

音声対話型コミュニケーション放送システムの開発

Development of Interactive Multimodal Broadcast System

鯨井 俊宏^{*1}
Toshihiro Kujirai

米澤 恵^{*2}
Kei Yonezawa

平野 義明^{*3}
Yoshiaki Hirano

(株)日立製作所(*¹ 中央研究所 *² ビジネスソリューション事業部 *³ 流通システム事業部)
Hitachi, Ltd. (Central Research Labs. / Business Solution Systems Division / Distribution Systems Division)

Terrestrial digital broadcasting systems enable us communicate interactively with a remote controller using BML (Broadcast Markup Language). We developed an architecture which enable us communicate interactively by speech dialog to realize the information barrier-free environment especially for people with disabilities. We also demonstrated effectiveness of our architecture in practice.

1. はじめに

デジタル地上波放送システムでは、放送用コンテンツ記述言語 BML(Broadcast Markup Language)[1]を用いて、リモコンによる双方向型のコミュニケーションが可能となっている。我々は、上記のシステムにおいて、情報弱者、特に視覚障害者や高齢者の満足度や使い勝手を向上させるために、音声対話による双方向型のコミュニケーションを実現するためのアーキテクチャを開発し、実証実験により提案方式の有効性を確かめた。

2. BMLによる双方向コミュニケーション放送

BMLはXHTMLをベースとしてARIB(Association of Radio Industries and Businesses)によって標準化された放送用コンテンツ記述言語である。デジタル地上波放送では、映像データとともにBMLデータを送信することによって、受信機に文字情報や画像などを表示し、ユーザがリモコンによってメニューを選択することで双方向のコミュニケーションを実現している。

図1はBMLを利用した放送の一例である。左上に表示される映像に加えて、L字型にBMLで記述されたコンテンツが表示されている。

3. VoiceXML

VoiceXMLは音声対話のシーケンスを記述するための言語であり、VoiceXML Forum[2]がバージョン1.0をW3C(World Wide Web Consortium)に提案し、現在バージョン2.0の標準化がW3Cで進められている[3]。VoiceXMLの用途は主に電話を入出力デバイスとした、インターネット情報へのアクセスである。VoiceXMLでは、ユーザに提示するガイダンス、ユーザから受け付けるコマンド、対話のシーケンスなどを、タグ付けされたテキストで記述する。



図1 BMLを用いた放送の一例(©(株)テレビマンユニオン)

連絡先: 鯨井 俊宏, 日立製作所中央研究所, 東京都国分寺市
東恋ヶ窪 1-280, 042-323-1111, kujira@crl.hitachi.co.jp

4. 音声対話型コミュニケーション放送

BMLの利用により、リモコンによる双方向型のコミュニケーションが可能になり、今後双方向型コミュニケーションの特長を活かしたサービスが開始されることが期待されている。

テレビ放送は公共性の高いメディアであり、全ての人がそこから等しく利益を得ることができるのが理想的である。しかし、受信システムの多機能化によるリモコンボタンの増加や操作の複雑化で、これらのサービスを利用できない情報弱者が生じることや、GUIを採用したことによって視覚障害者がこれらのサービスを利用できないといった問題が挙げられている。この問題を解決するために、我々は、データ放送で、BMLに加えてVoiceXMLによる音声対話情報を送信することにより、音声対話によって双方向型のコミュニケーションを実現するシステムを開発した。

図2に、本システムのご概念図を示す。送出センターでは、テレビ放送用の映像・音声に、BMLで記述されたデータ放送コンテンツと、VoiceXMLで記述された音声対話コンテンツを加えて圧縮多重化し、送出する。受信側では、受信機が映像・音声とデータ放送のコンテンツを分離し、PC上で開発した疑似STB(Set Top Box)でBMLとVoiceXMLの処理を行う。BMLブラウザは、BMLで記述されたコンテンツに従った画面を生成し、スキャンコンバータがBMLの画面とテレビ放送用の映像を合成してディスプレイに表示する。ユーザはリモコンによって、画面上に表示されたメニューを選択するなどの操作を行うことができる。一方、VoiceXMLブラウザは、VoiceXMLコンテンツに従ってユーザとの音声対話を行う。

5. BMLとVoiceXMLの連携

マルチモーダルインタフェースを備えたコンテンツを記述するための言語としては、SALT[4]やXHTML+Voice Profile[5]、

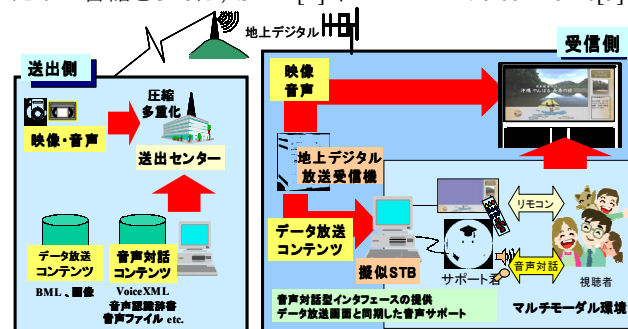


図2 対話型コミュニケーションシステム概念図

XISL[6]などがあるが、BML はすでに地上波デジタル放送のデータ放送用言語として使われることが決まっているため、その利用を前提条件とし、VoiceXML ブラウザと連携することでマルチモーダルインタフェースを実現した。また、従来の機器においても操作を可能とするため、VoiceXML ブラウザとの連携用にBML の記述方式を変更することはせず、ECMAScript[7]を利用した。さらに、BML と VoiceXML の連携を実現するための管理を行うモジュールとして全体制御部を実装した。

5.1 VoiceXML コンテンツの実行と同期制御

BML ブラウザと VoiceXML ブラウザの連携を実現するために、次の機能を実装した。

(1) BML ブラウザへの機能追加

- 全体制御部へのコマンド送信機能
- 全体制御部からコマンドを受信した場合に BML のイベントを発生させる機能

(2) VoiceXML ブラウザへの機能追加

- 全体制御部からのコマンドにより、コンテンツの実行・停止・再開を行う機能
- 音声認識完了時に全体制御部へ認識結果を送信し、全体制御部から戻ってきた値を真の認識結果とする機能

(3) 全体制御部の機能

- コマンドを中継する機能
- 各ブラウザから認識結果が送信された場合に、真の認識結果を決定し両方のブラウザへ送信する機能

図3を用いて BML ブラウザと VoiceXML ブラウザの連携方式について説明する。BML コンテンツが読み込まれた時点で、BML コンテンツは、ECMAScript を用いて VoiceXML コンテンツを実行するように全体制御部にコマンドを送信する。全体制御部は、このコマンドを VoiceXML ブラウザに中継し、VoiceXML コンテンツが実行される。ガイダンスに対して、ユーザが発声を行い認識結果が得られた場合、VoiceXML ブラウザは全体制御部に認識結果を送信する。逆に BML ブラウザ側で、リモコンによりメニューが選択された場合にも、同様に認識結果に相当するコマンドを BML ブラウザが全体制御部に送信する。全体制御部では、これらの認識結果を両方のブラウザに送信し、各ブラウザは全体制御部から送信された認識結果に基づいて、次の処理に移る。この時、VoiceXML ブラウザが実行する処理は VoiceXML の仕様で定められた手順に従う。一方、BML ブラウザ側では、認識結果を受け取った場合にイベントが発生するので、イベントに対するハンドラを ECMAScript で記述し、次のドキュメントに遷移するなど認識結果に応じた処理を記述しておく必要がある。

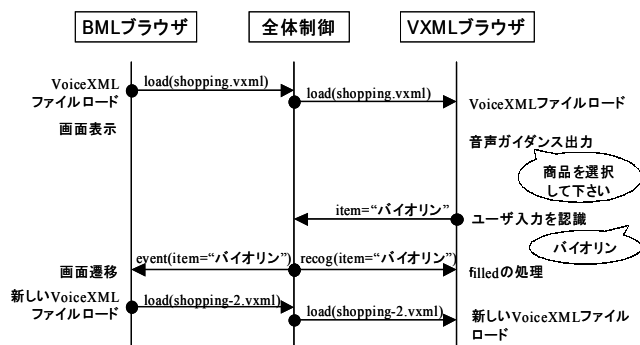


図3 BML と VoiceXML の連携

一定時間内に両方のブラウザから認識結果が全体制御部に対して送信された場合、全体制御部は次の処理を行う。

- 認識結果が同じ変数に対する値の場合、後から送信された認識結果は破棄する
- 認識結果が異なる変数に対する値の場合、それぞれの認識結果を各ブラウザに送信する

上記の処理により、ほとんど同時に矛盾した入力が入力された場合は、先に入力された情報だけを処理し、音声で商品名、リモコンで個数のように補完的な入力が行われた場合は、両方の入力情報を受理することができる。

5.2 外部サーバとの連携

本システムにおいては VoiceXML 側でのウェブサーバへの送信は行わず、BML の外部サーバ連携機能を利用した。この時、外部サーバから BML ブラウザが動的に生成された情報を取得した場合、VoiceXML 側でもその情報が必要になる場合が考えられる。BML 側で動的に VoiceXML コンテンツを生成する方式も考えられるが、本システムでは音声認識によって代入されることのない変数を持った VoiceXML コンテンツをあらかじめ用意しておき、BML 側から動的な情報を認識結果として送信することで対応した。

6. 実証実験

実証実験では、実際の送信設備を用いて BML/VoiceXML を含んだ番組を放送し、プロトタイプシステムで受信して行った。実験用に旅行番組とテレビショッピング番組の2種類を制作し、それぞれ番組の最後に、リモコン及び音声で操作可能なセッションを用意した。旅行番組では視聴者参加型クイズへの回答、テレビショッピングでは番組で紹介された商品の購買がタスクである。視覚障害を持つ被験者9名に各番組を試聴してもらい、音声操作によって上記タスクを実行してもらった後、本システムの有効性に関する評価を行った。評価は主観的な満足度や使い勝手を問うものである。その結果、中途失明の被験者では全員リモコンに比べて安心、便利、簡単であるという回答が得られた。一方、先天的な視覚障害の被験者でPC操作熟練度の高い被験者からは、慣れればリモコン操作の方が速いという回答もあった。

7. まとめ

デジタル地上波放送において、音声対話による双方向型コミュニケーションを実現するためのアーキテクチャを開発し、実証実験により特に中途失明で機械操作に慣れていない被験者に対する有効性が確かめられた。

*この研究は、通信・放送機構の行う通信・放送研究成果展開事業に基づく委託研究として(株)テレビマンユニオンと共同で行われた。

参考文献

- [1]電波産業会, デジタル放送におけるデータ放送符号化方式と伝送方式, ARIB STD-B24, 2000
- [2]VoiceXML Forum : <http://www.voicexml.org/>
- [3]Voice Browser Working Group : <http://www.w3.org/Voice/>
- [4]SALT Forum : <http://www.saltforum.org/>
- [5]X+V : <http://www-3.ibm.com/software/pervasive/multimodal/>
- [6]桂田 浩一他: “音声対話記述言語 VoiceXML と MMI 記述言語 XISL の比較, 情報処理学会研究報告, 2001-SLP-38, pp.49-54, 2001.
- [7]ECMA : <http://www.ecma-international.org/>