

日本語文章へのジェスチャー自動付与による 仮想世界アバターの実現

塚本 剛生[†] 中野 有紀子[†]

成蹊大学理工学部情報科学科[†]

1 はじめに

セカンドライフ (SL) 等の仮想世界が存在し、幅広いユーザに利用されつつある。このような仮想世界上で相手とコミュニケーションをする主な方法は、アバターと呼ばれる自身の分身を用いてチャットをさせることであるが、現状では、吹き出しにチャットテキストが表示される方式であり、対面場面での音声や身振りによる表現は非常に限定的である。

一方、対面コミュニケーションの研究では、非言語行動を加えることによって相手に情報をより正確に伝えることができることがわかっており、ユーザのチャットテキスト入力から、ジェスチャーつきのアバターアニメーションが自動生成できれば、仮想世界のコミュニケーションにおいて効果的であると期待される。しかし、ジェスチャー自動決定方式について、特に日本語に関しては十分な研究がなされていない。

そこで本研究では、SL のチャット機能を用いて日本語文章を入力すると、その文章に適したジェスチャーが自動付与され、日本語合成音声とタイミングを合わせてアバターアニメーションが自動で生成される機構を提案する。

2 ジェスチャー選定

2.1 ジェスチャー選定にかかわる要因

[Nakano,04]は、テキストが入力されると、エージェントのアニメーションスケジュールと、合成音声を用いて発話音声を作成するシステムを実装している。[Breitfuss,08]は、SL 内で英語の文章を入力すると、ジェスチャーアニメーションを実行するシステムを開発しており、一般的な場合はビートジェスチャー (腕や手の上下の運動) が実行される。しかし、これらの研究ではジェスチャーの形態決定については考慮されていない。

一方、[Tepper,04]は道案内の対話に焦点を当て、ジェスチャーの形態決定の方式を提案している。彼らの方式では、物の形の特徴をいくつかの次元における程度として表現している。例えば玄関は

水平方向に広がるものであり、横幅は大きく、奥行きは普通である。このように、言葉の概念を特徴づける次元とその値のセットとして単語を登録することにより、発話の意味表現からジェスチャーの形態を決定することができる。これにより、単語エントリとジェスチャーの形態の組み合わせをジェスチャー辞書としてあらかじめ用意することなく、より汎用性が高く、柔軟なジェスチャー選定が可能になる。そこで本研究では、英語を対象にした[Tepper,04]のジェスチャー決定方式を参考に、日本語のジェスチャー形態自動付与機構を実現する。また、SL をプラットフォームとして採用することにより、提案方式が仮想世界でのチャットシステムに応用可能であることを示す。

2.2 ジェスチャー選定アルゴリズム

本研究では[Tepper,04]を基に文中の単語に意味表現的な特徴を付与することによって、ジェスチャーを決定する。単語に 5 つの特徴 (“shape”, “position”, “width”, “depth”, “height”) を付与することで、特徴に合ったジェスチャーを選定することができる。単語の特徴を登録するための xml の具体例を図 1 に示す。この例では、「建物、ビル、デパート」の単語に横幅は普通、奥行きは普通、高さは大きいという特徴を記述している。この場合、横幅と奥行きは普通で、高さは大きいという特徴から「四角」という種類のジェスチャーが選定される。

```
<?xml version="1.0" encoding="SHIFT_JIS">
<database>
<word w="建物、ビル、デパート">
<feature shape="nospecify" position="nospecify" width="normal" depth="normal" height="large">
</database>
```

図 1: 単語の特徴例

3 ジェスチャー自動付与システム

図 2 のシステム構成図に示すように、本システムは、(1)Text Receiver, (2)ジェスチャー決定部、音声合成器, (3)Action-Voice Controller(AVC), の 3 つのモジュールから成る。プラットフォームには SL のオープンソース版である OpenSim を使用している。まず、OpenSim から Text Receiver にチャットテキストが送られると、Text Receiver はその日本

連絡先: 塚本剛生, 成蹊大学理工学部情報科学科,
〒180-8633 東京都武蔵野市吉祥寺北町 3-3-1
E-mail:us062071@cc.seikei.ac.jp

語文章をジェスチャー決定部に送信する。ジェスチャー決定部では受け取った日本語文章に適したジェスチャーのタイミングと形態を自動付与し、日本語合成音声とそれに合ったアニメーションのタイムスケジュールを作成する。AVC ではジェスチャー決定部で作成した合成音声を再生し、タイミングを合わせてジェスチャーアニメーションの実行命令を OpenSim 内のアバターに送信する。

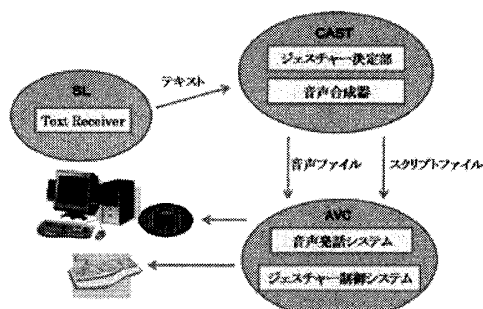


図 2: システム構成図

4 システム実装

システムの実装について、先に述べた 3 つのモジュールを基に説明する。まず、Text Receiver では SL 独自のプログラミング言語であるリンデンスクリプト言語 (LSL) を用いる。SL ではプリムと呼ばれるオブジェクトに LSL を組み込み、それをアバターに装着させることでアバターの制御ができる。この機能を用いて、ユーザが入力したテキストをジェスチャー決定部に送信する機能を実装した。

ジェスチャー決定部は CAST[Nakano,04]を拡張することにより実装した。CAST は、テキストを入力として受け取ると、日本語の形態素解析と統語解析を行い、入力文章の言語的特徴からジェスチャーを付与すべきタイミングを決定し、リップシンクのための viseme とあわせてアニメーションのタイムスケジュールを自動生成する。本研究での拡張により、この CAST システムに単語の意味に合ったジェスチャー形態を選定する機能を付加した。

AVC では、作成した日本語合成音声を再生し、タイムスケジュール通りにアバターにジェスチャーを適用する。具体的には、あらかじめ SL 内で F2 から F12 のキーにジェスチャーアニメーションを登録し、これらのキーをタイムスケジュールに合わせて自動的に作動させることにより AVC を実装した。ジェスチャーアニメーションは、より人間らしさを追求するために、モーションキャプチャソフトで実際の人間の動きを撮影し、その動きを基に BVH 形式のアニメーションファイルとして保存することにより作成された。以上の機構により、SL で日本語文章を入力すると、日本語音声合成が再生されると同時に、その文章に適したジェスチャーが出力される。

5 動作例

システムの動作例として、アバターに道案内をさせることを想定し、ジェスチャーの形やタイミングを検証した。簡単な例として、「あそこにビルがあります。」という一文を用いる。この文をチャットで入力すると、アバターの背中に付けている Text Receiver がテキストを読み込み、CAST に送信し、CAST は受け取った文を解析する。この文には「ビル」という単語が含まれているので、図 1 に記された特徴を参考に「四角」という種類のジェスチャーを選定する。その後、ジェスチャーの種類とタイミングが書かれたスクリプトファイルと日本語合成音声の音声ファイルが作成される。CAST で作成されたスクリプトファイルと音声ファイルを AVC が読み込み、音声ファイルを再生しつつ、あらかじめ OpenSim 内で登録した「四角」ジェスチャーのキーをスクリプトファイルに書かれたタイミングで自動押下することで、OpenSim 上のアバターでジェスチャーが実行される。その結果、図 3 に示すようなジェスチャーが生成される。

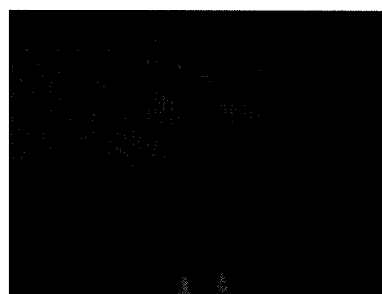


図 3: セカンドライフ内での動作

6 結論

SL において、チャットで日本語文章を入力すると、その文章に適したジェスチャーを自動付与し、日本語合成音声とタイミングを合わせてアバターアニメーションを生成する方法を提案した。

ジェスチャーの選定時に文脈を考慮することや、単語や文脈によってジェスチャーの優先度を調整することが今後の検討課題である。

<参考文献>

- [Nakano 04] Nakano, Y. et al. Converting text into agent animations: Assigning gestures to text. In *Proc. of HLT-NAACL 2004 Companion Volume*, pp. 153–156, 2004.
- [Breitfuss 08] Breitfuss, W. et al. Automatic generation of gaze and gestures for dialogues between embodied conversational agents. *Int'l J of Semantic Computing*, Vol. 2, No. 1, July 2008, pp 71-90
- [Tepper 04] Tepper, P. et al. (2004) "Content in Context: Generating Language and Iconic Gesture without a Gestionary" *Proc. of the Workshop on Balanced Perception and Action in ECAs at AAMAS '04*.