

【特許請求の範囲】**【請求項 1】**

音声に合わせて字幕を出力する字幕出力装置であって、
入力されたテキスト文を字幕の出力単位に分割することにより、字幕単位文を生成する字幕単位文生成手段と、
前記テキスト文を音声認識の処理単位に分割することにより、音声認識単位文を生成する音声認識単位文生成手段と、
前記音声認識単位文生成手段により生成された音声認識単位文の文節を音声認識するための認識候補の集合である認識候補単位を、前記音声認識単位文の先頭の文節に対応するものから順に連結することにより、音声認識ネットワークを生成する音声認識ネットワーク生成手段と、
前記テキスト文が発声された音声と、前記音声認識ネットワーク生成手段により生成された音声認識ネットワークを構成する認識候補単位との照合を先頭から逐次行うことにより、音声認識処理を行う音声認識手段と、
前記字幕単位文生成手段により生成された字幕単位文の少なくとも先頭の文節に対応する認識候補単位との前記照合が完了した時点で、前記字幕単位文を出力する字幕単位文出力手段と
を備えたことを特徴とする字幕出力装置。

10

【請求項 2】

前記字幕単位文出力手段は、
前記字幕単位文の先頭の文節に対応する認識候補単位を含む所定数の認識候補単位との前記照合が完了した時点で、前記字幕単位文を出力することを特徴とする請求項 1 に記載の字幕出力装置。

20

【請求項 3】

前記音声認識処理手段は、
前記音声認識処理を、前記音声認識ネットワーク生成手段により生成された 2 以上の音声認識ネットワークを用いて並列に行うことを特徴とする請求項 1 又は 2 に記載の字幕出力装置。

【請求項 4】

前記音声認識ネットワーク生成手段は、
前記字幕単位文の先頭の文節に対応する認識候補単位を少なくとも含む字幕先頭検出用ネットワークを生成する字幕先頭検出用ネットワーク生成手段を備え、
前記字幕先頭検出用ネットワーク生成手段は、
前記字幕先頭検出用ネットワークと、該字幕先頭検出用ネットワークと並列に音声認識処理を行う対象となる音声認識ネットワークと、の間のネットワーク間距離が所定の閾値以上となるように、前記字幕先頭検出用ネットワークを生成し、
前記字幕単位文出力手段は、
前記字幕先頭検出用ネットワークを構成する全ての認識候補単位との前記照合が完了した時点で、前記字幕単位文を出力することを特徴とする請求項 3 に記載の字幕出力装置。

30

【請求項 5】

前記音声認識ネットワーク生成手段は、
前記連結された各認識候補単位間に、誤認識を防ぐための特殊認識候補を挿入した上で、前記音声認識ネットワークを生成することを特徴とする請求項 1 から 4 の何れか 1 項に記載の字幕出力装置。

40

【請求項 6】

前記音声認識ネットワーク生成手段は、
前記認識候補単位に、誤認識を防ぐための特殊認識候補を含めた上で、前記音声認識ネットワークを生成することを特徴とする請求項 1 から 5 の何れか 1 項に記載の字幕出力装置。

【請求項 7】

50

前記特殊認識候補には、ポーズがないことを表す N U L L と、無音のポーズがあることを表す S I L と、任意の音を表す G a r b a g e と、の少なくとも 1 つが含まれることを特徴とする請求項 5 又は 6 に記載の字幕出力装置。

【請求項 8】

音声に合わせて字幕を出力する字幕出力装置であって、

入力されたテキスト文を字幕の出力単位に分割することにより、字幕単位文を生成する字幕単位文生成手段と、

前記テキスト文を音声認識の処理単位に分割することにより、音声認識単位文を生成する音声認識単位文生成手段と、

前記音声認識単位文生成手段により生成された音声認識単位文の文節を音声認識するための認識候補の集合である認識候補単位を、前記音声認識単位文の先頭の文節に対応するものから順に連結することにより、音声認識ネットワークを生成する音声認識ネットワーク生成手段と、

前記テキスト文が発声された音声の音声認識処理を、前記音声認識ネットワーク生成手段により生成された音声認識ネットワークのうち 2 以上の音声認識ネットワークを用いて、並列に行う音声認識手段と、

前記音声認識手段による音声認識処理の結果に基づいて字幕単位文を出力する字幕単位文出力手段と

を備えたことを特徴とする字幕出力装置。

【請求項 9】

音声に合わせて字幕を出力する字幕出力装置が実行する字幕出力方法であって、

入力されたテキスト文を字幕の出力単位に分割することにより、字幕単位文を生成する字幕単位文生成ステップと、

前記テキスト文を音声認識の処理単位に分割することにより、音声認識単位文を生成する音声認識単位文生成ステップと、

前記音声認識単位文生成ステップにおいて生成された音声認識単位文の文節を音声認識するための認識候補の集合である認識候補単位を、前記音声認識単位文の先頭の文節に対応するものから順に連結することにより、音声認識ネットワークを生成する音声認識ネットワーク生成ステップと、

前記テキスト文が発声された音声と、前記音声認識ネットワーク生成ステップにおいて生成された音声認識ネットワークを構成する認識候補単位との照合を先頭から逐次行うことにより、音声認識処理を行う音声認識ステップと、

前記字幕単位文生成ステップにおいて生成された字幕単位文の少なくとも先頭の文節に対応する認識候補単位との前記照合が完了した時点で、前記字幕単位文を出力する字幕単位文出力ステップと

を備えたことを特徴とする字幕出力方法。

【請求項 10】

音声に合わせて字幕を出力する字幕出力装置が実行する字幕出力方法であって、

入力されたテキスト文を字幕の出力単位に分割することにより、字幕単位文を生成する字幕単位文生成ステップと、

前記テキスト文を音声認識の処理単位に分割することにより、音声認識単位文を生成する音声認識単位文生成ステップと、

前記音声認識単位文生成ステップにおいて生成された音声認識単位文の文節を音声認識するための認識候補の集合である認識候補単位を、前記音声認識単位文の先頭の文節に対応するものから順に連結することにより、音声認識ネットワークを生成する音声認識ネットワーク生成ステップと、

前記テキスト文が発声された音声の音声認識処理を、前記音声認識ネットワーク生成ステップにおいて生成された音声認識ネットワークのうち 2 以上の音声認識ネットワークを用いて、並列に行う音声認識ステップと、

前記音声認識ステップにおける音声認識処理の結果に基づいて字幕単位文を出力する字

10

20

30

40

50

幕単位文出力ステップと

を備えたことを特徴とする字幕出力方法。

【請求項 11】

コンピュータに、

入力されたテキスト文を字幕の出力単位に分割することにより、字幕単位文を生成する字幕単位文生成ステップと、

前記テキスト文を音声認識の処理単位に分割することにより、音声認識単位文を生成する音声認識単位文生成ステップと、

前記音声認識単位文生成ステップにおいて生成された音声認識単位文の文節を音声認識するための認識候補の集合である認識候補単位を、前記音声認識単位文の先頭の文節に対応するものから順に連結することにより、音声認識ネットワークを生成する音声認識ネットワーク生成ステップと、

前記テキスト文が発声された音声と、前記音声認識ネットワーク生成ステップにおいて生成された音声認識ネットワークを構成する認識候補単位との照合を先頭から逐次行うことにより、音声認識処理を行う音声認識ステップと、

前記字幕単位文生成ステップにおいて生成された字幕単位文の少なくとも先頭の文節に対応する認識候補単位との前記照合が完了した時点で、前記字幕単位文を出力する字幕単位文出力ステップと

を実行させるためのプログラム。

【請求項 12】

コンピュータに、

入力されたテキスト文を字幕の出力単位に分割することにより、字幕単位文を生成する字幕単位文生成ステップと、

前記テキスト文を音声認識の処理単位に分割することにより、音声認識単位文を生成する音声認識単位文生成ステップと、

前記音声認識単位文生成ステップにおいて生成された音声認識単位文の文節を音声認識するための認識候補の集合である認識候補単位を、前記音声認識単位文の先頭の文節に対応するものから順に連結することにより、音声認識ネットワークを生成する音声認識ネットワーク生成ステップと、

前記テキスト文が発声された音声の音声認識処理を、前記音声認識ネットワーク生成ステップにおいて生成された音声認識ネットワークのうち 2 以上の音声認識ネットワークを用いて、並列に行う音声認識ステップと、

前記音声認識ステップにおける音声認識処理の結果に基づいて字幕単位文を出力する字幕単位文出力ステップと

を実行させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、共通の原稿をアナウンス用と字幕用との双方に利用してリアルタイム放送を行う場合に、アナウンスの音声に合わせて字幕を出力する字幕出力装置、字幕出力方法及びプログラムに関する。

【背景技術】

【0002】

近年、放送業界ではデジタル放送番組に字幕を表示することが推奨されていることもあり、字幕放送番組が増えつつある。字幕放送番組が録画番組の場合には、予め字幕を付加した放送番組データを作成して記録媒体に記録しておくことができるが、ニュース番組等のリアルタイム放送（生放送）番組である場合には、アナウンサーの発声するタイミングに合わせて字幕をリアルタイムで送出する必要がある。

図 12 は、ニュース等のリアルタイム放送番組において従来一般的に行われている字幕送出の仕組みを示す図である。同図に示すように、アナウンサーがマイクロホンに向かっ

10

20

30

40

50

てニュース原稿を読み上げている時に、字幕担当者がアナウンサーの音声を聞いて字幕の表示タイミングを判断し、表示タイミングとなった時に字幕切替装置のボタンを押す等の操作を行う。これにより、カメラで撮影された映像と、マイクロホンで収集された音声と、字幕切替装置において予め用意されている字幕とが、多重化機で多重化され、通信回線を介して受信機に送出される。

【0003】

このような字幕送出の仕組みでは、字幕担当者の技量に応じて、音声に対する字幕表示タイミングの遅延が少なくとも3～5秒程度発生する。このため、視聴者は音声を聴いてからかなりの間をおいた後に対応する字幕を見ることとなり、違和感を覚えてしまう。また、人手で字幕表示を行うため、操作ミスにより字幕が誤表示される危険性がある。

10

これに対して、ドラマなどの録画番組の場合には、音声と字幕との同期をとったデータを予め作成しておくことができるため、放送時に字幕表示の遅延や誤表示を防ぐことができる(例えば、特許文献1参照)。特許文献1に記載の自動字幕番組制作システムは、テキスト文から提示単位字幕文を生成し、提示単位字幕文毎にアナウンス音声との音声認識を行い、始点/終点タイミング情報を同期点として検出し、当該検出した始点/終点タイミング情報を提示単位字幕文毎に付与しておく。これにより、放送時には、付与したタイミング情報に基づいて音声と字幕との同期をとることが可能となる。

【特許文献1】特開2000-270263号公報

【発明の開示】

【発明が解決しようとする課題】

20

【0004】

特許文献1に記載の録画番組の技術をリアルタイム放送番組に適用した場合には、提示単位字幕文全体とアナウンス音声との音声認識を行い、始点/終点タイミング情報を同期点として検出した後に、提示単位字幕文を送出することとなる。つまり、提示単位字幕文に対応する音声アナウンスが終了してから当該提示単位字幕文が表示されることとなり、原理上、提示単位字幕文単位での大幅な遅延が生じることとなる。

また、特許文献1に記載の技術では、アナウンサーの息継ぎ等の無音区間(ポーズ、間)が想定通りに発生しなかった場合、アナウンサーが原稿を読み間違えた場合、読み飛ばした場合、雑音がいった場合等を想定していないため、音声に対応する提示単位字幕文が正しく認識されない場合が発生する。この場合、録画放送の場合には、放送前に修正することができるが、リアルタイム放送の場合には、修正する間もなく誤った提示単位字幕文が表示されてしまうという不具合が発生する。

30

【0005】

本発明は、上述した従来の問題点に鑑みてなされたものであり、リアルタイム放送において、音声に対して少ない遅延で字幕を出力することができる字幕出力装置、字幕出力方法及びプログラムを提供する。

また、音声に対応した字幕を誤りなく正確に出力することを可能とする字幕出力装置、字幕出力方法及びプログラムを提供する。

【課題を解決するための手段】

【0006】

40

上記問題を解決するために、請求項1に記載の発明は、音声に合わせて字幕を出力する字幕出力装置であって、入力されたテキスト文を字幕の出力単位に分割することにより、字幕単位文を生成する字幕単位文生成手段と、前記テキスト文を音声認識の処理単位に分割することにより、音声認識単位文を生成する音声認識単位文生成手段と、前記音声認識単位文生成手段により生成された音声認識単位文の文節を音声認識するための認識候補の集合である認識候補単位を、前記音声認識単位文の先頭の文節に対応するものから順に連結することにより、音声認識ネットワークを生成する音声認識ネットワーク生成手段と、前記テキスト文が発声された音声と、前記音声認識ネットワーク生成手段により生成された音声認識ネットワークを構成する認識候補単位との照合を先頭から逐次行うことにより、音声認識処理を行う音声認識手段と、前記字幕単位文生成手段により生成された字幕単

50

位文の少なくとも先頭の文節に対応する認識候補単位との前記照合が完了した時点で、前記字幕単位文を出力する字幕単位文出力手段とを備えたことを特徴とする字幕出力装置を提供する。

本発明によれば、字幕出力装置は、字幕単位文の少なくとも先頭の文節に対応する認識候補単位との照合が完了した時点で前記字幕単位文を出力するため、リアルタイム放送において、音声に対して少ない遅延で字幕を出力することが可能となる。

【0007】

請求項2に記載の発明は、請求項1に記載の字幕出力装置において、前記字幕単位文出力手段は、前記字幕単位文の先頭の文節に対応する認識候補単位を含む所定数の認識候補単位との前記照合が完了した時点で、前記字幕単位文を出力することを特徴とする。

10

本発明によれば、字幕単位文の先頭の文節に対応する認識候補単位を含む所定数の認識候補単位との前記照合が完了した時点で、前記字幕単位文を出力するため、リアルタイム放送において、音声に対して少ない遅延で字幕を出力することができる。

請求項3に記載の発明は、請求項1又は2に記載の字幕出力装置において、前記音声認識処理手段は、前記音声認識処理を、前記音声認識ネットワーク生成手段により生成された2以上の音声認識ネットワークを用いて並列に行うことを特徴とする。

本発明によれば、発話者の読み飛ばし等による誤認識を防ぎ、精度の高い音声認識結果に基づいて、音声に対応した字幕を少ない遅延で正確に出力することができる。

【0008】

請求項4に記載の発明は、請求項3に記載の字幕出力装置において、前記音声認識ネットワーク生成手段は、前記字幕単位文の先頭の文節に対応する認識候補単位を少なくとも含む字幕先頭検出用ネットワークを生成する字幕先頭検出用ネットワーク生成手段を備え、前記字幕先頭検出用ネットワーク生成手段は、前記字幕先頭検出用ネットワークと該字幕先頭検出用ネットワークと並列に音声認識処理を行う対象となる音声認識ネットワークとの間のネットワーク間距離が所定の閾値以上となるように、前記字幕先頭検出用ネットワークを生成し、前記字幕単位文出力手段は、前記字幕先頭検出用ネットワークを構成する全ての認識候補単位との前記照合が完了した時点で、前記字幕単位文を出力することを特徴とする。

20

本発明によれば、字幕の先頭文節が発声されたことを正確に検出するための字幕先頭検出用ネットワークを生成して音声認識を行うことで、字幕の出力タイミングの判定を正確かつ容易に行うことができる。

30

【0009】

請求項5に記載の発明は、請求項1から4の何れか1項に記載の字幕出力装置において、前記音声認識ネットワーク生成手段は、前記連結された各認識候補単位間に、誤認識を防ぐための特殊認識候補を挿入した上で、前記音声認識ネットワークを生成することを特徴とする。

本発明によれば、各認識候補文節間に、誤認識を防ぐための特殊認識候補を挿入することで、発話者の息継ぎの違い、読み間違い、言い直し、咳払い、雑音等に影響されずに、正確に音声認識を行うことができる。

【0010】

40

請求項6に記載の発明は、請求項1から5の何れか1項に記載の字幕出力装置において、前記音声認識ネットワーク生成手段は、前記認識候補単位に、誤認識を防ぐための特殊認識候補を含めた上で、前記音声認識ネットワークを生成することを特徴とする。

本発明によれば、認識候補単位に特殊認識候補が含まれることにより、発話者の読み間違い、雑音等に影響されずに、音声認識の誤認識を防ぐことができる。

請求項7に記載の発明は、請求項5又は6に記載の字幕出力装置において、前記特殊認識候補には、ポーズがないことを表すN U L Lと、無音のポーズがあることを表すS I Lと、任意の音を表すG a r b a g eと、の少なくとも1つが含まれることを特徴とする。

【0011】

請求項8に記載の発明は、音声に合わせて字幕を出力する字幕出力装置であって、入力

50

されたテキスト文を字幕の出力単位に分割することにより、字幕単位文を生成する字幕単位文生成手段と、前記テキスト文を音声認識の処理単位に分割することにより、音声認識単位文を生成する音声認識単位文生成手段と、前記音声認識単位文生成手段により生成された音声認識単位文の文節を音声認識するための認識候補の集合である認識候補単位を、前記音声認識単位文の先頭の文節に対応するものから順に連結することにより、音声認識ネットワークを生成する音声認識ネットワーク生成手段と、前記テキスト文が発声された音声の音声認識処理を、前記音声認識ネットワーク生成手段により生成された音声認識ネットワークのうち2以上の音声認識ネットワークを用いて、並列に行う音声認識手段と、前記音声認識手段による音声認識処理の結果に基づいて字幕単位文を出力する字幕単位文出力手段とを備えたことを特徴とする字幕出力装置を提供する。

10

本発明によれば、字幕出力装置は、テキスト文が発声された音声の音声認識処理を、2以上の音声認識ネットワークを用いて並列に行うため、発話者の読み飛ばし等による音声の誤認識を防ぐことができ、音声に対応した字幕を正確に出力することが可能となる。

【0012】

請求項9に記載の発明は、音声に合わせて字幕を出力する字幕出力装置が実行する字幕出力方法であって、入力されたテキスト文を字幕の出力単位に分割することにより、字幕単位文を生成する字幕単位文生成ステップと、前記テキスト文を音声認識の処理単位に分割することにより、音声認識単位文を生成する音声認識単位文生成ステップと、前記音声認識単位文生成ステップにおいて生成された音声認識単位文の文節を音声認識するための認識候補の集合である認識候補単位を、前記音声認識単位文の先頭の文節に対応するものから順に連結することにより、音声認識ネットワークを生成する音声認識ネットワーク生成ステップと、前記テキスト文が発声された音声と、前記音声認識ネットワーク生成ステップにおいて生成された音声認識ネットワークを構成する認識候補単位との照合を先頭から逐次行うことにより、音声認識処理を行う音声認識ステップと、前記字幕単位文生成ステップにおいて生成された字幕単位文の少なくとも先頭の文節に対応する認識候補単位との前記照合が完了した時点で、前記字幕単位文を出力する字幕単位文出力ステップとを備えたことを特徴とする字幕出力方法を提供する。

20

【0013】

請求項10に記載の発明は、音声に合わせて字幕を出力する字幕出力装置が実行する字幕出力方法であって、入力されたテキスト文を字幕の出力単位に分割することにより、字幕単位文を生成する字幕単位文生成ステップと、前記テキスト文を音声認識の処理単位に分割することにより、音声認識単位文を生成する音声認識単位文生成ステップと、前記音声認識単位文生成ステップにおいて生成された音声認識単位文の文節を音声認識するための認識候補の集合である認識候補単位を、前記音声認識単位文の先頭の文節に対応するものから順に連結することにより、音声認識ネットワークを生成する音声認識ネットワーク生成ステップと、前記テキスト文が発声された音声の音声認識処理を、前記音声認識ネットワーク生成ステップにおいて生成された音声認識ネットワークのうち2以上の音声認識ネットワークを用いて、並列に行う音声認識ステップと、前記音声認識ステップにおける音声認識処理の結果に基づいて字幕単位文を出力する字幕単位文出力ステップとを備えたことを特徴とする字幕出力方法を提供する。

30

40

【0014】

請求項11に記載の発明は、コンピュータに、入力されたテキスト文を字幕の出力単位に分割することにより、字幕単位文を生成する字幕単位文生成ステップと、前記テキスト文を音声認識の処理単位に分割することにより、音声認識単位文を生成する音声認識単位文生成ステップと、前記音声認識単位文生成ステップにおいて生成された音声認識単位文の文節を音声認識するための認識候補の集合である認識候補単位を、前記音声認識単位文の先頭の文節に対応するものから順に連結することにより、音声認識ネットワークを生成する音声認識ネットワーク生成ステップと、前記テキスト文が発声された音声と、前記音声認識ネットワーク生成ステップにおいて生成された音声認識ネットワークを構成する認識候補単位との照合を先頭から逐次行うことにより、音声認識処理を行う音声認識ステッ

50

ブと、前記字幕単位文生成ステップにおいて生成された字幕単位文の少なくとも先頭の文節に対応する認識候補単位との前記照合が完了した時点で、前記字幕単位文を出力する字幕単位文出力ステップとを実行させるためのプログラムを提供する。

【0015】

請求項12に記載の発明は、コンピュータに、入力されたテキスト文を字幕の出力単位に分割することにより、字幕単位文を生成する字幕単位文生成ステップと、前記テキスト文を音声認識の処理単位に分割することにより、音声認識単位文を生成する音声認識単位文生成ステップと、前記音声認識単位文生成ステップにおいて生成された音声認識単位文の文節を音声認識するための認識候補の集合である認識候補単位を、前記音声認識単位文の先頭の文節に対応するものから順に連結することにより、音声認識ネットワークを生成する音声認識ネットワーク生成ステップと、前記テキスト文が発声された音声の音声認識処理を、前記音声認識ネットワーク生成ステップにおいて生成された音声認識ネットワークのうち2以上の音声認識ネットワークを用いて、並列に行う音声認識ステップと、前記音声認識ステップにおける音声認識処理の結果に基づいて字幕単位文を出力する字幕単位文出力ステップとを実行させるためのプログラムを提供する。

10

【発明の効果】

【0016】

本発明によれば、字幕出力装置は、字幕単位文の少なくとも先頭の文節に対応する認識候補単位との照合が完了した時点で前記字幕単位文を出力するため、リアルタイム放送において、音声に対して少ない遅延で字幕を出力することが可能となる。

20

また、字幕出力装置は、テキスト文が発声された音声の音声認識処理を、2以上の音声認識ネットワークを用いて並列に行うため、発話者の読み飛ばし等による音声の誤認識を防ぐことができ、音声に対応した字幕を正確に出力することができる。

【発明を実施するための最良の形態】

【0017】

以下、本発明の実施形態について、図面を参照しつつ説明する。

図1は、本発明の実施形態に係る字幕出力装置10の機能構成を示すブロック図である。本実施形態では、ニュース等のリアルタイム放送番組の原稿が電子化された連続テキスト文と、当該原稿がアナウンサーにより読み上げられた音声とが、字幕出力装置10に入力されるものとする。これにより、字幕出力装置10から字幕単位文が出力され、当該字幕単位文は、図12に示す従来の方法で音声や映像と多重化された後に、受信機に送出されて表示されるものとする。

30

図1に示すように、本実施形態に係る字幕出力装置10は、形態素解析部11、文節推定部12、音声認識単位文生成部13、字幕単位文生成部14、ピタビネットワーク生成部15、音声認識部16、及び、字幕単位文出力部17を含んで構成される。これらの機能は、字幕出力装置10が備える図示せぬCPU(Central Processing Unit)が、ハードディスクやROM(Read Only Memory)等の記憶装置に記憶されたプログラムやデータ等のソフトウェアを読み出して実行することにより実現される機能である。

【0018】

(形態素解析部)

40

形態素解析部11は、光ディスク等の記録媒体や通信回線を介して字幕出力装置10に入力された連続テキスト文を、予め記憶装置に記憶されている文法のルールや品詞、読み仮名等の辞書データベースを用いて、形態素(Morpheme:品詞、単語等の言語で意味を持つ最小単位)に分割し、それぞれの品詞、読み等を判別する。

図2は、形態素解析結果の具体例を示す図である。同図には、連続テキスト文「民主党、社民党、国民新党の野党3党が提出した福田総理大臣に対する問責決議が参議院本会議で初めて可決されました。」を入力とした場合に、形態素解析により出力される表層語(連続テキスト文が分割された結果である各形態素)、基本形(活用語の終止形)、読み(表記上の仮名)、発音(表音上の仮名)、品詞名、活用形が示されている。

なお、図2においては、各表層語に対応する読みは1つずつ表示されているが、複数の

50

読みを持つ表層語については、複数の読みを得ることができる。例えば、図 2 では、「3」の読みは「サン」のみが示されているが、「ミ」、「スリー」の読みも得ることもできる。

【0019】

(文節推定部)

文節推定部 12 は、連続テキスト文中の句読点や形態素解析部 11 の解析結果による単語・品詞情報を、予め記憶装置に記憶されている文節推定ルールと照合することで、文節の単位(区切り位置)を推定する。なお、文節推定ルールとは、助詞、助動詞等の品詞種類や句読点の並び条件に基づいて、文節の単位を推定する公知のロジックである。なお、文節とは、名詞、動詞等の自立語に接語が接続された発音上の単位である。例えば、「あの人は私の甥です。」というテキスト文の文節は、「あの」、「人は」、「私の」、「甥です。」の 4 つとなる。

10

【0020】

(字幕単位文生成部)

字幕単位文生成部 14 は、所望の字幕単位文生成条件(例えば、画面に表示する字幕の文字数は 30 文字以内とする等の条件)に適合するように、入力された連続テキスト文を文節の区切りで分割することで、自然な箇所で区切られた字幕単位文を生成する。

【0021】

(音声認識単位文生成部)

音声認識単位文生成部 13 は、連続テキスト文の句読点や形態素解析部 11 による単語・品詞情報を、予め記憶装置に記憶されている公知の息継ぎ推定ルールと照合することによって、息継ぎによる無音区間を推定し、連続テキスト文を無音区間で区切ることにより、音声認識に適した処理単位である音声認識単位文を生成する。

20

図 3 は、形態素解析部 11 による解析結果に基づいて、文節推定部 12 により推定される文節と、字幕単位文生成部 14 により生成される字幕単位文と、音声認識単位文生成部 13 により生成される音声認識単位文の具体例を示す図である。

図 3 に示す原稿の連続テキスト文「民主党、社民党、国民新党の野党 3 党が提出した福田総理大臣に対する問責決議が参議院本会議で初めて可決されました。」は、形態素解析部 11 により形態素解析され、当該形態素解析された結果としての句読点や単語・品詞に基づいて、文節推定部 12 により図 3 に示す文節が推定され、字幕単位文生成部 14 により図 3 に示す字幕単位文が生成され、音声認識単位文生成部 13 により図 3 に示す音声認識単位文が生成されることとなる。

30

【0022】

(ビタビネットワーク生成部)

ビタビネットワーク生成部 15 は、原稿の連続テキスト文がアナウンサーにより読み上げられた場合の音声を認識するためのビタビネットワーク(Viterbi Network)を生成する。このビタビネットワークは、音声認識単位文生成部 13 により生成された音声認識単位文の文節を音声認識するための認識候補の集合である認識候補単位を、当該音声認識単位文の先頭の文節に対応するものから順に連結したものである。ここで、「認識候補」とは、文節が発声された音声を音声認識可能とするために、1 つの文節に対して形態素解析部 11 により得られた 1 又は複数の読みの発音記号列を例えば音素 HMM(Hidden Markov Model)に変換したものである。また、「認識候補単位」とは、1 つの文節に対する認識候補の集合である。したがって、「文節」と「認識候補単位」とは 1 対 1 に対応する。1 つの文節に対して得られた読みが複数の場合には、「文節」と「認識候補」、及び、「認識候補単位」と「認識候補」とは、1 対多の関係となる。1 つの文節に対して得られた読みが 1 つの場合には、「認識候補」と「認識候補単位」とは一致する。ビタビネットワーク生成部 15 は、このビタビネットワークを、音声認識単位文生成部 13 により生成された音声認識単位文の数だけ生成する。

40

【0023】

また、ビタビネットワーク生成部 15 は、連結された認識候補単位間に、誤認識を防ぐ

50

ための特殊認識候補を挿入する。ここで、「特殊認識候補」としては、「S I L」、「N U L L」、「G a r b a g e」等が存在する。「N U L L」は、ポーズ（間）がないことを意味しており、無音区間も不要語も発生しなかった場合を表現している。「S I L」は、無音のポーズ（無音区間）を意味しており、アナウンサーが発声の間を任意に取ることによって、ビタビネットワークの音声認識の尤度が低下するのを防ぐ機能を有する。「G a r b a g e」は、音声認識において期待していない語を意味し、不要語を吸収する機能を有する。不要語が挿入される場合としては、例えば、「福田そーり、ゲホ、総理大臣に対する・・・」といった咳き込みや、「もん、問責決議が」のような言い直しが発生した場合である。このように、認識候補単位の中にN U L L、S I L、G a r b a g e等の特殊認識候補を挿入することで、読み間違いや間のおき方の違いを吸収し、高精度の音声認識を行うことが可能となる。

10

【0024】

さらに、各認識候補単位を構成する認識候補中に特殊認識候補を含めることも可能である。例えば、形態素解析時に、読み仮名候補がない、或いは、英文字・記号などで読み方が不明又は不明瞭であると判定された文節については、G a r b a g eを並列な認識候補として、認識候補単位中に含めることができる。また、雑音などの理由による音声認識誤りを避けるためにG a r b a g eを認識候補単位中に含めることもできる。さらに、アナウンサーの読み飛ばし等による誤認識を避けるためには、N U L Lを並列な認識候補として認識候補単位中に含めることができる。なお、G a r b a g eは、全音素H M Mの並列な枝として構成される。

20

【0025】

図4には、3つの音声認識単位文及び文節から生成される3つのビタビネットワークの例を示す。なお、この例では、連続テキスト文の形態素解析時において、「3」の読み候補は、「サン」、「ミ」、「スリー」の3通りが存在し、「福田」の読み候補は「フクタ」、「フクダ」、「フグダ」の3通りが存在したため、同図に示すように、文節「3党が」の認識候補単位は認識候補「サン」、「ミ」、「スリー」で構成されており、文節「福田」の認識候補単位は認識候補「フクタ」、「フクダ」、「フグダ」で構成されている。また、この例では、文節「民主党」の認識候補単位は認識候補「ミンシュトー」、「N U L L」、「G a r b a g e」で構成されている。また、図4に示すビタビネットワークを構成する各認識候補単位を連結する矢印は、図5に示すように、N U L L、S I L、G a

30

【0026】

さらに、ビタビネットワーク生成部15は、図1に示すように字幕先頭検出用ネットワーク生成機能151を備えている。字幕先頭検出用ネットワーク生成機能151は、各字幕単位文の先頭の文節が発声されたことを検出するための字幕先頭検出用のビタビネットワーク（以下、「字幕先頭検出用ネットワーク」という）を、字幕単位文生成部14により生成された字幕単位文の数だけ生成する。この字幕先頭検出用ネットワークは、先頭部分が所定のビタビネットワークの先頭部分で構成され、終端部分が所定の字幕単位文の先頭部分に対応する認識候補単位で構成されている。なお、この字幕先頭検出用ネットワークの生成方法の詳細は後述する。

40

【0027】

（音声認識部）

音声認識部16は、原稿の連続テキスト文がアナウンサーにより発声された音声を、ビタビネットワーク生成部15で生成されたビタビネットワークを用いて音声認識する。

図6は、音声認識処理部16の詳細な機能構成を示すブロック図である。同図に示すように、音声認識処理部16は、音声特徴量抽出部161と、ビタビネットワーク比較評価部162と、事象発生判定部163とを含んで構成される。

音声特徴量抽出部161は、入力音声から音声特徴量を求める。

ビタビネットワーク比較評価部162は、ビタビネットワークを構成する各認識候補単位を構成する各認識候補及び各認識候補単位の中に挿入された特殊認識候補の音声特徴量

50

と、音声特徴量抽出部 161 で得られた音声特徴量との比較照合を逐次行い、ビタビネットワークで表される時系列的な音声特徴量変化が起こった尤度（確率）を逐次算出する。

【0028】

なお、ビタビネットワーク比較評価部 162 は、複数のビタビネットワークを並列に評価し、各ビタビネットワークの尤度を同時並行に算出する並列認識処理を行うことが可能である。並列評価を行う場合に並列評価対象となるビタビネットワークは、並列評価を行わない場合に認識対象となるビタビネットワーク（すなわち、現在アナウンサーが発声中の文節を含む音声認識単位文に対応するビタビネットワーク）に後続する 1 つ又は 2 つのビタビネットワークとしてもよいし、前後に隣接するビタビネットワークとしてもよい。また、並列評価対象となる字幕先頭検出用ネットワークは、上記認識対象となるビタビネットワークの先頭部分を含むネットワークとすることができる。これらの並列評価対象となるビタビネットワークの決定ルールは、予めプログラムやデータベースで定義しておくことができる。

10

事象発生判定部 163 は、ビタビネットワーク比較評価部 162 で算出された尤度に基づいて、複数のビタビネットワークで表される事象の何れか、もしくは、どれも発生していないことを任意の時点で判定し、事象検出結果を出力する。

【0029】

（字幕単位文出力部）

字幕単位文出力部 17 は、音声認識部 16 から得られた事象検出結果に基づいて所定の字幕単位文の出力タイミングを検出した時に、その字幕単位文を出力する。本実施形態では、字幕単位文出力部 17 は、字幕先頭検出用ネットワーク生成機能 151 で生成された字幕先頭検出用ネットワークで表される事象が発生したことを検出した時に、当該字幕先頭検出用ネットワークに対応する字幕単位文を出力する。

20

なお、音声認識部 16 は、字幕先頭検出用ネットワークで表される事象が発生したことを検出した後も、当該字幕先頭検出用ネットワークを構成する認識候補単位を先頭部分に有するビタビネットワークを続けて最後まで音声認識するため、次の字幕文が不要なタイミングを出力されるのを防ぐことができる。

【0030】

（字幕出力処理）

次に、図 7 に示すフローチャートを参照して、本実施形態に係る字幕出力装置 10 が実行する字幕出力処理について説明する。

30

まず、字幕単位文生成部 14 は、形態素解析部 11 及び文節推定部 12 による処理結果に基づいて、入力された原稿の連続テキスト文を字幕の出力単位に分割することにより、複数の字幕単位文を生成する（ステップ S101）。

次に、音声認識単位文生成部 13 は、形態素解析部 11 による処理結果に基づいて、入力された原稿の連続テキスト文を音声認識の処理単位に分割することにより、複数の音声認識単位文を生成する（ステップ S102）。

【0031】

次に、ビタビネットワーク生成部 15 は、音声認識単位文生成部 13 により生成された複数の音声認識単位文毎に、各文節に対応する認識候補単位を連結してビタビネットワークを生成する。また、ビタビネットワーク生成部 15 は、字幕先頭検出用ネットワーク生成機能 151 により字幕先頭検出用ネットワークを生成する（ステップ S103）。

40

次に、生放送中に、原稿の連続テキスト文がアナウンサーにより読み上げられて、リアルタイム音声が入力されると、音声認識部 16 は、入力音声と、ビタビネットワーク生成部 15 により生成された字幕先頭検出用ネットワークを含む複数の各ビタビネットワークを構成する認識候補単位とを、先頭から逐次並列に照合することにより、並列認識処理を行う（ステップ S104）。

字幕単位文出力部 17 は、字幕先頭検出用ネットワークで表される事象が発生したことを検出した時に、当該字幕先頭検出用ネットワークに対応する字幕単位文を出力する（ステップ S105）。

50

【 0 0 3 2 】

(字幕先頭検出用ネットワークの生成処理)

次に、図 8 に示すフローチャートを参照して、ビタビネットワーク生成部 1 5 の字幕先頭検出用ネットワーク生成機能 1 5 1 が実行する字幕先頭検出用ネットワークの生成処理について説明する。

前提として、「ネットワーク間距離」の算出方法を定義する。このネットワーク間距離は、ビタビネットワーク同士の類似度を表す指標となり、ネットワーク間距離が小さいほど 2 つのビタビネットワークを形成する音素同士が類似しており、誤認識が起こる確率が高いことを表す。例えば、ネットワーク間距離は、各ビタビネットワークに含まれる認識候補単位を形成する音素間距離を積算した値として定義できる。なお、ビタビネットワークが複数の経路を有する（つまり、ビタビネットワークに含まれる認識候補単位の中に複数の認識候補を含むものがある）場合は、例えば、比較対象となっているビタビネットワーク間の最近接部分の距離をネットワーク間距離として定義できる。

10

【 0 0 3 3 】

まず、字幕単位文生成部 1 4 で生成された字幕単位文のうち、字幕先頭検出用ネットワークの生成対象となる字幕単位文を 1 つ選択し、当該字幕単位文の先頭文節に対応する認識候補単位を含むビタビネットワーク（以下、「対象ビタビネットワーク」という）に対して、仮の字幕先頭検出用ネットワークを設定する。具体的には、対象ビタビネットワークの先頭の認識候補単位から字幕単位文の先頭文節に対応する認識候補単位までを、仮の字幕先頭検出用ネットワークとする（ステップ S 2 0 1 ）。

20

【 0 0 3 4 】

仮の字幕先頭検出用ネットワークと、対象ビタビネットワークと並列に音声認識されるビタビネットワークのうち字幕単位文の先頭文節に対応する認識候補単位を含まない各ビタビネットワークとの間のネットワーク間距離を各々算出する。算出したネットワーク間距離の中に予め定められた所定の閾値未満のものがある場合（ステップ S 2 0 2 ； N o ） 、 仮の字幕先頭検出用ネットワークに対して、字幕単位文の次の文節に対応する認識候補単位を追加していき（ステップ S 2 0 3 ） 、 ビタビネットワーク間距離が所定の閾値以上となり、他のビタビネットワークと十分な距離を確保できた場合に（ステップ S 2 0 2 ； Y e s ） 、 字幕先頭検出用ネットワークを決定する（ステップ S 2 0 5 ） 。 なお、仮の字幕先頭検出用ネットワークに認識候補単位を追加したときに、対象ビタビネットワークの終端に到達した場合、すなわち、仮の字幕先頭検出用ネットワークと対象ビタビネットワークとが同一となった場合は（ステップ S 2 0 4 ； Y e s ） 、 対象ビタビネットワーク全体を字幕先頭検出用ネットワークとして採用する。以上の字幕先頭検出用ネットワーク生成処理を、字幕単位文生成部 1 4 で生成された字幕単位文の数だけ行う。

30

【 0 0 3 5 】

以上のような手順で字幕単位文の先頭数文節を含む字幕先頭検出用ネットワークを生成し、字幕先頭検出用ネットワークで表される事象が発生したことを検出した時に当該字幕先頭検出用ネットワークに対応する字幕単位文を出力することで、字幕単位文の先頭数文節が発声された時に字幕単位文の出力を行うことができ、必要最小限の遅延で字幕単位文を出力することができる。また、並列に認識される他のビタビネットワークとのネットワーク間距離を十分にとることで、認識間違いをなくすることができる。

40

【 0 0 3 6 】

(字幕先頭検出用ネットワーク決定処理の具体例)

次に、字幕先頭検出用ネットワーク生成機能 1 5 1 が、図 9 に示す音声認識単位文に基づいて、同図に示す字幕単位文の先頭を認識するための字幕先頭検出用ネットワークを決定する処理の具体例について説明する。

この例では、現在発声中の文節を含む音声認識単位文に対応するビタビネットワークと、当該ビタビネットワークに後続するビタビネットワークと、の 2 つを並行して用いて音声認識処理を行うものとする。また、実際には、音声認識単位文に対応するビタビネットワークを構成する認識候補単位を用いて字幕先頭検出用ネットワークが生成されるが、こ

50

ここでは、「ビタビネットワーク」及び「認識候補単位」の代わりに、対応する「音声認識単位文」及び「文節」を用いて説明することとする。

【0037】

まず、字幕単位文1)の字幕先頭検出用ネットワークを決定するために、音声認識単位文(A)の先頭文節「別府へ」を、仮の字幕先頭検出用ネットワークとして設定する(図8のステップS201に対応)。この仮の字幕先頭検出用ネットワーク「別府へ」と、音声認識単位文(B)の先頭の文節「切符を」とのネットワーク間距離を計算すると、「ベップへ」と「キップオ」とのネットワーク間距離はかなり近いので(ステップS202; No)、音声認識単位文(A)の次の文節「行く」を仮の字幕先頭検出用ネットワークに追加する(ステップS203)。これにより、仮の字幕先頭検出用ネットワーク(音声認識単位文(A)の先頭から2文節「別府へ」+「行く」と、音声認識単位文(B)の先頭から2文節「切符を」+「買う」とのネットワーク間距離を十分に保つことができるため(ステップS202; Yes)、「別府へ」+「行く」を字幕単位文1)の字幕先頭検出用ネットワークとすることにより(ステップS205)、先頭2文節の発声で、音声認識単位文(A)が発声されていることを高精度に判定することができる。

10

【0038】

次の字幕単位文2)の字幕先頭検出用ネットワークは、上記と同様の処理手順により、「切符を」+「買う」となる。

次の字幕単位文3)は、音声認識単位文(B)の「チップを」+「渡した」まででは、音声認識単位文(C)の「チップを渡す」と十分な距離がとれないため、「ものか」までが接続され、音声認識単位文(B)の先頭文節から「チップを」+「渡した」+「ものか」までが、字幕単位文3)の字幕先頭検出用ネットワークとなる。

20

【0039】

(並列認識処理の具体例)

次に、図10及び図11を参照して、並列認識処理の具体例について説明する。

図10(a)は、原稿の連続テキスト文「民主党、社民党、国民新党の野党3党が提出した福田総理大臣に対する問責決議が参議院本会議で初めて可決されました。自民公明両党は対抗措置として・・・」から生成されたビタビネットワーク、(b)は上記連続テキスト文から生成された字幕単位文、(c)は(b)の字幕単位文1)、2)各々の下線部分を音声認識した時点で各字幕単位文を出力するための字幕先頭検出用ネットワークである。

30

【0040】

図11は、図10(a)に示すビタビネットワーク及び図10(c)に示す字幕先頭検出用ネットワークに基づいて音声認識部16が行う音声認識処理、及び、音声認識処理による事象検出結果に基づいて字幕単位文出力部17が行う字幕単位文の出力処理の具体例を示す図である。

まず、音声認識部16は、ビタビネットワーク生成部15が生成した図10(a)、(c)に示すビタビネットワークのうち、先頭のビタビネットワーク1Aと、並列認識処理対象となる次のビタビネットワーク2Aと、字幕先頭検出用ネットワーク1Bとを検出対象として入力する(ステップS301)。

40

音声1「みんしゅとうしゃみんとう」がアナウンサーにより発声された時に、音声認識部16は、字幕先頭検出用ネットワーク1Bの事象を検出する(ステップS302)。そして、音声認識部16は、検出対象から検出済みの字幕先頭検出用ネットワーク1Bを除外し、次の字幕先頭検出用ネットワーク2Bを追加する(ステップS303)。

【0041】

字幕単位文出力部17は、音声認識部16による事象検出結果に基づいて、字幕単位文1)を出力する(ステップS304)。

次に、音声2「こくみんしんとうのやとうさんとうがていしゅつした」が発声されると、音声認識部16は、ビタビネットワーク1Aの事象を検出する(ステップS305)。音声認識部16は、検出対象からビタビネットワーク1Aを除外し、次の並列認識対象の

50

ビタビネットワーク 3 A を追加する (ステップ S 3 0 6)。

次に、音声 3 「ふくだそうりだいじんにたいするもんせきけつぎが」が発声されると、音声認識部 1 6 は、ビタビネットワーク 2 A 及び字幕先頭検出用ネットワーク 2 B の事象を検出する (ステップ S 3 0 7)。音声認識部 1 6 は、検出対象からビタビネットワーク 2 A 及び字幕先頭検出用ネットワーク 2 B を除外し、次の並列認識対象のビタビネットワーク 4 A を追加する (ステップ S 3 0 8)。

【 0 0 4 2 】

字幕単位文出力部 1 7 は、音声認識部 1 6 による字幕先頭検出用ネットワーク 2 B の事象検出に基づいて、字幕単位文 2) を出力する (ステップ S 3 0 9)。

なお、ビタビネットワーク 2 A 及び字幕先頭検出用ネットワーク 2 B は同一であるため、一方を他方で兼用することも可能である。

以上説明したように、字幕出力装置 1 0 は、字幕単位文の少なくとも先頭の文節に対応する認識候補単位との音声の照合が完了した時点で字幕単位文を出力するため、リアルタイム放送において少ない遅延で字幕を出力することが可能となる。また、NULL、SIL、SIL、Garbage 等の特殊認識候補をビタビネットワークの構成要素とすることで、アナウンサーの読み間違いや間のおき方の違いを吸収し、高精度の音声認識を行うことが可能となる。

また、字幕出力装置 1 0 は、音声認識処理を、2 以上のビタビネットワークを用いて並列に行うため、アナウンサーの読み飛ばし等による音声の誤認識を防いだり、発声タイミングのずれを回復することができ、音声と対応した字幕を正確に出力することができる。

【 0 0 4 3 】

なお、本発明は、上述した実施形態に限定されることなく、特許請求の範囲に記載の技術的範囲内において、上述した実施形態に適宜の変形を加えて実施可能であることは言うまでもない。

例えば、上述した実施形態では、字幕単位文出力部 1 7 は、字幕先頭検出用ネットワークを用いて字幕単位文の出力タイミングを判定したが、これに限らず、例えば、字幕先頭検出用ネットワークを用いずに、字幕単位文に対応する音声の認識が開始されてからの時間で出力タイミングを判定してもよい。また、字幕単位文の先頭の数文節に対応する認識候補単位と入力音声との照合が完了した時点で字幕単位文を出力してもよい。「数文節」は予め定められた数であってもよいし、並列認識される他のビタビネットワークとの尤度の差が大きくなり事象発生が検出されるまでの数であってもよい。また、文節の代わりに音節や文字数を用いてもよい。

【 0 0 4 4 】

また、字幕先頭検出用ネットワークの決定方法は、上述した実施形態に限定されることはなく、最低限、字幕単位文の先頭文節が発声されたことを検出できるように、字幕単位文の先頭文節に対応する認識候補単位を少なくとも含むネットワークとなるように決定すればよい。

また、上述した実施形態では、音声認識の性能を高めるために、字幕単位文と音声認識単位文とを別々に生成し、音声認識単位文を字幕単位文と一致させなかったが、音声認識単位文を字幕単位文と一致させることも可能である。

【 0 0 4 5 】

また、形態素解析以外の解析ルール、分割ルール等を用いて、字幕単位文や音声認識単位文を生成してもよい。また、ビタビネットワーク以外の音声認識のためのネットワークを用いて音声認識処理を行ってもよい。

また、上述した実施形態では、生放送のニュース番組でアナウンサーのリアルタイム音声に合わせて字幕を出力する例について説明したが、共通の原稿をアナウンス用と字幕用との双方に利用するリアルタイム放送であれば、スポーツ中継であっても、生講演であってもよい。

【図面の簡単な説明】

【 0 0 4 6 】

10

20

30

40

50

【図 1】本発明の実施形態に係る字幕出力装置の機能構成を示すブロック図である。

【図 2】形態素解析結果の具体例を示す図である。

【図 3】形態素解析結果に基づいて、推定される文節及び生成される字幕単位文・音声認識単位文の具体例を示す図である。

【図 4】3つの音声認識単位文及び文節から生成される3つのビタビネットワークの具体例を示す図である。

【図 5】図 4 に示すビタビネットワークを構成する各認識候補単位を連結する矢印が意味する内容を説明するための図である。

【図 6】同実施形態に係る音声認識部の詳細な機能構成を示すブロック図である。

【図 7】同実施形態に係る字幕出力装置が実行する字幕出力処理の流れを示すフローチャートである。

10

【図 8】同実施形態に係る字幕先頭検出用ネットワーク生成機能を実行する字幕先頭検出用ネットワーク生成処理の流れを示すフローチャートである。

【図 9】同実施形態に係る字幕先頭検出用ネットワークを決定する処理の具体例を説明するための音声認識単位文及び字幕単位文を示す図である。

【図 10】同実施形態に係る並列認識処理の具体例を説明するための図である。

【図 11】同実施形態に係る並列認識処理の具体例を説明するための図である。

【図 12】従来におけるリアルタイム放送番組において一般的に行われている字幕送出の仕組みを示す図である。

20

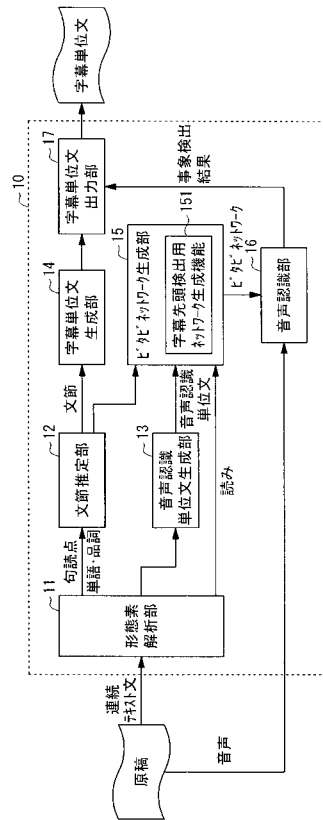
【符号の説明】

【0047】

- 10 字幕出力装置
- 11 形態素解析部
- 12 文節推定部
- 13 音声認識単位文生成部
- 14 字幕単位文生成部
- 15 ビタビネットワーク生成部
- 151 字幕先頭検出用ネットワーク生成機能
- 16 音声認識部
- 161 音声特徴量抽出部
- 162 ビタビネットワーク比較評価部
- 163 事象発生判定部
- 17 字幕単位文出力部

30

【図 1】



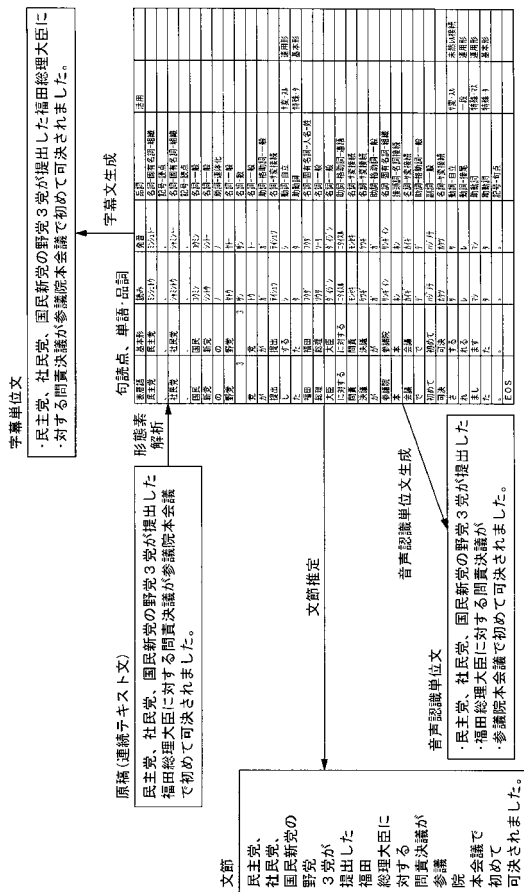
【図 2】

原稿(連続テキスト文)

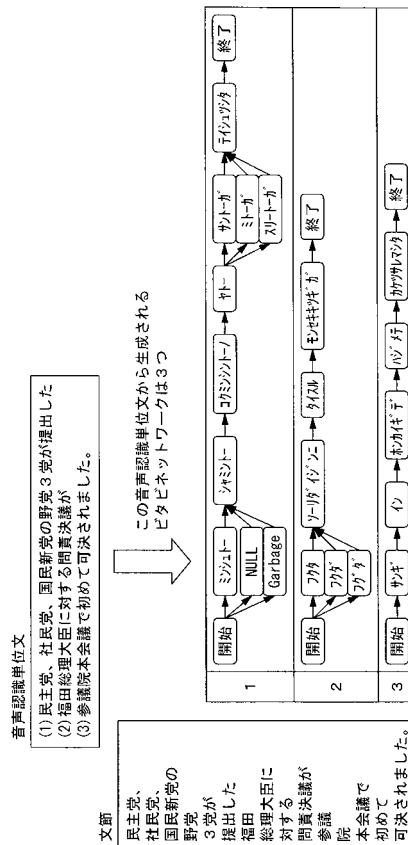
民主党、社民党、国民新党の野党3党が提出した福田総理大臣に対する問責決議が参議院本会議で初めて可決されました。

表層語	基本形	読み	発音	品詞	活用
民主党	民主党	シムシツ	シムシツ	名詞-固有名称-組織	
、	、	、	、	記号-読点	
社民党	社民党	シャミン	シャミン	名詞-固有名称-組織	
、	、	、	、	記号-読点	
国民	国民	コクミン	コクミン	名詞-一般	
新党	新党	シン	シン	名詞-一般	
の	の	ノ	ノ	助詞-連体化	
野党	野党	ヤウ	ヤウ	名詞-一般	
3	3	サン	サン	名詞-数	
党	党	トウ	トウ	名詞-一般	
が	が	ガ	ガ	助詞-格助詞-一般	
提出	提出	テイシュツ	テイシュツ	名詞-サ変接続	
し	する	シ	シ	動詞-自立	サ変・スル 連用形
た	た	タ	タ	助動詞	特殊・タ 基本形
福田	福田	フタダ	フタダ	名詞-固有名称-人名-姓	
総理	総理	ソウリ	ソウリ	名詞-一般	
大臣	大臣	ダイジン	ダイジン	名詞-一般	
に対する	に対する	ニタイスル	ニタイスル	助詞-格助詞-連語	
問責	問責	モンセキ	モンセキ	名詞-サ変接続	
決議	決議	ケツギ	ケツギ	名詞-サ変接続	
が	が	ガ	ガ	助詞-格助詞-一般	
参議院	参議院	サンギイン	サンギイン	名詞-固有名称-組織	
本	本	ホン	ホン	接頭詞-名詞接続	
会議	会議	カイギ	カイギ	名詞-サ変接続	
で	で	デ	デ	助詞-格助詞-一般	
初めて	初めて	ハジメテ	ハジメテ	副詞-一般	
可決	可決	カツ	カツ	名詞-サ変接続	
さ	する	サ	サ	動詞-自立	サ変・スル 未然形接続
れ	れる	レ	レ	動詞-接尾	一段 連用形
まし	ます	マシ	マシ	助動詞	特殊・マシ 連用形
た	た	タ	タ	助動詞	特殊・タ 基本形
。	。	。	。	記号-句点	
EOS					

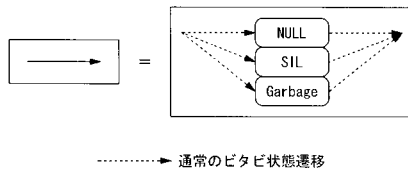
【図 3】



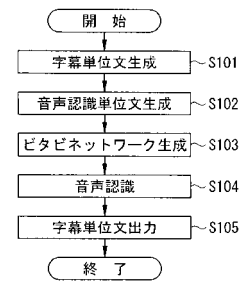
【図 4】



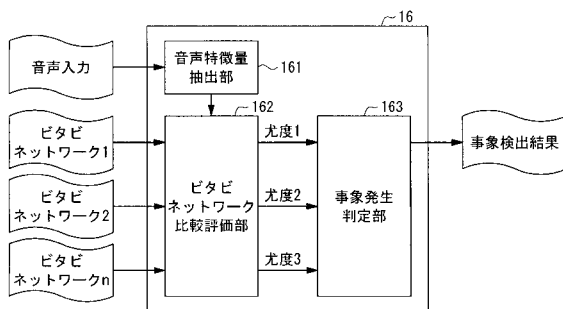
【図 5】



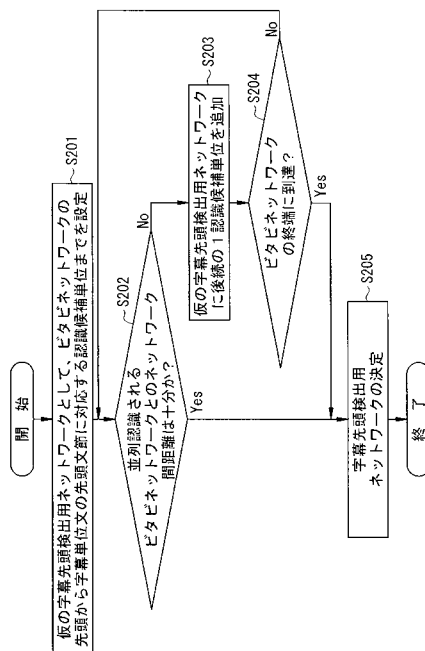
【図 7】



【図 6】



【図 8】



【図 9】

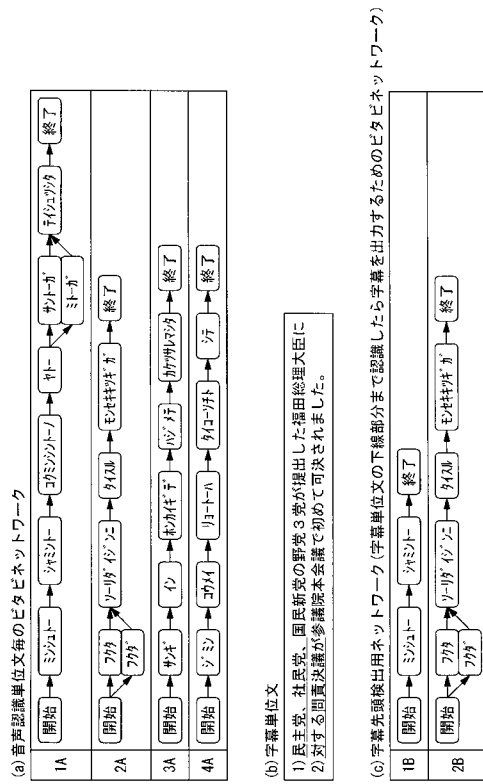
音声認識単位文

(A) 別府へ行く外国人旅行者を助けた。
(B) 切符を買い、世話をしてくれた代理店の人にチップを渡したのか、わからなくて右往左往していた。
(C) チップを渡す習慣は日本人にはないと説明をしなくてはならなかった。

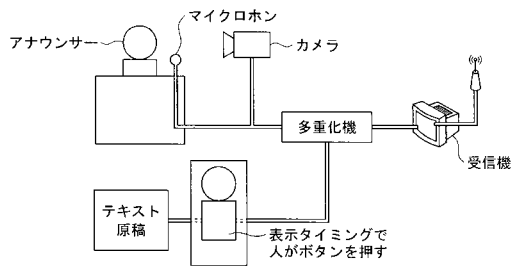
字幕単位文

1) 別府へ行く外国人旅行者を助けた。
2) 切符を買い、世話をしてくれた代理店の人に
3) チップを渡したのか、わからなくて右往左往していた。

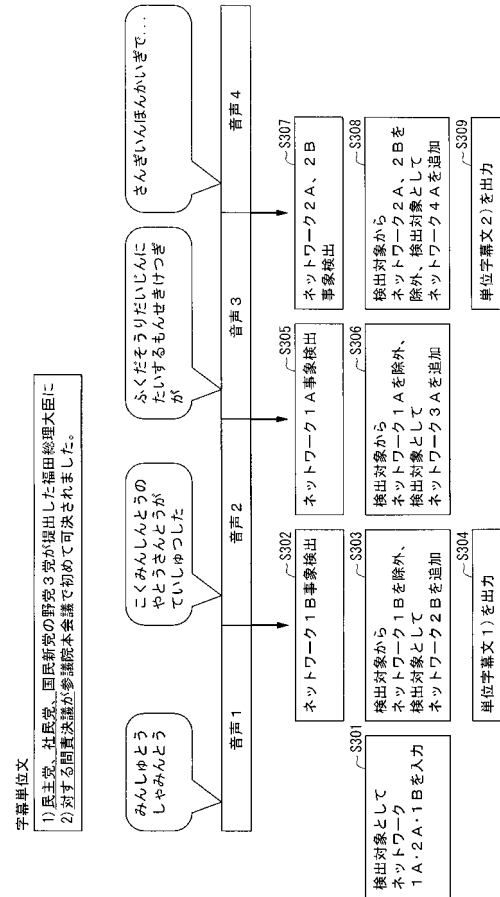
【図 10】



【図 12】



【図 11】



フロントページの続き

Fターム(参考) 5C164 FA12 GA04 MA06P MA07P PA44 SA41P SB04P SB10P
5D015 KK02