

非言語的合図に基づいて仮想都市を案内する対話エージェント

山中 信敏[†], 岡本 昌之[†], 中西 英之^{†, ††}, 石田 亨^{†, ††}

本稿の目的は仮想都市における周囲の人や物体など環境によって引き起こされる対話を扱う対話エージェントの提案である。従来の対話エージェントでは主に前後の文脈に依存するタスク指向対話を扱っているが、本稿では、仮想都市における重要な対話の一つである、環境の属性や非言語的合図を手がかりとする環境指向型対話に着目した。環境指向型対話はユーザの注視・言語的合図・場所・対象の特徴からなる手がかりを条件とし、エージェントの行動を出力とするルールを用いてモデル化される。また、我々は仮想都市において環境指向型対話を扱うエージェントをシナリオ記述言語によって実装した。

Conversational Agents Guiding Virtual Cities based on Nonverbal Cues

Nobutoshi YAMANAKA[†], Masayuki OKAMOTO[†], Hideyuki NAKANISHI^{†, ††}, Toru ISHIDA^{†, ††}

This paper proposes a new kind of conversational agents dealing with conversation which surrounding environment such as a human or a building triggers off. We focused on environment-oriented conversation in which properties of an environment or nonverbal signals are cues, while usual conversational agents mainly treat with task-oriented conversation which depends on conversational contexts. Environment-oriented conversation is modeled with cues consisting of observation, verbal signals, place, and features of objects as a condition part, and with an agent's behaviors as a series of actions. We also implemented conversational agents with a functionality of environment-oriented conversation in a 3-D virtual city with a scenario-description language.

1. はじめに

仮想都市は日常の生活空間を再現した環境である[1]。そこでは、多数の人間やエージェントが混在し、自由に出会い、そしてインタラクションが行われる。我々は、このような仮想都市においてアバターとして参加してくる人間と対話するエージェントの構築を試みる。仮想都市における訪問者は自由に環境内を移動できるため、タスク指向型対話だけでなく、周囲の人や物体など環境によって引き起こされる対話が頻繁に生じる。我々は、このような周囲の人や物体など環境によって引き起こされる対話を環境指向型対話と呼ぶ。環境指向型対話も、仮想都市における訪問者間の共有基盤[2]を形成する上で非常に重要である。そのため、仮想都市における対話エージェントには環境指向型対話を扱う能力が求められる。

ところが従来の具現化対話エージェント[3]は、タスク指向型対話に焦点を当ててきたため、このような対話は見過ごされてきた。

環境指向型対話を扱うためには、利用者の言動から対話の前提となる手がかりを抽出することが必要となる。仮想都市におけるこれらの手がかりは従来のようにタスクや対話の前後の文脈だけでなく、“誰がどこにいるのか”、“何がどこにあるのか”といった環境の属性や訪問者の関心がどこに向けられているのかを示す非言語的合図にあることが多い。

例えば、「あれは何?」といった質問発話や、立ち止まってある方向に視線を向けるという動作は環境への関心を表す言動である。これらの言動に対してタスクや前後の文脈をいくら解析しても応答したり対話を開始したりすることは難しい。しかし発話や動作を行っている人のいる場所や体の方向、視線の向きといった情報を用いることで対話を進めることが可能である。

我々は、環境指向型対話を扱うために、注視、言語的合図、場所、そして対象の特徴とという4つの手がかりに着目した。そして、それらの手がかりに基づいて環境指向型対話を扱いながら仮想都市を案内する役割を務める対話エージェントの構築を行った。

[†] 京都大学大学院情報学研究科社会情報学専攻
Department of Social Informatics, Kyoto University

^{††} JST CREST デジタルシティプロジェクト
JST CREST Digital City Project

表 1：環境指向型対話とタスク指向型対話

	環境指向型対話	タスク指向型対話
依存関係	発生場所	タスク
利用するリソース	環境のリソース	前後の文脈のリソース
対話開始の合図	言葉，非言語的合図，環境の空間的状況	主に言葉

2. 環境指向型対話

2.1 環境指向型対話の特徴

仮想都市においてアバターとエージェントは同じ環境を共有するため、アバターを操作している人間の環境への没入感は大きくなり、結果として仮想都市環境の中の状況に対してより多くの注意が向けられる。このような状況では、タスク指向型対話だけでなく、周囲の人や物体など環境によって引き起こされる対話が頻繁に生じる。我々は、このような対話を環境指向型対話と呼ぶ。環境指向型対話をタスク指向型対話と比較して特徴をまとめたのが表 1 である。タスク指向型対話はタスクに依存してある目的に向かって進むのに対して、環境指向型対話は発生場所に依存するため、その時その場所でしか意味を成さないことが多い。また、環境指向型対話は、言葉だけでなく、環境の空間的状況や非言語的合図も対話開始の合図となるという特徴を持っている。これは仮想都市空間におけるアバターも、関心表明のために発話だけでなくさまざまな動作を用いるためである。

2.2 環境指向型対話を引き起こす手がかり

2.1 で述べたように環境指向型対話は発話だけでなく環境の空間的状況、非言語的合図などによって開始される。我々は、このような対話開始の合図として以下の 4 種類を取り上げた。

1) 注視

現実世界における人間同士のコミュニケーションの多くが言葉以外によって成り立っている[4]のと同様に、仮想都市においてもアバターの発する非言語的合図はアバターの関心を認識するための情報として重要である。注視とは、このような非言語的合図の一つで対象を判断しようとする意志を表す体の向きや視線を対象に向ける動作[5]のことである。仮想都市における注視の対象となるものは、街や構造物などの物体や場所、あるいは参加者のいずれかである。特に、

仮想都市においては現実世界以上に視界情報に頼ることが多いため注視が環境指向型対話開始の手がかりとなる。

2) 言語的合図

参加者の発話、発話に含まれるキーワードといった言葉を中心とする情報を表す。場所や建物、方向、あるいは周囲の人や物体といった環境に関するキーワードが含まれる発話は環境指向型対話開始の手がかりとなる。

3) 場所

仮想都市における訪問者の位置、位置から特定される周囲の建物や物体に関する領域情報を表す。環境指向型対話は場所に依存するため、訪問者のいる場所に関する情報も環境指向型対話を開始する手がかりとなる。

4) 対象の特徴

仮想都市におけるエージェントやアバターといった参加者や物体などが持つ特徴を表す。例えば参加者の服装や建物の種類や形状などが挙げられる。対象の特徴によってアバターの興味を引いたり、引かなかったりするため、このような対象の特徴も対話開始の手がかりとなる。

実際には、これら 4 つの手がかりを単独で対話開始の手がかりとして用いるだけでなく組み合わせて利用する。

3. 環境指向型対話を扱うエージェント

3.1 ルールによる対話制御

従来の対話エージェントが対象としてきたタスク指向型対話は、ある目的に向かって進んでいくという特徴をもっているため、対話の流れをモデル化する必要があった。このようなモデルの例としては、言語行為理論とプランニングを組み合わせた対話モデル[6]や、発話を発話タイプというクラスに分けて、その系列を N-gram によってモデル化した対話モデル[7]などが挙げられる。

これに対して、環境指向型対話は、前後の文脈

との関係が希薄である，言葉だけでなく環境の状況や非言語的合図によって開始されるといった特徴から，対話の流れをモデル化するよりも，対話の前提となる手がかりをモデル化することが重要となる．そして，その手がかりに反応して，応答を返すルールに基づいて対話を進めるという方法が適している．本研究では，仮想都市において環境指向型対話の開始の手がかりを 2.2 で説明した注視，言語的合図，場所，対象の特徴という 4 つとする．

3.2 対話発生の流れ

ルールに基づいた対話発生の流れを図 1 に示す．まず建物・人の属性，参加者の位置，体の向き，視線の方向，発話といった仮想都市と参加者に関する入力パラメータから対話開始の合図を表す言語的合図，注視，場所，対象の特徴という手がかりを抽出する．次に，得られた手がかりとルールの条件部とを照合し発火するルールをすべて選択する．そして優先順位に基づいて実際に実行する 1 つのルールを決定し，実行するルールの出力部にある発話と動作の系列を実行することで対話が始まる．

以下，環境指向型対話開始の手がかりの抽出として入力パラメータからの注視の抽出と多数のルールから実行するルールを選択するために用いるルールの優先順位について説明する．

3.2.1 環境指向型対話開始の手がかりの抽出

我々は，抽出動作を建物への注視と人への注視という 2 種類に分けた．そして，それぞれの特徴に応じてアバターの一連の動作から注視の抽出を行った．図 2 は建物と人への注視の抽出方法を示したものである．

建物への注視は，アバターと建物の端点を結ぶ辺とのなす角を θ_1 ， θ_2 とした時，アバターの視線の方向が θ_1 と θ_2 の間に入るような動作を建物への注視とした(図 2(a))．人への注視は，対象が自由に動くことを考慮しアバターからの距離が distance 以内の参加者のみを対象として，アバターの視野角 θ にその参加者が入るような動作をその人への注視とした(図 2(b))．

3.2.1 ルールの形式と優先順位

条件が満たされた複数のルールから実際に実行する 1 つのルール選択を行うために用いるルール

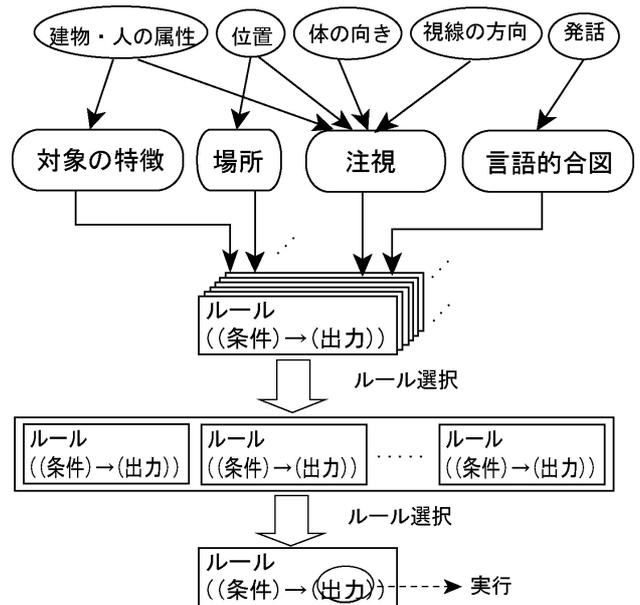


図 1：対話発生の流れ

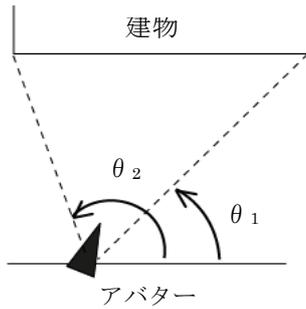
の優先順位は次の順番に従って決定される．

まず，条件部の要素数が多いほど条件の優先度は高くなる．要素数が同じ場合は，1)条件部に言語的合図に関連した要素を含む，2)参加者への注視に関連した要素を含む，3)物体への注視に関連した要素を含む，という順番に従って優先順位を決定する．これは，言葉は理解されることを前提として意図的に発せられるということ[4]，人間は動くものの方が静止しているものよりも注意を引きやすいということを考慮したものである．エージェントは，このようなルールを多数持ち，ルールに基づいて動作することで環境指向型対話を扱う．

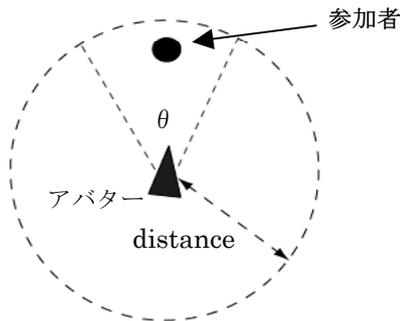
3.3 対話ルールの記述

実際の対話ルールの具体例を示す．対話ルールは，2.2 で示した 4 つの要因を条件部に持ち，出力部には条件に対応した発話と首振り，体の向きの変化，あるいは指差しといった動作の系列が記述される．

図 3 は，アバターの建物への注視に反応して対象を説明する対話ルールの例である．ルールの条件部には，注視を認識する手がかりが記述されており，出力部には条件が満たされたときに実行される発話と動作系列が記述されている．一行目の“?”から始まる記述にある coordinates や building といったパラメータに応じてアバターの建物への



(a)建物への注視



(b)人への注視

図 2 : 注視の抽出

の注視を認識する。2 行目以降の “!” から始まる記述は注視の認識が行われたときに実行する出力部である。出力部には、アバターの方に視線と体を向けてから話し始め、話し終わるとまた元の向きに戻るといった発話を含む一連の動作系列が記述されている。

図 4 は、人への注視に反応して対象となった人のプロフィールなど公開された属性を取得してアバターに伝達する対話ルールの例である。“?” から始まる 1 行目には人への注視を認識するための手がかりが記述されており、distance や view というパラメータに応じてアバターから一定距離内にいる他の参加者への注視を認識する。2 行目以降は、対象となった人を以前に紹介していないことを確認してから、その人の属性を取得しアバターに伝えるということが記述されている。

我々は、これらのルールの他にアバターの場所に関連した言語的動作に反応して対話を始めるルールと、アバターがいる場所に反応してその場所に関して対話を始めるというルールを用意した。

そして、そのようなルールをエージェントに持たすことで環境指向型対話を扱うことを実現した。

エージェントは、これらの対話ルートを仮想都市における街や建造物の種類に基づいて分割された

```
(?position :name avatar
:from coordinates :relative_angle buildings)
(!face :to avatar)
(!turn :to avatar)
(!speak to avatar
:sentence "This is a variety shop ...")
(!face :to forward)
(!turn :angle forward)
(!behave :gesture "standing")
(!!walk :route (list Destination))
```

図 3 : 建物への注視に反応する対話ルールの例

```
(?position :name $agent
:from Avatar :distance_range distance :angle_range view)
(if (memq $agent $done-list) (go scene next)
(begin
...
(!speak to avatar
:sentence (string-append "That person's nickname is"
(agent-attrib $nearest-agent "name") ".")
(agent-attrib $nearest-agent "open-profile")))
...
(!!walk :route (list Destination))))
```

図 4 : 人への注視に反応する対話ルールの例

領域に関連付けて管理する。これは各領域に応じて必要なルールや優先順位が異なると考えたためである。つまり、すべての動作に反応して対話を始めるのではなく、領域によってはアバターの動作に反応しないなど領域に応じた処理を行っている。

4 案内エージェントの構築

4.1 対話エージェントのモデル

仮想都市においては、タスク指向型対話と共にこれまで説明してきたような環境指向型対話が頻繁に生じる。そのため、仮想都市における対話エージェントは、タスク指向型対話が生じている間は、確率的モデルなどに従ってタスクに沿って対話を進め、タスク指向型対話が生じていない間はルールに基づいて環境指向型対話を進めるという 2 つの対話を扱うモデルが必要となる。

通常、タスク指向型対話はある目的に向かって進んでいく対話であるためタスク指向型対話と環境指向型対話の内容の競合が生じた場合は、タスク指向型対話を優先する。このため、環境指向型対話を扱うルールの実行は、タスクに影響しない

かどうかを考えて行う必要がある。

4.2 案内エージェントの構築

我々は、実際の仮想都市において道案内という役割を与え、その道中に環境指向型対話を扱うことができる案内エージェントを構築した。道案内の役割を与えたのは、歩いている最中に周囲の状況が常に変化するため、環境指向型対話が生じやすい、人間の姿をしたエージェントが行う役割として現実世界の規範から逸脱していないと考えたからである。仮想都市として3次元仮想都市環境FreeWalk[8]を、エージェント制御には、インタラクション記述言語Q[9]を用いた。構築した案内エージェントは、仮想都市においてアバターと出会い、挨拶してから目的地に連れて行く。目的地に向かう道中においてアバターがエージェントから離れようとすると一緒に歩くように促す発話を行う。道中でアバターがエージェントと一緒に歩いている場合はルールに基づいて環境指向型対話を行う。そして目的地につくとアバターに到着したことを説明して別れる。

エージェントは、1)導入部、2)目的地に向かう道中、3)終了部、という3つの状態を持っておりそれぞれの状態に道案内に必要な対話が組み込まれて入る。環境指向型対話を扱う対話ルールは2)の目的地に向かう道中に記述した。エージェントが2)の状態にいるときアバターがエージェントから離れていくとアバターの言動に関わらず一緒に歩くことを求める発話をするように設計した。

4.3 案内エージェントの動作

構築したエージェントの実際の仮想都市における動作について述べる。図5及び図6は目的地に向かって歩く道中において環境指向型対話が生じる様子を示したものである。

図5は図3の対話ルールの基づいてエージェントがアバターに話しかける様子である。エージェントはアバターといっしょに歩きながら常にアバターの言動や環境の属性から対話開始の手がかりを探す(図5(a))。アバターが建物の方向に視線を向け始める

とそれを観察する(図5(b))。そして、アバターの視線の方向がルールに記述されている方向と同じ角度になったことを確認するとそれに反応して出力部にある動作系列に従ってアバターの方向を向い

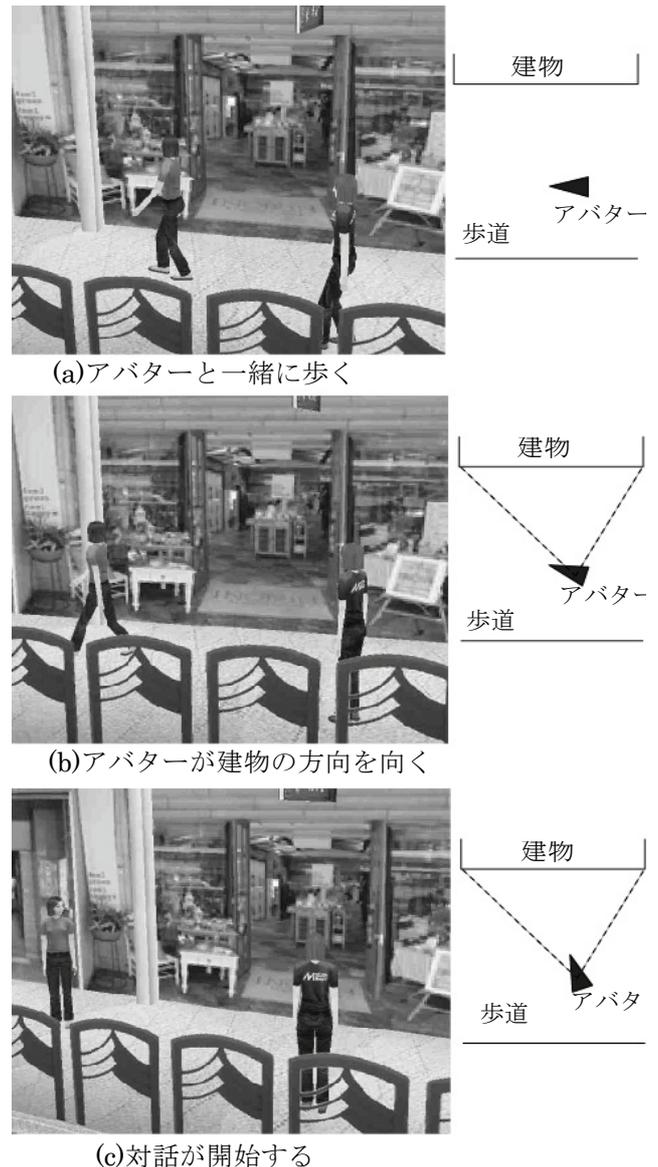


図5：建物への注視によって開始する対話

てから対象の建物の説明対話を開始する(図5(c))。

図6は図4の対話ルールに基づいてエージェントが話しかける様子を表したものである。エージェントはアバターといっしょに歩きながら、アバターから一定距離にいる参加者を探し出し(図6(a))、アバターがその参加者に気が付いてその方向に視線を向ける(図6(b))。エージェントはアバターの視線の角度にその参加者が入ったことを確認すると、ルールに従って対象となっている参加者の属性を取得し、アバターの方向を向いて属性情報を伝える(図6(c))。エージェントは、注視対象となった訪問者のプロフィール等の公開された属性をアバターに伝達する。

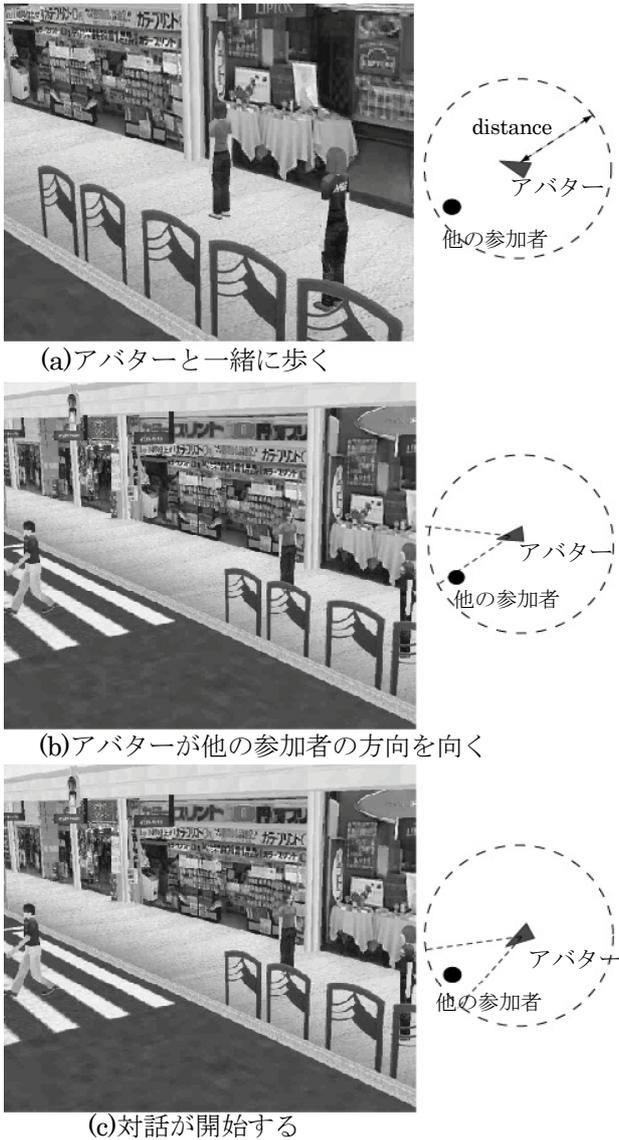


図 6：人への注視によって開始する対話

従来の対話エージェントは、エージェント主導で対話を進めるものがほとんどである。しかし、我々のエージェントは、常にアバターの言動から開始されるため主導権はアバターにある。また、単に発話だけでなくアバターの注視に反応して対話が始まるという点も特徴的である。

このようなエージェント構築するためには、仮想都市のどの場所でもどんなルールが必要となるのかが重要となってくる。そこで、実際の挙動を確認しながらパターン化されたルールをエージェントに組み込んでいく作業を支援するエージェント構築環境を開発中である。

5 おわりに

本稿では、従来の対話エージェントでは対象と

されてこなかった仮想都市における環境の属性や非言語的合図を手がかりとする環境指向型対話に着目した。そして、そのような対話をユーザの注視・言語的合図・場所・対象の特徴からなる手がかりを条件とし、エージェントの行動を出力とするルールを用いてモデル化した。また、実際の仮想都市において環境指向型対話を扱う対話エージェントをシナリオ記述言語によって実装した。

仮想都市における対話エージェントは、タスク指向型対話と環境指向型対話の両方を扱うことが求められる。今後は、本研究で実現した環境指向型対話をタスク指向型対話と組み合わせるより自然に対話を進めることができるエージェントを設計する予定である。タスク指向型対話と環境指向型対話の移行や対話リソースの競合解決などは今後の課題である。本研究の最終的な目標は仮想都市において人間同士のコミュニケーションを支援するエージェントの設計にある。

謝辞

本研究は科学技術振興事業団 CREST「デジタルシティのユニバーサルデザイン」プロジェクトから援助を受けて行われました。

参考文献

- [1] Ishida, T. Digital City Kyoto: Social Information Infrastructure for Everyday Life, *Commun. ACM*, Vol. 45, No. 7, pp. 76-81 (2002).
- [2] Clark, H.: *Using Language*, Cambridge University Press (1996).
- [3] Cassell, J., Sullivan, J., Prevost, S. and Churchill, E. Eds.: *Embodied Conversational Agents*, MIT Press (2000).
- [4] 黒川隆生: *ノンバーバルインタフェース*, オーム社 (1994).
- [5] マジョリー・F・ヴァーカス著, 石丸正訳: *非言語コミュニケーション*, 新潮新書, 新潮社 (1987).
- [6] Allen, J.: Analyzing Intension in Utterances, *Artificial Intelligence*, Vol. 15, pp. 143-178 (1980).
- [7] Nagata, M. and Morimoto, Y.: First steps towards statistical modeling of dialogue to predict the speech act type of the next utterance, *Speech Communication*, Vol. 15, pp. 193-203 (1994).
- [8] Nakanishi, H., Yoshida, C., Nishimura, T. and Ishida, T.: FreeWalk: A 3D Virtual Space for Casual Meetings, *IEEE MultiMedia*, Vol. 6, No. 2, pp. 20-28 (1999).
- [9] Ishida, T.: Q: A Scenario Description Language for Interactive Agents, *IEEE Computer*, Vol. 35, No. 11, pp. 42-47 (2002).