

# TVML を用いたノンビデオリアリスティック フェイシャルアニメーション

藤島 智子      藤代 一成

お茶の水女子大学 大学院人間文化研究科

〒 112 - 8610 東京都文京区大塚 2 - 1 - 1

tomo@imv.is.ocha.ac.jp, fuji@is.ocha.ac.jp

コンピュータグラフィックスにおいて、架空のキャラクターが人間らしく振る舞うヒューマンアニメーションは、エンターテインメントにおける需要も高く、長い間研究が進められてきた。しかし、現実感あふれるヒューマンアニメーションを制作するためには、専門的な知識と経験的な知識が必要となり、一般のユーザが魅力のあるアニメーションを制作することは容易なことではない。本研究では、簡易なレンダリング法を用いながらも、特徴的な動きを強調することで、より効果的な映像を受け手に与えることができるという考え方を発展させて、ノンビデオリアリスティックレンダリング (Non-Videorealistic Rendering : NVR) というコンセプトを提案する。本稿では実際に、発話状態における NVR に焦点を合わせ、TVML のキャラクタに舌を組み込む方法を提案する。

## Non-Videorealistic Facial Animation using TVML

Tomoko Fujishima      Issei Fujishiro

Graduate School of Humanities and Sciences  
Ochanomizu University

2-1-1 Otsuka, Bunkyo-Ku, Tokyo 112-8610, Japan

tomo@imv.is.ocha.ac.jp, fuji@is.ocha.ac.jp

A great demand for the entertainment industry has been an incentive to conducting active R and Ds in the area of human animation with fictitious characters. However, special knowledge and expertise are still required for common users to create fascinating human animation pieces. In this research, a focus is placed on the fact that a simple animation with ingeniously-emphasized feature human motions can provide impressive information that may not readily apparent in those produced by using sophisticated techniques. The fact can lead to a new paradigm termed non-videorealistic rendering (NVR). This article shows a first step toward the concept by presenting a method for rendering a tongue in an NVR manner to make a TVML character speak lively and expressively.

## 1 背景と目的

近年コンピュータグラフィックスの普及はめざましく、日常生活の中でも、映画やCMなどのさまざまなメディアを通して目にする機会が増えてきた。そのなかでも、架空のキャラクターが人間らしく振る舞うヒューマンアニメーションは、エンターテインメントでの需要がもっとも高い分野の一つである。

しかし既存のアニメーションは、テレビといった受動的なメディアを通して提供されているものが多く、一般のユーザ自らがアニメーションを制作する際には、以下のような三つの問題点が生じる：

- アニメーションを制作するには、モデリングやテクスチャリングといった専門的な知識と、カメラパスや照明効果といった経験的な知識が必要であり、一般のユーザが簡単に作品を制作することは困難である。
- キャラクターの動作やセリフを変更するときには、はじめからアニメーションを制作し直さなくてはならず、ユーザがアニメーションをインタラクティブに操作することができない。またインタラクティブな操作が許されているテレビゲームでも、決められたシナリオを無視した行動や選択肢以外の選択は許されない。
- アニメーションのファイルサイズは一般的に大きくなるため、ネットワーク上での配信には適していない。ユーザがアニメーションを制作したとしても、ネットワーク上で Web ページのように気軽に公開することは困難である。

ところで、人間は心に残る映像を記憶する際に、目で見ただけの映像をそのまま記憶するのではなく、特徴的な部分を強調して記憶する傾向がある。さらに、時間とともにその特徴はますます強調されていく。このような考えに基づいて、コンピュータグラフィックスのなかで生まれたのがノンフォトリアリスティックレンダリング (Non-Photorealistic Rendering: NPR) [1] である。ヒューマンアニメーションにおける NPR の代表例は似顔絵 [2] である。似顔絵を描くときには、特徴的な部分を誇張することで、より

効果的な画像を受け手に与えることができる。しかし、そのような NPR 画像のスナップショットを連ねることによって生成したアニメーションが、つねに同様の効果を与える保証はない。部品の静的な誇張に加えて、対象の特徴的な動きそのものに注目し、それを誇張することが必要になると考えられる。本研究ではこのような考えを、ノンビデオリアリスティックレンダリング (Non-Videorealistic Rendering: NVR) と定義し、ヒューマンアニメーションの分野に適用することを提案する。

本研究では、“ユーザが簡単に NVR アニメーションを制作すること”を目的とする。アニメーション制作には、作品の制作が容易な TVML (TV program Makeup Language) [3] を用いる。TVML は、台本 (スクリプト) によってテレビ番組を丸々一本記述することができるテキストベースの言語である。これを TVML プレイヤーとよばれるソフトウェアに読み込ませることで、テレビ番組をリアルタイム表示することができる。本稿では、TVML を用いて、特徴ある動きを効果的に強調する手法を提案する。

本稿は、以下のように構成されている。まず、次章で TVML の概要を述べ、第 3 章で関連研究について簡単に説明する。第 4 章では NVR を実現するための方針について詳しく述べ、第 5 章で実験結果を示す。最後に第 6 章で本論文をまとめ、今後の課題に言及する。

## 2 TVML

TVML [3] は、テレビ番組に必要な要素の多くを、誰でも理解しやすいような抽象度の高い表現で記述することができるテキストベースの言語である。例えば、“ズームイン”と記述することで、カメラがキャラクターにズームしていく様子を表現することができる。そのため、各設定に関する経験的な知識を必要としない。また、ループや条件分岐、GOTO 文は採用せず、時間の流れに従ってイベントを列挙していくスクリプト形式を採用しているため、コンピュータ言語に関する専門的な知識も必要としない。

TVML で設定することができる要素の例を以下に示す .

- キャラクタのセリフや動作
- スタジオのセット
- カメラ, 照明の設定
- 動画, テキスト情報の挿入
- オーディオファイルの再生
- ナレーション

図 1 に TVML スクリプトの例を挙げる .

```
camera: twoshot(name1=tomo, name2=shoko, adjusty=-0.2, adjustvangle=5)
character: look(name=tomo, what=camera)
character: look(name=shoko, what=camera)
character: talk(name=tomo, text="今日のお弁当はなににする?")
character: talk(name=shoko, text="春巻きかなあ?")
character: talk(name=tomo, text="おいしいもんね")
character: talk(name=shoko, text="じゃ決定!")
super: off()
```

図 1: TVML スクリプト例

一方, TVML プレイヤーは, TVML で書かれた台本を読み取り, 自動的にリアルタイムで番組映像音声を出力するソフトウェアである . TVML プレイヤーは, 音声合成装置をもったグラフィックワークステーション上で動作する . ユーザは, TVML スクリプトと動画やオーディオファイルといった各種データを与えることで, 簡単に番組を制作することができる . TVML プレイヤーは, TVML スクリプトによって記述されたイベントを一行ずつ解析し, リアルタイムに映像を表示するインタプリタとして動作する .

TVML を用いることで, 前節に掲げた一般のユーザがアニメーションを制作する際の問題点が以下のように解決される .

- テレビ番組を制作するために必要なコマンドが用意されているため, カメラパスや照明効果といった経験的な知識や, コンピュータ言語に関する専門的な知識が不必要であり, 一般のユーザが簡単に作品を制作することができる .

- スクリプトをインタラクティブに操作することができるため, 簡単にキャラクタの動作やセリフを変更することができる . また, TVML プレイヤーはインタプリタとして動作するため, 作品をリアルタイムに表示することができる .
- あらかじめユーザが, TVML プレイヤーとキャラクタの画像情報をもってれば, 新しい作品を制作したときに伝送するデータはテキスト形式ですむため, ネットワーク上での配信に適している .

TVML は, ニュース番組のような情報提供番組の制作を目的として開発されてきたため, キャラクタの動きは少なく, キャラクタの発話が中心となる . そのため, 特に顔を強調したキャラクタ (図 2) を用いて, 動きが効果的に誇張されている点が特徴的である .



図 2: TVML キャラクタ例

本研究では, TVML における発話状態に対して, その NVR の特徴を増強することを考える . 具体的には, 発話の特徴を強調するために, キャラクタに舌を組み込むことを提案する . これにより, 発話の聴覚情報に視覚情報が付加され, 効果的な映像を受け手に与えることができるからである . また, 舌というパーツを制御することにより, 既存のアニメーションにはない, 新しい表情を制作することもできると考えられる .

### 3 関連研究

人間は, 話をするとき相手の顔を見ることが多い . 話しているときに声が聞き取れなくとも, 顔の表情や口の動きを見ながら話を聞くことで, よりよく相手の話の内容を理解することができる . そのため顔は, 重要な役割をはたす

部分であり、コンピュータグラフィックスにおいてはフェイシャルアニメーションという分野で精力的に研究が進められてきた。

フェイシャルアニメーションで用いられている主な技法 [4] を以下に示す。

- 内挿法：[5-11]  
中間画像に補間計算を用いるため、アルゴリズムが容易である。しかし、表情に応じて画像データを用意しておかなくてはならないため、データ量が増えてしまう可能性がある。
- パラメタ法：[12]  
アルゴリズムが容易でリアルタイムに表示することができる。しかしパラメタごとに時間変化の値を設定したり、パラメタごとに制約を加える必要がある。
- Performance Driven 法：[13]  
実際の動きのデータをモーションキャプチャにより取得するため、本物らしさを表現することができる。また特別な計算を必要としないため、リアルタイムに表示することができる。しかしデータ量が膨大になり、データをとっていない表情は表現することができない。
- 筋肉モデル法：[14-16]  
解剖学的に正確で、顔の形状データが不必要である。しかし筋肉の時間変化の値を設定する必要があり、リアルタイムに表示することが難しい。
- 擬似筋肉法：[17] [18]  
顔の形状データが不必要で、リアルタイムに表示することができる。しかし形状の変化の順番によって、表情が変化してしまう。また自然な表情を合成することが難しい。

これらの技法のうち、どの技法が優れているかは一概には決められず、目的により適切な技法を選択することになる。TVML における表情のアニメーションでは、アルゴリズムが容易で、リアルタイム表示ができる内挿法を用いている。本研究では、TVML のキャラクタに舌を組み込んで、表情と舌を同期させたアニメーションを表示するため、補間計算にモーフィングを用いた内挿法を適用する。

## 4 キャラクタ制作の流れ

### 4.1 モデリング

モデリングには、エンターテインメント分野で頻繁に使われている *LightWave3D<sup>TM</sup>* (Ver. 7.5) \* [19] を用いる。顔や体の各オブジェクトは、パーツごとにモデリングする。複雑なパーツは、基本形状の変形と組み合わせにより製作し、各パーツを組み合わせることで、オブジェクトを製作する。TVML では、パーツの階層構造は内部的に決まってしまうため、骨の組み込みは必要ない。TVML のキャラクタには、BM 型と TK 型の 2 種類があり、階層構造の形式が異なる (図 3)。BM 型、TK 型のどちらにおいてもあらかじめ決められたパーツの中に“舌”というパーツは存在せず、舌を独自に制御することはできない。そこで、頭 (head) に舌を組み込んでモデリングを行うことを考える。

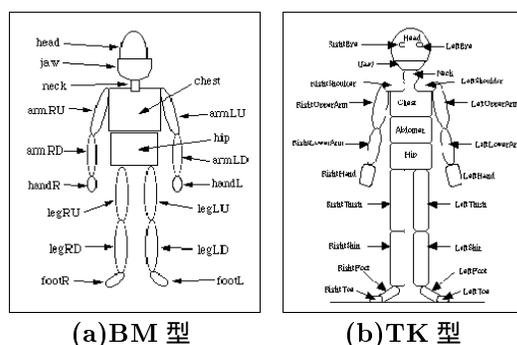


図 3: キャラクタの階層構造

現実の舌は、左右にはしる横舌筋、前後にはしるじゅう舌筋、上下にはしる垂直舌筋といった厚い筋肉でできていて、自由に動かすことができる [20]。舌の表面もざらざらした特有のテクスチャをもつ。しかし、本稿ではリアリスティックな舌のモデリングを目的としていないため、四角形ポリゴンを変形して、細分化することでモデリングを行う。

\*LightWave3D は以下の各社の商標または登録商標である。  
開発元：newtek 社  
日本総販売元：株式会社ディ・ストーム

## 4.2 アニメーション

基本的なアニメーションは，BM 型では構造記述ファイル，TK 型では標準動作設定ファイルにおいて設定する．これらのファイルで設定できる項目は，座る，立つ，歩く，お辞儀をする，左右を向くの5つである．標準動作以外のアニメーションは，ユーザが個々に設定したものを使用できる．

キャラクターの表情アニメーションは，モーフィングにより行う．表情のパターンをモーフターゲットとして用意し，モーフターゲットを設定するセグメントをピックアップしてモーフターゲットを追加する．ここで，各表情と組み合わせた舌のパターンを用意して，表情の変化と同時に舌の動きを実現することを考える．本稿では，舌を制御することで新しい表情が制作する．

## 5 実験

実験環境には，DELL Dimension XPS B933r (CPU: Intel PentiumPro<sup>§</sup>1GHz, RAM: 512MB, OS: Windows 98) を用いた．キャラクター制作には，*LightWave3D<sup>TM</sup>* (Ver. 7.5) を使用し，TVML プレイヤーには，Windows 版 Ver. 1.2 を用いて実験を行った．これは Windows 2000 上で動作を確認済である．

TVML のキャラクターには，TK 型を使用した．TK 型の頭の部分に舌を組み込み，舌を考慮したモーフィングによりアニメーションを制作した．その実験結果を図 4 に示す．従来のキャラクター図 5(a) と舌を考慮したキャラクター図 5(b) を比較すると，舌を考慮したキャラクターの方が，違和感が軽減され，効果的であることがわかった．また，図 5(c) のように，舌を制御することで，“あっかんべー” というような，新しい表情が表現できることがわかった．

## 6 まとめと今後の課題

本稿では，効果的に動きの特徴を強調するようなアニメーションを NVR と定義し，TVML を用いて，ユーザが簡単に NVR アニメーション



(a) 従来のキャラクター



(b) 舌を考慮したキャラクター



(c) 新しい表情

図 4: 実験結果

<sup>§</sup>PentiumPro は Intel 社の商標である．

を制作する手法を提案した。特に、発話状態での特徴を強調するためのパーツとして、舌をとりあげた。実験を行い、従来のアニメーションとの比較を行った結果、舌を考慮することで、現実感が効果的に強調されていることがわかった。また、従来のアニメーションでは表現できない新しい表情を表現することができた。

今後は、表情ではなく発話と同期させて、口の形と舌の位置を考慮していく予定である。また、効果的に特徴を強調するためのパーツとして、涙やほほの赤みも加える。さらに、一般的なユーザが簡単にNVRアニメーションを制作するためのユーザインタフェースに発展させていきたい。

## 謝辞

本研究を進めるにあたり、TVMLに関する貴重なご意見を数多くいただきました日立製作所中央研究所の栗原 恒弥 氏に深く感謝いたします。また舌のモデリングに関する有意義なコメントをいただきましたATR研究所の倉立 尚明 氏、九州工業大学の乃万 司 教授に深く感謝いたします。

## 参考文献

- [1] B. Gooch, A. Gooch, *Non-Photorealistic Rendering*, A K Peters, 2001.
- [2] <http://www.koshi-lab.sccs.chukyo-u.ac.jp/research/picasso/>
- [3] <http://www.strl.nhk.or.jp/TVML/>
- [4] 揚, 「バーチャル空間における顔表情の動的合成に関する研究」, 京都大学大学院 エネルギー科学研究科 エネルギー社会環境科学専攻修士論文, 2000年7月
- [5] T. Ezzat, "Example-Based Analysis and Synthesis for Images of Human Faces," *MIT EECS Masters Thesis*, Feb. 1996.
- [6] T. Ezzat, T. Poggio, "Facial Analysis and Synthesis Using Image-based Models," *Proceedings of the Workshop on the Algorithmic Foundations of Robotics*, Aug. 1996.
- [7] T. Ezzat, T. Poggio, "Facial Analysis and Synthesis Using Image-based Models," *Proceedings of the Second International Conference on Automatic Face and Gesture Recognition*, Oct. 1996.
- [8] T. Ezzat, T. Poggio, "Videorealistic Talking Faces: A Morphing Approach," *Proceedings*

*of the Audiovisual Speech Processing Workshop*, Sep. 1997.

- [9] T. Ezzat, T. Poggio, "MikeTalk: A Talking Facial Display Based on Morphing Visemes," *Proceedings of the Computer Animation Conference*, Jun. 1998.
- [10] T. Ezzat, T. Poggio, "Visual Speech Synthesis by Morphing Visemes," *MIT AI Memo No 1658/CBCL Memo No 173*, May 1999.
- [11] T. Ezzat, G. Geiger, T. Poggio, "Trainable Videorealistic Speech Animation," *In Proc. ACM SIGGRAPH 2002*, Jul. 2002.
- [12] T. Kuratate, H. Yehia, E. Vatikiotis-Bateson, "Kinematics-based synthesis of realistic talking faces," *International Conference on Auditory-Visual Speech Processing*, vol.4, no.7, pp.185-190, Dec. 1998.
- [13] 緒形, 中村, 森島, 「ワイヤフレームモデルを用いたマルチモーダル翻訳システムの構築」, *VCWS*, 2001年10月
- [14] Y.Lee, D.Terzopoulos, K.Waters., "Realistic face modeling for animation," *In Proc. ACM SIGGRAPH 1995*, pp.55-62, 1995.
- [15] K. Kahler, J. Haber, H.P. Seidel, "Geometry-Based Muscle Modeling for Facial Animation," *In Proc. GI '01*, pp.37-46, 2001.
- [16] I. Albrecht, J. Haber, H.P. Seidel, "Speech Synchronization for Physics-Based Facial Animation," *In Proc. WSCG 2002*, pp.9-16, Feb. 2002.
- [17] M. Cohen, J. Beskow, D. Massaro, "Recent developments in facial animation: An inside view," *International Conference on Auditory-Visual Speech Processing*, Dec. 1998.
- [18] D. Massaro, M. Cohen, J. Beskow, "From Theory to Practice: Rewards and Challenges," *Proceedings of the International Conference of Phonetic Sciences*, Aug. 1999.
- [19] <http://www.dstorm.co.jp>
- [20] <http://plaza3.mbn.or.jp/iso/structure.htm>