

科学研究費助成事業 研究成果報告書

平成 27 年 9 月 27 日現在

機関番号：37115

研究種目：若手研究(B)

研究期間：2011～2014

課題番号：23700260

研究課題名(和文)学習者の顔形状に合わせた発音練習のための口唇動作CGアニメーション自動生成

研究課題名(英文)Automatic Generation of CG Lip Movements Based on Individuals Face

研究代表者

河野 央 (Kono, Hiroshi)

久留米工業大学・工学部・教授

研究者番号：60437746

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：本研究は、コンピュータグラフィックスを応用し、聴覚障がい児のための口唇動作学習システムにおける口唇動作CGアニメーション教材の自動作成手法を開発した。本研究の結果、学習者の顔の違いに影響されずにモデリングとアニメーションの基礎部分について自動作成することが可能となった。モデリングについては、シュリンクラップ方式により、学習者の顔形状を作成する時間を短縮した。アニメーションについては、母音の口唇動作を定義するために、複数の話者モデルの口唇動作の基準点を計測し、それらの変化を2次元数理モデルとして表現した。またその数理モデルを3D-CGに適用することで、自動的にアニメーションを生成した。

研究成果の概要(英文)：This study develops the automatic generation method of lip movements CG animation for teaching materials in the learning system of the hearing impaired children. The results shows that automatically generate the face and lip movements is possible, without affecting by the differences between the learner's face. For modeling, we reduce the time to create a learner's face shape by the shrink-wrap system. For animations, in order to define the lip movements in vowel, we measure the reference points of the lip movements of some of speaker model and express their changes as a two-dimensional mathematical model. In addition, mathematical model is unaffected the difference of the face shape of each learner, because it is relatively expressed as a reference to the site of the face. Finally, mathematical model by applying the mathematical model to the 3D-CG, generate the lip movements animation automatically.

研究分野：CGによる視覚情報デザイン

キーワード：コンピュータグラフィックス 口唇動作 数理モデル アニメーション

1. 研究開始当初の背景

聴覚障がい児の発音練習はお手本となる先生の口の動きを真似て、口唇動作を矯正することにより発音を改善する。また、幼児も大人の口の動きを見真似し発音を学ぶ。コンピュータグラフィックスを利用した聴覚障害児用見真似発音練習システムの開発(平成20年度科学研究費補助金(基盤研究(C)研究代表:小田まりこ)の見真似発音練習システムを通じた成果により、CGを用いた口唇動作アニメーションが、理想的な口唇動作を表現でき、そのアニメーションを見ながら見真似発音することにより発音や口唇動作を矯正することに非常に効果があることが分かった。その中で、学習者の顔にそっくりなCGモデルをお手本のアニメーションに用いることが、更なる学習意欲の向上や口唇動作の矯正に大きく影響することも分かり、実際に学習者、特に子供のニーズが高いことも分かっている。しかしながら、個々の学習者の顔の特徴に合わせて、お手本となる口唇動作のCGアニメーション教材をソフトウェアを用いて手入力で作成する方法は、作成者の技量に制作時間や質が依存してしまう問題があった。そこで、学習者に合わせた口唇動作モデルを3DCGで作成し、それをアニメーション化する工程を自動化することで、従来の手作業の問題点を解消できると考え、学習者一人一人に合わせた口唇動作教材を自動作成することで、この問題を解決することにした。

2. 研究の目的

本研究では、コンピュータグラフィックス(CG)を応用し、個人対応型発音練習システムのための口唇動作お手本CGアニメーションを自動作成する方法を開発する。

発音練習するためのCGアニメーションによる教材を自動的に作成することで、従来のソフトウェアを利用した手入力による作成で生じていた制作時間の長期化や質のばらつきの問題を解消する。

3. 研究の方法

自動化の対象となる工程は、モデリングとアニメーションの2つに分類される。

モデリングでは、3Dスキャナを用いて学習者の顔形状を取得し、それを基にアニメーションに適した顔形状を自動作成する。

アニメーションでは、始めに、正しい発音と同定された話者モデルの実写の口唇動作を画像解析により計測する。そして、この計測値から、話者モデルらの共通性を抽出し、数理モデル化する。この際に、数理モデルで動かされる口唇部位の動作量は個人によって異なるため、学習者の顔の部位を基準とすることで、例えば口が大きくても小さくてもそのような個人差まで含めて対応できるモデル化を行う。

次に、この数理モデルを、モデリングで作成した顔形状に適用し、パラメータ入力によ

り口唇動作のアニメーションを自動作成する。

最後に、個々の学習者に使用してもらい、自動作成されたCGアニメーションの検証を行う。

4. 研究成果

(1) 学習者の顔形状の生成

学習者の顔形状は、口唇動作のアニメーション設定時において、顔形状が適切に変形する必要もある。そのため、作成した顔形状はアニメーションに適した構造を持っていないなければならない。

はじめに、学習者の顔をデジタイザで取得し、アニメーションに対応できるトポロジーを持った顔マスクを用意する。(図1)。

次に、顔マスクをデジタイザデータに密着させる(図2)。この時、顔マスクの頂点とデジタイザデータの頂点との距離が最短となる場所を求めて、密着させる。この方法はシュリンクラップ方式と呼ばれる。

最後に、密着させた顔マスクを取り出すと、適切な構造を持った学習者の顔形状が取得できる(図3)。



図1: デジタイザデータと顔マスク

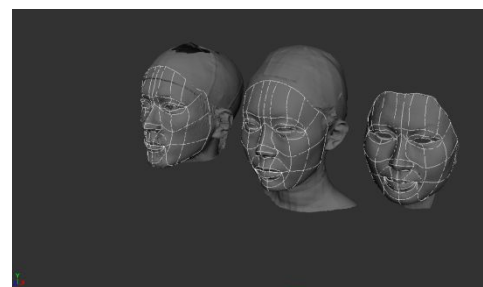


図2: 顔マスクをデジタイザデータに自動密着



図3: 取り外した顔マスク

(2)口唇動作の計測

本研究では、個々の学習者ごとにお手本となる口唇動作 CG アニメーションを自動作成する。母音発音時に、口をどのようにどれくらい動かせばよいのかという点については定義されていない。そのため、数理モデルで定義し、それを CG に適用する。

口唇動作の数理モデル作成において考慮すべき点は、口唇動作を行う顔形状モデルは個々の学習者別に異なり、それぞれが特徴を持った形状ということである。様々な形状に対応できる数理モデルでは、数値を絶対値で扱わずに、学習者の顔の部位を基準とし、それに対する相対値で扱う必要がある。

本研究では(1)母音の口唇動作の計測(2)計測結果の分析に基づく数理モデル化(3)数理モデルに基づいた 3DCG アニメーションの自動生成、の3つの過程で行う。始めに、各母音「あいうえお」の口唇動作が正同定された話者の画像を用意する。正同定とは、第3者に口唇動作を正しく同定される、例えば、話者による「あ」の動作が第3者によって「あ」と認識される割合が100%のことをいう。ここでは、口唇動作「あ」を用いて、説明する。

(3)母音「あ」における分析

図4は、正同定モデルの例である。同じ共通の動作であっても、顔の形状はそれぞれ異なるために、動作量には個人差がある。従って、正同定モデルの計測を通して、共通性を数量的に抽出する必要がある。



図4：口唇動作「あ」の正同定モデルの例

計測にあたっては、美術解剖学で用いられる口唇の7つの計測基準点(図5および図6)を基本とし、母音毎に計測を行う。

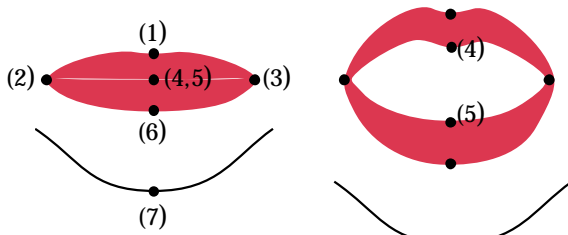


図5:通常時

図6:最大変化時

これらの計測基準を基に、8人の正同定モデルの口唇動作「あ」を計測した結果が表1である。この実測値は、個人差も含まれ、絶対値として見ると共通性を把握することができないが、口唇動作をする前の通常時を基準とし、口唇動作における最大の変化時を相対的に表せば、個人の顔の部位の違いは無視することができるかと仮定した。そこで、通常時と最大変化時を相対値で表現したところ、表2のように動作の共通性が数値的に把握す

ることが可能となった。このように、他の母音についても、通常時の距離や顔の部位の大きさを基準として相対表現を行うことで、通常時に対してどのように変化するのか、個人差に影響を受けないような共通の定義を行った。

表1:話者モデルの口唇動作「あ」の実測値

話者モデル	計測時間	口の幅	上の唇の高さ	下の唇の高さ	口の中心の長さ	口の中心の幅	口の中心の高さ	口の中心の長さ	口の中心の高さ
A	通常時	80	7.8	9.5	20				69.9
	最大変化時	80	11.6	13.8	70	42.3	20.6	23.1	85.4
C	通常時	58.8	7.4	6.7	15.9				55.7
	最大変化時	57.2	7.1	11.3	39.9	20.1	9.9	11.8	58.9
D	通常時	68.2	10.2	8.8	22.2				66.0
	最大変化時	72.1	9.5	12.7	51.9	26.5	16.1	10.2	74.4
E	通常時	61.9	9.2	9.9	21.9				54.0
	最大変化時	66.6	8.8	14.8	45.9	21.9	7.8	14.1	62.4
G	通常時	67	14.5	14.1	30.7				67.7
	最大変化時	69	14.5	14.8	56.8	27.2	16.2	11.5	70.9
H	通常時	60	11.6	15.2	27.2				62.8
	最大変化時	60.4	9.9	14.1	54.3	27.2	14.3	13.8	68.1
J	通常時	61.5	10.6	10.6	22.6				65.6
	最大変化時	61.1	11.3	12	46.2	23.6	8.5	14.6	69.5
K	通常時	73.7	12	12	24.3				68.4
	最大変化時	65.1	10.6	14.1	47.6	23.3	11.8	11.3	74.8

表2: 相対表現による変化量

モデル	横幅の倍率	縦の倍率	顎の倍率
A	1.000	3.500	1.221
C	0.972	2.509	1.057
D	1.057	2.337	1.127
E	1.075	2.095	1.155
G	1.029	1.850	1.047
H	1.006	1.996	1.084
J	0.993	2.044	1.059
K	0.883	1.958	1.093
平均	1.002	2.286	1.105
中央値	1.003	2.070	1.088

変化量を抽出する際に、データのばらつきを考慮し平均値もしくは中央値を選択している。

(4)計測結果に基づく数理モデル化

次に、計測結果に基づいて、口唇の計測基準点の変化を数理モデル化する。例えば、図3の上唇点(1)について、通常時の座標を(X1, Y1)、最大変化時を(X'1, Y'1)とする。他の計測点についても同様に座標値を設定すると、表2の倍率を使って、表3のように最大変化時の計測基準点の位置を表現できる。なお、表3のMHは図3(1)と図3(6)の距離、すなわち通常時の口の高さ、JHは図3(2)と図3(3)を結ぶ直線と(7)の距離を示す。

表3: 計測点の数式モデルの例

計測点	横座標の式	縦座標の式
右の口角点(3)	$X'3=X3$	$Y'3=Y3$
左の口角点(2)	$X'2=X2$	$Y'2=Y2$
上の口点(4)	$X'4=X4$	$Y'4=Y4+0.5*MH$
下の口点(5)	$X'5=X5$	$Y'5=Y5-0.5*MH$
上唇点(1)	$X'1=X1$	$Y'1=Y1+0.5*MH$
下唇点(6)	$X'6=X6$	$Y'6=Y6-0.5*MH$
オトガイ点(7)	$X'7=X7$	$Y'7=Y7-0.1*JH$

(5)数理モデルによる 3DCG アニメーションの自動生成

はじめに、図7のような口唇動作の基本形状となる顔のパラメトリックデータを用意する。このデータは、先の4.で得られた顔マスクと同じものである。

次に、このパラメトリックデータの各制御点を、計測点に合わせてグループ化し、そのグループを口唇動作の数理モデルに従って、縦横の方向に対して自動的にアニメーションさせる。他の母音

「い・う・え・お」についても、同様に計測基準の座標値を数式化で定義した後、6.で得られた数理モデルを 3DCG モデルに適用する。本研究では Autodesk 社 Maya 及びスクリプト言語 MEL にて実装を行った。

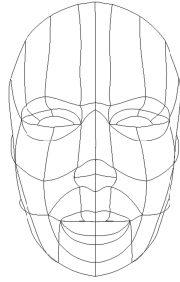


図7: 顔のデータ

(6)自動生成結果

(5)により自動作成されたアニメーション結果を表4に示す。従来は手入力アニメーションを行っていたため、制作者の経験度に応じた時間がかかっていたが、数理モデルを用いることで口唇動作を数式として記述できるため、アニメーションの設定は数秒で完了する。大幅な制作時間の短縮が可能となった。

表4: 自動生成されたアニメーション

通常時	
あ	
い	

う	
え	
お	

また、他の学習者でも同じように適用できるかどうかを確認するため、異なる顔形状においても同じ数理モデルを適用してアニメーションさせた(表5)。この結果から、顔形状の個人差に左右されずに口唇動作の自動アニメーションの作成が可能であると考えられる。

表5: 複数モデルにおける数理モデル適用例

口唇動作	あ	い	う	え	お
IKEMATSU モデル					
SHIMURA モデル					

(7)検証

(6)で得られた CG アニメーションについては、お手本教材としての精度、つまり、自動作成された口唇動作アニメーションが実際に正しく第3者に認識されるのかについて検証する必要がある。そこで、生成した「あ・い・う・え・お」の CG アニメーションを各3回表示することをランダムに行い、被験者10人に同定させた。その実験結果において、最も高かった正同定率を表6に示す。

表 6：生成したアニメーションの正同定率

	モデル A	モデル B	モデル C
あ	100%	80%	100%
い	50%	60%	40%
う	60%	90%	80%
え	30%	60%	20%
お	70%	60%	70%

正同定率に注目すると、母音「あ」は 100% も出ているが、「い」や「え」については、低い結果となった。これは、今回の研究対象が口唇表面の形状のみであったが、実際には歯の見え方が大きく影響していたためだと考えられる。そのため、口唇だけではなく歯の見え方をモデル化し、それを CG アニメーション教材に取り入れる必要があると考えられる。

(8)まとめ

本研究では、コンピュータグラフィックスを応用し、聴覚障がい児のための口唇動作学習システムにおける口唇動作 CG アニメーション教材の自動作成を目指した。

モデリングについては、シュリンクラップ方式により、学習者の顔形状を作成する時間を大幅に短くした。アニメーションについては、母音の口唇動作を定義するために、複数の話者モデルの口唇動作を計測し、それらから共通性を抽出し、数理モデルとして表現した。またその数理モデルを 3DCG に適用することで、アニメーションを自動的に作成した。

最後に、自動作成されたアニメーションの評価および検証を行った。その結果、正同定率が 100%に到達せず、ユーザーアンケートでは口唇だけではなく、歯の見え方も要因として大きいことが挙げられ、これらについても数理モデル化を行う必要があることが分かった。

従来、口唇動作の教材は、学習者ごとに手作業で作成していたが、本研究の結果、モデリングとアニメーション自動作成の基礎部分については確立できたと考える。しかしながら、更に口唇動作の再現性を高める必要があり、正同定に影響する他の要素のモデル化が今後の課題である。

5 . 主な発表論文等

〔雑誌論文〕(計 1 件)

知的障害児のための CG アニメーションを用いた教育支援ソフトウェアの開発
田口浩太郎, 小田まり子, 河野央, 小田誠雄, 新井康平, 教育システム情報学会ジャーナル, 31 巻 1 号, 査読有, 2014, 48-56.

〔学会発表〕(計 4 件)

田口浩太郎, 小田まり子, 河野央, 小田誠雄, 知的障害を持つ肢体不自由児のための教育支援ソフトウェアと入力機器の開発, 電子情報通信学会福祉情報工学研究会, 2014

数理モデルによる口唇動作 3DCG の生成手法, 河野央, 小田まり子, ADADA Japan 学術大会発表概要集 B4-7, 2014.

口唇動作の数理モデルによる 3DCG アニメーションの自動生成

河野央, 小田まり子, 秋山侑也, 日本デザイン学会誌第 60 回研究発表大会概要集 CD-ROM, 2013.

シュリンクラップモデリングによる 3 次元顔形状の生成

秋山侑也, 河野央, 小田まり子, 日本デザイン学会誌第 59 回研究発表大会概要集 CD-ROM, 2012.

6 . 研究組織

(1)研究代表者

河野 央 (KONO, Hiroshi)

久留米工業大学・工学部・情報ネットワーク工学科・教授

研究者番号：60437746