音声対話による Web フォーム入力システム: FormsTalk

辻野 克彦　　チャールズ リッチ

三菱電機(株) 先端総研
システム基盤技術部

Mitsubishi Electric
Research Laboratories

１．はじめに

音声により Web ページへの入力を行うためには、通常の HTML ページに OS レベルの音声 FEP 機能などを使って入力する方法、VoiceXML などの音声ページ記述言語を用いる方法、SALT などの音声拡張機能を実装したマルチモーダルページを構成する方法などが考えられる。

音声 FEP を用いる方法は、通常の表示系コンテンツがそのまま利用可能である点で優れているが、入力位置の指定や修正などの際にはマウスやキーなどの操作を必要とする場合が多い。

VoiceXML は、電話を使った CTI 用途などに特化したシステム主導の音声アクセスを中心としたコンテンツを簡易に記述できるという特徴があるが、表示系との共用は想定されていない。

SALT は表示系ブラウザに音声処理機能を付加するものであり、VoiceXML より複雑な音声コントロールが可能であるが、コンテンツ中に多くの手続き的スクリプト記述が必要となる。

このようにコンテンツ側で音声機能を準備しておく方法に対して、コンテンツ記述は極力モーダルに依存しない論理的記述にとどめ、それをブラウザ側で解釈することのより、マルチモーダル機能を提供する方法も考えられる。

XForms はそのような用途に対応するページ記述の１つであり、入力フォームで入力されるべきデータの論理タイプや入力フィールドの説明として表示されている注釈ラベルとの対応関係などが構造化されているため、ブラウザ側でこれらの解釈が可能である。

２．FormsTalk

本稿で述べる FormsTalk は標準的な Web ブラウザである Internet Explore 6.0 と連携して動作する Applet および Java アプリケーションとして実装されており、通常の HTML コンテンツを基本とした Web ページの表示と連動して、表示内容に応じてユーザとシステムが適切に対話の主導をとることができる混合主導の音声対話により Web フォーム入力を実現するものであり、次のような特徴を備えている；

(1) 表示ページ内に現われている項目およびそれらの関係（ラベル名と入力項目との対応関係、データタイプなど）に基づき、コンテンツ読み込み時に音声認識すべき項目（辞書や文法）を動的に決定し音声認識の性能を高めることが可能。

(2) 表示項目に関する上記情報に基づき発話内容を解釈することにより、表示ページ内の項目について、自由な順番・柔軟な発話パタンによるユーザ主導の入力処理が可能。

(3) 複数項目に関する組合せ発話から、対応する項目への自動展開が可能。

(4) 発話の曖昧性が生じた場合には、それを解消する質問を音声で出力することにより、緩やかなシステム主導の対話制御（混合主導）に移行することが可能。

図１に FormsTalk の全体構成図を示す。現在のところ、コンテンツは専用のグラフィックエディタで生成しているが、基本的には後述する仮想デバイスアプレット及びこれとブラウザとの

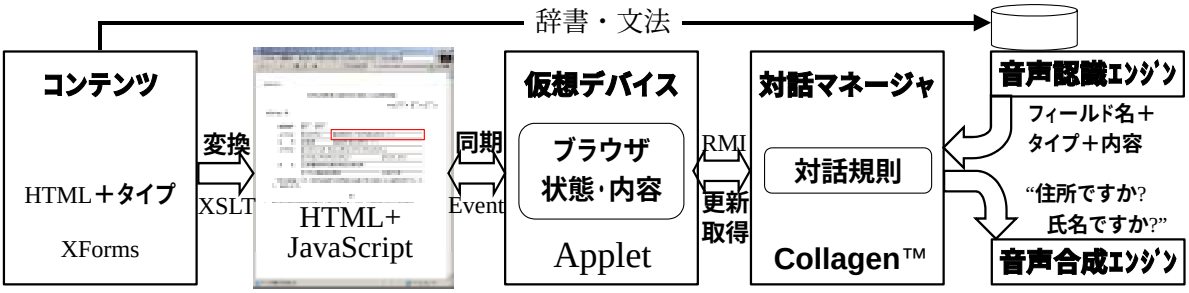

図１ FormsTalk の全体構成図

間で表示状態の同期を行うための JavaScript と入力項目のデータタイプを指定するための

FormsTalk: Web Form-Filler by Speech

Katsuhiko Tsujino, *MITSUBISHI electric corp.*
Charles Rich, *Mitsubishi Electric Research Laboratories.*

SemanticType 属性の追加を行い、入力項目と対応する表示ラベルを同一のクラス名でテーブル組みした HTML ファイルであるため、タイプ属性のみ人手で指定すれば、既存の HTML ファイルからの変換および市販の HP ビルダなどでの新規作成も比較的容易に行うことができる。

一方で XForms にはこれらの情報が含まれているため、XForms 対応のブラウザを使うか、前述のような HTML に変換するスクリプトを用意すれば、XForms コンテンツを利用することができると考えられる。現在は前述の HTML コンテンツを IE6.0 に読み込ませて利用している。

仮想デバイスはブラウザに表示されている入力項目への入力データ、フォーカス、入力項目と対応付けられるラベル内容などを、後述の対話マネージャからアクセスできる形式で保持すると共に、ユーザによるブラウザ操作および対話マネージャからの変更設定が起きた際に、ブラウザ内容と内部状態を同期させる機能をもっている。コンテンツの読み込み時に読み込まれる Applet として実装されており、ブラウザとは JavaScript により、対話マネージャとは Java ベースの RMI により連携して動作する。

音声認識エンジンはコンテンツに現われる単語と文法に基づき音声認識を行う。文法には「電話番号」や「郵便番号」などのように特定の単語列からなる汎用のデータタイプに対応するのもの(タイプ情報と呼ぶ)に加えて、入力項目と組で表示されるラベル(“勤務先電話”など)を使った代表的な発話パタンに対応するコンテンツ依存のもの(項目情報と呼ぶ)を用意した。項目情報の利用により「(私の)勤務先(の電話番号)は 012-345-6789(です)」という発声に対して、単なる「電話番号」タイプについての発話でなく「勤務先電話」という特定の入力項目についての発話であると認識することができる。音声認識エンジンからはこれらの組(“項目情報”,”タイプ情報”,”発話内容”)のリストが認識結果として得られる。

対話マネージャは音声認識エンジンから得られた音声認識結果に基づき、仮想デバイスに反映されているブラウザ表示内容を参照しながら、ブラウザへの入力操作などを仮想デバイスに対して行う。

また音声認識結果に曖昧性がある場合には、音声出力エンジンを使って曖昧性を解消するための質問を発声させると同時に仮想デバイスを介してブラウザに質問への返答を催すフレームを表示させる。この状態でユーザは質問に答えても良いし、別の項目について別の発言を行っても良い。

表1に対話マネージャで用いられている対話規則の例を示す。この規則はフォームを埋めるタスクに共通するものであり、表示コンテンツおよび言語には非依存である。

表1 対話モデルの例

- 「項目情報」が一致する入力項目があればそこを「発話内容」で埋めよ。
- 「タイプ情報」が一致する入力項目が１つだけしかなければそこを「発話内容」で埋めよ。
- 「タイプ情報」が一致する空の入力項目が複数あれば、埋め先の項目情報を尋ねよ。
- 項目の確認中に「項目情報」のみの発話があれば、その項目を確認中の「発話内容」で埋めよ。
- 項目の確認中でない場合に、「項目情報」のみの発話があればその項目にフォーカスを移動させよ。
- フォーカス中の項目と「タイプ情報」が一致する発話があれば、その項目を「発話内容」で埋めよ。

３．現状と課題

予稿完成時時点では英語版のデモシステムが稼動しており、並行して日本語版の作成を進めつつある。

また現状では確認ダイアログが出ている場合でも、確認事項以外の発話も受けつけているが、そのことが(一般に発話単語が短い)確認事項に対する返答の認識率を下げる原因となっている。辞書に重みをつけるなどの方法により、確認事項以外の発話を受け付けながらも確認事項の認識性能を重視できる方法について検討を進める。

さらに表示情報をより積極的に活用し、マルチモーダル性を高めることを目指して、例えば表組みされている入力項目について「三人目の参加者の電話番号は…です」などといった処理を可能とするべく、コンテンツ(の表示状態)から項目情報同定用の発話パタンを自動生成する方法などについて検討を進めて行きたい。

４．おわりに

Web ブラウザと連携して混合主導の音声対話によりフォーム入力を行う対話システム FormsTalk について述べた。

XForms 相当のページ記述に基づき、ページ中の項目について自由な順番・発話パタンで入力可能なこと、単一の発話で複数の項目に展開入力が可能なこと、表示項目と対応付けても曖昧性がある場合に質問を発する機能があることなどが特徴である。

今後、表示情報および音声対話上の文脈を有効活用することにより、認識率の向上および柔軟な対話を実現する方法について検討を進める。