

音声に合う口の動きを実現した吹替え映像を容易に作成するソフトウェア —Mouth Motion Generator 音から創るアニメ—

1. 背景

近年、日本のアニメ文化の発展は目覚ましいものがあり、大衆的なコンテンツとなるまでに至った。また、その勢いは海外にまで広がっており、日本を代表する文化の一つとなっている。それに伴って、キャラクタの動きをより高品質に制作する傾向が一層強まっている。キャラクタの動きの中でも特に、口の動きと声優の音声とが同期しているか否かに対して視聴者は敏感である。そのため、両者を同期させることはアニメ作品の品質を大きく左右する要素である。しかしながら、一般にアニメキャラクタの口の動きを生成することは熟練のアニメータの感覚的な技術を基に、試行錯誤をしながら作成することから、非常に難しい作業となっている。実際、商業向けのアニメ作品の 1 話分(放送時間:約 30 分)の制作過程において、修正すべき箇所は約 200 箇所程度存在し、その内の約 30%は「ロパク」関係の修正となっている。この修正によって、更なる制作時間と制作コストが必要となっているのが現状である。以上をまとめると、一般にキャラクタの口の動きと声優の音声との同期は映像の質を左右する上で重要なものである一方、高品質なロパクアニメを生成するには時間・労力・コストが大きくかかることが現状のアニメ制作における大きな問題として挙げられる。

2. 目的

本プロジェクトでは音声のみから容易にロパクアニメーションを生成するシステムの開発に取り組む。本システムでは、ユーザにとっての利便性を考慮し、音声の性別を選択するだけで、高速かつ高精度にロパクアニメーションを生成することを目指す。また、アニメータが制作フローで一般的に利用するツールのプラグインとして開発することによって、現行のアニメータの制作フローへの容易な導入を実現する。さらに、ユーザがインタラクティブに編集できる機能を実装することで、アニメータの制作支援を実現する。

3. 開発の内容

本プロジェクトで開発したシステム「Mouth Motion Generator」は任意の人物の音声ファイル(wav 形式, サンプル周波数 44100kHz 以上, モノラル形式推奨)から高速かつ高精度にそれに同期するアニメキャラクタの口の動きを計算することが可能である。本システムはアニメ制作で一般的に用いられる Adobe After Effects のスクリプトとして実装されており、現行のアニメ制作フローへの導入が容易となっている。図 1 に本システムの GUI を示す。本 GUI ではユーザは使用する音声は男性のものか女性のものかを選択し、アニメーションを出力したいコンポジション及び使用したいキーフレーム画像シーケンス(口の開閉度合ごとにユーザが用意した画像, 例:開口, 中口, 閉口のイラスト)をドロップダウンリストより選択して実行をするだけで高速にアニメーションを生成することが可能である。本システムの使用画面を図 2 に示す。

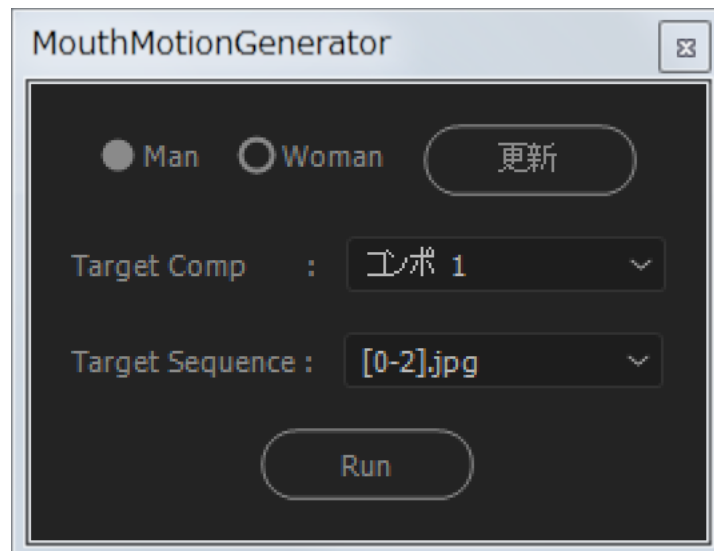


図 1 GUI

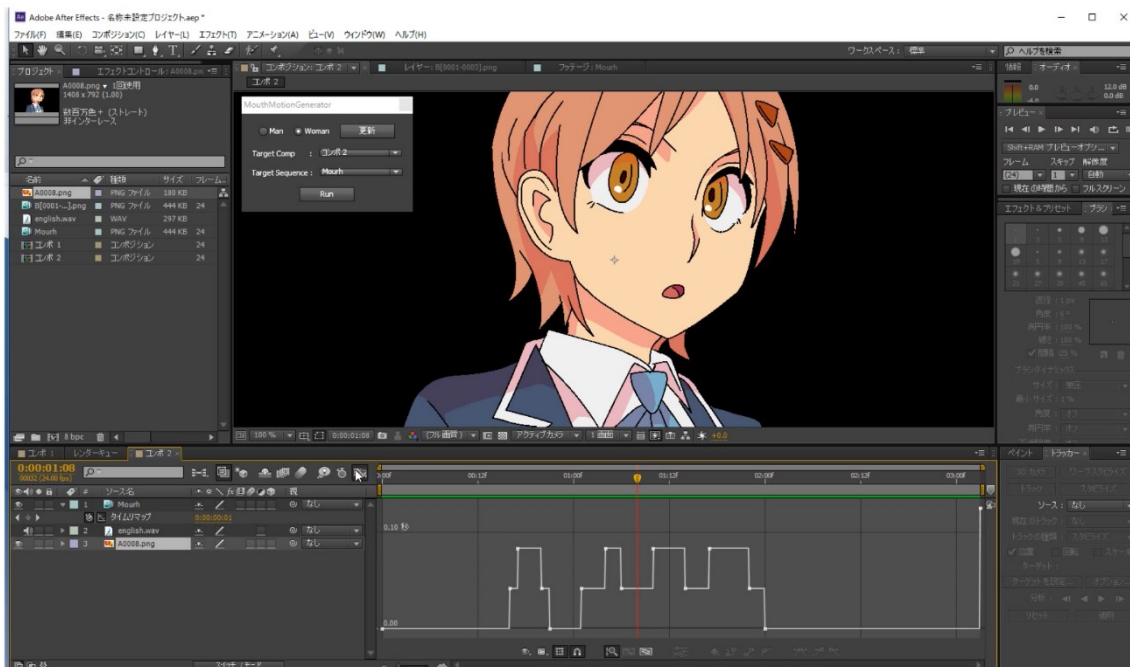


図 2 システム使用画面

本システムの特徴として以下の点が挙げられる。

- アニメらしさの実現

実写人物の口の動き情報を単に削減し、アニメキャラクタの口の動きに量子化した場合には、不自然な口の動きをすることが確認される。これはアニメ作品におけるキャラクタの口の動きのように、限られた枚数の画像から如何に自然な動きを実現するかといった“アニメらしい”動き情報が存在するためである。本システムはこの点に初めて着目し、音声のみから数値的に“アニメらしい”口の動きを計算することが可能となっている。

- 多言語への適用可能性

本システムは言語に依存せずにアニメキャラクタの口の動きを推定することが可能である(例:日本語, 英語, 中国語など). そのため, 一次制作における制作支援はもちろんのこと, 吹替え作品の制作といった二次制作支援も実現することが可能である.

- タイムリマップ形式の出力

本システムではキャラクタの口の動きを計算した後に, アニメータが容易にかつインタラクティブに編集を行えるように「タイムリマップ形式」で出力されるように実装した(図3). タイムリマップ形式とは口の動きの遷移を 0, 1, 2 の値に基づいてグラフ的に表示したものであり, 口の動きの遷移がアニメータに理解しやすい形で可視化されている. また, タイムリマップ形式にはキーフレームとなる点が打たれており, これをマウス操作することでユーザは容易に口の動きを編集することが可能である. そのため, 本システムによって自動で生成された口の動きの遷移に基づいて部分的に開閉値を強調したり, 開口のまま発話シーンを終わらせたりするといった編集も容易であり, アニメータの制作支援を実現できる.

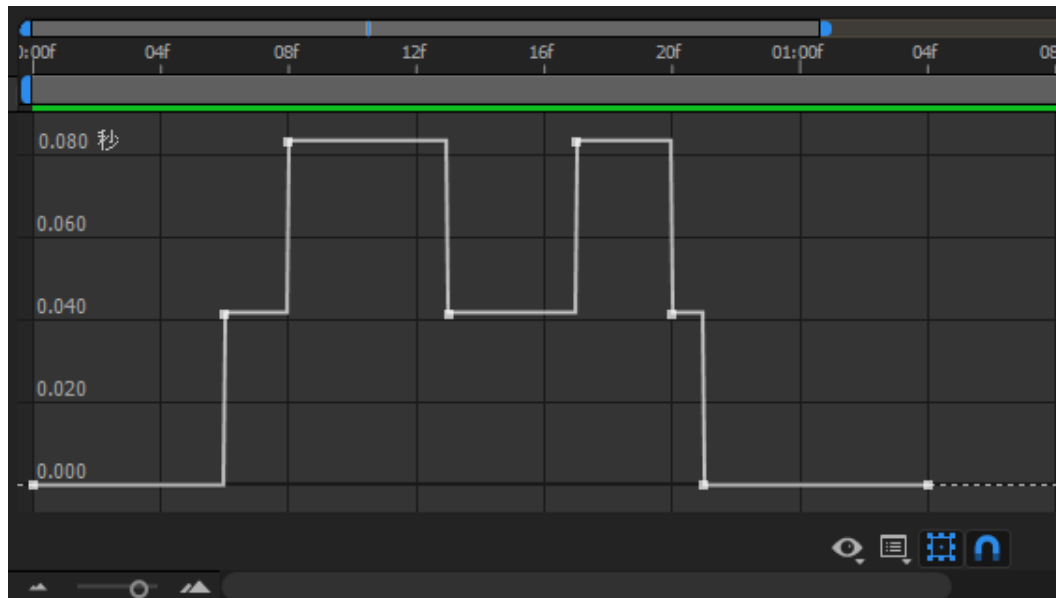


図 3 タイムリマップ形式

4. 従来の技術(または機能)との相違

- 3次元キャラクタの口の動きの制御

これまでに音声情報を用いて, 3次元キャラクタの口の動きを制御する方法が考案されてきた. しかし, これらの手法では2点問題がある. 1点目は音声認識ベースの処理を行うことであり, 言語ごとに認識モデルを構築する必要があるため, 手間がかかるという問題である. 2点目は3次元キャラクタを対象としているため, そのまま日本におけるような2次元キャラクタの口の動きの生成に適用できないという点が挙げられる.

- 2次元キャラクタの口の動きの制御

3次元キャラクタの場合と同様に音声認識ベースで2次元キャラクタの口の動きを生

成するツールが存在する。しかし、これは時間軸方向を考慮しておらず、認識した母音ごとにキーフレームを割り当てているため、“アニメらしい”動きを表現することはできない。

こういった背景からこれまでアニメータは自身の手作業によるトライ＆エラーを繰り返すことで、アニメキャラクタの口パクアニメーションを生成してきた。実際、プロジェクト期間を通して自主制作や商業アニメ制作に携わっているアニメータに取材をしたところ、これまでに口パク自動生成ツールを使ったことがあるものはいなかった。

一方で、本システムは“アニメらしい”口の動きに着目し、初めて音声のみから言語に依らずに数値的に口の動きを計算することを実現した。また、Adobe After Effects といった現行の制作フローに沿ったツール上で動作し、タイムリマップ形式を導入した点でアニメータフレンドリーなシステムの開発を実現した。

5. 期待される効果

以下の領域で大きな効果が期待される。

- アニメ業界

本システムはアニメータフレンドリーに口パクアニメーションの制作を支援することができる。そのため、従来の制作において必要となっていた労力・時間・コストを大きく削減できることはもちろんのこと、アニメータの創作力を抽出することができる。これはつまり、自動計算された口パクの遷移に対して、「ここはもっと口を開けさせたい」といったように編集を加えることで、アニメータの表現力を高める（演出考案を支援する）ことができるということである。また、経験の浅いアニメータに対して自然な口の動きを提示することで、教育手段としても活用が期待される。

- ゲーム業界

近年ではソーシャルゲーム等でキャラクタの静止画が置かれている画面上で音声の流れるといったシーンが多く見受けられる。これは多数のキャラクタと多数の音声が存在するため、口パクアニメーションを生成することが時間・労力・コストの面で不可能であるためである。これに対して本システムを適用して口の動きを自動生成することで、キャラクタの生き生きとした表現を実現することができ、ゲーム業界における一層の発展を支援することができる。

6. 普及（または活用）の見通し

今後はより多くのアニメータに使用してもらい、得られたフィードバックをその都度反映して開発を進めていきたい。また、プロのアニメータと密に連携してアニメータ向けの製品化を進めていく。

7. クリエータ名（所属）

古川 翔一（早稲田大学大学院）