

視覚障害者のための日本語音声ブラウザの標準的モデルの提案

井上 謙次*¹ 持田 徹*² 煙山 薫*³ 市丸 朋史*⁴ 石黒 稚文*¹ 中嶋 鴻毅*¹

Proposal for a Standard Model of Japanese Voice Browser for Visually Impaired

Kenji Inoue,*¹ Toru Mochida,*² Kaoru Kemuriyama,*³
Tomofumi Ichimaru,*⁴ Masafumi Ishiguro*¹ and Hirotake Nakashima*¹

Abstract – This paper describes the whole figure of a standard voice browser which is used by the visually impaired including blinds, by demonstrating what kind of components and functionality will be needed for them. It figures out that the functionality includes auditory rendering engine which utilizes the information relations and structures, natural language processing to search a Japanese keyword in the text and reformatting engine for dynamic contents will make the usability of Web more efficient for visually impaired. It also refers some underlying accessibility issues.

Keywords : Auditory Rendering, Earcon, Usability, Visually Impaired, Voice Browser

1. はじめに

情報障害とも呼ばれる視覚障害を持つ者にとって、広く流通している墨字本（インク印刷された本）は情報の入手において重要な手段であるが、それが困難あるいは不可能なこともある。このため、別の手段として World Wide Web（Web）から情報を入手することは障害者自らが知りたい情報を能動的に獲得できるため重要な位置を占める。

しかしながら、一般的に利用されている Web コンテンツや Web ブラウザは視覚を使った提示や操作を前提として作成されているため、視覚障害者にとって使いやすい Web ブラウザとは言えず、またアクセシブルでない Web コンテンツも多いのが現状である。そこで、音声読み上げを主体とした音声ブラウザが研究・開発され、音声ブラウザを構成する各機能や要素の研究は進んでいるが、音声ブラウザが有すべき機能や現状の問題点の全体像を概観したものは少ない。本稿では、特に日本語を日常的に使う全盲の視覚障害者を対象として、Web 利用における問題点を列挙し、それに対処可能な理論的ベースとなるモデルを示すことにより、音声ブラウザの発展のために議論・研究すべき内容に関する指針を与えることを目指す。

2. 視覚障害の特性

視覚障害の症状は多様であり、例えば弱視、視野狭窄、色覚異常などがある。各障害に合わせたユーザインタフェースが求められるが、ここでは最悪のケース、すなわち視覚情報が全く利用できない全盲者をベースとし、視覚情報に頼らないインタフェースを考える。他の視覚障害者に対しては、各障害に対応した視覚的インタフェースを全盲者用インタフェースに加えればよいだろう。

視覚情報が利用できない場合は聴覚または触覚に対して情報を提示する必要がある。音声ブラウザはそのうち聴覚に対して情報を提示するものである（点字出力が可能な音声ブラウザも存在する）。また、視覚情報の提示を行えないということは、単にコンピュータからの出力だけではなくマウスに代表されるポインティングデバイスのような入力手段も利用できないことを意味する。

従って、以下の2点について考える必要がある。

- どのように聴覚・触覚に情報を提示するか
- どのように視覚に頼らない操作システムを構築するか

3. Web の利用目的

視覚障害者の主な Web の利用目的をまとめると以下ようになる（「視覚障害者の Windows パソコン及びインターネット利用・学習状況」を参考にした^[1]）。

- 情報獲得：情報の検索と入手
- 電子商取引：ショッピングや各種予約
- コミュニケーション：チャットや掲示板

このうち、情報獲得は基本的に Web サイトとのインタラクションがないのに対して、電子商取引やコミュ

*1: 大阪工業大学 情報科学部

*2: Accessible Contents Research Initiative

*3: 奈良県立盲学校

*4: 大阪大学大学院 工学研究科

*1: Faculty of Information Science and Technology, Osaka Institute of Technology

*2: Accessible Contents Research Initiative

*3: Nara School for the Blind

*4: Graduate School of Engineering, Osaka University

ニケーションなどの Web サイト上のサービスを利用する場合は入力フォームへの記入やボタンを押すなどのインタラクションが必要であり、また現在の状況が動的に変化・通知されるなどの特徴を有する。そのような動的コンテンツは静的な Web ページと同様に扱うことが難しいため、音声ブラウザでアクセスする場合は特別な考慮が必要となる。

4. 情報獲得

一般的に人々は Web ページを流し読み（スキャン）するとされるが^[2]、視覚障害者も同様に流し聞きすると言われる^[3]。これは、ユーザはまず欲しい情報を探索し、求める情報のみを時間をかけて読むことを意味していると言える。従って単に存在する情報を読むようにするだけではなく、より効率的に求める情報を探索できるようなシステムが必要である。

では視覚障害者を含め我々はいかにして情報を探索しているのだろうか。ユーザがどのように Web をブラウジングしているかに関する調査レポートを読むと、以下の 3 点が見出せる。

- 情報の構造を利用する
- 特徴量を用いた検索（キーワード検索）
- 文章を順次読む

ここで述べる情報の構造とは、例えば章見出しの中に節見出しが包含されるというような関係や、段落がひとまとまりの情報を指示することといった、情報間の関係を表す広範な概念を意図する。

特徴量の検索は、例えば検索エンジンによる多数ページの横断検索や Web ブラウザのページ内検索のように、自然言語でテキスト検索を行って求める情報を探すキーワード検索に代表される方法である。

もし情報の構造も特徴量を用いた検索も利用できない場合、単純に文章を最初から（あるいは最後から）一語一句読んでいくような方法で探索を行うしかないだろう。例えば、見出しも段落もなく、改行もないようなテキストの羅列を想起していただきたい。これは非常に効率の悪い方法であるが、ほとんどの視覚障害者はこれに近い方法で Web ページを探索しているのが現状である。このような場合では、音声読み上げの速度をどれだけ上げることができるかが探索速度を決める重要な点となる。

以下、情報の構造を利用した場合と特徴量検索を用いた場合の探索について、それぞれ詳しく検討する。

4.1 情報の構造を利用した探索

Web の情報空間全体は非常に広大であり、情報の構造を頼りにその中から得たい情報をいかに効率的に見つけるかということは非常に興味深い分野である。ディレクトリ型と呼ばれる検索エンジンが Web サイ

表 1 情報の関係の種類
Table 1 An example of information relation types

関係の種類	関係の意味
ISA 関係	A は B である
包含関係（階層関係）	A は B を含む
順列・並列関係	A と B は並んでいる
2 項関係	{A, B} は C である

トをカテゴリ別に分類しているのはその一つの試みと言えよう。このような大きく一般的な問題を音声ブラウザが一手に引き受けるべきではないから、既存の検索エンジンなどを利用できるような環境を整えることが音声ブラウザの役目と言える。音声ブラウザが扱うべきなのは、たかだか Web サイトや Web ページといった単位の大きさの構造であろう。

さて、今まで情報の構造という言葉を使ってきたが、ここで改めて本稿で用いる情報の「関係」と「構造」という用語を定義しておく。

まず情報とは何らかの意味を持つ概念か、何らかの関係によって結び付けられた複数の情報としておく。ここで述べたいことは、情報が集まったものも単に情報と記述していることを語弊がないようにしたいからであって、情報の意味の再定義を行いたいという目的ではない。むしろ情報という言葉は直感的な意味合いで使用している。

そして情報の関係とは、ある情報から見て別の情報との間に成り立つ意味上のつながり方である。Web に関して有用と思われる情報の関係には例えば表 1 に示すようなものが考えられる。

次に情報の構造とは、ある特定のコンテキスト（環境）において使われる情報の関係である。また、構造はある情報を別の情報と区別する機能（ブロック化）を有すると言える。

簡単な例を示そう。例えば「見出し」という構造は、ある情報（見出し文）と他の情報（本文）とを区別し、（意味の上で）見出し文が本文を包含するという関係を有すると考えられる。この関係は章見出しと節見出しとも同じであるが、節見出しは章見出しで区別された本文中というコンテキストに現れるという違いがある。逆に言えば、同様の構造を特定のコンテキストにおいて章や節という名前を付けたに過ぎず、本質的な情報の関係（あるいは構造が有する機能）は同じである。

また、一般的に節 1 の見出し文と節 2 の見出し文は順列関係（節 2 は節 1 に依存する）にあると言える。リスト構造においては、項目 1 と項目 2 は順列関係または並列関係にあると考えられる。順列か並列かは意味上の違いであるが、もし順列関係であるか並列関係であるかを特定できるような情報が記述されていなか

見出し

文章1

文章2

図 1 意味的に不明瞭な構造

Fig. 1 Semantically-uncleared Structure.

れば、これらを自動的に区別するのは難しい。人間の読み手であれば判別することも可能であるが、必ずしも可能であるとは言えない。

一般的に構造の関係は、そのテキストの記述者がどのように考えているかに依存する恣意的なものである。もし記述者の考えを再現可能な構造がそのテキストにマークアップされていなければ、構造の解釈は一意に定まらない。このことは、特に使用する構造セットの体系に左右される。例えば、図 1 のように見出し、文章 1、水平線、文章 2 というテキストを HTML で記述した場合、文章 2 が見出しと同じ階層なのか文章 1 と同じ階層なのかという意味上の関係が不明瞭である (HTML の見出しは名前の通りセクションではなく見出しであり、範囲を指定することができないため)。単にそれだけの記述であれば、前者の場合と後者の場合で音声読み上げ時に空ける間を変更するといった提示の仕方ができない。もしそのような違いを設けて提示したい場合、テキストの記述者が CSS (Cascading Style Sheet) などを用いて音声レンダリングの方法を指示しなければならないだろう。

実際に Web ページ内の典型的な構造がどのような情報の関係を有するかについて列挙する。ただし、ここで述べる構造は必ずしもこのように解釈すべきという意図ではない。

タイトル ページ全体を包含する。

見出し 同レベルの見出しが現れるまでの情報を包含する。通常、同レベルの見出しとは順列関係にある。

ヘッダ・コンテンツ・フッタなど 特定の意味内容の情報を包含する。それぞれ順列・並列関係にある。

リスト 順列・並列関係にある情報を包含するためのコンテナである。それぞれの項目は互いに順列・並列関係にある。

段落 情報 (主に文章) を包含する。通常、段落は互いに順列関係にある。

強調 指定した範囲の情報は強調される (ISA 関係)。

表 2 項関係にある情報を表現するためのコンテナ

表 2 情報の構造のレンダリング手法の対応関係

Table 2 Auditory rendering methods of information structure corresponding to visually rendering methods.

視覚的レンダリング	聴覚的レンダリング
空間的な配置	時間的な配置 (間など)
フォントの変化	声質・語調などの変化
シンボルや図形	サイン音や BGM

である。列、行とその値が 2 項関係にある。

また、Web ページ間の構造としてはリンクやフレームなどがある。その他、Web サイトという単位の構造も考えることができるだろう。

このような構造を、その構造が有する関係を理解しやすい形で提示することで、より効率的な探索を行うことができるだろう。

視覚への情報提示の場合、情報を 2 次元空間上にうまく配置・装飾することにより記述されている情報の構造の把握を容易にしておき、画面の上下スクロールとリンクを辿るといった基本的な操作で効率的に探索できるようになっている。そのさい、言語は文字で、情報の構造は空間的な配置やフォントの変更の他、リスト項目の「・」や表の罫線のようなシンボルや図形を用いて視覚的なレンダリングを行っている。

それに対して聴覚への情報提示の場合、必ず時間軸上に音を配置しなければならず、また同時に提示できる情報が非常に限られているため、全体的な見通しや構造を把握することが困難である。聴覚的なレンダリングにおいても、視覚的なレンダリングと同様に言語は音声で、情報の構造は時間的な間の配置や声質・語調の変化などの他、サイン音 (イアコン: earcon) や BGM などのシンボリックな要素を用いることによって構造を理解しやすくできるだろう。情報の構造の提示における視覚的なレンダリングとの対応関係を表 2 に載せる。

聴覚的なレンダリング手法ははまだ一般的なコンセンサスが得られておらず、発展途上である。音訳や音読の経験知識やサイン音などを活かし、より効率的な情報探索を研究する必要がある。

また、ナビゲーションのスキップなど、従来よりも構造の特徴を活かした操作系も重要であろう。しかしナビゲーションを示すような記述は HTML では用意されていないため、何らかの手法を用いてそのような構造を抽出する方法が必要である。

4.2 特徴量検索を利用した探索

ロボット型と呼ばれる検索エンジンにおけるキーワード検索のように、Web 上に散在する実データの特徴をパターンマッチングで検索することによって求める情報をすばやく探すことができる。多数の Web

ページの横断検索には既存の検索エンジンを利用できるため、音声ブラウザは Web ページ内といった規模の検索をサポートすればよい。

テキスト検索における日本語特有の問題として、漢字表記が異なる同音異義語が多いことが挙げられる。音声読み上げ環境では検索キーワードの入力時に漢字を特定するのが難しいため、Migemo のように仮名や音素記号（ローマ字）を用いて読み仮名や発音を検索できるようにすると便利である^[4]。ただし、その場合は漢字仮名混じり文を読み仮名や発音の表記に変換する必要がある（音声合成エンジンが行っているのと同様の処理である）。

5. 動的コンテンツ

ショッピングやチャットといった Web 上のサービスは多く存在しているが、このような動的コンテンツを音声ブラウザで扱うのは難しく、必要な情報を適切にユーザに提示するということがなかなかできない。

なぜ難しいのかというと、そのような動的コンテンツの結果やインタフェースとして HTML が使用されているが、そもそもこのことが誤用なのである。HTML が想定している構造は論文のような文章を書くための構造であり、ブラウザに返される結果やチャットの発言といった意味の論理構造を表すことができない。従って、想定しているレンダリング結果によく合致するような構造を当てはめてしまっているのが現状で、その結果動的コンテンツが本来ブラウザに返す情報の構造と実際の構造が乖離してしまっている。

従って、根本的な解決方法としては、一つにはサービス提供者がその動的コンテンツ用のプログラムやユーザインタフェースをその場でブラウザに送信する方法がある。しかし、ActiveX などの既存の仕組みでは音声化は不十分であり、汎用的なユーザインタフェース記述言語を送受信できるようなシステムを導入する必要があるだろう。もう一つには、ある程度分野を限った構造を有するフォーマットで（例えば XML を用いて）記述し、それを送信する手法が考えられる。それには Web サービスや RSS などが活用できるが、現状では常に利用可能なわけでもなく、当然フォーマットごとに対応する必要がある。ただしこのようなフォーマットは HTML と併用して提供されることがあるため、HTML の代替フォーマットとしては現状でも有用であると言える。

そこで現状では、個別ページ（サービス）ごとに返される HTML にタグを追加するか他のフォーマットに変換して、そのフォーマットに対する操作システムを提供するという方法で対処するのが良策であろう。

6. その他

上記の流れで述べられなかった点を本章でいくつか補足しておく。

画像自体を提示するには、それを解説するようなテキストを付与する以外には、視覚情報を完全に聴覚情報に変換して提示できるような装置が本質的に必要になるだろう。

しかしながら、表やグラフ、単純な図のように、言語情報に情報の関係を表す線や図形が結びついたものについては、上述のように音声や音などの代替手段を用いて表現することが原理的に可能であると言える。

7. 標準的モデル

以上のことをまとめると、音声ブラウザとして必要な機能には、おおまかには次のようにまとめられる。

- 構造解析・構造変換エンジン
- 聴覚的レンダリングエンジン
 - 音声再生部
 - イアコン再生部
- 自然言語処理エンジン

8. おわりに

本稿で導入した情報の関係と構造は、ソシユールの記号論を言語とシンボルだけに留めずに、空間における位相情報とも言えるような情報を表す空間配置や線、図などの記法にまで拡張するという考えを基にしている。これにより、どのような情報をどのような手段で表せばよいのか、ということとともに、どのような情報であれば聴覚的な情報提示が可能であるのか、ということの考え方のベースを説明できると考える。

最後に、本稿を基にして実際の音声ブラウザ開発に着手する予定である。

参考文献

- [1] 独立行政法人国立特殊教育総合研究所：視覚障害者の Windows パソコン及びインターネット利用・学習状況，http://www.nise.go.jp/kenshuka/josa/kankobutsu/pub_d/d-190.html (2003)。
- [2] Nielsen, J.: How Users Read on the Web, <http://useit.com/alertbox/9710a.html> (1997)。
- [3] Redish, J., Theofanos, M.: Observing Users Who Listen to Web Sites; Usability Interface, **Vol.9**, No.4 (2003). (available at <http://www.stcsig.org/usability/newsletter/0304-observing.html>)
- [4] Migemo, <http://namazu.org/satoru/migemo/>
- [5] Cohen, M., Giangola, J., Balogh, J.: Voice User Interface Design, Addison-Wesley (2004)。