

科学研究費助成事業 研究成果報告書

平成 30 年 5 月 25 日現在

機関番号：37115

研究種目：若手研究(B)

研究期間：2014～2017

課題番号：26730179

研究課題名(和文)3次元数理モデルによる口唇動作CG教材制作システムの開発

研究課題名(英文)Development of CG Teaching Material for Lip Motion by Three-Dimensional Mathematical Model

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) 2,700,000円

研究成果の概要(和文)：本研究は、口唇動作の3次元数理モデルをCGへ応用し、聴覚障がい児のための口唇動作学習システムにおける教材である口唇動作CGアニメーションを自動生成した。口唇動作は3Dスキャナを用いて口唇動作を時系列的に連続な形状として記録・計測し、口唇動作における計測点の移動量を抽出した。さらに、移動量の共通性を抽出し、奥行き方向の動作を数理モデルとして定義した。この数理モデルは、個人の顔形状の違いにも対応できるように顔の部位を基準として相対表現した。また、その数理モデルを3DCGに適用し自動生成されたアニメーションを評価検証した。その結果、最大正同定率は前研究の数理モデルと比較して最大で62%向上した。

研究成果の概要(英文)：In this research, we applied a three - dimensional mathematical model of lip movement to CG and automatically generated lip motion CG animation which is a teaching material in a lip motion learning system for hearing impaired children. Lip motion was recorded and measured as a continuous shape in time series of lip movement using a 3D scanner and the amount of movement of measurement points in lip movement was extracted. Furthermore, we extracted commonality of the movement amounts, and defined the movement in the depth direction as a mathematical model. This mathematical model is expressed as a relative reference to the facial part which as to be able to cope with differences in individual face shapes. In addition, we applied the mathematical model to 3DCG and evaluated and verified the automatically generated animation. As a result, the maximum correctly identified rate improved by up to 62% compared with the mathematical model of the previous study.

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。著者らは、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効果を得た。その中で、学習者の顔にそっくりなCGモデルをお手本教材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、そのモデルのアニメーションを自動生成することで、従来の手作業の問題点を解消できると考えた。

筆者は、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(平成23年度科学研究費補助金(若手研究(B)研究代表:河野央)において、口唇動作の2次元数理モデルを定義した。このモデルにより、様々な3DCG顔形状の口唇動作を自動生成することが可能になったが、このモデルは縦横方向の移動に着目したものであり、奥行き方向の変化は取り扱っていない。そのため、本研究では奥行き方向の口唇動作についてもモデル化し、3次元数理モデルの構築を試みた。

2. 研究の目的

本研究では、口唇動作の3次元数理モデルを定義し、そのモデルをコンピュータグラフィックス(CG)に応用し、個人対応型発音練習システムのための口唇動作CGアニメーションを自動作成する方法を開発する。

3. 研究の方法

はじめに、1秒間に複数回対象形状を記録できる3次元スキャナを用いて、複数の話者モデルの口唇動作における時系列的な変化を形状として記録する。

次に、記録された形状データを元に解剖学的な計測点の位置の時間的変化(移動量)を計測する。また、計測点の移動量を顔の部位を基準とした相対値で表すことができるか検討する。これに基づき、3次元数理モデルを定義する。

最後に、3次元数理モデルを用いて口唇動作CGアニメーションを自動生成し、正同定率の評価検証を行う。

4. 研究成果

(1)口唇動作の3次元形状採取

本研究では、Artec社製3Dスキャナを用いて、口唇動作を15fps(frame per second)でポリゴンデータとして計測した(図1)。解像度はおよそ0.5mmである。

口唇動作をポリゴン形状として記録することで、3次元空間内における顔形状の計測が可能となり、計測点の変化について、計測のし易さも向上し精度を高めて計測することが可能となる。ここでは、2名の話者モデルを対象に口唇動作の3次元形状を採取した。

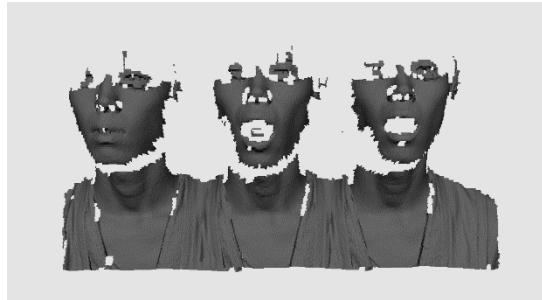


図1: 口唇動作の形状データの例(抜粋)

(2)口唇動作の計測

次に、(1)で得られた口唇動作の形状をAutodesk社Mayaに読み込み3次元空間内で計測した。時間経過と共に移動する5つの解剖学的な計測点(図2-1および2-2)を追跡記録し、フレーム毎の座標値(x, y, z)を抽出した(表1)。

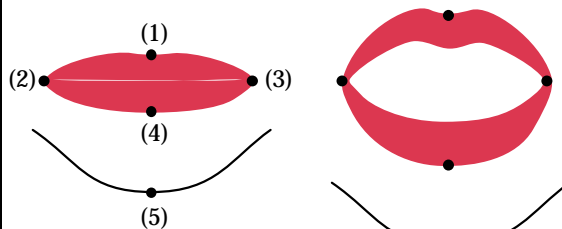


図2-1: 通常時

図2-2: 動作時

表1: 計測値の例(抜粋)[mm]

	1frame	2frame	3frame	4frame	5frame
x	-4.022	-4.405	-4.788	-4.651	-4.513
y	85.473	84.149	82.364	81.887	81.749
z	72.130	71.467	70.542	70.156	70.033

(3)口唇動作の分析

(2)の計測データ(実測値)は絶対量である。同じ共通の動作、例えば発音「あ」であっても、学習者の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、実測値を相対的に扱う必要がある。前研究では、顔の部位、例えば口の幅等を基準として移動量を相対量として表現した。口唇動作「う」のデータを用いて説明すると、図4は

口角点右(2)の計測結果から、奥行き方向 z 値の通常時と最大変化時の座標値および口の幅 MW に対する相対量を抜粋したものである。

表 2: 相対量での表現例
(母音「う」における右の口角点)

	話者モデルA	話者モデルB
最大変化時の平均	43.34863	72.89602
通常時の平均	31.84287	62.82104
移動量	11.50576	10.07498
口の横幅 MW	45.25478	43.87395
移動量 (MW 比)	0.25424	0.22963

表 2 の口唇動作時における移動量をみると話者モデル間の差は約 9mm であり、同じであるとは言えないが、それぞれの話者モデルの口幅を基準に移動量を表現すると約 0.25 倍および 0.23 倍となり、平均化してもその差は 1mm 未満であるので考慮しなくてもよい誤差となると考えた。

このように個人差があってもモデル化できるように、顔の各部位を用いて計測点の移動量を相対量に変換し、共通の動きとして数値モデル化することを行った。

(4) 口唇動作の 3 次元数値モデル

(3) の計測データを基に各計測点の奥行き方向の変化についての特徴をまとめたものが表 3 である。さらに、顔の部位に対する相対量も基に、表 4-1 および 4-2 のような奥行き方向 z の数式として表現した。なお、顔モデルから見て、奥が $+z$ 方向であり、空欄部分は共通動作が抽出できず数式表現ができない箇所である。

表 4-1 および 4-2 における MW は口の横幅 (計測点(2)と(3)の距離) MH は口の高さ (計測点(1)と(4)の距離) JH はオトガイの高さ (計測点(2)と(3)を結ぶ直線と計測点(5)の距離) を表す。

表 3: 奥行き方向の変化の特徴

	あ	い	う	え	お
上唇(1)	N/A	z 変化なし	z 手前方向	z 変化なし	z 手前方向
口角点右(2)	z 変化なし	z 奥方向	z 手前方向	z 奥方向	z 手前方向
口角点左(3)	z 変化なし	z 奥方向	z 手前方向	z 奥方向	z 手前方向
下唇(4)	N/A	z 奥方向	z 手前方向	z 奥方向	z 手前方向
オトガイ(5)	z 奥方向	N/A	z 手前方向	N/A	z 奥方向

表 4-1: 計測点の奥行き方向の数値モデル化 (あ・い・う)

	あ	い	う
上唇(1)			$z' = z + 0.17 * MW$
口角点右(2)		$z' = z - 0.2 * MW$	$z' = z + 0.3 * MW$
口角点左(3)		$z' = z - 0.2 * MW$	$z' = z + 0.3 * MW$
下唇(4)		$z' = z - 0.3 * MH$	$z' = z + 0.17 * MW$
オトガイ(5)	$z' = z - 0.05 * JH$		$z' = z + 0.1 * JH$

表 4-2: 計測点の奥行き方向の数値モデル化 (え・お)

	え	お
上唇(1)		$z' = z + 0.3 * MH$
口角点右(2)	$z' = z - 0.2 * MW$	
口角点左(3)		$z' = z + 0.4 * MH$
下唇(4)	$z' = z + 0.36 * MH$	
オトガイ(5)		$z' = z - 0.3 * JH$

(5) 3 次元数値モデルによる 3DCG アニメーションの自動生成

(4) で得た z 方向の数値モデルを前研究の xy 方向の 2 次元数値モデルと組み合わせ、3DCG 顔モデルに適用した。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。

3DCG 顔モデルは、3 次元数値モデルによる口唇動作アニメーションを検証するために、前研究の 3 つの顔モデルを用いた (図 3)。これらのモデルは同じトポロジーを持つ NURBS データを基にポリゴン変換して構成している。そのため、異なる顔形状でも同じ頂点数や頂点番号を有しており、同一の 3 次元数値モデルを適用することができる。

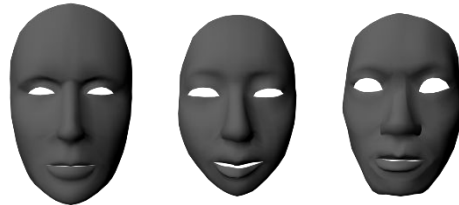
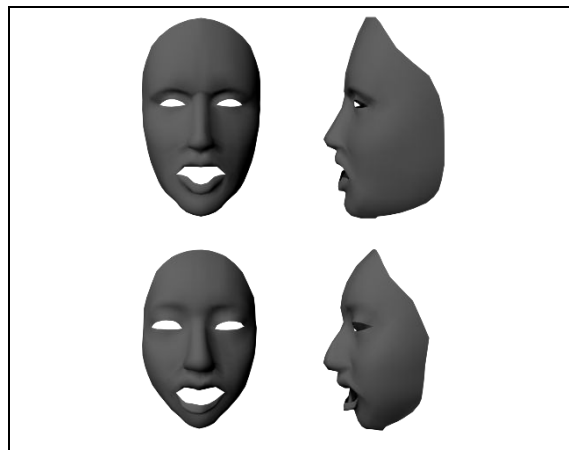
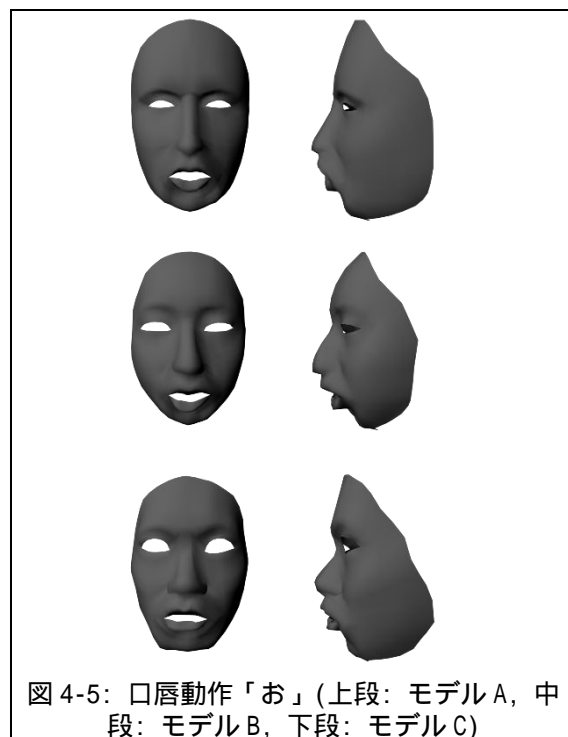
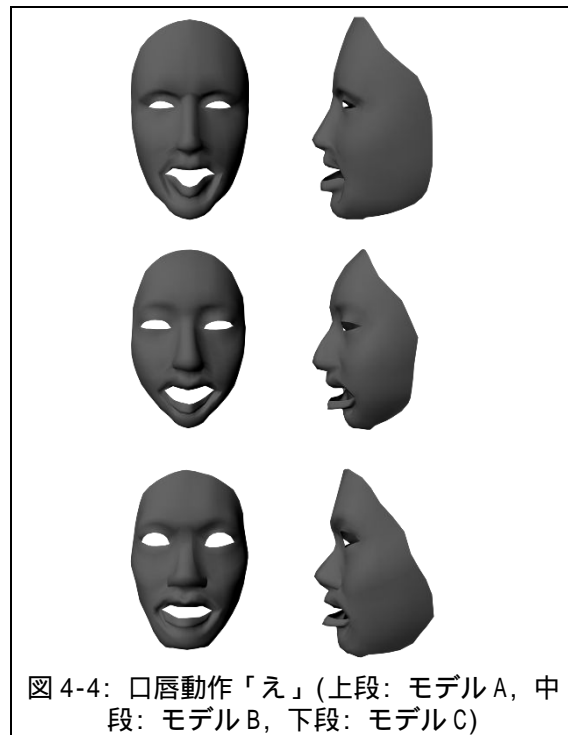
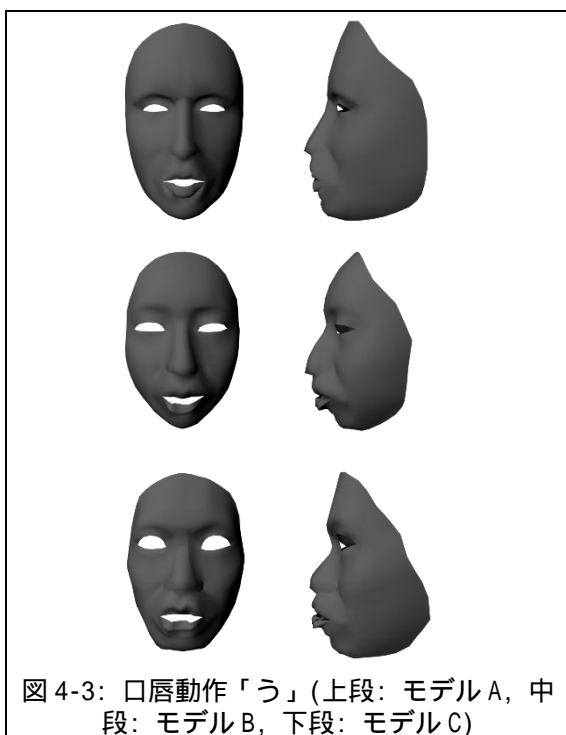
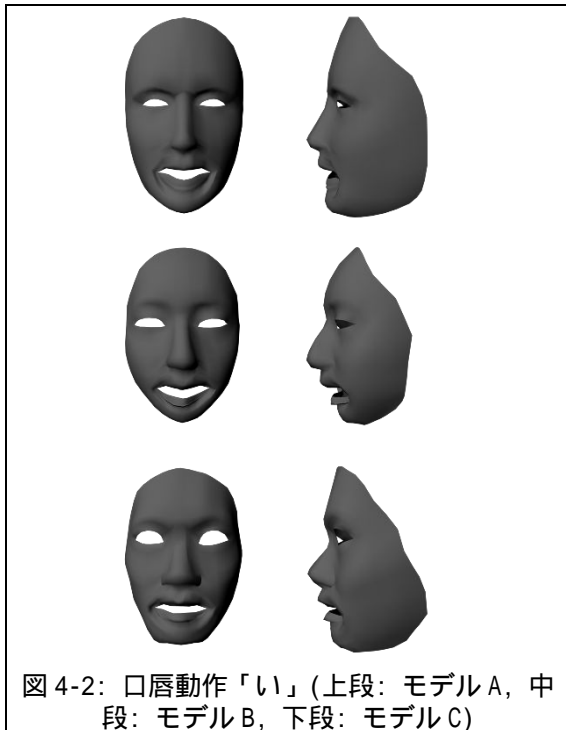
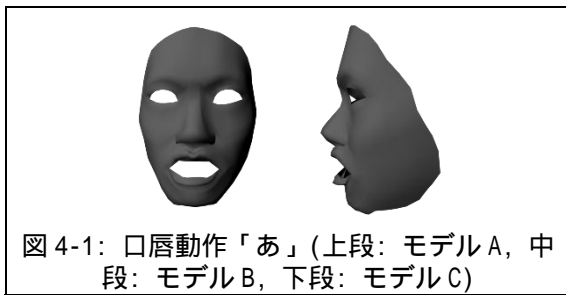


図 3: 3 次元顔形状モデル
(左:モデルA, 中:モデルB, 右:モデルC)

自動作成された発音「あ・い・う・え・お」の口唇動作アニメーション結果を以下に示す。





(6) 検証

(5)で自動生成により得られた口唇動作 CG アニメーションについて、第3者に実際に正しく認識されるのか、つまり、お手本教材としての精度について検証する必要がある。そこで、モデルA/B/Cそれぞれについて、生成した「あ・い・う・え・お」の口唇動作 CG アニメーションをランダムな順番で各3回ずつ合計15回表示し、被験者11人に同定させた。その実験結果を図5-1から5-3に示す。

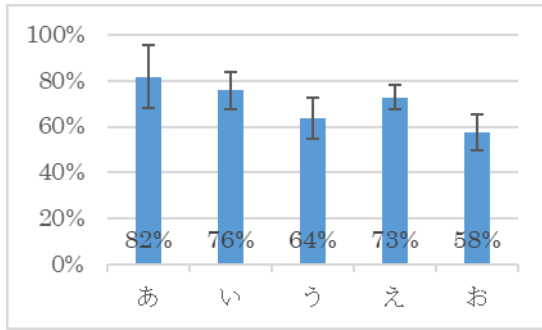


図 5-1: 平均正同定率(モデル A)

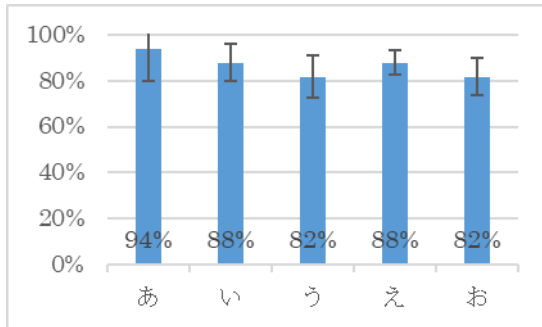


図 5-2: 平均正同定率(モデル B)

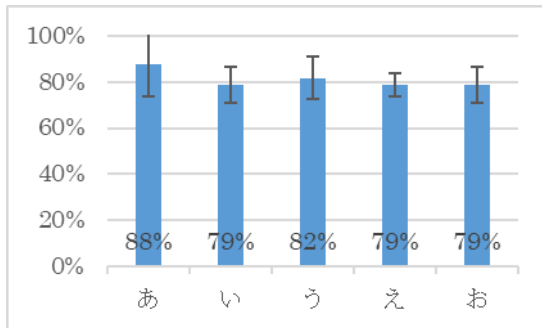


図 5-3: 平均正同定率(モデル C)

図 5-1 から 5-3 は同じ 3 次元数理モデルを異なる顔形状(モデル A/B/C)に適用し、3 回の同定における正同定の平均を表したものである。正同定とは、「あ」の口唇動作 CG アニメーションが正しく「あ」と同定されることをいう。また、標準誤差を併せて提示した。

母音「あ」については数理モデルからみても最も単純な動きであり、どのモデルも最も高い正同定率となっているが、母音「お」については、他の母音と比較して低い結果となった。本研究の 3 次元数理モデルでは個人毎の異なる顔形状に対応できない部分があるということが考えられる。

次に、前研究の 2 次元数理モデルに比べ、3 次元数理モデルの正同定率がどのように向上したかについて検証するため、2 次元及び 3 次元数理モデルにおける最大正同定率を図 6-1 から 6-3 に示す。

最大正同定率とは、3 回の同定における最も好成績なものを抽出したものである。

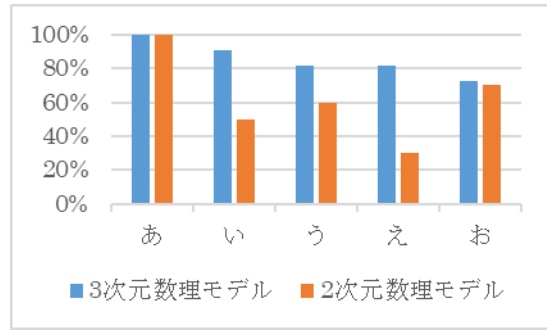


図 6-1: 最大正同定率の比較(モデル A)

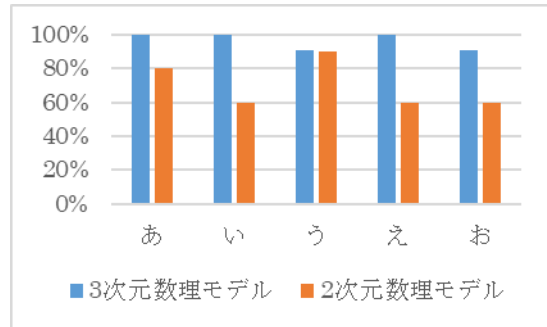


図 6-2: 最大正同定率の比較(モデル B)

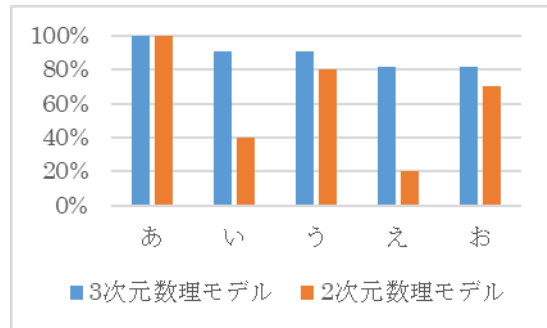


図 6-3: 最大正同定率の比較(モデル C)

2 次元数理モデルの結果と比較して、3 次元数理モデルの結果と比較して、3 次元数理モデルの最大正同定率は、すべてのモデル、すべての母音で向上した。特に、母音「い・え」については大幅な改善が見られた。例えば、モデル C では「え」の最大正同定率が 20%から 82%に向上した。これは、3 次元数理モデルを適用したことにより、奥行き動作の口唇動作が加わり、より強調した口唇動作になったことや、横方向の視点からも口唇動作を確認可能になったことで、口唇動作を判断しやすくなったことが要因であると考えられる。

(7)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 アニメーションの自動作成のための、3 次元数理モデルの構築を行った。

前研究では 2 次元数理モデルの定義までとなっており、奥行き方向の口唇動作を扱って

いなかった。そのため、本研究では、3D スキャナを用いて口唇動作を時系列的に連続な形状として採取し、計測を行うことで奥行き方向の口唇動作を数量化した。さらに、計測データから共通性を抽出し、奥行き方向の動作を数理的に定義した。数理モデルは、個人の顔形状の違いに対応させるため、顔の部位（口の横幅や口の高さ等）を基準として相対的表現とした。また、その数理モデルを 3DCG に適用することで、奥行き方向の変化を持った口唇動作 CG アニメーションを自動的に生成した。

最後に、自動生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、最大正同定率は前研究の 2 次元数理モデルと比較して最大で 62%向上する等大幅に改善した。しかしながら、平均正同定率は全てのモデルおよび母音で 100%には到達していないため、更なる改善が必要である。

今後は、モデル化するのではなく口唇動作の特徴を深層学習させ、新たな顔モデルを提示した際にそのモデルに最適な口唇動作量を生成することにより本研究の内容をさらに向上させることを目標とする。

5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕(計 2 件)

河野央、秋山侑也、小田まり子、母音の口唇動作の 2 次元数理モデルによる CG アニメーションの自動生成、久留米工業大学研究報告 No.39、2017、pp.44-54

小田まり子、小田誠雄、河野央、佐塚秀人、高橋雅人、サービスマーケティングによる地域特別支援学校のための工学的・教育的支援、教育システム情報学会研究報告 29 巻、pp.115-120

〔学会発表〕(計 1 件)

河野央、小田まり子、数理モデルによる口唇動作 3DCG の生成手法、第 1 回 ADADA JAPAN 学術大会、2014-09-05、九州大学

〔図書〕(計 0 件)

〔産業財産権〕

出願状況 (計 0 件)

取得状況 (計 0 件)

〔その他〕

なし

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