携帯端末を用いた音声読みあげブラウザ@ｔａｌｋ*

5C−07

海老名　毅　　大野　裕之†
通信総合研究所§

1 はじめに

i-mode や sky-web など携帯端末を用いたwebブラウザ（以下，携帯webブラウザ）が登場して久しい．これら携帯webブラウザは，PC上で動作するwebブラウザに比べて機能が制約されているため，利用方法や対象ユーザに制限がある．例えば，PCでは，視覚障害者や高齢者は音声読みあげ可能なwebブラウザを用いて，画面に頼らずにwebにアクセス可能であった．しかし，携帯電話に代表される携帯端末にはマルチタスク処理機能や音声合成機能がなく，またあっても処理能力やメモリの機能的制約が大きいため，今までwebコンテンツの音声出力を実現できなかった．

われわれは，サーバクライアント型の音声web読みあげシステム@talk(あっとトーク)を開発した．このシステムは，負荷の高い処理をサーバが分担することで，機能的制約の大きい携帯端末でwebの音声読みあげを実現する．われわれが開発した音声webブラウザの操作イメージ図を，図1に示す．

2 PCにおけるWebブラウジング

PCで用いられる音声Webブラウザは，音声合成エンジンを使いWebコンテンツを音声に変換出力するソフトウェアである．コンテンツを非視覚化することで，ユーザは画面を見ることなくWebページにアクセスできる．しかし，単に画面上のテキストを音声に変換するだけでは，Webコンテンツにアクセスできたことにならない．例えば，音声をそのまま読むと時間がかかるため，音声Webブラウザでは時間を短縮するために読みあげ位置をすばやく移動する機能や，画面を音声で聞きとりやすいよう，ページの余分な要素を省略する機能が実装されている[1]．しかしながら，PC音声Webブラウジングにはいくつか制約がある．

図1:音声読みあげブラウザの操作イメージ

最初の制約は，ページレイアウトに関する制約である．Webページは視覚的な効果を狙ってレイアウトされていることが多いため，重要な

* Voice speech browser @talk using mobile terminal
† Tsuyoshi Ebina, Hiroyuki Ohno
§ Communications Research Laboratory

情報が最初の方に現れるとは限らない．したがって，Webページを音声で聞きとる場合，ユーザがページ全体からレイアウトを推測して，情報の重要性を判断しなければならない．

2つめの制約は，必要なweb操作の複雑性である．Webコンテンツを音声で効率的に聞き取りながら読み進めようとすると，複数の読みあげモードや読みあげスキップ機能が必要になる．このため，webコンテンツを効率的にアクセスしようとすると，操作方法はある程度複雑にならざるを得ない．

3 携帯端末における音声webブラウザ

携帯端末でweb コンテンツの読みあげ機能を用いる利点について述べる．

まず利点の1つめとして，操作の簡便性がある．先に述べたように，webコンテンツは複雑なレイアウトを持っていることが多いため，複雑なwebコンテンツにアクセスするには，操作方法が複雑にならざるを得ない．例えば，ある種の音声 web ブラウザではフレーム情報を伝えるために，フレームの開始および終了地点の情報を読みあげる．ユーザは，メッセージを聞いて画面のレイアウトを想像しつつ操作を行う．一方，i-modeに代表される携帯端末ブラウザは，ブラウザの画面が小さく表示できる情報も少ないため，i-mode対応コンテンツのレイアウトは，通常のwebのそれに比べ，単純に設計されている．よって，携帯端末向けコンテンツへのwebアクセスの方が，理解がより容易である．

2つめの利点は，操作に必要なボタンの数である．従来のwebブラウザでは，効率的にコンテンツをアクセスするために，多くの機能が必要であった．そのため，通常操作に必要なキーが多く割り当てられているか，または多くの読みあげモードを持っていた．一方，携帯端末向けのwebブラウザでは，必要な機能の数はきわめて少なくて済むので、アクセスに必要なキーの数も少なくてすむ．

3つめの利点は，他人にコンテンツを伝える補助手段としての利用である．携帯端末はパソコン画面と異なり画面を複数の人が同時に見ることは困難であり，またそのような利用方法も想定していない．しかし，他人に携帯端末を渡すと電話帳など他人に知られたくない情報まで知られてしまう可能性がある．このようなときに，指定した部分だけを音声で読みあげることで，自分の周囲の人間に情報を同時に伝えるとともに，住所録など他人に知られたくない情報を見られずに済む，といった利用形態が考えられる．

4 携帯音声webブラウザ@talk

4.1 @talkの構成

われわれは，携帯端末上でwebを音声読みあげするシステム@talkを開発した．
携帯端末では，CPU能力やメモリ容量，アプリケーションがきわめて少ないため，現在の機能の携帯端末では，単体で音声出力することができない．そこで，テキスト音声変換，音声フォーマット変換などの比較的重い処理をサーバに受け持たせ，携帯側のiアプリにユーザとのインターフェース機能を持たせることで，webコンテンツの音声出力を実現した．

@talk のシステム構成を図1に示す．@talkサーバはサンマイクロシステムズのマシン240R 上に実装されている．また，本システムを用いて利用可能な携帯端末は， 今のところNTT docomo のFOMA端末 N2001である．ただし，十分な通信帯域がとれれば，音声出力部分等を他機種に対応させることによって，今後他機種や i-mode 以外の機種で利用することも可能である．

@talk を構成するソフトウェアは，携帯端末上のアプリケーションであるiアプリおよび，@talk サーバ上のアプリケーションである

@talk アプリケーションから構成される．ユーザがiアプリを@talk サーバからダウンロードし実行すると，携帯端末は@talk サーバと接続可能な状態になる．ユーザがIアプリで URL を入力すると，URL が@talk サーバに送られる．@talk サーバは URL を取得すると対象となる HTML コンテンツをインターネット上からロードする．ロードされた HTML データはiアプリに送られ表示される．ユーザが読みあげボタンを押すと，フォーカス位置情報が@talk サーバに送られる．@talk サーバは，送られた位置情報からフォーカスされているテキストを抽出し，音声データに変換し，さらにiメロディ形式のデータに変換してiアプリ側に返す．iアプリは，受け取ったiメロディデータを再生する．

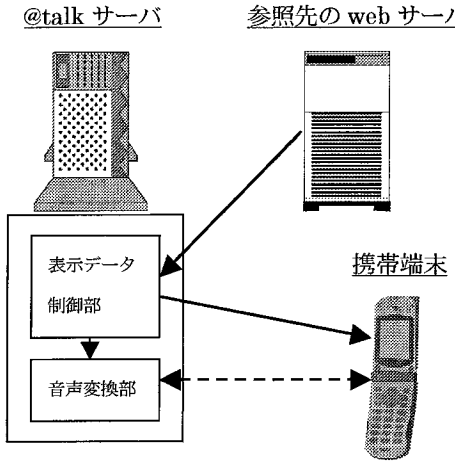

図2 @talk のシステム構成

4.2 @talk の操作

@talk の操作方法について述べる．最初に@talk アプリを@talk サーバからダウンロードし，実行する．次に対象 URL を入力すると，コンテンツが表示される．フォーカスの移動操作はi-mode の web ブラウザである DOJA と同じである．音声を出力するときには，左上のアドレスキーを押すと，音声が再生される．図3に，キーの割り当てを示す．

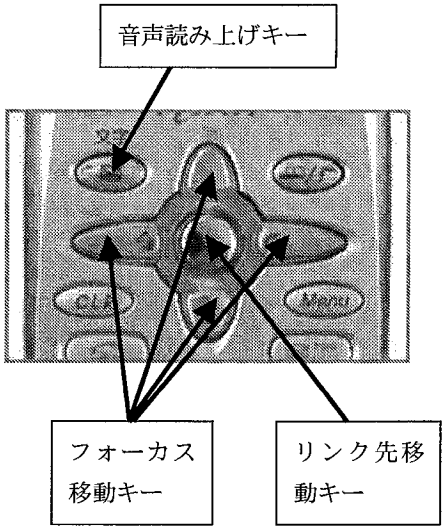

図3： 操作キーの割り当て

4.3 VoiceXML による記述

従来型の音声 web ブラウザでは，コンテンツを部分ごとに区切って，またはコンテンツ全部を順番に音声化できた．しかし，現在の携帯端末では，利用できるヒープスペースが非常に限られているため，2，3秒分のiメロディデータしか1度にバッファできない．そのため，現在のシステムでは，コンテンツを逐次読むことになるため，何らかの手段で読むテキストを指定する必要がある．

また，web コンテンツからみた場合，コンテンツの全ての情報が重要なわけではなく，重要な情報は一部であることが多い．そこで，ページ中の web コンテンツを@talk サーバ側で全て音声に直すのではなく，あらかじめ音声化すべき部分をコンテンツ側で指定することにより，より効率的な web アクセスを実現する．

XML の音声拡張として，VoiceXML[2]がある．

そこで，本システムでは，VoiceXML で記述された web ページを対象として，読みあげるべき部分を指定しておくことにより，音声読み上げを実現する．例えば以下の VoiceXML 記述

```
<audio src="1.mld">
  あっとトークについて
</audio>
```

は，携帯端末の@talk アプリで解釈され，1.mld という音声データを@talk サーバからダウンロードして再生する．

VoiceXML を採用することで，音声読みあげの曖昧性を排除できるとともに，他のボイスアプリケーションとの互換性を保つことができる．

5 考察

@talk のプロトタイプシステムを使い，サンプルのコンテンツの読みあげを行ってみた．サンプルのコンテンツは，災害時の利用を想定し，防災情報を表示した web コンテンツである．

ボリュームを最大にして音声出力キーを押すと，１m離れていても内容を聞きとることができた．ただし，現在のシステムは，音声フォーマット変換部分に不具合があるため，音声出力時にノイズが付加されるため，内容は聞きとりづらかった．

また，音声出力キーを押してから音声が再生されるまでの時間は，最初にキーを押した場合と，２回目以降にキーを押した場合とで大きく異なる．１回目の操作では数秒程度時間がかかることがあったが，２回目以降はほぼ実時間で再生が行えた．これは，１回目のアクセスでテキストを音声変換するために時間がかかったことと，FOMA でサーバに接続する時間がかかったためである．

使用してみたところ，上記の２点が今後の改善点としてあげられたが初期の目的である音声ブラウジングは実現できた．今後，上記の点を改善しつつ，システムの拡張を行うことを検討中である．

次に，コスト面を考察する．本システムは，サーバと音声データのやり取りを行うことから，テキストのみをやりとりするのに比べて通信料金が多くかかる．データ量を調べてみたところ，１行分の音声データは約 7kBytes であった．よって，通信コストの最も安いコースで計算した場合，１行あたりの通信コストは約 1.1 円であり，音声出力機能の利用によって発生する通信コストはユーザの許容範囲内におさまる．

6 まとめ

本稿では，携帯端末で web コンテンツを音声化するシステム@talk の概要を述べ，プレ評価を行った．その結果，まだ改善すべき点は残っているものの，機能的制約の大きい携帯端末で web コンテンツを音声化するシステムを構築できることが示された．携帯端末上で web コンテンツを音声でインタラクティブに確認できるシステムは世界的にみてもほとんどなく，本システムは携帯電話で web を非視覚的に利用できる今までにないシステムといえる．

これまで携帯端末のアプリケーションは，ゲームなど若者にフォーカスしたものが多かった．本システムは，高齢者や視覚障害者など情報弱者向けの携帯端末アプリケーションであり，また周囲に情報を伝えられるという意味では，グループで利用できるといった特色を持つ．

今後は，本システムを拡充してゆくとともに，外部にライセンスするなどの方法で，展開していくことを予定している．

参考文献

[1] Ebina.T, Igi.S,& Miyake.T (2000, November). Fast Web by Using Updated Content Extraction and a Bookmark Facility. ACM press, Proceedings of Assistive Technologies2000, pp.64-71.

[2] http://www.voicexml.org/