

TR-H-310

MRIに基づく声道形状の抽出
と実体モデル作成

藤田 覚

2001.2.7

ATR人間情報通信研究所

〒619-0288 京都府相楽郡精華町光台2丁目2-2 TEL: 0774-95-1011

ATR Human Information Processing Research Laboratories
2-2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan

Telephone: +81-774-95-1011
Fax : +81-774-95-1008

MRI に基づく声道形状の抽出

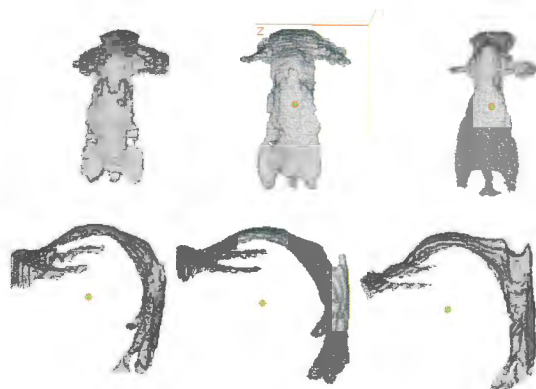
と実体モデル作成

MRI-based Extraction of Vocal Tract Shape
and Construction of a Physical Vocal Tract Model

藤田 覚

ATR 人間情報通信研究所

〒619-0288 京都府相楽郡精華町光台 2-2



1	概要.....	3
2	MRIデータに基づく声道形状の抽出.....	3
	2.1 声道のMRIデータの特徴.....	3
	2.2 MIMICSによる声道形状の抽出.....	4
3	声道形状の特徴.....	6
	3.1 5母音の声道形状.....	6
	3.2 男性話者3名の声道形状の比較.....	9
4	実体モデル作成.....	12
	4.1 積層光造形装置 E-DARTS.....	12
	4.2 積層データ作成.....	12
	4.3 E-DARTSの使用手法.....	13
5	実体モデルを使った音声合成実験.....	14
6	まとめ.....	16
7	参考文献.....	16

1 概要

音声合成は単純な声道形状を仮定して行われるが、実物の声道の形状は複雑である。そこで磁気共鳴画像（MRI）データより声道の3次元形状を復元し、それを解析すればより精密な音声合成の知見が得られると考えられる。また、実体モデルを作成し、人工音源で駆動して音声合成することも考えられる。さらに、実体モデルに変形などの解析操作を施すことが可能であり、音声合成の人間らしさあるいは人間の発声機構の解明に役立つと考えられる。

この報告では頭頸部のMRIを元にした声道の3次元形状データの抽出方法を述べ、母音の形状の特徴、および男性話者3名の声道について比較した形態的特徴を述べる。また声道の変形操作による合成音比較実験についても簡単に述べる。

2 MRI データに基づく声道形状の抽出

MRI データは静磁場強度 1 T をもつ MRI 装置（島津製作所 SMT-100GUX）により記録された資料であり、spin echo 法により撮像野(FOV)を 250x250mm、撮像スライス間隔を 3.5mm として矢状面方向に記録されたものを用いた。これを各方向について 2 倍に補間して約 0.5 x 0.5x 1.75mm のボクセルデータとした。被験者は厚さ約 1mm の歯冠造影プレートを装着し、MRI 装置の中で仰向けの状態で発声を行った。撮像時間は母音毎に約 3 分を要した。

2.1 声道のMRIデータの特徴

MR 画像に於いて、声道の空洞部は他の隣接部位に比べ以下の特徴がある。

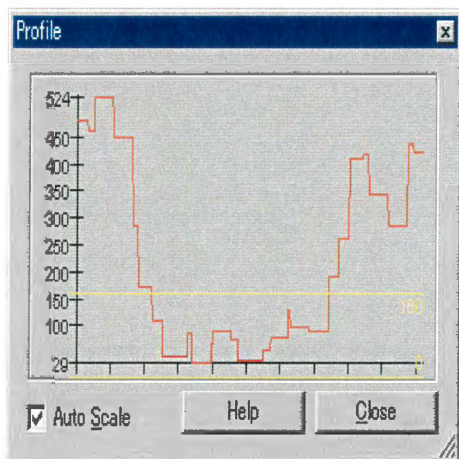
- a) データ(濃度)値が小さい。
- b-1) 大部分の隣接部位との境界は濃度値の急激な低下により抽出しやすい。
- b-2) ただし、口唇の開口部は外気と接続している為、濃度値の境界はない。
- b-3) また、歯との境界は歯冠造影法を用いれば抽出できる。

(これには造影剤をサンドイッチ状に挟み込んだ歯型のプレートをはめて撮影する。)



左図は男性が/e/を発声中のMRIデータで正中付近の画像であるが舌、喉頭壁、喉頭蓋が明瞭に現れている。

歯も歯冠造影プレートの造影剤が見えているが途切れている箇所がある。また声帯の付近は撮影時の被験者の動きにより明瞭ではない部分がある。硬口蓋も一部鼻腔との境界が明瞭でない部分がある。



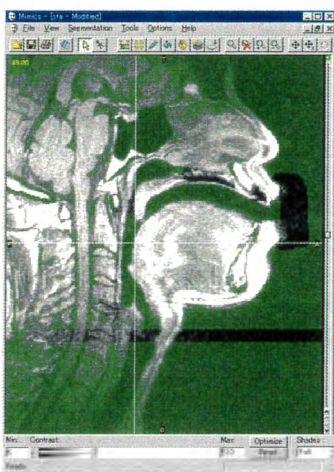
左図は口蓋垂から舌にかけて（前図で赤い線で示されている部分）の濃度値の変化グラフで、空洞部分と内壁の境界部分では空洞部側で濃度値が急激に小さくなっているのが分かる。濃度値の変化が急激なので内壁の明瞭な部分は閾値処理の閾値変化に頑健である。150 から 300 まで閾値を変化させた場合でも閾値境界の内壁側への移動は 2,3pixel に留まる。（1pixel はこの場合 0.5mm 角である。）

閾値は声帯付近の分離を目視して決めるが、対象のデータでは 150 から 300 の範囲が適当であった。

2.2 MIMICSによる声道形状の抽出

声道の抽出には Materialise 社製 医療用画像処理ツール「MIMICS」を使用した。理由は閾値処理、目視による修正、region growing などの機能が揃っているため、および積層造形システムへの入力機能が揃っているためである。但し、目視による修正時に対象となるスライスしか見られず、他のスライスが参照できない点は改善を要する。また、処理の中心となる「マスク」の仕様が公開されておらず他の画像処理ソフトとの連携ができない点も改善が期待される。またMRI装置のデータ形式から汎用のデータ形式の変換にはMRI Toolbox を使用した。（参考文献1）

閾値処理



左の図は正中付近のスライスを閾値処理したものである。処置したあとのデータは2値化され MIMICS では「マスク」と呼ばれる。ここではマスクの on の部分は緑色で示されている。閾値は声帯付近および梨状窩周辺の抽出結果に従って設定した。（声帯付近はMRデータにより明瞭度が異なり同一の閾値ではうまく抽出できなかった。声道と内壁との境界では濃度値の変化が急激なのでそこでの閾値変化の影響は少ないとはいえ閾値の決定方法は改善されるべきであろう。）

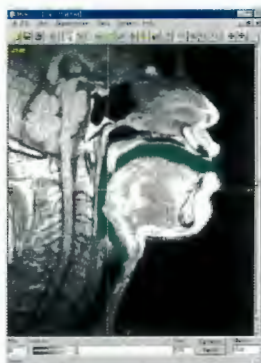
画像から分かるように声道は全て緑色になっている。また低濃度値の鼻腔、骨、体外空気も緑色になっている。

マスク修正

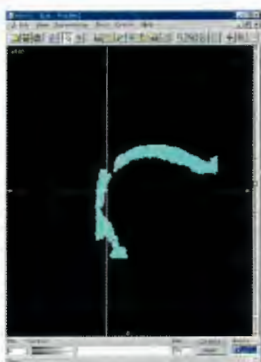
閾値処理では正しく分離されない部分は手作業で修正した。修正はMIMICSで用意されている「マスク編集」で行った。声帯付近と歯の分離は現段階では手作業を必要とする。

MIMICSのマスク編集では「pen」の大きさと「erase,draw,threshold」のモードが選択できるので比較的効率よく編集できる。

region growing



region growing はあるボクセルを指定してその隣接する on のマスク部分を逐次領域に追加し、連結領域として抽出する方法である。声道の抽出においては、大部分の境界は明瞭な内壁であり閾値処理で分離でき、手作業を要する部分は少ないので、region growing のよる声道抽出法が効率的であると考えられる。前処理としてマスク修正において唇の先端付近で外気と分離し、声帯で気管と分離しておく。鼻腔が連結している場合も今回は分離を行った。



声道内の1ボクセルを指定し region growing を実行した結果は左の図のようになる（青のマスク）。左図では1スライスだけの表示であるが全スライスにわたって（3次元的に）この処理は施されている。声道の分離が完全でなかった場合には結果のマスクが他の領域を含んでしまうので「マスク修正」あるいは「閾値処理」から再度行うことが必要である。

(左図はマスクだけの表示)

3次元表示

region growing により声道が連結領域として抽出されたならばMIMICSを用いて3次元表示し確認する。



回転などを行って各部を点検し、不具合があったら上記の閾値処理やマスクの修正をやり直す。

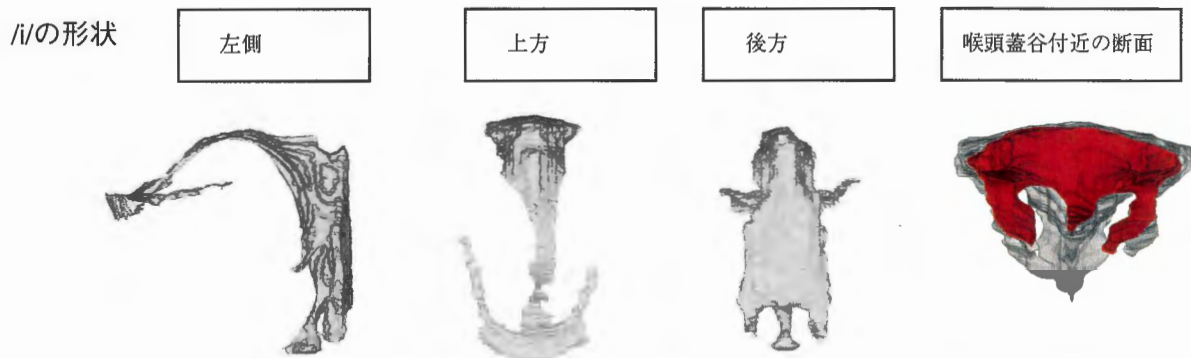
(図は抽出された声道の3次元表示 左側から見たものと(左)と上から見たもの(右))

3 声道形状の特徴

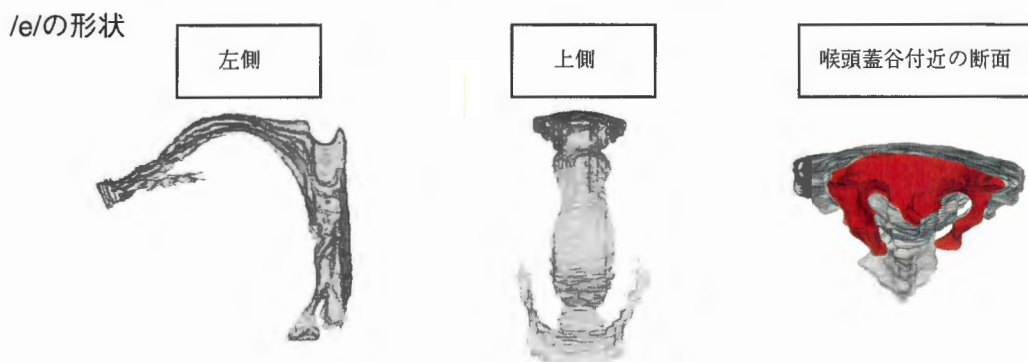
抽出された声道の3次元形状データをもとに、男性話者1名の5母音の声道形状及び男性話者3名の声道形状の比較をした。尚、撮像は仰向けの状態でさらに歯冠造影プレートを装着して行った。

3.1 5母音の声道形状

以下に成人男性Hの5母音発声時の声道形状を図示する。

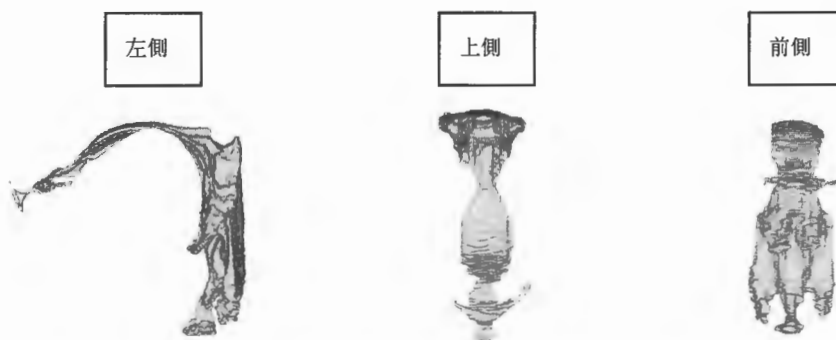


/i/は音声学分類上では張唇前舌狭母音となっているが、抽出された声道形状でも唇が横方向にひろがっているのが分かり、前舌部が上方に上がり硬口蓋に近づいていることが分かる。これに伴って舌根部の前方への移動が起こり、この部分で声道が広がっている。このデータでは歯列間隙が観測される。これは、頬と歯、舌で囲まれた空間で声道の分岐管であり、/i/において最も長くなっている（参考文献2）。また、喉頭蓋谷も小さな分岐管になっている。喉頭蓋谷付近の声道断面を見てわかるように、その付近の断面（赤の面）は複雑な形状をなす。また声道の下端に位置する梨状窩も分岐管をなしている。



/e/は張唇前舌半狭母音に分類されているが、抽出されたデータでも唇は横方向にひろがり、前舌部が/i/ほどではないが硬口蓋に近づいている。口腔における声道断面積の変化が小さい。また歯列間隙が最も明瞭に見られる。後方から見た形状は/i/と大差なかった。喉頭蓋谷は分岐管を成しているが前方への突き出しは小さい。

/u/の形状



/u/は円唇奥舌狭母音と分類されているが、唇の開口部は小さく円い。この母音は、多くの場合（特に関東方言話者では）、中舌の/w/と発音される。舌は舌背（軟口蓋と硬口蓋の境界付近）で盛り上がり声道を狭くしている。口腔の前側も上下方向には比較的狭く歯の付け根の奥で特に狭い。横方向には硬口蓋付近で広がっている。したがって口腔の断面積は全般的には小さいが声道全体の断面積の変化は大きい。歯列間隙も小さいが観測される。梨状窩、喉頭蓋谷は分岐管をなしている。

/a/の形状



/a/は張唇奥舌広母音と分類されている。図では唇の開口部は横に広がっている。口腔の前部は断面積が広く、軟口蓋から咽頭にかけて狭くなっている。声道断面積の変化は大きい。このデータでは歯列間隙が観測されるがこれは歯冠造影プレートの影響であると考えられ、本来この部分は口蓋付近の空間と一体になると考えられる。（後出する他話者の/a/のデータを参照）

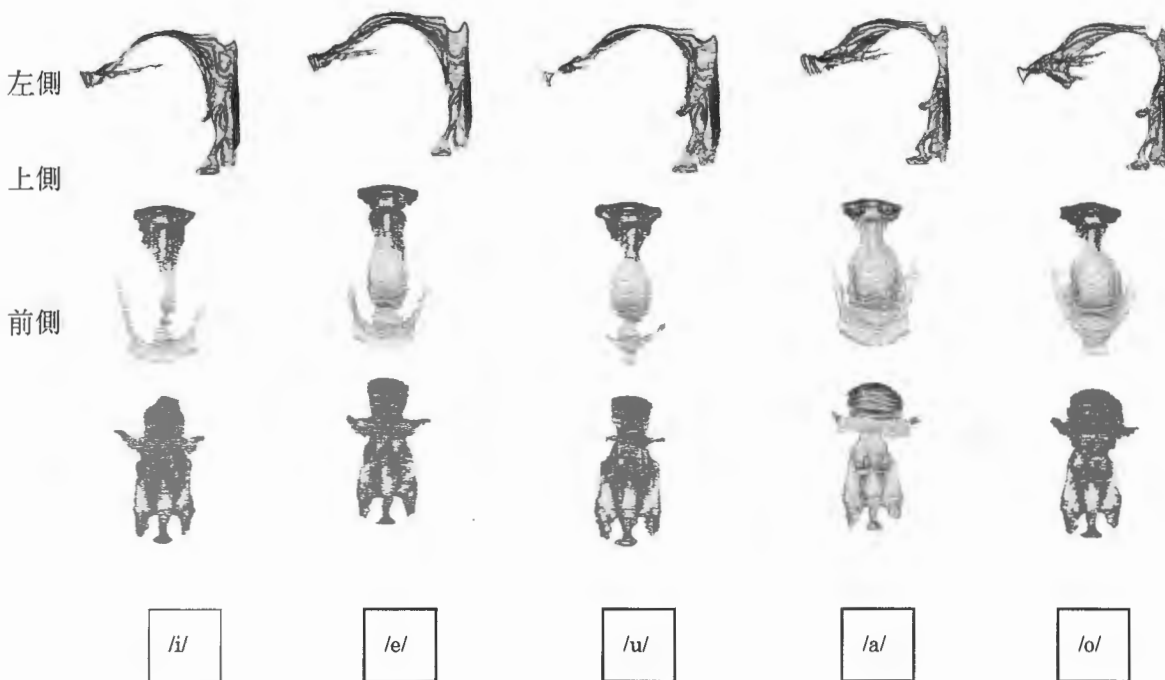
/o/の形状



/o/は円唇奥舌半狭母音と分類されている。図でも唇の開口部は縦横方向とも狭く円い。口腔内前側の断面積は大きい軟口蓋付近で小さい。唇の開口部以外は/a/と似ていて声道断面積の変化も大きい。またこのデータでは歯列間隙が観測されるがこれは歯冠造影プレートの影響と考えられ本来は口蓋付近の空間と一体になると考えられる。

まとめ

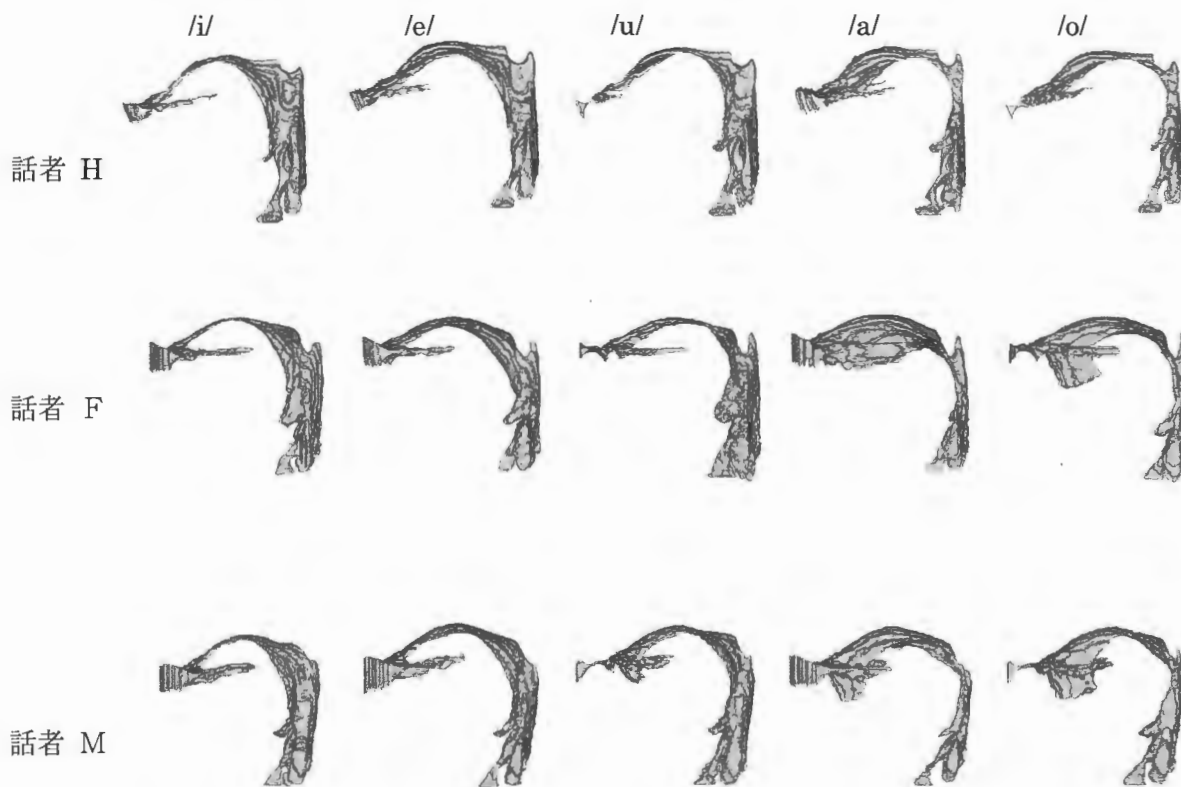
音声学的分類の特徴が抽出された声道形状データでも観測された。喉頭蓋谷と梨状窩は常に分岐管を成し、/i/ /e/ /u/ では歯列間隙による分岐管がある。断面の形状は複雑であり、部位による変化と調音による変化がある。以下に5母音の3方向から見た形状図をまとめる。



3.2 男性話者3名の声道形状の比較

下の図は男性話者3名の母音発声時の声道形状である。左側面図、上面図、後面図をそれぞれ示す。

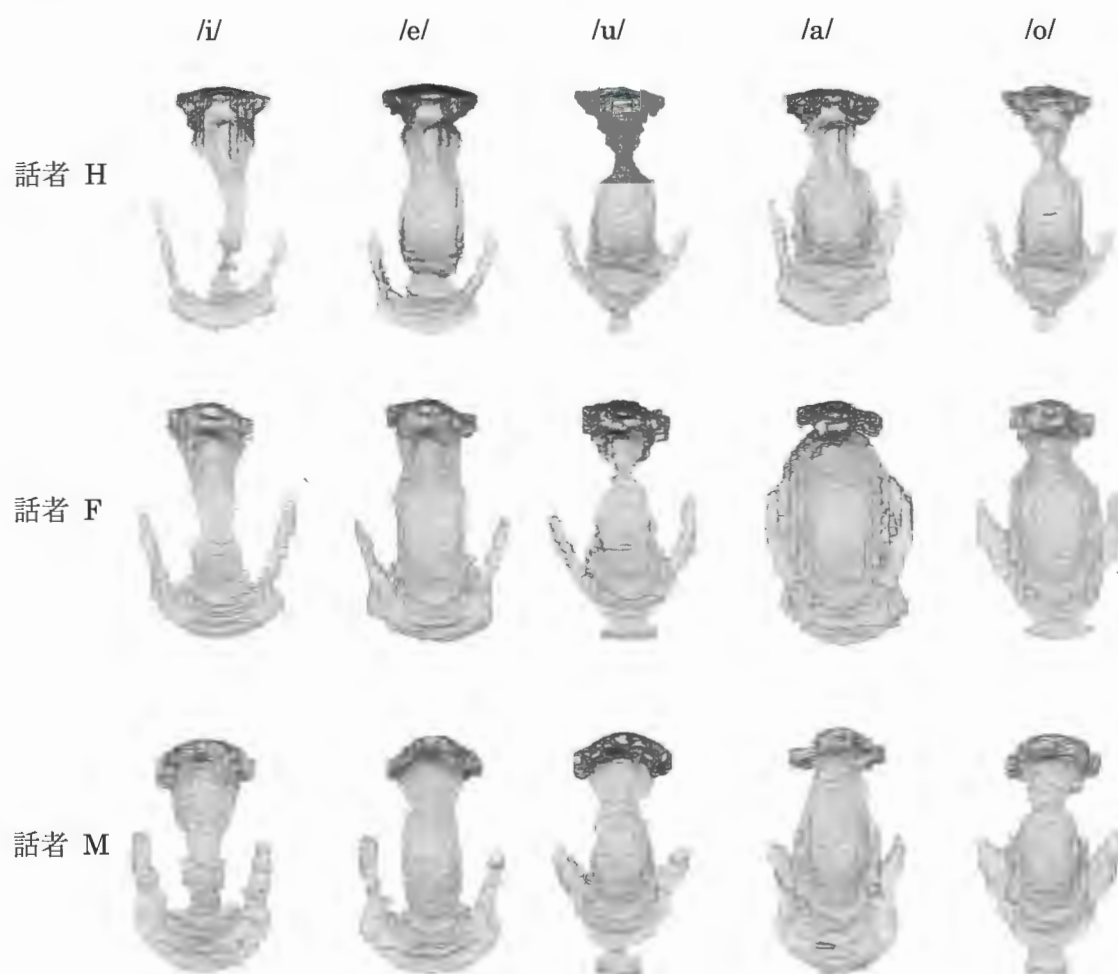
左側面図



それぞれの調音の傾向には共通性があるが、発話器官形状と細かい調音に個人差が見られる。まず口腔の長さや咽頭の長さの比に違いが見られる。また口蓋から咽頭の粘膜へつながる形にも個人差が見られ、鼻腔方向への分岐管の形成にも個人差が見られる。喉頭蓋谷による分岐管の形成も個人差がある。

/a/と/o/において前舌下方に空間がみられるものとみられないものがあるが、これは仰向け姿勢による長時間の発声による前舌の後方への移動に影響していることが考えられる。同被験者による/a/のデータで前舌下方に空間がある場合と無い場合とが見られた。/e/において歯列間隙が明瞭に形成される点は共通して認められる。

上面図



上方から見ても口腔断面の横方向の変化には共通性があり調音の傾向が似ていることが分かる。但し断面の横方向の大きさには個人差がある。

歯列間隙の形には大きな差があるが歯冠プレートの影響が大であろう。話者 F の/a/の発声では歯列間隙が生じていないが奥舌の調音の場合にはこれが本来の形状であると考えられる。

後面図

M /e/ (後方より)

F /e/ (後斜め上方より)

K /e/ (後方より)



後方から見た場合には母音による違いはあまり無いので各話者の /e/ だけを表示する。
また話者Fでは分岐を示すため少し上方からの視点にしてある。

後方から見た形状は個人差が大きい。

- ・ 左右対称性に個人差がある。
- ・ 梨状窩の大きさ、形状に個人差がある

喉頭蓋の咽頭粘膜への接触により分岐を生じる場合がある。但し、仰向けの状態であることが影響していると考えられる。

- ・ 鼻腔方向への分岐管の形状に差がある。

まとめ

話者の各母音の声道形状は類似していて、その調音は類似していると考えられる。
一方、話者の声道形状には個人差があり、声の個人性が声道形状にも由来していることは明らかであろう。特に左右対称性や分岐管の形状、大きさに違いが見られる。

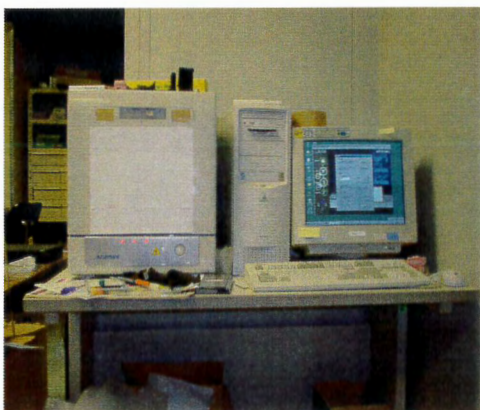
今回の考察は定性的なものがあるため今後定量的な解析が望まれる。

4 実体モデル作成

上で作成した声道の3次元形状データを実体化した。観察用に声道の形状（空洞の形状）をそのまま、音声合成用に管状のもの（実体モデル）を造形した。ここでは樹脂製の固いモデルを作る装置と方法を説明する。造形装置には、機械的に削りだしを行うタイプと積層タイプがある。削りだしタイプは声道の様に複雑なものは造形できない。特に管状の内部に空間があるものは難しい。積層タイプには光硬化性樹脂を用いたもの、紙を切って積層するもの、粘性の高いインクジェット方式のもの、澱粉を塗り固めるものなどがある。今回は計画時点で手に入る低価格なものを選んだ。造形精度は、表示上はどの製品も大差なかった。(0.1mm程度)

4.1 積層光造形装置 E-DARTS

E-DARTS は光硬化性樹脂が入ったトレーの下からレーザービームをXYプロッタの様に照射して硬化させ1層を形成し、造形物を上昇させることによって積層していく装置である。



仕様：

最小硬化単位：XY軸方向0.08mm Z軸方向 1.00mm

最小移動単位：XY軸方向0.02mm Z軸方向 0.01mm

ソフトウェア分解能：XY軸方向0.01mm

Z軸方向0.01mm

積層厚： Z軸方向 0.1 ~ 0.4mm

造形領域：幅 200mm 奥行 200mm 高さ 200mm

4.2 積層データ作成

E-DARTS専用の編集ソフト(AutoEdit)を積層データ作成に用いる。入力となるSTL形式の3次元形状データはMIMICSのCTM機能を使用する。

注意すべき点は

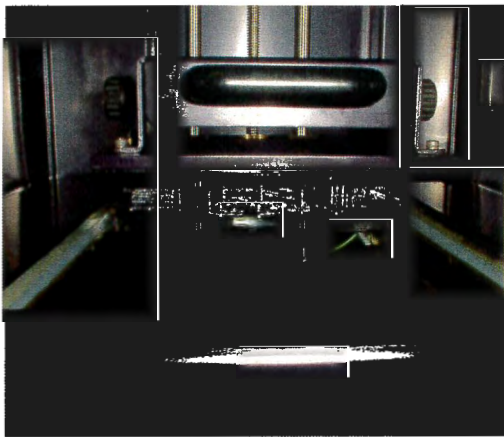
- ・ 引き上げ式なので局所的に高い部分(ピーク)はサポートを要する。
- ・ 喉頭蓋付近の複雑な部分はサポートを作成できないことがあるので、回転をしてみてピークができない様にする。
- ・ ピークが複数あり回転では対処できない場合は分割を考える。分割にはMaterialize社製CADの「MAGICS」の分割機能が使える。
- ・ 歯列間隙は細いのでサポートを多く付ける。

- ・ 唇の先端部分は細かなピークがたくさんできるのでサポートを大量に付ける。
- ・ 空洞データ（実体モデル）は声道の形状データ（空洞の形状データ）からMAGICSの中空化の機能で作成する（hollow）。中空化の厚みは5mmとした。（積層厚0.1mmでは造形に10日程かかる。）これ以上の厚さだとさらに造形時間を要し、また薄いとひび割れが発生する。
- ・ 実体モデルの歯列間隙部分は縦に造形した方が精度がよい。斜めの造形だと引き上げによると考えられる変形が大きい。また喉頭蓋が内部でピークになるため口蓋垂付近で分割し、それぞれピークの回避と精度の確保をした方がよい。
- ・ 積層厚は、時間はかかるが0.1mm（最小値）が精度を確保するためにはよいと思われる。

4.3 E-DARTSの使用方法

ここでは声道の作成に関する注意を述べる。

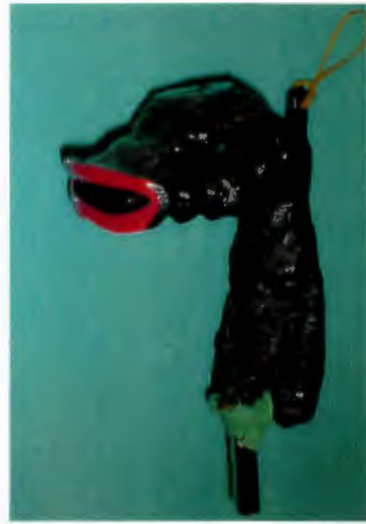
- 1) 光硬化性樹脂と硬化剤を混合して使用するが、マニュアル通りの比率（1%）が良い。
- 2) 0点設定は毎回行った方が確実である。繰り返して造形した時に0点がズレ、引き上げできずに、造形物がトレーに固着したケースがあった。
- 3) 造形中はethernet cableをはずしておく。「ネットワークサービスを停止せよ」とメーカーの指示があるが、この処置で問題は起きてない。
- 4) 声道の造形には7～10日位要するが引き上げに失敗するケースがままあり、そのままにしているとトレーに固着し、トレーが破損することが多い。一日一回は一時停止させて造形状態を確認した方がよい。
- 5) またレーザの出力低下も起こりやすい。この場合には引き上げ面全体が同時に引き上げ失敗し、他の原因と区別される。



E-DARTS内部：下部が樹脂の入ったトレー、
上部は造形プレートで一層ずつ引き上げる



上図は製作された声道の形状モデルの例



実体モデルの例：

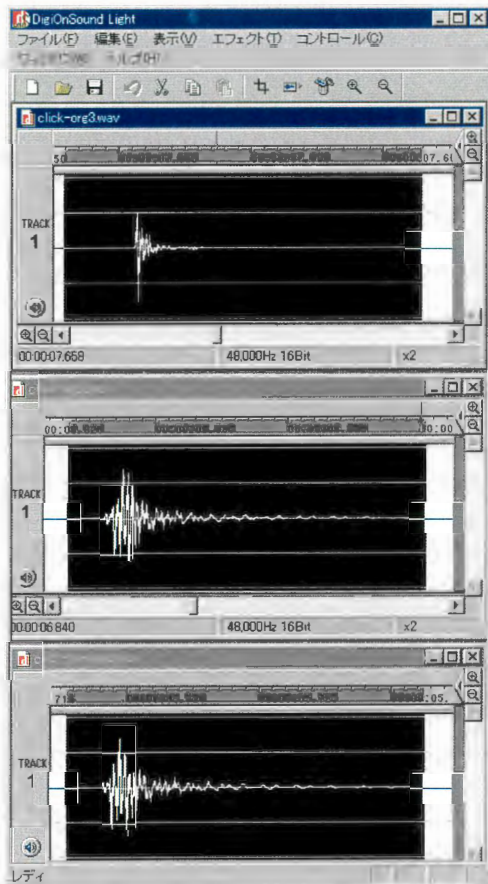
チェロの調子笛を装着し着色してある

5 実体モデルを使った音声合成実験

被験者を使った実験では難しいが、声道実体モデルでは部分的に変形・削除することによってその部分の音響的機能を解析することができる。ここでは、歯列間隙を充填による合成音への影響を調べた結果を述べる。実験には歯列間隙が明瞭に見られる/e/の実体モデルを使用し、歯列間隙をシリコンゴムで塞いで合成した音と塞がないで合成した音を比較した。音源はクリック音を用いた。また口唇部にはアクリル製バッフル板（200x135mm）を取り付けた。



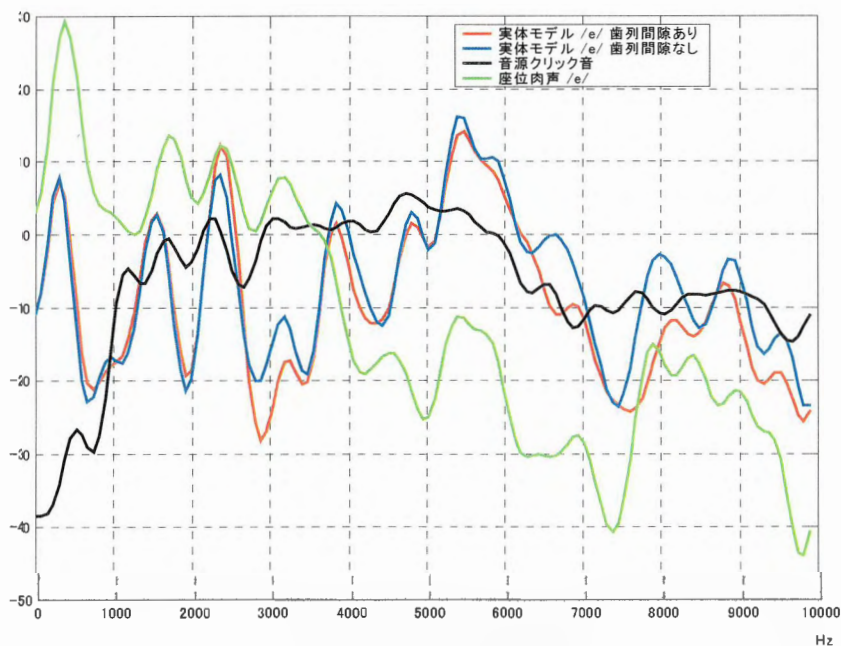
左図は声道実体モデルにバッフル板を付けたものである。実験では実体モデルをカメラ用の防音袋で覆い音源からの周り込み音を防いだ。マイクはSONY製ECM-MS957を指向性120度にして用い、それをONKYO USB DIGITAL AUDIO PROCESSOR SE-U77を介してノートパソコンに直接、sampling 周波数 48KHz 16bit PCMモノラル録音した。



左図は録音されたデータの波形で上から「click音のみ」、「歯列間隙ありの実体モデルの合成音」、「歯列間隙を塞いだ実体モデルの合成音」である。それぞれの click 音部を含んだ 650 サンプルの間を黒の背景で示してある。

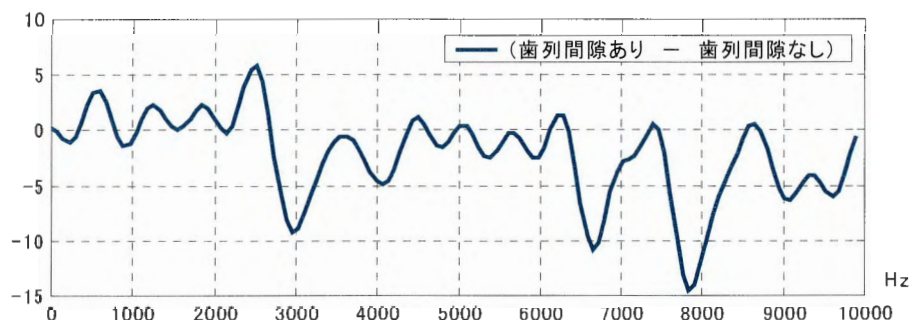
周波数解析は窓長 650 (約 13.5msec 区間) として HAMMING 窓を用いたフーリエ解析の後、ケプストラム包絡を求めた。リフターのカットオフは 2 msec とした。

録音したデータのクリック音部分に対するケプストラム包絡線が下図である。それぞれ 3 区間分を平均してある。歯列間隙は零点を形成すると考えられるが、このデータでは実体モデルの歯列間隙によるとみられる零点の移動が 2800Hz 付近にある。歯列間隙による影響の程度を調べるには変形量を調節するなどの実験が今後必要であろう。座位による肉声との比較では F 1, F 2, F 3 とみられるピークがあり周波数特性に類似があると考えられる。しかし F 1, F 2 の低域への移動がみられるなど違いがあり、今後の課題である。



ケプストラム包絡線

下図に歯列間隙ありとなしの差を示した。零点を構成するマイナス方向への移動が低域では3000Hz付近にみられる。



6 まとめ

母音の声道形状を抽出し、実体モデルを作成した。歯列間隙の変形操作による音声の変化を実験したが、声道の形状の声への寄与を確かめるためには変形量の調整が必要であることがわかった。また、今回製作した実体モデルは固い素材で出来ているため定常音しか発声できない。柔らかい実体モデルならば変形量が調整可能であり、また動的な発声調節も可能なので子音も生成可能である。それを機械制御するならば人間の発声機構の解明にさらに貢献するであろう。

7 参考文献

1. MRI toolbox、 Mark K. Tiede、 ATR Technical Report TR-H-262
2. 上下歯列間隙により生じる声道の分岐管 本多清志 藤田覚 Mark Tiede 和久本雅彦 党建武、日本音響学会講演論文集 1999年9月～10月 p.235
3. MIMICS マニュアル (株) マテリアライズ
4. MAGICS マニュアル (株) マテリアライズ
5. E-DARTS マニュアル (株) アウストラータ