

音声インターフェースを用いたWeb新聞へのアクセス

近藤 玲史 稲垣 敬子 磯 健一 三留 幸夫

NEC 情報メディア研究所

World Wide Web 上で公開されている新聞記事を、音声を用いて簡単にアクセスするためのアプリケーションを試作したので報告する。

Web 上の情報にアクセスするためには、通常ブラウザが用いられるが、マウスやキーボードを使うことのできない状況や、それらの操作に習熟していない人がアクセスしたい場合には、音声の利用が便利である。そこで、対象を Web 上の新聞に限定して、記事をアクセスするアプリケーションを試作した。このアプリケーションは、音声を記事の選択と読み上げに用いており、音声インタフェースを効果的に用いることで、利用者の使い勝手を向上させている。

Speech Interface for Access to the Newspapers on the Web

KONDO Reishi, INAGAKI Keiko, ISO Ken-ichi, MITOME Yukio

NEC Corporation

This paper describes an application program using speech interface to access newspapers on the World Wide Web.

People use a Web browser to get information from the Web. The browser cannot be used when the mouse and keyboard are unavailable. It is also inconvenient for those who can not handle them well. Speech interface is applicable in these cases. Thus, we developed an application program to access the newspapers on the Web by using speech recognition and speech synthesis. As the speech is used in an effective way, the human interface of the program is much improved.

1 はじめに

インターネットの普及に伴って、World Wide Web で情報を公開する団体や個人が増加している。このような情報サービスの中でも、新聞社各紙が公開している新聞記事や速報は、人気のあるメニューの一つとして知られている (図 1)。

一般に、Web 上の情報を閲覧するためには、Web ブラウザが用いられる。その際、URL の入力などの操作は、全てマウスあるいはキーボードを使って行われ、情報はディスプレイに表示される。そのため、他の作業を行っている場合や、物理的にキーボードやマウスを設置できない状況では、Web ブラウザを使うことは困難である。また、マウスやキーボードの操作に習熟していない人にとっても、Web ブラウザの使用は難しい。

上記のような状況でも、音声インタフェースを用いることにより、情報にアクセスすることが容易になる [1]。これまで、音声による Web 上の情報のアクセスとして、Web ブラウザに表示される文字情報を全て読み上げる方法や、音声認識した単語に予め関連づけられた URL をアクセスする方法、メニューを読み上げて利用者に発声を促し、候補を音声認識で選択させ

る方法 [2] などが検討されてきた。しかし、これらはまだ音声の特徴を十分に生かしているとは言えない。

音声の特徴の一つとして、情報の伝達がシーケンシャルに行われること、および伝達速度が遅いことが挙げられる。音声を利用するに当たって、伝達される情報の特性がこれらの特徴に適していることが重要である。例えば、新聞ニュースや手紙文などの長文の伝達は、受け手の理解がシーケンシャルに行われ、また音声の伝達速度が受け手の理解に相当であるため、音声の利用が適している。また、メニューにおいて、利用者が項目名を発声して選択することは、伝達速度が問題とならないため音声の利用が適しているが、システムがメニューの全項目を読み上げて利用者に提示することは、伝達する情報が多くて時間がかかること、かつメニュー形式の情報は基本的に受け手がランダムにアクセスしたい情報であることから、音声の利用は適していない。

今回は、アクセスする情報を Web 上の新聞記事に限定し、記事の選択や記事の本文を読み上げる場面で効果的に音声入出力を用いることで、利用者の使い勝手を向上させたアプリケーションを試作した。

2 Web 上の新聞記事の特徴

本アプリケーションの対象は、各新聞社が Web 上で公開している新聞記事である。これらの新聞は、一般の Web ブラウザで表示されることを想定して作成されており、以下のような特徴を持つ。

視覚的効果の重視

新聞のロゴや画像、広告を入れたり、フレームを使用するなど、ブラウザで表示する際の視覚的効果を重視した構成になっている。

新聞毎の独自フォーマット

新聞毎にタイトル表示に用いるタグが異なるなど、それぞれ異なった HTML フォーマットで書かれている。

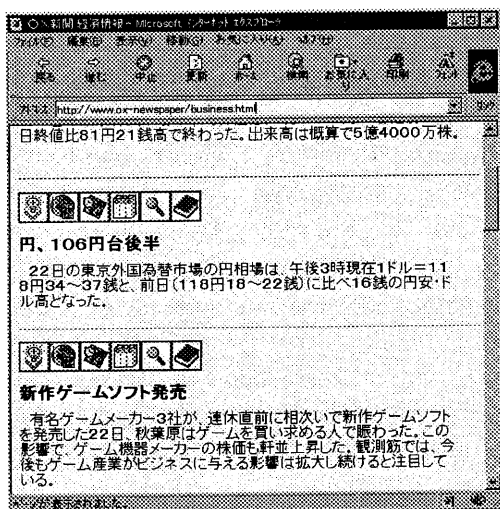


図 1: 新聞記事のブラウザによる表示例

情報の更新

多くの場合、記事は高い頻度で更新される。

新聞特有の文体

新聞には、一般の文章とは異なる特有の文体が使われることが多い。例えば、一般の文と比べて一文が長く、多くの情報が詰め込まれている点や、振りがな等のために括弧を多用する点などが挙げられる。

これら Web 上の新聞記事の特徴は、音声によるアクセスにおいて問題となる場合がある。そこで、解決策については次のアプリケーションの実装とあわせて解説する。

3 新聞記事アクセスアプリケーション

新聞記事アクセスアプリケーションでは、音声入出力の機能を実現するために、音声認識合成プラットフォームを使用している[1]。ここでは、音声認識合成プラットフォームが提供する機能と特徴について説明し、その後、新聞記事アクセスアプリケーションの実装について述べる。

3.1 音声認識合成プラットフォーム

本アプリケーションで使用した音声認識合成プラットフォームは、図 2 で示すように、マルチタスク OS 上で動作する音声認識エンジンと音声合成エンジンで構成される。これらのエンジンが、アプリケーションに対して音声認識・合成の機能を API (Application Programming Interface) として提供する。それぞれのエンジンの基本仕様を表 1 に示す。

音声認識エンジンは、独自の半音節モデルを用いた混合ガウス分布 HMM により、不特定話者の連続単語認識が可能となっている。半音節モデルの使用によって、認識辞書をかな文字列で記述できるので、あらかじめ認識語彙を発声して学習させる必要は無い[4]。

音声合成エンジンは、入力されたかな漢字混じりのテキストを解析して発音記号列に変換し、その後、韻律を制御して音声波形を生成す

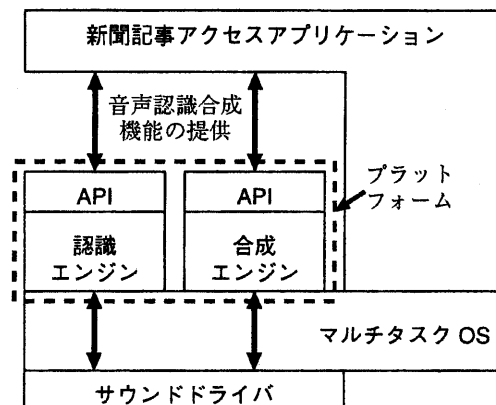


図 2: 音声認識合成プラットフォーム

表 1: 音声認識合成エンジンの基本仕様

音声合成エンジン	
波形生成話者	波形編集方式 男、女 (通常、老人、ロボット)
標準化周波数設定機能	11.025kHz, 8kHz テンポ、ピッチ、アクセント 話者、読みの区切り文字 数字・記号の読み方
音声認識エンジン	
話者	不特定話者 (学習可能)
認識対象	連続単語
認識単語数	最大 200
辞書登録	テキスト (かな文字列)

る。発声速度に依存しない韻律的特徴を用いることで、自然なイントネーションやリズムを実現している。また、音声生成に波形編集方式を用いることで、少ない演算量で高い明瞭性を得ている[3]。

音声認識エンジン・音声合成エンジン共に、OS の標準的な呼び出し手段に従った API の実装やマルチタスク・マルチスレッド対応などの機能強化も行っており、アプリケーションから使いやすいものとなっている。

以下で述べるアプリケーションの実装においては、かな文字列で登録できる音声認識エンジ

ンの特徴を生かして、動的な認識辞書の登録を行っている。また、この際に必要なかな文字列は、音声合成エンジンのテキスト解析機能を用いて生成している。

3.2 アプリケーションの実装

音声を用いて Web 上の新聞記事をアクセスするアプリケーションを実装するにあたって、マウスやキーボードの利用が困難な状況や、それらの操作が苦手な人でも快適に使えるよう、音声を用いて必要十分な情報アクセスが可能であることを目標とした。そのための基本機能として、利用者が得たい記事を音声で指定することにより、該当する記事を読み上げることを考える。実際には、記事の指定を音声認識によって行い、記事の読み上げを音声合成によって行う。ただし、全ての操作を音声のみで行うと利用者の使い勝手を損なう場面もあるので、ディスプレイの表示を併用することにした。具体的には、記事のタイトル一覧を利用者に提示する場合は、その全てを音声で読み上げるのではなく、ディスプレイ上に一覧を表示した。

本アプリケーションを起動すると、図 3 に示すような画面が立ち上がり、最初に記事のタイトル一覧が表示される。表示されたタイトルには、それぞれ自動で付与されたキーワードが、括弧で括られて、記事番号と共に表示される。利用者は、このタイトルの一覧を見て、得たい記事のキーワードをマイクロフォンに向かって発声する。キーワードが認識され、該当する記事がディスプレイに表示されると共に、本文の読み上げが開始される。記事の指定は、キーワード以外にも、タイトルに振られた記事番号を発声することによっても行うことができる。また、経済面や社会面などの記事のカテゴリ名が発声されると、そのカテゴリのタイトル一覧に移動する。

本アプリケーションは図 4 に示す構成を持つ。起動されると、Web サイトから該当する記事一覧と記事本文の含まれる HTML ファイルが取得され、解析されて表示と内容の読み上げに用いられると同時に、タイトルからキーワードが抽出されて認識辞書に登録される。

以下、Web 上の新聞記事特有の問題点を解

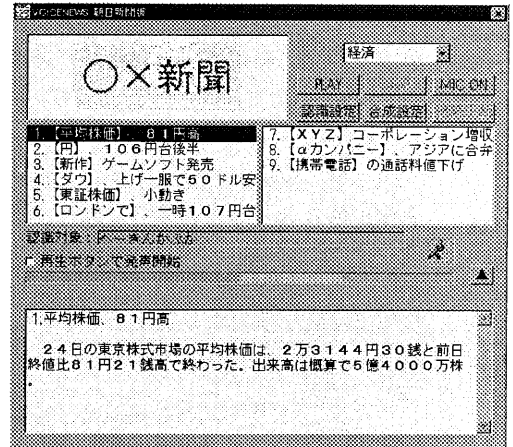


図 3: 実行画面 (読み上げ時)

決するための処理を説明する。

HTML ファイルの解析とリンクの解決 新聞記事は HTML で記述されているが、記事のタイトルと本文以外に、視覚的効果を実現するための情報が多く、また新聞毎に独自フォーマットになっている。そのため、一般的な記事の HTML ファイルから、タイトルと本文を取り出すことは困難である。

これを解決するため、HTML ファイルの解析は、新聞毎に専用の解析ルーチンを用意している。解析部は、HTML ファイル内の特定のタグを目印にして、タイトルと本文を探す。新聞によっては、タイトルと本文の HTML ファイルが別になっている場合もあるが、この場合にはタイトルと該当するリンク先を解析しておいて、記事が選択された時点でリンク先のファイルを読んで本文の解析を行う。

また、一文が複数行に渡る場合は一行にまとめるが、本文中に改行などのタグが入っている場合は指定された整形を行う。

キーワードの付与 新聞記事は情報の更新が早く、アプリケーションがタイトルを含む HTML ファイルを読み込むまではタイトルの内容が判らない。そのため、HTML ファイルを読み込む毎に、解析部で得られた各タイトルからキーワードを抽出する。

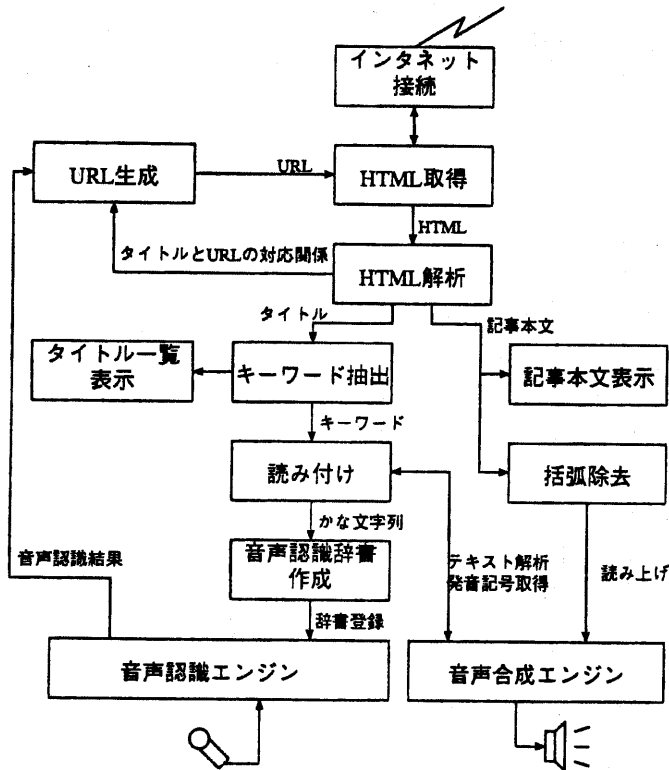


図 4: アプリケーションの構成

音声認識辞書の動的な作成 情報の更新に伴って前述のキーワードが変化するため、音声認識辞書にその都度登録する必要がある。今回使用した音声認識エンジンの認識辞書は、認識語彙をかな文字列で入力するため、キーワードに読みを付けてから登録を行う。

音声合成エンジンは、入力された文字列に対してテキスト解析を行い、発音記号列を出力する機能を持っている。そこで、キーワードをテキスト解析し、出力された発音記号列を規則によってかなに変換し、認識辞書を作成する。これにより、動的に変化するキーワードの認識も可能となる。

音声認識辞書には、これらのキーワードの他、「経済面」や「社会面」などの記事カテゴリ、「3番」「7番」などの記事番号も一緒に登録する。これにより、記事カテゴリの移動や、番号による記事の指定も音声で行うことができ

る。

記事の読み上げ 新聞記事には、特有の文体が存在する。特に、振りがなの表記や所属の紹介などで括弧が多用されるが、これらは明らかに視覚的効果を期待したもので、読み上げの際には不要であるか、あるいは意味がわかりづらくなるものが多い。そこで、予め本文中の括弧の中を省略してから読み上げを行うことで対処する。

4 まとめ

Web上の新聞記事を簡単にアクセスすることができるアプリケーションを試作した。このアプリケーションは音声認識と音声合成の機能を備えており、キーワードの音声認識による記事の選択や、本文の読み上げを実現している。これにより、マウスやキーボードを使うことの

できない環境や、あるいはそれらを使うことに慣れていない人などでも、情報に簡単にアクセスすることが可能となった。

参考文献

- [1] 近藤,岩田,磯,畑崎,三留,渡辺,水野: 「パソコン向け音声認識合成プラットフォームの構築とアプリケーションの試作」 情報処理学会第53回全国大会, 2-363 (1996)
- [2] 西本,小林,新美: 「WWW 上のデータベース検索のための汎用音声インタフェース」 日本音響学会平成9年度春季研究発表会講演論文集, pp.179-180 (1997)
- [3] 高橋,岩田,三留,永野,森谷: 「パソコン向け音声合成ソフトウェア」 情報処理学会第47回全国大会, 2-377 (1993)
- [4] 篠田,坂井,磯,畑崎,渡辺,水野: 「パソコンソフト連続音声認識」 情報処理学会第50回全国大会, 2-465 (1995)