

ウェアラブル環境における 音声データのハイパーリンク型ナビゲーションシステム

庄 司 武[†] 中 村 聡 史^{††}
塚 本 昌 彦[†] 西 尾 章 治 郎[†]

本論文では、ウェアラブル環境下における情報閲覧システムとして、音声データをハイパーリンク形式で構造化することで直観的な閲覧を可能とし、そのように構造化された音声データの作成もサポートするシステムの提案および実装を行う。音声データには、リンクの開始・終了時間などのリンク情報を記述した HTML 形式の構造ファイルを添付する。システムはリンク情報に従い、音声データ再生中にリンク音を重ねて鳴らし、リンクの存在をユーザに通知する。ユーザは、リンク音が鳴っている間にマウスのボタンをクリックすることでリンクをたどることができる。

A Hyperlink-Style Navigation System of Voice Data in Wearable Computing Environment

TAKESHI SHOJI,[†] SATOSHI NAKAMURA,^{††} MASAHIKO TSUKAMOTO[†]
and SHOJIRO NISHIO[†]

In this paper, we propose a system for browsing and creating structured voice data in wearable computing environment. In this system, voice data is structured by a hypertext-style link. A linked part in voice data is specified by the start/end time in an attached HTML file. While playing voice data, each link is notified to a user by a link sound such as a beep. A user can get the linked information by clicking a mouse button while the link sound is being played.

1. はじめに

近年、コンピュータの小型化、高性能化にともない、コンピュータを持ち歩いて利用するモバイルコンピューティングが広まっている。近い将来、高速無線通信ネットワークが整備されると、ユーザは持ち歩いているコンピュータから随時ネットワーク上の情報にアクセスし、必要な情報を閲覧できるようになる。今後、ユーザにとってコンピュータは必要不可欠なものとなり、日常生活の中でコンピュータをつねに身につけ利用するのが一般化していくと考えられる。このようなスタイルのコンピューティングは、ウェアラブルコンピューティング⁷⁾と呼ばれている。

ウェアラブルコンピューティングにおける情報閲覧には、HMD (Head Mounted Display) の利用が一

般的であり、インターネットのブラウジングやメールの読み書きなど、画面を用いた作業を可能にしている。しかし、HMD を使用した情報閲覧システムでは、ユーザは画面に注目しなければ閲覧が困難であることが多く、一般に他の作業との並行性が低い。たとえば、車の運転中、目的地までの道順を知るために近隣の地図を HMD に表示することは可能だが、運転をしながら画面を見るのは難しく、危険もともなう。また、レシピを HMD に表示することで、調味料の割合などを確認しながら料理できるが、画面を注目している間に食材を焦がしてしまうなど、大きな失敗をする可能性がある。一方、車の運転や料理をしながらラジオで渋滞情報や料理番組を聞くことは十分可能であることから、他の作業を中断せずに並行して情報閲覧を行うには、音声を用いることが有効であると考えられる。

音声を用いた情報閲覧システムに関しては多くの研究がなされており、既存のウェブページを合成音声で読み上げユーザに提示するシステムや、デジタル録音した音声情報を相互に関連付けし、情報の閲覧を容易化するシステム、音声入出力のみでメールを閲覧・作

[†] 大阪大学大学院情報科学研究科
Graduate School of Information Science and Technology,
Osaka University

^{††} 大阪大学大学院工学研究科
Graduate School of Engineering, Osaka University

成できるシステムなどがある。しかし、これまでのシステムでは個々の情報間の関連付けが固定で、異なるコンテンツ間で自由に関連付けを設定できないなど、構造化された音声情報を直観的に閲覧することが困難だった。

一般に個々の情報を相互に関連付け、直観的な閲覧を可能とする基礎技術としてハイパーリンクが知られており、WWW (World Wide Web) などで広く用いられている。このハイパーリンクの概念を音声情報システムに導入することで、音声データの閲覧が容易になると考えられる。

そこで本研究では、ウェアラブルコンピューティングにおいて音声データの手軽な閲覧が行え、作成も容易にできるハイパーリンク型ナビゲーションシステムを構築することを目的として SoundWeb システムを提案し、プロトタイプシステムの実装を行う。

以下、2章で関連研究について説明し、3章で提案する SoundWeb システムについて述べる。4章でプロトタイプシステムの実装について説明し、5章で本研究についての考察を行う。6章で SoundWeb システムの応用について述べ、最後に7章でまとめと今後の課題について述べる。

2. 関連研究

音声を用いた情報閲覧システムに関して、さまざまな研究がなされている。

出力に合成音声を用いるシステムの研究として、i-Shadow²⁴⁾がある。このシステムでは、テキストデータを音声合成エンジンを利用して読み上げ、画面にリンク部分のみを表示することにより、携帯電話やPDAなど表示画面に制約のある携帯機器を利用して、情報が豊富な既存の Web ページを閲覧することを可能とする。この研究の拡張として、リンク部分が画面に表示されるたびに音を鳴らし、ユーザにリンクの存在を通知する方法²⁶⁾も提案されている。合成音声を利用した読み上げシステムには、既存のコンテンツをそのまま利用できるというメリットがある。しかし、合成音声はユーザにとって聞きにくいものであるため、長時間の利用はユーザにとってストレスの原因となる¹⁶⁾。また、リンクをたどるために表示されたリンク部分を確認し選択する必要があるため、このシステムの場合、画面なしでの利用は困難である。

出力にデジタル録音した音声を再生するシステムの研究として、Hyperspeech^{1),2)}がある。このシステムでは、会議やミーティングなどで録音した音声データを、すべてテキストに書き起こしたうえで意味的な区

切りで分割する。その後分割した音声データどうしを内容に基づいて関連付けを行う。音声データどうしが関連付けされていることで、音声データの階層的な閲覧が可能となる。また、デジタル録音した音声を使用することにより、合成音声やテキストデータでは表現しにくい言葉の微妙なニュアンスや話し手の個性などをユーザに提示することができる。しかし、Hyperspeech では、音声データの書き起こしや関連付けなどのデータ作成をすべて手作業で行う必要があるため、データ作成が困難となる。また、音声データ間は、話のまとまりごとに関連付けされているため、データ間の関連付けが固定で、内容に関連した情報に対して手軽にアクセスして閲覧を行うことは難しい。

ニュースや講義などを音声で閲覧するシステムとして、NewsComm^{10),11)}がある。このシステムでは、あらかじめ設定しておいた趣向に応じたニュースなどの音声データをネットワーク上から自動で取得し、再生提示する。趣向に応じたデータがユーザに提示されるため、ユーザは効率良く音声情報の閲覧を行うことができる。しかし、取得される音声データはシステムが自動で決定するため、必ずしも欲しい情報を閲覧できるとはいえない。

音声入出力を用いて情報閲覧を行うシステムとして、Nomadic Radio¹³⁾や AVoN calling⁹⁾、SpeechWear^{12),15)}がある。Nomadic Radio は、音声入出力によるテキストメール・ボイスメールの閲覧・作成を可能にするシステムで、ウェアラブルコンピューティング環境における利用に適している。AVoN calling は XML (eXtensible Markup Language) で記述されたデータを音声入力で操作するシステムである。SpeechWear は、アメリカ海兵隊の水陸両用車両の点検を行うウェアラブルシステムで、入力を音声で行える。これらのシステムは音声入力を行うことでハンズフリーでの入力を可能としているが、音声入力には周りの雑音により誤認識が発生するなどの問題があり、システムの使用環境が限定され、ウェアラブルコンピューティングで常時使用するのは難しい。

音声ベースの Web ページを作成するための記述言語として VoiceXML²¹⁾がある。また、W3C (World Wide Web Consortium) は VoiceXML で作成される音声 Web ページのブラウザ VoiceBrowsers²⁰⁾に必要とされる機能を規定している。VoiceBrowsers では電話機によるデータ閲覧も想定し、入力として音声と電話機のプッシュボタンを利用することが考慮されている。出力としては、合成音声によるテキストの読み上げやあらかじめ録音しておいた音声データの再生をサ

ポートしている。ユーザはこうした入出力を利用することにより、音声ベースによる情報閲覧を対話的に行うことができる。しかし、個々の情報間を細かく関連付けすることができないため、既存の音声情報サービス以外への応用が難しい。

3. SoundWeb システム

本章では、SoundWeb システムの提案を行う。

SoundWeb システムでは、音声を用いた情報提示を行うことで、ウェアラブルコンピューティング環境においてユーザが他の作業と並行して情報閲覧を行えるようにする。音声データの閲覧については、従来の Web ページを閲覧するように直観的に行え、関連情報に対して手軽にアクセスし閲覧することを可能とする。また、データ作成を容易に行えるようにすることで、さまざまな人がデータ作成を行いネットワーク上で公開できるようにする。

以上にあげた目的を実現するため、本研究では以下の 4 つのアプローチをとる。

3.1 データとしてデジタル録音した音声を使用

音声を用いた情報閲覧システムでは、新たに音声を録音する必要がないという理由から、合成音声による読み上げが多く用いられている。しかし、合成音声はユーザにとって聞き取りにくいものであり、長時間の利用はユーザにストレスや不快感を与えてしまう。ウェアラブルコンピューティングでの常時使用を考えると、閲覧の際にユーザにできるだけストレスや不快感を与えないことが求められる。

そこで SoundWeb システムでは、データとしてデジタル録音した音声を使用する。録音した音声を用いることで、ユーザにストレスや不快感を与えないだけでなく、合成音声やテキストデータでは表現しにくい喋っている人の特徴や言葉の微妙なニュアンス、感情などをユーザに伝えることが可能になると考えられる。また、いろいろな人の声を聞くことができるため、バリエーションが豊かになると考えられる。

3.2 音声データをハイパーリンクの形式で構造化

録音した音声データを閲覧する場合、ユーザは音声の早送りや巻き戻しを用いて欲しい情報を探すが、その間録音した音声を聞き続けなければならない。閲覧を容易にするため、音声データどうしを相互に関連付けする方法が考えられているが、これまで実現されている方法では、自由度が低く、直観性の高い関連付けが行えないなどの問題があった。各情報を相互に関連付け、直観的な閲覧を可能とする基礎技術としてハイパーリンクが知られており、WWW など広く利用

されている。このハイパーリンクの概念を音声情報システムに導入することで、音声データの閲覧が容易に行えるようになると考えられる。

そこで SoundWeb システムでは、音声データをハイパーリンクの形式で構造化する。ユーザは音声データの途中に埋め込まれたリンクを選択したとすることで、その部分に関連した情報の手軽な閲覧が可能になると考えられる。

3.3 副音声を用いてリンク部分を強調

音声データをハイパーリンクの形式で構造化する場合、音声中に埋め込まれたリンクをユーザにどのようにして提示するかということについて考慮する必要がある。i-Shadow では、リンク部分が再生されるたびに画面にリンク情報を表示し、ユーザにリンクを提示する方法が考えられている。しかし、この方法では、リンクをたどるためにユーザは画面に注目する必要があり、他の作業と並行して情報閲覧を行うのが困難となる。

そこで SoundWeb システムでは、テキストベースの Web ブラウザにおいてリンク部をアンダーライン表記することで強調していることに注目し、副音声として他の音を重ね合わせて再生することで、音声中のリンク部を強調することを考える。これを「音のアンダーライン」と呼ぶ。音のアンダーラインによって、ユーザは画面を利用することなく音声中のリンク部分を認識し、リンクを順にたどることができるため、音声のみで情報を閲覧することが可能になると考えられる。

3.4 データは音声の録音とボタン操作で作成

本研究では、音声をデジタル録音し、その音声データをハイパーリンクの形式で構造化する。この構造化などのデータ作成作業を、ユーザにできるだけ負担をかけずに簡単化する必要がある。

そこで SoundWeb システムでは、データ作成を音声の録音とボタン操作だけで行い、音声データどうしに関連付けなど構造化の作業を可能な限りシステムで自動化する。現在はデータ作成方法として、画面を見ずに作成を行えるオーディオベースの方法と、GUI (Graphical User Interface) ベースの方法を想定している。画面なしでデータ作成を可能とすることで、ウェアラブルコンピューティング環境に適した作成ツールを実現できると考えられる。

4. 実 装

4.1 システム概要

SoundWeb システムはアプリケーション本体と、

SoundWeb システムが動作可能なクライアントマシン、そして音声データが保存されているサーバマシンの3つからなる。ユーザはネットワーク上のサーバマシンから音声データを取得し、再生することでヘッドフォンを通して情報の閲覧を行う。閲覧に関する操作はマウスのボタン入力で行う。

4.2 必要要件

SoundWeb システムを実現するために、以下の要件を満たす必要がある。

音声データのリンク情報の保持：

ハイパーリンクの形式で構造化するために、音声データのリンク情報を保持する必要がある。

音声データの取得：

音声データ取得のため、ネットワーク上のデータに対してアクセスできる必要がある。

音声の多重再生：

リンク部分を音のアンダーラインを用いて強調するために、音声を多重再生できる必要がある。

SoundWeb システムのプロトタイプ実装において必要要件を満たすために以下のアプローチをとる。

データ構成：

音声データに直接リンク情報を埋め込むことも考えられるが、データ加工が困難になることが予想される。そこで、別に音声データの構造を記述する構造定義ファイルを用意し、その中へリンク情報も書き込む。

通信方式：

SoundWeb システムとサーバ間の通信には HTTP (HyperText Transfer Protocol) を使用する。これにより、Apache や IIS など既存の HTTP サーバを使用することができる。

再生方式：

音声の再生に Windows 標準の音声再生用 API を使用すると、音声を多重で再生することができない。そこで、音声データの再生には音声の多重再生を可能とする Microsoft DirectX 8.0 の DirectX Audio を使用する。

4.3 副音声を用いたリンクの提示

ユーザは、リンクをたどることにより音声データの閲覧を行う。音声データの再生位置が構造定義ファイル内でリンクとして設定されている部分にさしかかると、音のアンダーラインとして副音声を重ねて再生し、リンク部をユーザに提示する。音のアンダーラインとして再生する音を「リンク音」と呼ぶ。図1にリンク部分でリンク音が重ねて再生されている様子を示す。ユーザは、リンク音からリンクの存在だけでなく、

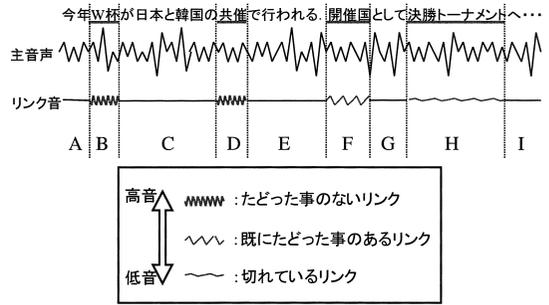


図1 リンクの通知
Fig. 1 Notification of a link.

リンクの状態も知ることができる。リンクの状態には、以前にたどったことがある、たどったことがない、リンクが切れているという3つがある。これらリンクの状態は、リンク音の音の高さを変化させることでユーザに提示する。リンクの状態を提示することによって、ユーザは同じ情報や切れているリンクへアクセスしてしまうのを防ぐことができる。図1には、リンクの状態によってリンク音の音の高さが変化する様子も示されている。リンク音の変化により、Fの部分(開催国)に設定されているリンクはユーザが以前にたどったことがあり、Hの部分(決勝トーナメント)に設定されているリンクは切れていることが分かる。なお、図において時間軸の長さが実際に発音される長さとも一致していないところがあるが、見やすさを考慮し「日本」のように漢字による表記も使用した。

4.4 構造定義ファイル

今回のプロトタイプシステムの実装では、ハイパーリンクの概念を表現するのに適している点を考慮し、音声データの構造を記述する構造定義ファイルに HTML (HyperText Markup Language) 形式を採用している。また、音声データのファイル形式としては、作成および再生の容易さを考慮し、WAVE 形式を用いている。

構造定義ファイルでは、まず、META タグを用いて SoundWeb システムのファイルであることを宣言する。次に、TITLE タグで再生する音声データの情報を記述する。

BGSOUND タグでは、再生する音声データを指定でき、以下にあげる SRC 属性と END 属性を利用することができる。

SRC 属性 (必須):

再生する音声データの URL (Uniform Resource Locator) を記述する。システムは SRC 属性に指定されているアドレスに従い、音声データの取得を行う。

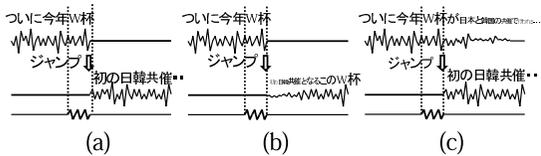


図2 CHANGE属性の指定結果

Fig. 2 The result of selecting the 'change' property.

END 属性：

音声の再生終了時の処理方法を指定する。デフォルトでは、停止状態となる。END 属性に PLAY を指定すると、最初から再生を再開することができる。また、BACK を指定しておくことで、自動でリンク元へ戻るようにできる。さらに、JUMP を指定し、続けて URL を記述しておくことで、再生終了時に自動でその URL へ移動させることができる。

A タグでは、リンクに関する設定を行うことができ、以下にあげる HREF 属性、ALT 属性、OPTION 属性、END 属性を利用することができる。A タグの開始タグと終了タグの間にリンクが設定されている位置を、再生開始からの秒数で指定する。

HREF 属性（必須）：

リンク先の SoundWeb システムデータの URL を指定する。dummy.html を HREF 属性として指定すると、リンク先が未設定となる。

ALT 属性：

リンク先の情報を記述する。リンク先の情報は、画面を使用できる際に表示するなどして、ユーザの情報閲覧を補助するために使用する。

OPTION 属性：

リンク音の種類を指定する。デフォルトでは、リンク音としてトーン音が再生される。OPTION 属性に WATER, DOG, CAT を指定すると、リンク音を流れる水の音、猫の鳴き声、犬の鳴き声にそれぞれ変えることができる。その他、OPTION 属性に WAVE ファイルへのリンクをつけておくことで、意図する音声データをリンク音として指定できる。

CHANGE 属性：

リンク先へ移動するときの再生方法を指定する。デフォルトでは、再生中の音声を停止してから、リンク先の音声を再生する(図2(a))。CHANGE 属性に FADE-IN を指定するとリンク先の音声をフェードインしながら再生し始めることができる(図2(b))。また、FADE-OUT を指定すると、再生中の音声をフェードアウトで停止することが

```

1: <html>
2: <head>
3: <meta http-equiv="Content-Style-Type" content="sweb">
4: <title>日本サッカー</title>
5: </head>
6: <body>
7: <bgsound src="http://www-nishio.ist.osaka-u.ac.jp/~shoji/soccer_japan.wav"
  end="play">
8: <a href="worldcup.html" alt="W杯" option="http://www-nishio.ist.osaka-
  u.ac.jp/~shoji/cheer.wav">12.7-14.24</a>
9: <a href="kyousai.html" alt="日韓共催" option="http://www-nishio.ist.osaka-
  u.ac.jp/~shoji/cheer.wav">30.4-32.8</a>
10: </body>
11: </html>

```

図3 構造定義ファイルの例

Fig. 3 An example of associated information.

表1 情報閲覧時の操作方法

Table 1 Button operations.

	左ボタン（進む）	右ボタン（戻る）
クリック	リンク先へ移動	リンク元へ戻る
プレス	早送り	巻き戻し
ダブルクリック	次のリンクの頭出し	前のリンクの頭出し

きる(図2(c))。

構造定義ファイルの例を、図3に示す。3, 4行目に meta タグと title タグがあり、SoundWeb ファイルの宣言と音声データの説明を記述している。7行目にある bgsound タグでは src 属性の URL で再生する音声データを指定しており、さらに end 属性で play を指定して、再生終了時に初めから再生を再開するよう設定している。8, 9行目には2つの a タグがありそれぞれリンク部の設定をしている。alt 属性でリンク先の説明をしており、option 属性には WAV ファイルの URL が指定されており、リンク音として主音声にあわせて歓声の WAV ファイルが重ねて再生される。

4.5 情報の閲覧方法

SoundWeb システムは、マウスのボタン入力による操作で音声データの閲覧を行う。情報閲覧時のボタン操作の詳細を表1に示す。基本的に左ボタンで進む操作を、右ボタンで戻る操作を行う。情報閲覧時に利用するモードを、アクセスモードと呼ぶ。

4.6 データ作成

SoundWeb システムのデータは、作成時に画面を必要としないオーディオベースの方法と、GUI ベースの方法で作成することができる。

4.6.1 オーディオベースでのデータ作成

録音とリンク設定

まず、音声データ作成のため音声の録音を行い、録音中にリンクを設定したい部分でボタンをプレスしリンク区間の指定を行う。リンク部分の設定を行った後、それぞれのリンクについてリンク先の設定を行う。リンク先は、あらかじめリンク先の候補となる SoundWeb データをいくつか登録しておき、順番に再



図4 GUIベースのデータ作成ツール

Fig.4 A GUI-based data creation tool.

生しその中から選択する．リンク先をすべて設定すると，ユーザの決定に基づいた構造定義ファイルが，システムによって自動的に生成される．この音声の録音とリンクの設定を行うモードを録音モードと呼ぶ．録音する際の WAVE ファイルのフォーマットは，今回の実装ではモノラルのサンプリング周波数 22,050 Hz で，量子化ビット数は 8 とした．

リンク先の再設定

録音によりデータ作成を行う際，適当なリンク先の候補が登録していないことは十分考えられる．そのようなときには音声の録音とリンク部の設定だけをしておき，後でリンク先の再設定を行うことができる．このリンク先の再設定を行うモードを，エディットモードと呼ぶ．エディットモードでは，再生されている音声の後ろでつねに小さな音を再生し，現在エディットモードであることをユーザに提示する．

4.6.2 GUIベースでのデータの作成

録音とリンクの設定

GUIベースでデータを作成する際も，オーディオベースでのデータ作成と同様に，まずは音声データの録音とリンク部分の設定を行う．録音を終わると，ウィンドウ上に音声データの波形が自動で表示される．このとき図4に示したように，リンクが設定されている部分は色が変わっている．リンク部分をマウスで選択し，URLを入力することにより直接リンク先を指定することができる．また，END属性やOPTION属性といった属性を指定し，システムの動作指定を行うこともできる．録音する際の WAVE ファイルのフォーマットは，今回の実装ではモノラルのサンプリング周波数 22,050 Hz で，量子化ビット数は 8 とした．

リンク先の再設定

GUIベースの作成方法では，リンク先の再設定だけでなく，リンク部に設定されている範囲の変更や，リンク部の追加および削除を行うことができる．リンク先の再設定は URL を入力し直すことで行う．範囲変

表2 実装環境

Table 2 Implementation environment.

使用計算機	東芝 DynaBook SS PORTEGE 3440/1J2
使用 OS	Microsoft Windows 2000
開発言語	Microsoft Visual Basic 6.0



図5 SoundWeb プロトタイプシステム

Fig.5 The SoundWeb prototype system.



図6 プロトタイプシステムの装着例

Fig.6 A style using the prototype system.

更やリンク部の追加および削除は，マウスのドラッグアンドドロップで行う．ドラッグアンドドロップでリンクまたは範囲を指定した後，メニューを表示してリンクの追加や削除などの処理を選択する．

4.7 実装環境

今回は表2の環境のもとで SoundWeb システムのプロトタイプシステムを実装した²³⁾．ソースコードの量は SoundWeb システムで約 3,400 行，GUIベースのデータ作成ツールで約 2,000 行となった．

SoundWeb システムには，音声データの閲覧を行うアクセスモード，データ作成を行う録音モード，リンク先の再設定を行うエディットモードの3つのモードがある．システムの立ち上げ時はアクセスモードになっている．モードは，マウスの左右ボタンを同時押しすることで変更できる．GUIベースのデータ作成ツールは SoundWeb システムのサポートツールとして提供している．

実装したプロトタイプシステムを利用する場合の構成を図5に，実際に装着して利用している様子を図6

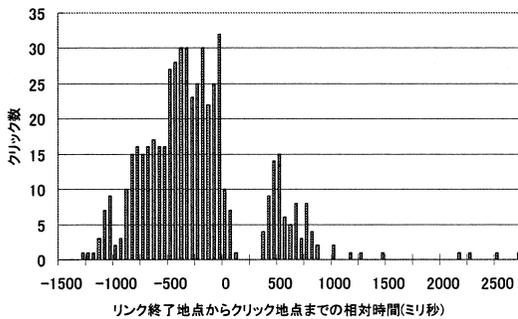


図7 リンク終了地点からクリック地点までの相対時間

Fig. 7 Relative time from a link end point to a click point.

に示す。図5にあるように、携帯利用可能な小型のトラックボールを使用することで、ウェアラブルコンピューティングにおける操作が可能となる。

5. 考 察

5.1 操作性

本節では、データ閲覧時の操作性、データ作成時の操作性、ハンズフリーの操作方法について考察する。

5.1.1 データ閲覧時の操作性

SoundWeb システムでは、リンク音が鳴っている間に左ボタンをクリックするという直観的な操作で、ユーザはリンクをたどることができる。しかし、新しいリンクを聞いたときなど、リンク部分の内容を確認するためには、ユーザは1通りその部分を聞く必要がある。リンクをたどる場合は、1通り聞いた後で巻き戻しやリンクの頭出し操作を行う必要がある。そこで、リンク部分終了後にある程度のクリック可能な時間をシステム側で設定することで、巻き戻しなどの操作をしなくともリンクをたどれるようになると考えられる。

実装した SoundWeb システムの操作性を調べるため、10問の問題の解答を総時間約15分のコンテンツを閲覧して探す簡易実験を行った。被験者は14名(平均年齢22.6歳)で全員が SoundWeb システムの初心者である。なお、リンク部分が過ぎてからクリックを行うまでの時間を調べるため、リンク終了部分から次のリンクまでもクリック可能としている。

実験の結果を図7に示す。図7の横軸は、リンク終了部分からクリック地点までの相対時間を表している。縦軸は、横軸の値を50ミリ秒ごとに分けたそれぞれの範囲に入るクリックの回数の和を表している。

実験結果から、ユーザは8割近くの操作においてリンク音再生中にクリックしており、リンク部分を過ぎ

た場合でも一部を除き1,000ミリ秒以内でクリックできていることが分かる。このことから、リンク部分を聞いてからクリックまでにかかる時間は、1,000ミリ秒程度であると考えられる。つまり、システム側でリンク部分終了後の1,000ミリ秒間をクリック可能にすることで、その間は巻き戻しせずにクリックできるようになり、現在の情報の提示方法でも、コンテンツ閲覧の効率を上げることができると考えられる。

一方、次のリンクとの間隔が400ミリ秒以下のリンクをクリックするときに、クリックが間に合わず、次のリンクをクリックしてしまうという操作が多く見られた。Cardらの人間情報処理モデル⁴⁾によると、音を聞いてからユーザがクリックするまでには、最低で240ミリ秒の時間遅れが生じることが分かっている。本システムの操作では、クリックする前にユーザがリンクの内容を把握する必要があるため、クリックするまでに400ミリ秒程度時間がかかってしまい、結果としてリンク間がつまっているときに操作ミスが発生したと考えられる。つまり、コンテンツを作成するときにリンクどうしの間隔をできるだけ広くとることで、データ閲覧時の操作性を向上できると考えられる。

今回の被験者は、全員が日頃からコンピュータに触れており、ボタンのクリックには慣れていて、クリックに慣れていないユーザは、リンク部分を聞いてからクリックをするまで余計に時間がかかってしまうが、システム側でリンク終了後のクリック可能時間を延長することで対応できる。そのため、さまざまなユーザが SoundWeb システムを利用可能であると考えられる。

今回の実験中は、問題の解答を探す以外は特に並行して作業は行わなかった。SoundWeb システムと並行して行う作業としては、業務や日常生活での作業など、さまざまなものを想定している。並行して行う作業によってクリックまでの時間が変化すると考えられるが、今回得られた結果を基にクリック可能時間を調整することで、それぞれの作業に対応可能と考えられる。

システムを利用する時間はコンテンツに大きく依存する。今回の実験のように構造化されたコンテンツを閲覧する場合は、クリックなどの操作を頻繁に行う必要があるが、選曲のための音声メニューとして利用する場合はユーザが操作する時間はほんのわずかである。

SoundWeb システムは、通常の Web と同様幅広い用途を想定している。しかし、今回の実験に対して、複数回リンクをたどると以前に聞いていた内容の把握が困難になるなどのコメントが出ており、複雑なものに利用するには現在のシステムではまだ機能が不十分で、今後拡張していく必要がある。

5.1.2 データ作成の操作性

プロトタイプシステムではデータ作成方法として、オーディオベースと GUI ベースの 2 つの方法を実装している。オーディオベースの作成方法は、作成の際に画面を見る必要がなく、ウェアラブルコンピューティング環境において、即座に短いデータを作成したりするのに適している。たとえば、散歩をしながらちよつとしたメモをとったり、ニュースのように鮮度が求められる情報を事件があったその場で作成したりすることができる。一方、GUI ベースのデータ作成方法では、ファイル管理を GUI を用いて行えるため、大量のファイルを扱うことが容易である。また、リンク部分の範囲を音声の波形を見ながら設定したり、OPTION などの各属性を指定したりできるなど、細かな設定を行うのに適している。

構造を持つ音声データの作成および閲覧に関する研究として Audio Notebook^{17),18)} や VoiceNotes¹⁶⁾ がある。これらの研究では、音声を録音する際に、録音時間をノート上の位置と関連付けたり、内容に応じたカテゴリに分けて保存するなどして、データの作成や閲覧を容易にしている。こうした方法を用いることで、構造定義ファイルの作成やリンク先の設定などの作業をより簡単にできると考えられる。

5.1.3 ハンズフリーの操作方法

現在 SoundWeb システムの操作には、マウスのボタン入力を用いている。ボタン入力で操作を行うことで、ユーザは周りの環境に左右されずに確実に操作を行うことができるが、操作時に片手が塞がるという問題がある。ハンズフリーで操作を行う方法としては音声認識による入力が広く知られている。

音声入力では、直接操作コマンドを発話するなどと言語情報を利用して入力を行う方法⁸⁾ が一般的で、音声入力を用いて Web ページの閲覧を行うシステムに関する研究も行われている^{19),25)}。言語情報を用いない音声入力に関しては、Igarashi ら⁵⁾ が音本来の持つ音響信号としての側面に注目した入力方法を提案している。これらの音声入力方法を用いることで、SoundWeb システムをハンズフリーで操作できると考えられる。ただし、音声認識は周りの雑音に強く影響を受けるうえ、周囲の人に迷惑をかけてしまう場所では利用できないなど使用環境は限定されてしまう。そのため、現在の操作方法との使い分けが重要になると考えられる。

5.2 情報の提示

今回の実験後に得られたコメントから、複数回リンクをたどると以前聞いていた内容の把握が困難になることや、音声の再生時間がまだどの程度残っているか

分からないことなど、5.1.1 項で述べたこと以外の問題が明らかになってきた。こうした問題を解決するため、カクテルパーティ効果³⁾などの音本来が持つ特性を生かした新しい情報の提示方法について考えてみる。

音の特性を生かした情報提示方法に関する研究として、音声提示型インタフェース²²⁾や AudioStreamer¹⁴⁾、Dynamic Soundscape⁶⁾がある。音声提示型インタフェースでは、カクテルパーティ効果を利用して複数の選択候補を 1 度に再生しユーザに提示する手法を提案しており、その評価結果から複数の候補を聞き分け選択することができるということが分かっている。AudioStreamer では、3 つの音声データを同時に聞き、頭を振るというジェスチャ入力により音声データの 1 つに注目できる。1 つの音声データに注目することで、音声データを聞き分け閲覧することができる。Dynamic Soundscape は音の再生する位置を仮想の 3 次元空間の中で時間的に移動させることによって、音の時間軸を空間軸に置き換えて音声データの閲覧を行うシステムである。音声の再生位置が仮想の 3 次元空間を左から右へ半円を描いて移動したとすると、ユーザは以前に聞いた内容へ、たとえば左 45 度の付近で再生されていた、というように空間的にアクセスすることができる。

これら音の特性を生かした情報提示を SoundWeb に適用し、たどってきた音声を同時に再生することで、これまでの履歴情報を提示できると考えられる。また、音声の再生位置を仮想の 3 次元空間の中で移動させることで、場所に応じた現在の再生時間を提示できると考えられる。これらの音の特性を生かした情報提示方法が本当に有効かどうか、今後検討を行っていく予定である。

5.3 データ形式

SoundWeb システムのプロトタイプシステムで使っているデータ形式について考察する。

5.3.1 音声データ

プロトタイプシステムでは音声データの形式として、作成および再生の容易さを考慮し、WAVE 形式をとっている。しかし、WAVE 形式はデータ容量が大きくなってしまいうえ、無線 LAN や Bluetooth などの無線ネットワークで利用する場合、ダウンロードに時間がかかると考えられる。今後、ダウンロード時間短縮のために、リンク先の音声データを音声再生の裏で先読みすることや、MP3 (MPEG-1 Audio Layer-3) など比較的データ容量が小さい他のファイル形式へ対応することが考えられる。

5.3.2 構造定義ファイル

構造定義ファイルに HTML 形式を採用しているため、SoundWeb データは Internet Explorer や Netscape Navigator など広く利用されている Web ブラウザでも聞くことが可能である。ただし、HTML 形式では bgsound タグと a タグの間に包含関係がないなど、音声の構造を十分記述できていない。今後、SoundWeb システムの機能を追加していくにあたり、HTML より柔軟に構造を記述できる XML (eXtensible Markup Language) を用いることについて考えていく予定である。

6. 応 用

SoundWeb システムの応用として多くのサービスが考えられる。

6.1 音声案内サービス

SoundWeb システムを利用することで、従来からある携帯電話などの音声案内サービスをより閲覧の快適なものにできる。たとえば、従来のサービスでは「現在の天気予報を聞きたいときは 1 番のボタンを、今月のイベント情報を聞きたいときは 2 番のボタンを押してください」といった案内を 1 通り聞いた後、指定のボタンを押して聞きたい情報を選択する。一方 SoundWeb システムを用いた音声案内サービスでは、「コンテンツとしては、現在の天気予報や、今月のイベント情報があります」という音声の 現在の天気予報と今月のイベント情報の部分にリンクが付けられており、ユーザはクリックすることで聞きたい情報を即座に選択することができる。リンク選択がリンクが登場したその場で行えるため、特に目的の情報が明確な場合に、無駄な情報を聞く時間を省くなどの効果が期待できる。また、初めから音声のみで閲覧することを想定してデータ作成を行うため、視覚障害者向けに音声で情報提供するサービスを構築する際に効果が期待できる。

6.2 道案内サービス

カーナビゲーションなどの道案内サービスは、GPS (Global Positioning System) からの位置情報と地図情報を用いてサービスを行っている。SoundWeb システムを用いると、地域を限定することにより、位置情報を利用せずに道案内サービスを実現できる。

例として、図 8 にこれまでに実現した大阪大学の最寄駅である北千里駅から大阪大学までの道案内サービスのフローチャートを示す。北千里駅に着いて駅周辺の道案内サービスへアクセスすると、「改札を出て右手の歩道橋を渡ってください」といった案内を聞くこと

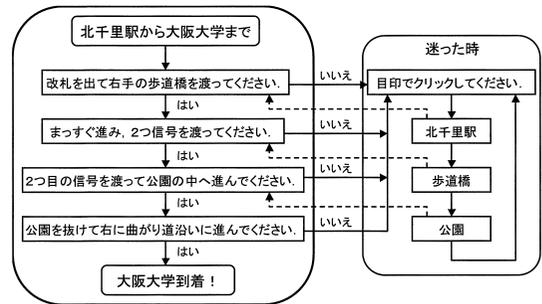


図 8 道案内サービスのフローチャート
Fig. 8 A flowchart of guidance service.

ができる。ユーザは、目的地まで目印となる建物などの情報を順番に聞きながら進んでゆく。目印を見失ったり、道を間違えたりしたときは「近くに見えるものがあればそこでリンクをたどってください。北千里駅が見えますか？ 歩道橋が…」のように道案内サービスが目印となる建物などを順番に読み上げてくれ、それぞれリンクをたどることで正しい道までの案内を聞くことができる。案内の内容は実際に道を歩きながら作成することで、聞いている人の立場に立った案内をすることができる。道に迷ったとき以外はリンクのクリックにリアルタイム性はそれほど要求されないため、道案内サービスは一般ユーザに対して有用と考えられる。

6.3 遊園地内での混雑情報の配信サービス

現在、アトラクションの混雑情報の配信サービスを、ラジオ放送など用いて行っている遊園地があるが、すべての情報がそのまま流されておりユーザが欲しい情報を取得するまでに待ち時間が発生してしまう。SoundWeb システムを応用した混雑情報の配信サービスを考えると、配送する情報を構造化することでユーザは音声メニューなどのリンクをたどって欲しい情報へアクセスできるようになる。また、興味がない情報を聞く必要もなくなる。混雑情報が更新された際は、リンク先が変更になりリンク音がたどったことのないリンクへと変わるため、ユーザは情報の更新を容易に知ることができる。音声メニューではリンク部分となるキーワードが並べられて、リンクどうしの感覚が狭くなることが考えられる。そのため、サービス対象者としては、クリック操作に慣れているユーザが望ましい。図 9 に遊園地案内の情報配信サービスのイメージを示す。図は、展望台は混んでいており、滑り台が空いているという現在の混雑情報をユーザが聞いている様子を表している。

6.4 街角情報案内サービス

SoundWeb システムを用いることで、商店街などの



図9 遊園地案内の情報配信サービス

Fig. 9 An image of the guidance system for an amusement park.



図10 街角情報案内サービス

Fig. 10 An image of guidance service of town information.

街角情報案内サービスも実現することができると考えられる。このサービスでは、街角放送を用いて駅周辺の情報を配信する。コンテンツとしては、たとえば、駅周辺にある店の宣伝や観光スポットの情報、6.2節で述べた道案内などが考えられる。本サービスは、コンテンツの各情報へはトップの音声メニューからリンクを選択することでアクセスし、そのアクセスも、目的となる情報が明確であればスムーズに行えるなど、6.1節や6.3節で述べた要素が含まれている。SoundWebシステムに慣れたユーザであれば、本サービスを十分に活用でき、ウェアラブル環境下での音声による情報閲覧を行えると期待できる。図10に街角情報案内サービスのイメージを示す。図では、本屋の宣伝を聞いたユーザが詳しい内容を閲覧した後、道案内を聞きながらその店へ向かっている様子が示されている。

6.5 インタラクティブなサウンドノベル

SoundWebシステムをエンターテインメントの分野に応用すると、インタラクティブなサウンドノベルが実現できると考えられる。

サウンドノベルとは、音と音声为主体として進行する小説のことであり、物語は登場人物のセリフと周りの状況を表す音が同時に再生されて進行する。SoundWebシステムでサウンドノベルを実現する場合、セリフや周りの音にリンクを付け、そのリンクをたどることでユーザはインタラクティブに物語へ介入することができるものとする。たとえば、ミステリー小説の場合、探偵や刑事、容疑者などの登場人物のセリフと周りの状況を表す音に付加されたリンクのたどり方で、事件を解決できたり、迷宮入りさせてしまったりする。リンクが付けられた音の再生タイミングを変化させたり、問題解決の鍵となるリンクをわざとクリックしにくく設定したりするなどによりゲーム性を高められ、特にゲーム好きの一部のユーザに対して効果が期待できる。

7. まとめ

本論文では、ウェアラブルコンピューティングにおいて、音声データを手軽に閲覧でき、作成も可能なハイパーリンク型ナビゲーションシステムを構築することを目的として、SoundWebシステムを提案し、プロトタイプシステムの実装を行った。ハイパーリンクの概念を導入し、音声データの構造化と副音声を利用した音声のハイライト手法により、個々の音声データ間の細かい関連付けが可能となり、関連情報の閲覧が容易化された。データ作成のためGUIベースとオーディオベースの2種類の方法を実装し、散歩しながらなどさまざまな状況での作成が可能となった。

今後の課題としては音の特性を生かした情報提示方法の検討や、SoundWebシステムの運用実験によるより本格的な評価などがあげられる。また、音声データの先読みなどを用いることで、リンクをたどる際のタイムラグの低減を行うことも課題としてあげられる。

謝辞 未筆ながら、本研究を進めるにあたって西尾研究室の諸氏には有益なコメントを多数いただいた。ここに衷心より感謝の意を表す。なお、本研究は、文部科学省21世紀COEプログラム(研究拠点形成費補助金)と、文部科学省振興調整費「モバイル環境向P2P型情報共有基盤の確立」、文部科学省特定研究領域(C)「Grid技術を適応した新しい研究手法とデータ管理技術の研究」(プロジェクト番号:13224059)によっている。ここに記して謝意を表す。

参考文献

- 1) Arons, B.: Authoring and transcription tools for speech-based hypermedia, *Proc. 1991 Conference, American Voice I/O Society*, pp.15-20

- (1991).
- 2) Arons, B.: Hyperspeech: navigating in speech-only hypermedia, *Proc. 3rd Annual ACM Conference on Hypertext*, San Antonio, TX, New York, pp.133-146 (1991).
 - 3) Arons, B.: A review of the cocktail party effect, *Journal of the American Voice I/O Society*, Vol.12, pp.35-50 (1992).
 - 4) Card, S.K., Moran, T.P. and Newell, A.: *The Psychology of Human-Computer Interaction*, Lawrence Erlbaum Assoc. (1983).
 - 5) Igarashi, T. and Hughes, J.F.: Voice as sound: using non-verbal voice input for interactive control, *Proc. 14th Annual Symposium on User Interface Software and Technology, ACM UIST'01*, Orlando, FL, pp.155-156 (2001).
 - 6) Kobayashi, M. and Schmandt, C.: Dynamic soundscape: mapping time to space for audio browsing, *Proc. CHI '97 on Human factors in computing systems*, pp.194-201 (1997).
 - 7) Mann, S.: Definition of wearable computer, *Keynote Address for The First International Conference on Wearable Computing ICWC '98*, Fairfax VA (1998).
 - 8) Morley, S., Petrie, H., O'Neill, A.M. and McNally, P.: Auditory navigation in hyperspace: design and evaluation of a non-visual hypermedia system for blind users, *Proc. 3rd international ACM conference on Assistive technologies*, pp.100-107 (1998).
 - 9) Rollins, S. and Sundaresan, N.: AVoN calling: AXL for voice-enabled web navigation, *Proc. 9th International World Wide Web Conference, Computer Networks*, pp.533-551 (2000).
 - 10) Roy, D.K. and Schmandt, C.: NewsComm: a hand-held interface for interactive access to structured audio, *Proc. CHI '96 on Human Factors in Computing Systems*, New York, pp.173-180 (1996).
 - 11) Roy, D.K.: Speaker indexing using neural network clustering of vowel spectra, *International Journal of Speech Technology*, Vol.1, No.2, pp.143-149 (1997).
 - 12) Rudnicki, A.I., Reed, S.D. and Thayer, E.H.: SpeechWear: a mobile speech system, *Proc. IC-SLP'96*, Philadelphia, PA, Vol.1, pp.538-541 (1996).
 - 13) Sawhney, N. and Schmandt, C.: Nomadic radio: scaleable and contextual notification for wearable audio messaging, *Proc. CHI '99 conference on Human factors in computing systems*, Pittsburgh, PA USA, pp.96-103 (1999).
 - 14) Schmandt, C. and Mullins, A.: AudioStreamer: exploiting simultaneity for listening, *Conference companion on Human factors in computing systems*, pp.218-219 (1995).
 - 15) Smailagic, A. and Siewiorek, D.P.: Modalities of interaction with CMU wearable computers, *IEEE Personal Communications*, Vol.3, No.1, pp.14-25 (1996).
 - 16) Stifelman, L.J., Arons, B., Schmandt, C. and Hulteen, E.A.: VoiceNotes: a speech interface for a hand-held voice notetaker, *Proc. CHI '93 conference on Human factors in computing systems*, pp.179-186 (1993).
 - 17) Stifelman, L.J.: Augmenting real-world objects: a paper-based audio notebook, *Proc. CHI '96 conference companion on Human factors in computing systems*, pp.199-200 (1996).
 - 18) Stifelman, L.J., Arons, B. and Schmandt, C.: The audio notebook: paper and pen interaction with structured speech, *Proc. SIG-CHI on Human factors in computing systems*, pp.182-189 (2001).
 - 19) Tan, C.L. and Pathmasuntharam, J.S.: Web navigation in small communication/computing devices using voice recognition, *Proc. ACTS Mobile Communications Summit*, Aalborg, Denmark, pp.519-524 (1997).
 - 20) VoiceBrowsers.
<http://www.w3.org/TR/NOTE-voice/>
 - 21) VoiceXML Forum.
<http://www.voicexml.org/>
 - 22) 梅本あずさ, 柴尾忠秀, 水口 充, 浦野直樹: 音声提示型インタフェースの実装と評価, インラクティブシステムとソフトウェア VII: 日本ソフトウェア学会 WISS '99, pp.83-92 (1999).
 - 23) 庄司 武, 中村聡史, 塚本昌彦, 西尾章治郎: SoundWeb: ウェアラブル環境下でハイパーリンク型音声データのブラウジングを行うシステム, インラクティブシステムとソフトウェア IX: 日本ソフトウェア学会 WISS '01, pp.11-16 (2001).
 - 24) 惣田一幸, 角谷和俊, 上原邦昭: 携帯情報端末における音声を用いた Web ナビゲーション, 電子情報通信学会データ工学ワークショップ論文集 (2001).
 - 25) 西本卓也, 新美康永: ネットサーフィン支援のための音声対話システム, 第 12 回人工知能学会全国大会, S7-03, pp.141-142 (1998).
 - 26) 平松治彦, 角谷和俊, 上原邦昭: 音声読み上げとアンカー同期表示を用いた携帯情報端末のための Web ナビゲーション, 分散, 協調とモバイルシンポジウム DICOMO '01, pp.405-410 (2001).

(平成 14 年 10 月 2 日受付)

(平成 15 年 3 月 4 日採録)



庄司 武

2002年大阪大学工学部情報システム工学科卒業。現在、同大学大学院情報科学研究科マルチメディア工学専攻博士前期課程在籍。ウェアラブルコンピューティングに興味を持つ。



中村 聡史

1999年大阪大学工学部情報システム工学科卒業。現在、同大学大学院工学研究科情報システム工学専攻博士後期課程在籍。インタフェース、ウェアラブルコンピューティング、P2Pに興味を持つ。ヒューマンインタフェース学会会員。



塚本 昌彦 (正会員)

1987年京都大学工学部数理工学科卒業。1989年同大学大学院工学研究科博士前期課程修了。同年、シャープ(株)入社。1995年大阪大学大学院工学研究科情報システム工学専攻講師を経て、1996年同専攻助教授、2002年より同大学大学院情報科学研究科マルチメディア工学専攻助教授となり、現在に至る。工学博士。ウェアラブルコンピューティング・ユビキタスコンピューティングに興味を持つ。ACM, IEEE等8学会の会員。



西尾章治郎 (正会員)

1975年京都大学工学部数理工学科卒業。1980年同大学大学院工学研究科博士後期課程修了。工学博士。京都大学工学部助手、大阪大学基礎工学部および情報処理教育センター助教授を経て、1992年大阪大学大学院工学研究科情報システム工学専攻教授。2002年より同大学院情報科学研究科マルチメディア工学専攻教授となり、現在に至る。2000年より大阪大学サイバーセンター長を併任。この間、カナダ・ウォータールー大学、ピクトリア大学客員。データベース、知識ベース、分散システムの研究に従事。現在、ACM Trans. on Internet Technology, Data & Knowledge Engineering, Data Mining and Knowledge Discovery, The VLDB Journal等の論文誌編集委員。本学会フェロー含めACM, IEEE等9学会の会員。