Vol. 41 No. 4 情報処理学会論文誌 Apr. 2000

キャラクタエージェント制御機能を有するマルチモーダル・プレゼンテーション記述言語ＭＰＭＬ

筒井貴之[†,☆] 石塚満[†]

WWW上での効果的な情報提供形態として，また新しい情報コンテンツとして，キャラクタエージェントとの音声対話を含むインタラクティブなマルチモーダル・プレゼンテーションは注目に値する対象である．我々はそれを誰でも簡単に作成，流通，視聴することを可能にするために，マルチモーダル・プレゼンテーション記述言語MPML (Multimodal Presentation Markup Language)を開発した．MPMLはXML規格に準拠したマークアップ言語であり，特定のシステムに依存せず，エージェントの動作命令などを容易に記述できるほか，メディア同期，音声指示等によるプレゼンテーション制御機能をサポートしている．本論文では，MPMLの紹介を行うとともに，MPMLを用いたプレゼンテーションを行うために開発したツールに関して報告する．

A Multimodal Presentation Markup Language MPML with Controlling Functions of Character Agent

Takayuki Tsutsui[†,☆] and Mitsuru Ishizuka[†]

As a new style of effective information presentations and a new multimodal information-content production on the WWW, multimodal presentations using interactive lifelike agents with verbal conversation capability appear to be very attractive and important. For this purpose, we have developed MPML (Multimodal Presentation Markup Language) which allows many users to write attractive multimodal presentations easily. MPML is a markup language conformed to XML (Extensible Markup Language) and supports functions for controlling verbal presentations and scripting agent behaviors. In this paper, we present MPML and its related tools for producing multimodal presentations.

1. はじめに

ここ数年の間にWWW (World Wide Web)は急速に発展し，今や社会の重要な情報発信・流通のインフラストラクチャになりつつある．WWWは誰でも容易に情報発信できることが大きな特色であり，非均質だが膨大な情報空間を形成するようになってきている．しかしWWW上で提供される情報が多様化，複雑化，大規模化するのに伴い，情報洪水といった状況も呈しつつある．このような中で，情報利用者は「より効率的な情報へのアクセス」，「情報の無駄のない選択」，「情報の効果的な活用」を求めるようになっており，そのための技術開発も数多く行われている．情報コンテンツの質は内容と表現形式の両面で重要であり，魅力のない情報は淘汰されてしまう傾向にある．情報提供者にとって，効果的かつアトラクティブな情報」を「誰にでも，分かりやすく視聴できる形で提供する」ことが重要な課題になってきている．

我々はWWW上での新しい情報表現形態として，キャラクタエージェントを用いるマルチモーダル・プレゼンテーションは魅力があり，重要になると考えている．現在のプレゼンテーション・ツール (Power Pointなど)は説明用画面の作成・表示機能を提供し，人間のプレゼンタがこれを操作し，音声と動作でプレゼンテーションを行う形式である．これはマルチモーダルな情報チャネルの使用，そしてインタラクティブ性により，単に文書，本などによる情報提供に比べて情報伝達量が多く，我々人間の理解に適していることから，現代の代表的プレゼンテーション形式になってきている．本論文でのキャラクタエージェントを用いるマルチモーダル・プレゼンテーションは，人間のプレゼンタの代わりをキャラクタエージェントに果たさせ，人間のプレゼンタがいなくても同様な効果を達成しようとするものであり，同時にWWWの新しいマ

† 東京大学工学系研究科電子情報工学専攻
School of Engineering, University of Tokyo
☆ 現在，株式会社ＮＴＴデータ
Presently with NTT DATA Corp.

ルチモーダルな情報コンテンツを可能にしようとするものである．これによって，時間，空間を越えて享受できる，我々人間にとって親しみやすく，分りやすいマルチモーダル・情報コンテンツの流通を可能にしようとするものである．

キャラクタエージェント・システムの登場や音声合成/認識システムの性能向上により，このようなプレゼンテーションの実現は可能になってきている．しかし，そのコンテンツの作成には個々のシステム毎に固有の方法で（固有のコマンド系列，スクリプト言語などで多くの場合は細部に至るまで）記述しなければならず，容易ではない段階である．これを広く普及するには，HTML レベルの共通的な記述言語を定めることにより，誰でもがマルチモーダル・コンテンツの記述，作成を容易にできるようにすることが重要となる．

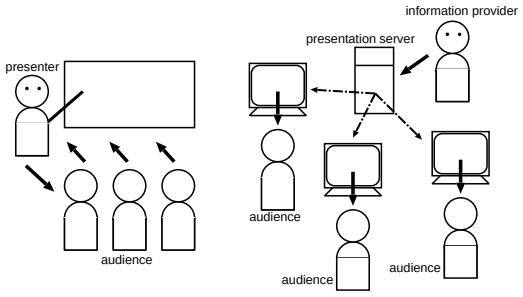
本論文ではこのような目的で，キャラクタエージェントを用いるマルチモーダル・プレゼンテーションの共通的記述言語として開発した MPML（Multimodal Presentation Markup Language）について記す．MPML は XML 規格に準拠したマークアップ言語であり，特定のキャラクタシステムに依存せず，エージェントの動作命令などを容易に記述できるほか，SMIL に準拠したメディア同期機能，音声指示等によるプレゼンテーション制御機能をサポートしている．MPML は各キャラクタエージェント・システムに対応するコンバータを介して，各々のシステムでマルチモーダルなプレゼンテーションを実行する．

2. プレゼンテーションエージェント

2.1 現在の WWW 上のマルチメディア・プレゼンテーション

昨今の WWW ページには，テキストやグラフィクスだけでなく，アニメーションや音楽，さらには音声まで多様に組み合わせたマルチメディア・プレゼンテーションの形態をとっているものが登場するようになってきた．情報提供者（presenter）はコンテンツを作成してインターネット上に公開しておくことで，そのプレゼンテーションを誰にでも視聴可能にすることができる（図 1 (b)）．

それらは魅力的なものも多いが，ほとんど一方的な情報伝達しか行われず，ユーザが行うことのできる操作は，マウス操作で，与えられたリンク構造やブラウザのボタンを用いてページをジャンプさせたり，ウィンドウを閉じたりといった「次ページへの移行と購読の停止」を目的としたものにほぼ限られている．また，人間のプレゼンタによるプレゼンテーション（図 1 (a)）



(a) REAL Presentation (b) Multimedia Presentation on the WWW

図 1 現在のプレゼンテーションのスタイル
Fig. 1 Present styles of presentations.

と異なり，質問をするなどして視聴者（audience）の意志を情報提供者へ伝えることはできない．

我々はこうした一方通行のシステムを改良し，人間のプレゼンタによる親しみやすく分かりやすいプレゼンテーションに近い形式を可能として，WWW 上の魅力あるマルチモーダル情報コンテンツを作成するために，キャラクタエージェント・システムを利用する方法に注目した．

2.2 擬人化エージェントシステム

現在主流のインタフェースの形式は GUI であり，ユーザは “See and Point” 型を主体とするインタラクションを行う．これに対し，マルチメディア機能の向上，コンピュータへの指示内容の複雑化に対応して，音声による “Ask and Tell” 型のインタラクションも導入したマルチモーダル・インタフェースの研究開発が進んでいる．この中で，日常生活の “Face-to-Face” 型の親しみやすいコミュニケーション形態を実現するものとして，CG キャラクタの表情や動作等の擬人的表現を使用する擬人化エージェント・インタフェースがある．このようなマルチモーダルの擬人化エージェント・インタフェースとしては，東芝の TOSBURG II[1]，Sony CSL[2] のシステム，我々の Web ブラウザと結合した VSA システム[3)〜5)]，電総研のシステム[6)] などが開発されている．

一方，単なるインタフェースとしてだけでなく，魅力あるマルチモーダル情報コンテンツの作成，流通も含めてシステムやツールを考えることが今後重要な観点になると思われる．キャラクタエージェントを用いるマルチモーダルな情報コンテンツとして，最も期待できる形態の一つが，プレゼンテーションという場であろう．既にペンシルベニア大の Human Presenter[7)] や，ドイツ人工知能研究センター（DFKI）の WWW 連携も行った WebPersona[8)] の研究例があり，プレゼンテーションにキャラクタエージェントを用いる有意

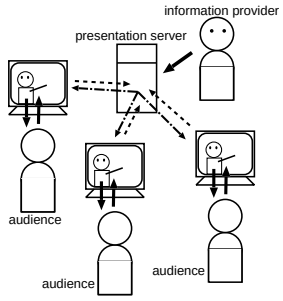


図 2 キャラクタ・エージェントによる WWW 上のプレゼンテーション
Fig. 2 Multimodal presentation by character agent on the WWW.

性は示されている．これらの例ではユーザインタラクションに関してはあまり言及していないが，本研究では多少のインタラクションを考慮した“マルチモーダル・プレゼンテーション”を取り扱う．

マルチモーダル・プレゼンテーションは現在の音声認識技術レベルからも好ましい応用の形態といえる．即ち，音声合成については十分実用的に使えるレベルになっているといえるが，音声認識については進歩してきているものの，認識率の点でまだ必ずしも十分なレベルではない．誤認識は音声インタフェースの不自然さ，煩わしさを増し，評価を悪くしてしまう．シンポジウム等学術発表の場や商品の紹介などのプレゼンテーションでは，音声についてはその大部分が出力であり，入力は数%に過ぎないので，現状の音声技術レベルにも適合した応用形態といえる．

2.3 キャラクタエージェントによる WWW 上のプレゼンテーション

WWW 上での情報提供は情報提供者の都合に関係なく，時，場所を選ばずに行うことができることが大きな利点である．その内容は年々多彩にはなってきているが，人間によるプレゼンテーションと比較すると，分かり易さ，人を引き付ける魅力，インタラクティブ性といった点で劣っているといえよう．そこで前節に記したキャラクタエージェントによるプレゼンテーションを Web 上で可能にすることは，魅力的で広域に流通する今後のマルチモーダル情報コンテンツになるものと考えられる（図 2）．

動作する CG キャラクタは従来以上に視聴者の注意を引き付ける効果をもち，また音声に加えてキャラクタの表情や動作等の付加的情報によって，視聴者の認知的負担を軽減させ，伝達すべき情報内容をより良く伝えることができる．自由対話というまでにはいかないが，視聴者はキャラクタエージェントとの対話を通じてプレゼンテーションに注文をつけたり，(限定はあるものの) 質問をはさんだりすることも可能になる．また，こうした視聴者の反応をフィードバックして情報提供者はプレゼンテーションを改善していくことが可能になる．

このようなキャラクタエージェントを用いたプレゼンテーションを WWW 上で広く利用できるようにするための一つの大きな課題は，個々のキャラクタエージェント・システム等に依存しない，記述しやすいコンテンツ記述言語を普及させることであろう．我々はこのような目的に向けての第一歩として，今後の Web 上の情報コンテンツ記述の標準として位置づけされるマークアップ言語 MPML (Multimodal Presentation Markup Language) の設計，開発を行った．

3. マルチモーダル・プレゼンテーション記述言語 MPML の特徴

MPML はキャラクタエージェントを用いたマルチモーダル・プレゼンテーションを容易に記述できるように設計，開発されたマークアップ言語である[9),10)]．以下にその特徴及び仕様設計の経緯を記す．

システム非依存 ： 情報提供者がプレゼンテーションを配信する際には，通常は OS，ブラウザ，リソース等の視聴者の環境を考慮しなくてはならない．MPML は特定のブラウザやシステムに依存する言語ではなく，MPML で記述されたプレゼンテーション・コンテンツを様々なツールによって再生，視聴することを目的として設計された．したがって，プレゼンテーションの作成者は視聴者の環境を考慮することなくコンテンツを記述，作成することができる．

容易な記述性 ： たとえば VRML 等でキャラクタエージェントを用いたプレゼンテーションコンテンツを作成するのも可能ではあるが，作成には高度な知識とテクニックが必要である．MPML は XML (Extensible Markup Language)[11)] の仕様に準拠しており，誰でも簡単にスクリプト記述を行えるように設計された．タグについては極力簡略化を目指し，現在の MPML バージョン 1.0 では高々19 種類となっている．これにより，HTML スクリプトを書いてホームページを作成できる程の知識がある人なら，簡単にエージェントを用いたプレゼンテーションを作成できる．また熟練者は様々な属性値を変化させることによって，より詳細にわたってプレゼンテーションを設定できるようにもなっている．

メディア同期 ： 魅力的なプレゼンテーションにおいては，音楽，画像，動画等の様々なマルチメディアを

効果的に再生することが必要である．この点に関してW3Cより，WWW上で複数のメディアデータの表示を制御するための記述言語であるSMIL (Syncronized Multimedia Integration Language)[12]の勧告が1998年に公開された．MPMLはこの仕様を参考にしてメディア同期を図るための機能を有している．

キャラクタエージェントの制御 ： プレゼンテーションというほぼ場面を限定した下でのコミュニケーションであるため，エージェントが行う動作は，おおむね人間がプレゼンテーションの場で行う動作と同じで事足りる．MPMLでサポートする動作の設計にあたっては，文献6)〜8)，14)などからプレゼンテーションで用いる動作を抽出し，また代表的な表情，表現などを収集した上で，機能が複雑にならないようそれらを約70種類程度まで取捨選択した．それらの中には「おじぎ」，「指さし」，「説明」などの動作や，「笑顔」，「困惑」などの表情があり，メディア再生と同期をとり，記述することができる．また，たとえばキーワードgreet, bowのどちらを記述しても「おじぎ」動作を実行するなど，仕様に若干の柔軟性を持たせている．

エージェントは画面上もしくはウィンドウ中を自由に移動し，さまざまな動作を交えながら音声やふきだしを用いて視聴者にアピールを行う．また，音声認識機能を用いてユーザの意図を認識することも行う．情報提供者はこれらを簡単なタグで記述することができる．

インタラクティブなプレゼンテーションの制御 ： またMPMLではハイパーリンク機能もサポートしており，システムの音声認識機能と組み合わせることで，エージェントとの音声対話などによるインタラクションによって，視聴者を効果的にナビゲートすることが可能になる．

4. MPMLの言語仕様

4.1 MPMLドキュメント

本章ではMPMLの言語仕様について簡単な解説を行う[10]．各要素に関する詳しい定義が記載されているMPML1.0の仕様書，及びDTD (Document Type Definition)はMPMLホームページ[13] (http://www.miv.t.u-tokyo.ac.jp/mpml.html)において公開されている．

MPMLドキュメント全体を表す構造木は図3のように表すことができる． 全ての要素の祖先要素になるのがタグ要素 <mpml> である．タグ要素 <mpml> はその属性として id を持つ．idは要素を識別するために用いられる識別値であり，ほとんどの要素に記載

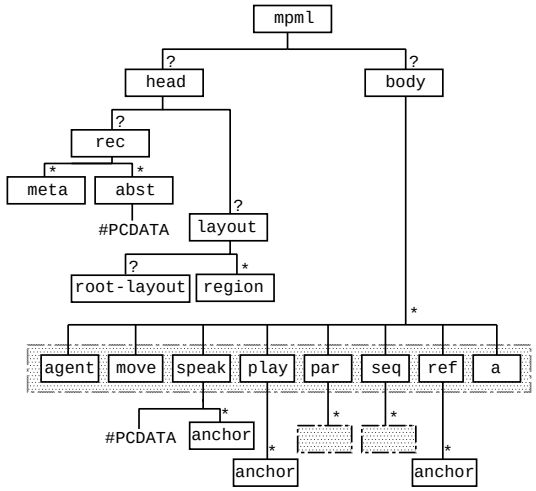


図3 MPMLの構造木
Fig. 3 MPML structure tree.

することができる．また，子要素としては <head> と <body> を持つことができる．以下では各要素に関して順次説明する．

4.2 ドキュメントのヘッダ

<head>...</head>で囲まれたドキュメントのヘッダには，プレゼンテーションに関するメタデータ及び，レイアウトに関する情報を記述することができる．メタデータはタグ要素<rec>に，レイアウトはタグ要素<layout>に記述される．

メタデータ記述

MPML記述者は，タグ要素<rec>内の<meta>と<abst>に，作成したプレゼンテーション全般に関するデータを記載しておくことができる．

タグ要素 <meta> は要素内容を持たない空要素であるが，属性を用いてそのタグ一つに一組のメタデータを記載しておくことができる．

タグ要素<abst>はテキストデータを要素内容に持ち，プレゼンテーションの概要が記載できる．

レイアウト記述

タグ要素<layout>のコンテンツにはプレゼンテーションのレイアウト情報が置かれる．タグ要素内容は任意であり，任意の書式でレイアウト記述ができるが，デフォルトではMPML特有の書式になり，<root-layout>及び<region>を子要素に持つ．

タグ要素<root-layout>では，プレゼンテーションを行うルートウィンドウに関する性質が設定される．

またタグ要素 <region> の属性において「点」，もしくは「長方形型の範囲」としてリージョン情報を記述できる．一つのタグで一つのリージョンを設定できる．

4.3 ドキュメントのボディ

<body>...</body>で囲まれたドキュメントのボディには，プレゼンテーションの実行に関わるコンテンツ群が置かれる．タグ要素<body>は暗黙にタグ要素<seq>の定義を含むため，他の指定がなければその子要素の実行はシーケンシャルに行われる．

エージェント定義

タグ要素<agent>で，プレゼンテーションを行うエージェントキャラクタの設定を行うことができる．<agent>で定義したエージェントをタグ要素<move>，<speak>，<play> などの動作記述で参照する．

複数の<agent>を定義して，それぞれの id 値を逐次参照することで，複数のキャラクタを用いたマルチエージェントによるプレゼンテーションが可能になる．

エージェントの動作記述

タグ要素<move>でエージェントキャラクタを移動させることができる．すでに設定したリージョンへの移動のほか，移動位置を座標指定しての移動も可能である．

タグ要素<speak>のコンテンツにはテキストが置かれる．プレイヤは音声合成エンジンなどを用いて，その内容をエージェントキャラクタに発話させる．

また，タグ要素<play>によって，エージェントキャラクタに様々な動作をさせることができる．MPML では挨拶などの基本動作のほか，指定したリージョンを参照しての指さしをさせたり，発話しながら，まばたきしながら等の「ながら動作」も記述できるようになっている．

以上がエージェント動作記述を行うタグ要素であり，記述できる属性などのもう少し詳しい仕様を表 1 に示す．

メディア同期

タグ要素<par>の子要素として置かれたコンテンツは，そのドキュメント中の順序に関係なく，並列に実行される．例えば以下のスクリプトでは，図 4 (a) に示した動作モデルのように，エージェントは挨拶動作の実行を開始してから 2 秒後に発話を行う．

```
<par>
  <play act="greet" />
  <speak begin="2s">
    hello!
  </speak>
</par>
```

また，タグ要素<seq>の子要素としておかれたコンテンツは，そのドキュメント中の順序にしたがって，シーケンシャルに実行される．例えば以下のスクリプトでは，図 4 (b) に示した動作モデルのように，エージェントは挨拶動作を完了してから 2 秒後に発話を行う．

```
<seq>
  <play act="greet" />
  <speak begin="2s">
    hello!
  </speak>
</seq>
```

また，メディアオブジェクトの再生はタグ要素<ref>を用いてロケーションを指定することによって行われる．

ハイパーリンク/プレゼンテーション制御

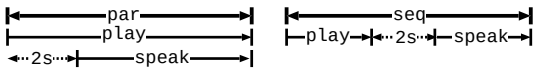
タグ要素<a>を用いて，マウス操作及び音声入力によるプレゼンテーション制御のためのキーを設定する．プレゼンテーションのジャンプ，中断，再開等の基本的な制御を行うことができる．

タグ要素<a>は要素属性として次のものをとる．

id, title, href, show, mode, begin,

表 1 エージェント動作記述のタグ要素
Table 1 Tag elements for agent behavior description.

タグ要素名	要素属性	機能
move	id	識別子
	agent	移動させる <agent> の id 値
	region	移動目標のリージョンの id 値
	location	移動目標の座標指定
	stand	移動目標に対する立ち位置
	speed	移動の早さ
speak	id	識別子
	agent	発話させる <agent> の id 値
	lang	発話させる言語
	voice-type	発話音声のタイプ
	speed	発話のスピード
	begin	発話の開始タイミング
	end	発話の終了タイミング
	dur	発話時間
	alt	音声非対応時の代替メッセージ
play	id	識別子
	agent	動作させる <agent> の id 値
	act	動作内容
	parts	動作させる部位
	object	動作の対象オブジェクトの id 値
	object-loc	動作の対象の座標指定
	degree	動作の強さの程度
	speed	動作のすばやさ
	begin	動作の開始タイミング
	end	動作の終了タイミング
	dur	動作時間
	track	動作対象移動時の追跡の有無
	point-gesture	指示の際の手の形



(a) Parallel Play (a) Sequential Play

図 4 MPML のメディア同期再生のモデル
Fig. 4 Model of synchronized play in MPML.

end, dur, region, listen, lang,
key, confidence

この一部について簡単に説明する．mode はインタラクションが起こった際のプレゼンテーションの挙動を設定するために用い，例えば <a mode="jump" href="link1"> では指定されたリンクにジャンプする．href は HTML と同様にリンクを指定する．show は他リンクへのジャンプが行われた場合などに，プレゼンテーションの挙動を制御するのに用いられる．key は受け付ける音声の指定であり，選択記号"|"，省略記号"[]"，省略記号"..."，解釈優先記号"()"などを使用できる．例えば key="...[say](hello|hi)" の指定では，"say hell"，"please hello"，"Hi"などにマッチする．confidence はマッチする音声の信頼度の下限を指定するのに使用する．listen は音声入力の受け付けを示し，begin, end, dur で音声入力を受け付ける時間を指定することができる．region は指定リージョンでマウスクリックが行われた場合に指定の制御を行うのに使用する．

また，タグ要素 <anchor> は以下の二通りに用いられる．

- メディアオブジェクトの実行やエージェント動作の間の時間的なポイントに，リンクの終点となるアンカーを設定する
- メディアオブジェクトの空間的なポイントに，エージェントの動作対象などになるアンカーを設定する．

アンカーとして指定された識別子を <a> などで指定するジャンプ先の URI の後に付与することで，メディアオブジェクトやエージェントの発話などの途中へジャンプすることも可能になる．例えば下の例では，要素属性 id で示された"anc1"へジャンプして入ってきた場合にはそれ以降の部分のみが発話されるようになる．

```
<speak>
  This part will not be read
  if jump-started from "anc1".
  <anchor id="anc1" />
  Read from here!
</speak>
```

代替コンテンツの提示

タグ要素<switch>には，プレイヤがコンテンツの再生に適さない場合に備えて，代替コンテンツを複数置いておくことができる．コンテンツ作製者は最も視聴してもらいたいコンテンツから順番に指定し，テキスト表示などの最も安全なコンテンツを最後に指定することが望ましい．

4.4 サンプルスクリプト

MPML で記述されたプレゼンテーションのサンプルスクリプトを以下に示す．

```
<mpml>
 <head>
  <layout>
   <root-layout id="root" width="800" height="600" />
   <region id="spot1" location="500, 300" />
   <region id="spot2" location="20%, 50%" />
  </layout>
 </head>
 <body>
  <ref region="root"
   src="http://www.miv.t.u-tokyo.ac.jp/mpml.html" />
  <agent region="spot1" />
  <par>
   <play act="point" region="spot2" />
   <speak>
    This is MPML Home Page!
   </speak>
  </par>
 </body>
</mpml>
```

このスクリプトによって，以下のプレゼンテーションが行われる．

> <ref>で指定された URL のページを背景に，エージェントキャラクタが spot1 として指定された場所に出現する．エージェントは spot2 として指定された位置を指さしながら，<speak>のコンテンツを発話する．

5. MPML によるプレゼンテーション記述，実行のソフトウェア

5.1 MPML 関連ソフトウェアツールの概観

本研究の目的は，マークアップ言語 MPML を利用することによって，誰でも簡単にマルチモーダル・エージェントを用いた WWW 上のプレゼンテーションを作成，流通，視聴することを可能にすることである．情報提供者が MPML を介して発した情報を，視聴者 (audience) が受け取る際の概念モデル図を図 5に示す．この図に示された要素ソフトウェアツールはすべて開発されているという訳ではなく，現在はコンバータが 2 種開発されているのみである．開発中，または計画段階にあるものに関して，詳しくは 7章で説明することとし，ここでは概観を行う．

MPML プレイヤ

作成された MPML スクリプトを読み込み，記述内容通りに直接実行するツールが MPML プレイヤ (MPML Player) である．MPML プレイヤは MPML スクリプトに従って，独自のキャラクタエージェント群を用い，メディア再生を含めたプレゼンテーショ

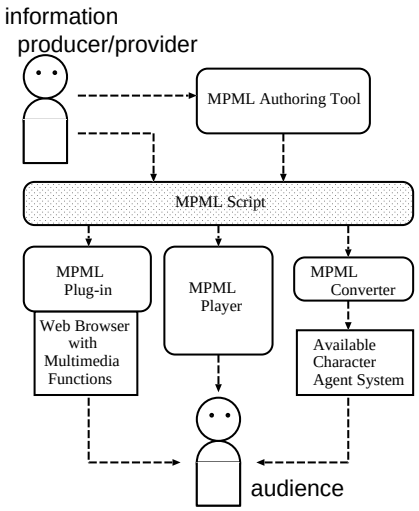


図 5 MPML の関連ソフトウェアツール
Fig. 5 MPML-related software tools.

ンを実行するものである．

コンバータ / プラグイン

MPML プレイヤを使用できない場合に対しても，すでに流通しているエージェントシステムやメディアブラウザを利用して，MPML 記述のプレゼンテーションを視聴してもらうために，たとえば Netscape や InternetExplorer のような既存のブラウザに対してプラグインを適用する方法によって，ブラウザに MPML スクリプトを認識させたり，また別の方法として，MPML スクリプトにコンバータを適用し，各種存在する既存のエージェントシステムに対応したスクリプトを生成する方法がある．

オーサリングツール

GUI を用いたオーサリングツールによって，情報提供者がより簡便に MPML スクリプトを作成することを可能とする．

5.2 現在の開発状況 :2 種のコンバータ

現時点では MPML の効用を実証するために，MPML 記述のプレゼンテーションを実行するツールとして，MPML コンバータが 2 種類開発されている．

一つは Microsoft Agent システム [14] を利用して，Web と連携したプレゼンテーションスクリプトを作成する *MPML2MSAScript* である．これは Perl を用いて MPML で記述されたスクリプトを，MicrosoftAgent を動作させることのできる VBScript のプレゼンテーションスクリプトにコンバートし，HTML ファイル中に埋め込むものである．システムの都合上，MPML の機能の全てはサポートされないが，以下のような基本的な機能を使用することが可能になっている．

図 6 MPML2MSAScript を用いて作成したプレゼンテーション例
Fig. 6 Presentation snapshot generated by MPML2MSAScript.

- 疑似的に同期をとったメディア再生
- 任意の場所へのキャラクタエージェントの移動や動作
- 英語音声の合成と認識
- 音声やマウス操作によるプレゼンテーションの制御
- リンクされた Web ページの閲覧

図 6は，4.4に示したサンプルスクリプトをこの *MPML2MSAScript* コンバータを介して実行したもののプレゼンテーション画面例である．

MPML コンバータのもう一つの例は MPML スクリプトによって，手と指の微妙な動作が可能な CG ツール Ashow[6] を動作させるマスタ / スレーブシステムである．マスタにおいて MPML スクリプトを解析し，ソケット通信を行ってスレーブである Ashow に様々な動作をさせるものである．図 7にその実行例を示す．

MPML で記述可能な動作と，利用するキャラクタエージェント・システムで実現可能な動作との間に不整合が生ずることがあるが，これはコンバータで近い動作にマッピングするようにしている．

これらのコンバータ・ツールを用いることで，複雑なスクリプトを書かなければならないエージェントシステム，メディアプレゼンテーションにおいても，MPML の簡易なタグ記述によって簡単にプレゼンテーションを実現することが可能になった．

6. 考　　察

6.1 他のマークアップ言語との比較

スクリプトに記述できる機能について，MPML と他のマークアップ言語 (SMIL, HTML) について表 2に比較を行った．

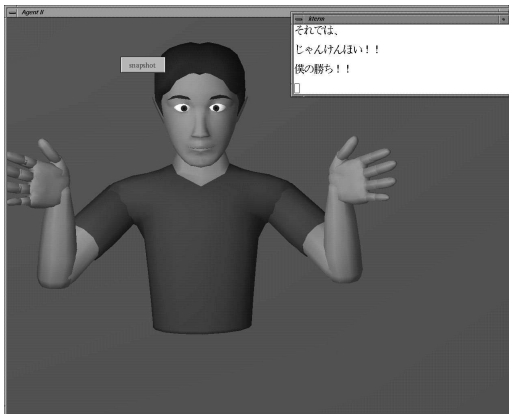


図 7 MPML2Ashow を用いて生成されたプレゼンテーション画面の例
Fig. 7 Presentation snapshotgenerated by MPML2Ashow.

表 2 MPML の他の WWW マークアップ言語との比較
Table 2 Comparison of MPML with other WWW markup languages.

記述できる機能	MPML	SMIL	HTML
Web での配信	可能	可能	可能
他 URI とのリンク	可能	可能	可能
データ記述	標準書式あり	自分で定義	基本は不可
レイアウト記述	可能	可能	基本は不可
メディア同期	必要最小限	十分に対応	不可
エージェント動作記述	可能	不可	不可
マウスによる制御	可能	可能	可能
音声による制御	可能	不可	不可
テキスト音声出力	可能	不可	不可
現在の利用者数	非常に少ない	少ない	非常に多い
ツール	少ない	約 10 種類	非常に多い
タグの種類の数	約 20	約 20	約 80

どの言語も Web での配信を第一に考えたマークアップ言語であるが，その特徴は微妙に異なっている．たとえば，SMIL はメディアを同期させて実行することが主目的であるため，メディアを再生するためのレイアウト記述や，タイミング記述に関しては強固な仕様を備えている．

一方，MPML の言語目標は簡単にキャラクタエージェントを用いたマルチメディア・プレゼンテーションを実現するというものであるため，最低限のメディア同期とレイアウトの記述が可能な仕様になっている．また，キャラクタエージェントとの対話を可能にするため，音声での入力と TTS (Text-To-Speech) に対応した記述が可能になっている．MPML を適用するために開発されたツールは現在前述の 2 種類だけであり，まだ適用範囲が広いとはいえないため，今後の関連ツールの充実が期待される．

6.2 評価と考察

定量的な評価は難しいが，定性的な評価及び考察を記しておく．マークアップ言語 MPML を用いると，他の言語を用いて記述するよりははるかに簡単にキャラクタエージェントの動作を含むマルチモーダル・プレゼンテーション・コンテンツの記述が可能になる．

たとえば Microsoft Agent システムを動作させることのできる VBScript の場合，言語構造を理解していない一般ユーザがスクリプト記述するのは大変に困難である (1997 年に Microsoft Agent が発表されて以来，そのキャラクタエージェント・システムの画期性とは裏腹に一般への浸透度がはかばかしくないのは，プログラミングの困難さが一因ではないだろうか)．

これからのネットワーク上のプレゼンテーションは，新しく他と違った面，つまり魅力ある新奇性が求められる．今後のキャラクタエージェントを用いたプレゼンテーションにおいても，他とは異なる複雑さを求められるようになるであろう．しかしながら，こうしたマルチモーダル・プレゼンテーションのスクリプト記述を既存の言語で行うのは Microsoft Agent の例にもあるように容易ではない．一方，MPML はこの目的専用の言語であるため，簡易な仕様でありながら，必要とされる機能を備えている．すなわち，既存の言語では複雑で難解なスクリプト記述となるプレゼンテーションでも，MPML で記述し *MPML2MSAScript* のようなコンバータツールを使用することで，広く一般のユーザでも比較的容易に作成することができるようになると考えられる．

WWW 上のマルチモーダル・プレゼンテーションの配信においては，プレゼンテーションを再生するリソースやシステムは視聴者に依存するため，情報の送り手は通常それらを考慮して様々な方法を提供する必要が生じる．しかし，MPML を用いてプレゼンテーション・コンテンツを記述しておけば，一種類の方法で効果的なプレゼンテーションの配信が実現できる．

たとえばユーザが学会の発表論文に興味を持った場合，論文資料だけでは伝わりにくいニュアンス等も，発信者 (著者) がプレゼンテーションを MPML で記述しておくことで，ユーザは発表資料のテキストと共に発表時の音声，OHP の画像等が視聴できる．そして，キャラクタエージェントがそれらを用いて発表者の発表に似たプレゼンテーションを行うことで臨場感があり，味わいのある情報伝達が可能となろう．

7. まとめと今後の課題

7.1 ま と め

WWW 上での新しい情報コンテンツとして有望と考える，キャラクタエージェントを用いるマルチモーダル・プレゼンテーションを，誰でもが容易に作成し流通させることを可能にするために開発した記述言語 MPML について記した．MPML は XML 規格に準拠したマークアップ言語であり，エージェントの動作や音声指示などと共に，SMIL に準拠するメディア同期機能もサポートしている．

MPML は現在，2 種のコンバージョン・ツールにより既存のキャラクタエージェント・システムを駆動してマルチモーダル・プレゼンテーションを実行できるようになっている．

7.2 ツールの充実

MPML に課せられた課題は多く，以下のような関連ツールを充実させていかなければならない．

MPML プレイヤ

5章に記した MPML を直接実行する MPML プレイヤが必要とされる．この実現には以下に挙げるような技術が必要となる．

- キャラクタエージェントのマルチ制御
- 日本語 / 英語を含めた音声認識・音声合成エンジンの利用
- 同期をとることのできるメディア再生
- プラットフォームに依存しない動作
- Web と連携した情報授受

このように，必要とされる技術は多く，その実装や統合には多くの努力を要するが，鋭意取り組んでいきたい．

コンバータ / プラグイン

より多くのシステムに対応できるよう，コンバータ，プラグインの充実も必要であり，これらを用いて，できるかぎり MPML スクリプトに忠実なプレゼンテーションを実行するように設計する必要がある．

オーサリングツール

オーサリングツールは，たとえば XML のタグエディタのように，入れ子構造のタグを間違いなく記述でき，タグ挿入などを簡易に行えるようなものを基本とする．応用的なものとして，現在あるマルチメディアオーサリングツールの使用目的をベースとし，ツール上でのマウス操作によってほとんどのコードを作成してくれるようなビジュアルエディタの開発が計画されている．たとえばリージョンを画面上で指示したり，動作内容をプレビューを見ながらメニューボックスから選択したりといったことができる．

これらの機能は，ユーザのタイピングの手間を省くと同時に，作成したプレゼンテーションを視覚的に把握しやすくすることで，スクリプトを作成しては実行試験を繰り返すという試行錯誤の回数を減らす効果も期待される．

7.3 言語仕様のバージョンアップ

MPML の仕様については，現在のバージョン 1.0 では視聴ユーザによるプレゼンテーション制御機能や双方向インタラクション機能は必要最小限のものしか与えていない．プレゼンテーションという一方向性が強い場面を対象にした情報コンテンツではあるが，応用の幅を広げる意味でも，技術発展に併せて，この面の機能の充実を図りたいと考えている．

現在，MPML の仕様と開発したコンバータは WWW 上で公開しているが[13]，使用環境を向上させて広く使用してもらえる WWW 上の共通言語となるように努力したい．

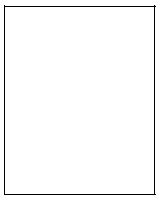
謝辞　マルチモーダルエージェント等について教示，協力いただいた同僚の土肥浩氏，Ashow システムの使用の便宜を図っていただいた長谷川修氏（電総研）に感謝致します．

参 考 文 献

1) 竹林洋一: 音声自由対話システム TOSBURG II –ユーザ中心のマルチモーダルインタフェースの実現に向けて, 信学論, Vol. J77-D-II, No. 8, pp. 1417–1428 (1994).
2) Nagao, K. and Takeuchi, A.: Speech Dialogue with Facial Displays: Multimodal Human-Computer Conversation, *Proc. 32nd Annual Conf. of ASSOC of Computational Lingustics*, pp. 102–109 (1994).
3) 土肥浩, 石塚満: WWW/Mosaic と結合した自然感の高い擬人化エージェントインタフェース, 信学論, Vol. J79-D-II, No. 4, pp. 585–591 (1996).
4) Dohi, H. and Ishizuka, M.: A Visual Software Agent: An Internet-Based Interface Agent with Rocking Realistic Face and Speech Dialog Function, *AAAI-96 technical report of Internet-Based Information Systems, WS-96-06*, pp. 35–40 (1996).
5) 土肥浩, 石塚満: Face-to-Face 型擬人化エージェント・インタフェースの構築, 情処論, Vol. 40, No. 2, pp. 547–555 (1999).
6) Hasegawa, O. and Sakaue, K.: A CG Tool for Constructing Anthropomorphic Interface Agents, *Proc. IJCAI-97 Workshop on Animated Interface Agents: Making Them Intelligent*, Nagoya, Japan, pp. 23–26 (1997).

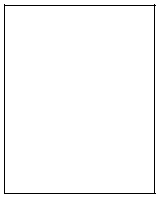
7) Noma, T. and Badler, N.: A Virtual Human Presenter, *Proc. IJCAI-97 Workshop on Animated Interface Agents: Making Them Intelligent*, Nagoya, Japan, pp. 45–51 (1997).
8) André, E., Rist, T. and J. Müller: WebPersona: A Life-Like Presentation Agent for the World-Wide Web, *Knowledge-Based Systems*, Vol. 11, No. 1, pp. 25–36 (1998).
9) 筒井貴之, 李寛容, 土肥浩, 石塚満: マルチモーダルプレゼンテーションエージェントの作成, 情処全大 (57), No. 1N-9 (1998).
10) 筒井貴之, 土肥浩, 石塚満: エージェントキャラクタによるWWW連携プレゼンテーション–マルチモーダルプレゼンテーション記述言語MPML–, マルチメディア 仮想環境研究技術報告, MVE98-90, 信学会 (1999).
11) XML HomePage: . http://www.w3.org/TR/REC-xml/.
12) SMIL HomePage: http://www.w3.org/AudioVideo/.
13) MPML HomePage: http://www.miv.t.u-tokyo.ac.jp/mpml.html.
14) Clark, D. and Stuple, S. J.(eds.): *Developing for Microsoft Agent*, Microsoft Press (1998).

(平成11年5月13日受付)
(平成12年2月 4 日採録)



筒井 貴之

1997年東京大学工学部電子情報工学科卒業．1999年同大学修士課程修了．同年 (株)NTTデータ入社，現在に至る．在学中にマルチモーダル・プレゼンテーションツールMPMLの研究を行い，現在，エレクトロニック ・バンキングの開発に従事．映像情報メディア学会会員．



石塚 満 (正会員)

1971年東京大学工学部電子工学科卒業．1976年同大学院博士課程修了．工学博士．同年NTT入社，横須賀研究所．1978年東京大学生産技術研究所助教授，同教授を経て，1992年より工学部電子情報工学科教授．研究分野は人工知能，知識処理，マルチモーダル擬人化エージェント，知能的ネットワーク化情報環境．IEEE，AAAI，電子情報通信学会，人工知能学会，映像情報メディア学会，画像電子学会各会員．