

デジタルウェーブガイドメッシュを用いた 3次元MRIデータからの声道断面積関数の抽出*

井上謙次 (大阪工業大院・情), 竹本浩典 (ATR-Promotions/ATR-CIS),
北村達也 (甲南大), 正木信夫 (ATR-Promotions), 中嶋鴻毅 (大阪工業大・情)

1 はじめに

近年, 磁気共鳴画像法 (Magnetic Resonance Imaging: MRI) を利用した発話器官の形態・動態の観測技術の発達により, 発話中の精密な3次元声道形状が得られるようになってきた [1, 2]。しかし, 声道形状に基づく実時間で音声生成シミュレーションといった工学的応用では, 計算量の制約から現状では直接3次元声道形状をモデルとして扱うことは困難であり, シミュレーションに適した1次元の声道断面積関数を3次元の声道形状から推定することが行われる。

従来の Story ら [3] や Kröger ら [4] の手法では, 声道の中心線をグリッド断面の幾何学的中心から求めている。しかし, 音波の声門上部の伝播経路は必ずしも幾何学的中心を通らないことが示されている [5, 6]。中井ら [7] の手法では, 有限要素法を用いて第一フォルマント周波数付近での音響インテンシティから求められた音波の伝播経路から声道断面積を推定している。

しかしながら, 各推定手法の比較・検討は少なく, 標準的な手法は確立されていない。そこで, 本稿では, 波動方程式を時間領域で解くことによって音波の伝播経路を求め, その伝播経路に沿った中心線から声道断面積を推定するアルゴリズムとして, 波動方程式の離散化にデジタルウェーブガイドメッシュ(DWM)[8, 9]を用いる手法を提案する。本提案手法を用いてATR母音発話MRIデータ[10]から推定された声道断面積のフォルマント周波数を求め, 北村ら[11]の製作した声道模型の音響特性と比較し, 検討を行った。

2 声道断面積関数の抽出

本稿では, 声道領域の中心線を求めるアルゴリズムとして, 提案手法であるDWMによる方法の他, マンハッタン距離による方法[12], 領域

二分法[3], エッジ間最短距離点による方法[13]の4種類を実装し比較を行った。

中心線計算以外の前処理や中心線からの断面積関数の算出処理などは同一の処理を行った。本稿で用いた3次元MRIデータからの断面積関数の抽出手順を以下に示す。

1. ノイズ除去フィルタを適用
2. 閾値処理で空気と組織に分割
3. 声門および口唇の位置を目視にて指定
4. 声道領域および分岐管(鼻腔・梨状窩)領域を抽出
5. 声道領域の中心線を計算
6. 中心線に対してスプライン補間でスムージング処理を行う
7. スムージング処理された中心線に沿ってグリッド線を求める
8. 各グリッド線に対して3次元MRIデータをスライスする
9. 声門を始点として, スライスされた声道領域から気道領域のボクセル数を数え上げ, 断面積関数を求める

本節では, まず特別の説明を要する声道領域と分岐管領域の分離アルゴリズムについて述べ, その後中心線計算アルゴリズムについて詳述する。

2.1 声道領域と分岐管領域の分離

本稿で用いる中心線の計算法では, 空気から体組織を分離するだけでなく, 鼻腔, 梨状窩, および喉頭蓋谷の領域に関しても非声道領域としてあらかじめ除去しておく必要がある。本稿における実験ではこれらの分岐管領域を以下の手法で検出した。

まず, 後述するマンハッタン距離を用い, 声門からの距離マップ (distance map) を計算する。距離マップ $dm(p)$ は点座標 p における声門から

*Extraction of vocal tract area function from 3D MRI data using digital waveguide mesh. by INOUE, Kenji (Osaka Institute of Technology), TAKEMOTO, Hironori (ATR-Promotions/ATR-CIS), KITAMURA, Tatsuya (Konan Univ.), MASAKI, Shinobu (ATR-Promotions), NAKASHIMA, Hirotake (Osaka Institute of Technology).

のステップ距離を表す。次に、音波が声門から口唇まで後戻りせずに届く領域が声道領域であると考えられるので、以下のアルゴリズムで口唇から声門へと伝播経路を逆に辿ることで声道領域を抽出する。

1. $t = 0$ とし、口唇の点座標のみを持つ集合を $R(t) = R(0)$ とする。
2. 要素数 n の集合 $R(t)$ に属する全ての点 $p_i, 0 \leq i < n$ に対して、集合 $Q_i = \{q \mid q \in FN(p), q \notin R(t), dm(q) \leq dm(p_i), dm(q) \geq 0\}$ を求め、 $R(t+1) = R(t) \cup Q_1 \cup \dots \cup Q_n$ とする。ただし $FN(p)$ は点 p の 4-近傍を表す。
3. $R(t+1) = R(t)$ となった時点でアルゴリズムを終了し、 $R(t)$ を声道領域を構成する点集合とする。

また、閾値処理によって判別できなかった、声道領域内に孤立して存在する非声道領域は、声門につながっている非声道領域部分からの連結性を判定して削除した。

2.2 中心線の算出アルゴリズム

本稿では、音波の伝播経路に沿った中心線を算出するアルゴリズムとして、音波伝播の波動方程式を時間領域で解く DWM による提案方法と、音波伝播の近似として計算量の少ないマンハッタン距離による方法を用いた。

また、声道領域の幾何学的中心点から中心線を算出するアルゴリズムとして、領域二分法とエッジ間最短距離点による方法を用いた。

2.2.1 デジタルウェーブガイドメッシュ (DWM) による方法

DWM は時間領域有限差分法 (FDTD 法) の実装の一つであり、音波伝播の波動方程式を時間領域で効率的に解くことが可能な数値解析手法である。直交座標系における音圧 p に関する 2 次元もしくは 3 次元の波動方程式

$$\frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} = \frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} \quad (1)$$

に対して、各座標軸に沿ってメッシュを区切って解く (c は音速)。

本稿では、画像の各ボクセルに対してジャンクションを設定した (Fig. 1)。各ジャンクションはそれぞれの 4-近傍に対してウェーブガイドを持

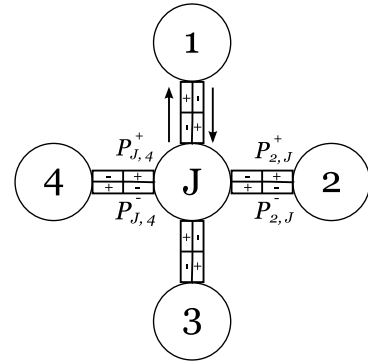


Fig. 1 4 接続のウェーブガイドジャンクション

つ。M 個のウェーブガイドを持つジャンクション J において、 k 番目のウェーブガイドへの出力 p_k^- は、 k 番目のウェーブガイドからの入力 p_k^+ 、単位ステップ時刻前における i 番目のウェーブガイドからの入力 $p_{J,i}^+$ に対して、

$$p_k^- = r_k p_k^+ + \sum_{i=1, i \neq k}^M (1 + r_i) p_{J,i}^+ \quad (2)$$

として定式化される。ここで r_k は k 番目のウェーブガイドに対する反射係数であり、

$$r_k = \frac{2Y_k - \sum_{i=1}^M Y_i}{\sum_{i=1}^M Y_i} \quad (3)$$

で与えられる。ここで Y_k はウェーブガイド k におけるアドミッタンスを表す。また、インピーダンス Z_k はアドミッタンス Y_k の逆数として次のように定義される。

$$Z_k = \frac{1}{Y_k} \quad (4)$$

入力には、時刻 $t = 0$ において、声門を構成する全ての各ボクセルに音圧 1.0 を設定し、それ以降はすぐに解放した。ただし、各隣接ウェーブガイド (2 次元では 4-近傍) への出力は等しく音圧 1.0 を接続ウェーブガイド数で割った値とした。

シミュレーションでは、声道領域内の各ボクセルについて、そのボクセルをその時刻までに通過したエネルギーの総和がある閾値 T_{dm} を初めて超えたステップ時間を距離として記録し、距離マップを求めた。閾値を適切に設定すれば、この手法で声門からの各距離における等音圧面を得ることができる。

中心線は、求めた距離マップに対して各距離ごとに重心点を計算することで得られる。ただし、そのままでは小さな変動が多いため、後述するスムージング処理を行った。

2.2.2 マンハッタン距離による方法

Takemoto ら [12] によって使われているマンハッタン距離による手法では、マンハッタン計量 (L_1 計量) を用いて声門からの距離マップを作成する。この方法は音波の伝達の近似解として考えることができる (Fig. 2)。

本稿では、スムージング処理に関して、DWM による方法に対して適用した後述のアルゴリズムをこの方法にも適用した。

2.2.3 領域二分法

領域二分法は、Story ら [3] によって使われている手法であり、幾何学的中心点を通る中心線を求める手法である。

2.2.4 エッジ間最短距離点の探索による方法

エッジ間最短距離点の探索による方法 [13] は、幾何学的中心点を通る中心線を求める手法である。

3 実験と評価

各種アルゴリズムの実験と評価には、ATR 母音発話 MRI データ [10] に収録されている成人男性 1 名の日本語 5 母音発話時の 3 次元 MRI データを使用した。これらの MRI データに対しては、北村ら [11] によって声道模型が製作され、その音響特性が分かっている。様々な要因が影響している実音声ではなく、声道模型のスペクトルと比較することで、3 次元の声道形状と推定された声道断面積関数の音響特性の差異を正確に評価可能である。

本稿では、母音/e/に関して声道断面積関数を抽出した実験について報告する。

まず、DWM による提案手法において、どのようなパラメータを用いれば精度良く声道断面積を推定できるかを調べた。距離マップを求めるための閾値 T_{dm} は固定とし、シミュレーションのステップ時刻 t に関して $T_{dm}(t) = 0.01/(t^2 + 1)$ に設定した。空気のインピーダンス Z_{air} は 1 に固定し、声道壁のインピーダンス Z_{wall} を 1, 10, 100, 1000, 10000 と変動させ、算出された距離マップの等距離線が等音圧線として利用できるかどうかを確認した。スムージング距離は、そのステップ距離ごとに重心をとって点を間引くための定数であり、距離を 10, 20, 30, 40, 50, 60 と変動させてシミュレーションを行った。

声門から引く中心線の終端は口唇付近から 1cm

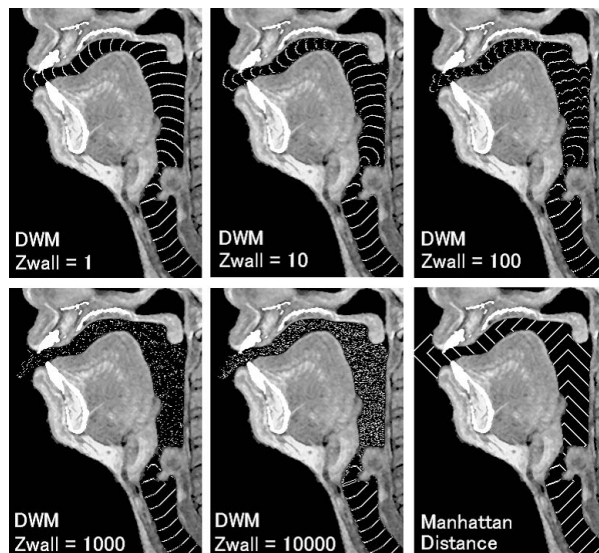


Fig. 2 母音/e/の等距離線 (上段: 左から DWM $Z_{wall} = 1, 10, 100$, 下段: 左から DWM $Z_{wall} = 1000, 10000$, 下段右: マンハッタン距離)

ほど離れた位置に設定した。そのため、声道断面積関数を求める際には、口唇側の端から無条件に 1cm 分を削除し、またさらにそこから 2.5cm^2 以上の領域を断面積関数から削除した。

パラメータを変更して得られた声道断面積関数から、Adachi and Yamada [14] によって提案された伝送線路モデルに Causseé ら [15] による放射インピーダンスの計算を組み入れたモデルを用いて伝達関数を計算した。ただし、北村らの声道模型に合わせるために剛壁とした。得られた伝達関数からピークピッキングにより第 1 ~ 第 4 フォルマントを求め、声道模型と比較し、最適なパラメータを決定した。

最後に、その最適なパラメータを用いて DWM により得られた声道断面積関数と他の 3 手法によって得られた声道断面積関数に対して上述の方法でフォルマント周波数を求め、声道模型の音響特性から得られたフォルマント周波数と比較した。

4 結果と考察

DWM とマンハッタン距離による手法を用いて得られた等距離線を図示する (Fig. 2)。 $Z_{wall} = 1000, 10000$ においては、DWM は波動伝播を示す等音圧線を算出することができなかったため、これらの条件は以降の実験では省いた。

声道壁のインピーダンス $Z_{wall} = 1, 10, 100$ の

Table 1 各手法と声道模型との母音/e/の第1～第4フォルマント周波数の比較 (Δ は声道模型との相対誤差 (%))

	声道長	F1	F2	F3	F4	$\Delta 1$	$\Delta 2$	$\Delta 3$	$\Delta 4$
声道模型		370	1629	2193	2684				
DWM	17.75 cm	410	1746	2262	2684	10.8	7.2	3.2	0
マンハッタン距離	18.5 cm	398	1699	2262	2719	7.6	4.3	3.2	1.3
領域二分法	16.75 cm	410	1594	1992	2754	10.8	2.2	9.2	2.6
エッジ間最短距離点	16.75 cm	422	1852	2332	2813	14.1	13.7	6.3	4.8

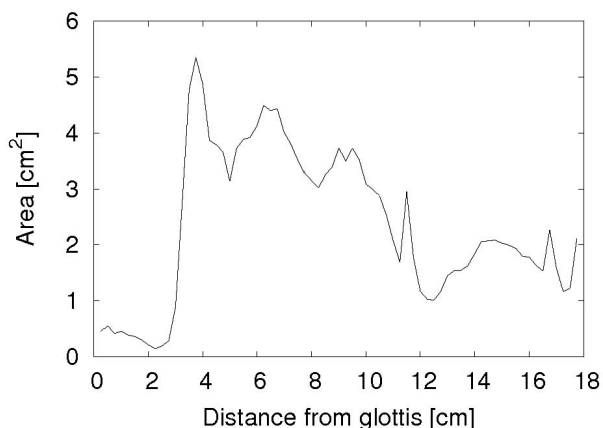


Fig. 3 DWMによって得られた母音/e/の声道断面積関数 (スムージング距離 10, $Z_{wall} = 100$)

各々に対して, スムージング距離を 10, 20, 30, 40, 50, 60 と変化させて断面積関数および伝達関数を求めた。その結果では, 声道壁のインピーダンスよりもスムージング処理の方がフォルマント周波数に影響する傾向が見られた。スムージング距離が高いところでは, スムージングを大きくかけ過ぎた結果, 声道長が短く算出されすぎている可能性がある。また, スムージング距離 10, $Z_{wall} = 100$ のときフォルマント周波数が最もよく一致している。このときの声道断面積関数を Fig. 3 に示す。

DWM以外の3手法も含めて, 声道模型のフォルマント周波数と比較した表を Table. 1 に示す。

5 おわりに

本稿では, DWM を用いて声道内の音波伝播の経路を計算し, その経路に沿った中心線から母音/e/の声道断面積関数を得た。声道模型の実測値と比較した結果, 元の3次元MRIデータの第1～第4フォルマント周波数とよく一致すること

が示された。

謝辞 本研究で使用したMRIデータは, ATR 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『ATR 母音発話MRIデータ』の一部である。本データの使用および成果の発表は, 株式会社ATR-Promotions との使用許諾契約に基づいている。

参考文献

- [1] 北村, 正木, 音響誌, 62(5), 385-390, 2006.
- [2] 正木他, 音講論(秋), 453-456, 2007.
- [3] Story *et al.*, JASA, 100(1), 537-554, 1996.
- [4] Kröger *et al.*, Proceedings of SPS 5, 333-336, 2000.
- [5] 中井他, 信学技報, HIP98-43, 31-38, 1999.
- [6] 鈴木他, 音講論(春), 247-248, 2001.
- [7] 中井他, 信学技報, SP2002-57, 1-4, 2002.
- [8] Van Duyne, S. and Smith, J., ICMC'93, 40-47, 1993.
- [9] Mullen, "Physical Modelling of the Vocal Tract with the 2D Digital Waveguide Mesh," Ph.D. thesis, University of York, UK, 2006.
- [10] ATR-Promotions, "ATR 母音発話MRIデータCD-ROM説明書第4版," 2007.
- [11] 北村他, 信学技報, EA2007-89, 19-24, 2007.
- [12] Takemoto *et al.*, JASA, 119(2), 1037-1049, 2006.
- [13] Mochizuki, K. and Nakai, T., Acoust. Sci. & Tech., 28(5), 346-348, 2007.
- [14] Adachi, S. and Yamada, M., JASA, 105, 2920-2932, 1999.
- [15] Caussé *et al.*, JASA, 75, 241-254.