

## 三次元計測顔形状を利用した話者適応唇モデル生成

倉立尚明\* Kevin Hsu†\* Marcia Riley‡\*

\*ATR 人間情報科学研究所

†Department of Computer Science, University of Victoria, CANADA

‡College of Computing, Georgia Institute of Technology, USA

E-mail: kuratate@atr.jp,

hsuk@{uvic.ca, his.atr.co.jp}, mrily@{coc.gatech.edu, atr.jp}

顔アニメーションにおいて、実人物のリアルな発話表現を行うためには口腔内を含めた形状表現が重要となる。しかしレーザーレンジファインダなどの三次元形状装置では唇内部の構造までは計測することが不可能で、何らかのモデル表現が必要とされる。しかし、唇も顔の一部として個性を示す重要な役割を担っており、単純なモデルでは個々の人物に適応した唇形状を生成することは容易ではない。本研究では、三次元計測装置では取得できない唇の内側を含めたサーフェースモデルを、話者の唇特徴を保持したまま生成する手法を解説する。

### Creating Speaker Specific 3D Lip Models Using 3D Range Data

TAKA AKI KURATATE\* KEVIN HSU†\* MARCIA RILEY‡\*

\*ATR International, Human Information Science Laboratories

†Department of Computer Science, University of Victoria, CANADA

‡College of Computing, Georgia Institute of Technology, USA

E-mail: kuratate@atr.jp,

hsuk@{uvic.ca, his.atr.co.jp}, mrily@{coc.gatech.edu, atr.jp}

When creating realistic talking head animations of real people, accurate modeling of speech articulators is important for making the speech perceptible. A 3D laser range finder helps us acquire accurate face models from real subjects, but cannot capture articulator data such as the inner mouth for all speech postures, as they are not always visible. In this work we focus on a method for more accurate and speaker-specific lip representation. Lips are an important articulator that also express individuality. Previous methods such as simple numerical lip models focus on creating a general lip model without incorporating lip speaker variations. In this paper, we propose an adaptive numerical lip model created using 3D range data that retains the individual characteristics of a speaker's lips, including the inner part of the mouth. By adjusting spline parameters automatically to minimize the error between node points of the lip model and the raw 3D data surface, new 3D lips are created quite easily. The resulting lip model will be used in our talking head animation system to evaluate auditory-visual speech perception and to analyze our 3D face database for relevant lip features.

## 1 はじめに

近年のパーソナルコンピュータとグラフィックアクセラレータの高速化と低価格化により、映像産業におけるコンピュータグラフィックスの応用や、個人レベルでのコンピュータアニメーションの生成が盛んとなってきている。中でも人物のCGアニメーションの利用は顕著であり、フルCGによるキャラクターが実際の俳優とともに映画やTV画面中に現れることはごく普通のこととなりつつあり、さらにインターフェースとしての利用も増えつつある。このように、実際の人物と違和感なく存在したり、またより自然なインターフェースとし人物アニメーション

を生成するためには、特に発話においてはより実物の人物に近い発話にともなう自然な顔面運動を生成することが重要となる [1]。

このような自然な顔アニメーションを実人物のリアルなモデルで生成する場合、特に口腔内を含めた形状表現が重要となり、なかでも唇は常に視覚的に特徴ある動きを舌や歯以上に明確に提示するため、異和感の無い形状表現および運動表現が必要となる [2, 3]。しかしビデオ映像を用いた唇モデル生成技術や [4]、実人物のモデル取得に用いられるようなレーザーレンジファインダなどの三次元形状装置では唇内部の構造までは計測することが不可能で、何らかのモデル表現が必要とされる。また唇の形状は個人

の顔の特徴を示す役割も担っており、より自然な顔生成を目的とする場合には、唇のモデル化を行う際に単純な数学的なモデルではなく、個々の人物の特徴を反映できる必要がある。

本論文では、これまでに開発してきた唇輪郭情報にもとに口腔内を含めた唇のサーフェースモデルを自動生成するモデルに対し、レーザーレンジファインダによる三次元顔形状を利用することにより個々の人物の唇の計測形状に適応できるように改良したので、それらについて報告する。

## 2 唇モデル開発の背景

これまでに我々は、発話を知覚できる顔アニメーション生成を目標として、Cyberware 4020RGB/PS, 3030RGB/PS(Cyberware 社製)およびVIVID700(ミノルタ社製)のレーザーレンジファインダより得られた複数の三次元顔形状を用いたシステムを開発してきた [5, 6, 7]。計測された複数の顔形状は、一般化メッシュによる適合により [8]、同一のメッシュ構造へと変換されることで形状の分析やアニメーション生成へと利用することができる。これには一般化メッシュに定義されている目・眉・唇の輪郭や鼻・顎・髪のはえぎわなどの特徴線を計測顔形状においても定義し、それらをもとに Cyberware のデータ形式である円柱座標系の  $(\theta, z)$  空間中において一般化メッシュ適合させることにより実現されている。

しかし、発話の際に視覚的な発話関連情報を与える唇内部の構造まではこれらの計測装置では取得することができず、また唇領域の表面形状の変化に対して解像度が低く、さらに被検者によっては唇表面の皮膚の状態により、正確に唇形状が取得できないという問題があり、特にこのような唇形状データに基いてアニメーションを生成しても、視覚的に異和感を与えるような結果しか得られなかった [5]。これを解決するため、まず我々は次に示す基本モデルを開発した。

### 2.1 基本唇モデル

基本的な唇のモデル化を行うため、唇の外側輪郭線をもとにスプライン曲線を利用して唇断面の形状を決定し、唇表面のポリゴンを生成するモデルを構築した。より具体的には図 1 に示す以下の手順でモデルを生成する [7]。

**(a) 唇の外側輪郭線の取得:** 唇の外側輪郭線は唇表面の特性にかかわらず、三次元計測装置でも安定した結果が得られる。従ってモデル生成の基準としてこの輪郭線を利用する。我々のシステムでは、この外側輪郭線は一般化メッシュの適合前にすでに特徴線として定義されており、またアニメーションを生成す

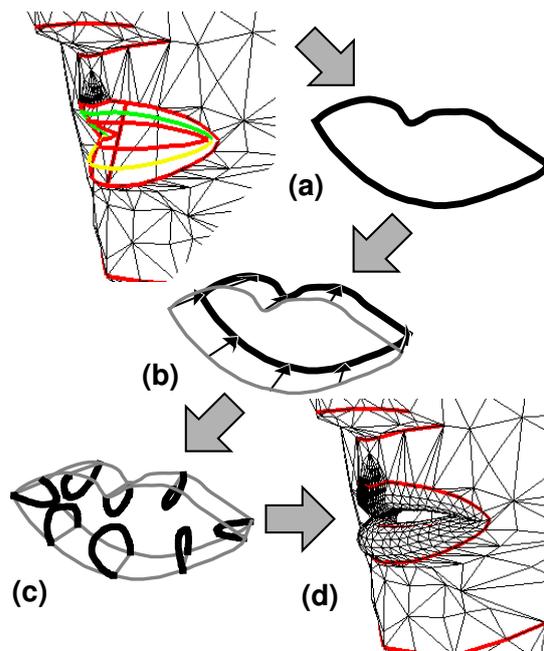


図 1: これまでに開発した唇の外側輪郭に基く唇モデルの基本処理方法

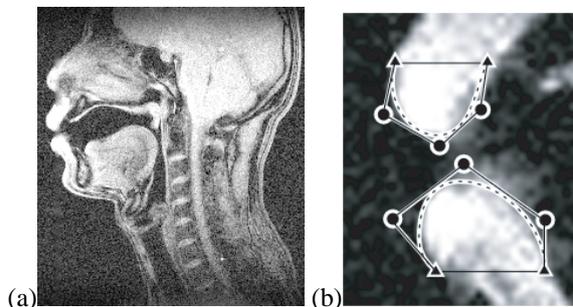


図 2: (a) MRI 正中断面画像例、(b) 唇領域の拡大図と制御点を用いた B スプラインによる唇断面モデル (概念図)

る場合でも、生成された顔形状においてもこの情報は保持されているため、容易に利用可能である。一般化メッシュ中には唇の内側輪郭線も定義はされているが、通常はノイズが多く、また顔表情間で取得できる輪郭線が同じ場所である保証はなく、形状としての信頼性は低いため、本モデルでは後述の厚み補正時および表面テクスチャの取得時に用いている。

**(b) 口腔内境界線の定義:** 得られた唇の外側輪郭線を元に、口腔内部方向に対して輪郭線を投影し、モデル生成における口腔内の境界線とする。投影には唇の外側輪郭線が円柱上に存在すると仮定した場合の円柱座標系における中心線方向に向かい、唇中央における奥行き方向の距離(ここではこれを深さと呼ぶ)を  $d_c(\text{mm})$  とした時に以下の式により定義される

深さモデルに従って行われる。

$$d(x) = d_c - 20.0(x - 0.5)^2 \quad (1)$$

ここで  $x$  は正規化された唇輪郭の水平方向の位置を示し、 $x = 0.5$  が唇中央を示し、 $x = 0.0$  および  $x = 1.0$  が唇両端を示す。この式は数名の被検者の実測値と図 2(a) に示すような MRI 画像の複数の断面データをもとに擬似的に定義したものであるが、上唇・下唇ともに同じ深さとし、また  $d_c$  は  $8.0 < d_c < 20.0$  程度の範囲にあると仮定している。

**(c) スプラインによる断面の設定:** 続いて外側輪郭線と口腔内境界線をそれぞれ予め定義した分割数に分割する。そして、外側輪郭線と口腔内境界線の対応する点どうしを結ぶ線分を含み、外側輪郭線と垂直をなす平面を断面平面として、この平面中でスプラインによる断面輪郭を定義する。我々の唇モデルは、開発の基盤となる共同研究先であったフランス ICP (Institut de la Communication Parlée, Grenoble, FRANCE) 提供のモデル構造を継承して、下唇を 20 分割、上唇では中央部分を密とする 40 分割した断面から構成されている [4]。ただし唇両端では上下の輪郭線が不連続に交わるため、双方から得られる面の中間の面を断面とし、また両端近傍では断面方向が滑らかに変化するように補正を加えている。図 1(c) においては、簡略化のため一部の断面輪郭のみを示している。

具体的な輪郭生成には、始点・終点をそれぞれ外側輪郭線と口腔内境界線とする三次の B スプラインを用い、3つの制御点を用いて図 2(a) に示すような MRI の正中断面が得られる被検者のデータを参考に、上下それぞれの唇に対し唇中央の始点・終点から相対的に制御点の設定を行った。図 2(b) に示す唇領域の拡大図に示すように、黒丸で示す点を制御点として操作し、ド・ブーアのアルゴリズムを用いて点線で示すような B スプライン曲線を生成する [9]。

**(d) 唇ポリゴンの決定:** 最終的には生成された各断面輪郭となる B スプライン曲線をあらかじめ規定してある数の線分で分割し、隣接する線分の頂点を結ぶ三角形ポリゴンを定義することにより唇の表面モデルを得る。

口腔内部の表面のテクスチャについては、一般化メッシュで定義されている唇の内側輪郭線に対応する唇モデル上の点を境界として、唇の外側輪郭線にむかうテクスチャを折り返して口腔内部に用いている。ただし、単純な鏡像では不自然な場合があるため、口腔内部へはテクスチャを引き延すようにマッピングを行っている。

## 2.2 唇モデルの補正

一部の被検者については上記の唇モデルでも十分な視覚的効果が得られるが、単純に唇の外側輪郭情

報のみに頼りきってしまうと、場合によっては不自然な唇形状が得られてしまうことがある。特に、上唇中央の曲線が極端な二重弓形に変化するような人物の場合や、笑顔のように口元における唇輪郭の変化が激しくなる表情などにおいては、この外側輪郭線のみを用いると唇の内側輪郭まで同じような変化を持つことになってしまう。また、顔アニメーションを生成する際に口の閉じ具合が強い場合に、上下の唇が重なりあってしまうことがある。わずかな重なり程度では視覚的には問題はないが、特に極端に重なりあう場合は明かに異和感を生じることとなる。そこでこれらの問題を解決するため、以下に示すような二種類の補正を行った。

**(1) 断面方向と厚み補正:** まず最初に唇の内側が滑らかなようになるように、断面を定義する際の方向と、スプライン生成時の制御点の補正を行った。これには当初のモデルでは利用しなかった唇の内側輪郭線を利用した。内側輪郭線はノイズが多いとはいえ、ほとんどは奥行き方向のみのノイズである。すなわち円柱座標系  $(r, \theta, z)$  においては  $r$  方向が計測上のノイズを含んでおり  $(\theta, z)$  平面上においては比較的安定した輪郭を定義することができる。

そこで、外側輪郭の分割点から断面方向を決定する際に  $(\theta, z)$  平面上の内側輪郭に最も近い点を対応分割点として求め、この対応分割点が存在する方向を断面平面として三次元空間上で定義する。そして外側輪郭の分割点から対応分割点との距離に比例して図 2(b) に示す中央の制御点の高さを調整する。これは、現在の制御点によるスプライン制御では生成断面位置が制御点よりも低い位置となるため、経験的に高め的位置に制御点を設定し、スプライン生成後の内側輪郭に対応する点をこの対応分割点と同様の位置に生成させるためである。

さらに、この厚み補正を行ったのち、隣接する断面間でスプラインの形状が極端に変化しないように、近傍の断面間でパラメーターのスムージングを行っている。

**(2) 接触判定と厚み補正:** 接触判定は、顔アニメーション生成時のフレーム毎の連続した顔形状データに対してのみ行っている。この場合、処理するフレーム数が多くなるため、各断面ごとに接触判定を行うと膨大な計算時間を要することとなる。このため、まず上下それぞれの唇を横方向に四つのグループに分割し、そのグループごとのバウンディングボックスを用いて大まかな垂直方向の重なりが発生するフレームを判定することで高速に接触判定を行っている。

重なりが発生した場合は、重なり対象となるグループ内の各断面について、垂直方向に重なりの中間となる高さにおいて  $(\theta, z)$  平面上に擬似的な内側輪郭線を設定し、この線に対して (1) と同様の厚み補正を行うことにより、双方の唇表面を適切な位置に生成させることができる。

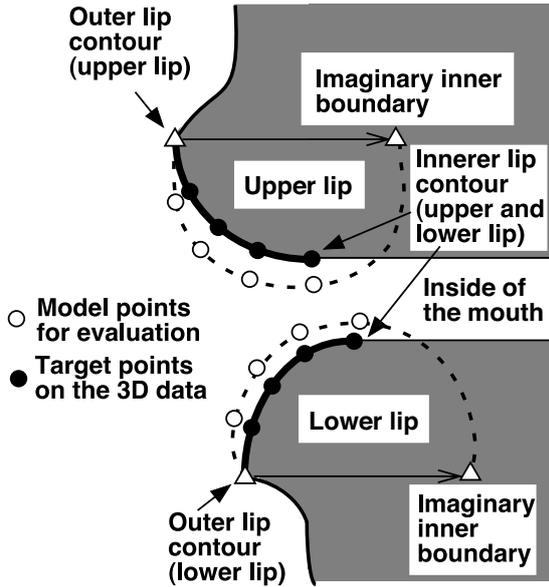


図 3: 話者適応モデル概念図: 三次元データ (太線) 上の目標頂点 (黒丸) と唇モデル (点線) 上の評価点 (白丸) との誤差を最小とする B スプラインを決定するパラメータを多変数の最小化問題を利用して求める

### 3 話者適応唇モデルへの改良

前述のモデルでは、外側輪郭線を主として唇モデルを容易に生成できるため、主成分分析結果を用いた線型推定結果からでも唇ポリゴンをすることが可能で、顔アニメーションに適用して多くの知覚実験用刺激映像を作成する手助けとなった。しかし、唇生成結果に影響を与えるスプライン生成のパラメータは対象となる顔ごとに、その人物らしく見えるように手作業で制御点の設定を行う必要があった。特定人物の顔データのみを利用する場合はそれでも十分であるが、今後様々な人物のデータを扱うこととなると、常に MRI 画像を取得できるとは限らず、このような設定を自動で処理することが望まれる。

そこでもとの三次元計測形状を利用し、スプラインの生成パラメータを調整して最も良く三次元計測形状と一致し、かつ唇の口腔内の構造を持つような話者に適応した唇モデルへの改良を試みた。

#### 3.1 話者適応目標頂点の設定

話者適応唇モデルを得る上で、モデルを構成する頂点の目標となる三次元計測形顔状上の頂点を設定する必要がある。このため、断面を定義する平面は厚み補正で求められた平面を用い、さらにこの時に利用した外側輪郭の分割点と内側輪郭線上の対応分割点を目標点設定に用いる。

図 3 に示すように、外側輪郭の分割点と内側輪郭線上の対応分割点を結ぶ三次元計測データ上の線分

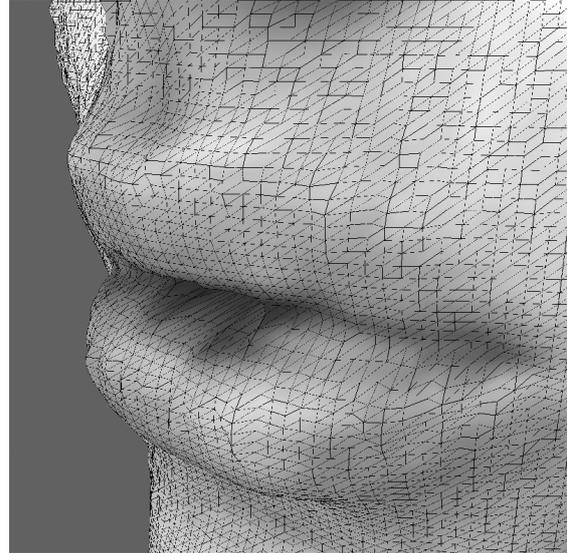


図 4: レーザレンジファインダの計測結果から生成した三次元顔形状メッシュ

上 (太線) において、対応する唇モデル上の点数 (白丸) に応じてこの線分を等分する点を目標点 (黒丸) とする。

これら得られた目標点と、外側輪郭の分割点と内側輪郭線上の対応分割点のみからスプライン生成も可能であるが、これまでに利用したモデルにおいては MRI 映像より得たデータとの整合性も良いことから、現状のモデルを利用して次に示す手法により、最適なパラメータを各断面ごとに求めることとした。

#### 3.2 最適値の決定

目標頂点に対して、唇モデル上の対応点が最も近くなるようなスプラインの生成パラメータが決定できれば、唇の内部表面を擬似的に持ちつつ、もとの三次元データと見かけ上の表面形状が類似した唇を生成することができる。この最適値の決定には、断面平面上にて三点の制御点の座標を決定する 6 個のパラメータを入力値に、目標頂点と唇モデル上の対応点との誤差を戻り値に持つ関数を定義することにより、多変数関数の最小値問題として解くことができる。

今回は Mathworks 社の MATLAB に実装されているシプレックス法を用いた多変数関数の最小値問題を解く `fminsearch()` 関数を利用することにより、目的とする最適値を得ることとした。`fminsearch()` は評価関数と多変数の初期値を定義するだけで、設定した収束条件下における評価関数を最小化する多変数の解を求めることができる。

各断面についてスプラインを生成する最適値を求めることにより、最終的に唇前面においてはもとの

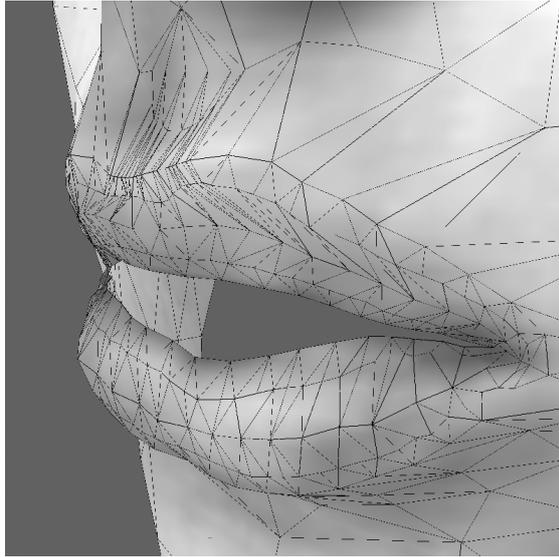


図 5: 計測結果に対し唇領域を含む一般化メッシュを適合した結果

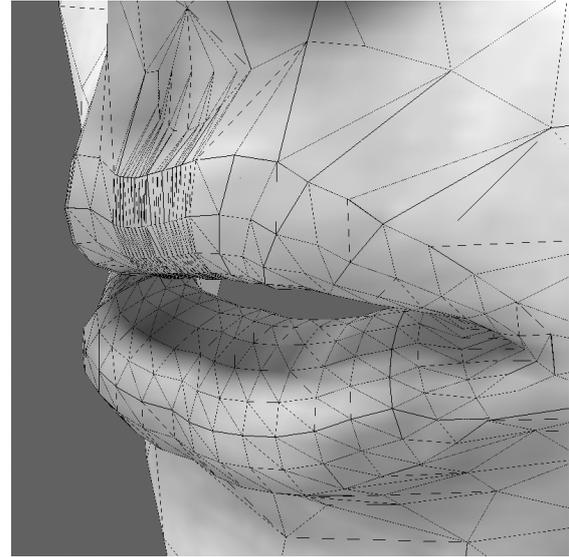


図 7: 唇を含まない一般化メッシュ適合結果と厚み補正を行った唇モデル

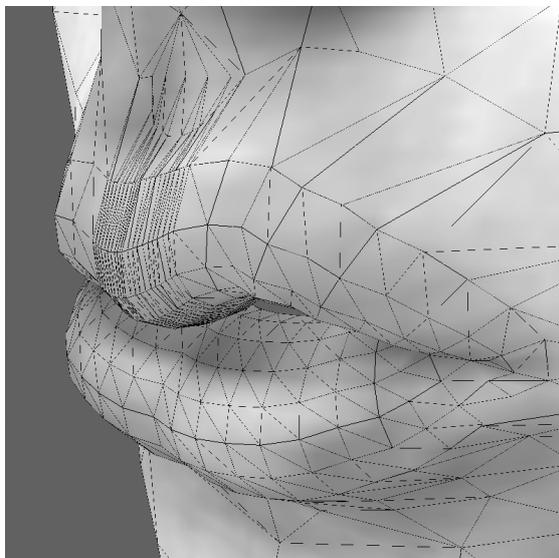


図 6: 唇を含まない一般化メッシュ適合結果と話者適合を行わない初期の唇モデル

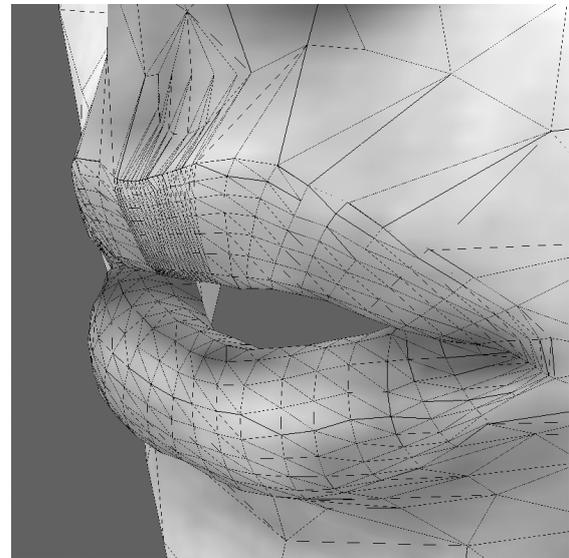


図 8: 唇を含まない一般化メッシュ適合結果と本手法により話者適合を行った唇モデル

三次元計測顔形状にほぼ一致したうえで、口腔内構造も持つ新たな唇形状を得ることができる。

#### 4 話者適応唇モデル生成結果

図 4 に Cyberware 4020RGB/PS レーザーレンジファインダを用いて計測した白人女性被検者の三次元顔形状について三角ポリゴンとワイヤーフレームの両方を用いて生成した唇周辺の画像を示す。このデータに対し、唇領域について単純なメッシュ構造を持つ一般化メッシュを適合した結果を図 5 に示す。唇

の内側領域はノイズが多いため、口腔内に至る領域の比較的形状が安定した部分に対して内側輪郭を定義しているため、この図におけるような視点から観測した場合、唇表面が突然切れてしまい、不自然な印象を与えてしまう。

これに対し、唇以外の一般化メッシュの適合結果から得られる外側輪郭を用いて、単純なスプラインを利用して生成した唇モデルと一般化メッシュ適合結果を示したものが図 6 である。唇の口腔内部領域のメッシュ構造を有することで、先の様な不自然な印象は無いものの、このモデルでは話者適合を行っていないため、この話者としての唇形状の特徴が失

われてしまっている。さらにこの話者の場合、上唇中央の輪郭線の二重弓状の変化が大きいため、唇の内側にまで大きく影響を与えている。

この外側輪郭のみから得たモデルに対して厚み補正を行った結果を図7に示す。これより、上唇の内側における外側輪郭の影響が無くなっていることがわかるが、まだ元の形状と違いがあることがわかる。

これに対して話者適合を行った結果が図8である。図6、7同様に、口腔内部領域のメッシュにより不自然な印象を与えず、さらにこの話者に適応した唇表面構造が得られていることがわかる。以上のように話者適合を行うことで、より視覚的に適切な唇形状を生成することが可能である。

## 5 まとめと今後の課題

本論文では、レーザーレンジファインダなどの三次元計測装置では計測できない唇の口腔内部にいたる構造を有し、さらに計測された唇領域のデータを利用して、話者に適応した唇モデルを生成する手法を提案した。今後の課題としては、より多くの被検者データに適応し、得られた結果の主観評価実験をおこなうこと、顔アニメーションに応用し、発話の知覚にどれだけ効果を与えるかを調査することが挙げられる。さらに、唇生成のスプライン曲線生成法をサンプル点経由の手法への改良を行う予定である。また、本モデルを応用し、三次元顔形状データベースの分析の一環として、唇領域についての人物間の共通性や個人性に関係するパラメータの特定なども今後の課題の一つである。

## 6 謝辞

本研究は通信・放送機構の研究委託「人間情報コミュニケーションの研究開発」により実施したものである。

## 参考文献

- [1] Hani C. Yehia, Philip E. Rubin, and Eric Vatikiotis-Bateson. Quantitative association of vocal-tract and facial behavior. *Speech Communication*, 26:23–44, 1998.
- [2] Michael Cohen and Dominic Massaro. Synthesis of visible speech. *Behavior Research Methods: Instruments & Computers*, 22:260–263, 1990.
- [3] Dominic W. Massaro. *Perceiving talking Faces*. MIT Press, Cambridge, MA, 1998.
- [4] Lionel Reveret and Christian Benoit. A new 3d lip

model for analysis and synthesis of lip motion in speech production. *AVSP'98*, pages 207–212, 1998.

- [5] 倉立尚明, Frédérique Garcia, Hani Yehia, and Eric Vatikiotis-Bateson. 三次元運動学に基づく顔面アニメーション. *情報処理学会 グラフィクスと CAD 研究会 研究報告*, 97-CG-87:31–36, 1997.
- [6] 倉立尚明, Guillaume Vignali, and Eric Vatikiotis-Bateson. 大規模三次元顔形状データベースと cg アニメーション応用. *Visual Computing グラフィックスと CAD 合同シンポジウム*, pages 105–110, 2003.
- [7] Takaaki Kuratate, Saeko Masuda, and Eric Vatikiotis-Bateson. What perceptible information can be implemented in talking head animations. *IEEE International Workshop on Robot and Human Interactive Communication (ROMAN2001)*, pages 430–435, 2001.
- [8] Yuencheng Lee, Demetri Terzopoulos, and Keith Waters. Realistic modeling for facial animation. *Proceedings of SIGGRAPH95*, pages 55–62, 1995.
- [9] Gerald Farin 著, 木村文彦監修, and 山口泰監訳. *CAGD のための曲線・曲面理論*. 共立出版, 1991.