

マルチモーダル対話システムにおける 動的環境情報を利用した応答文のスタイリングの検討*

吉田 昌弘[†] 桂田 浩一[†] 入部 百合絵[‡] 新田 恒雄[†]

[†] 豊橋技術科学大学 大学院工学研究科 知識情報工学専攻

[‡] 豊橋技術科学大学 情報メディア基盤センター

1. はじめに

近年、携帯電話や PDA・タブレット PC 等ユーザ端末の多様化に伴い、Web アプリケーションのマルチモーダル対話 (MMI) への応用が活発に議論されている。MMI では様々な入出力手段が用いられるが、その中でも、音声は数多くのシステムで利用されており、中核技術の一つといえる。従来の音声対話システムでは、例えば VoiceXML [1] ベースのシステムのように、予め固定した対話シナリオを用意し、それに従った対話進行を行うのが一般的である。しかしこのようなシステムでは、どのようなユーザがどのような環境で発話しても、同じ応答を返すことしかできないため、対話が単調になってしまうという問題が生じる。本研究では、ユーザや環境に適応して応答文をスタイリングすることで、より自然な対話を実現する MMI システムの開発を目指す。

以上に述べたシステムを開発する上で、端末情報や、環境情報、ユーザプロフィール等の情報（これらを動的環境情報と呼ぶ）の利用は必要不可欠である。Web 技術の標準化団体である W3C では、動的環境情報を web ベースシステムで利用するための枠組みとして、Delivery Context Interfaces (DCI) [2] の検討を行っている。これまでに我々は、動的環境情報を利用するため、MMI システム[3]に DCI を参考に設計した動的環境情報管理部を導入してきた[4]。本稿では、動的環境情報を用いた応答文の適応的スタイリング法、および評価について報告する。

2. MMI システム

2.1. システム構成

我々がこれまで開発してきた MMI システムの構成を図 1 に示す。MMI システムは、入出力を管理するフロントエンド部、複数の入力モダリティを統合する入力統合部、ユーザへの出力を生成する出力生成部、シナリオに沿って対話を制御する対話制御部、動的環境情報を管理・利用するための動的環境情報管理部からなる。

入力統合部・出力生成部は、それぞれに用意された統合ルール・生成ルールを用いて統合・生成を行う。また、対話シナリオはこれまで我々が検討してきた XISL2.0 (eXtensible Interaction Scenario Language) [5] によって記述する。XISL は対話制御部の要求に従ってドキュメントサーバからダウンロードされる。

動的環境情報管理部では、フロントエンド部から取得する端末情報、環境情報、ユーザプロフィールデータベースからダウンロードするユーザプロフィール等をアプリケーション毎に管理する。

2.2. 動的環境情報

動的環境情報は図 2 に示すように階層的に管理され、ルートから端末情報、ユーザプロフィールという様にツリーが分化している。ツリーを構成する各ノードをプロパティと呼び、子プロパティはその機能や性質ごとに分類されている。例えば、ユーザプロフィール (User) の子プロパティは情動 (emotion) と性別 (sex)、年齢 (age) といったように構成されている。

動的環境情報管理部では、例えばマイクが ON から OFF に切り替わるというような動的な変化が起きた場合、フロントエンド部から送信された情報と、ツリー内のプロパティを比較し

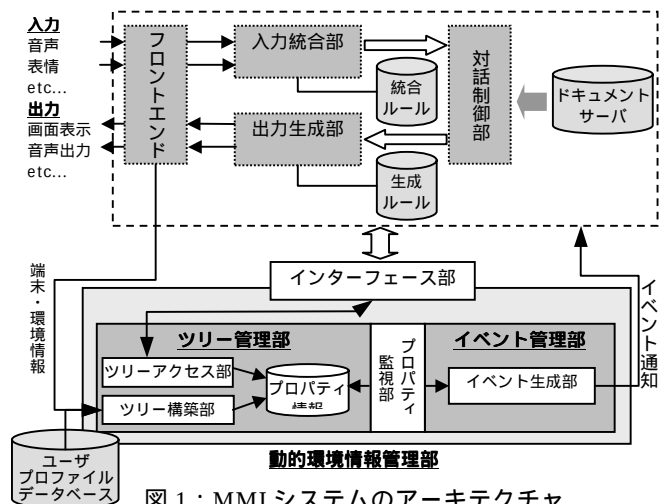


図 1: MMI システムのアーキテクチャ

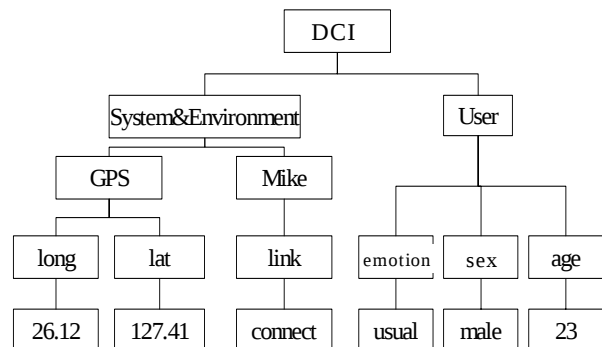


図 2: 動的環境情報の木構造の例

て値を更新する。この時、同時にプロパティの変化を他のモジュールに通知するためにイベントを発生させる。

他のモジュールから動的環境情報を参照する場合は、DOM (Document Object Model) を用いてプロパティツリーのノードを辿る方法の他に、プログラミング言語の関数のように、予め定義した予約語を指定することによって情報を取得する方法も提供している。前者の場合はツリーの構成を熟知している必要があるが、後者ではツリーの構成を把握する必要がないため情報の参照は比較的容易となる。

3. 応答文のスタイリング

これまで多くの対話システムでは、使用環境やユーザに関わらず、同じ応答を返すことしかできなかった。しかし、様々な使用環境やユーザが想定されるシステムでは、それぞれの使用状況に適応した応答を返すことが望ましい。また、MMI では、場面による出力メディアの切り替えや、メディアごとに表現方法を切り替えることが理想である。本研究では、そのための前段階として、音声出力において、動的環境情報を利用し、使用状況に適応した応答文のスタイリングを行うことを目指す。

ここでスタイリングとは、例えばシステム利用環境の雑音が大きい場合、出力音声の音量を上げる、ユーザが子供であった場合は難しい単語を利用することは避けるといった処理を行う

*On output styling based on dynamic-environment-information in MMI systems.

A.Yoshida[†], K.Katsurada[†], Y.Irube[‡], T.Nitta[†]

[†] Graduate School of Engineering, Toyohashi Univ. of Tech.

[‡] Information and Media Center, Toyohashi Univ. of Tech

ことを指す。以下、具体的手法について述べる。

3.1. スタイリングの種類

スタイリングの具体的内容としては、話者の変更、言い回しの丁寧さ、方言への言い換えなど多様なものが考えられるが、これを全て実装するのは困難である。そこで、今回は動的環境情報として、ユーザの年齢と、雑音状況のみを使用し、音声出力に対するスタイリングのみを検討した。ユーザ年齢による語彙、理解力、聴覚機能の違い、雑音状況による発話音声の聞き取り難さに対応するため、以下の各項目をコントロールするスタイリングを実装した。

- 1) 発話音量の変更
- 2) 発話速度の変更
- 3) 使用単語の難易度の変更

発話音量の変更は、予め音量の数値を3段階に定め、スタイリングを適用する際に動的環境情報によってそのいずれかを指定することとした。

発話速度の変更も、音量と同様に3段階の数値を指定する。この際、単純に発話レートを下げるのでは、発話が間延びするだけであって聞き取り難くなってしまうという問題が発生する。そのため、発話速度を遅くする場合は、レートを下げるだけでなく、文節間に無音を挟むという処理を行うことで実現した。

使用単語の難易度は、小学校から中学校の国語教科書出現語彙を難易度の判断基準としている。具体的には小学校6年生以下、中学校3年生以下、それ以外の3段階に分類している。応答文中に指定された難易度以上の単語が存在した場合、指定の難易度以下の同義の単語と置き換える。同義語は言語学研究所「デジタル類語辞典」を参考に、前述の難易度とともにまとめた辞書を自作した。また、固有名詞のように置き換えが望ましくない単語は入力文中においてダブルクォート(“)で囲むことで表現し、置き換えの対象から外すことにした。

3.2. スタイルの指定

より柔軟なスタイリングを行うには、システムが自動的にスタイリングを行うだけでなく、アプリケーション開発者がどのようなユーザ、環境の場合にどのようなスタイリングを行うかということ指定できることが必要である。そこで、スタイリングの条件、パターンを記述するためのフォーマットを導入した。図3に記述例を示す。

スタイリングの内容は<style>タグに記述する。type属性に

<pre><styles> <style type="volume"> <case cond="SE.user.age > 60">2</case> : <case></case> </style> <style type="pattern" id="xxx"> <volume>1~3</volume> <rate>1~3</rate> <difficulty>1~3</difficulty> <target> <users> <user>user1</user> : <user>userx</user> </users> <conditions> <case cond="SE.user.age <= xx" /> : <case /> </conditions> </target> </style> </styles></pre>		
		各スタイリング に対する条件
		スタイリング パターン
		パターンを 適用するユーザ
		パターンを 適用する 動的環境情報

図 3：スタイルの記述例

表 1：評価結果

条件	評価
大人	4.44
子供	4.38
老人	4.30

入力文		京都のホテルを検索します
想定 ユーザ	子供	<VOLUME LEVEL="1.5">京都のホテルを 探します。</VOLUME>
	老人	<VOLUME LEVEL="1.5"><RATE SPEED="1.5"> 京 都 の<SILENCE MSEC="5" />ホ テ ル を <SILENCE MSEC="5" />探 します。</RATE> </VOLUME>

図 4：スタイリング例

は"volume (音量)"、"rate (速度)"、"difficulty (難易度)"、"pattern (パターン)"のいずれかを記述する。それぞれ記述されたスタイリングを適用する条件とその数値を<style>タグ以下に記述する。"pattern"が指定された場合、スタイリングのパターンを記述し、<target>タグ以下にそのスタイリングパターンを適用するユーザ(<users>タグ以下)や、動的環境情報(<conditions>タグ以下)を記述する。

4. 評価実験

本システムの有用性を示すために評価実験を行った。以下に実験概要と結果を示す。

4.1. 実験概要

実験では、被験者にシステムを利用してもらい、スタイリングの有用性についてアンケートを取った。対話タスクは旅行プラン作成タスクで、対話数は 30 ターンである。被験者は研究室の学生 6 人である。各被験者には 3 種類のスタイリング(ユーザが大人、小学校高学年程度の子供、70 歳程度の老人であると想定したもの)を示して評価させた。スタイリング結果が想定したユーザに対して適切であるかどうかを、「適切」を 5、「不適切」を 1 とする 5 段階で評価させた。出力例を図 4 に示す。システムの音声合成には GalateaGtalk[6]を使用しており、出力文中に見られる<VOLUME>タグ、<RATE>タグ、<SILENCE>タグは、それぞれ音量、速度、無音を指定するものである。

4.2. 結果

結果を表 1 に示す。数値はそれぞれの条件に対する評価の平均である。

結果より、概ね適切にスタイリングされていると言える。ただし、いくつかの応答文に関しては不適切と評価されたものがあり、変換語彙の不足によるもの、単語選択が適切でなかったもの、発話速度を遅くしたためにかえって聞き取りにくくなったものなどが指摘された。

これらを解決するために、単語辞書の充実、アルゴリズムの改善、各条件に対する適切な発話速度・音量の再検討が必要である。

5. おわりに

より自然な対話システムの実現のために、応答文のスタイリングの検討を行い、その有用性を示した。

今後は、上記の問題の解決に取り組むとともに、さらに様々な条件下におけるスタイリング手法の検討と実装を進めていきたい。

参考文献

- [1] <http://www.w3.org/TR/voicexml21/>
- [2] <http://www.w3.org/TR/DPF/>
- [3] 桂田浩一，他：“多様な端末からのアクセスを可能にする MMI アーキテクチャ”，情処研報 2002-SLP-40，pp.51-56(2002)
- [4] 桂田浩一，他：“静的・動的情報を利用した MMI システムの設計と実装”，情処研報 2006-SLP-62，pp.87-92(2006-7)
- [5] 青木一峰，他：“異なる端末環境から利用可能な MMI アプリケーション開発における記述負担の軽減”，信学技報，SP2004-126，pp.25-30，(2004-12)。
- [6] <http://hil.t.u-tokyo.ac.jp/~galatea/>