

# VoiceXML をベースにした音声対話管理方式の開発

○ 畑岡信夫、渡邊純一郎(日立中研) 赤堀一郎、△立石雅彦(デンソー基礎研)  
△ Eric Nyberg、△Teruko Mitamura、Yasunari Obuchi (CMU)

## 1 はじめに

音声対話は、気の利いた、人間的な HMI (Human Machine Interface)を構築するために、必須技術である。音声対話は、単なる音声認識・合成技術だけではなく、言語処理との融合の結果、可能となる。一方、音声対話システムを実現する記述言語として、VoiceXML[1]等の標準化提案がされている。

本稿で提案する音声対話システムは、VoiceXML を対話シーケンス記述言語とし、①文法、辞書の動的な生成、②言語処理の有機的な融合、の2つの方式的な課題への対応を目指す。具体的なサービスとしては、音声対話を HMI とした車載情報サービスを取り上げ、対話内容(タスク) 遷移にもスムーズに対応できる対話管理方式の実現を狙った。今回は、3 階層構造の対話管理方式を提案する[2]。

## 2 VoiceXML での対話シーケンス実現の課題

スムーズな対話管理を実現するためには、対話シーケンスの表現能力と共に、タスク遷移(例えば、ナビゲーション・タスクやルート計画のタスク等)に追従する柔軟な対話管理が必須となっている。従来の対話システムは、階層的なメニューや状態遷移グラフが用いられていた。また、2, 3 の対話タスクに対処できるシステムも提案されている[3,4]が、対話中でのタスク遷移への柔軟な対応は実現できていない。

W3C(World Wide Web)標準の VoiceXML は、実際の対話シーケンスを自由に組み込める、いわゆる form-filling 型の記述言語である。しかし、単純な VoiceXML では、タスク遷移への柔軟な対応はできないという問題がある。そこで、我々は、push&jump 型の構成を取り入れた拡張型 VoiceXML による対話管理方式を開発している。ScenarioXML と DialogXML の2階層を組み込んだ

### Dynamic Speech Dialog Management based on VoiceXML

N.Hataoka, J.Watanabe (Hitachi Cent. Res. Lab.)  
I.Akahori, M.Tateishi (Denso Res. Labs)  
E.Nyberg, T.Mitamura, Y.Obuchi (Carnegie Mellon U.)

だ、いわゆる 3 階層の対話管理アーキテクチャである(後述図2)。この結果、対話タスクが遷移した場合も、遷移情報により、現状態のタスクを一時保持しながら、次のタスクへと遷移し、その後、元のタスクへ戻る事が可能となる。さらに、言語処理を有機的に融合して、遷移先タスクの文法と語彙辞書を自動的に生成して、VoiceXML による対話シーケンス処理が可能となる[5,6]。

## 3 全体システム構成と3階層による対話管理

応用としては、車載情報サービスを取り上げ、車載端末とネットワークにて結合されたセンターとのやり取りにより、インターネット上の各種サービスを、車内でも享受できることを狙う。いわゆる、音声 HMI によるテレマティクスの実現である。VoiceXML は、ネットワーク型での端末での処理を考慮した記述言語ではないが、組み込み型応用でも対話シーケンスによるシステム実現を可能としている。

図1は、VoiceXML をベースとした音声対話システムの全体構成を示している。将来的には自然言語処理(例えば、KANTOO[7])による対話応答の意味処理も含めた対話管理を目指すのが、当面は、事前に対話タスクを解釈して、オフラ

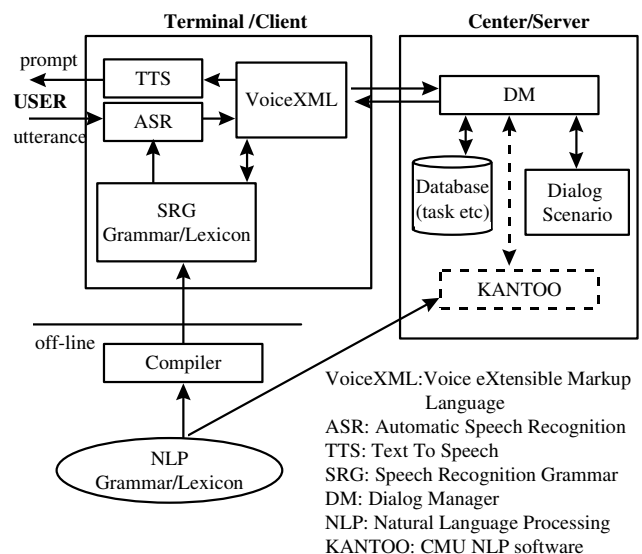
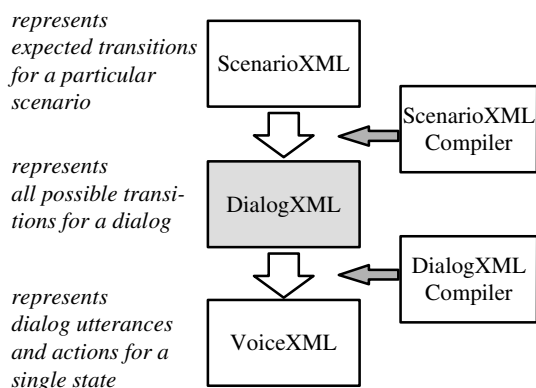


図1 全体システム構成 (ASR/DM/NLP 連携)

インで得られたタスク文法(Task Grammar)と語彙辞書(Task Lexicon)を基に、対話実施フェーズにおいて、VoiceXML が自動生成され、音声対話シーケンス処理を実行することを狙う。

対話管理(DM)は、図 2 に示すように、3 階層構成となっている。ScenarioXML は、対話タスクの管理、DialogXML は具体的な対話内容の管理を行い、システムは VoiceXML にて記述された対話シーケンスを実行する。自然言語処理を十分に活用するためには、文法(grammar)を自由に扱える必要がある。すなわち、状態(state)と状態遷移(state-transition)が自由に表現できることである。この点を考慮して、図 2 に示すように、DialogXML では、対話内容を arc 要素で表現し、その中で grammar の規定と push による遷移が記述できる。



```
<DialogXML>
.....
<arc>
  <grammar src="directions_request.pat"
            type="kantoo/patrick"/>
  <push dialog="directions-request.jsp"/>
</arc>
<arc>
  <grammar src="parking_request.pat"
            type="kantoo/patrick"/>
.....
```

図 2 3 階層による対話管理と DialogXML の例

## 4 音声対話タスクと対話コーパスの評価

### 4.1 音声対話タスク

具体的には、車載情報サービスにおけるルートガイダンス(道案内)、観光案内、施設案内等のタスクである。本提案の 3 階層による対話管理方式により、図 3 に示すような道案内タスク中に、駐車場に関する問い合わせがあった場合でも、柔軟(スムーズ)な対応が可能となる。現在、英語テキストを入力としたプロトを開発し、機能評価を行うとともに、日本語音声入力

版への拡張を行っている。

### 4.2 対話コーパスの評価

最終的には、被験者 250 名にて、上記対話タスクで一人 4 回、計 1000 対話を収集し、対話コーパスの解析を行う。現在、約 250 名計 523 対話に関して評価した。システム(オペレータ)34,745 発話、ユーザ 33,888 発話があり、ユーザ発話解析では、異なり文章 14,630 個、100 回以上出現する文章として、「はい」12,450 回、「え～」749 回、「あ、そうですか」406 回であった。また、異なり語 5,050 語のうち、意味的に完全に異なっているのは 4,387 語であった。

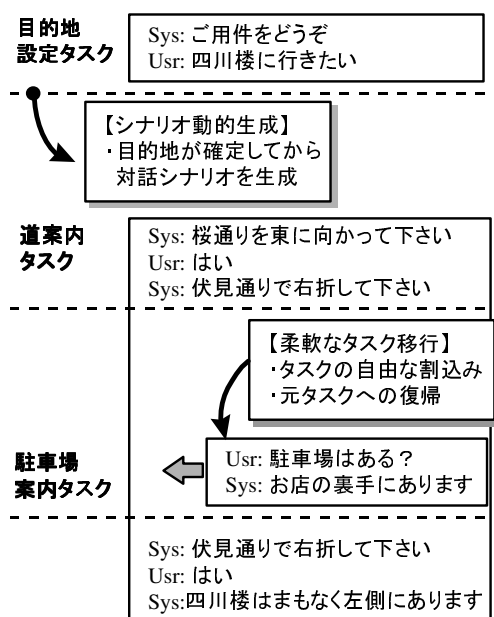


図 3 対話シーケンスの例と特長機能

### 参考文献

- [1] VoiceXML2.0, 04/24/02, <http://www.w3.org>
- [2] E.Nyberg, T.Mitamura, P.Placeway, M.Duggam and N.Hataoka, DialogXML: Extending Voice-XML for Dynamic Dialog Management, Proc. of HLT-2002 (2002)
- [3] P.Price, Evaluation of Spoken Language Systems: The ATIS domain, Proc. of 3rd DARPA Workshop on Speech and Natural Language (1990)
- [4] S.Seneff, et al., Organization, Communication, and Control in the Galaxy-II Conversational System, Proc. of Eurospeech99 (1999)
- [5] M.Araki, et al., Dialogue scenario generation from XML-based database, Proc. of 1st NLP and XML Workshop (2001)
- [6] 安達史博、他、VoiceXML の自動生成に基づく協調的な電話自動応答システム、音響学会論文集 1-5-18 (2002 年 3 月)
- [7] E.Nyberg, T.Mitamura, The KANTOO Machine Translation Environment, Proc of AMTA (2000)