

## Bilingual Emacspeak — A universal speech interface with GNU Emacs —

渡辺隆行<sup>1a</sup>、井上浩一<sup>2</sup>、坂本貢<sup>3</sup>、本多博彦<sup>4</sup>、釜江常好<sup>5</sup>

湘南工科大学情報工学科<sup>1</sup>、(株)リコーソフトウェア研究所<sup>2</sup>、国立職業リハビリテーションセンター<sup>3</sup>、  
東海大学理学部<sup>4</sup>、広島大学大学院理学研究科<sup>5</sup>

Bilingual Emacspeak は、Emacs エディタからスピーチサーバーを制御して Emacs を音声利用するシステムである。このシステムは Raman 博士の Emacspeak の拡張であり、Lisp 系のプログラム言語を用いて情報の性質に応じて聴覚に最適化した形で情報を処理する。つまりカレンダーならば 2 次元構造を保つような方法で情報を出だし、プログラムソースならば文の種類に応じて声を使い分ける。また日米両言語を自動的に切り替えて正確に発音する。このように本システムは既存の音声化システムと比較してはるかに豊かな情報処理能力と表現力を持っており、視覚障害者の高度なコンピュータ利用や晴眼者とコンピュータの新しいインターフェースとして利用できる。Windows で動作する本システムを使って、音声による Emacs 利用の可能性を示す。

## Bilingual Emacspeak — A universal speech interface with GNU Emacs—

T. Watanabe<sup>1</sup>, K. Inoue<sup>2</sup>, M. Sakamoto<sup>3</sup>, H. Honda<sup>4</sup>, and T. Kamae<sup>5</sup>

Dept. Information Science, Shonan Inst. Tech.<sup>1</sup>, Software Research Center, RICOH Co. Ltd.<sup>2</sup>,  
National Vocational Rehabilitation Center for the Disabled<sup>3</sup>, Tokai University<sup>4</sup>, Hiroshima University<sup>5</sup>

The Bilingual Emacspeak is the self-speaking Emacs editor that controls a bilingual, Japanese and English, speech server. The aim of the current system is to enhance Japanese visually impaired graduate students, researchers, and programmers to use computers at high levels. The current system, which is implemented as an extension of Dr. Raman's Emacspeak, processes information by Lisp to provide it in the best way for the sense of hearing. The current system has richer expressive power than the conventional speech systems and can be used as an effective speech interface to sighted users as well as visually impaired ones. The potential of Emacs with a dedicated speech feedback will be shown by the Bilingual Emacspeak for Windows.

---

\*takayuki@la.shonan-it.ac.jp

# 1 はじめに

パソコンの普及によって、視覚障害者が点字以外の文字を書き、電子情報を盲人者と共有できるようになった。しかし現代の高機能な OS を視覚障害者が利用しようとすると、Windows では GUI が障害となり [1]、UNIX では音声化ソフトウェアが開発されていないことが障害となる [2]。我々が取り組んでいる Bilingual Emacspeak Project (BEP)[2] は、Raman 博士が開発した Emacspeak を元にしてこのような状況を打開することを目指している。つまり文字ベースである Emacs エディタを共通の基盤として利用し、Emacs が扱う情報を聴覚に最適化して音声化することで、Emacs が提供する様々な機能を音声利用する。つまり、既存のスクリーンリーダーが OS を音声化しているのに対し、本システムは Emacs を音声化する。

GNU Emacs は UNIX を利用する研究者や技術者の間で広く利用され標準となっているエディタで、UNIX や Windows をはじめとする多くの OS で動作する。Emacs はテキスト編集以外にも、仮想端末、ファイラー、Web ブラウズ、電子メール、ftp、telnet、辞書検索、PIM、プログラム編集、ゲームなどのいろいろな機能をエディタ内から実行できる。また一般アプリケーションでは実装されていない先端の機能がいち早く実装されるアプリケーションでもある。さらに欧米語、アジア言語などのほとんどの文字コード体系をサポートしている。この Emacs に効率的で分かりやすい音声出力インターフェースを追加すれば、視覚が利用できない状況にあるすべてのユーザーがユニバーサルに Emacs を利用できるようになる。

本報告では、最初に Emacspeak の特徴を述べ、次に本システムが追加する日本語特有の機能について述べ、最後に本システムがもたらす音声インターフェースの可能性を議論する。

## 2 Emacspeak

図 1 に示すように、Emacspeak<sup>1</sup>は Auditory User Interface (AUI)[3, 4] を使って Emacs を音声化す

<sup>1</sup><http://emacspeak.sourceforge.net/> が公開サイト。

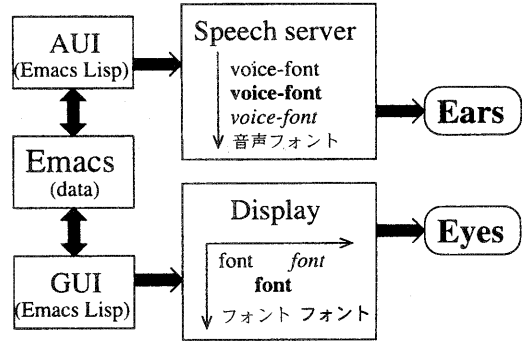


図 1: Emacspeak は、2次元の画面に情報を表示する GUI とは独立した AUI が、1次元のシーケンシャルな音声ストリーム専用に情報を処理してユーザーに提供する。

る。音声表示用の AUI は画面表示用の GUI とは独立に情報の論理構造にアクセスして必要な情報を取り出し、聴覚に適した形で音声化する。Emacspeak は buffer の mode に応じた最適な AUI を Emacs Lisp で実装しており、calendar-mode ではカレンダーに適した機能を供給でき、W3-mode では CSS2 (Cascading Style Sheets, level 2)<sup>2</sup>の音声スタイルシートに対応した Web ブラウザとなる。

GUI 画面に表示される文字には複数のフォントがあり、さらに太字やイタリックなどの修飾をしたり大きさを変えたりできる。音声においても、ピッチなどの声の個性を変えたり、男女を変えたり、年齢を変えたり、ボリュームを変えたりすることで、テキストの読み上げに対して多様な修飾ができる。Emacspeak には audio formatting や voice-lock という機能があり、プログラムソースを編集するときにはコメント、キーワード、関数名、文字列、などの構成要素ごとに異なった声を割り当てたり、大文字の場合は短いピープ音を鳴らして大文字であることを示したり声の調子を変えることで、どこが大文字か耳で聞いただけでわかるようになっている。

このように Emacspeak は、全盲の大学院生であった Raman 氏が自分自身の博士論文執筆のために開発した極めて高機能な音声化システムである。しか

<sup>2</sup><http://www.w3.org/TR/REC-CSS2/> 参照。CSS2 では、画面、音声出力装置、プリンタ、点字出力装置などの異なった出力装置ごとにスタイルシートを指定できる。

しながら、Emacs を知らなければ使えない、UNIX 以外の OS を考慮していない、英語以外の言語を全く考慮していない、スピーチサーバーに対するコマンドが DECTalk という特定の音声合成装置に特化している、などの限界があり、米国においても一般ユーザーには普及していない。

### 3 Bilingual Emacspeak

豊かな音声表現力を持つ Emacspeak を英語と日本語を扱えるように拡張し、Windows 及び Linux の両方で利用できるようにしようというのが本システムの目的である。本システムは、buffer 上の情報を処理する Emacs Lisp のパッケージ (AUI) と、受け取ったテキストを音声化してからサウンドデバイスに出力するスピーチサーバーからなる。スピーチサーバーは日本語及び英語の音声合成 (text-to-speech, TTS) エンジンとそれを制御するプログラムからなり、Windows と Linux 用の 2 種類のスピーチサーバーを開発している。AUI となる Emacs Lisp のほとんどは Windows、Linux の両 OS で共通する。2 つの OS をターゲットにしているのは、個人レベルでは Windows が最も普及していて、仕事では Linux などの UNIX を使っている視覚障害者が多いからである。

本システムは Emacs が好きな人間が集まり、Emacspeak の多言語化及び Windows 用及び Linux 用の日米対応スピーチサーバーの開発にオープンに取り組んでいる。Windows 版の Bilingual Emacspeak は少人数のユーザーテストを実施するところまで開発が進んでおり、一般公開の準備中である [5]。

#### 3.1 日本語の音声表現

Emacspeak に日本語を処理させるためには、AUI 部、スピーチサーバー部ともに多言語対応化 (multilingualization, m17n) しなければならない。この作業はソースコードをマルチバイト文字に対応させる作業と、日本語の音声出力に適した機能の実装を含む。さらに日本語の音声出力に適した表現方法には、以下の 2 種類がある。

##### 説明読み

漢字は表意文字なので複数の同音異義語があり耳で聞いただけでは漢字を同定できないので、既存の日本のスクリーンリーダーは注釈をつけて漢字を説明する詳細読み機能を備えている。漢字の詳細読みが必要になるのは、かな漢字変換時と、カーソルを動かして一文字ずつ文字を調べるカーソル読み時である。本システムでは、かな漢字変換時の詳細読み時に、だれが聞いても漢字を同定できる分かりやすい読み方と、慣れた人ならすぐに同定できる早い読み方を切り替えられるようにする予定である。またカーソル読み時は短い説明で読み上げないと、時間がかかってわずらわしいので、グラスルーツなどのスクリーンリーダにならった短い説明の辞書を用意する。

日本語にはひらがなの他にカタカナと半角カタカナによる表現があるが、読み上げるとこれらはすべて同じ音になる。音声でこれを区別する方法のひとつは、「カタカナの あ」というふうに注釈をつけることであるが、読み上げに時間がかかる。全角と半角の英数字を区別すべき場合もある。また、日本語だけの問題ではないが、記号などを読み上げる際はどの記号であるかという説明をつけて読み上げた方が 良い場合と、無視して音声化しない方が 良い場合がある。アルファベットの太文字と小文字の区別が必要になる場合もある。

以上からわかるように、日本語を音声化するときは操作の状況とユーザーの要求にしたがって、読み上げ方を使い分ける必要がある。本システムではこの読み上げの切り替えをある程度自動化した上でユーザーが自由に切り替えることができるようにする。例えば、一文字読み上げをする際は、カーソル移動に伴う読み上げ時は短い説明で読み上げ、一文字読み上げのコマンドによる場合は 1 回目は簡単な詳細読みを使い、引き続いて 2 回同じコマンドが実行されたときは、詳しい詳細読みをする予定である。

##### 音声フォント

わかりやすい音声化には、以上で示した説明読みの他に、声を使い分けて読み上げる方法があり、これは既存の日本のスクリーンリーダーでは限定的にしか使われていない本システム独自の機能である。

ひらがなとカタカナ、大文字と小文字などを識別する際に、文字の種類に応じて声のピッチや種類などの声の性質を使い分けることにより、注釈をつけなくても文字の種類を識別できる (実装中)。同じこと

は全角と半角の英数字にも言えるが、半角の英数字は英語のエンジン、全角の場合は日本語のエンジンが読み上げれば簡単に識別できる。なお Emacspeak では大文字と小文字のアルファベットを識別するのに、大文字の所だけアクセントをつける方法、大文字の前に短い音 (audio queue) を鳴らして識別する方法、注釈をつける方法などを、場合に応じて使い分けている。

また本システムは、Emacspeak が備えている、コンテキストに応じた声の使い分け機能を日本語でも使用できる。これによって、プログラムソースや Web ページや電子メールなどの読み上げ時に文の種類に応じた声を使い分けることで、データの構造をわかりやすく音声表現できる。Windows 版は英語と日本語それぞれに対し、10 種類近い声を持っている。

### 3.2 ユーザーテスト

Windows 版 Bilingual Emacspeak は、BEP メーリングリスト<sup>3</sup>参加者によるテストが行われている。そこで報告された本システムの特徴をまとめる。

- 今までは音声化した DOS 端末などからログインして使うしかなかったが、使いたれた Emacs を直接 Windows でも使用できるようになった。
- MUA の Mew<sup>4</sup>を使えるようになったので、電子メールの処理を Bilingual Emacspeak 上の Mew に乗り換えることができた。
- コンテキストに応じた音声修飾による表現力の豊かさに驚いた。耳で聞いていてもプログラムの構造が良くわかる。
- ftp も Emacs の directory editor モードで使えるので便利。
- Web ブラウズも Emacs の W3 モードでかなり使える。また音声スタイルシートを使えるので、リンク以外の構造も聞き分けることができる。
- News などでも使えそうなので期待している。

<sup>3</sup>BEP メーリングリストの参加方法は BEP の Web サイト; <http://www.argv.org/bep/> 参照。

<sup>4</sup><http://www.mew.org/>

● Emacs の中だけでかなりの作業ができるので便利。

● 英語は極めてレスポンスが良く、日本語のレスポンスもかなり良い。

● 数単語程度の短い英語ならば日本語でも聞けるが、パラグラフ単位の長い英語は日本語エンジンの発音では聞くに耐えなかった。しかし本システムは言語に応じてエンジンを切り替えるので、長文がとでも聞きやすい。

● 英語のヒアリングの練習になる。

● ドキュメントを整備しないと、Emacs を知らない人には使えない。

● Windows 用のスクリーンリーダーとは全く違う機能を持っているので、スクリーンリーダーと比較するアプリケーションではない。スクリーンリーダーと共存すべきアプリケーションである。

● 晴眼者にとっても、Emacs の使いやすさを向上させるシステムである。耳で聞いていると文章の流れの悪さやスペルミスなどが良くわかるし、ながら作業ができる。

この他にも改良を求めるたくさんの要望やアイデアが出て、開発に参加するメンバーも増えており、一般公開の準備を進めている。

### 3.3 視覚が使えない環境などへの応用

本システムは Emacs の優れた機能を活用して、視覚障害者がコンピュータを高度に利用することを目指している。しかし晴眼者でも視覚が使えない環境でコンピュータを利用したい時がある。運転中は手を使ったコンピュータへの情報入力も視覚を経る情報出力も行えない。座れない通勤電車の中では片手キーボードなどを使った情報入力は可能であるが、通常のディスプレイを使った情報出力は困難であるし、ヘッドマウントディスプレイを使うのはかなりの勇気が必要であり画面の大きさも制約される。そのほかにも電話による情報処理や何らかの作業をしながらの情報処理では視覚を使えない。

視覚障害者が実際に使っている音声化システムは、訓練さえ積めば晴眼者でもある程度使用できるは

ずである。両者が共有できるシステムにするためには、一般社会で使用されているアプリケーションに音声インターフェースを追加する形で実装することが最も重要である。さらに、音声化速度を晴眼者用の遅めの速度から視覚障害者用の極めて速い速度まで自由に変更できるようにする、聴覚で文字情報を取得することに不慣れな晴眼者用に分かりやすく短いメッセージで情報提示するかあるいは高度な能力を持つ視覚障害者用に最短時間で本質的な情報だけを提供することもできるようにする、などの工夫が必要になる。本システムも、ユーザーのレベルに適合したインターフェースを切り替えることができるようにして、聴覚による情報取得能力の差によらずに利用できるような音声インターフェースに仕上げる予定である。このようなシステムを開発する上で、Lisp 系言語を使った高度なカスタマイズが可能な GNU Emacs をベースにしている、ソースを公開しているの自分が必要とする機能を追加しやすいオープンソースプロジェクトであるという本システムの特徴が有利になる。

本システムは情報の出力のみに音声を使っており、入力にはキーボードを想定している。これは、Emacs を使おうというようなユーザーならばキーボードに慣れており、タッチタイプをするのに視力の有無は関係ないと考えたからである。しかし、データやコマンドの音声入力を本システムと併用することも可能であり、簡単なテストも実施している。音声入出力を活用した場合、視覚障害者だけではなく肢体障害者やモバイルコンピューティングやテレフォニーなどの分野に応用することもできる。

## 4 まとめ

我々は、Emacspeak を日米 2ヶ国語に対応するように拡張し、2ヶ国語を協調して制御できるスピーチサーバーを Windows 及び Linux 用に開発している。

本システムは、言語を判別して米語と日本語の TTS エンジンを実動的に切り替えて音声化し、さらに日本語に対してもシンタックスに応じて声を変えることができる初の音声化システムである。本システムは Emacs の内部だけを音声化しているの、スクリーンリーダーとは相補的な関係にある。本システム

は Emacspeak の優れた機能をそのまま受け継ぐように実装している。さらに日本語をわかりやすくそして効率よく音声化するために、日本語の読み上げ方の使い分けなどの既存の日本語スクリーンリーダーの優れた機能も実装しつつある。

本システムによって Emacs が提供する機能のかなりの部分を、Windows 上でも視覚障害者が利用できるようになった。また本システムを一般公開すれば、世界の視覚障害者も Emacspeak を Windows で使用できるようになる。Linux 用の 2ヶ国語対応スピーチサーバーの開発作業が終了すれば、Linux 上でも日本の視覚障害者が Emacs の優れた機能を利用できるようになる。また、本システムは視覚障害者だけではなく、聴覚への情報出力が主になる環境への応用も視野にいれて開発に取り組んでいる。

Windows 版の Bilingual Emacspeak はすでに ARGV (Accessibility Research Group for the Visually impaired) のメンバーなどが日常生活でテストしており、彼らの要望を受けながら重要な部分を改善し、ドキュメントの整理などの準備を終えた後、以下の Web サイトでソース付きで一般公開する予定である。

### ARGV の BEP のページ:

<http://www.argv.org/bep/>  
(BEP のメインサイトだが回線が細い。)

### 湘南工科大学の渡辺研究室:

<http://www.shonan-it.ac.jp/each.science/info/nabeken/voice/>

### SFC の渡辺のページ:

<http://www.sfc.keio.ac.jp/~wata7be/bep/>

### ITRC の Voice Windows のページ:

<http://www.itrc.net/voicewindows/>

本プロジェクトはだれでも自由に参加できる。また本システムの一般公開時には、オープンソースとして配布する。

## 謝辞

GNU やオープンソースという言葉によって代表される知的財産の共有文化に感謝する。本システムは

Raman 博士の Emacspeak や Gary Bishop 教授のスピーチサーバーのソースを自由に使用できなければ開発できなかったし、彼らのソースコードをハックすることが良い勉強になった。

本プロジェクトは、ITRC (日本学術振興会産学協力会 163 委員会; インターネット技術研究委員会) の「視聴覚障害者のインターネット利用分科会」のメンバーと、視覚障害者のコンピュータ利用について広く議論し UNIX サーバーを立ち上げて活動している ARGV のメンバーの出会いから本格的なスタートを切った。また、ARGV のサーバーで BEP メーリングリストと、ソースコードなどの管理用の CVS サーバーを立ち上げ、メーリングリスト参加者による活発な意見交換と共同開発が行われている。インターネットがもたらしたこれらのテクノロジーと、それを生かした分散型でオープンなシステム開発の形態に感謝する。こうしたバザール型のシステム開発により、志のあるユーザーが開発初期から参加しユーザーが必要とする機能を取り入れたり、初期メンバーが不慣れな技術を新たなメンバーが助成したりすることが容易になった。

米 2ヶ国語の豊かな音声表現力を持つ Emacs”, 電子情報通信学会 音声研究会 (2000 年 8 月 24 日、北海道大学).

## 参考文献

- [1] 石川 准: “GUI用スクリーンリーダーの現状と課題 - 北米と欧州の取り組みを中心に -”, 情報処理, Vol. 36, No. 12, (1995) pp. 1133~1139.
- [2] 渡辺隆行、井上浩一、鳥原信一、飛岡正人、坂本貢、釜江常好: “Bilingual Emacspeak Project - Windows 及び Linux 用の日英 2ヶ国語対応 Emacspeak の開発 -”, 電子情報通信学会技術研究報告 WIT99 - 1 22 [福祉情報工学], (2000) pp. 71~76.
- [3] T.V. Raman: “Audio System for Technical Readings”, Ph. D thesis, Cornell University (1994).
- [4] T.V. Raman: “Auditory User Interfaces - Toward the Speaking Computer-”, Kluwer Academic Publishers (1997).
- [5] 渡辺隆行、井上浩一、坂本貢、本多博彦、釜江常好: “Bilingual Emacspeak for Windows; 日