音声入力 Web システムによる音声認識アプリケーションの構築技術

西村 竜一† 三宅 純平‡ 河原 英紀† 入野 俊夫†
† 和歌山大学 システム工学部 ‡ 奈良先端科学技術大学院大学 情報科学研究科

1 はじめに

一般的な Web ブラウザ上で動作する Web システムに，音声入力の機能を付加する枠組みである w3voice システムを開発した [1]．我々は，本システムの基本コンポーネントをフリーのソフトウェアとして公開しており，Web 開発者が自らの Web サイトに音声入力インタフェースを容易に追加できるようツール整備を行っている．

同時に，我々のプロジェクトが運用する Web サイト (http://w3voice.jp/) では，音声入力・音声認識をフロントエンドとする Web サービスの公開試験を行った．また，インターネットユーザが実際に音声認識や自動対話の音声 Web アプリケーションにアクセスした際に得られるログや発話の記録を行っている．これにより，これまでの据え置き型の公共型音声情報案内システム「たけまるくん」[2] を用いたフィールドテストでは観察することのできなかった，家庭や職場等のパーソナル空間での音声認識の利用実態を調査，分析することが可能になる．

本発表では，w3voice システムの概要と公開試験に用いたアプリケーションについて紹介する．また，ユーザからのフィードバックに基づき改良した w3voice の次期バージョンについて述べる．

2 システムの概要

本システムを用いて試作したインターネット上のフルーツショップサイトの動作画面を図 1 に示す．このサイトでは，例えば，「ミカンを 10 個ください.」「バナナとリンゴをお願いします.」等の利用者自身の発話を認識し，注文を受理できる．図 2 のように，本システムを適用した音声入力 Web ページには，通常のコンテンツの他に，音声入力パネルが表示される．音声入力パネルを使った操作の手順は以下のようになる．(1) マウスカーソルを音声入力パネルに移動，(2) マウスボタンを押している間，録音，(3) マウスボタンを離して録音終了，(4) しばらくすると対話結果が Web ブラウザに表示される（応答の合成音声も再生される）．利用者の発話をサポートするため，(2) のとき，音声入力パネルは，正しく発話が録音できているかを視覚的に確認できるレベルメータとして動作する（赤い領域が入力レベルを示す）．また，(3) の際も，送信・処理中の状態を示す視覚的フィードバックを提供する．

これまで，VoiceXML や SALT 等，音声入力 Web システムはさまざまな方法で提案がなされてきた．これらと異なり，本システムは，利用者に負担を与えない，やさしいシステムとしての以下の特徴を有する．

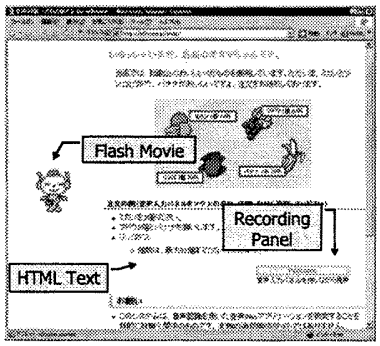

図 1: 対話型音声 Web アプリケーションの動作画面

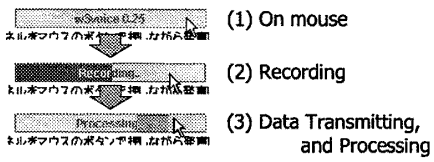

図 2: 音声入力パネル部分の拡大

・導入がやさしい

システムを利用する準備として，特別なソフトウェアの導入や設定を利用者に要求しない．図 3 に示すようにサーバサイド型アーキテクチャを採用することで実現した．クライアント PC では Java アプレットとして動作する音声入力パネルが録音とデータ送信のみを担っている．このため，該当のページにアクセスするだけで，普段のブラウザのままで利用することができる（Windows, Mac, Linux で動作する IE, firefox の主要 Web ブラウザで動作確認した）．

・PC 環境にやさしい

音声認識や対話処理等のシステムのコアは Web サーバ側で CGI プログラムとして動作する（図 3）．このため，従来の音声認識システムとは異なり，クライアント PC 与える負荷が少ない．また，録音された音声は，PCM の波形信号データのまま，HTTP の POST メソッドを用いて Web サーバに送信される．結果は，HTTP のレスポンスとして受け取り，HTTP Redirection または Ajax like な動的 HTML によってブラウザに反映される．このように，標準的なプロトコルのみで本システムのメッセージ交換は完結している．このため，通常の Web ブラウジングが可能ならばハードウェアや通信環境（ファイヤウォール等）による利用制限を受けないことも特徴となっている．

A New implementation technique for building ASR applications based on voice-enabled Web systems.
Ryuichi NISIMURA†, Jumpei MIYAKE‡, Hideki KAWAHARA†, Toshio IRINO†
† Faculty of Systems Engineering, Wakayama University
‡ Graduate School of Information Science, Nara Institute of Science and Technology

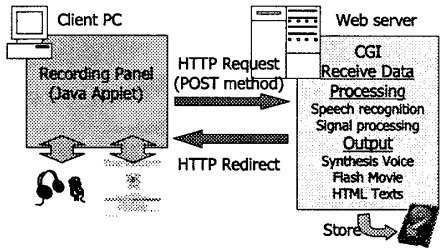

図 3: w3voice システムのアーキテクチャ: 録音，データ送信を担う Java アプレット（クライアント PC 側）と CGI プログラム（Web サーバ側）から構成する．

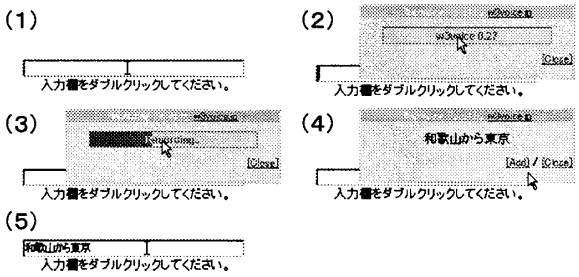

図 4: w3voiceIM.js を利用したフォームへの文字入力

3 応用事例

`http://w3voice.jp/` において公開中のアプリケーションの中から，前述の対話型以外の 2 つを紹介する．

・音声認識 Javascript ライブラリ w3voiceIM.js

w3voiceIM.js は，通常の Web ページにあるテキスト入力のフォームに対して，音声認識による文字の入力を提供する Javascript ライブラリである [3]．本ライブラリをロードするだけで，あらゆる Web ページ上で音声認識を利用できるようになる．また，Javascript を用いることで，検索エンジンサービスや Web API 等との連携が可能である．加えて，既存の Web サイトの HTML に一行を追加するだけで導入できる手軽さが特徴である．図 4 に動作例を示す．

・STRAIGHT ボイスチェンジャ

また，本システムは，PCM 録音された波形信号をサーバサイドで処理するため，音声認識等の特定の用途に限らず，さまざまな音アプリケーションのフロントエンドとして適用することが可能である．その一例として，入力音声に対し，声道長及びピッチを変化させた合成音を出力するボイスチェンジャプログラムを試作した．この中では，リアルタイム STRAIGHT[4] を音声分析・合成のエンジンとして用いた．

4 次期バージョンにおける改善

利用者から頂いた意見を参考に，w3voice の次期公開バージョンに向けた以下に述べる改良を行った．

・GZIP による信号圧縮

波形を無圧縮で送信するのは非効率であると指摘があったため，信号を圧縮するよう変更した．ただし，信号処理等の歪みとなるため非可逆な圧縮方式は利用せず，GZIP を採用することにした．

・録音と並行したストリーミング送信

これまでの音声入力パネルは，録音が終了した後に信号の送信を開始する仕様だった．これを，録音をしながら送信も並行するようにストリーミング化した．実現には，HTTP/1.1 以降で定義されている chunked 転送コーディングを用いた．

w3voice システムの利用者から寄せられる意見には，応答の遅さに対する改善の要望が多い．ネットワークを経由したデータ交換を行っているため，完全なリアルタイム処理は困難だが，応答スピードの改善は急務である．そこで，上記の改良による効果を確かめるため，実験により応答時間を調査した．w3voiceIM.js を用いてテキスト入力フォームに「和歌山から東京まで」と入力する場合の，発声開始から認識結果が表示されるまでの時間を調べた．時間情報は Javascript を用いて取得した．各 10 回試行の平均を図 5 に示す．従来の w3voice システム（original）と比較して，GZIP による信号圧縮（GZIP Comp.）による応答時間の改善は得られなかった．しかし，ストリーミング送信（GZIP + Stream）によって，応答時間の改善を得られることを確認した．

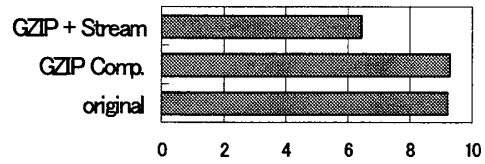

図 5: 応答時間の比較 (sec.)

5 まとめ

Web システムに音声入力・認識のフロントエンドインタフェースを提供する w3voice システムについて述べた．冒頭で述べたように，我々は，本システムの開発ツールキットをフリーソフトウェアライセンスの下で配布している（`http://w3voice.jp/skeleton/`）．アーカイブには，動作に必要なコンポーメントと，音声認識と連動するサンプルの CGI 及び PHP プログラムを収録した．本システムが音声認識アプリケーションの利用拡大の一助になれば幸いである．

謝辞 本研究は文科省 e-Society プロジェクトの支援を受けた．

参考文献

[1] 西村 他: 音声入力・認識機能を有する Web システム w3voice の開発と運用，情報処理学会研究報告，2007-SLP-68-3, 2007.

[2] 西村 他: 実環境研究プラットホームとしての音声情報案内システムの運用，電子情報通信学会論文誌，Vol.J87-D-II, No.3, pp.789-798, 2004.

[3] 西村 竜一: 音声入力 Web システムを用いた辞書共有型音声認識サービス，日本音響学会 2007 年秋季研究発表会講演論文集，pp.61-62, 2007.

[4] H. Banno et al.: Implementatioin of realtime STRAIGHT speech manipulation system: Report on its first implementation, *Acoustic Sci. and Tech.*, Vol.28, No.3, pp.140-146, 2007.