

静岡大学 博士論文

有限要素法を用いた三次元声道画像からの
面積関数の推定

2007年12月

大学院電子科学研究科

電子応用工学専攻

望月圭太

要旨

声道面積関数は、調音の様子を直感的に捉えることができ、音素の遷移時の声道形状の変化の様子もわかり、音声スペクトルへの変換も容易であることから、音声生成の研究にとって非常に有用である。そのため、音声生成に関わる数多くの研究で解析に用いられている。しかし声道形状の複雑さから適切な声道面積関数の導出は困難であり、声道面積関数の正確さについてはこれまで意識されていなかった。そこで、本論文では、声道内の等音圧面を用いることでより正確な声道面積関数を推定する手法を提案した。提案法は核磁気共鳴画像 (MRI) より得られる顔断面画像から声道領域を抽出し、声道形状から有限要素法による数値計算で等音圧面を求め、その等音圧面を用いて声道分割を行い、それぞれの分割領域の厚さと体積から声道面積関数を得る。これまでの方法の多くは声道を平面で分割していたため、咽喉部で隣接する二つの分割面が交差しやすいという問題があったが、等音圧面は交差しないため、提案法ではより自然な声道分割が期待できる。比較手法として声道を平面で分割する方法を用意し、提案法との比較により評価を行った。有限要素法より得られるホルマント周波数を基準値として、成人男性 2 名の声道データを提案法と比較手法にそれぞれ適用して得られた声道面積関数からホルマント周波数を求め両者を比較した結果、比較手法では基準値に対する誤差率は平均 6.5%、標準偏差 3.3% であったのに対し、提案法では平均 1.7%、

標準偏差 1.4% となり，提案法の有効性を示した．

また，等音圧面の形状はどの周波数で音圧分布を求めるかにより異なるため，いくつかの周波数の等音圧面から声道面積関数を計算し，ホルマント周波数を比較することで提案法にとって最適な等音圧面を形成する周波数を検討した．提案法で声道面積関数を計算するためには等音圧面の形状は帯状である必要があるが，第 1 ホルマントを超える周波数領域では等音圧面の形状が帯状とならないため，それ以上の周波数の等音圧面を用いて声道面積関数を計算することは困難である．従って 1Hz から第 1 ホルマント周波数付近までの等音圧面を用いて声道面積関数を作成し，ホルマント周波数の比較を行った．その結果，第 1 ホルマント周波数近傍の等音圧面を用いることが最適であることを示した．

目次

第1章	はじめに	1
1.1	音声とは	1
1.2	背景と目的	3
1.3	本論文の構成	5
第2章	音響解析手法	6
2.1	はじめに	6
2.2	M R Iによる顔断面の撮像	6
2.3	有限要素法	8
2.4	音響等価回路	11
2.5	線形予測法によるホルマント周波数検出	14
第3章	声道断面積関数の導出	16
3.1	はじめに	16
3.2	前処理及び可視化	16
3.2.1	声道領域の抽出	17
3.2.2	中心線の検出	22
3.2.3	各種音響パラメータの可視化	27
3.2.4	等音圧面の可視化	29
3.3	声道領域分割	32

3.4	音響等価回路への適用	34
3.5	比較手法	37
第4章	結果と考察	40
4.1	3次元声道データ	40
4.2	声道壁インピーダンスを剛壁とした際の周波数特性	43
4.3	声道壁インピーダンスを有限値とした際の周波数特性	44
4.3.1	はじめに	44
4.3.2	近似式によるホルマント周波数の補正	57
4.3.3	損失を考慮した音響等価回路を用いたホルマント周 波数の導出	57
4.4	口唇付近の補正の効果の検証	61
4.5	音圧分布導出のための周波数と面積関数との関係	63
第5章	まとめ	93
付録A	3次元画像に対する細線化アルゴリズム	98

第1章 はじめに

1.1 音声とは

音声は、我々人類にとって情報伝達のための最も重要な手段である。音声の種類を大別すると母音と子音の2種類となる。母音は、声帯で発生させた三角波状の体積速度波が声道を通り、口唇から放射されることで生成される。主に声道の形状により母音の種類が決定される。一方子音は、声道の一部を一時的に閉鎖する、または狭めることで発生する瞬発的または持続的な音を主な音源としている音である。声帯の振動の有無、及び声道の狭めの位置により子音の種類が変化する。音源により生成された振動が空洞を通過し、特定の周波数の共振を伴い外部に出力される。音声における発音メカニズムを音響管モデルにあてはめたとき、声道を音響管の一種とみなすことができる。音声は声帯の特性、声道形状による特性、及び口唇の放射特性の相互作用により生成される。これらの特性をそれぞれフィルタとして独立に表現することができる。それぞれのフィルタは図1.1に示した特性を有している。

声道伝達特性の共振周波数の位置は、声道形状にのみ依存する。この共振周波数のことをホルマント周波数と呼ぶ。ホルマント周波数は母音の種類を決定付ける要素であることが知られている。

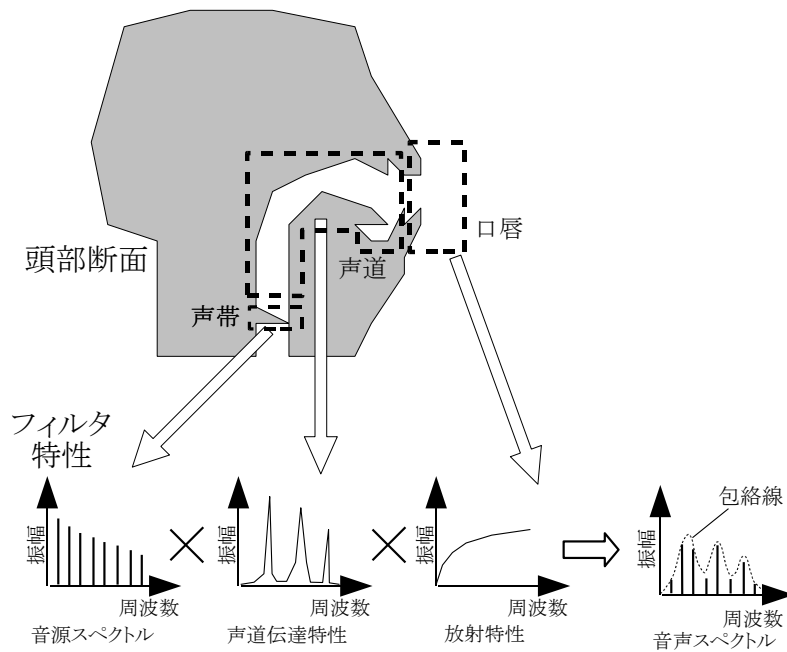


図 1.1: 声帯，声道形状，口唇のフィルタ特性

音声スペクトルは，声帯や狭めの位置で生成される音源スペクトルと，声道伝達特性，及び口唇の放射特性の掛け合わせである．母音における音声スペクトルの包絡線の極大値の位置は，声道伝達特性の共振周波数により決定される．包絡線が極大値となるこの共振周波数のことをホルマント周波数と呼ぶ．

1.2 背景と目的

音声発声時の物理現象を手がかりに声道形状を推定し，その声道形状から声道伝達特性を推定することで，その音声の声質と同等の音声を生成することができる．声道形状を調査する方法として，発話音声や口唇インパルス音から得られる音声スペクトルを用いて声道形状を推定する方法 [1–6] と，音声発声時の声道形状を撮像により直接取得する方法 [7–24] の2通りに分けることができる．前者のアプローチでは，1つの音声スペクトルから複数の声道形状が得られる可能性があるため，発話音声または口唇インパルス音から声道形状を一意に推定することは困難であることがある．一方，後者のアプローチでは声道形状から音声スペクトルは一意に定まるため，前者の問題を考慮する必要はない．声道形状を撮像し，音声スペクトルを得るアプローチは千葉，梶山 [7] の研究，及び Fant [8] の研究に端を発する．これらの研究では，X線撮像により声道形状を取得し，これを直線的な音響管に近似することで解析を行っているが，X線撮像では3次元顔断面画像を取得できないことや，被爆を伴うため繰り返し使用することは生理学的に困難といった問題があった．近年，核磁気共鳴画像法による顔断面の撮像が可能となり，X線撮像の問題を回避できるようになったことで，同様のアプローチによる数多くの解析が行われるようになってきた [10–22]．これらの手法のいくつかは，直管への近似により得られた音響管から，声帯からの距離を横軸に，横軸で示された位置に対応する断面積を縦軸にとった声道面積関数を作成し，声道面積関数を等価回路モデルに置き換えてこれを解析することで音声スペクトルを推定する方法を採っている．声道面積関数は，調音の様子を直感的に捉えることができ，また/a/から/i/のような音素の遷移による声道

形状の変化の様子がわかり，声道面積関数から音声スペクトルが推定できるといった，音声生成の研究にとって非常に有用な特長を持っている．しかし，声道形状の複雑さから声道形状と声道面積関数を一意に結びつける，または直線的な音響管を一意に推定することは難しいため，声道形状から声道面積関数を正確に求めることはあまり意識されていなかった．例えば，文献 [19] で示された結果では，声道面積関数から推定したホルマント周波数と，録音音声から推定したホルマント周波数との間で 10% 以上の誤差が発生し，さらに誤差のばらつきについても一貫性が無いため，ずれの要因を容易に説明できない．一方，計算機演算能力の大幅な進歩により，声道形状を 3 D モデル化し，有限要素法などの物理シミュレーションにより声道伝達特性を解析することが可能となってきている [23–26]．有限要素法では，共振周波数を高い推定精度で求めることができることが示されており [27, 28]，声道形状から声道伝達特性を正確に求めるために有用である．しかし，

- 直線的な音響管への近似を伴わないため，調音の様子を直感的に把握することが困難である．
- 音素の遷移を伴う声道伝達特性を観測する場合，遷移過程を含む 3 次元声道動画を用意する必要があるが，現在の撮像技術では動的かつ 3 次元の顔断面画像を取得することが困難であるため，音素遷移の解析には不向きである．

といった問題がある．

本研究の目的は，核磁気共鳴画像 (MRI) より得られる 3 次元声道画像から，より正確な声道面積関数を推定する手法を考案することである．子音を含む音声を解析するためには遷移について考慮する必要があるた

め、本研究では遷移を伴わない母音を解析対象とする。声道形状から正確な声道面積関数を推定するためには、声道分割方法を検討する必要がある。これまでの方法の多くは声道を平面で分割し、その断面積を並べて声道面積関数としていた。この方法では、断面の角度の設定方法により得られる声道面積関数が大きく変化したり、咽喉部のように曲がりの大きな箇所では2枚の隣接する断面が重なり合うという問題点があった。提案法では、声道分割面として、平面の代わりに等音圧面を用いる。声道内音圧の傾斜の向きは音の進行方向に近いので、それぞれの等音圧面が声道を輪状に分割する形となる。さらに、それぞれの等音圧面は互いに重なり合うことがない。これらの特性により、平面を用いるよりも良好な声道分割が期待できる。等音圧面を得るためには声道内音圧分布を得る必要がある。MRIデータから音圧分布を取得するために、提案法では有限要素法を用いる。

1.3 本論文の構成

第2章で、本研究を行うために必要な音響解析手法について述べる。第3章では、有限要素法により得られた音圧・音響インテンシティを用いて声道面積関数を推定する提案法の詳細(発表論文[2,3,4,6,7])を述べる。また、声道データの作成及びデータの正当性を検証するための声道内部の音響特性の可視化(発表論文[1])、及び、声道面積関数導出のための比較手法についても述べる。第4章では、提案法により得られる声道面積関数の評価、及び声道面積関数導出の条件を変化させたときの結果の比較(発表論文[2,3,5])について述べる。最後に第5章で本論文のまとめと今後の課題について述べる。

第2章 音響解析手法

2.1 はじめに

本章では、本研究を行うために必要な音響解析手法について述べる。2.2では、声道形状の導出に用いる顔断面画像を取得する手法である核磁気共鳴画像法(MRI)、及び研究に用いるMRIデータについて述べる。2.3では、声道データを基にして音圧・音響インテンシティの分布をシミュレーションする手法である有限要素法について述べる。2.4では、声道面積関数からホルマント周波数を計算する手法の1つである音響等価回路について述べる。2.5では、録音音声からホルマント周波数を推定するためによく用いられる線形予測法について述べる。

2.2 MRIによる顔断面の撮像

核磁気共鳴画像法(MRI: Magnetic Resonance Imaging)は、撮像対象を強い磁界の中に置き、水素原子核の状態を観測することで体内組織の断面構造を画像化する手法である。これは、体内の水分子の濃度分布を画像化することと等しい。媒体としてX線の代わりに磁界を用いることで、非侵襲的な観測が可能である。MRIを用いることで任意方向の断面画像を取得することが可能であるが、本研究で用いる顔断面画像は全て矢状面(横方向から見たときに垂直となる面)方向である。

話者がMRI装置内で仰向けになり，同期撮像信号に合わせて発音を行う，または発音時と同じ声道形状を30秒程度維持し，その間に撮像を行う．撮像時に息を吸うと声道形状，特に軟口蓋の形状が変化する．これを防止するために，撮像時間中は息止めが必要になる．1つの母音に対し，これを数十回繰り返し行うことで3次元画像を合成する．

撮像時に発声音声を録音し，MRIデータと比較検討できることが好ましいが，MRI装置で撮像する際に非常に大きな音響的かつ磁氣的ノイズが発生するため，撮像と同時に録音することは非常に困難である．従って，比較検討のための音声データは，撮像後に音声の録音を別途行うことで取得する．

本研究では成人男性2名の5母音分，すなわち計10セットの顔断面MRIデータを用いる．以降，2名の被験者をそれぞれ被験者A，被験者Bと記す．被験者Aのデータは，ATR人間情報科学研究所が公表している『ATR母音発話MRIデータ』である．一方，被験者Bのデータは本研究室で所有しているMRIデータである．2者のデータはそれぞれ異なる経路で取得したため，諸元が異なる．それぞれの諸元を表2.1に示す．両者のデータ間での諸元の違いは主に原画像解像度と，歯列充填の有無である．歯は水分を殆ど含まないため，MRIでは歯が映らない．被験者Aのデータは，母音発音時の撮像の他に口にブルーベリージュースを含んだ状態でMRI撮像を行い，水分濃度の低い部分を歯列として検出し，得られたそれぞれのMRIデータの合成により歯列充填を行っている．2名の被験者のデータを同様に扱うため，解像度が両者ともに $1.5 \times 1.5 \times 1.5\text{mm}^3$ となるように拡大・縮小処理（面積平均法）を行い，処理後の各画素を1要素として以降の処理に用いる．画素（要素）の大きさの設定の根拠は2.3で述べる．

表 2.1: M R I データの諸元

被験者	原画像解像度 (画素サイズ)	各母音の断面枚数	歯列充填
A	$0.5 \times 0.5 \times 0.5\text{mm}^3$	161 枚	あり
B	$1.0 \times 1.0 \times 2.5\text{mm}^3$	26 枚	なし

被験者Aのデータは、ATR人間情報科学研究所が公表している『ATR母音発話MRIデータ』である。一方、被験者Bのデータは本研究室で所有しているMRIデータである。2者のデータはそれぞれ異なる経路で取得したため、諸元が異なる。両者のデータ間での諸元の違いは主に原画像解像度と、歯列充填の有無である。被験者Aのデータは、口にブルーベリージュースを含んだ状態でMRI撮像を行い、水分濃度の低い部分を歯列として扱うことで歯列充填を行っている。

2.3 有限要素法

有限要素法は、有限の大きさを持つ要素に分割した対象に対して微分方程式を適用することで各種の物理現象をシミュレーションする手法である。

本項では手法の詳細を説明する。声道模型に対するホルマント周波数を分析するために、声道領域 V 及び境界条件 $S_1, S_2, \text{ or } S_3$ として式 (2.1) を適用する。

$$\begin{aligned}
 \nabla^2 p + k^2 p &= 0 && \text{in } V \\
 p &= \bar{p} && \text{on } S_1 \\
 \partial_n p &= -j\rho\omega\bar{v}_n && \text{on } S_2 \\
 \partial_n p &= -j\rho\omega Y_n p && \text{on } S_3
 \end{aligned} \tag{2.1}$$

ただし、 p は音圧、 $k = \omega/c$ は波数、 ω は角周波数、 ρ は密度、 c は音速、 \bar{p} は音圧値が一定であることを、 $\bar{v}_n = \text{m/s}$ は粒子速度の定数、 Y_n はアド

ミッタンスをそれぞれ示す．なお，添え字の n は法線方向であることを示している．例えば境界条件 S_2 ，または S_3 の面が X-Z 面に対し平行であるときは， $\partial_n p$ は $\partial_y p$ と等価である．提案法では 36，1 気圧のときの条件 $\rho = 1.14\text{kg/m}^3$, $c = 353\text{m/s}$ で解析を行う．入力面 (声帯) の境界条件として， S_2 (ノイマン条件) を適用する．入力面の粒子速度は $\bar{v}_n = 1.0\text{m/s}$ とする．また，出力放射面 (口唇) の境界条件として， S_3 を適用する．放射アドミッタンスの周波数特性は，放射面の形状が単純であっても非常に複雑な式となるために，任意の放射面の形状から正しい周波数特性を得ることは難しい．そこで，放射面を円筒形に近似したときのアドミッタンス Y_n を求め，これを出力放射面に適用する． Y_n は，式 (2.2) のように定義される [29] ．

$$Y_n = \left\{ \rho c \left(1 - \frac{J_1(2kr)}{kr} \right) + j \frac{\rho c}{2k^2 r^2} K_1(2kr) \right\}^{-1} \quad (2.2)$$

ただし， J_1 はベッセル関数， K_1 は Struve 関数， r は口唇半径である．その他の声道壁は剛壁 (S_1 : p の壁に垂直な方向の微分が 0) として扱う． S_1 は一意に決定されるが，他の条件 V, S_2, S_3 は位置や周波数により変化するため，一意には決定できない．ここで，誤差関数 R_V, R_{S_2}, R_{S_3} をそれぞれ次のように定義する．

$$R_V = \nabla^2 p^* + k^2 p^* \quad (2.3)$$

$$R_{S_2} = -\partial_n p^* - j\rho\omega c \quad (2.4)$$

$$R_{S_3} = -\partial_n p^* - j\rho\omega Y_n p^* \quad (2.5)$$

ただし， p^* は音圧の近似解である．これらの誤差を，重み付け残差法の 1 つであるガラーキン法を用いて縮小することで，目的の音圧を計算する．エレメント i の重み関数 W_i に対する誤差を減少するために用いられ

る．エレメント i における重み関数 W_i は次の方程式を満足させる関数である．

$$G_i^*(p) = \int_V W_i R_v dV + \int_{S_2} W_i R_{S_2} dS + \int_{S_3} W_i R_{S_3} dS = 0 \quad (2.6)$$

領域全体を考慮すると，次のようになる．

$$\sum_{i \in E} G_i^*(p) = 0 \quad (2.7)$$

ただし， E は声道画像を構成しているエレメントの集合である．エレメント i における近似解 p_i^* は，次の式によって近似音圧値 p_j に置き換えられる．

$$p_i^* = \sum_{j \in i} N_j p_j \quad (2.8)$$

ただし， j はエレメント i に含まれるノードで， N_j は形状関数である．これらとともに，エレメントの形状により決定される．ガラーキン法は次に示すような重み関数 W_i を用いる．

$$W_i = \sum_{j \in i} N_j \quad (2.9)$$

誤差関数，近似解 p_i^* ，及び重み関数 W_i を式 (2.7) に代入することで，次の式を得ることができる．

$$([K] + j\rho\omega[C] - \omega^2[M])\{P\} = -j\rho\omega\{F\} \quad (2.10)$$

ただし，行列 $[K]$ は音響スティフネス，行列 $[C]$ は音響減衰，行列 $[M]$ は音響質量，とそれぞれ呼ばれている．また， $\{P\}$ はそれぞれのノードの音圧ベクトル， $\{F\}$ は音響力 (acoustic force) である． $[K]$ ， $[C]$ ， $[M]$ の大きさはそれぞれ $n \times n$ (n : エレメント内のノード数) である． $[K]$ ， $[C]$ ，

$[M], \{F\}$ は次式によって表される .

$$K_{ij} = \int_V \nabla N_i \times \nabla N_j dV \quad (2.11)$$

$$C_{ij} = \int_{S_3} Y_n N_i N_j dS \quad (2.12)$$

$$M_{ij} = \int_V N_i N_j / c^2 dV \quad (2.13)$$

$$F_i = \int_{S_2} \bar{v}_n N_i dS \quad (2.14)$$

ただし, K_{ij} は行列 $[K]$ の i 行 j 列目の要素を表している . 同様に, C_{ij} は $[C]$ の要素, M_{ij} は $[M]$ の要素, F_i はベクトル $\{F\}$ の i 番目の要素である .

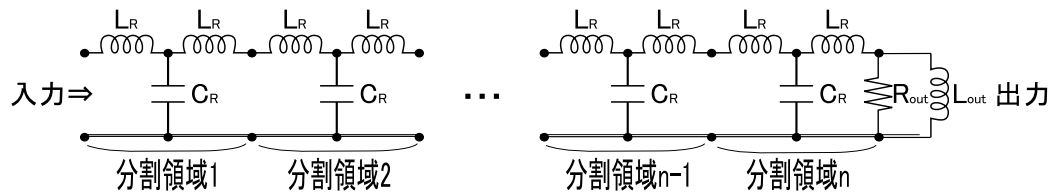
有限要素法のソルバアプリケーションとして LMS SYSNOISE ver. 5.5 を用いる . また, 画素と要素を 1 対 1 に対応させるために, それぞれの要素の形状は立方体とする . 要素の大きさによって, 解析可能な上限周波数が変化する . 各要素のサイズが大きいほど, FEM で正しく計算できる周波数の上限が低くなる . 一方, 各要素のサイズを小さく設定すると, FEM モデル 1 つあたりの要素数が多くなり, 計算時間がかかる . 使用した PC での計算可能な要素数の限界サイズである長さ 1.5mm^3 の立方体要素の場合は, 上限周波数 (誤差 0.5%) はおよそ 16kHz となることが示されている [30] . 母音の特性を示す第 1 ~ 第 3 ホルマント周波数が存在する周波数領域の上限は 3kHz であるため, 誤差は十分小さいといえる .

2.4 音響等価回路

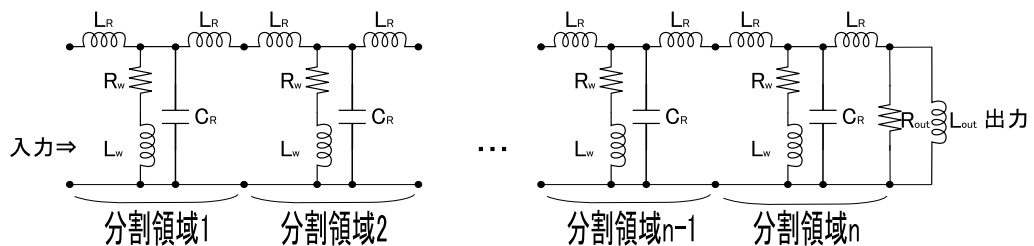
音は, 圧力や粒子速度の時間変化により生じる . それらの関係を物理式で表したとき, 電気回路における電圧, 電流などの関係と同様な式となる [29] . このとき音圧は電圧に, 粒子速度は電流にそれぞれ対応する . そのため, 音響管の周波数特性などを求める際, 電気回路に関する式を

そのまま適用することが可能である．音響解析を行うために作成する電気回路のことを音響等価回路と呼ぶ．本論文では母音音声のみを扱うため，定常状態のみを扱えばよい．その場合，声帯からの入力は常に等しい粒子速度を持つとみなせるため，音響等価回路の入力端は定電流源とすればよい．

音響管内部は分布定数回路とみなすことができるため，インダクタとコンデンサにより構成される素子を複数つなげることで音響間を模擬できる(図 2.1)．同図 (a) は声道壁からの損失を考慮しない，つまり無損失の条件に，同図 (b) は FEM において声道壁に損失を与えたときの条件にそれぞれ対応している．それぞれの素子の値は音響管の部分的な断面積と長さ，媒質の密度，媒質の音速から計算することができる． C_R , L_R , R_{out} , L_{out} の具体的な式については 3.4 に示す．



(a) 損失がない場合



(b) 損失がある場合

図 2.1: 音響等価回路

音響管内部は分布定数回路とみなすことができるため、無損失である場合 (図 (a)) はインダクタとコンデンサを図のように接続していくことで音響管を模擬できる。一方、損失がある場合 (図 (b)) は、 $R_w L_w$ 直列素子が C_R と並列に接続される回路により模擬できる。これは、FEM において声道壁に有限インピーダンス (損失) を与えたときの条件に対応している。インダクタ L_R 、及びコンデンサ C_R の値は音響管の部分的な断面積、長さ、媒質の密度、音速を基に求めることができる。また、 $R_w L_w$ 直列素子の値は生体組織の密度、および声道壁と空洞との境界の表面積から計算できる。それぞれの素子の値は、分割領域ごとに異なる。入力端は声帯を模擬しており、解析の際は定電流源を接続する。口唇の放射アドミッタンスを与えるために、出力端にはインダクタ L_{out} と抵抗 R_{out} を並列に接続する。

2.5 線形予測法によるホルマント周波数検出

音声波形からホルマント周波数を抽出する方法として、線形予測法 [31, 32] がよく用いられる。

口唇から発声される音声波 $x(t)$ は、声帯により生成される音源 $g(t)$ と、声道の伝達特性のインパルス応答 $v(t)$ の畳み込みとして表すことができる。

$$x(t) = \int_0^t g(\tau)v(t - \tau)d\tau \quad (2.15)$$

両辺をラプラス変換すると

$$X(s) = G(s) \cdot V(s) \quad (2.16)$$

となる。 $V(s)$ は声道の伝達関数で、 s 平面上の極とゼロを用いると次のように表される。

$$V(s) = A \frac{\prod_{i=1}^{N/2} (s - n_i)(s - n_i^*)}{\prod_{i=1}^{P/2} (s - p_i)(s - p_i^*)} R(s) \quad (2.17)$$

ここで、 A は振幅係数、 n_i, n_i^* は s 平面上のゼロ点、 p_i, p_i^* は s 平面上の極、 N はゼロ点の数、 P は極の数、上付きの $*$ は共役複素数、 $R(s)$ は放射特性をそれぞれ表す。放射特性は $R(s) \simeq s$ と近似できる。音声の解析対象を母音に制限すると、定常状態のみを扱えばよいので、 s を $z (= \exp(j\omega T))$

ただし、 T : サンプリング周期) と置き換えることができる。すなわち、

$$V(z) = A \frac{\prod_{i=1}^{N/2} (z - \nu_i)(z - \nu_i^*)}{\prod_{i=1}^{P/2} (z - \gamma_i)(z - \gamma_i^*)} z \quad (2.18)$$

$$= Az^{-(P-N-1)} \frac{\prod_{i=1}^{N/2} (1 - \nu_i z^{-1})(1 - \nu_i^* z^{-1})}{\prod_{i=1}^{P/2} (1 - \gamma_i z^{-1})(1 - \gamma_i^* z^{-1})} \quad (2.19)$$

となる．ここで， ν_i, ν_i^* は z 平面上のゼロ点， γ_i, γ_i^* は z 平面上の極である．母音の特徴を決定するものは主にホルマント周波数であり，極とホルマント周波数が対応しているため，ホルマント周波数を推定するためには式の分母のみを考慮すればよいことになる．つまり， $V(z)$ はより簡単に，

$$V(z) = \frac{A'}{\prod_{i=1}^{P/2} (1 - \gamma_i z^{-1})(1 - \gamma_i^* z^{-1})} \quad (2.20)$$

$$= \frac{A'}{1 + \alpha_1 z^{-1} + \alpha_2 z^{-2} + \cdots + \alpha_P z^{-P}} \quad (2.21)$$

と表現することができる． $Z^{-1}(V(z)z^{-i}) = x_{n-i}$ ，すなわち逆 Z 変換を施すと，この式は

$$x_n = A' - \alpha_1 x_{n-1} - \alpha_2 x_{n-2} - \cdots - \alpha_P x_{n-P} \quad (2.22)$$

となる．これは， n 番目のサンプル値 x_n がそれ以前の $P - 1$ 個のサンプル値の線形結合で予測できることを示している．この $\alpha_1, \alpha_2, \cdots$ のことを線形予測係数と呼ぶ．線形予測係数 $\alpha_k (k = 1, 2, \cdots, P)$ の推定ができれば，式 (2.21) の分母を因数分解することで， $\gamma_l (l = 1, 2, \cdots, P/2)$ を計算でき，結果としてホルマント周波数を求めることができる．

第3章 声道断面積関数の導出

3.1 はじめに

本章では，第2章で述べた音響解析手法を利用して，声道面積関数をより正確に導出するために提案する手法について述べる．以下に手順を示す．

- (1) MRI で撮像した画像から声道領域を抽出する．
- (2) 声道領域画像を基に有限要素モデルを作成する．
- (3) 2.3 に示した手法を用いて音響解析を行う．
- (4) 声道領域画像と音圧，音響インテンシティ分布を基に声道分割を行う．
- (5) それぞれの分割領域の体積，長さから声道面積関数を導出する．
- (6) 2.4 に示した方法を用いて声道面積関数の周波数特性を計算する．

これらの詳細は以下の節で述べる．

3.2 前処理及び可視化

3次元データは立体的な構造をしているのに対し，ディスプレイやマウスは平面的なデータを操作する装置であるために，3次元データを2次

元データと同様に扱うためには工夫を要する．ここでは，Windows GDI 及び OpenGL を用いて，3次元声道解析のための描画，編集システムを構築した．

3.2.1 声道領域の抽出

MRI は，撮像対象の水素原子の濃度分布を示した画像データである．この画像データから声帯及び口唇の位置や声道領域内外の境界といったより高次の情報を得るためには，画像データに対し解析を行うか，画像を目視で確認しながら手作業で情報を付加する必要がある．本研究では成人男性2名の5母音分，すなわち計10セットの声道データの解析を行う．拡大・縮小処理後の断面画像の枚数は1母音あたり54枚（被験者A）または44枚（被験者B），その中で声道を含む画像が約20枚ずつ存在する．10セットで約200枚の2次元画像データを処理するため，特に手作業が必要な部分において効率化が必須となる．本項ではその方法を述べる．

空洞部と生体組織部の分離

MRI データでは，画像中の輝度値の大小で生体組織の密度を表現している．従って，輝度値に対し適切な閾値を設定できれば，空洞部と生体組織部の分離が実現できる．そして分離後の2値画像に対しエッジ検出を施すことで，声道領域と声道外を分離する境界の大部分を得ることができる．本システムでは，空洞（空気）と生体組織を分離するために判別分析法 [33] を用いて閾値を計算する．判別分析法は，画像の輝度値のヒストグラム $L(t)$ の形状を基に閾値を決定する手法である．この手法では，

ヒストグラムが双峰性を持っている場合，2つの山を互いに分離する位置が閾値となる．また，双峰性の検出を直接行う方法ではないため，双峰性を持たない場合でも自動的に閾値を決定できるという特長を有する．

本研究で用いる MRI データは空気が黒く，生体組織が白い 8bit グレイスケール画像である．すなわち，輝度値が 0 に近いほど空洞である可能性が高く，255 に近いほど生体組織である可能性が高い．空洞となる群 (以下，群 1) が輝度値 $0 \sim k - 1$ ，生体組織となる群 (以下，群 2) が輝度値 $k \sim 255$ とすると，群 1 の画素数は $n_1 = \sum_{t=0}^{k-1} L(t)$ ，平均輝度値は $\bar{k}_1 = \frac{\sum_{t=0}^{k-1} tL(t)}{n_1}$ ，群 2 の画素数は $n_2 = \sum_{t=k}^{255} L(t)$ ，平均輝度値は $\bar{k}_2 = \frac{\sum_{t=k}^{255} tL(t)}{n_2}$ ，全画素の平均輝度値は $\bar{k} = (n_1\bar{k}_1 + n_2\bar{k}_2)/n$ となる．ただし， $n = n_1 + n_2$ である．それらの結果からそれぞれの群の分散 σ_1^2, σ_2^2 を求める．それぞれの分散は次式で求められる．

$$\sigma_1^2 = \frac{\sum_{t=0}^{k-1} n_t \cdot (t - \bar{k}_1)^2}{n_1} \quad (3.1)$$

$$\sigma_2^2 = \frac{\sum_{t=k}^{255} n_t \cdot (t - \bar{k}_2)^2}{n_2} \quad (3.2)$$

次に，群内分散 σ_W^2 及び群間分散 σ_B^2 を求める．

$$\sigma_W^2 = \frac{n_1\sigma_1^2 + n_2\sigma_2^2}{n} \quad (3.3)$$

$$\sigma_B^2 = \frac{n_1(\bar{k}_1 - \bar{k})^2 + n_2(\bar{k}_2 - \bar{k})^2}{n} \quad (3.4)$$

分散比 σ_B^2/σ_W^2 を $0 \leq k \leq 255$ のそれぞれについて計算し，最も大きくなるときの k を求める．その k が閾値となる．ただし， $\sigma_T^2 = \sigma_B^2 + \sigma_W^2$ (σ_T^2 は全分散) の関係を用いて，

$$\frac{\sigma_B^2}{\sigma_W^2} = \frac{\sigma_B^2}{\sigma_T^2 - \sigma_B^2} \quad (3.5)$$

$$= \frac{\sigma_T^2}{\sigma_T^2 - \sigma_B^2} - 1 \quad (3.6)$$

と導出できる． σ_T^2 は k に依存しないため，分散比が最大となる k を求めることは， σ_B^2 が最大となる k を求めることと等価となる．従って，実際はそれぞれの k について， σ_B^2 のみを計算すればよい．

この閾値を用いて2値化を行い，さらに8近傍エッジオペレータによるエッジ検出を行うことで2値エッジ画像が得られる(図3.1)．このエッジを生体組織部と空洞部の境界として用いる．

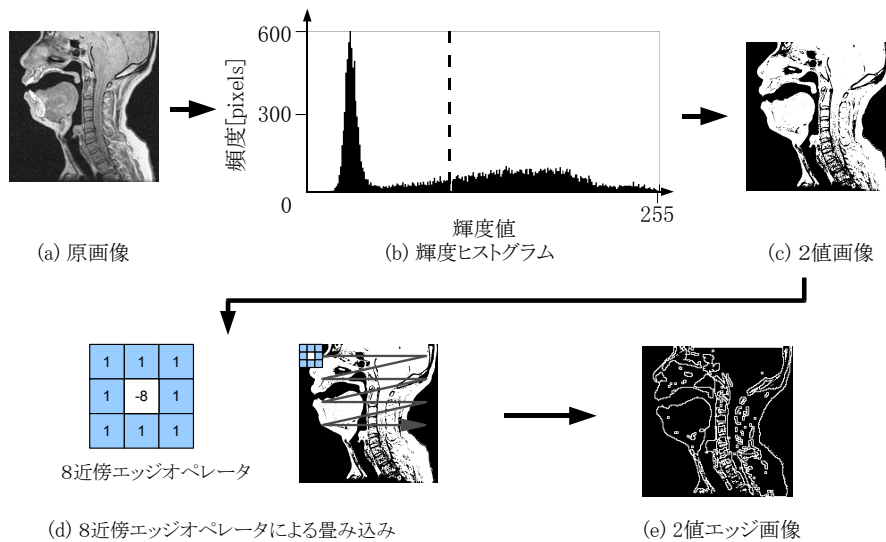


図 3.1: 判別分析法を用いた空洞部と生体組織部の分離

図 (a) の顔断面画像から得られる輝度ヒストグラム (b) に対して，判別分析法を適用して得られる閾値は 104(図 (b) 縦の点線) である．この閾値よりも大きい輝度を持つ領域を白，小さい輝度を持つ領域を黒で表現すると，図 (c) のような2値画像が得られる．さらに，8近傍エッジオペレータのラスタースキャンによる畳み込み(図 (d))を行うことで図 (e) のエッジ画像が得られる．

声帯面，及び口唇面の指定

MRI データから声道領域を決定するためには，輝度情報としてほとんど含まれない声帯面と，全く含まれない口唇面をそれぞれ指定する必要がある．声帯面，及び口唇面を高精度に検出するアルゴリズムを考案することは容易ではないため，本システムでは手作業による指定を支援する方法を採用した．声帯面，口唇面指定の手順を以下に示す．

(1) 矢状面方向の断面画像を表示する．

ユーザはキー操作により断面の位置を前後に移動することができる．表示する断面画像として，原画像と 8 近傍エッジ画像のどちらかを選択できる．

(2) 声帯面を指定する．

ユーザが声帯付近の 1 点を断面画像に対して指定すると，システムはその点を通り地面に対し平行な面を声道内に作成する．その面が入力面となる．

(3) 口唇面を指定する．

ユーザが口唇の左上と右下にそれぞれ 1 点ずつ指定すると，システムはその 2 点を通る線分を対角線に持つ長方形を作成し，その長方形の内部の形状から口唇の位置を検出し，上唇と下唇を直線で結ぶ．その面が口唇面となる．

声帯面，口唇面設定例を図 3.2 に示す．



(a) 断面画像 (エッジ画像)



(b) 声帯面設定例



(c) 口唇面設定例



(d) 声道領域設定例

図 3.2: 声帯面, 口唇面の設定

(a) は矢状面方向の断面画像の例である。(b) は声帯面を、(c) は口唇面をそれぞれ指定している例である。(b) は声帯の左側をマウスで指定すると、指定した点から右側に向かって直線を引く。この線と声道内領域が接する部分が入力面となる。(c) は、上唇の左側と下唇の右側を指定すると、その2点を結ぶ線分を対角線に持つ長方形を作成し、その矩形領域内の形状から口唇を検出し、直線を引く。その線と声道内領域が接する部分が口唇面となる。(d) の灰色の部分が声道領域となる。声帯面、口唇面、および鼻腔への接続部を切断した後、声道領域内の1点を指定すると、その点を起点として塗りつぶしを行う。塗りつぶされた領域が声道領域となる。この処理を全ての断面に対し行うことで、3次元画像を構築することができる。

3.2.2 中心線の検出

声道の中心線は，声道画像に対し細線化を施し，細線化した画像から分岐を除去することで求める．2次元画像に対する細線化画像は広く知られている Hilditch の細線化アルゴリズム [34] を用いることで非常に良好に導出できる．しかし，3次元画像に対しては，任意の画像に対応する細線化画像を求めることは困難であることが知られている [35]．本システムでは声道の中心線を検出するために，3次元画像に対し比較的良好な細線化を施すことができるアルゴリズム [36] を採用する．このアルゴリズムは，距離変換の性質を利用している．距離変換とは，2値画像（0：背景画素，1：対象画素）の各対象画素について，最も距離が近い背景画素との距離を求め，その値と画素値とする処理を表す．2値画像に対して距離変換を施すと，対象領域の中心の画素値が極大値となる．従って，距離変換後の画像の極大値となる画素を検出し，その位置情報などを利用して画像の中心を計算できる．距離変換においてよく用いられる距離尺度として，マンハッタン距離，チェスボード距離，ユークリッド（2乗）距離などが挙げられる（図 3.3）．このアルゴリズムは，ユークリッド 2 乗距離による距離変換を用いている．距離変換の例を図 3.4 に示す．アルゴリズムの詳細を付録 A に示す．

図 3.5(b) に示したように，声道 2 値画像から得られる細線化画像には声帯から口唇までを繋ぐ線の他に，枝状の線も含まれている．細線化画像から中心線を求めるためには，これら枝状の線を除く必要がある．手順を以下に示す．

- 細線化画像のそれぞれの非 0 画素の隣接画素同士を 1 つの群として，ラベリングを行う．

- 画素数が最大となる群に属する画素のうち，声帯部に最も近いノードを探索開始点，放射部に最も近いノードを探索終了点としてダイクストラ法による最短経路探索を行う．

より確実に中心となる線を検出できるようにするために，3次元声道画像全体に対してこの処理を行う代わりに，図 3.5(c) に示した領域のみを残した声道画像に対して中心線検出処理を行う．図 3.5(d) はこれらの手順によって導出される中心線の例を示している．

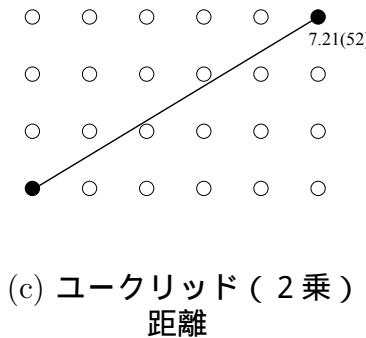
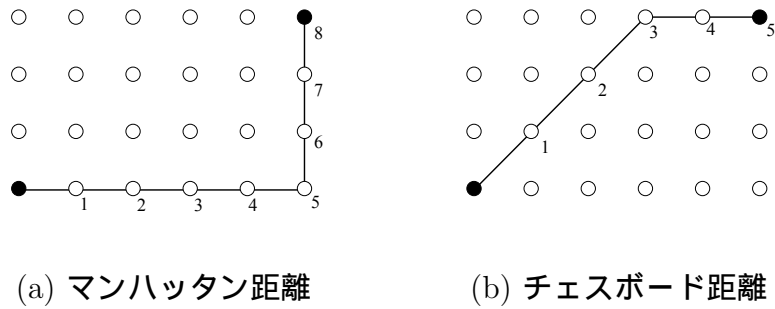


図 3.3: 距離尺度

図中の線は左下の格子点から右上の格子点までのそれぞれの距離尺度における最短経路を示している。格子点の脇の数字は、左下の格子点からその格子点までの距離を示している。(c)において括弧外の数字はユークリッド距離を、括弧内の数字はユークリッド2乗距離をそれぞれ示している。(a)は、上下左右の格子点との距離が1となり、斜め方向の隣接格子点との距離は2となる。従って、左下から右上までの距離は8となる。(b)は、上下左右に加えて斜め方向の隣接格子点との距離が1となる。従って、左下から右上までの距離は5となる。(a)及び(b)においては、最短経路は複数存在する。(c)は、2点を線分で結んだときの線分の長さが左下から右上までの距離となる。従って、ユークリッド2乗距離は52、ユークリッド距離は7.21となる。

1	1	1	1	1
1	2	2	2	1
1	2	2	1	
1	2	3	2	1
1	1	2	1	1
		1		

(a) マンハッタン距離変換

1	1	1	1	1
1	2	2	1	1
1	2	2	1	
1	2	2	1	1
1	1	1	1	1
		1		

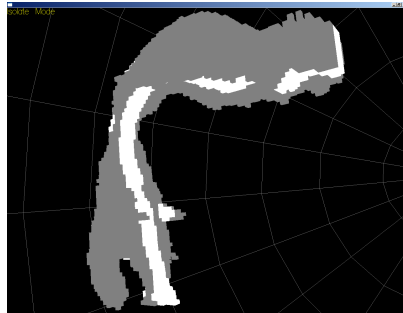
(b) チェスボード距離変換

1	1	1	1	1
1	4	4	2	1
1	4	4	1	
1	4	5	2	1
1	1	2	1	1
		1		

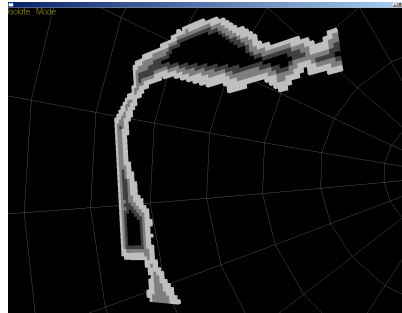
(c) ユークリッド2乗距離変換

図 3.4: 距離変換の例

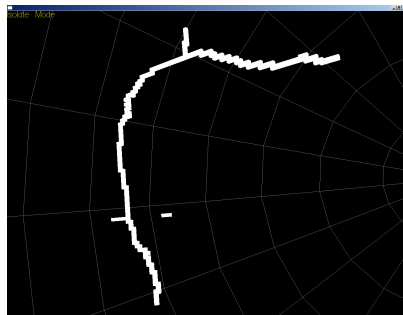
灰色の格子は対象画素を表す．格子中の数字は，その格子と最も近い背景との距離を表す．(c)において，それぞれの数値の平方根をとるとユークリッド距離変換となる．背景と隣接している画素の値は小さく，中央に近い画素の値は大きくなるのがわかる．



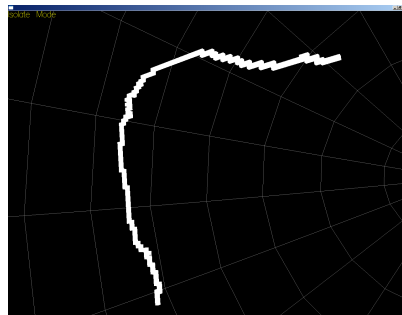
(a) 中心線検出に用いる領域



(b) ユークリッド 2 乗距離変換



(c) 細線化



(d) 中心線

図 3.5: 中心線の検出

(a) 中心線をより良好に検出するために，声道領域 (図中灰色の領域) のうち，正中矢状面付近の領域 (図中白の領域) のみを用いて処理を行う。(b) は 3 次元声道画像に対しユークリッド 2 乗距離変換を行ったときの正中矢状面を示している。白い部分は値が小さく，黒い部分は値が大きいことを示している。声道領域の中心が最も黒いことがわかる。(c) は，(b) の距離変換画像を基に細線化を行ったものである。(d) は，(c) の各点を経路として，入力点から出力点までの最短経路を求めたものである。

3.2.3 各種音響パラメータの可視化

本システムは、SYSNOISE によって計算された音圧分布、及び音響インテンシティといった各種音響パラメータを読み込み、それらを表示することができる (図 3.6)。

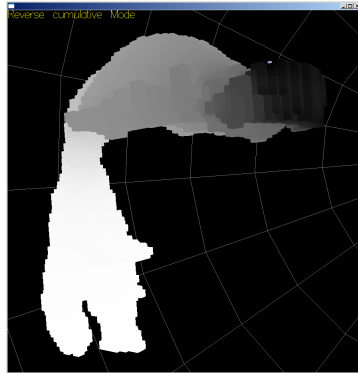
図 3.6(a), (b) は有限要素法より得られる各ノードの音圧を、音圧の最大値を白、最小値を黒として描画した例である。また、音圧分布の様相をよりよく表現するために、音圧分布から等音圧面を検出する。エレメントの大きさは 1 辺あたり 1.5mm であるが、等音圧面を求めるには大き過ぎるため、細分化を行う。顔断面画像より作成されたメッシュをさらに横方向、縦方向にそれぞれ 24 分割、エレメントの数にして 13824 分割する。細分化によって新たに作成されるノードの音圧、音響インテンシティは分割前のノードのそれぞれの値を線形補間することで求める。1 次元データ列 p_k における線形補間は次式によって実現できる。

$$p'_s = \frac{1}{n} \{sp_i + (1-s)p_{i+1}\} \quad (3.7)$$

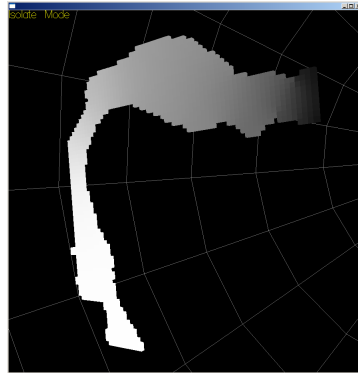
ただし、 p_k は補間前の k 番目の値、 p'_s は $p_i \sim p_{i+1}$ を n 個に分割したときの t 番目の値である。これを 3 次元に拡張すると次式のようなになる。

$$p'_{stu} = \frac{1}{n_x n_y n_z} \left\{ stup_{ijk} + (1-s)tup_{(i+1)jk} + s(1-t)up_{i(j+1)k} + (1-s)(1-t)up_{(i+1)(j+1)k} + st(1-u)p_{ij(k+1)} + (1-s)t(1-u)p_{(i+1)j(k+1)} + s(1-t)(1-u)p_{i(j+1)(k+1)} + (1-s)(1-t)(1-u)p_{(i+1)(j+1)(k+1)} \right\} \quad (3.8)$$

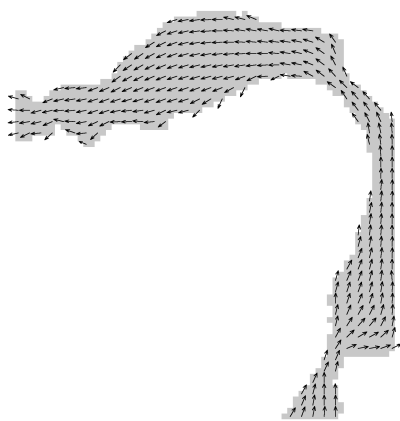
ただし、 n_x, n_y, n_z はそれぞれ横、縦、奥行き方向の分割数で、 s は横軸方向、 t は縦軸方向、 u は奥行き方向の配列番号をそれぞれ表す。



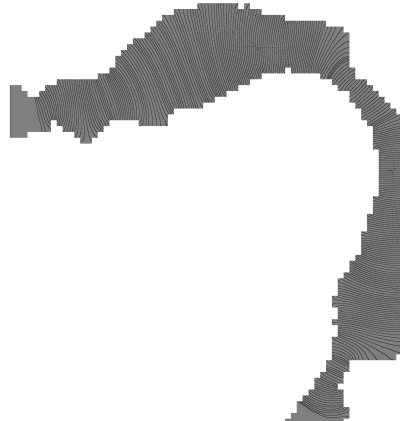
(a) 音圧分布



(b) 正中矢状面の音圧分布



(c) 正中矢状面の音響インテンスィティ分布



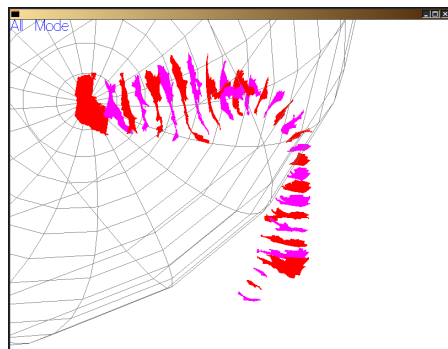
(d) 等音圧面

図 3.6: 音響パラメータの可視化

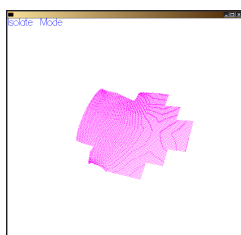
有限要素法 (SYSNOISE) で求めたそれぞれのノードの音圧を基に、それぞれのボクセルに対する音圧分布を求める。(a) は声道とそれ以外の領域との境界の音圧分布を、(b) は正中矢状面の声道領域の音圧分布をそれぞれ示している。(c) は (b) と同様に正中矢状面の声道領域の音響インテンスィティの分布を示している。(d) は音圧分布から等音圧面を検出した例である。

3.2.4 等音圧面の可視化

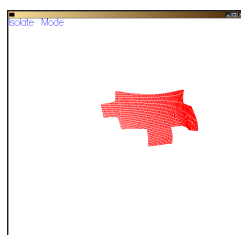
本システムは、等音圧面を3次的に表示することができる。図3.7は、計算により得られた等音圧面を一定間隔ごとに計30枚表示した例である。



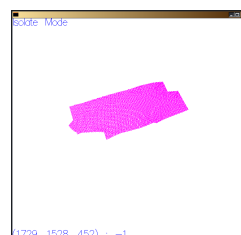
(a) 全体像



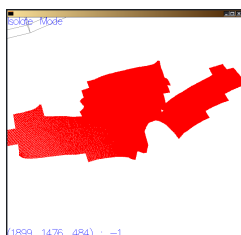
(b) 1枚目



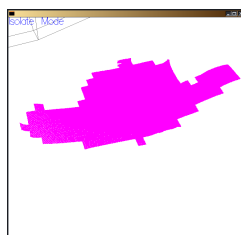
(c) 2枚目



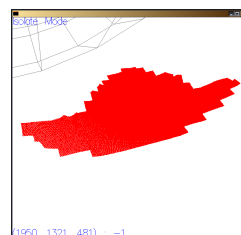
(d) 3枚目



(d) 4枚目

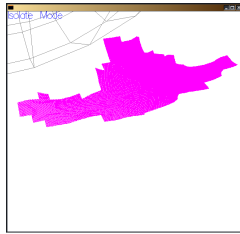


(e) 5枚目

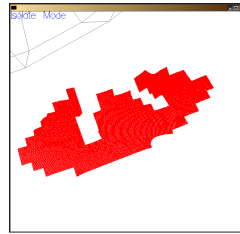


(f) 6枚目

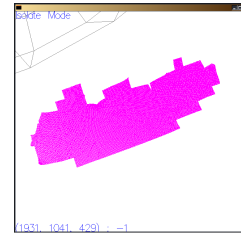
(次ページへ続く)



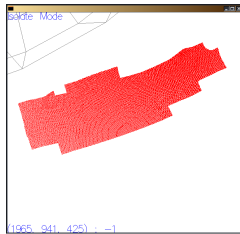
(g) 7 枚目



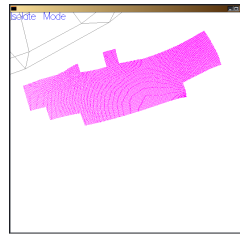
(h) 8 枚目



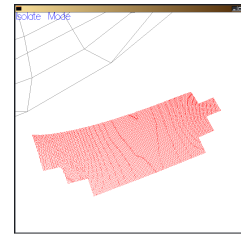
(i) 9 枚目



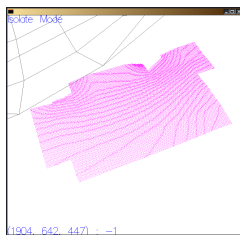
(j) 10 枚目



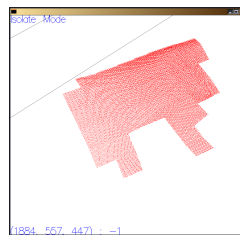
(k) 11 枚目



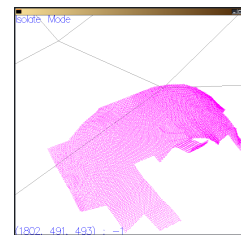
(l) 12 枚目



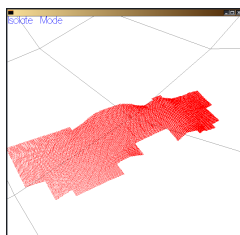
(m) 13 枚目



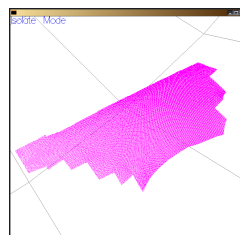
(n) 14 枚目



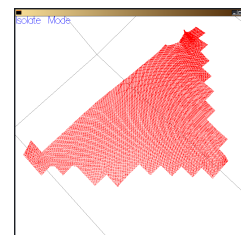
(o) 15 枚目



(p) 16 枚目



(q) 17 枚目



(r) 18 枚目

(次ページへ続く)

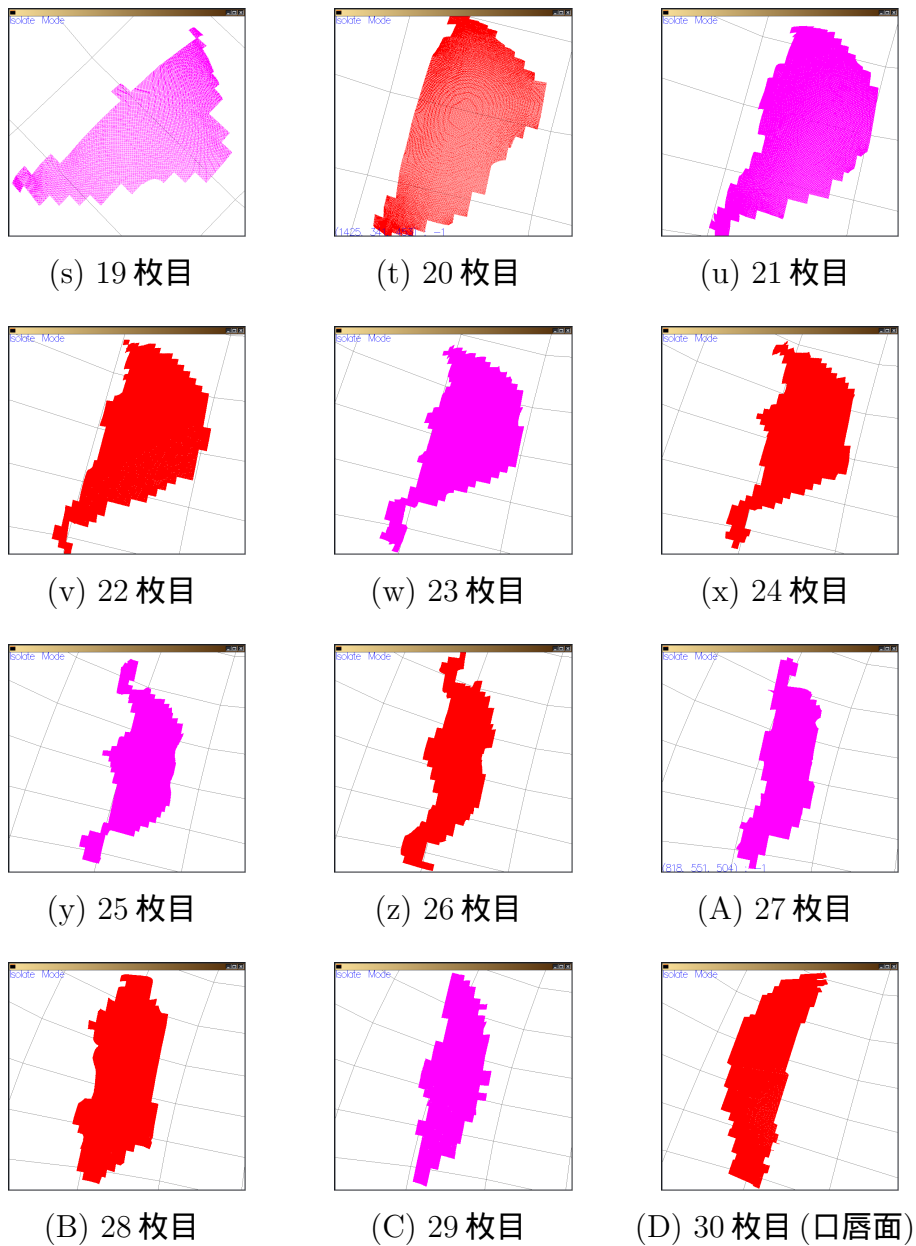


図 3.7: 等音圧面表示例

被験者 A の /a/ 発音時の声道データから計算した等音圧面を 3 次的に表示した例である。(a) は全体像を、(b) 以降は等音圧面を声帯に近い順に 1 枚ずつ表示した例である。(D) は口唇面である。

3.3 声道領域分割

声道領域を等音圧面で分割することで，咽喉部のように急に曲がる箇所においても隣接する2つの分割面が互いに重なることなく，自然な分割が可能となる．声道領域を入力から出力まで等間隔に分割するためには，声道長を求める必要がある．声道形状は複雑であるために，声道長を一意に決定することはできないが，例えば細かい間隔で等音圧面を求め，隣接する2つの等音圧面間の距離が計算できるとすると，その総和を求めることで声道長を得ることができる．しかし，2つの曲面間の距離を一意に決定することはできない．提案法では，音のエネルギーの流れを利用して2つの等音圧面間の距離の決定を試みる．音のエネルギーの流れは，音響インテンシティの実部から求めることができる．

図3.8に処理の流れを示す．まず，2.3で示した方法を用いて音圧，及び音響インテンシティの分布をそれぞれ求める．また，等音圧面検出の起点を作成するために，3.2.2で示した方法を用いて声道の中心線を検出する．等音圧面をできる限り高精度に検出するために，音圧，および音響インテンシティの分布を線形補間(式(3.7))を用いてより細かく求める．中心線を一定間隔に区切ることで作成されるそれぞれの点を起点として，その点と等しい音圧となる面を検出する．そして，2つの隣接する等音圧面間の距離を計算し，セクション分割を行う．それぞれのセクションの長さや体積から面積を求め，それを入力から出力まで並べることで面積関数を導出する．等音圧面を検出するために，声道領域の2値画像に対し3.2.2の方法を適用して中心線を求める．求めた中心線を0.5mm間隔に分割し，それぞれの分割点を通る等音圧面を入力から出力まで検出する．そして，2つの隣接する等音圧面と声道壁で囲まれた領域（以降，

サブセクションと呼ぶ)の厚さを求める．

- (1) 声帯に近い側の等音圧面 C_a に属するそれぞれのノード s_i に対する音響インテンシティの実部を正規化したベクトル I を計算する．ただし, i は C_a 中のノード番号．
- (2) C_a の次の等音圧面 C_b に属するそれぞれのノード s_j と s_i からベクトル $E_{ij} = L(s_i) - L(s_j)$ を計算する． $L(s)$ はノード s の3次元座標を表す．
- (3) 等音圧面 C_b 中のノードのうち, ベクトル E_{ix} が最も I に近くなるノード s_x を選択する．
- (4) s_i と s_x との距離 d_i を計算する．
- (5) 等音圧面 C_a 中の全てのノードについて d_i を計算し, それらを平均したものを $C_a - C_b$ 間の距離 $\text{thickness}(C_a, C_b)$ とする．

口唇面は一般的に等音圧面とはならないために, 最後のサブセクション, つまり, 最も口唇側に近いサブセクションの厚さの補正を行うことが望ましい．最後のサブセクションに関しては, 最後のサブセクションの体積 V_f を最後の等音圧面の面積 A_{f-1} で割ることで, 厚さを良好に推定できることが解っている [23]． A_{f-1} は, $V_{f-1}/\text{thickness}(C_{f-2}, C_{f-1})$ とすることで導出できる．声道長はサブセクションの厚さを合計することで導出できる．

セクション分割する際に, それぞれのセクションの長さが完全に等間隔となるように分割できれば望ましい．しかし, セクションを構成するそれぞれのサブセクションの厚さが厳密には異なるために, 全てのセク

クションの長さを完全に均一にすることはできない．そこで，それぞれのセクションの長さができる限り等間隔となるようにセクション境界となる等音圧面を選択する．つまり，分割後のセクション数を n ，声道長を l_v としたとき，完全に等間隔に分割できた場合のそれぞれのセクションの長さ $l = l_v/n$ を基準にして分割面にする等音圧面を選択する (図 3.9)．

それぞれのセクションの長さは声道長の導出方法と同様に，そのセクションに含まれるサブセクションの厚さを合計することで導出する．セクション k の面積はそのセクションの体積からセクションの長さを割ることで導出できる．セクションの体積は，そのセクションを構成しているエレメントの数に，エレメント 1 つあたりの体積 ($1/16\text{mm} \times 1/16\text{mm} \times 1/16\text{mm}$) を掛けることで導出できる．声道面積関数は，横軸が入力からの距離，縦軸が断面積となるように入力側のセクションから出力側のセクションまでの面積を並べることで得ることができる．本研究ではセクション数は 30 とする．すなわち，入力側はセクション 1，出力側はセクション 30 となる．

3.4 音響等価回路への適用

前項で求めることができる声道面積関数を，2.4 で述べた音響等価回路に適用することで，周波数特性を得ることができる．周波数解像度を十分に確保するために，声道 1 セクションの長さを，6 領域に等分割し，等価回路に適用する．その場合，図 2.1 中のそれぞれの分割領域の値は次式により与えられる．

$$L_{R(i-1) \times 6+j} = \frac{\rho}{A_i} \times l_i/6 \quad (j = 1, 2, 3, 4, 5, 6) \quad (3.9)$$

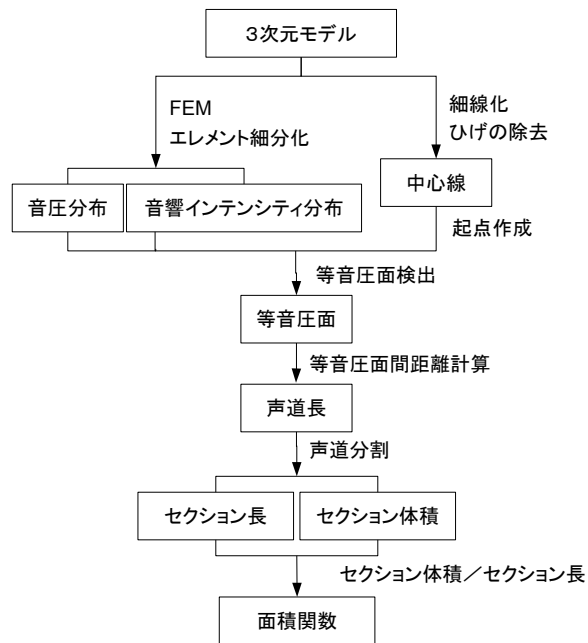


図 3.8: 処理の流れ

まず、3次元モデルをFEMで音響解析し音圧、および音響インテンシティの分布を求め、線形補間による細分化を施す。また、等音圧面検出の起点を作成するために中心線も検出する。中心線を一定間隔に区切ることで作成されるそれぞれの点を起点として、その点と等しい音圧となる面を検出する。そして、2つの隣接する等音圧面間の距離を計算し、セクション分割を行う。それぞれのセクションの長さから体積を求め、それを入力から出力まで並べることによって面積関数を導出する。

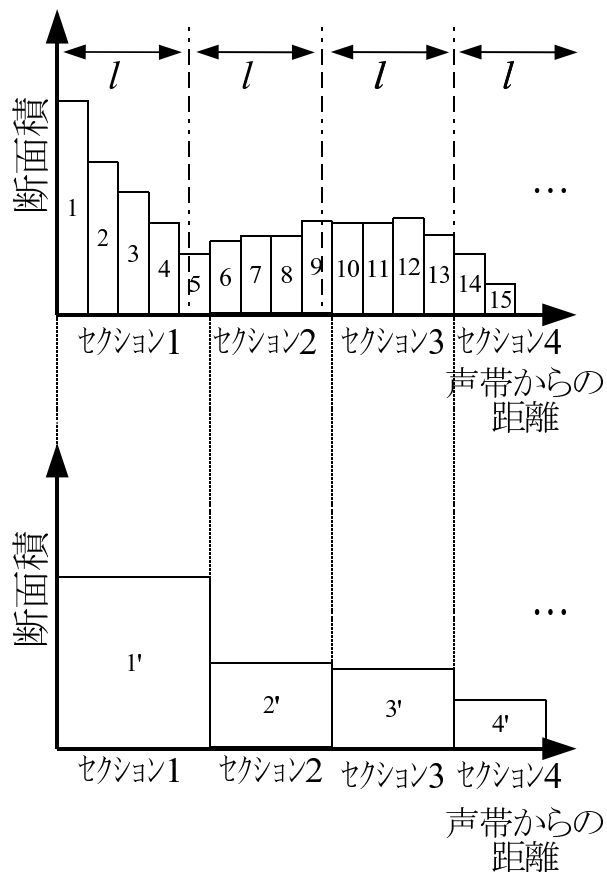


図 3.9: セクション境界面選択方法

上側のグラフの長方形はサブセクションを，下側のグラフの長方形はセクションをそれぞれ示している．上下のグラフ間をつなぐ破線がセクション境界面である．それぞれのセクションの長さができる限り等間隔となるようにセクション境界面を選択する．

$$C_{R(i-1) \times 6+j} = \frac{A_i}{\rho c^2} \times l_i/6 \quad (j = 1, 2, 3, 4, 5, 6) \quad (3.10)$$

$$L_{w(i-1) \times 6+j} = 14/(S_i/6) \quad (j = 1, 2, 3, 4, 5, 6) \quad (3.11)$$

$$R_{w(i-1) \times 6+j} = 20000/(S_i/6) \quad (j = 1, 2, 3, 4, 5, 6) \quad (3.12)$$

$$L_{out} = \frac{8\rho}{3\pi \cdot \pi r} \quad (3.13)$$

$$R_{out} = \frac{128\rho c}{9\pi^2 \cdot \pi r^2} \quad (3.14)$$

ただし, ρ は密度, c は音速, i は声道セクション番号 (1 ~ 30) である. 本研究ではそれぞれ, $\rho = 1.14\text{kg/m}^3$, $c = 353\text{m/s}$ とする. l_i は i 番目の声道セクションの長さを, A_i は i 番目の声道セクションの断面積をそれぞれ表している. また, L_R , および C_R についている数字は, 回路側の分割領域の番号を表している (L_{R1} : 分割領域 1 の L_R).

L_w の被除数の 20000 は, 機械抵抗 (単位: $\text{kg} \cdot \text{s}^{-1} \cdot \text{m}^{-2}$) を, R_w の被除数の 14 は生体組織 (主に頬) 単位面積あたりの重量 (単位: $\text{kg} \cdot \text{m}^{-2}$) をそれぞれ表している. これらの値は, 文献 [37] の実験より得られた結果に基づいて決定した.

3.5 比較手法

提案法の有効性を確認するため, 声道を等音圧面の代わりに平面で分割する手法を採用し, 声道面積関数の推定を行う. 提案法とこの手法との比較を行うことで, 等音圧面を用いる効果を評価することができる. 手順を以下に示す.

- (1) 中心線を 3.2.2 の方法を用いて検出する.
- (2) 中心線を B スプラインを用いて補間する.

- (3) 中心線を 30 個に等分割する分割点を計算する．その結果 29 個の分割点を得られる．
- (4) (3) で求めた 29 個の分割点のそれぞれについて，分割点を通る中心線に対して垂直な面を計算する．
- (5) 断面積が最小となるように断面の角度を ± 5 度の範囲で調整する．
- (6) 隣接する面が声道内で互いに交差しないように手動で調整する．
- (7) 2 つの隣接する断面で囲まれた声道領域からセクション体積を計算する．
- (8) 2 つの隣接する断面の重心のユークリッド 2 乗距離を求めることでセクション長を計算する．

最終セクションの長さに関しては，提案法と同様，(8) の方法を用いる代わりに最終セクションの体積と 1 つ前のセクションの断面積から計算する．図 3.10 はこの手法を用いて声道を分割した例を示している．

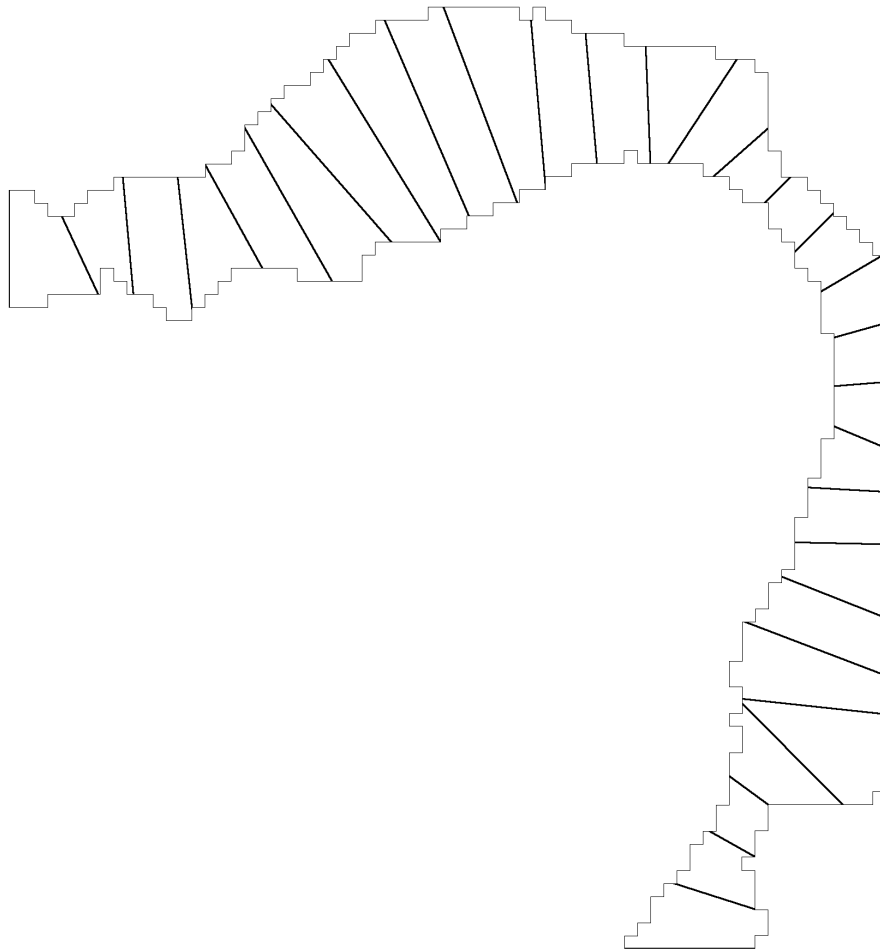


図 3.10: 比較手法を用いて声道分割を行った例

/a/を発声している3次元画像に対し、比較手法を適用して声道分割を行った例を示している。声道領域内に描かれている黒線は2つの隣接するセクションの境界面である。

第4章 結果と考察

4.1 3次元声道データ

図4.1～4.5に被験者A,およびBの顔断面画像から抽出した3次元声道データを示す.また,MRI撮像後に別途録音した日本語5母音から2.5で示した方法を用いてホルマント周波数を計算した結果を表4.1に示す.

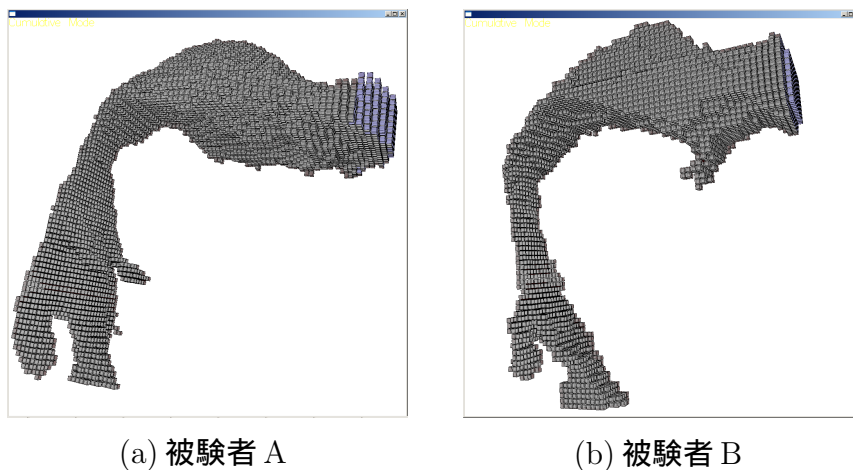
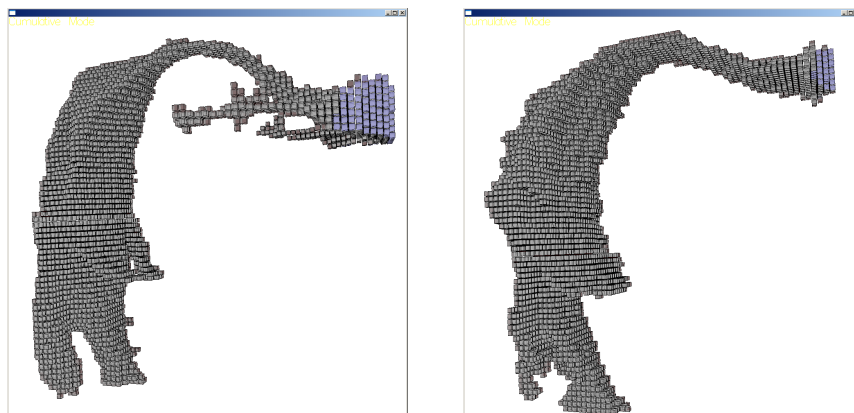


図 4.1: /a/発声時の3次元画像データ

被験者A,およびBの/a/発声時の3次元MRIデータから抽出した声道領域を3次元レンダリングしたものを示している.

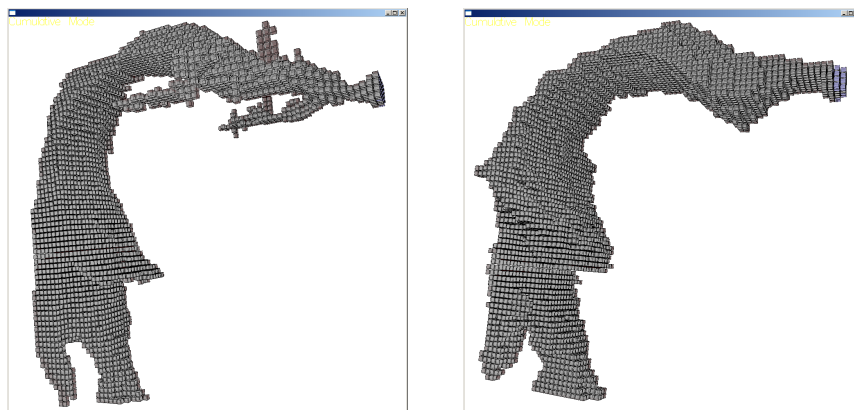


(a) 被験者 A

(b) 被験者 B

図 4.2: /i/発声時の 3 次元画像データ

被験者 A, および B の /i/ 発声時の 3 次元 MRI データから抽出した声道領域を 3 次元レンダリングしたものを示している。

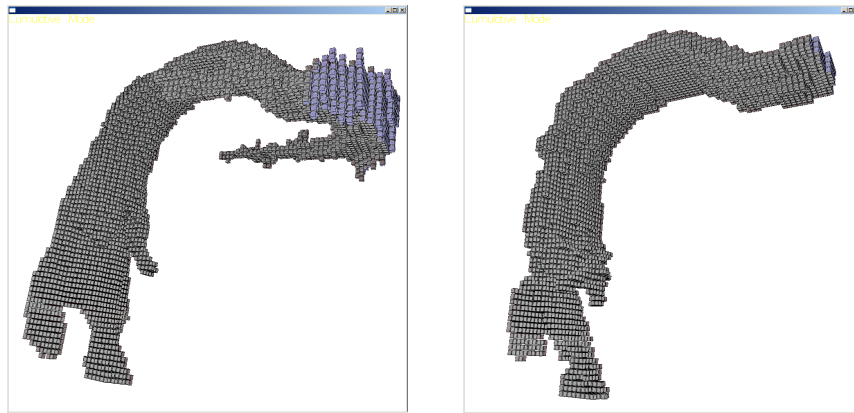


(a) 被験者 A

(b) 被験者 B

図 4.3: /u/発声時の 3 次元画像データ

被験者 A, および B の /u/ 発声時の 3 次元 MRI データから抽出した声道領域を 3 次元レンダリングしたものを示している。

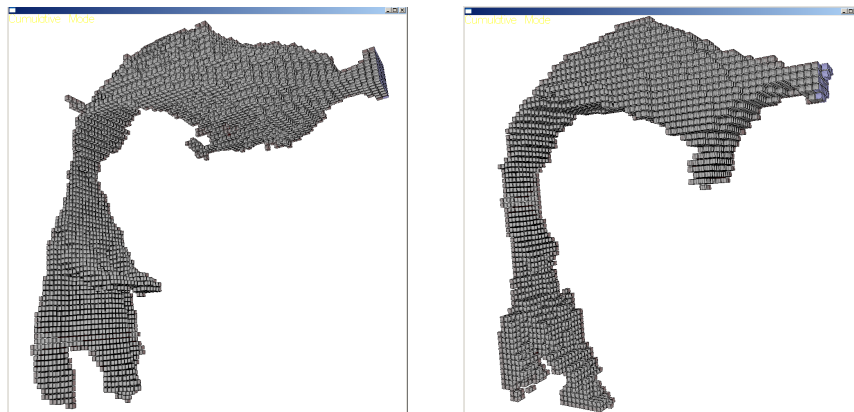


(a) 被験者 A

(b) 被験者 B

図 4.4: /e/発声時の 3 次元画像データ

被験者 A, および B の /e/ 発声時の 3 次元 MRI データから抽出した声道領域を 3 次元レンダリングしたものを示している.



(a) 被験者 A

(b) 被験者 B

図 4.5: /o/発声時の 3 次元画像データ

被験者 A, および B の /o/ 発声時の 3 次元 MRI データから抽出した声道領域を 3 次元レンダリングしたものを示している.

表 4.1: 録音音声から抽出したホルマント周波数 (単位は Hz)

母音	被験者 A			被験者 B		
	$F1$	$F2$	$F3$	$F1$	$F2$	$F3$
/a/	643	1122	2652	713	1121	2414
/i/	279	2015	2842	281	2131	3073
/u/	312	1273	2134	355	1018	2357
/e/	420	1682	2375	502	1863	2478
/o/	413	791	2384	543	863	2258

被験者 A, および被験者 B の日本語 5 母音発声中の録音音声 を 12 kHz にダウンサンプリングし, 15 次の線形予測係数で解析を行った際に得られたホルマント周波数を示している.

4.2 声道壁インピーダンスを剛壁とした際の周波数特性

3次元声道データに対し, 提案法, および比較手法を用いて解析を行った結果を表 4.2, 4.3 及び図 4.6 ~ 4.15 に示す. 声道面積関数から得られたホルマント周波数の正確さを評価するために, FEM により得られるホルマント周波数 (F_f) を基準として用いる. FEM の計算結果を基準とすることで, 声道壁インピーダンスや鼻腔放射等の要因によるホルマント周波数のずれの問題を無視できる. 提案法において, 声道面積関数の導出に用いた音圧分布の周波数は $F1$ である. その他の周波数を用いた結果は 4.5 で述べる. 図 4.6 ~ 4.10 は被験者 A に対する結果を, 図 4.11 ~ 4.15 は被験者 B に対する結果をそれぞれ示している. それぞれの手法により得られた声道面積関数の形状は類似しているが同一ではないことがわか

る．表 4.2 は FEM より求めたホルマント周波数と提案法の声道面積関数より求めたホルマント周波数を比較している．表 4.3 は FEM より求めたホルマント周波数と比較手法の声道面積関数より求めたホルマント周波数を比較している．表 4.2 では，ほとんどの条件について誤差率の大きさが 1% 台，またはそれ以下を示しているのに対し，表 4.3 では，多くの条件について誤差率の大きさが 2% 台を超えている．この結果から，提案法の方が比較手法よりもより良好なアルゴリズムであると考えられる．しかし，ホルマント周波数を表 4.1 と比較すると，提案法，比較手法ともに特に F_1 が低く推定されていることがわかる．これは主に声道壁を剛壁として扱っていることが原因であることがわかっている [38]．この問題に関して，次項で考察する．

4.3 声道壁インピーダンスを有限値とした際の周波数特性

4.3.1 はじめに

4.2 では，有限要素法，および，声道面積関数からホルマント周波数を求める際，声道壁を剛壁とした．そのため，前項の結果では録音音声のホルマント周波数と比較して低い周波数を示している．ここでは，声道壁にインピーダンスを与えたときのホルマント周波数を計算し，録音音声のホルマント周波数と比較する．

表 4.2: FEM より得られた 第 1 ~ 第 3 ホルマント周波数 ($F1 - F3$) と提案法 (声道壁は剛壁) より得られた $F1 - F3$ の比較

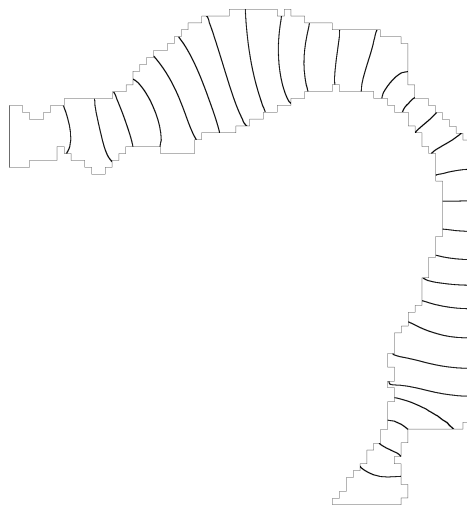
母音	$F1$			$F2$			$F3$		
	提案法 [Hz]	FEM [Hz]	誤差率 [%]	提案法	FEM	誤差率	提案法	FEM	誤差率
/a/ _A	466	465	0.22	1044	1061	1.60	2668	2652	0.60
/i/ _A	184	181	1.66	2039	2089	2.39	2913	2987	2.48
/u/ _A	238	239	0.42	1109	1092	1.56	2315	2266	2.16
/e/ _A	391	387	1.03	1680	1650	1.82	2439	2406	1.37
/o/ _A	315	319	1.25	688	698	1.43	2519	2476	1.74
/a/ _B	607	599	1.34	1169	1158	0.95	2331	2290	1.79
/i/ _B	206	204	0.98	1928	1919	0.47	2805	2750	2.00
/u/ _B	280	280	0.0	1175	1168	0.60	2127	2007	5.98
/e/ _B	410	412	0.49	1573	1521	3.42	2132	2008	6.18
/o/ _B	438	429	2.10	851	834	2.04	2352	2313	1.69

FEMにより得られる口唇部の音圧から $F1 - F3$ を求め、提案法により得られた $F1 - F3$ と比較している。声道面積関数を音響等価回路に適用する際、声道壁は剛壁として扱っている。誤差率は $|F_a - F_f|/F_f$ とすることで得られる。ただし、 F_f はFEMによるホルマント周波数、 F_a は提案法によるホルマント周波数である。/a/_A は被験者 A が/a/を発声しているときの結果を、/a/_B は被験者 B が/a/を発声しているときの結果をそれぞれ表している。

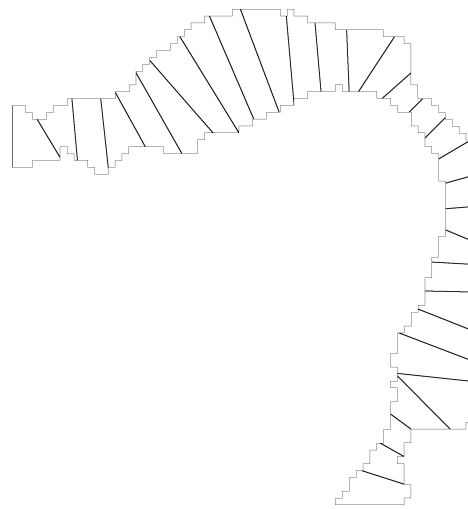
表 4.3: FEM より得られた $F1 - F3$ と比較手法 (声道壁は剛壁) より得られた $F1 - F3$ の比較

母音	$F1$			$F2$			$F3$		
	比較手法 [Hz]	FEM [Hz]	誤差率 [%]	比較手法	FEM	誤差率	比較手法	FEM	誤差率
/a/A	475	465	2.15	1046	1061	1.41	2791	2652	5.24
/i/A	190	181	4.97	2169	2089	3.83	3195	2987	6.96
/u/A	261	239	9.21	1188	1092	8.79	2377	2266	4.90
/e/A	414	387	6.98	1789	1650	8.42	2565	2406	6.61
/o/A	346	319	8.46	767	698	9.89	2629	2476	6.18
/a/B	621	599	3.67	1196	1158	3.28	2404	2290	4.98
/i/B	213	204	4.41	2151	1919	12.1	2880	2750	4.73
/u/B	275	280	1.78	1178	1168	0.86	2263	2007	12.8
/e/B	431	412	4.61	1649	1521	8.41	2208	2008	9.96
/o/B	480	429	11.9	922	834	10.6	2443	2313	5.62

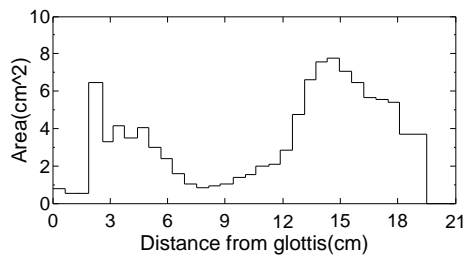
FEM により得られる口唇部の音圧から $F1 - F3$ を求め、比較手法により得られた $F1 - F3$ と比較している。声道面積関数を音響等価回路に適用する際、声道壁は剛壁として扱っている。誤差率は $|F_a - F_f|/F_f$ とすることで得られる。ただし、 F_f は FEM によるホルマント周波数、 F_a は提案法によるホルマント周波数である。/a/A は被験者 A が/a/を発声しているときの結果を、/a/B は被験者 B が/a/を発声しているときの結果をそれぞれ表している。



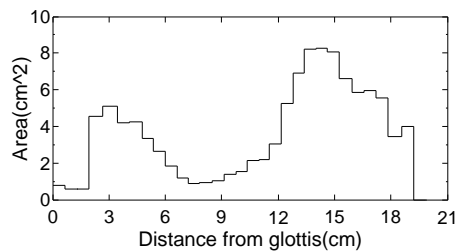
(a) 提案法による声道分割



(b) 比較手法による声道分割



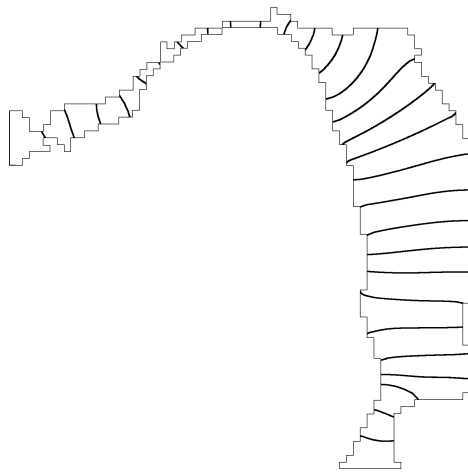
(c) 提案法による声道面積関数



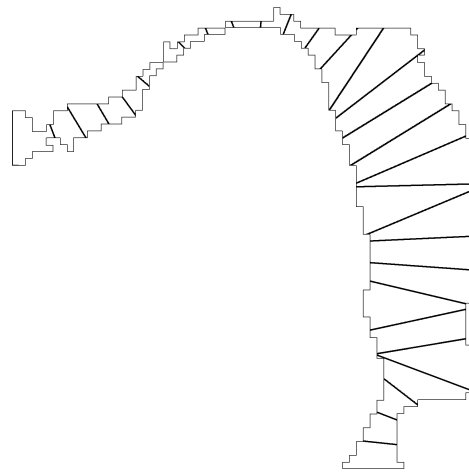
(d) 比較手法による声道面積関数

図 4.6: /a/A の声道分割画像と声道面積関数

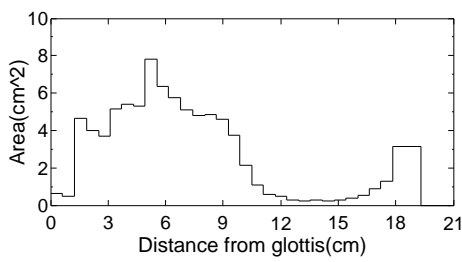
被験者 A が /a/ を発音している声道画像に対する声道分割の様子と、声道面積関数を、提案法と比較手法のそれぞれについて示している。



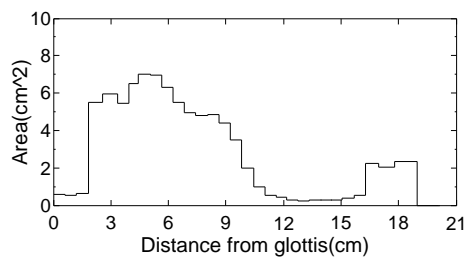
(a) 提案法による声道分割



(b) 比較手法による声道分割



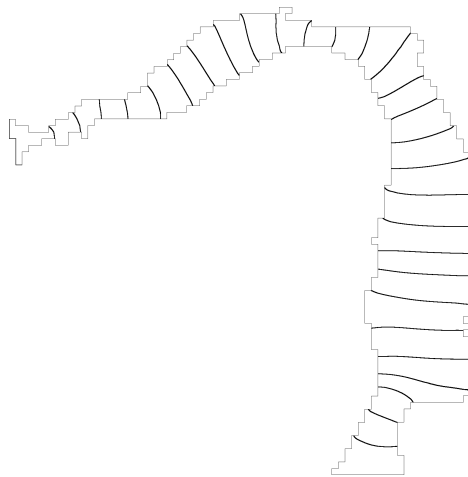
(c) 提案法による声道面積関数



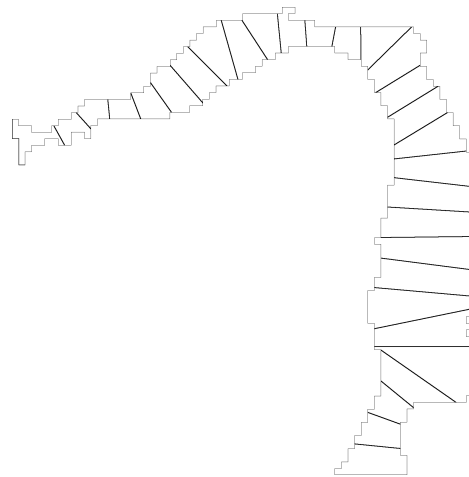
(d) 比較手法による声道面積関数

図 4.7: /i/_A の声道分割画像と声道面積関数

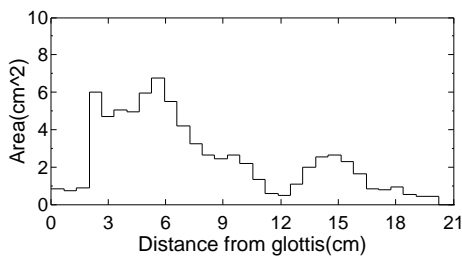
被験者 A が /i/ を発音している声道画像に対する声道分割の様子と、声道面積関数を、提案法と比較手法のそれぞれについて示している。



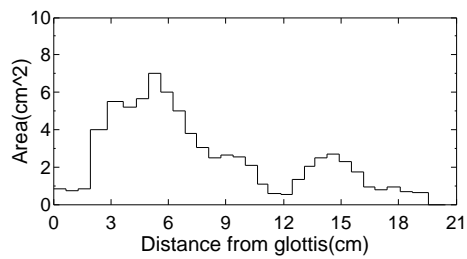
(a) 提案法による声道分割



(b) 比較手法による声道分割



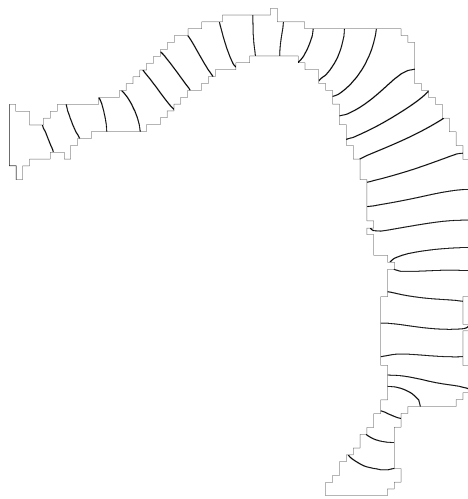
(c) 提案法による声道面積関数



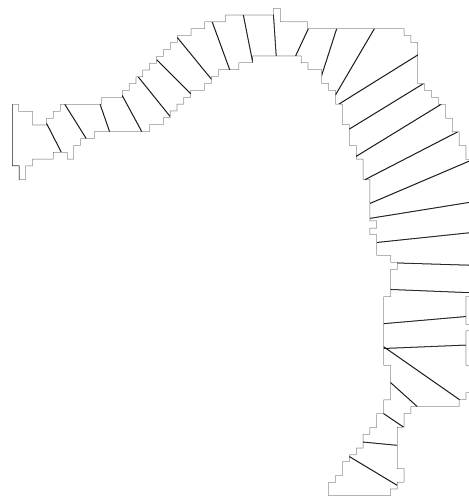
(d) 比較手法による声道面積関数

図 4.8: /u/_A の声道分割画像と声道面積関数

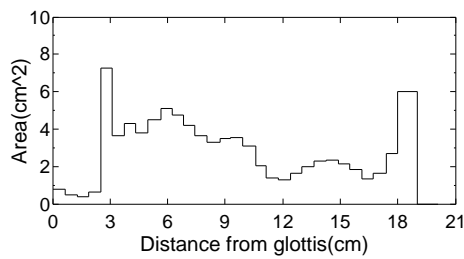
被験者 A が /u/ を発音している声道画像に対する声道分割の様子と、声道面積関数を、提案法と比較手法のそれぞれについて示している。



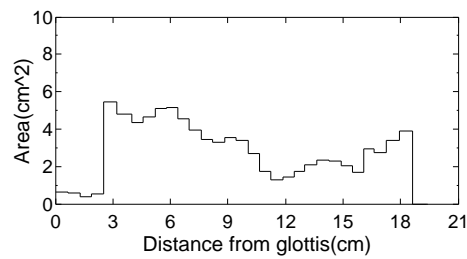
(a) 提案法による声道分割



(b) 比較手法による声道分割



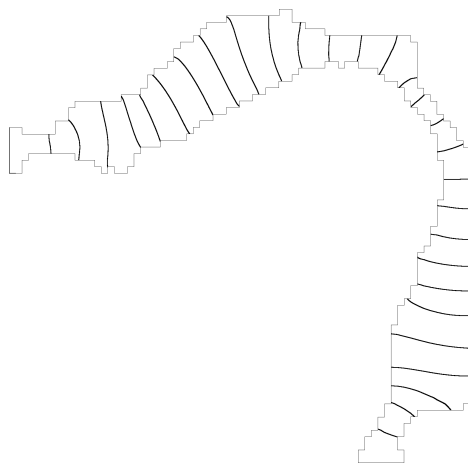
(c) 提案法による声道面積関数



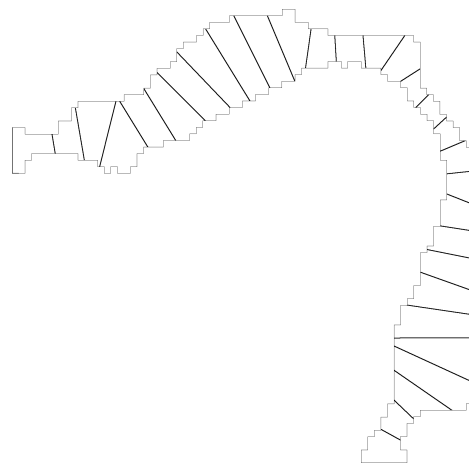
(d) 比較手法による声道面積関数

図 4.9: /e/_A の声道分割画像と声道面積関数

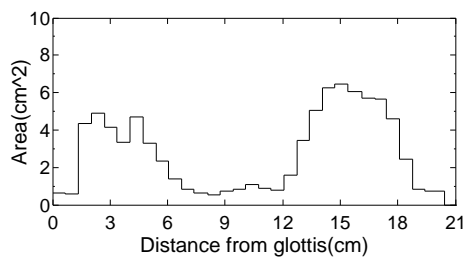
被験者 A が /e/ を発音している声道画像に対する声道分割の様子と，声道面積関数を，提案法と比較手法のそれぞれについて示している．



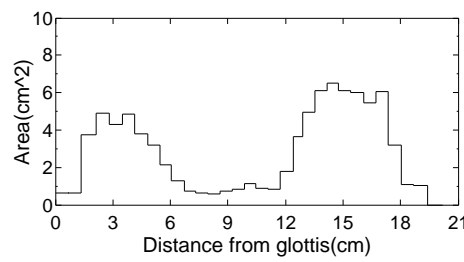
(a) 提案法による声道分割



(b) 比較手法による声道分割



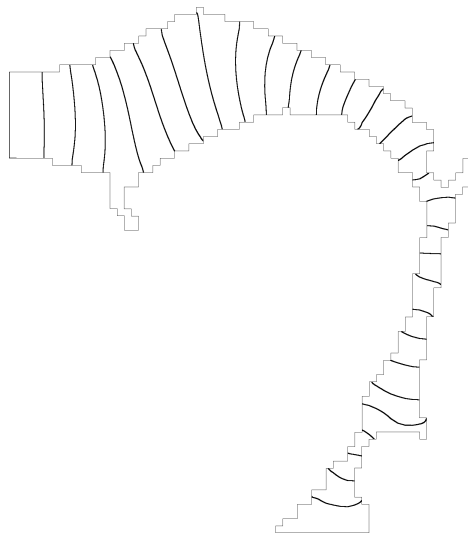
(c) 提案法による声道面積関数



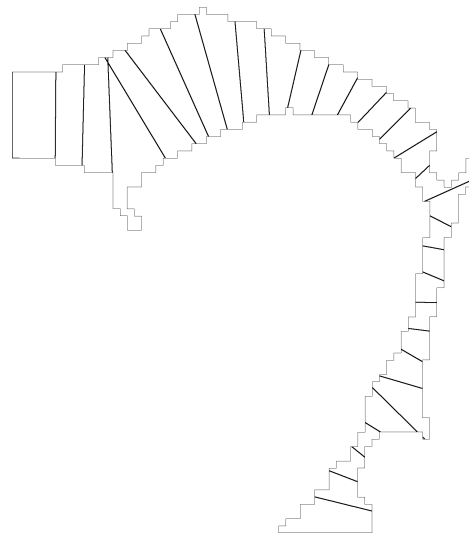
(d) 比較手法による声道面積関数

図 4.10: /o/_A の声道分割画像と声道面積関数

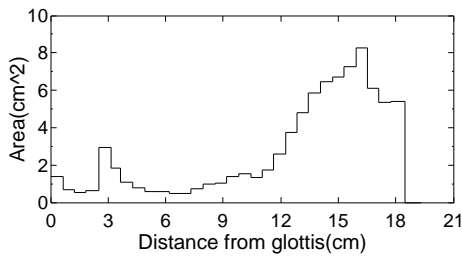
被験者 A が /o/ を発音している声道画像に対する声道分割の様子と，声道面積関数を，提案法と比較手法のそれぞれについて示している．



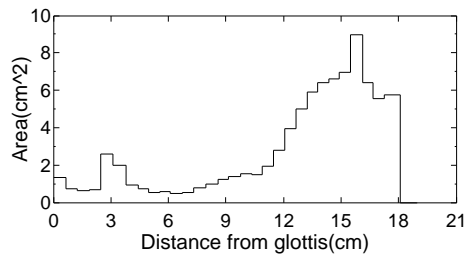
(a) 提案法による声道分割



(b) 比較手法による声道分割



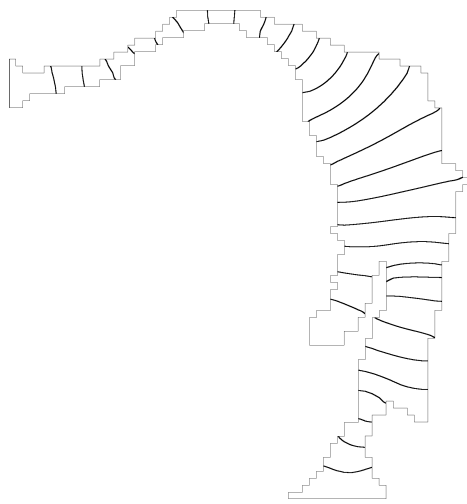
(c) 提案法による声道面積関数



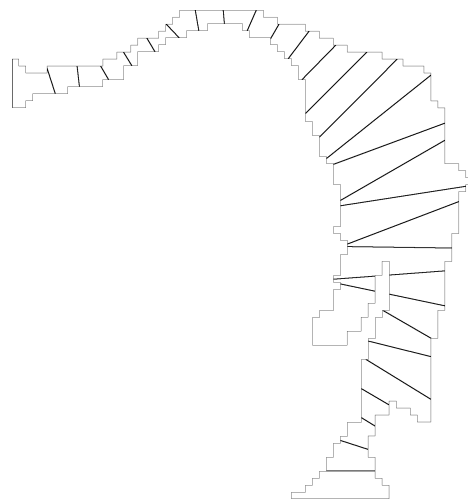
(d) 比較手法による声道面積関数

図 4.11: /a/_B の声道分割画像と声道面積関数

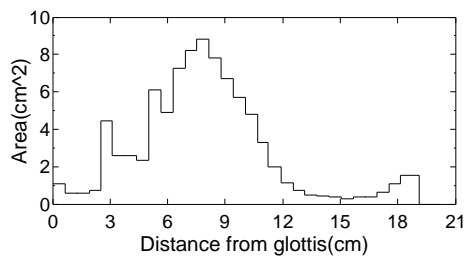
被験者 B が /a/ を発音している声道画像に対する声道分割の様子と、声道面積関数を、提案法と比較手法のそれぞれについて示している。



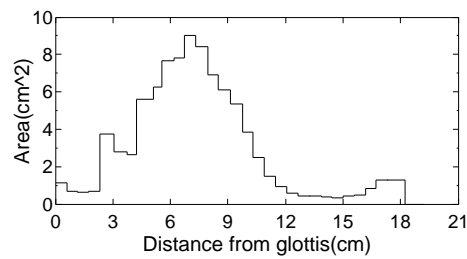
(a) 提案法による声道分割



(b) 比較手法による声道分割



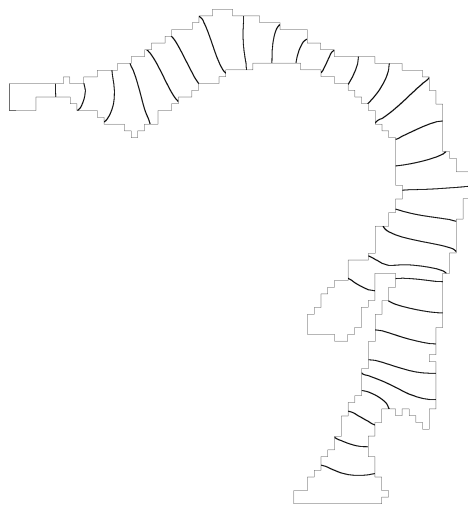
(c) 提案法による声道面積関数



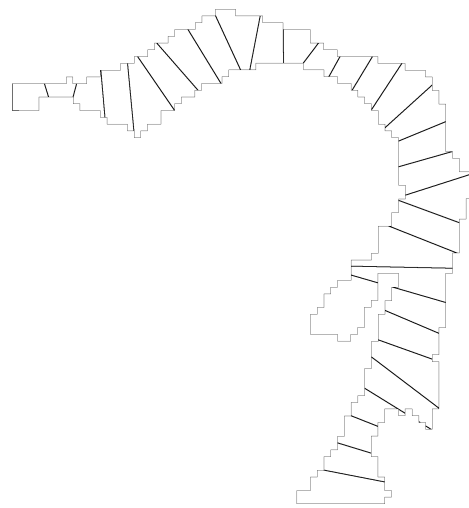
(d) 比較手法による声道面積関数

図 4.12: /i/_B の声道分割画像と声道面積関数

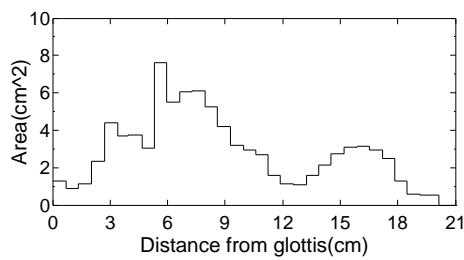
被験者 B が /i/ を発音している声道画像に対する声道分割の様子と，声道面積関数を，提案法と比較手法のそれぞれについて示している．



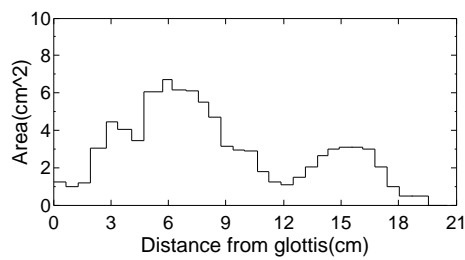
(a) 提案法による声道分割



(b) 比較手法による声道分割



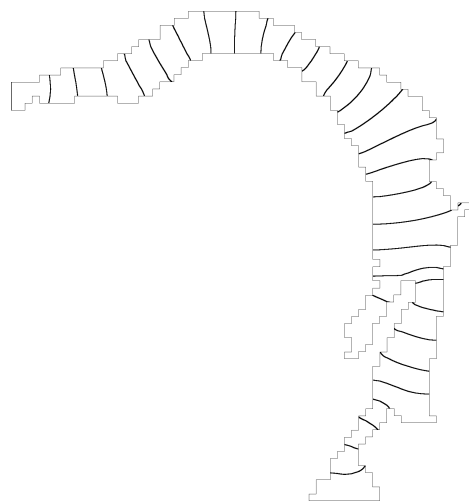
(c) 提案法による声道面積関数



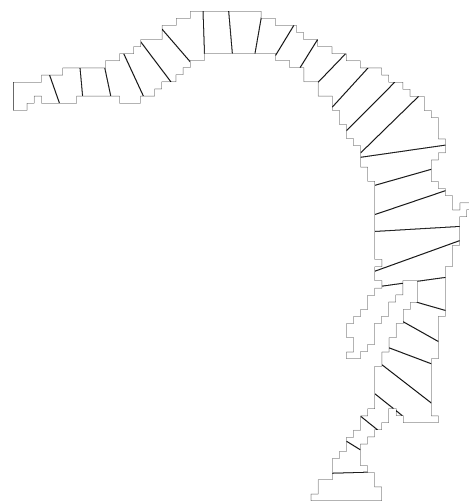
(d) 比較手法による声道面積関数

図 4.13: /u/_B の声道分割画像と声道面積関数

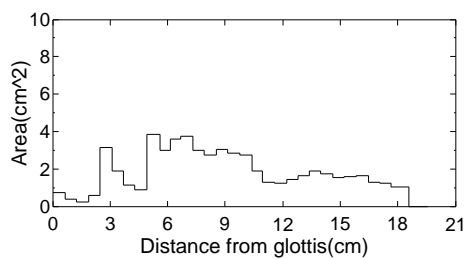
被験者 B が /u/ を発音している声道画像に対する声道分割の様子と、声道面積関数を、提案法と比較手法のそれぞれについて示している。



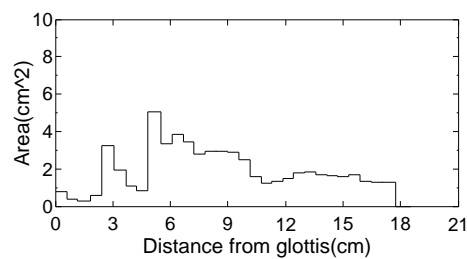
(a) 提案法による声道分割



(b) 比較手法による声道分割



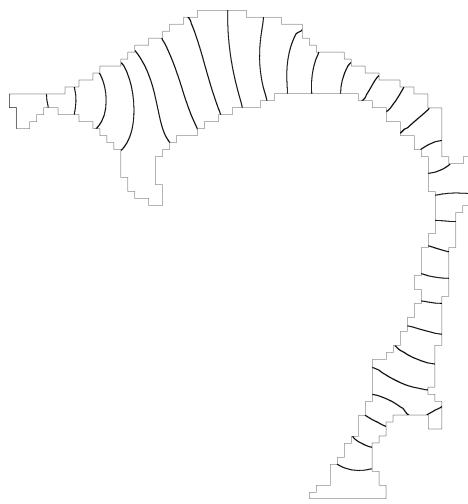
(c) 提案法による声道面積関数



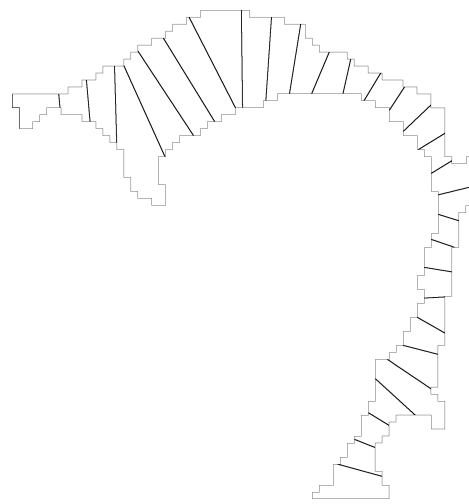
(d) 比較手法による声道面積関数

図 4.14: /e/ の声道分割画像と声道面積関数

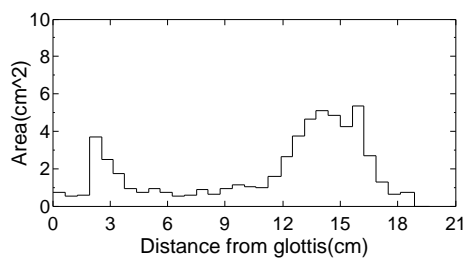
被験者 B が /e/ を発音している声道画像に対する声道分割の様子と，声道面積関数を，提案法と比較手法のそれぞれについて示している．



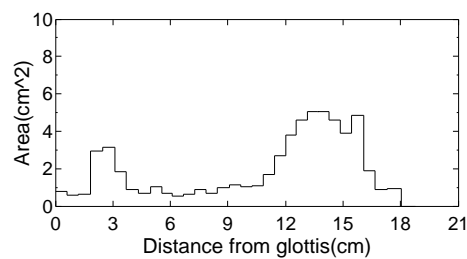
(a) 提案法による声道分割



(b) 比較手法による声道分割



(c) 提案法による声道面積関数



(d) 比較手法による声道面積関数

図 4.15: /o/_B の声道分割画像と声道面積関数

被験者 B が /o/ を発音している声道画像に対する声道分割の様子と，声道面積関数を，提案法と比較手法のそれぞれについて示している．

4.3.2 近似式によるホルマント周波数の補正

Sondhi ら [38] は，声道壁を剛壁とした際のホルマント周波数と閉口時の基本共振周波数から声道壁にインピーダンスを与えたときのホルマント周波数を推定できることを示した．次式は3者の関係を示している．

$$\omega^2 = \hat{\omega}^2 + \omega_c^2 \quad (4.1)$$

ただし， ω は声道壁にインピーダンスを与えた時のホルマント周波数， $\hat{\omega}$ は声道壁が剛壁であるときのホルマント周波数， ω_c は閉口時の基本共振周波数である．この式において， ω_c を与えることができれば， $\hat{\omega}$ から ω を推定することができる．つまり，声道壁を剛壁とした際のホルマント周波数を，声道壁にインピーダンスを与えた際のホルマント周波数に補正することができるようになる．文献 [38] では， $\omega_c = 406\pi\text{rad/sec}(= 203\text{Hz})$ としている．この値を用いて，提案法のホルマント周波数に対し補正を行ったものと録音音声と比較した結果を表 4.4 に示す．補正前と比較すると，特に $F1$ に関しては録音音声のホルマント周波数に近くなっているが，それでもまだ広母音では誤差が大きい．

4.3.3 損失を考慮した音響等価回路を用いたホルマント周波数の導出

4.2 で示した声道面積関数を，3.4 に示した損失を考慮した音響等価回路に適用し，ホルマント周波数を求めた結果を表 4.5 に示す．表 4.4 と同様に，録音音声より得られたホルマント周波数と比較している．表 4.4 では，誤差 1% 台またはそれ以下の数は 3 例，5% 台またはそれ以下の数は 13 例，10% 台またはそれ以下の数が 23 例あるのに対し，表 4.5 では，誤

差 1% 台またはそれ以下の数は 1 例，誤差 5% 台またはそれ以下の数は 6 例，誤差 10% 台またはそれ以下の数は 19 例となり，表 4.4 と比較して表 4.5 の方が若干誤差の程度が大きいことがわかる．このように誤差が大きくなる要因として，以下の理由が考えられる．

- 壁のインピーダンスを声道全体で均一として計算したが，実際の声道では不均一である．
- 軟口蓋から鼻腔への漏れによりホルマント周波数が変化する．
- 撮像時に発声された音声と録音音声が異なる．
- 声道の周囲長は一般に円筒近似で行われているが，図 3.7 で示したように円筒とは大きく異なっている．

表 4.4: 4.1 式で補正を行った提案法の第 1 ~ 第 3 ホルマント周波数 ($F1 - F3$) と録音音声より得られた $F1 - F3$ の比較

母音	$F1$			$F2$			$F3$		
	補正後 [Hz]	録音音声 [Hz]	誤差率 [%]	補正後	録音音声	誤差率	補正後	録音音声	誤差率
/a/A	508	643	20.9	1064	1122	5.21	2676	2652	0.89
/i/A	274	279	1.80	2049	2015	1.69	2920	2842	2.75
/u/A	313	312	0.26	1127	1273	11.44	2324	2134	8.90
/e/A	441	420	4.89	1692	1682	0.61	2447	2375	3.05
/o/A	375	413	9.26	717	791	9.31	2527	2384	6.01
/a/B	640	713	10.2	1187	1121	5.84	2340	2414	3.07
/i/B	289	281	2.92	1939	2131	9.03	2812	3073	8.48
/u/B	345	355	2.58	1192	1018	17.13	2137	2357	9.35
/e/B	457	502	8.86	1586	1863	14.87	2142	2478	13.6
/o/B	483	543	11.1	875	863	1.38	2361	2258	4.55

表 4.2 で示した提案法のホルマント周波数を式 4.1 を用いて補正したものと録音音声を比較した結果を表している。被験者 A, B とともに、多くの条件で誤差率の大きさが 10% を超えていることがわかる。

表 4.5: 損失を考慮した音響等価回路より得られた第 1 ~ 第 3 ホルマント周波数 ($F1 - F3$) と録音音声より得られた $F1 - F3$ の比較

母音	$F1$			$F2$			$F3$		
	等価回路 [Hz]	録音音声 [Hz]	誤差率 [%]	等価回路	録音音声	誤差率	等価回路	録音音声	誤差率
/a/A	568	643	13.2	1066	1122	5.25	2670	2652	0.67
/i/A	327	279	14.68	2084	2015	3.31	2941	2842	3.37
/u/A	368	312	15.22	1148	1273	10.89	2324	2134	8.18
/e/A	493	420	14.81	1733	1682	2.94	2472	2375	3.92
/o/A	463	413	10.80	762	791	3.81	2551	2384	6.54
/a/B	767	713	7.04	1203	1121	6.81	2377	2414	1.56
/i/B	330	281	14.85	1974	2131	7.95	2868	3073	7.15
/u/B	385	355	7.79	1216	1018	16.28	2151	2357	9.58
/e/B	541	502	7.21	1632	1863	14.15	2181	2478	13.62
/o/B	429	543	26.57	834	863	3.47	2313	2258	2.38

声道面積関数を損失を考慮した音響等価回路に適用し、録音音声と比較した結果を表している。表 4.4 では、誤差 1% 台またはそれ以下の数は 3 例、5% 台またはそれ以下の数は 13 例、10% 台またはそれ以下の数が 23 例あるのに対し、表 4.5 では、誤差 1% 台またはそれ以下の数は 1 例、誤差 5% 台またはそれ以下の数は 6 例、誤差 10% 台またはそれ以下の数は 19 例となり、表 4.4 と比較して表 4.5 の方が若干誤差の程度が大きいことがわかる。

4.4 口唇付近の補正の効果の検証

提案法では、口唇面は一般的に等音圧面とはならないために、最後のサブセクションの厚さの補正を行っている。本項ではこの口唇付近の補正の効果がどの程度であるかを検証するために、口唇付近の補正を行った場合と、口唇付近の補正を行わない場合のそれぞれの声道面積関数を求め、ホルマント周波数の誤差率により比較する。補正を行う場合は、3.3で示した通り、最後のサブセクションの厚さを一つ前のサブセクションの断面積と最終サブセクションの体積から求める。補正を行わない場合は、口唇面を3.3の C_b とし、 C_b の一つ前、すなわち最も口唇面に近い位置にある等音圧面を C_a として C_a-C_b 間の距離を求め、これを最終サブセクションの厚さとする。それぞれの方法で得られたホルマント周波数をFEMのホルマント周波数と比較した結果を表4.6に示す。 $/a/A$ の $F2$ 、 $/i/A$ の $F1, F3$ 、 $/e/B$ の $F3$ 、 $/o/B$ を除いては口唇付近の補正有りの方が誤差率が小さい、または等しいことがわかる。誤差率の平均値に関して比較を行うと、補正有りとした場合は1.72%であるのに対し、補正無しとした場合は2.41%であり、口唇付近の補正を行った方が有意に小さい($t(29) = -2.65, p < 0.05$)。これらのことから、口唇付近の補正を行った方が声道面積関数をより良好に推定できることがわかる。

表 4.6: 口唇付近の補正の有無によるホルマント周波数の比較

母音・被験者	ホルマント	FEM	口唇付近の補正有り		口唇付近の補正無し	
		周波数 [Hz]	周波数 [Hz]	誤差率 [%]	周波数 [Hz]	誤差率 [%]
/a/A	F1	465	466	0.22	470	1.08
	F2	1061	1044	1.60	1078	1.60
	F3	2652	2668	0.60	2690	1.43
/i/A	F1	181	184	1.66	183	1.10
	F2	2089	2039	2.39	2037	2.49
	F3	2987	2913	2.48	2946	1.37
/u/A	F1	239	238	0.42	253	5.86
	F2	1092	1109	1.56	1166	6.78
	F3	2266	2315	2.16	2332	2.91
/e/A	F1	387	391	1.03	395	2.07
	F2	1650	1680	1.82	1698	2.91
	F3	2406	2439	1.37	2444	1.58
/o/A	F1	319	315	1.25	326	2.19
	F2	698	688	1.43	717	2.72
	F3	2476	2519	1.74	2524	1.94
/a/B	F1	599	607	1.34	607	1.34
	F2	1158	1169	0.95	1169	0.95
	F3	2290	2331	1.79	2331	1.79
/i/B	F1	204	206	0.98	206	0.98
	F2	1919	1928	0.47	1935	0.83
	F3	2750	2805	2.00	2832	2.98
/u/B	F1	280	280	0.00	285	1.79
	F2	1168	1175	0.60	1192	2.05
	F3	2007	2127	5.98	2124	5.83
/e/B	F1	412	410	0.49	416	0.97
	F2	1521	1573	3.42	1579	3.81
	F3	2008	2132	6.18	2127	5.93
/o/B	F1	429	438	2.10	436	1.63
	F2	834	851	2.04	849	1.80
	F3	2313	2352	1.69	2350	1.60

口唇付近の補正を行った場合と行わなかった場合の声道面積関数からそれぞれホルマント周波数を求め、FEMのホルマント周波数との誤差率を比較している。/a/AのF2, /i/AのF1, F3, /e/BのF3, /o/Bを除いては口唇付近の補正有りの方が誤差率が小さい、または等しいことがわかる。誤差率の平均値は、補正有りとした場合は1.72%であるのに対し、補正無しとした場合は2.41%であり、口唇付近の補正を行った方が有意に小さい($t(29) = -2.65, p < 0.05$)。

4.5 音圧分布導出のための周波数と面積関数との関係

4.2では、等音圧面、およびセクション長さを導出するために、それぞれの声道のFEMにおける $F1$ の音圧分布・音響インテンシティを用いた。FEMの分析で用いる周波数を変えれば、音圧分布・音響インテンシティの様相も変化するため、等音圧面・セクション長、およびそれらによって得られる声道面積関数や提案法のホルマント周波数も変化する可能性がある。本項では、解析に用いる周波数を変化させたときに音圧分布・音響インテンシティが変化することで、最終的に得られる声道面積関数、およびホルマント周波数がどのように変化するかを調べた。被験者A及びBの声道画像に対し、さまざまな周波数の音圧分布、音響インテンシティを用いて声道面積関数を求め、さらにそれらの声道面積関数からホルマント周波数を求めた。

音圧・音響インテンシティ分布が変化すると等音圧面の形状も変化する。図4.16は被験者Aが/u/を発音している声道画像に対してFEMを行い、1092 Hzの音圧分布及び音響インテンシティから等音圧面を検出した例である。図4.16(a)は音響インテンシティの実部ベクトルの向きを、図4.16(b)は音圧分布を、図4.16(c)は中心線に沿って音圧分布を取得した際のグラフを、図4.16(d)は検出した等音圧面をそれぞれ示している。図4.16(a)において音がほぼ進行方向に沿って伝播しており、進行方向に対して垂直な方向への音の伝播は大きくない。ところが、図4.16(c)のような極大値・極小値が発生するため、その結果帯状ではない等音圧面が検出される。図4.16(d)の(1)、(2)のように帯状ではない面が作成される可能性がある。提案法では、等音圧面は声道を進行方向に対し垂直な方向

に分断している必要がある，つまり，等音圧面の形状は帯状である必要があるため，帯状ではない等音圧面は除外した後，声道面積関数を計算した． $F1$ を超える周波数領域では，波長の $(2n - 1)/4$ (n は整数) の定在波が発生するため，閉口端から開口端までの間に，1つ，またはそれ以上の数の極小値(節)と極大値(腹)が発生する(図 4.17)．周波数が高くなるほど，節及び腹の数が多くなる．従ってこのような等音圧面は，周波数が $F1$ を超えるあたりから現れはじめ，周波数が高くなるほど帯状ではない等音圧面の比率が大きくなり，高い周波数領域に関しては適正に計算できなくなる．よって，1Hz から $F1$ 付近までの周波数範囲について計算し，その誤差率を評価する．

声道面積関数を図 4.18 ~ 4.37 に，得られたホルマント周波数との誤差率の比較を示したグラフを図 4.38 ~ 4.47 に示す．図 4.38 ~ 4.47 より，/i/ の $F2, F3$ を除いては音圧・音響インテンシティに用いる周波数が $F1$ に近くなると誤差率が小さくなる，または周波数によらずに誤差率の変化が少ないことがわかる．一方， $F1$ を超えると急激に誤差が大きくなる． $F1$ より大きな周波数帯では図 4.48 のような等音圧面が発生するために，これらを除外して声道面積関数の計算を行うことになる．このとき，このような等音圧面が存在しない $F1$ 以下の周波数による結果と比べて情報の欠如が発生するため，それが誤差の増加に影響していると考えられる．/i/に関しては， $F1$ の時点で同図(b)の丸で示した位置の等音圧面が現れているため， $F1$ において誤差が大きくなると考えられる．それぞれの結果を総括すると，多くの場合， $F1$ 付近で最小値となるが，声道面積関数の形状が複雑な場合， $F1$ より少し低い周波数で最小値となることがある，すなわち， $F1$ ，または $F1$ より少し低い周波数の音圧，及び音響インテンシティの分布を用いることが最適であることを示している．

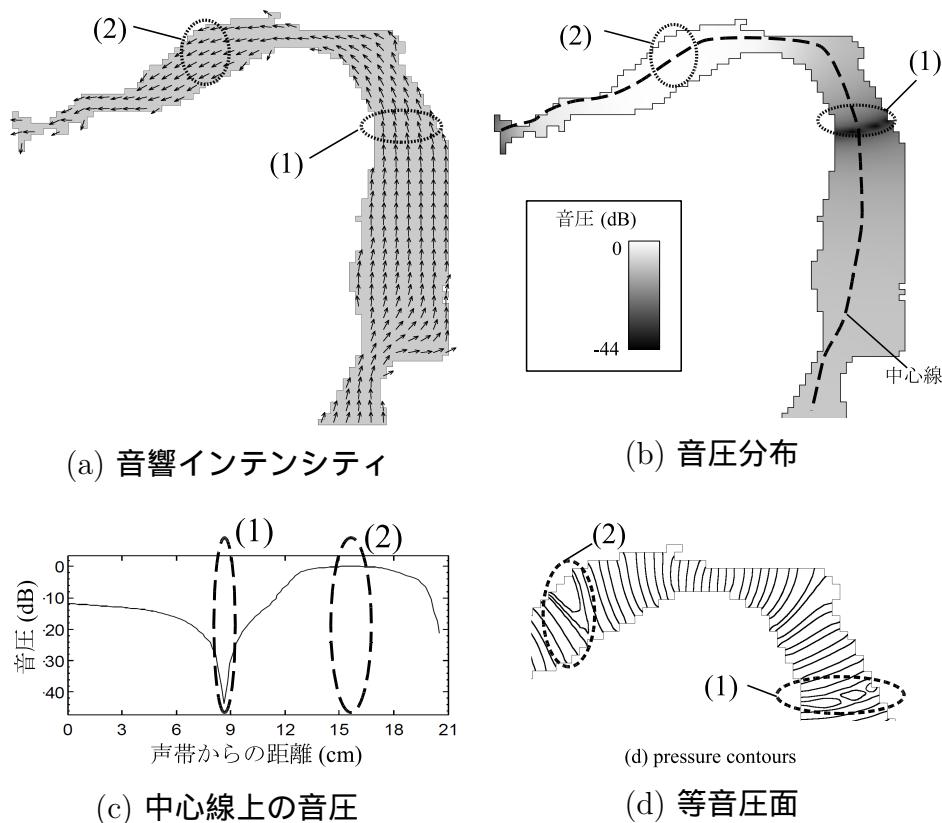


図 4.16: 帯状ではない等音圧面が発生した例 /u/A 1092 Hz (a),(b),(d) の断面は正中矢状面

被験者 A が /u/ を発音している声道画像に対して FEM を行い, 1092 Hz の音圧分布及び音響インテンシティから等音圧面を検出した例である. (a) は音響インテンシティの実部ベクトルの向きを, (b) は音圧分布を, (c) は中心線に沿って音圧分布を取得した際のグラフを, (d) は検出した等音圧面をそれぞれ示している. (a) において音がほぼ進行方向に沿って伝播しており, 進行方向に対して垂直な方向への音の伝播は大きくない. ところが, (c) のような極大値・極小値が発生するため, (d) のように帯状ではない等音圧面が検出されることがある. 提案法ではこのような等音圧面は除外して計算せざるを得ないため, このことが誤差の増加に影響を及ぼす可能性がある.

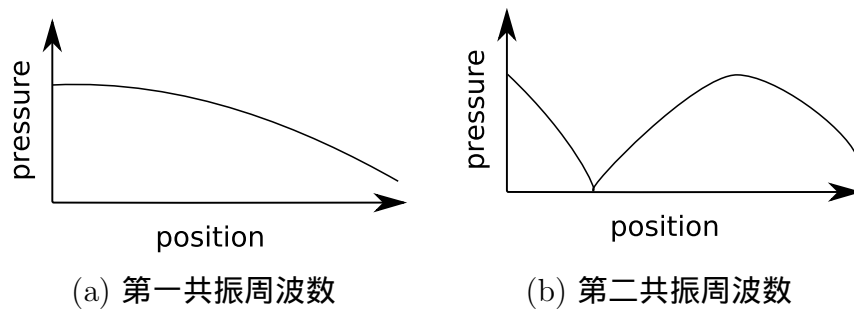
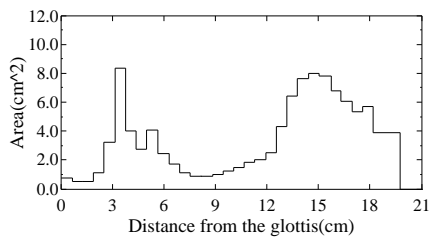
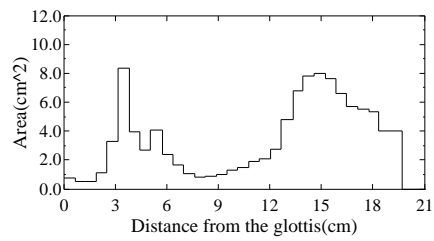


図 4.17: 共振周波数において一次元音響管内に発生する定在波と音圧分布の関係

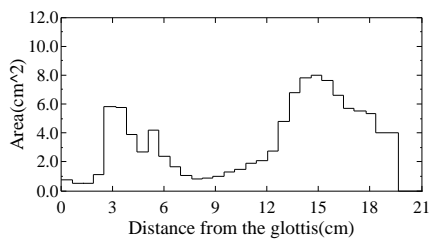
一端閉，一端開の音響管では，第一共振周波数のとき波長がおよそ $1/4$ の定在波が発生する．このとき，閉口端から開口端までの音圧変化は単調である．一方，第二共振周波数では，波長がおよそ $3/4$ の定在波が発生し，閉口端から開口端までの間に極小値（節）と極大値（腹）がそれぞれ1つずつ存在する．節や腹となる領域では，音の進行方向に対する音圧の傾斜は0となる．このような領域で等音圧面を強引に検出しようとする時，進行方向に対して垂直な方向に等音圧面が形成される可能性が高くなる．



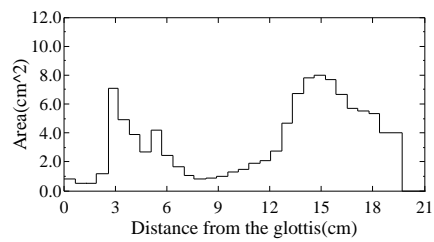
(a) 1Hz 声道面積関数



(b) 100Hz 声道面積関数



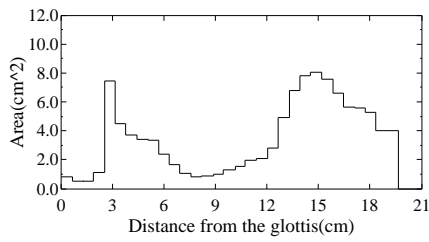
(c) 200Hz 声道面積関数



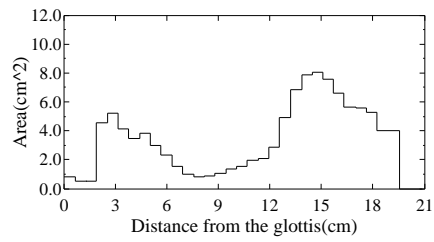
(d) 300Hz 声道面積関数

図 4.18: /a/A の声道画像の 1Hz, 100Hz, 200Hz, 300Hz の音圧・音響インテンシティ分布を用いたときの声道面積関数

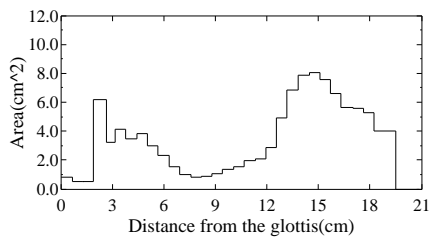
被験者 A が/a/を发声している声道画像に対し FEM を用いて 1Hz, 100Hz, 200Hz, 300Hz の音圧分布・音響インテンシティを求め、その音圧分布・音響インテンシティを用いて提案法を適用したときの声道面積関数を示している。



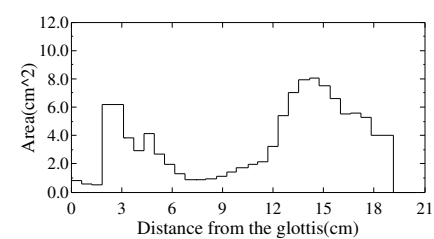
(a) 400Hz 声道面積関数



(b) 460Hz 声道面積関数



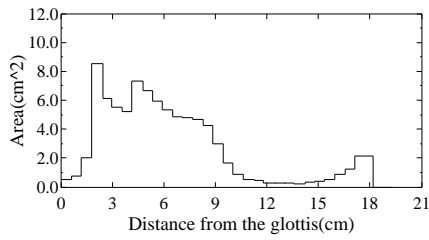
(c) 466Hz (F_1) 声道面積関数



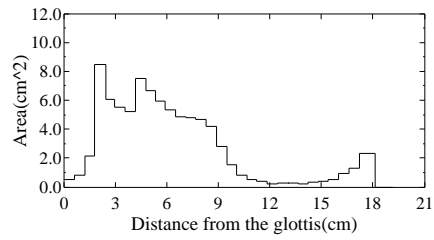
(d) 500Hz 声道面積関数 (等音圧面除去あり)

図 4.19: /a/ の声道画像の F_1 付近の音圧・音響インテンシティ分布を用いたときの声道面積関数

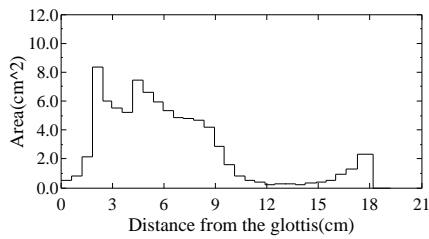
被験者 A が /a/ を発声している声道画像に対し FEM を用いて第 1 ホルマント付近の音圧分布・音響インテンシティを求め、その音圧分布・音響インテンシティを用いて提案法を適用したときのおよび声道面積関数を示している。500Hz において声帯付近で帯状ではない等音圧面が発生したため、(d) はその等音圧面の除去を行った結果を示している。



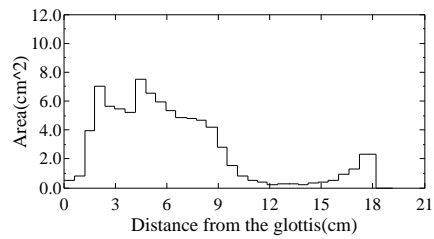
(a) 1Hz 声道面積関数



(b) 50Hz 声道面積関数



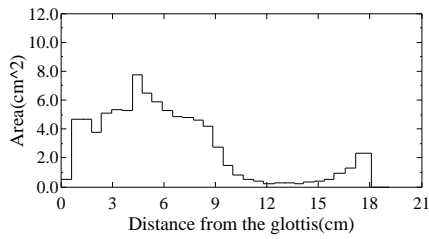
(c) 100Hz 声道面積関数



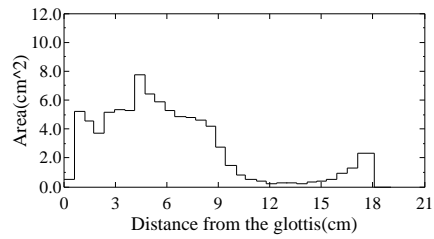
(d) 150Hz 声道面積関数

図 4.20: /i/_A の声道画像の 1Hz, 50Hz, 100Hz, 150Hz の音圧・音響インテ
ンシティ分布を用いたときの声道面積関数

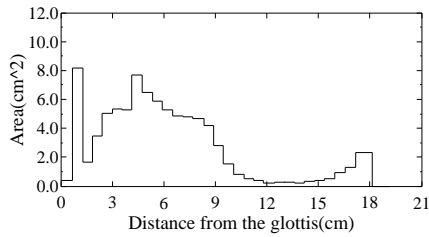
被験者 A が /i/ を発声している声道画像に対し FEM を用い
て 1Hz, 50Hz, 100Hz, 150Hz の音圧分布・音響インテンシ
ティを求め、その音圧分布・音響インテンシティを用いて
提案法を適用したときの声道分割画像、および声道面積関
数を示している。



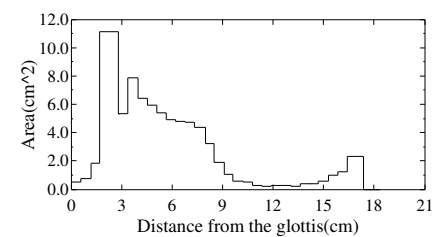
(a) 179Hz 声道面積関数



(b) 180Hz 声道面積関数



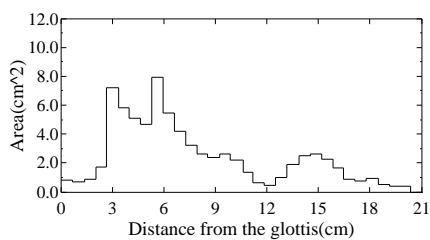
(c) 181Hz(F_1) 声道面積関数
(等音圧面除去あり)



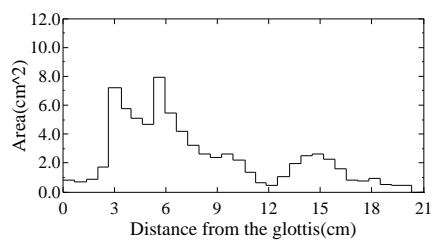
(d) 200Hz 声道面積関数 (等音圧面除去あり)

図 4.21: /i/_A の声道画像の F_1 付近の音圧・音響インテンシティ分布を用いたときの声道面積関数

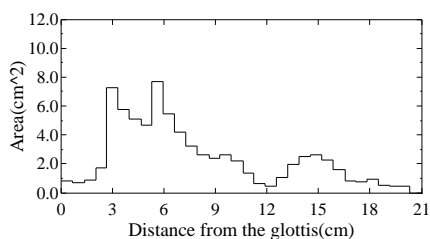
被験者 A が /i/ を発声している声道画像に対し FEM を用いて第 1 ホルマント付近の音圧分布・音響インテンシティを求め、その音圧分布・音響インテンシティを用いて提案法を適用したときの声道面積関数を示している。181Hz, 200Hz において声帯付近で帯状ではない等音圧面が発生したため、(c) 及び (d) はその等音圧面の除去を行った結果を示している。



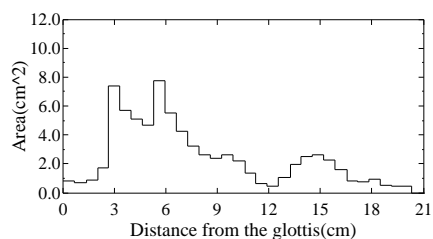
(a) 1Hz 声道面積関数



(b) 50Hz 声道面積関数



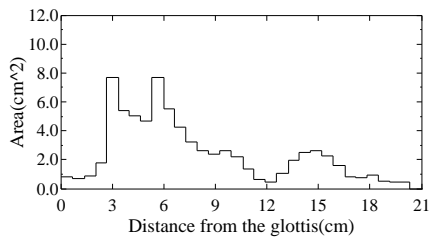
(c) 100Hz 声道面積関数



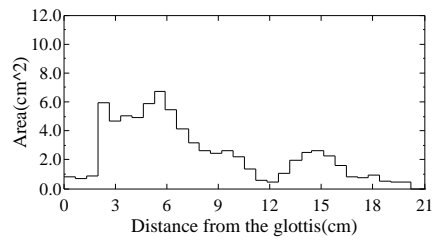
(d) 150Hz 声道面積関数

図 4.22: /u/A の声道画像の 1Hz, 50Hz, 100Hz, 150Hz の音圧・音響インテンシティ分布を用いたときの声道面積関数

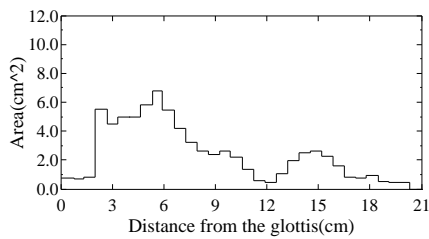
被験者 A が /u/ を発声している声道画像に対し FEM を用いて 1Hz, 50Hz, 100Hz, 150Hz の音圧分布・音響インテンシティを求め、その音圧分布・音響インテンシティを用いて提案法を適用したときの声道面積関数を示している。



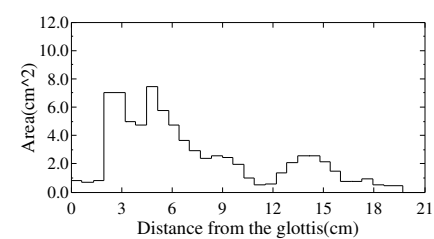
(a) 200Hz 声道面積関数



(b) 239Hz(F_1) 声道面積関数



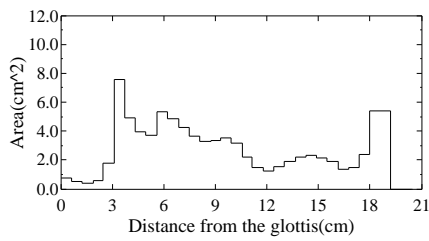
(c) 240Hz 声道面積関数



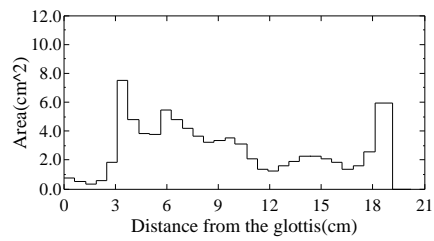
(d) 250Hz 声道面積関数 (等音圧面除去あり)

図 4.23: /u/_A の声道画像の F_1 付近の音圧・音響インテンシティ分布を用いたときの声道面積関数

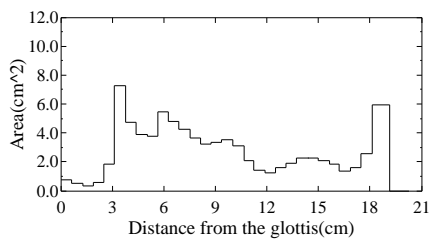
被験者 A が /u/ を発声している声道画像に対し FEM を用いて第 1 ホルマント付近の音圧分布・音響インテンシティを求め、その音圧分布・音響インテンシティを用いて提案法を適用したときの声道面積関数を示している。250Hz において声帯付近で帯状ではない等音圧面が発生したため、(d) はその等音圧面の除去を行った結果を示している。



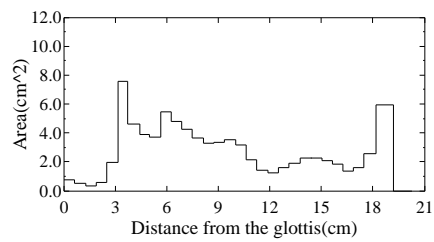
(a) 1Hz 声道面積関数



(b) 100Hz 声道面積関数



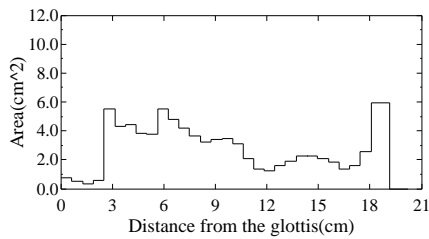
(c) 200Hz 声道面積関数



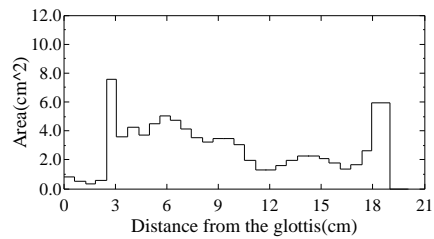
(d) 300Hz 声道面積関数

図 4.24: /e/_A の声道画像の 1Hz, 100Hz, 200Hz, 300Hz の音圧・音響インテンシティ分布を用いたときの声道面積関数

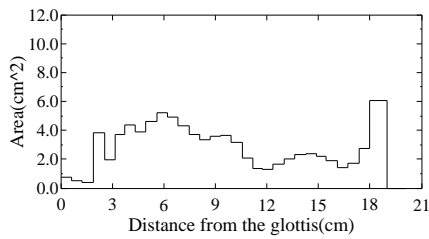
被験者 A が/e/を発声している声道画像に対し FEM を用いて 1Hz, 100Hz, 200Hz, 300Hz の音圧分布・音響インテンシティを求め、その音圧分布・音響インテンシティを用いて提案法を適用したときの声道面積関数を示している。



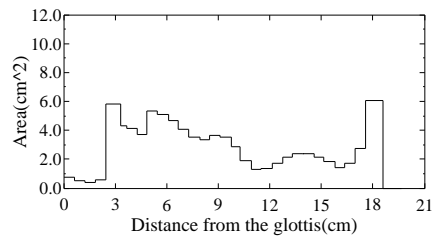
(a) 350Hz 声道面積関数



(b) 387Hz(F_1) 声道面積関数



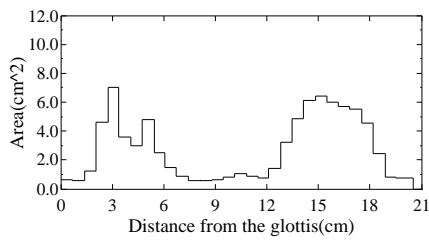
(c) 392Hz 声道面積関数 (等音圧面除去あり)



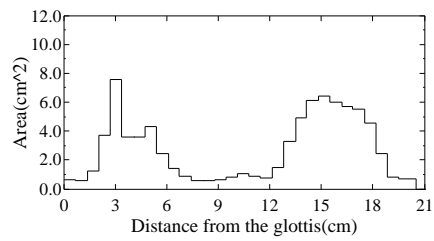
(d) 400Hz 声道面積関数 (等音圧面除去あり)

図 4.25: $/e/A$ の声道画像の F_1 付近の音圧・音響インテンシティ分布を用いたときの声道面積関数

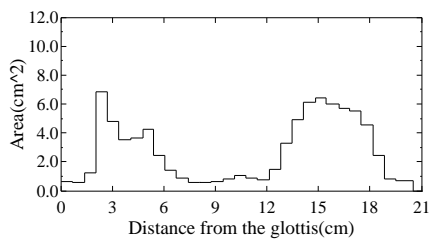
被験者 A が $/e/$ を発声している声道画像に対し FEM を用いて第 1 ホルマント付近の音圧分布・音響インテンシティを求め、その音圧分布・音響インテンシティを用いて提案法を適用したときの声道面積関数を示している。392Hz, 400Hz において声帯付近で帯状ではない等音圧面が発生したため、(c) 及び (d) はその等音圧面の除去を行った結果を示している。



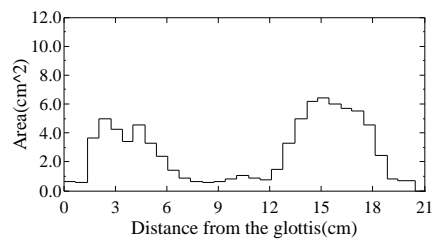
(a) 1Hz 声道面積関数



(b) 100Hz 声道面積関数



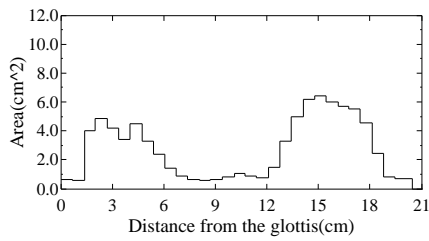
(c) 200Hz 声道面積関数



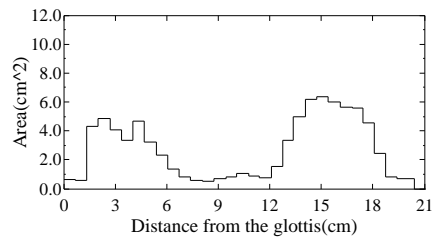
(d) 300Hz 声道面積関数

図 4.26: /o/A の声道画像の 1Hz, 100Hz, 200Hz, 300Hz の音圧・音響インテンシティ分布を用いたときの声道面積関数

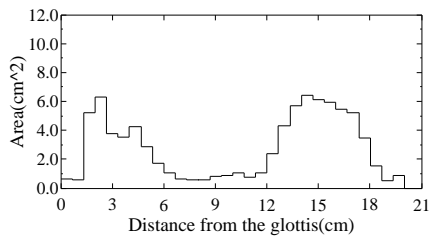
被験者 A が /o/ を発声している声道画像に対し FEM を用いて 1Hz, 100Hz, 200Hz, 300Hz の音圧分布・音響インテンシティを求め、その音圧分布・音響インテンシティを用いて提案法を適用したときの声道面積関数を示している。



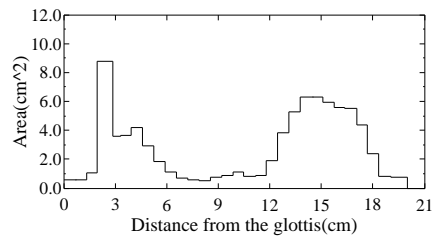
(a) 310Hz 声道面積関数



(b) 319Hz(F_1) 声道面積関数



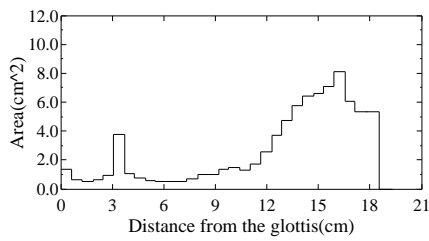
(c) 330Hz 声道面積関数 (等音圧面除去あり)



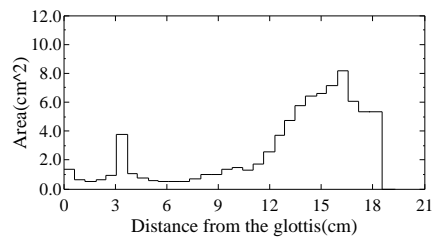
(d) 350Hz 声道面積関数 (等音圧面除去あり)

図 4.27: /o/_A の声道画像の F_1 付近の音圧・音響インテンシティ分布を用いたときの声道面積関数

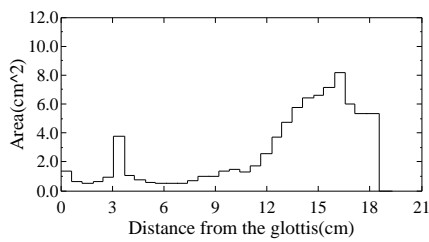
被験者 A が /o/ を発声している声道画像に対し FEM を用いて第 1 ホルマント 付近の音圧分布・音響インテンシティを求め、その音圧分布・音響インテンシティを用いて提案法を適用したときの声道面積関数を示している。330Hz, 350Hz において声帯付近で帯状ではない等音圧面が発生したため、(c) 及び (d) はその等音圧面の除去を行った結果を示している。



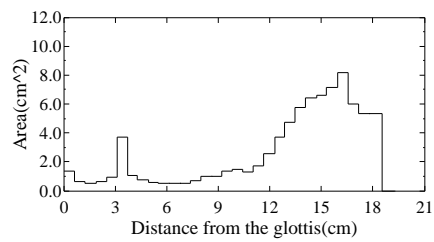
(a) 1Hz 声道面積関数



(b) 100Hz 声道面積関数



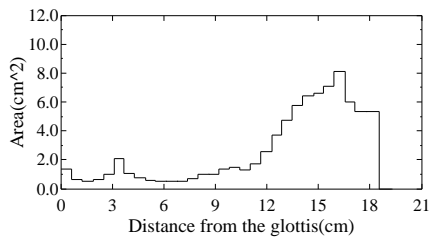
(c) 300Hz 声道面積関数



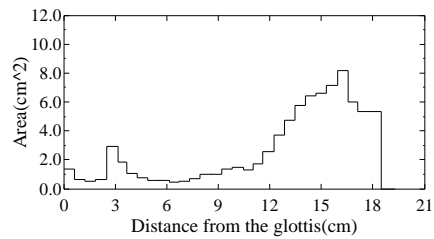
(d) 400Hz 声道面積関数

図 4.28: /a/_B の声道画像の 1Hz, 100Hz, 300Hz, 400Hz の音圧・音響インテンシティ分布を用いたときの声道面積関数

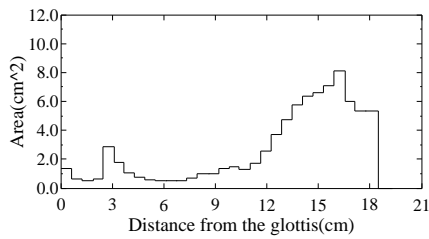
被験者 B が /a/ を発声している声道画像に対し FEM を用いて 1Hz, 100Hz, 300Hz, 400Hz の音圧分布・音響インテンシティを求め、その音圧分布・音響インテンシティを用いて提案法を適用したときの声道面積関数を示している。



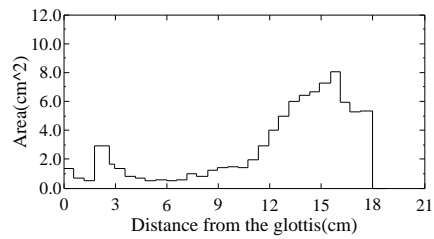
(a) 550Hz 声道面積関数



(b) 599Hz(F_1) 声道面積関数



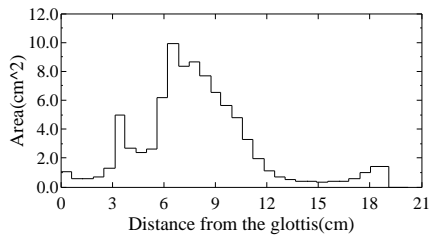
(c) 630Hz 声道面積関数 (等音圧面除去あり)



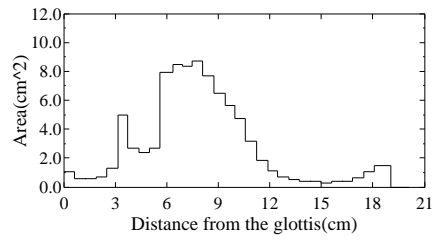
(d) 700Hz 声道面積関数 (等音圧面除去あり)

図 4.29: $/a/_B$ の声道画像の F_1 付近の音圧・音響インテンシティ分布を用いたときの声道面積関数

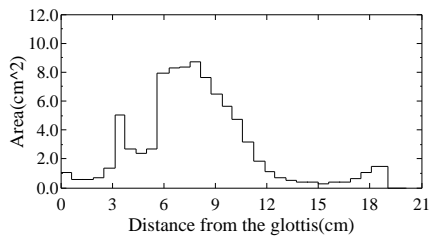
被験者 B が $/a/$ を発声している声道画像に対し FEM を用いて第 1 ホルマント付近の音圧分布・音響インテンシティを求め、その音圧分布・音響インテンシティを用いて提案法を適用したときのおよび声道面積関数を示している。630Hz, 700Hz において声帯付近で帯状ではない等音圧面が発生したため、(c) 及び (d) はその等音圧面の除去を行った結果を示している。



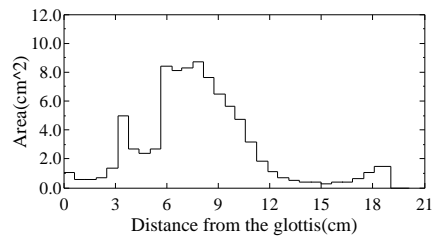
(a) 1Hz 声道面積関数



(b) 50Hz 声道面積関数



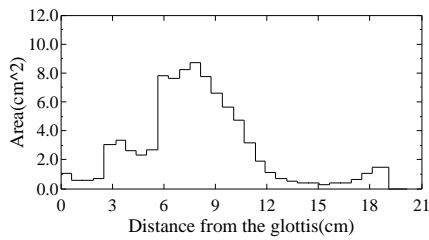
(c) 100Hz 声道面積関数



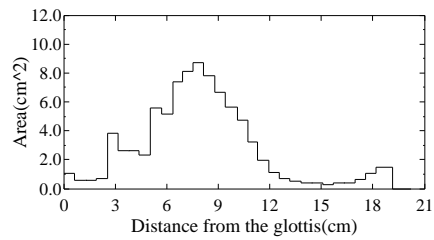
(d) 150Hz 声道面積関数

図 4.30: /i/_B の声道画像の 1Hz, 50Hz, 100Hz, 150Hz の音圧・音響インテ
ンシティ分布を用いたときの声道面積関数

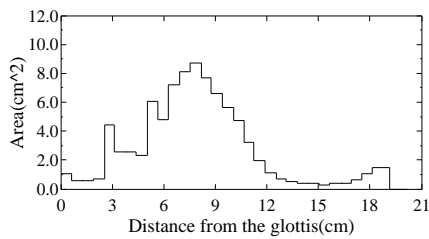
被験者 B が /i/ を発声している声道画像に対し FEM を用いて 1Hz, 50Hz, 100Hz, 150Hz の音圧分布・音響インテ
ンシティを求め、その音圧分布・音響インテ
ンシティを用いて
提案法を適用したときの声道分割画像、および声道面積関
数を示している。



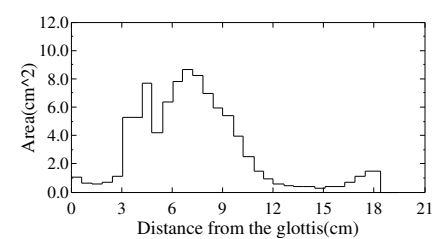
(a) 180Hz 声道面積関数



(b) 200Hz 声道面積関数



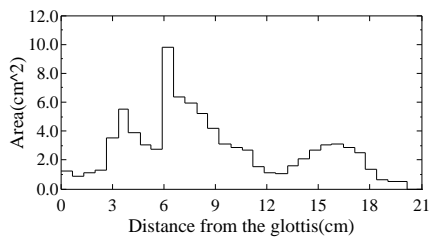
(c) 204Hz(F_1) 声道面積関数



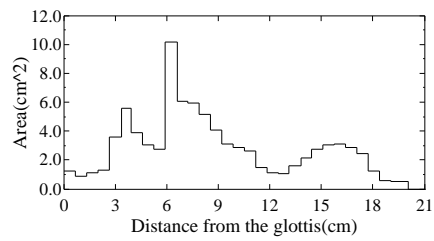
(d) 220Hz 声道面積関数 (等音圧面除去あり)

図 4.31: $/i/B$ の声道画像の F_1 付近の音圧・音響インテンシティ分布を用いたときの声道面積関数

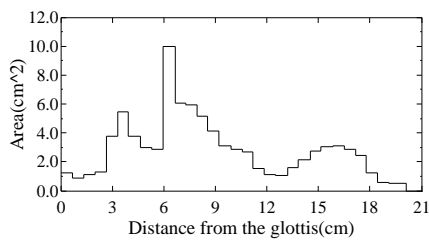
被験者 B が $/i/$ を発声している声道画像に対し FEM を用いて第 1 ホルマント周波数付近の音圧分布・音響インテンシティを求め、その音圧分布・音響インテンシティを用いて提案法を適用したときの声道面積関数を示している。220Hz において声帯付近で帯状ではない等音圧面が発生したため、(d) はその等音圧面の除去を行った結果を示している。



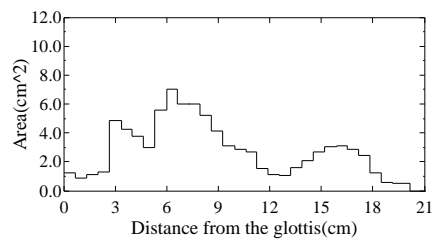
(a) 1Hz 声道面積関数



(b) 100Hz 声道面積関数



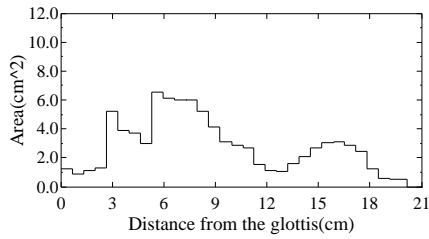
(c) 240Hz 声道面積関数



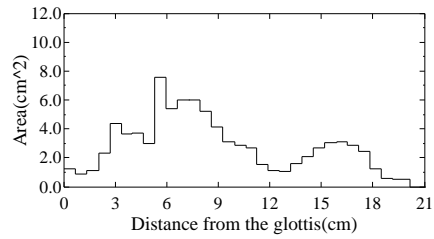
(d) 250Hz 声道面積関数

図 4.32: /u/_B の声道画像の 1Hz, 100Hz, 240Hz, 250Hz の音圧・音響インテンシティ分布を用いたときの声道面積関数

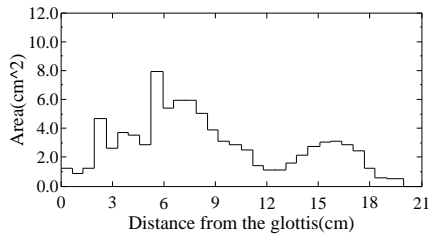
被験者 B が /u/ を発声している声道画像に対し FEM を用いて 1Hz, 100Hz, 240Hz, 250Hz の音圧分布・音響インテンシティを求め、その音圧分布・音響インテンシティを用いて提案法を適用したときの声道面積関数を示している。



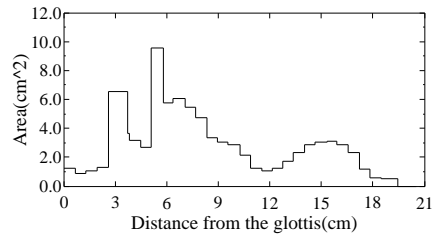
(a) 265Hz 声道面積関数



(b) 280Hz(F_1) 声道面積関数



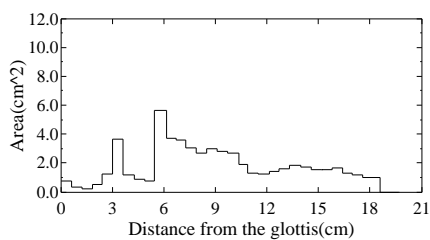
(c) 285Hz 声道面積関数 (等音圧面除去あり)



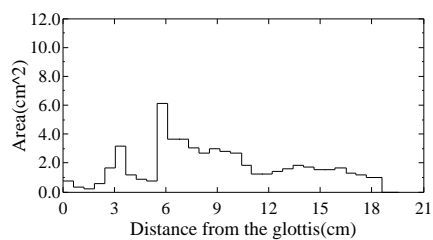
(d) 300Hz 声道面積関数 (等音圧面除去あり)

図 4.33: $/u/_{B}$ の声道画像の F_1 付近の音圧・音響インテンシティ分布を用いたときの声道面積関数

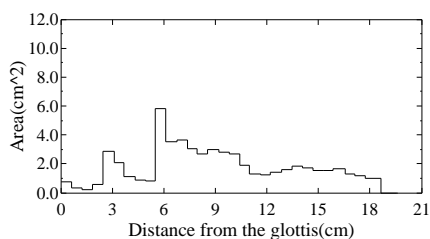
被験者 B が $/u/$ を発声している声道画像に対し FEM を用いて第 1 ホルマント付近の音圧分布・音響インテンシティを求め、その音圧分布・音響インテンシティを用いて提案法を適用したときの声道面積関数を示している。285Hz, 300Hz において声帯付近で帯状ではない等音圧面が発生したため、(c) 及び (d) はその等音圧面の除去を行った結果を示している。



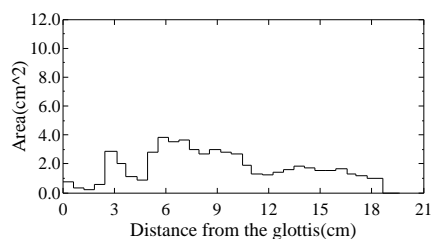
(a) 1Hz 声道面積関数



(b) 200Hz 声道面積関数



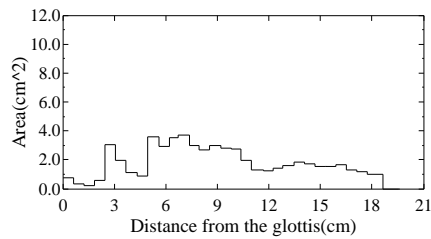
(c) 360Hz 声道面積関数



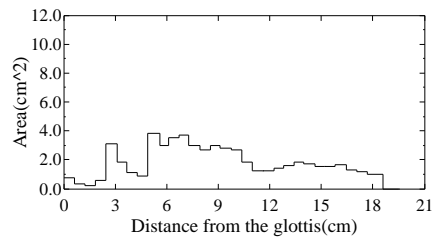
(d) 380Hz 声道面積関数

図 4.34: /e/_B の声道画像の 1Hz, 200Hz, 360Hz, 380Hz の音圧・音響インテンシティ分布を用いたときの声道面積関数

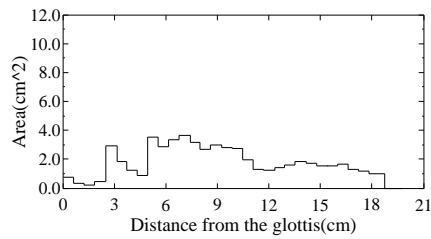
被験者 B が /e/ を発声している声道画像に対し FEM を用いて 1Hz, 200Hz, 360Hz, 380Hz の音圧分布・音響インテンシティを求め、その音圧分布・音響インテンシティを用いて提案法を適用したときの声道面積関数を示している。



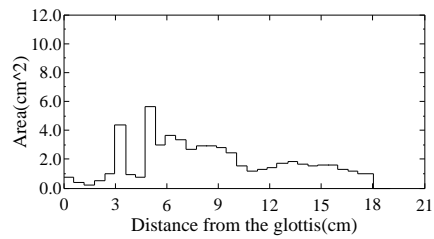
(a) 400Hz 声道面積関数



(b) 412Hz(F_1) 声道面積関数



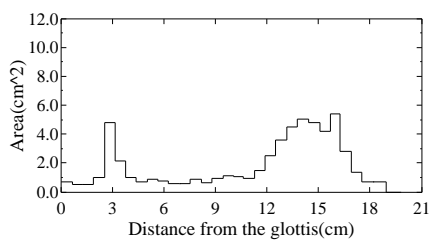
(c) 418Hz 声道面積関数 (等音圧面除去あり)



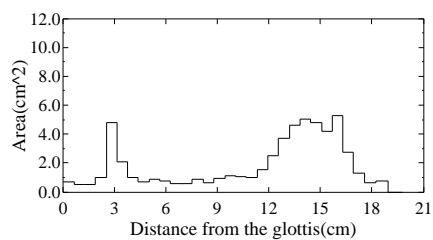
(d) 450Hz 声道面積関数 (等音圧面除去あり)

図 4.35: $/e/_B$ の声道画像の F_1 付近の音圧・音響インテンシティ分布を用いたときの声道面積関数

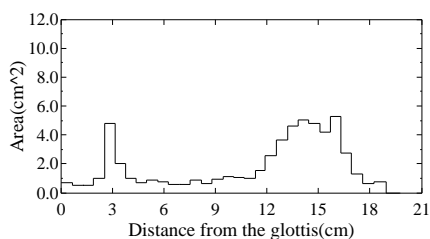
被験者 B が $/e/$ を発声している声道画像に対し FEM を用いて第 1 ホルマント 付近の音圧分布・音響インテンシティを求め、その音圧分布・音響インテンシティを用いて提案法を適用したときの声道面積関数を示している。418Hz, 450Hz において声帯付近で帯状ではない等音圧面が発生したため、(c) 及び (d) はその等音圧面の除去を行った結果を示している。



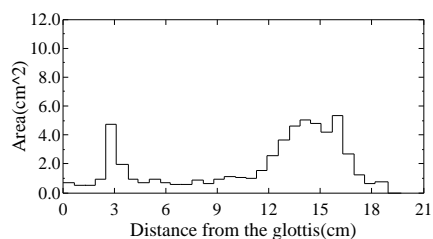
(a) 1Hz 声道面積関数



(b) 100Hz 声道面積関数



(c) 200Hz 声道面積関数



(d) 300Hz 声道面積関数

図 4.36: /o/_B の声道画像の 1Hz, 100Hz, 200Hz, 300Hz の音圧・音響インテンシティ分布を用いたときの声道面積関数

被験者 B が /o/ を発声している声道画像に対し FEM を用いて 1Hz, 100Hz, 200Hz, 300Hz の音圧分布・音響インテンシティを求め、その音圧分布・音響インテンシティを用いて提案法を適用したときの声道面積関数を示している。

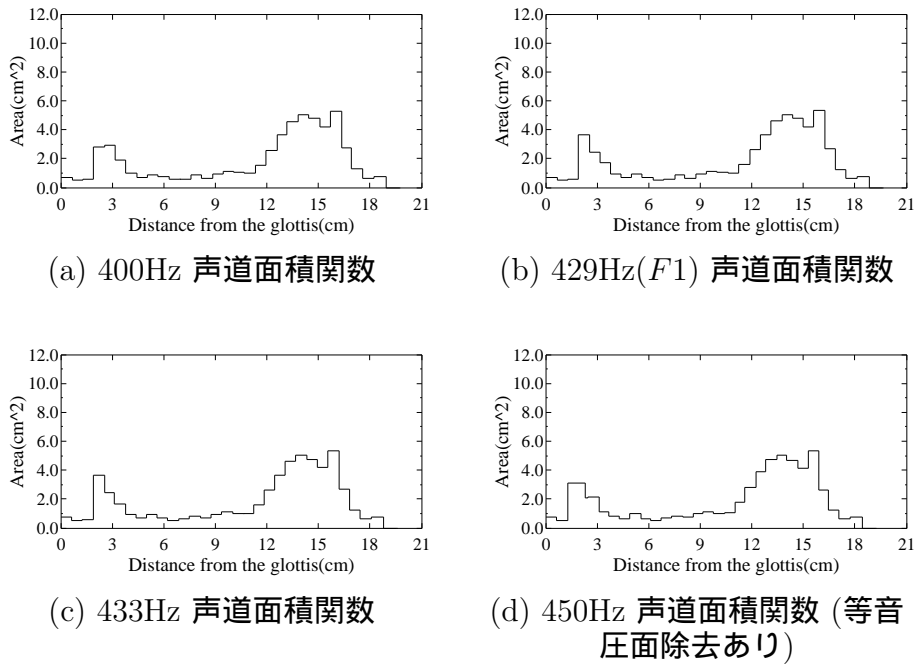


図 4.37: /o/_B の声道画像の F_1 付近の音圧・音響インテンシティ分布を用いたときの声道面積関数

被験者 B が /o/ を発声している声道画像に対し FEM を用いて第 1 ホルマント付近の音圧分布・音響インテンシティを求め、その音圧分布・音響インテンシティを用いて提案法を適用したときの声道面積関数を示している。450Hz において声帯付近で帯状ではない等音圧面が発生したため、(d) はその等音圧面の除去を行った結果を示している。

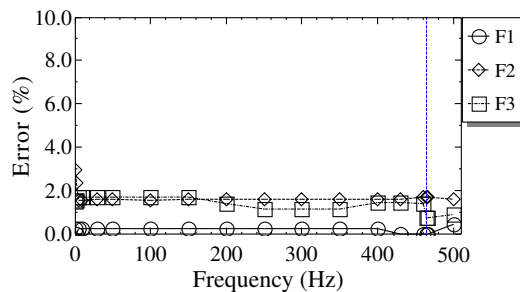


図 4.38: 被験者 A /a/ における音圧・音響インテンシティ分布導出に用いた周波数と誤差率

FEM より得られる $F1$ と声道面積関数より得られる $F1$ の誤差率を縦軸に，音圧・音響インテンシティ導出に用いた周波数を横軸にとったグラフである．グラフ中の縦線は FEM によって得られた $F1$ の値を示している．

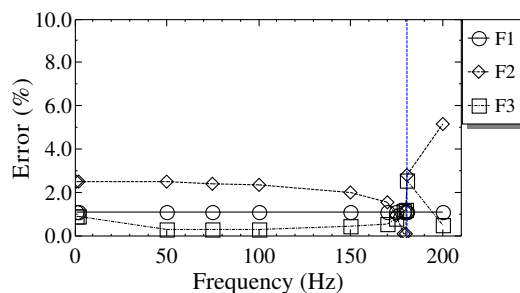


図 4.39: 被験者 A /i/ における音圧・音響インテンシティ分布導出に用いた周波数と誤差率

FEM より得られる $F1$ と声道面積関数より得られる $F1$ の誤差率を縦軸に，音圧・音響インテンシティ導出に用いた周波数を横軸にとったグラフである．グラフ中の縦線は FEM によって得られた $F1$ の値を示している．

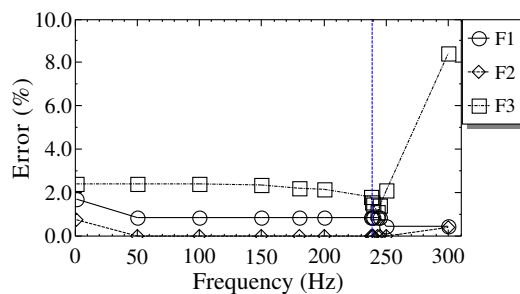


図 4.40: 被験者 A /u/ における音圧・音響インテンシティ分布導出に用いた周波数と誤差率

FEM より得られる $F1$ と声道面積関数より得られる $F1$ の誤差率を縦軸に, 音圧・音響インテンシティ導出に用いた周波数を横軸にとったグラフである. グラフ中の縦線は FEM によって得られた $F1$ の値を示している.

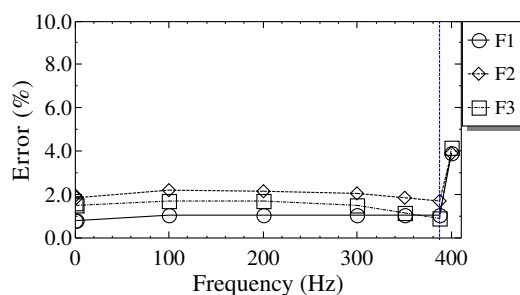


図 4.41: 被験者 A /e/ における音圧・音響インテンシティ分布導出に用いた周波数と誤差率

FEM より得られる $F1$ と声道面積関数より得られる $F1$ の誤差率を縦軸に, 音圧・音響インテンシティ導出に用いた周波数を横軸にとったグラフである. グラフ中の縦線は FEM によって得られた $F1$ の値を示している.

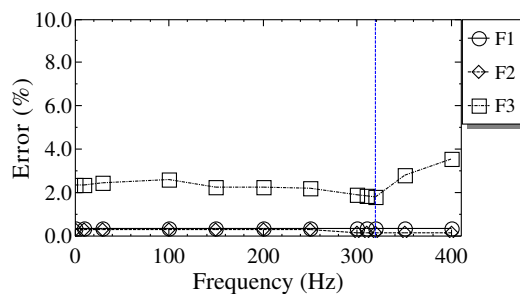


図 4.42: 被験者 A /o/ における音圧・音響インテンシティ分布導出に用いた周波数と誤差率

FEM より得られる $F1$ と声道面積関数より得られる $F1$ の誤差率を縦軸に, 音圧・音響インテンシティ導出に用いた周波数を横軸にとったグラフである. グラフ中の縦線は FEM によって得られた $F1$ の値を示している.

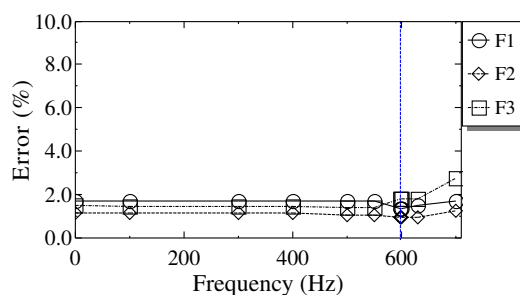


図 4.43: 被験者 B /a/ における音圧・音響インテンシティ分布導出に用いた周波数と誤差率

FEM より得られる $F1$ と声道面積関数より得られる $F1$ の誤差率を縦軸に, 音圧・音響インテンシティ導出に用いた周波数を横軸にとったグラフである. グラフ中の縦線は FEM によって得られた $F1$ の値を示している.

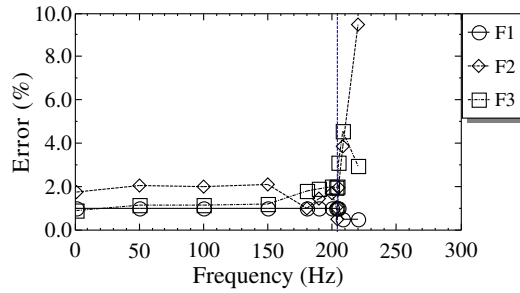


図 4.44: 被験者 B /i/ における音圧・音響インテンシティ分布導出に用いた周波数と誤差率

FEM より得られる $F1$ と声道面積関数より得られる $F1$ の誤差率を縦軸に、音圧・音響インテンシティ導出に用いた周波数を横軸にとったグラフである。グラフ中の縦線は FEM によって得られた $F1$ の値を示している。

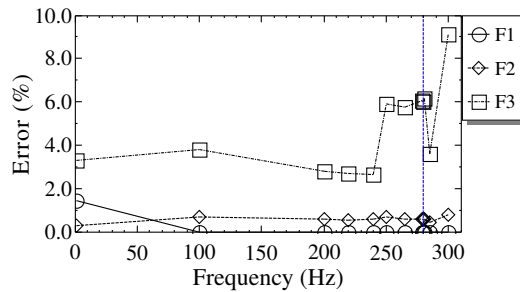


図 4.45: 被験者 B /u/ における音圧・音響インテンシティ分布導出に用いた周波数と誤差率

FEM より得られる $F1$ と声道面積関数より得られる $F1$ の誤差率を縦軸に、音圧・音響インテンシティ導出に用いた周波数を横軸にとったグラフである。グラフ中の縦線は FEM によって得られた $F1$ の値を示している。F3 の誤差率が急激に上昇している 250 Hz と その直前の 240 Hz の声道面積関数を図 4.32 に示す。

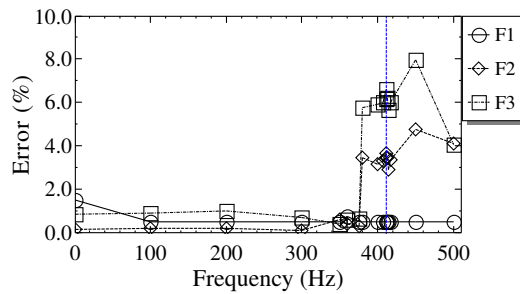


図 4.46: 被験者 B /e/ における音圧・音響インテンシティ分布導出に用いた周波数と誤差率

FEMより得られる $F1$ と声道面積関数より得られる $F1$ の誤差率を縦軸に、音圧・音響インテンシティ導出に用いた周波数を横軸にとったグラフである。グラフ中の縦線はFEMによって得られた $F1$ の値を示している。F2, F3の誤差率が急激に上昇している 380 Hz と その直前の 360 Hz の声道面積関数を図 4.34 に示す。

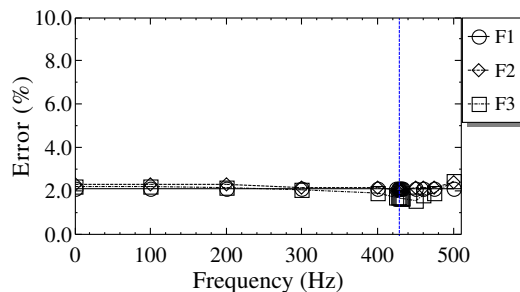


図 4.47: 被験者 B /o/ における音圧・音響インテンシティ分布導出に用いた周波数と誤差率

FEMより得られる $F1$ と声道面積関数より得られる $F1$ の誤差率を縦軸に、音圧・音響インテンシティ導出に用いた周波数を横軸にとったグラフである。グラフ中の縦線はFEMによって得られた $F1$ の値を示している。

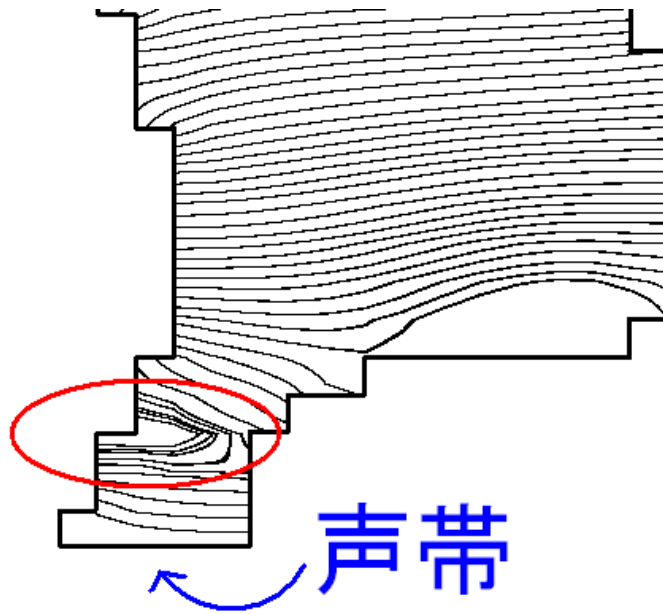


図 4.48: /i/_A における F_1 の音圧・音響インテンシティ分布を用いたときの等音圧面の形状

一般的に，丸で示した等音圧面の形状は F_1 より大きな周波数において観測されるが，今回用いた声道データの /i/ に関しては F_1 の時点ですでにこのような形状の等音圧面が現れる．

第5章 まとめ

本論文では，MRIデータより得られる3次元声道画像から，より正確な声道面積関数を推定する手法について述べた．提案法は，以下の特長を持つ．

- (1) 声道領域と入力面，出力面を与えることで自動的に声道面積関数を計算できる．
- (2) 等音圧面を用いて声道分割を行うため，2つの隣接する分割面が互いに交差しない．
- (3) FEMより得られる音圧分布を用いて声道面積関数を導出するため，FEMの解析結果と同等の声道伝達特性を持つ声道面積関数を推定できる．

この提案法の有効性を評価するために，まず，成人男性2名がそれぞれ日本語5母音/a/,/i/,/u/,/e/,/o/を発音しているMRIデータから声道領域を抽出し，提案法及び比較手法を適用してそれぞれの声道面積関数を求め，声道面積関数を音響等価回路に適用することで得られるホルマント周波数とFEMより得られるホルマント周波数との誤差を比較したところ，平面で分割する比較手法のホルマント周波数の誤差率は平均6.5%，標準偏差3.3%であったのに対し，提案法の誤差率の平均は1.7%，標準偏差1.4%となり，誤差率の大きさ，ばらつきともに良好な結果が得られ

ることが示された。また、口唇付近における補正の効果を確認するために、口唇付近の補正を施す場合と施さない場合について同様に比較を行ったところ、補正を施した場合の誤差率の平均は 1.7%、標準偏差 1.4% であるのに対し、施さない場合の誤差率の平均は 2.4%、標準偏差 1.6% となり、補正を施した方が良好な結果が得られることが示された。さらに、提案法において、どの周波数の音圧、及び音響インテンシティの分布を用いて声道面積関数を導出することが最適であるかについて検討を行った。1Hz から $F1$ 付近までの周波数で声道面積関数を計算し、得られるホルマント周波数と FEM から得られるホルマント周波数との誤差が最も小さくなる周波数を調査したところ、多くの場合、 $F1$ 付近で最小値となるが、複雑な形状では、 $F1$ より少し低い周波数で最小値となる、すなわち、 $F1$ 、または $F1$ より少し低い周波数の音圧、及び音響インテンシティの分布を用いることが最適であることが示された。

本論文で示した研究成果を利用することで、今後、声道面積関数を用いた調音点の移動を観測することが可能となる。これは、時間変化を伴う MRI データから提案法を用いて各時点の声道面積関数を求め、それぞれの声道面積関数から調音点を計算することで実現できる。また、現在までに得られている音声波形から推定されている断面積関数について比較検討ができ、さらに直接的に音声の生成での断面積関数ができるため、音声生成の研究に役立つと期待される。

今後の課題

今後の課題として以下の項目が挙げられる。

- (1) 損失を考慮した音響等価回路から得られるホルマント周波数と録音音声のホルマント周波数との誤差が大きくなる要因の調査とその改善

(2) 調音点の観測と音声生成モデルの改善

それぞれの項目の詳細を以下に示す。(1)については、正確な声道面積関数の導出をテーマとしたとき、最終的な目標は声道面積関数から得られるホルマント周波数と実音声のホルマント周波数が一致することである。その目標と現状との距離を把握するために、4.3.3において提案法により得られた声道面積関数を損失を考慮した音響等価回路に適用し、得られたホルマント周波数と録音音声のホルマント周波数とを比較したところ、誤差率が1%台、またはそれ以下となるデータが30個のうちの3個であり、誤差率も平均8.8%、標準偏差5.8%となり、誤差が大きい。実音声のホルマント周波数との誤差をより小さくするために、誤差を生じる要因をより詳細に調査し、それに基づき改善する必要がある。例えば、実際の声道は部位により声道壁インピーダンスが異なるため、各部位で適切な声道壁インピーダンスを設定することで誤差を改善できる可能性がある。また、広母音では口腔内において頬の面積が大きくなるのに対し、狭母音では小さくなるため、これに伴う声道壁インピーダンスの差異を考慮することで誤差を改善できる可能性がある。(2)については、本論文で示した研究成果を利用することで、今後、声道面積関数を用いた調音点の移動を観測することが可能となる。これは、時間変化を伴うMRIデータから提案法を用いて各時点の声道面積関数を求め、それぞれの声道面積関数から調音点を計算することで実現できる。ただし、時間変化を伴う場合、現段階では撮像技術の限界により、3次元MRIデータを取得することが困難である。2次元MRIデータから自動的に声道面積関数を精度良く求める手法の開発も必要である。

謝辞

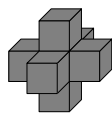
本研究の遂行，ならびに本論文の執筆に関して以下の各氏に心より感謝致します．まずはじめに，筆者の指導教員である静岡大学工学部，中井孝芳教授には，日々の研究や本論文の執筆にあたり，終始懇切なるご指導を賜り，ご激励を頂きました．また，静岡大学深林太計志教授，北澤茂良教授，下平美文教授にはご多忙の中，貴重な時間を割いて査読，ご助言を頂きました．静岡大学工学部 大橋剛介准教授には，本論文の査読，研究に関するご助言の他，研究以外の様々な問題に関する数多くの支援を頂きました．静岡大学工学部技術部技術専門職員 石田敬三氏には，計算機環境のメンテナンスや博士審査申請の提出準備等，数多くの支援を頂きました．また，静岡大学中井研究室のみなさまには，日々の研究に関しまして多大なるご協力を頂きました．そして，筆者の家族には，博士課程進学への理解と応援，その他様々な支援を受けました．以上の方々に改めて感謝の意を表します．

M R I データの一部は，(株) A T R の 正木信夫氏，本多清志氏のご好意により提供していただきました．ここに感謝します．また，本研究で使用した M R I データの一部は，A T R 人間情報科学研究所が独立行政法人情報通信研究機構からの研究委託「人間情報コミュニケーションの研究開発」に基づいて収録し，公表した『A T R 母音発話 M R I データ』

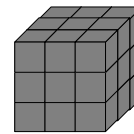
の一部です。本データの使用および成果の発表は、A T Rとの使用許諾
契約に基づいております。

付 録 A 3次元画像に対する細 線化アルゴリズム

本論文で述べた，3次元画像に対し細線化を施すアルゴリズム [36] の詳細を示す．このアルゴリズムでは，着目画素の他に，6近傍，及び26近傍画素の情報を利用している．着目画素を中心としたときのそれぞれの近傍画素の位置を図 A.1 に示す．



(a) 6近傍



(b) 26近傍

図 A.1: 3次元画像における近傍画素

(a) は着目画素に対する6近傍画素を示している．(b) は着目画素に対する26近傍画素を示している．灰色の立方体はそれぞれの近傍画素を表している．

処理手順を以下に示す．

(1) 配列，及びリスト変数を用意する．

list : 3次元座標と画素値を格納できるリスト．

$f(\text{locate})$: 2 値化画像 . 1 が物体で 0 が背景 . ただし , locate は 3 次元座標 .

$dmin$: list の画素値の最小値 . ただし , 画素値 0 は除外 .

$dmax$: list の画素値の最大値 .

- (2) 2 値化画像に対し , ユークリッド 2 乗距離変換を行い , 全ての非背景画素に 20 を加算する .

$f(\text{locate})'$: 距離変換を行い , $f(\text{locate}) = 0$ となる画素値は 0 を , それ以外の画素は距離変換後の画素値に 20 を加算した値を格納した画像 . locate は 3 次元座標 . 図 3.5(a) は距離変換後の画像の例である .

- (3) 初期境界画素を検出する .

全画素について , 6 近傍画素のうち 0 がある画素の座標 , 及び画素値を list に加える .

- (4) 永久保存点 (削除されない点) を検出する . また , 境界画素の分類を行う .

```
foreach item in list
  if (item.画素値 <= dmin)
    if (item.座標の画素が消去不可である)
      f(item.座標)' 16
    else if (f(item.座標) が端点である)
      item を list から削除
    else
      f(item.座標)' int(
        item.画素の 2 6 近傍の非 0 画素数/3) + 7
    endif
  endif
endif
end
```


(5) 画素を消去する .

```
for bt = 7 to 15
  foreach item in list
    if (item.画素値 == bt)
      if (item.座標の画素が消去不可である)
        f(item.座標)' = 16
      else if (item.座標の画素が端点である)
        item を list から削除
      else
        f(item.座標)' = 0 (画素の消去)
        item を list から除く
        foreach 座標 in item.画素の6近傍
          if (f(座標)' > 20)
            list に座標と画素値 f(座標)' を加える .
          else
            f(座標)' = 1
          endif
        end
      endif
    endif
  end
end
```

(5) 終了条件を判定する .

list の現在の画素の最小値 (19 以下の画素値は除外) よりも *dmax* の方が大きい, または *list* の画素値が 16 でない要素が一つでも含まれるときは (4) に戻る .

画素が消去不可であるかどうかを判定方法は [39] に記述されている . また, 端点であるかどうかの判定は, 着目画素の 2 6 近傍に非 0 画素がただ 1 つだけであるとき, 端点とする .

参考文献

- [1] 武田昌一，浅川吉章，市川薫，“残差音源型規則合成における女声音質改善方式の検討”，電子通信学会論文誌，vol.J73-A，pp.700-708，1990.
- [2] 武田昌一，浅川吉章，市川薫，“残差音源利用分析合成方式とマルチパルス法の基本特性の比較検討”，電子通信学会論文誌，vol.J73-A，pp.1735-1742，1990.
- [3] 濱上知樹，橋本新一郎，“非定常スペクトル推定による高品質音声合成”，日本音響学会講演論文集，vol.3-P-23，pp.337-338，1991.
- [4] 濱上知樹，“音源波形形状を高調波位相により制御する音声合成方式”，日本音響学会誌，vol.54，no.9，pp.623-631，1998.
- [5] 持田岳美，誉田雅彰，“口唇インパルス応答に基づく声道断面積関数の推定”，日本音響学会誌，vol.55，no.3，pp.147-155，1999.
- [6] 堤一男，加川幸雄，“声道断面積関数の音声信号からの推定 内部損失及び声道壁インピーダンスの影響とその補正”，日本音響学会誌，vol.56，no.1，pp.22-31，2000.
- [7] T. Chiba, and M. Kajiyama, The Vowel: Its Nature and Structure, Tokyo-Kaiseikan, Tokyo, 1941.

- [8] G. Fant, *The Acoustic Theory of Speech Production*, Mouton, Gravenage, 1960.
- [9] J. Heinz, and K. Stevens, "On the derivation of area functions and acoustic spectra from cineradiographic films of speech," *The journal of the Acoustical Society of America*, vol.37, p.1037, 1964.
- [10] T. Baer, J.C. Gore, L.C. Gracco, and P.W. Nye, "Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels," *J.Acoust. Soc. Am.*, vol.90, no.2, pp.799-828, 1991.
- [11] 松村雅史, 杉浦淳, "核磁気共鳴影像法 (MRI) による声道形状の計測," *電子情報通信学会論文誌 A*, vol.J75-A, no.7, pp.1115-1123, 1992.
- [12] C.S. Yang, H. Kasuya, and S. Kano, "Considerations on accurate measurement of vocal tract shapes using magnetic resonance imaging," *IEICE Trans.*, vol.J77-A, no.10, pp.1327-1335, 1994.
- [13] J. Dang, and K. Honda, "Morphological and acoustical analysis of the nasal and the paranasal cavities," *J.Acoust. Soc. Am.*, vol.96, no.4, pp.2088-2100, 1994.
- [14] S.S. Narayanan, and A.A. Alwan, "An articulatory study of fricative consonants using magnetic resonance imaging," *J.Acoust. Soc. Am.*, vol.98, no.3, pp.1325-1347, 1995.

- [15] B.H. Story, I.R. Titze, and E.A. Hoffman, "Vocal tract area functions from magnetic resonance imaging," *J. Acoust. Soc. Am.*, vol.100, no.1, pp.537-554, 1996.
- [16] M.K. Tiede, and H. Yehia, "A shape-based approach to vocal tract area function estimation," In *Proceedings of the ASA-ASJ 3rd Joint Meeting*, pp.861-866, 1996.
- [17] H. Yehia, and M. Tiede, "A parametric three-dimensional model of the vocal-tract based on **MRI** data," *Proceedings of the 1997 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp.1619-1622, 1996.
- [18] C.Y. Espy-Wilson, "Acoustic modeling of american english /r/," *J. Acoust. Soc. Am.*, vol.108, no.1, pp.343-356, 2000.
- [19] B.H. Story, I.R. Titze, and E.A. Hoffman, "The relationship of vocal tract shape to three voice qualities," *J. Acoust. Soc. Am.*, vol.109, no.4, pp.1651-1667, 2001.
- [20] S. A., L. V., M. T., and D. D., "Mid-sagittal cut to area function transformations: Direct measurements of mid-sagittal distance and area with **MRI**," *Speech Communication*, vol.36, pp.169-180, 2002.
- [21] J. Brunner, S. Fuchs, and P. Perrier, "The influence of the palate shape on articulatory token-to-token variability," *ZAS Papers in Linguistics*, vol.42, pp.43-67, 2005.

- [22] P. Badin., and A. Serrurier , “Three-dimensional modeling of speech organs: Articulatory data and models ” 電子情報通信学会技術研究報告 , vol.36 , no.5 , pp.421-426 , 2006.
- [23] C.X. Lu , T. Nakai, and H. Suzuki , “A numerical simulation of the effects of the actual lip geometry on acoustic fields by a three-dimensional F E M ” IEICE Trans. , vol.E77-A , no.2 , pp.422-428 , 1994.
- [24] 堤一男 , 加川幸雄 , “声道の音響間モデル近似に関する二 , 三の考察 その2 有限要素モデルとの比較 ” 日本音響学会誌 , vol.54 , no.1 , pp.101-110 , 1998.
- [25] T. Nakai , S. Morikawa , Y. Suzuki, and K. Ishida , “Estimation of effective cross sectional area from 2-d vocal tract using F E M ” The 18th International Congress on Acoustics ICA 2004 , pp.III-2306-2309 , Kyoto , 4 2003.
- [26] 松崎博季 , 元木邦俊 , 三木信弘 , “有限要素法における 3次元声道モデルの曲がり及び断面形状の簡略化に関する検討 ” 日本音響学会誌 , vol.59 , no.8 , pp.440-449 , 2003.
- [27] 山崎憲 , 瀬谷浩一郎 , “開放型段付き管内音圧分布の有限要素計算と実験 ” 日本音響学会誌 , vol.46 , no.9 , pp.711-715 , 1990.
- [28] 北村達也 , 西本博則 , 藤田覚 , 本多清志 , “音響計測、電気回路理論および有限要素法により求めた声道モデルの伝達関数の比較 ” 電子通信学会技術研究報告 , vol.103 , no.26 , pp.37-42 , 2003.

- [29] 西山静男，池谷和夫，山口善司，奥島基良，音響振動工学，コロナ社，1979.
- [30] 高橋信之，中井孝芳，鈴木久喜，“有限要素法による音響管の要素分割と計算精度” 静岡大学大学院電子科学研究科研究報告第 18 号，vol.18，pp.41-44，1996.
- [31] P. Elias, “Predictive coding,” IRE transactions on information theory, vol.1, no.1, pp.16-24, 1955.
- [32] 電子通信学会，デジタル信号処理，コロナ社，1975.
- [33] 大津展之，“判別および最小 2 乗基準に基づく自動しきい値選定法” 電気通信学会論文誌，vol.J63-D，no.4，pp.349-356，1980.
- [34] 八木伸行，井上誠喜，林正樹，中須英輔，三谷公二，奥井誠人，鈴木正一，金次保明，C 言語で学ぶ実践画像処理，オーム社，1992.
- [35] 高木幹雄，下田陽久，新編画像解析ハンドブック，東京大学出版会，2004.
- [36] 斉藤豊文，森健策，鳥脇純一郎，“ユークリッド距離変換を用いた 3 次元デジタル画像の薄面化および細線化の逐次型アルゴリズムとその諸性質” 電子情報通信学会論文誌 D-II，vol.J79-D-II，no.10，pp.1675-1685，1996.
- [37] 党建武，中井孝芳，鈴木久喜，“声道内音圧及び頬の振動加速度による頬インピーダンスの測定” 日本音響学会誌，vol.48，no.9，pp.621-628，1992.

- [38] M.M. Sondhi, "Model for wave propagation in a lossy vocal tract,"
J. Acoust. Soc. Am., vol.55, no.5, 1974.
- [39] 米倉達彦, 横井茂樹, 鳥脇純一郎, 福村晃夫, "3次元デジタル空間における1-要素の消去可能性と図形収縮について," 電子情報通信学会論文誌 D-II, vol.J65-D-II, no.12, pp.1543-1549, 1982.

本研究に対する発表論文

論文

- [1] 望月圭太, 中井孝芳, “声道解析のための音響パラメータの可視化”, 静岡大学大学院電子科学研究科研究報告, 第 27 号, pp.55–58 (Mar. 2006)
- [2] K.Mochizuki, T.Nakai, “Estimation of area function from 3-D magnetic resonance images of vocal”, Acoustical Science and Technology, Vol.28, No.5, pp.346-348 (Sep. 2007)
- [3] 望月圭太, 中井孝芳, “等音圧面形状に基づく声道面積関数の推定”, 日本音響学会誌, 採録決定

国際会議

- [4] K.Mochizuki, T.Nakai, “Estimation of Area Function for Three-dimensional Vocal Tract Using Finite Element Method”, 4th Joint Meeting of ASA/ASJ, p.3372, Hawaii, USA(Dec. 2006)
- [5] T.Nakai, K.Mochizuki, “Estimation of area function using pressure contours and sound intensity in 3-D vocal tract”, The 19th Interna-

tional Congress on Acoustics ICA 2007, p.152, Madrid, Spain(Sep. 2007)

国内口頭発表

- [6] 望月圭太, 中井孝芳, “有限要素法を用いた三次元モデルの声道面積関数の推定”, 電子情報通信学会技術報告, vol.106, No.177, SP2006-22, pp.11–14 (Jul. 2006)
- [7] 望月圭太, 中井孝芳, “有限要素法を用いた三次元声道モデルにおける声道面積関数の推定”, 日本音響学会 2006 年秋季研究発表会講演論文集, pp.163–164 (Sep. 2006)
- [8] 望月圭太, 中井孝芳, “F 1 以外の周波数の音圧分布と声道面積関数との関係”, 日本音響学会 2007 年春季研究発表会講演論文集, pp.259–260 (Mar. 2007)