容とする音声が再生される確度をいう。例えば、信頼度算出部１６０は、文字列検出部１２０のみにより一致する文字列が検出された認識データに対応付けて、文字列検出部１２０及び文字検出部１３０により一致する文字列が検出された認識データと比較して、より高い信頼度を算出してもよい。そして、信頼度算出部１６０は、内容データ２０の各文字列に信頼度を対応付けて信頼度取得部１６５に送る。

【００２１】

信頼度取得部１６５は、各文字列に対応付けて、その文字列を表示すべき時刻、即ちその文字列に一致する認識データとして音声認識された音声の再生時刻を、文字列検出部１２０を介して音声認識部１１０から取得する。更に、信頼度取得部１６５は、その再生時刻にその文字列を内容とする音声が再生される確度である信頼度を、信頼度算出部１６０から取得する。

【００２２】

表示設定部１７０は、内容データ２０に含まれる文字列の各々を表示させる表示タイミ

ングを、その文字列に一致する認識データとして音声認識された音声の再生時に設定する。例えば、表示設定部１７０は、文字列を表示するべき時刻をその文字列に対応付けた表示タイミング情報４０を、字幕行表示装置１５に出力してもよいし、字幕行表示装置１５から参照可能なデータベース等に格納してもよい。更に、表示設定部１７０は、表示するべき字幕の設定情報を字幕行表示装置１５から取得し、取得したその設定情報に基づいて文字列を連結してもよい。設定情報とは、例えば、字幕行表示装置１５の表示部において１行に表示可能な文字数、又は、句点若しくは読点において字幕を改行するか否かを示す情報をいう。

【００２３】

図２は、内容データ２０の一例を示す。内容データ２０は、音声の内容として、漢字、仮名、アルファベット、及び句読点により構成される文字列を含む。内容データ２０を用いることにより、音声認識の結果をそのまま表示する技術とは異なり、発音されない記号等も含めて適切な字幕を生成できる。以降、内容データ２０が、「近年ＩＴを取り巻く環境は著しく変化いたしております。ここで流れをおさらいします。」という文章である場合について説明する。

【００２４】

図３は、認識データ３０－１～Ｎの一例を示す。音声認識部１１０は、認識データ３０－１～Ｎの各々を、その認識データが認識された認識時刻に対応付けて生成する。ここで、認識時刻とは、例えば、音声の再生開始時から、その認識データとして認識される部分の再生時点までに経過した時間である。一例として、音声認識部１１０は、認識データ３０－１として「金ねうん」の文字列を生成し、認識データ３０－１に対応付けて「０２．１０３」の認識時刻を生成する。

【００２５】

図３を図２と比較することにより、音声認識部１１０が、音声認識の結果、内容データ２０とは異なる誤った文字列を生成していることが分かる。本実施例における字幕設定装置１０は、音声認識の結果がこのように誤っている場合であっても、適切な表示タイミングを設定することができる。

【００２６】

図４は、表示タイミング情報４０の一例を示す。表示設定部１７０は、表示タイミング情報４０として、表示するべき文字列に、その文字列を表示するべき表示時刻を対応付けた情報を生成し、字幕行表示装置１５に設定する。ここで、表示時刻とは、音声の再生開始時から文字列を表示すべるべき時刻までの時間である。一例として、表示設定部１７０は、「近年ＩＴを取り巻く環境は」の文字列を、表示時刻の一例であるタイムスタンプ「０２．１０３」に対応付けた情報を生成する。

【００２７】

ここで、「近年ＩＴを取り巻く環境は」の文字列は、認識データ３０－１～４の組に対応する。このように、表示設定部１７０は、表示すべき時刻を設定するのみならず、必要に応じて文字列を連結する処理を行う。

【００２８】

図５は、字幕設定装置１０が表示タイミングを設定する処理の動作フローを示す。内容データ取得部１００は、音声の内容を示す内容データ２０を取得する（Ｓ５００）。ここで、内容データ取得部１００は、取得するべき内容データを利用者からの指示に基づいて選択してもよい。音声認識部１１０は、再生される音声を音声認識し（Ｓ５１０）、音声認識した文字データを分割して認識データ３０－１～Ｎを生成する（Ｓ５１５）。例えば、音声認識部１１０は、予め定められた単語数、例えば1から3単語毎に分割してもよいし、予め定められた再生時間、例えば１．５秒毎に分割してもよい。

【００２９】

これに代えて、音声認識部１１０は、無音状態が所定の長さ、例えば１００ミリ秒以上継続した部分を、認識データの境界と判断してもよい。好ましくは、音声認識部１１０は、文字データを、できるだけ短い意味のまとまり、例えば文より短い単語、句、又は節等

の文字列毎に分割する。これにより、タイムスタンプの生成精度を高めることができる。音声認識部１１０によるこの分割の処理を、細分音声認識行処理と呼ぶ。

【００３０】

また、好ましくは、音声認識部１１０は、音声認識した認識データ３０－１～Ｎが、再生される音声の内容と一致する可能性を示す音声認識確信度を、音声認識データ毎に更に生成する。更に、音声認識部１１０は、音声認識した複数の文字データの中から、分割するべき文字データを、利用者からの指示に基づいて選択してもよい。

【００３１】

続いて、文字列検出部１２０は、認識データ３０－１～Ｎの各々に一致する文字列を内容データ２０から検出する（Ｓ５２０）。この処理の中で、必要に応じて、文字検出部１３０は、認識データに含まれる各文字に一致する文字を内容データ２０から検出してもよい。また、認識データ３０－１～Ｎのうち文字検出部１３０により一致する文字が検出されなかった各文字について、その文字の読みに含まれる音素に一致する音素を内容データ２０から検出してもよい。詳細は後述する。

【００３２】

表示設定部１７０は、文字列検出部１２０により何れの文字列にも一致しない認識データがあるか否か判断する（Ｓ５３０）。そのような認識データがある場合に（Ｓ５３０：ＹＥＳ）、表示設定部１７０は、文字列検出部１２０により内容データ２０において一致する文字列が検出されなかったその認識データを、その認識データとして音声認識された音声の再生時に表示させるべく、表示すべき文字列に追加する（Ｓ５４０）。

【００３３】

更に、表示設定部１７０は、表示するべき字幕の設定情報を字幕行表示装置１５から取得し、取得したその設定情報に基づいて文字列を連結する（Ｓ５５０）。表示設定部１７０は、この連結処理を、字幕表示に先立って予め行ってもよいし、字幕を表示すべき指示を字幕行表示装置１５から受けた場合に動的に行ってもよい。これにより、表示設定部１７０は、字幕行表示装置１５の表示部のサイズ等に応じて、適切な字幕を生成することができる。以降、表示設定部１７０によるこの連結処理を、最適字幕行表示処理と呼ぶ。そして、表示設定部１７０は、内容データ２０に含まれる文字列の各々を表示させる表示タイミングを、その文字列に一致する認識データとして音声認識された音声の再生時に設定する（Ｓ５６０）。

【００３４】

図６は、Ｓ５２０における処理の詳細を示す。文字列検出部１２０は、まず、認識データ３０－１～Ｎの各々に含まれる各文字に一致する文字を内容データ２０から検出する（Ｓ６００）。従来、この検出処理として効率的な方法であるＤＰマッチングが、提案されている（非特許文献１参照。）。文字列検出部１２０は、例えば、ＤＰマッチングによりこの検出処理を行ってもよいし、他の手法によりこの検出処理を行ってもよい。文字列検出部１２０は、検出結果として、内容データ２０中の文字を順次縦軸に配列し、認識データ３０－１～Ｎを順次横軸に配置した座標軸において、内容データ２０中の各文字と認識データ３０－１～Ｎ中の各文字が一致する点を順次経由するグラフである最適マッチング路を生成してもよい。

【００３５】

そして、文字列検出部１２０は、認識データ３０－１～Ｎの各々に一致する文字列を再度検出するべく、以下の処理を各認識データについて繰り返す（Ｓ６１０）。本実施例において、文字列検出部１２０は、複数の認識データに対して、認識された順に、即ち音声として再生された順に、以下の処理を行う。これに代えて、文字列検出部１２０は、音声認識確信度が高い順に、即ち、音声認識確信度がより高い認識データに対して、その認識データと比較して音声認識確信度が低い認識データに先立って、以下の処理を行ってもよい。

【００３６】

文字列検出部１２０は、その認識データに一致する文字列を内容データ２０から検出す

る（Ｓ６２０）。続いて、文字列検出部１２０は、一致する文字列を検出できなかった認識データについて、その認識データに含まれる文字と一致する文字を検出する処理が必要か否か判断する（Ｓ６３０）。処理が必要な場合に（Ｓ６３０：ＹＥＳ）、文字検出部１３０は、その認識データに含まれる文字に一致する文字を検出する（Ｓ６４０）。なお、このように、一致する文字列が検出できなかったデータあるいは単純に一括した文字列に対して、そのデータに含まれる部分データについて一致する文字を順次再帰的にスケールレベルを遷移させながら検出する処理を、本実施例においては、アップスケーリング処理と呼ぶ。

【００３７】

続いて、信頼度算出部１６０は、認識データ３０－１～Ｎの各々が文字列に一致する確度である信頼度を、認識データ毎に算出する（Ｓ６５０）。例えば、信頼度算出部１６０は、文字列検出部１２０のみにより一致する文字列が検出された認識データに対応付けて、文字列検出部１２０及び文字検出部１３０により一致する文字列が検出された認識データと比較して、より高い信頼度を算出してもよい。

【００３８】

また、信頼度算出部１６０は、音素検出部１５０により一致する音素が検出された文字を含む認識データに対応付けて、音素検出部１５０により一致する音素が検出されることなく文字検出部１３０により一致する文字が検出された認識データと比較して、より低い信頼度を生成する。即ち、アップスケーリング処理の段階が増加するのに応じてより低い信頼度を生成する。これにより、異なる文字で音素が偶然一致したようなエラーを含み得る認識データに対しては、より低い信頼度を生成することができる。

【００３９】

字幕設定装置１０は、以上の処理を各認識データについて繰り返す（Ｓ６６０）。

本図に示すように、文字列検出部１２０は、各認識データに一致する文字列を内容データ２０から検出する処理に先立って、検出精度の高いＤＰマッチング等により、文字単位の一致を判断する。そして、文字列検出部１２０は、ＤＰマッチングにより既に一致する文字が検出された認識データについて、その認識データと一致する文字列を内容データ２０から再度検出する。これにより、一致する文字を検出する精度を高められると共に、その一致が認識データ単位で判断されるものであるか、又は文字単位で判断されるものであるか判断できる。これに代えて、文字列検出部１２０は、ＤＰマッチング等の文字単位の一致を判断しなくともよい。

【００４０】

図７は、Ｓ６２０における処理の詳細を示す。文字列検出部１２０は、まず、検出対象の認識データに一致する文字列を内容データ２０から検出する。更に、文字列検出部１２０は、検出対象の認識データの検出結果のみならず、その前に検出する対象であった認識データ、及び更にその前に検出する対象であった認識データの検出結果に基づいて、更に以下の処理を行う。

【００４１】

本図における丸印は、認識データに一致する文字列が検出されたことを示す。一方、Ｘ印は、認識データに一致する文字列が検出されていないことを示す。例えば文字列検出部１２０は、１つ前の認識データ及び対象の認識データの各々に一致する文字列を検出した場合には、２つ前の認識データの検出結果に関わらず、対象の認識データに一致する文字列を検出したと判断する。

【００４２】

文字列検出部１２０は、１つ前の認識データに一致する文字列を検出し、かつ対象の認識データに一致する文字列を検出できなかった場合には、２つ前の認識データの検出結果に関わらず、対象の認識データについての処理を保留し、Ｓ６２０の処理を終え、次の認識データについての処理に移る。

【００４３】

１つ前の認識データに一致する文字列が検出できない場合において、文字列検出部１２

０は、以下の処理を行う。
文字列検出部１２０は、２つ前の認識データ及び対象の認識データの各々に一致する文字列を検出した場合には、当該１つ前の認識データが、２つ前の認識データ及び対象の認識データの各々に一致する各文字列の間の文字列に一致すると判断する。
【００４４】
文字列検出部１２０は、２つ前の認識データに一致する文字列を検出し、対象の認識データに一致する文字列を検出できない場合に、対象の認識データについての処理を保留し、Ｓ６２０の処理を終え、次の認識データについての処理に移る。但し、対象の認識データが、検出対象の最後の認識データである場合には、文字列検出部１２０は、１つ前の認識データ及び対象の認識データを連結したデータを対象として、文字検出部１３０により一致する文字を検出させる。即ちこの場合、Ｓ６３０において、文字列検出部１２０は、一致する文字を検出する処理が必要と判断する。
【００４５】
文字列検出部１２０は、２つ前の認識データに一致する文字列を検出できず、対象の認識データに一致する文字列を検出できた場合に、２つ前の認識データ及び１つ前の認識データを連結したデータを対象として、文字検出部１３０により一致する文字を検出させる。文字検出部１３０の検出結果に基づく認識データの信頼度が、予め定められた基準信頼度未満であれば、文字列検出部１２０は、その認識データに対応付けて、その認識データが信頼度の低い旨を示す低信頼データである旨の再評価フラグを付す。本図においてはこのフラグを三角印で表す。
【００４６】
文字列検出部１２０は、２つ前、１つ前、及び対象の認識データの何れにも一致する文字列を検出できなかった場合に、１つ前、２つ前、及び更にその前の認識データを結合したデータを対象として、文字検出部１３０により一致する文字を検出させる。
【００４７】
一方、文字列検出部１２０は、２つ前の認識データに一致する文字列を検出できず、１つ前の認識データに再評価フラグが付されている場合においては、以下の処理を行う。
文字列検出部１２０は、対象の認識データに一致する文字列を検出した場合に、１つ前の認識データから再評価フラグを取り除くことにより、当該１つ前の認識データに一致する文字列を検出したと判断する。即ち三角印を丸印に変更する。一方、対象の認識データに一致する文字列を検出できなかった場合に、文字列検出部１２０は、１つ前の認識データから再評価フラグを取り除くことにより、当該１つ前の認識データに一致する文字列を検出できないと判断する。即ち、三角印をバツ印に変更する。
【００４８】
このように、文字列検出部１２０は、複数の認識データの各々について、その認識データに完全に一致する文字列のみならず、その認識データの前後の認識データが一致した場合に、その認識データについても一致したと判断する。より正確には、文字列検出部１２０は、第１の認識データに一致する第１の文字列及び第２の認識データに一致する第２の文字列を検出した場合に、第１の認識データに後続しかつ第２の認識データに先行する認識データに一致する文字列として、第１の文字列に後続し第２の文字列に先行する文字列を検出する。即ち、一致とは、完全一致のみならず、前後の認識データが一致したことに基づくこの一致を含む。以降、この一致を、挟み打ち処理による一致と呼ぶ。
これにより、文字又は音素単位の一致を検出する処理をできるだけ減少させて、処理の効率を高めることができる。更に、文字単位の一致を検出する必要がある場合であっても、検出範囲を限定することができるので、効率がよい。
【００４９】
図８は、Ｓ６４０における処理の詳細を示す。文字検出部１３０は、検出対象の認識データに含まれる各文字について、以下の処理を繰り返す（Ｓ８００）。まず、文字検出部１３０は、その文字に一致する文字を内容データ２０から検出する（Ｓ８１０）。そして、文字検出部１３０は、一致する文字を検出できなかった認識データについて、その認識

データに含まれる文字に含まれる音素と一致する音素を検出する処理が必要か否か判断する（Ｓ８２０）。

【００５０】

処理が必要な場合に（Ｓ８２０：ＹＥＳ）、音素検出部１５０は、文字検出部１３０により一致する文字が検出されなかった各文字について、その文字の読みに含まれる音素に一致する音素を、内容データ２０の読みの中から検出する（Ｓ８３０）。文字検出部１３０は、以上の処理を各文字について繰り返す（Ｓ８４０）。

【００５１】

図９は、Ｓ８１０における第１の処理の詳細を示す。文字検出部１３０は、検出対象の文字が認識データの末尾の文字でない場合に、本図の処理を行う。まず、文字検出部１３０は、検出対象の文字に一致する文字を内容データ２０から検出する。更に、文字検出部１３０は、検出対象の文字の検出結果のみならず、その前に検出する対象であった文字、及び、検出対象の認識データ（認識データの組を含む）の先頭の文字の検出結果に基づいて、更に以下の処理を行う。

【００５２】

文字検出部１３０は、１つ前の文字に一致する文字を検出し、かつ対象の文字に一致する文字を検出した場合には、Ｓ８１０における処理を終了し、次の文字に対する検出処理に移る。一方、文字検出部１３０は、１つ前の文字に一致する文字を検出し、かつ対象の文字に一致する文字を検出できなかった場合には、対象の文字についての検出処理を保留して、次の文字についての処理に移る。

【００５３】

文字検出部１３０は、１つ前の文字に一致する文字を検出できなかった場合においては、他の条件に応じて以下の処理を行う。

文字検出部１３０は、先頭の文字及び対象の文字の各々に一致する文字を検出した場合に、文字の一致を検出すべき対象の認識データ全体が、内容データ２０の文字列に一致したと判断する。即ち、文字検出部１３０は、同一の認識データ内の文字については、複数の文字についても挟み打ち処理による一致の判断を行う。このように、文字の一致とは、文字の完全一致のみならず、前後の文字が一致したことに基づく一致を含む。

【００５４】

一方、文字検出部１３０は、対象の文字に一致する文字を検出できなかった場合には、対象の文字についての検出処理を保留して、次の文字についての処理に移る。文字検出部１３０は、先頭の文字に一致する文字を検出できず、対象の文字に一致する文字を検出できた場合には、Ｓ８１０における処理を終了し、次の文字に対する検出処理に移る。

【００５５】

図１０は、Ｓ８１０における第２の処理の詳細を示す。文字検出部１３０は、検出対象の文字が認識データの末尾の文字である場合に、本図の処理を行う。具体的には、文字検出部１３０は、検出対象の認識データ（認識データの組を含む）の先頭の文字の検出結果、及び、検出対象の次の認識データの文字列検出部１２０による検出結果に基づいて、以下の処理を行う。

【００５６】

文字検出部１３０は、先頭の文字に一致する文字を検出し、かつ次の認識データに一致する文字列が検出されている場合に、対象の認識データ全体を一致と判断する。一方、文字検出部１３０は、先頭の文字に一致する文字を検出し、かつ次の認識データに一致する文字列が検出されていない場合に、先頭の文字に後続する文字列の音素を検出対象として、音素検出部１５０により一致する音素を検出させる。

【００５７】

文字検出部１３０は、先頭の文字に一致する文字を検出せず、かつ次の認識データに一致する文字列が検出されている場合に、次の認識データに先行する文字列の音素を検出対象として、音素検出部１５０により一致する音素を検出させる。一方、文字検出部１３０は、先頭の文字に一致する文字を検出せず、かつ次の認識データに一致する文字列が検出

されていない場合に、文字を検出する対象の認識データ全体を検出対象として、音素検出部１５０により一致する音素を検出させる。

【００５８】

図１１は、Ｓ８３０における第１の処理の詳細を示す。音素検出部１５０は、検出対象の音素が認識データの末尾の文字でない場合に、本図の処理を行う。まず、音素検出部１５０は、検出対象の音素に一致する音素を内容データ２０の所定の読み候補から検出する。音素検出部１５０は、一致する音素が検出できなかった場合には、検出対象の音素を読まれる可能性が高い順に複数の読み候補の各々と比較するべく、次に可能性の高い読み候補と比較する。何れの読み候補にも一致しない場合には、音素検出部１５０は、その音素についての処理を保留して、次の音素の処理に移る。

【００５９】

続いて、検出対象の音素に一致する音素を検出した場合には、音素検出部１５０は、検出対象の認識データ（認識データの組を含む）の先頭の文字の検出結果、及び、検出対象の１つ前の音素の検出結果に基づいて、以下の処理を行う。

先頭文字に一致する文字が検出されている場合に、文字検出部１３０は、検出対象の認識データ内において一致する文字が検出されていない各文字について、一致する文字が検出されたと判断する。このように、音素の一致に基づいて、文字についての挟み打ち処理による一致を判断してもよい。一方、音素検出部１５０は、対象の音素に一致する音素を検出したその他の場合には、対象の音素についての処理を終了して、次の音素についての処理に移る。

【００６０】

図１２は、Ｓ８３０における第２の処理の詳細を示す。音素検出部１５０は、検出対象の音素が認識データの末尾の文字である場合に、本図の処理を行う。具体的には、音素検出部１５０は、先頭文字に一致する文字の検出結果と、次の認識データに一致する文字列の検出結果、又は、対象の認識データが最後の認識データ（例えば認識データ３０－Ｎ）であるか否かとに基づいて、以下の処理を行う。

【００６１】

まず、次の認識データに一致する文字列が検出されていない場合と、検出対象の認識データが最後の認識データでない場合とにおいて、音素検出部１５０は、Ｓ８３０の処理、即ち一致する音素を検出する処理を終了する。この結果、文字列検出部１２０は、次の認識データについての処理に移る。一方、次の認識データに一致する文字列が検出されている場合、又は、検出対象の認識データが最後の認識データである場合においては、以下の処理を行う。

【００６２】

音素検出部１５０は、先頭文字に一致する文字が検出されている場合には、検出対象の認識データ内の不一致文字を一致するものと判断する。一方、音素検出部１５０は、先頭文字に一致する文字が検出されていない場合には、音声の内容に関わらず音声の長さ又は文字の長さに基づいて一致する音素を検出する強制割り振り処理を行う。信頼度算出部１６０は、この強制割り振り処理により一致する音素が検出された認識データに対応付けて、この強制割り振り処理によらず一致する文字が検出された認識データと比較して、更に低い信頼度を算出する。

【００６３】

図１３は、Ｓ５５０における処理の詳細を示す。表示設定部１７０は、利用者から入力された、表示するべき字幕の設定情報を字幕行表示装置１５等から取得する（Ｓ１３１０）。例えば、表示設定部１７０は、設定情報を示すコマンド等をパースすることにより、設定情報の内容を解析する（Ｓ１３２０）。設定情報とは、字幕行表示装置１５の表示部の１行に表示させる文字数であってもよいし、字幕を句読点で改行するか否かの指示であってもよい。

【００６４】

信頼度取得部１６５は、内容データ２０に含まれる複数の文字列の各々に対応付けて、

その文字列を表示すべき時刻、及びその時刻にその文字列を内容とする音声が再生される確度である信頼度を、信頼度算出部１６０から取得する（Ｓ１３２５）表示設定部１７０は、各認識データと一致する内容データ２０内の各文字列について、その文字列が、設定情報の条件を満たすか否か判断する（Ｓ１３３０）。

【００６５】

満たしていない場合に（Ｓ１３３０：ＮＯ）。表示設定部１７０は、複数の文字列を連結する（Ｓ１３４０）。具体的には、表示設定部１７０は、複数の文字列のうち連続した２つの文字列について、先に表示すべき文字列に対応する信頼度が、後に表示すべき文字列に対応する信頼度より高い場合に、先に表示すべきその文字列の末尾に後に表示すべきその文字列を連結した文字列を、先に表示すべき文字列を表示すべき時刻に表示させる設定を行う。

【００６６】

そして、表示設定部１７０は、Ｓ１３３０に処理を戻して判断を繰り返す。この結果、表示設定部１７０は、設定情報の条件を満たすまで、文字列の連結を繰り返す。例えば、表示設定部１７０は、Ｓ１３４０の処理を行う直前において先に表示すべき文字列に対応する信頼度が、後に表示すべき文字列に後続する後続文字列に対応する信頼度より高い場合に、Ｓ１３４０において連結した文字列の末尾にその後続文字列を更に連結した文字列を、先に表示すべきその文字列を表示すべき時刻に表示させる設定を行ってもよい。

このように、表示設定部１７０は、表示デバイスの機能・特徴に応じて文字列を連結することにより、最適なユーザビリティ（可用性）を利用者に提供することができる。

【００６７】

以上、図１から図１３において説明したように、字幕設定装置１０は、音声認識した文字データに対して表示タイミングを設定する場合には、細分認識行処理により文字データを１から３単語程度の長さの認識データ毎に分割して、内容データ２０内の文字列との一致を判断する。これに対して、字幕行を表示する場合には、表示部等の特徴に基づいてこれらの認識データを適切に連結する。即ち、表示タイミングの設定と、字幕行生成とでは、異なるサイズのデータを処理対象とする。これにより、双方の処理を効率的に行うことができる。また、字幕設定装置１０は、音声認識の結果を用いて表示タイミングの設定を行うので、音声認識技術の進歩に伴い、表示タイミングの設定精度を向上させることができる。

【００６８】

本実施例における字幕設定装置１０により行った実験結果を次に示す。本実験において、字幕設定装置１０は、アドリブに基づく３２行分の音声と、台本の定められた８６行分の音声とを入力とする。ＤＰマッチングによって、全ての行の一致が判断された結果、そのうち１２％の文においてタイムスタンプに誤りが生じている。そして、文字列検出部１２０により６６行分の文字列が検出され、文字検出部１３０により３６行文の文字が検出された結果、タイムスタンプの誤りは一切生じていない。音素検出部１５０により６行分の文字の音素が検出された結果、２％の文字においてタイムスタンプに誤りが生じている。更に、音素検出部１５０により強制割り振りが処理された結果、４％の文字においてタイムスタンプに誤りが生じている。このように、従来効率が高いアルゴリズムとして知られているＤＰマッチングと比較して、より高い精度で表示タイミングを設定することができる。

【００６９】

また、この実験において、表示設定部１７０は、文字検出部１３０により検出した文字を含む文字列のうち２つを、他の文字列に連結して表示させ、音素検出部１５０により音素が検出された文字を含む文字列のうち３つを、他の文字列に連結して表示させた。更に、強制割り振りにより検出された文字を含む文字列のうち１０の文字列を、他の文字列に連結して表示させた。このように、アップスケーリング処理の段階が進むのに応じて低い信頼度を生成することにより、誤っている可能性の高いタイムスタンプを有する文字列を

、他の文字列に連結して表示させる。この結果、字幕行の表示タイミングの精度を高めることができる。

【００７０】

図１４は、字幕設定装置１０として機能するコンピュータのハードウェア構成の一例を示す。字幕設定装置１０は、ホストコントローラ１４８２により相互に接続されるＣＰＵ１４００、ＲＡＭ１４２０、グラフィックコントローラ１４７５、及び表示装置１４８０を有するＣＰＵ周辺部と、入出力コントローラ１４８４によりホストコントローラ１４８２に接続される通信インターフェイス１４３０、ハードディスクドライブ１４４０、及びＣＤ－ＲＯＭドライブ１４６０を有する入出力部と、入出力コントローラ１４８４に接続されるＲＯＭ１４１０、フレキシブルディスクドライブ１４５０、及び入出力チップ１４７０を有するレガシー入出力部とを備える。

【００７１】

ホストコントローラ１４８２は、ＲＡＭ１４２０と、高い転送レートでＲＡＭ１４２０をアクセスするＣＰＵ１４００及びグラフィックコントローラ１４７５とを接続する。ＣＰＵ１４００は、ＲＯＭ１４１０及びＲＡＭ１４２０に格納されたプログラムに基づいて動作し、各部の制御を行う。グラフィックコントローラ１４７５は、ＣＰＵ１４００等がＲＡＭ１４２０内に設けたフレームバッファ上に生成する画像データを取得し、表示装置１４８０上に表示させる。これに代えて、グラフィックコントローラ１４７５は、ＣＰＵ１４００等が生成する画像データを格納するフレームバッファを、内部に含んでもよい。

【００７２】

入出力コントローラ１４８４は、ホストコントローラ１４８２と、比較的高速な入出力装置である通信インターフェイス１４３０、ハードディスクドライブ１４４０、及びＣＤ－ＲＯＭドライブ１４６０を接続する。通信インターフェイス１４３０は、ネットワークを介して外部の装置と通信する。ハードディスクドライブ１４４０は、字幕設定装置１０が使用するプログラム及びデータを格納する。ＣＤ－ＲＯＭドライブ１４６０は、ＣＤ－ＲＯＭ１４９５からプログラム又はデータを読み取り、ＲＡＭ１４２０を介して入出力チップ１４７０に提供する。

【００７３】

また、入出力コントローラ１４８４には、ＲＯＭ１４１０と、フレキシブルディスクドライブ１４５０や入出力チップ１４７０等の比較的低速な入出力装置とが接続される。ＲＯＭ１４１０は、字幕設定装置１０の起動時にＣＰＵ１４００が実行するブートプログラムや、字幕設定装置１０のハードウェアに依存するプログラム等を格納する。フレキシブルディスクドライブ１４５０は、フレキシブルディスク１４９０からプログラム又はデータを読み取り、ＲＡＭ１４２０を介して入出力チップ１４７０に提供する。入出力チップ１４７０は、フレキシブルディスク１４９０や、例えばパラレルポート、シリアルポート、キーボードポート、マウスポート等を介して各種の入出力装置を接続する。

【００７４】

字幕設定装置１０に提供されるプログラムは、フレキシブルディスク１４９０、ＣＤ－ＲＯＭ１４９５、又はＩＣカード等の記録媒体に格納されて利用者によって提供される。プログラムは、入出力チップ１４７０及び/又は入出力コントローラ１４８４を介して、記録媒体から読み出され字幕設定装置１０にインストールされて実行される。

【００７５】

字幕設定装置１０にインストールされて実行されるプログラムは、内容データ取得モジュールと、音声認識モジュールと、文字列検出モジュールと、文字検出モジュールと、読付モジュールと、音素検出モジュールと、信頼度算出モジュールと、信頼度取得モジュールと、表示設定モジュールとを含む。各モジュールが字幕設定装置１０に働きかけて行わせる動作は、図１から図１３において説明した字幕設定装置１０における、対応する部材の動作と同一であるから、説明を省略する。

【００７６】

以上に示したプログラム又はモジュールは、外部の記憶媒体に格納されてもよい。記憶

媒体としては、フレキシブルディスク１４９０、ＣＤ－ＲＯＭ１４９５の他に、ＤＶＤやＰＤ等の光学記録媒体、ＭＤ等の光磁気記録媒体、テープ媒体、ＩＣカード等の半導体メモリ等を用いることができる。また、専用通信ネットワークやインターネットに接続されたサーバシステムに設けたハードディスク又はＲＡＭ等の記憶装置を記録媒体として使用し、ネットワークを介してプログラムを字幕設定装置１０に提供してもよい。

【００７７】

図１５は、文字列検出部１２０による処理の一例を説明する図である。文字列検出部１２０は、第１の認識データである「取り巻く」及び第３の認識データである「変化いたして」の各々に一致する文字列を内容データ２０から検出する。一方、文字列検出部１２０は、第１及び第３の認識データの間の認識データである「緩急は著しく」に完全に一致する文字列を検出できない。このような場合には、文字列検出部１２０は、挟み打ち処理により、認識データである「緩急は著しく」に一致する文字列として、内容データ２０における「取り巻く」及び「変化いたして」の間の文字列である「環境は著しく」を検出することができる。同様に、文字列検出部１２０は、認識データである「流れをお洗い」に一致する文字列として、「流れをおさらい」を検出することができる。

【００７８】

図１６は、音素検出部１５０による第１の処理の一例を示す。音素検出部１５０は、文字検出部１３０により一致する文字が検出されなかった各文字を、その文字の読みに含まれる音素に展開する。本例において、検出対象の文字が日本語であるので、音素検出部１５０は、漢字及び仮名の混じった文字列を、その文字列の読みを示す平仮名に展開する。即ち、音素検出部１５０は、「金ねうん」及び「愛ティーを」を、「きんねうん」及び「あいてぃーを」に展開する。

【００７９】

一方、読付部１４０は、内容データである「近年ＩＴを」の読み方の候補の１つとして、「きんねんあいてぃーを」を生成する。この結果、文字列検出部１２０は、内容データにおいて、認識データである「愛ティーを」に一致する文字列として、「愛ティーを」の音素に一致する音素を含む文字列である「ＩＴを」を検出することができる。更に、この検出結果に基づいて、文字列検出部１２０は、認識データである「金ねうん」に一致する文字列として、「近年」を検出してもよい。

【００８０】

図１７は、音素検出部１５０による第２の処理の一例を示す。音素検出部１５０は、文字検出部１３０により一致する文字が検出されなかった各文字を、その文字の読みに含まれる音素に展開する。本例において、検出対象の文字が日本語であるので、音素検出部１５０は、漢字及び仮名の混じった文字列を、その文字列の読みを示す平仮名に展開する。即ち、音素検出部１５０は、「逆しすせ」及び「五人を」を、「ぎゃくしすせ」及び「ごにんを」に展開する。なお、本例においては、図１６と比較して音声認識処理による認識率が低い。

【００８１】

図１６と同様に、読付部１４０は、内容データである「近年ＩＴを」の読み方の候補の１つとして、「きんねんあいてぃーを」を生成する。しかしながら、音素検出部１５０は、「ぎゃくしすせ」及び「ごにんを」に音素が一致する文字又は文字列を、「きんねんあいてぃーを」の中から検出することができない。

【００８２】

この場合、音素検出部１５０は、音声の内容に関わらず音声の長さ又は文字の長さに基づいて一致する音素を検出する強制割り振り処理を行う。例えば、「ぎゃくしすせ」として認識された音声の再生時間と、「ごにんを」として認識された音声の再生時間との比率に基づいて、「ぎゃくしすせ」の音素が、「きんねん」の音素に一致すると判断してもよいし、「ごにんを」の音素が、「あいてぃーを」の音素に一意すると判断してもよい。

【００８３】

図１８は、本実施例による処理の概要をまとめた図である。内容データ取得部１００は

、内容データ２０、例えば、「アクセシビリティについて」という文字列を取得する。音声認識部１１０は、音声認識処理により認識データ３０－１～３、例えば「汗しびれ」、「地位」、及び「について」を生成する。文字列検出部１２０は、まず、ＤＰマッチングにより、「アクセシビリティについて」及び「汗しびれ地位について」を比較して、認識データ３０－１～３の各々に含まれる各文字に一致する文字を内容データ２０から検出する。

【００８４】

この結果、文字列検出部１２０は、認識データ３０－１に一致する文字列として「アクセシビ」を検出し、認識データ３０－２に一致する文字列として「リティ」を検出する。この検出結果に基づいてそのまま字幕を作成した場合には、２つの問題がある。１つ目の問題は、「地位」として音声認識された時間に表示するべき文字列が「ティ」であるにも関わらず、「リティ」が表示されてしまうことである。即ち、文字列「リティ」を表示するべき時間を示すタイムスタンプが誤っている。

【００８５】

２つ目の問題は、「アクセシビリティ」という１つの単語が、音声認識処理の辞書などに登録されていないので、２つの文字列「アクセシビ」及び「リティ」に分割されて検出されていることである。これにより、字幕行において「アクセシビ」及び「リティ」の間で改行される恐れがある。

【００８６】

文字列検出部１２０は、認識データ３０－１～３の各々に一致する文字列を内容データ２０から再度検出する。そして、信頼度算出部１６０は、認識データ３０－１～３の各々が内容データ２０における文字列を検出する確度である信頼度を算出する。この結果、信頼度算出部１６０は、認識データ３０－１である「汗しいびれ」に対応付けて、認識データ３０－２である「地位」と比較して高い信頼度を算出する。

【００８７】

表示設定部１７０は、字幕の１行に表示可能な文字数が１０文字である旨の設定情報を取得する。この場合、表示設定部１７０は、「アクセシビ」及び「リティ」を連結して「アクセシビリティ」を生成するが、「アクセシビリティ」を「について」に連結しない。その結果、表示設定部１７０は、表示タイミング情報４０として、「アクセシビリティ」及び「について」の各々を、「４１．５」及び「５０．５」等の所定の時刻に表示すべき旨の情報を生成することができる。

【００８８】

以上、本発明を実施の形態を用いて説明したが、本発明の技術的範囲は上記実施の形態に記載の範囲には限定されない。上記実施の形態に、多様な変更または改良を加えることが可能であることが当業者に明らかである。その様な変更または改良を加えた形態も本発明の技術的範囲に含まれ得ることが、特許請求の範囲の記載から明らかである。

【００８９】

以上に示す実施例によると、以下の各項目に示す設定装置、プログラム、記録媒体、及び設定方法が実現される。

（項目１）　内容が予め定められた音声の再生に同期して前記内容を表示する表示タイミングを設定する設定装置であって、前記音声の内容を示す内容データを取得する内容データ取得部と、再生される前記音声を音声認識した文字データを分割して複数の認識データを生成する音声認識部と、前記複数の認識データの各々に一致する文字列を前記内容データから検出する文字列検出部と、前記文字列検出部により一致する文字列が検出されなかった各認識データについて、当該認識データに含まれる各文字に一致する文字を前記内容データから検出することにより、当該認識データに一致する文字列を前記内容データから検出する文字検出部と、前記内容データに含まれる文字列の各々を表示させる表示タイミングを、当該文字列に一致する認識データとして音声認識された音声の再生時に設定する表示設定部とを備える設定装置。

（項目２）　前記認識データのうち前記文字検出部により一致する文字が検出されなかっ

た各文字について、当該文字の読みに含まれる音素に一致する音素を、前記内容データの読みの中から検出する音素検出部を更に備え、前記文字検出部は、前記認識データのうち前記音素検出部により一致する音素が検出された文字に一致する文字として、前記内容データにおいて当該音素を含む文字を検出する項目１記載の設定装置。

【００９０】

（項目３）　前記内容データの読み方の候補である読み候補を複数生成する読付部を更に備え、前記音素検出部は、前記認識データのうち前記文字検出部により前記内容データに一致する文字が検出されなかった各文字について、当該文字の読みに含まれる音素に一致する音素を、前記読付部により生成された複数の前記読み候補の何れかの中から検出する項目２記載の設定装置。

（項目４）　前記読付部は、前記内容データにおける複数の前記読み候補の各々を、当該読み候補により読まれる可能性を示す情報に対応付けて生成し、前記音素検出部は、前記認識データに含まれる文字の読みに含まれる音素を、読まれる可能性が高い順に、前記複数の読み候補の各々と比較する項目３記載の設定装置。

（項目５）　前記複数の認識データの各々が文字列に一致する確度である信頼度を算出する信頼度算出部を更に備え、前記文字列検出部は、予め定められた基準信頼度未満の信頼度の認識データである低信頼データについて、当該低信頼データに後続する認識データに一致する文字列を検出できなかった場合に、当該低信頼データに一致する文字列は検出できないと判断する項目１記載の設定装置。

（項目６）　前記複数の認識データの各々が文字列に一致する確度である信頼度を算出する信頼度算出部を更に備え、前記表示設定部は、前記内容データにおける複数の前記文字列のうち連続した２つの文字列について、先に表示すべき文字列に対応する信頼度が、後に表示すべき文字列に対応する信頼度より高い場合に、先に表示すべき前記文字列の末尾に後に表示すべき前記文字列を連結した文字列を、先に表示すべき前記文字列を表示すべき時刻に表示させる設定を行う項目１記載の設定装置。

【００９１】

（項目７）　前記信頼度算出部は、前記文字列検出部により一致する文字列が検出された認識データに対応付けて、前記文字検出部により一致する文字列が検出された認識データと比較して、より高い信頼度を算出する項目６記載の設定装置。

（項目８）　前記認識データのうち前記文字検出部により一致する文字が検出されなかった各文字について、当該文字の読みに含まれる音素に一致する音素を、前記内容データの読みの中から検出する音素検出部を更に備え、前記文字検出部は、前記認識データのうち前記音素検出部により一致する音素が検出された文字に一致する文字として、前記内容データにおいて当該音素を含む文字を検出し、前記信頼度算出部は、前記音素検出部により一致する音素が検出された文字を含む認識データに対応付けて、前記音素検出部により一致する音素が検出されることなく前記文字検出部により一致する文字が検出された認識データと比較して、より低い信頼度を生成する項目６記載の設定装置。

（項目９）　前記音声認識部は、音声認識した前記複数の認識データが、再生される音声の内容と一致する可能性を示す音声認識確信度を、認識データ毎に更に生成し、前記文字列検出部は、音声認識確信度がより高い認識データに一致する文字列を、当該認識データと比較して音声認識確信度が低い認識データに先立って検出し、第１の前記認識データに一致する第１の文字列及び第２の前記認識データに一致する第２の文字列を検出した場合に、前記第１の認識データに後続しかつ前記第２の認識データに先行する認識データに一致する文字列として、前記第１の文字列に後続し前記第２の文字列に先行する文字列を検出する項目１記載の設定装置。

【００９２】

（項目１０）　前記表示設定部は、前記文字列検出部により前記内容データにおいて一致する文字列が検出されなかった認識データを、当該認識データとして音声認識された音声の再生時に表示させる設定を行う項目１記載の設定装置。

（項目１１）　内容が予め定められた音声の再生に同期して前記内容を表示する表示タイ

ミングを設定する設定装置であって、再生される前記音声の内容を示す内容データに含まれる複数の文字列の各々に対応付けて、当該文字列を表示すべき時刻、及び、当該時刻に当該文字列を内容とする音声が再生される確度である信頼度を取得する信頼度取得部と、前記複数の文字列のうち連続した２つの文字列について、先に表示すべき文字列に対応する信頼度が、後に表示すべき文字列に対応する信頼度より高い場合に、先に表示すべき前記文字列の末尾に後に表示すべき前記文字列を連結した文字列を、先に表示すべき前記文字列を表示すべき時刻に表示させる設定を行う表示設定部とを備える設定装置。

（項目１２）　前記表示設定部は、先に表示すべき前記文字列に対応する信頼度が、後に表示すべき前記文字列に後続する後続文字列に対応する信頼度より高い場合に、連結した前記文字列の末尾に前記後続文字列を更に連結した文字列を、先に表示すべき前記文字列を表示すべき時刻に表示させる設定を行う項目１１記載の設定装置。

【００９３】

（項目１３）　内容が予め定められた音声の再生に同期して前記内容を表示する表示タイミングを設定する設定装置として、コンピュータを機能させるプログラムであって、前記コンピュータを、前記音声の内容を示す内容データを取得する内容データ取得部と、再生される前記音声を音声認識した文字データを分割して複数の認識データを生成する音声認識部と、前記複数の認識データの各々に一致する文字列を前記内容データから検出する文字列検出部と、前記文字列検出部により一致する文字列が検出されなかった各認識データについて、当該認識データに含まれる各文字に一致する文字を前記内容データから検出することにより、当該認識データに一致する文字列を前記内容データから検出する文字検出部と、前記内容データに含まれる文字列の各々を表示させる表示タイミングを、当該文字列に一致する認識データとして音声認識された音声の再生時に設定する表示設定部として機能させるプログラム。

（項目１４）　内容が予め定められた音声の再生に同期して前記内容を表示する表示タイミングを設定する設定装置として、コンピュータを機能させるプログラムであって、前記コンピュータを、再生される前記音声の内容を示す内容データに含まれる複数の文字列の各々に対応付けて、当該文字列を表示すべき時刻、及び、当該時刻に当該文字列を内容とする音声が再生される確度である信頼度を取得する信頼度取得部と、前記複数の文字列のうち連続した２つの文字列について、先に表示すべき文字列に対応する信頼度が、後に表示すべき文字列に対応する信頼度より高い場合に、先に表示すべき前記文字列の末尾に後に表示すべき前記文字列を連結した文字列を、先に表示すべき前記文字列を表示すべき時刻に表示させる設定を行う表示設定部として機能させるプログラム。

（項目１５）　項目１３又は項目１４に記載のプログラムを記録した記録媒体。

【００９４】

（項目１６）　内容が予め定められた音声の再生に同期して前記内容を表示する表示タイミングを設定する設定方法であって、コンピュータにより、前記音声の内容を示す内容データを取得する内容データ取得段階と、再生される前記音声を音声認識した文字データを分割して複数の認識データを生成する音声認識段階と、前記複数の認識データの各々に一致する文字列を前記内容データから検出する文字列検出段階と、前記文字列検出段階において一致する文字列が検出されなかった各認識データについて、当該認識データに含まれる各文字に一致する文字を前記内容データから検出することにより、当該認識データに一致する文字列を前記内容データから検出する文字検出段階と、前記内容データに含まれる文字列の各々を表示させる表示タイミングを、当該文字列に一致する認識データとして音声認識された音声の再生時に設定する表示設定段階とを備える設定方法。

（項目１７）　内容が予め定められた音声の再生に同期して前記内容を表示する表示タイミングを設定する設定方法であって、コンピュータにより、再生される前記音声の内容を示す内容データに含まれる複数の文字列の各々に対応付けて、当該文字列を表示すべき時刻、及び、当該時刻に当該文字列を内容とする音声が再生される確度である信頼度を取得する信頼度取得段階と、前記複数の文字列のうち連続した２つの文字列について、先に表示すべき文字列に対応する信頼度が、後に表示すべき文字列に対応する信頼度より高い場

合に、先に表示すべき前記文字列の末尾に後に表示すべき前記文字列を連結した文字列を、先に表示すべき前記文字列を表示すべき時刻に表示させる設定を行う表示設定段階とを備える設定方法。

【図面の簡単な説明】

【００９５】

【図１】図１は、字幕設定装置１０の機能ブロック図である。

【図２】図２は、内容データ２０の一例を示す。

【図３】図３は、認識データ３０－１～Ｎの一例を示す。

【図４】図４は、表示タイミング情報４０の一例を示す。

【図５】図５は、字幕設定装置１０が表示タイミングを設定する処理の動作フローを示す。

【図６】図６は、Ｓ５２０における処理の詳細を示す。

【図７】図７は、Ｓ６２０における処理の詳細を示す。

【図８】図８は、Ｓ６４０における処理の詳細を示す。

【図９】図９は、Ｓ８１０における第１の処理の詳細を示す。

【図１０】図１０は、Ｓ８１０における第２の処理の詳細を示す。

【図１１】図１１は、Ｓ８３０における第１の処理の詳細を示す。

【図１２】図１２は、Ｓ８３０における第２の処理の詳細を示す。

【図１３】図１３は、Ｓ５５０における処理の詳細を示す。

【図１４】図１４は、字幕設定装置１０として機能するコンピュータのハードウェア構成の一例を示す。

【図１５】図１５は、文字列検出部１２０による処理の一例を説明する図である。

【図１６】図１６は、音素検出部１５０による第１の処理の一例を示す。

【図１７】図１７は、音素検出部１５０による第２の処理の一例を示す。

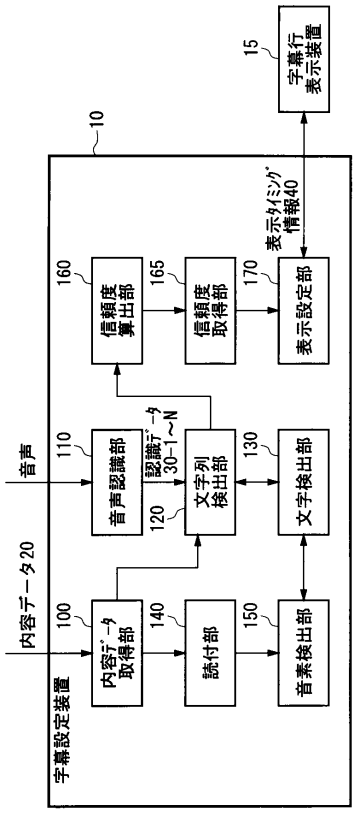
【図１８】図１８は、本実施例による処理の概要をまとめた図である。

【符号の説明】

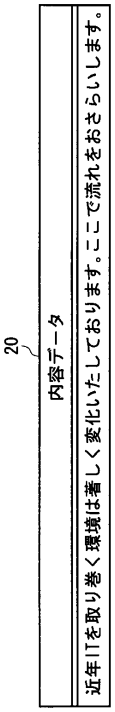
【００９６】

１０　字幕設定装置
１５　字幕行表示装置
２０　内容データ
３０　認識データ
４０　表示タイミング情報
１００　内容データ取得部
１１０　音声認識部
１２０　文字列検出部
１３０　文字検出部
１４０　読付部
１５０　音素検出部
１６０　信頼度算出部
１６５　信頼度取得部
１７０　表示設定部

【図１】



【図２】



【図３】

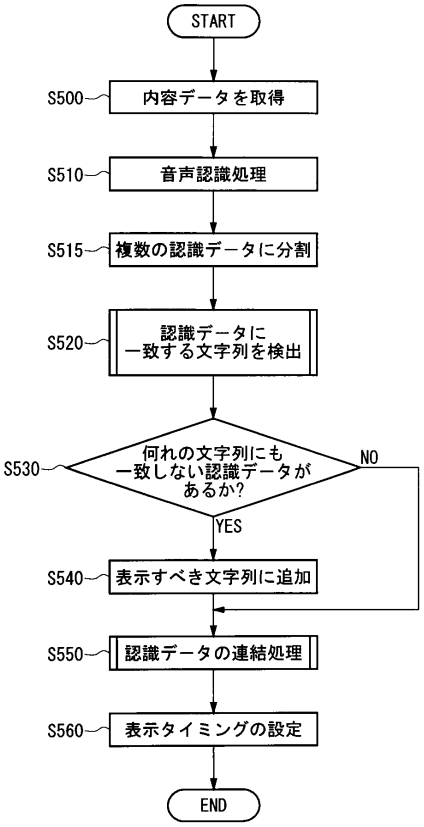
認識時刻	認識データ	
02.103	金ねうん	30-1
03.211	愛ティーを	30-2
05.235	取り巻く	30-3
06.456	緩急は	30-4
08.387	著しく変化いたして	30-5
11.125	おります。	30-6
16.887	ここで	30-7
17.001	流れをお洗い	30-8
18.051	致します。	30-9
・・・	・・・	
XX	XX	30-N

【図４】

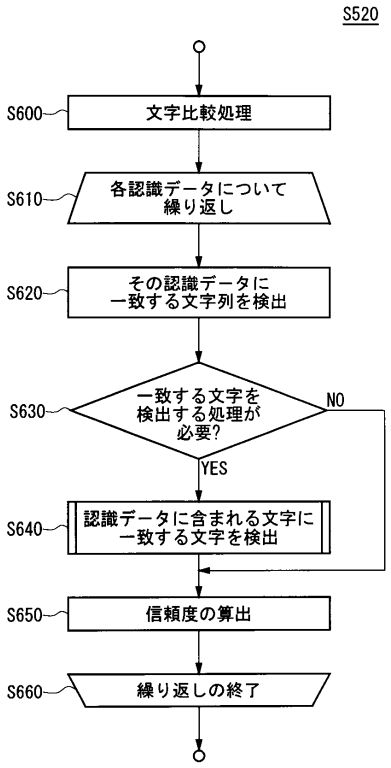
40

表示時刻	文字列
02.103	近年ITを取り巻く環境は
08.387	著しく変化いたしております。
16.887	ここで流れをおさらいします。

【図５】



【図６】

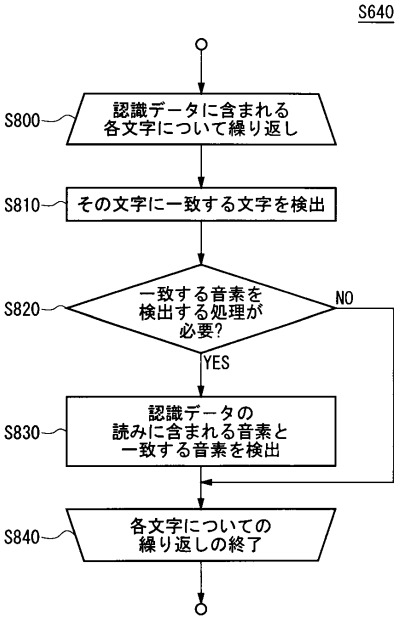


【図７】

S620

2つ前の認識データ	1つ前の認識データ	対象の認識データ	処理内容
○/×	○	○	対象の認識データを一致と判断
		×	保留
○	×	○	1つ前の認識データを一致と判断
		×	保留,但し最後の認識データならば,末尾確定文字比較
×		○	2つ前,1つ前の認識データに含まれる文字と一致する文字を検出(範囲確定) 信頼度が基準信頼度未満ならば,再評価フラグ(△)
		×	2つ前,1つ前の認識データに含まれる文字と一致する文字を検出(範囲不確定)
×	△	○	1つ前の認識データを一致と判断
		×	1つ前の認識データを一致しないと判断

【図８】



【図９】

S810

先頭文字	前の文字	対象文字	処理内容
○/×	○	○	対象文字を一致と判断
		×	保留
○	×	○	認識データ内の不一致文字を一致と判断
		×	保留
×		○	対象文字を一致と判断
		×	保留

【図１０】

S810

先頭文字	次の認識データ	処理内容
○	○	対象の認識データを一致と判断
	×	先頭文字確定、次の認識データ不確定音素検出処理
×	○	先頭文字不確定、次の認識データ確定音素検出処理
	×	先頭文字不確定、次の認識データ不確定音素検出処理

【図１１】

S830

先頭文字	前の音素	対象音素	処理内容
○/×	○	○	音素一致と判断
		×	保留、又は、次の読み候補
○	×	○	認識データ内の不一致文字を一致と判断
		×	保留、又は、次の読み候補
×		○	音素一致と判断
		×	保留、又は、次の読み候補

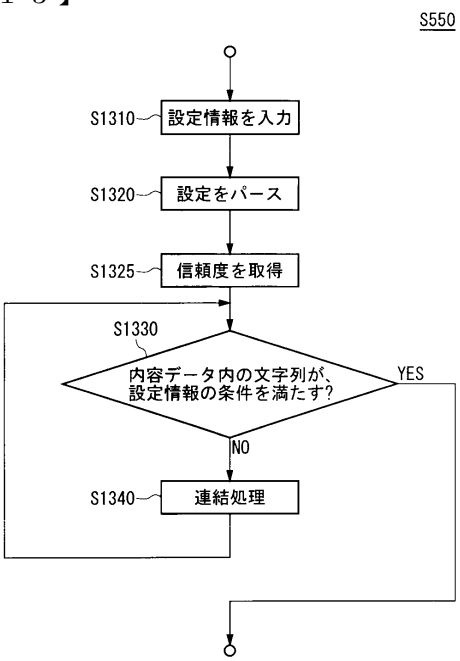
【図１２】

S830

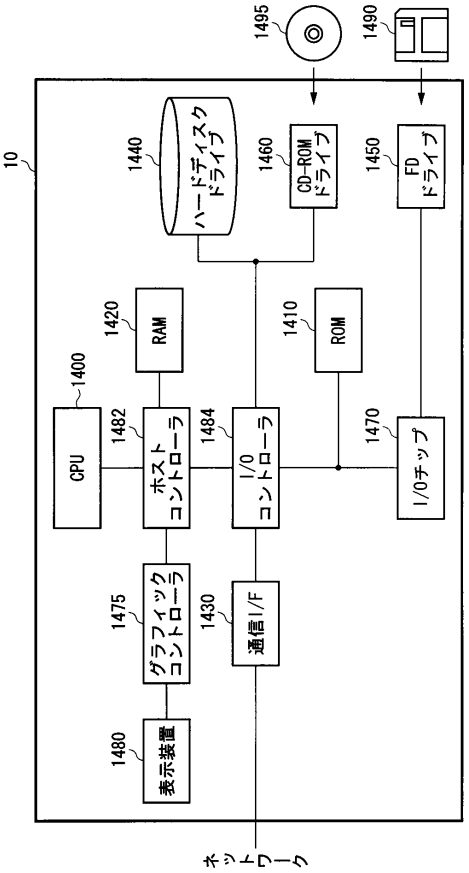
先頭文字	次の認識データ、又は最後の認識データ	処理内容
○	○	認識データ内の不一致文字を一致と判断
	×	保留
×	○	強制割り振り
	×	保留

【図１３】

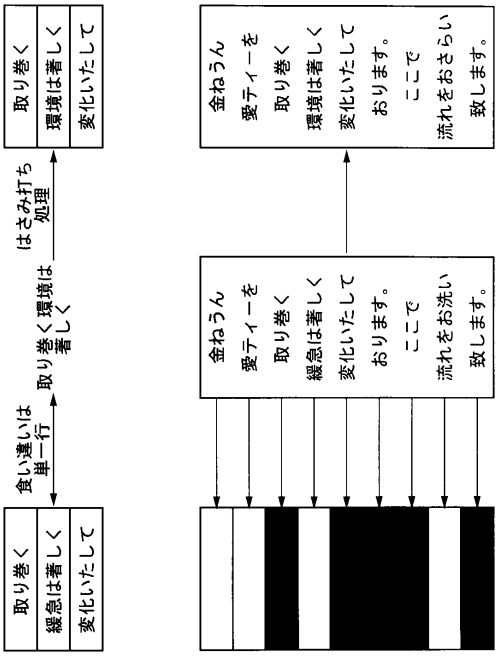
S550



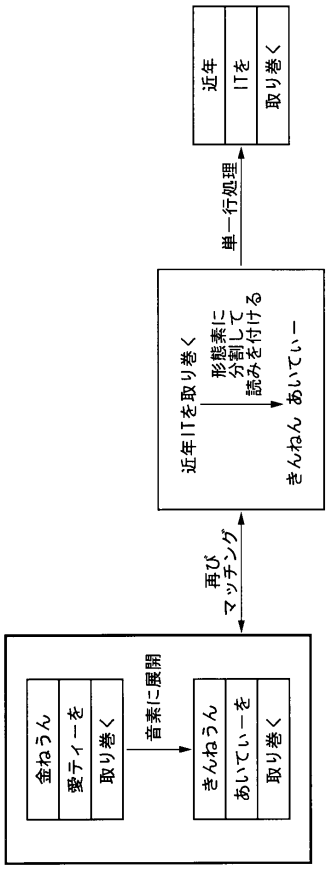
【図１４】



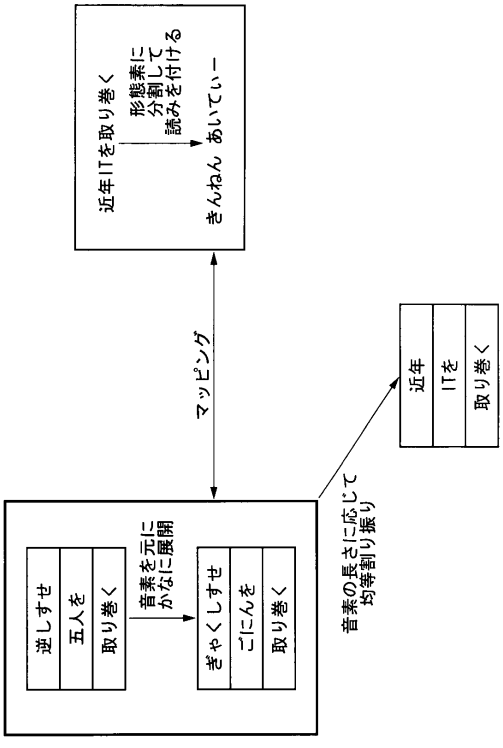
【図１５】



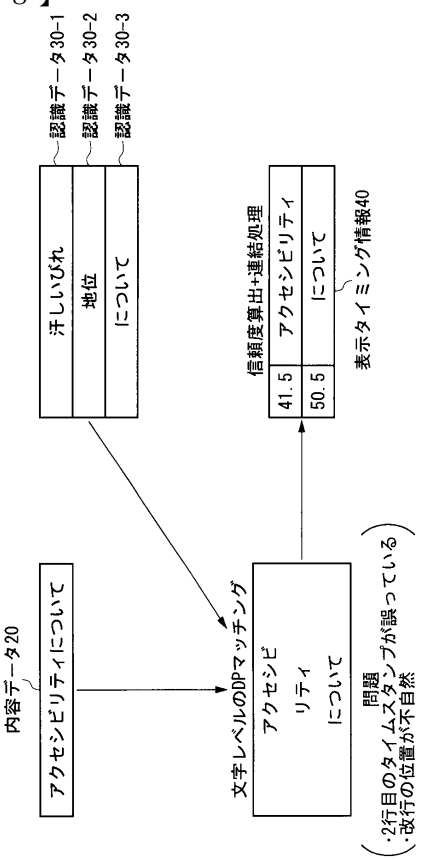
【図１６】



【図１７】



【図１８】


認識データ30-1
認識データ30-2
認識データ30-3
汗しいびれ
地位
について
信頼度算出+連結処理
41.5
アクセシビリティ
50.5
について
表示タイミング情報40
内容データ20
アクセシビリティについて
文字レベルのDPマッチング
アクセシビ
リティ
について
問題
・2行目のタイムスタンプが誤っている
・改行の位置が不自然

フロントページの続き

(72)発明者　宮本　晃太郎
神奈川県大和市下鶴間１６２３番地１４　日本アイ・ビー・エム株式会社　東京基礎研究所内
(72)発明者　東海林　みどり
神奈川県大和市下鶴間１６２３番地１４　日本アイ・ビー・エム株式会社　東京基礎研究所内

審査官　樫本　剛

(56)参考文献　特開平１０－３０８８８７（ＪＰ，Ａ）
特開２００１－０３４１５１（ＪＰ，Ａ）
特開２００５－０４５５０３（ＪＰ，Ａ）
丸山他，字幕送出タイミング検出におけるワード列ペアモデルの構成検討，日本音響学会平成１０年度秋季研究発表会講演論文集，日本，１９９８年　6月２４日，Vol.1，p.25-26

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ１０Ｌ　１５／００－１５／２８
ＪＳＴＰｌｕｓ（ＪＤｒｅａｍ2）
ＩＥＥＥ