

立体内視鏡とハイスピードカメラによる声門面積関数の計測

今川 博*・榊原健一***・徳田 功***・
大塚満美子****・田山二郎*****

Estimation of Glottal Area Function Using Stereo-Endoscopic High-Speed Digital Imaging

Hiroshi IMAGAWA*, Ken-Ichi SAKAKIBARA***, Isao T. TOKUDA***,
Mamiko OTSUKA**** and Niro TAYAMA*****

SUMMARY: Glottal area function provides important information in clarifying physical mechanisms of vocal fold vibration and investigating voice qualities in a quantitative manner, whereas its estimation has been technically difficult. In this paper, estimation of glottal area function was conducted in vivo by stereo-endoscopic measurement of the larynx combined with a high-speed digital imaging technique. In the present data from a female speaker, the high-speed camera captured images at an image resolution of 768 (horizontal) × 352 (vertical) with a frame rate of 3750 fps for the sample duration of 10.12 s. Glottal length, width, and area of the female participant were estimated in three different fundamental frequencies (F_0 s).

キーワード：声門面積関数，ステレオ内視鏡，高速度デジタル撮像，喉頭，女声

1. はじめに

喉頭音源は、音声における非言語的情報である声質を規定する重要な要因である。喉頭音源の音響的特性は、通常、声帯振動様式によってほぼ決定され、したがって、声帯振動様式を生理学的方法により観察することが様々な声質の生成メカニズムを明らかにする上で必要である。

声帯振動様式の直接的な観察方法として、内視鏡を用いて声帯の振動を高速度で撮影し、発話中の話者における声門面積の時間的変化（声門面積関数）を計測する方法があるが、実際の計測には様々な課題があった。通常の内視鏡による像は2次元画像であり、したがって、声門の面積は画

像内では、内視鏡と喉頭との距離および内視鏡の光軸の傾きにより線型に変形し、映像から得られる声門面積関数は相対的な時間変化を示すものの、絶対値としての実測面積は求めることが不可能である。従来研究で、高速度映画撮影を用いて求められた声門面積関数は、時間変化する声門面積における最大値を基準として正規化された相対値、あるいは相対値に適当に設定した基準値としての面積を乗じたものであった (Timcke et al. 1958, Flanagan 1958)。

内視鏡を用いた高速度撮像により得られた喉頭像から声帯の長さや声門の面積を実測する方法として、これまで2つの方法が報告されている。一つは、2本の内視鏡を組み合わせたステレオ内視

* 東京大学医学部耳鼻咽喉科学教室 (Department of Otolaryngology, Faculty of Medicine, The University of Tokyo)

** 北海道医療大学心理科学部言語聴覚療法学科 (Department of Communication Disorders, Health Sciences University of Hokkaido)

*** 北陸先端科学技術大学院大学情報科学研究科 (School of Information Science, Japan Advanced Institute of Science and Technology)

**** クマダ・クリニック (Kumada Clinic)

***** 国立国際医療研究センター病院耳鼻咽喉科・頭頸部外科 (Department of Otolaryngology, Head and Neck Surgery, National Center for Global Health and Medicine)

鏡を用い、喉頭の立体計測を行う方法である。較正用の撮影画像をもとに立体視の較正を行い、対物レンズと被写体との距離を測り、被写体のサイズを測定する。この方法は喉頭の上下動や声帯の長さを計測する目的で実施されてきた (Sawashima et al. 1974, Fujimura et al. 1979, Honda et al. 1980)。もう一つは、内視鏡の視軸に沿って並行する複数のレーザービームを喉頭に投射しビーム間の距離を参照して測定する方法である (Schuster et al. 2005)。

ステレオ内視鏡を用いた喉頭の3次元観測は、1980年代に澤島らによりステレオ内視鏡を用いて行われ、歌唱時の声の高さによる声帯長の変化をフィルムにより観察した研究が報告されている (Sawashima et al. 1981, Sawashima et al. 1983, 西澤1989)。しかしながら、内視鏡立体視を声帯振動の高速度撮像に応用した例はこれまでほとんどなく、著者らにより最近着手されたところである (今川ほか2009)。

本稿では、ステレオ内視鏡と高解像度の高速度カメラとを組み合わせたステレオ視高速度デジタル撮影システムの概要、および、女性1名に関し声門面積関数の絶対値の計測を試みた結果を、報告する。

2. 装置および計測方法

2.1 ステレオ側視型硬性内視鏡

喉頭の立体視を目的に、臨床の場で通常使用されている外径9mmの側視型硬性内視鏡を左右に2本束ねたステレオ内視鏡を1980年代に澤島らが永島医科器械に製作させた(図1)。この内視鏡先端に対物レンズとプリズムがあり、70度下方の側視像が得られる。視野角は40度であり、2本の内視鏡の視軸間の距離は10mmである(図2)。照明光はキセノン光源からライトガイドでそれぞれの内視鏡先端に導かれている。このステレオ内視鏡を市販の焦点距離50-mm CCTVレンズに装着した後、高速度カメラに取り付けて喉頭の撮影を行った。



図1 計測に用いたステレオ側視型硬性内視鏡 (永島医科器械)

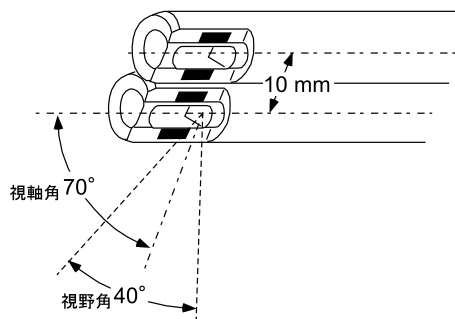


図2 内視鏡の視軸角、視野角、および視軸間隔

2.2 高速度カメラと立体視喉頭像

今回、高速度デジタル撮像カメラとしてFastcam 1024PCI (Photron)を用いた。このカメラのイメージセンサは17.4 mm × 17.4 mm、フルフレーム1024 × 1024 pixelsの解像度を持ち、フルフレーム撮影時の最高撮影速度は1000 fps、濃度諧調はモノクローム8 bitである。イメージ格納用メモリは12 GBで、1024 × 1024 pixels、1000 fps撮影の条件で9600フレームの画像を格納できる。

この高速度カメラは、イメージセンサの一部のみを使用し撮影領域を狭くすることにより、撮影速度をより高速に設定することが可能である。ステレオ内視鏡の2つの対物レンズから得られた喉頭像は高速度カメラのイメージセンサ上に左右に並んだ形で結像する。喉頭立体視像の撮影は768 (H) × 352 (V) pixels、3750 fpsの条件で行った¹⁾。この撮像条件では最長10.12秒間の撮影が可能である。図3に撮影された喉頭像の例を示す。

本稿で用いた3次元計測における座標の計算方法は従来研究と同様である (Fujimura et al. 1979)。

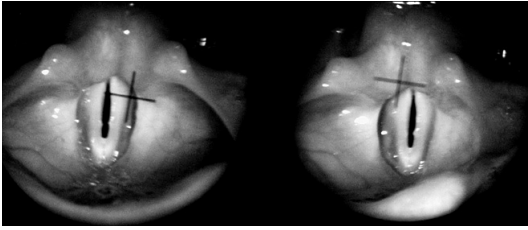


図3 ステレオ側視型硬性内視鏡による喉頭像

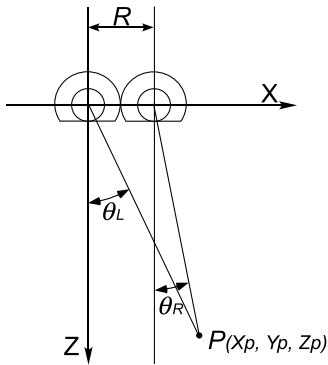


図4 3次元計測の幾何学

計測対象点 P の3次元座標 (x_p, y_p, z_p) は、内視鏡先端の視軸間距離 R と X 軸方向の対象点角 θ_R と θ_L および図には示されていない紙面を貫く方向である Y 軸方向の対象点角により決定される(図4)。図4において、3次元座標の原点は左の内視鏡の先端に設定されている。

図5に、撮影された対象点 P の左右内視鏡像における位置を示す。 D_L は左内視鏡像の点 P の X 座標、 D_V は Y 座標であり、 D_R は右内視鏡像の点 P の X 座標である。対象点 P の3次元座標は、理想的な光学系の場合に式(1)～(3)で計算することができる。

$$z_p = \frac{1}{k_1(D_L - D_R) + k_2} \quad (1)$$

$$x_p = k_3 z_p D_L \quad (2)$$

$$y_p = k_3 z_p D_V \quad (3)$$

2.3 較正と計測誤差

式(1)～(3)における複数の係数の値を決定するために、方眼紙を対象物とし、対物レンズか

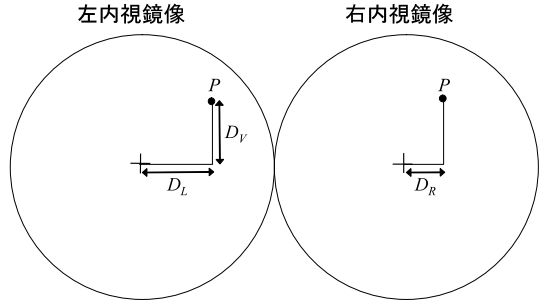


図5 内視鏡像における対象点の3次元座標

らの距離を変化させて撮影を行い、対象点座標の算出に必要な較正を行った。距離は14 mmから84 mmまで5 mmステップで変化させた。撮影された像を用い、撮影画像上の方眼紙の5 mm幅の交点の座標を左右視野の両方で調べ、最小自乗法によって、3つの計算式の k の値を求めた²⁾。方眼紙の交点の実際の3次元座標と[注]2)の式(4)～(6)より計算した交点座標との誤差を調べた結果、喉頭像の実際の撮影における対物レンズと声帯との距離として想定される59 mmまでは、 XY 軸の誤差は Z 軸の誤差の16%以下であり、3次元ユークリッド距離としての位置推定誤差は中央値0.55 mm (5 percentile: 0.15 mm, 95 percentile: 2.96 mm)であった。

3. 声門面積関数の計測

前節で述べた座標計算の方法は、声帯長計測に必要な対象点の座標を求める際に用いられたものと同様であるが、本稿における声門面積の計測には、3次元空間上、いくつかの幾何学的な補正を必要とする。

2次元喉頭像における声門は、急激に画素輝度が暗くなる両側声帯縁間の空間として定義される。しかしながら、3次元喉頭像においては、声門は、ひだ状の立体構造を持つ両側声帯の間の3次元の空間として実現されており、また、左右の内視鏡の2次元喉頭像における声帯遊離縁は必ずしも同一の曲線ではなく、また、水平断面に平行な同一の超平面上に存在するとも限らない。

3次元における声門は、2次元の領域としてではなく、3次元の領域として定義すべきであり、たとえば、声帯振動の上下の位相差を考慮して、3次元領域の声帯上部における断面、声帯下部における断面、それぞれの領域を別々に計測すべきである。しかし、喉頭上部からの立体視では、上部の狭窄は下部を視野から隠し、実際には、声帯の最内側部の遊離縁および視野に依存した片側声帯の上下方向の一部のみが観察可能である。

したがって、本稿では、3次元喉頭像における声門面積の算出において、声門を水平断面と平行な超平面上にある2次元領域にある点の集合とみなし、声門面積を計算することとする。

3.1 声門面積計測方法

声門面積を計測するために、まず、撮影した左右それぞれの視野における喉頭像の左右声帯遊離縁のエッジ抽出を行う。画像の水平走査線それぞれについて、声帯部分でありかつ輝度値が声門部の輝度値に近く、また、輝度微分値が最大となる点を声帯遊離縁の点と定める。自動抽出ができない場合は、視察によりエッジ位置の補足を行う。また、声帯遊離縁は前後方向に滑らかな曲線をなすと仮定し、前後方向に移動平均によるスムージング処理を行う。左右視野像において、内視鏡の光軸が声門の領域を定義する平面と直交しない側では、声門の一部が披裂部の陰になって見えない場合がある。その場合は、左右視野の両方で見えている部分に声門の領域を限定し、それぞれのエッジ点を同数に設定して「左右視野対応点」とみなして[注]2)の式(4)～(6)に従い3次元座標を計算する。

本方法の妥当性を確かめるために、図6に示すように0.25 mm厚の紙に横2.2 mm、縦13.2 mm、(面積29.0 mm²)の大きさに穴を開けた静止した直方体の領域を声門のモデルとして、面積の計測を行った。図7(a)に声門モデルの輪郭情報として計算された3次元図を示す。面積の計測結果は35.6 mm²であり、実値(29.0 mm²)とは相違が認められた。理由としては、左右の像において対応

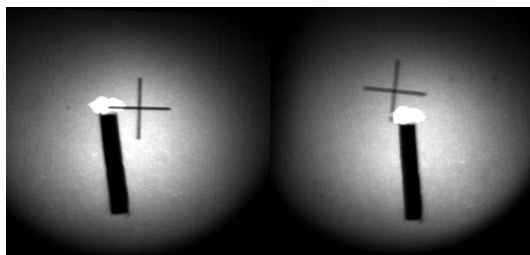
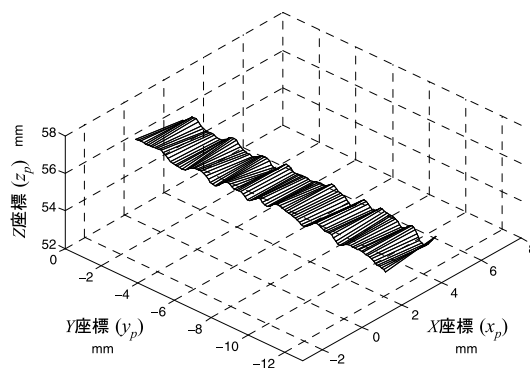
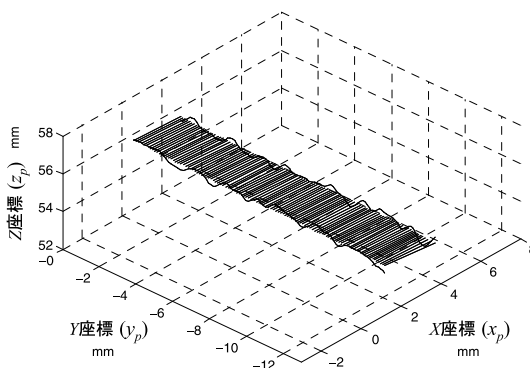


図6 声門のスリットモデルの撮像画像



(a)



(b)

図7 声門モデルの計測結果 (a) 平滑化した声帯遊離縁の抽出結果, (b) 声門傾斜を補正した後の声帯遊離縁の抽出結果

点であると推定された点には実際には同一点ではない場合があり、その場合、計算される3次元座標のZ方向の誤差を生じると考えられる。そこで、左右声帯遊離縁の midpoint (声門 midpoint) を声門長軸に沿って座標を求め、声門 midpoint に関しZ座標とY座標との回帰直線を求め、その直線により声門 midpoint のZ座標を決定する。次に、左右声帯遊離

縁を結んだ直線の側方傾斜を声門長軸に沿って求め、その平均値を求める（以下、この値を声門傾斜³⁾と呼ぶ）。最後に、声門中点の近似Z座標と声門傾斜を用いて左右声帯遊離縁Z座標を再計算し、声門を平坦な面として表現する（図7(b)）。これらの補正により、声門モデルの面積として、真値に近い29.5 mm²が得られた。

上記の方法を用いて、発声中の声帯振動の連続画像において声門面積を計算する場合、声門の開大が大きい場合に、声門傾斜が大きくなる傾向が見られた。声門傾斜が大きくなる理由は、声門開大時に、左右の喉頭像において、声門の法線方向と光軸との差が大きい方で、声帯遊離縁が厚みをもったものとして写り、推定された左右の遊離縁が上下方向に異なる平面上にあるからだと考えられた。そこで、各フレームの声門傾斜を記憶しておき、数回の声帯振動、あるいは異なる発声など複数フレームに渡る全体的な平均声門傾斜を計算する。最終的に、フレーム毎の回帰直線で近似した声門中点のZ座標と左右声帯遊離縁の全体的な平均声門傾斜を用いて左右声帯遊離縁Z座標を再計算し、声門を平坦な面に近似し、声門面積を計算した。

3.2 計測対象

1名の女性被験者（年齢：30代）が地声による通常の高さの声（ $F_0 = 230$ Hz）、高い声（ $F_0 = 450$ Hz）、低い声（ $F_0 = 100$ Hz）で母音を持続発声したときの喉頭像を撮影した。内視鏡挿入下での発声なので、/e/に近い母音となっている。高い声は地声声区又は中声区、低い声はフライ声区の聴覚印象であった。高速度画像データから、3ないし4周期分を選んで切り出し、声門面積、声門膜様部中央付近の最も声門幅が大きい場所の声門幅、声門前後長の時間変化を計算した。図8に3つの異なる F_0 での発声中の喉頭像を示す。計測区間内の平均声門傾斜³⁾は-0.3であった。なお、内視鏡像では安静呼吸時には声帯全貌を観測できるが、発声時は披裂軟骨の内転によって後方の声門軟骨部は披裂部に覆われるために観察す

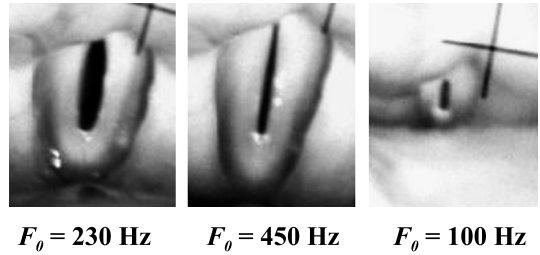


図8 異なる F_0 での女性の発声中の喉頭像

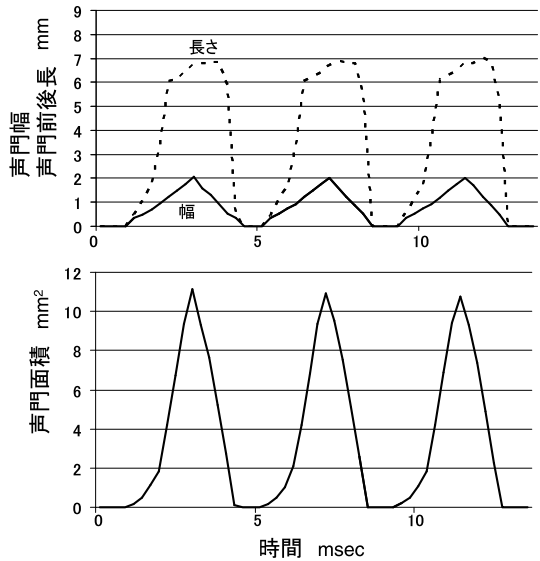


図9 $F_0 = 230$ Hzにおける声門前後長と声門幅（上）および声門面積（下）の時間波形

ることができない。内視鏡の構造上披裂上部よりの観察となるので、今回の観測例でも振動に関わる声門膜様部のみを観察、計算の対象とし、視界から隠れた声門軟骨部の開閉については声門部から除外した。

3.3 計測結果

図9に通常の高さ（ $F_0 = 230$ Hz）の発声における声門面積、最大声門幅と声門前後長の時間変化波形を示す。声帯振動1周期は画像16フレームで、画像から得られる声門開放時間率（OQ）は77%、声門面積の最大値は約11 mm²であった。また、声門幅の最大値は約2 mmであり、声門開大時に比べ声門閉鎖時の速度が若干速い三角形となっている。

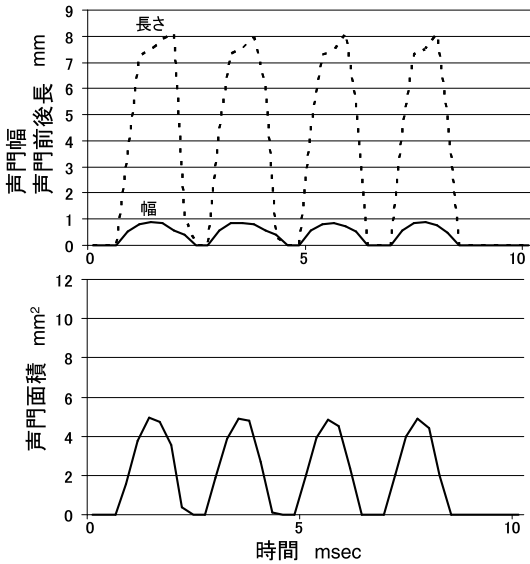


図10 $F_0 = 450$ Hzにおける声門前後長と声門幅（上）および声門面積（下）の時間波形

図10に高い声 ($F_0 = 450$ Hz) 発声時の声門面積、声門幅、声門前後長を示す。声帯振動1周期は画像8フレームで、画像から得られるOQは69%であり、通常の高さの声と比較してやや小さかった。声門面積の最大値は約 5 mm^2 で、 $F_0 = 230$ Hzの場合の1/2以下となっている。一方、声門前後長は1 mmほど長くなっており、また声門幅の時間変化波形は $F_0 = 230$ Hzの波形とは異なる丸い形状を示した。

低い声 ($F_0 = 100$ Hz) 発声時の声門面積、声門幅、声門前後長を図11に示す。声帯振動1周期は画像32フレームで、OQは16%であった。声門面積の最大値は約 1.3 mm^2 と非常に小さい値であった。

4. 考察

両眼視画像の立体構築処理において、左右視野対応点がないという状態が起こることがあり、両眼視におけるオクルージョンの問題といわれる。手前にある物体は背後にある物体を隠すので、片方の視野では見えているが、もう一方の視野では

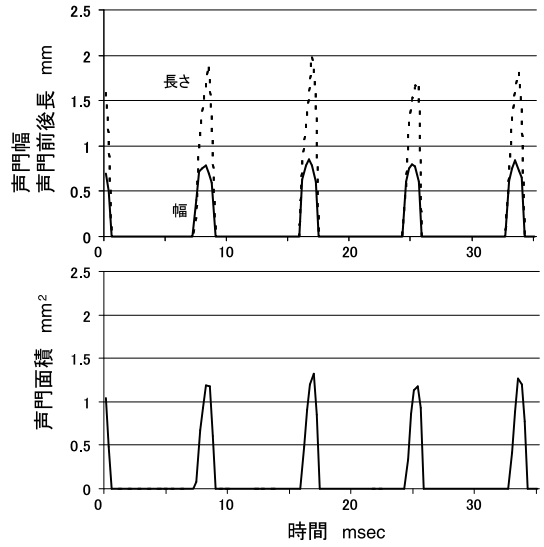


図11 $F_0 = 100$ Hzにおける声門前後長と声門幅（上）および声門面積（下）の時間波形

同じ物体が見えないという状態である。硬性内視鏡による喉頭の撮影において、内視鏡接眼部から声門までの距離は著者らの経験より、概ね 40 mm から 60 mm と考えられる。それに対して現在のステレオ内視鏡の視軸間隔は 10 mm と広く、声帯の上下方向の厚さと声带上唇下唇の振動の位相差によるオクルージョンがoccurり易い。そのため左右の声帯遊離縁の位置が異なり誤差の要因となる。同じ理由で、仮声帯と声帯のレベル差は容易に両眼立体視が可能であるが、声带上唇下唇レベルでの両眼立体視は困難である。

図12に、図9に示した通常の高さの発声の、左右声帯遊離縁の傾きの平均値（声門傾斜）の変化と声門面積波形を示す。声門面積が大きい時はこの傾きが大きい。これは上述のように撮影時のステレオ内視鏡の対物レンズの位置により片側の視野では声帯上方のエッジを、他方の視野では声帯下方のエッジを抽出してしまうことが原因と考えられた。フレーム毎の声門傾斜を使用して近似した声門面積を図12の破線で示す。一方、全フレームで同一平均声門傾斜を使った場合の近似結果を実線で示す。これは図9と同じものである。

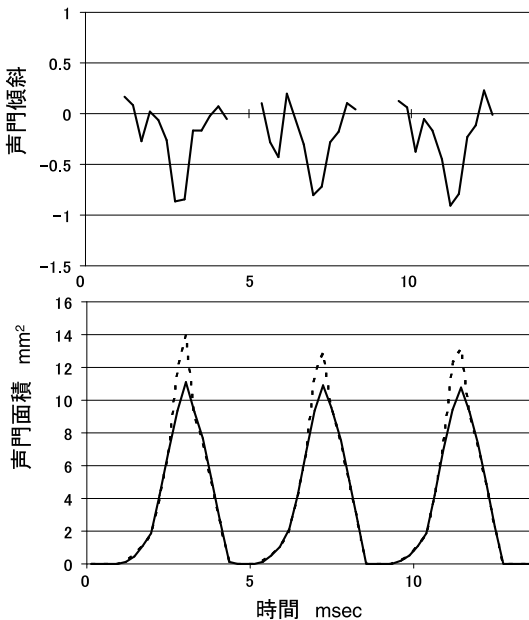


図 12 $F_0 = 230$ Hz の発声における左右声帯遊離縁の傾き (上図: 声門傾斜) と声門面積波形 (下図: 実線は全フレームにおける平均声門傾斜を, 破線はフレーム毎の声門傾斜を用いて算出したもの)

破線の最大声門面積は実線の約 20% 程度大きな値を示している。左右両視野の対応点をできるだけ多く撮影できるように、ステレオ内視鏡を喉頭内の最も良い位置に置くことが重要であるが、被験者により、狭い喉頭内では困難なときもある。したがって視軸間隔の狭いステレオ内視鏡の使用が望ましい。

西澤 (1989) によりステレオ内視鏡像の写真撮影により調べられた声の高さと声帯長の関係においては、4 名の女子の声帯長 (前連合から声帯突起までの長さ) の最小値は 6.0 mm から 7.3 mm, 最大値は 9.2 mm から 12.6 mm の範囲であった。西澤の計測では裏声も含まれており、すべての被験者で声の高さが高くなれば声帯長は長くなる傾向があったと述べている。本研究の計測結果では、通常の高さの発声において視認された部分の声門前後長は約 7 mm, 高い声では約 8 mm であり、西澤のデータ範囲内の値となっている。

Flanagan (1958) は高速度フィルムによる声帯

振動の観測を行い、男性 2 名の声門面積波形を求めた。これらの声門面積波形では声門閉小期と声門開大期のはっきりした時間差は見られない。また最大開口面積は 12 mm^2 から 20 mm^2 の範囲にあり、今回の計測結果よりもかなり大きい、男女差によるものと考えられる。

また OQ に関しては、低い声で著明に小さくなったが、これは、低い声がフライの聴覚印象に聞こえることから、フライ声区に特徴的な OQ の低値を示していると考えられる。

5. 結 語

ステレオ内視鏡と高速度カメラを用いたシステムにより、声門面積関数の絶対値の計測を試みた。従来研究で知られる声帯長、声門面積と比較して妥当な値を示した。

今後、ステレオ内視鏡の光学的精度を高め、数値処理にも誤差を最小化する方法を考案する必要があると考える。また、多くのデータを蓄積するために、画像処理の自動化も考慮する必要があると考える。本システムの将来に向けた応用としては、(1) 病変部位の大きさの計測 (臨床的応用); (2) 声門下圧, (口唇部) 体積流など他の物理パラメータとの同時計測; (3) 物理モデルのシミュレーションへの実測値の適用, などが今後の課題である。

謝 辞

本研究に関してご助力頂く、東京大学の横西久幸氏、二藤隆春講師、山嵜達也教授に感謝する。また、本研究に関して、ご議論頂いた国際医療福祉大学 新美成二教授に感謝する。本研究の一部は、総務省戦略的情報通信研究開発推進制度 (SCOPE: 071705001), 科研費 (20500161) および JAIST Grant for Exploratory Research の助成を受けた。なお、本論文の一部は INTERSPEECH 2010-ICSLP (11th International Congress on Spoken Language Processing, September 2010) において発表した。

〔注〕

- 1) 画像の解像度は対象物までの距離によって変わる。XY (左右, 前後) 平面およびZ軸 (奥行き) 方向の分解能はそれぞれ, 内視鏡先端対物部からの距離 44 mm で XY:0.09 mm, Z:0.4 mm, 距離 59 mm で XY:0.12 mm, Z:0.72 mm であった。
- 2) 現状のステレオ内視鏡の光学系には無視できない画像の歪みなどがあり, 式(1)~(3)に補正の式を加えた。それを式(4)~(6)に示す。右視野画像は左視野に比べ6%ほど画像が小さかったので式(1)において, 定数 c_1 を 1.06 とした。また同一 z_p 平面において, Y 軸方向の位置により $D_L - c_1 D_R$ の値が異なっていたため, $c_2 D_V$ の項でそれを補正する。 D_L が x_p に比例し, z_p に反比例 (z_p^{-1} に比例) する理想的な光学系を前提としたのが式(2)であるが, 実際は $z_p^{-0.91}$ に比例していた。式(5)の定数 c_7 はこれを補正するためのもので, 値は 0.91 である。補正のための定数 c を決定した後, 改めて式(4)~(6)の k の値を最小自乗法により求めた。

$$z_p = \frac{1}{k_1(D_L - c_1 D_R + c_2 D_V) - k_2} \quad (4)$$

$$x_p = k_3 f(z_p, D_L) + k_4$$
 where $f(z_p, D_L) = \frac{D_L - c_3 z_p^2 + c_4 z_p + c_5}{c_6 z_p^{-c_7}} \quad (5)$

$$y_p = k_5 z_p D_V + k_6 \quad (6)$$
- 3) 声門傾斜は「声帯遊離縁のZ軸方向の左右差/左右声帯遊離縁間のX軸方向の距離」で計算される。この値は正接関数値であり, 例えば -0.3 のときは $\tan^{-1}(-0.3) = -16.7^\circ$ に対応する。

参考文献

今川博・二藤隆春・山唄達也・榊原健一・木村美和子・田山二郎 (2009) 「ステレオ内視鏡と高速度ディジ

タルカメラによる喉頭立体視撮影システム」『音声言語医学』50: 1, 61-62.

西澤典子 (1989) 「ステレオ側視鏡による喉頭像の観察—呼吸時および定常発声時における声帯長変化—」『日本耳鼻咽喉科学会会報』92, 1239-1252.

Flanagan, J. L. (1958) "Some properties of the glottal sound source." *J. Speech and Hearing Res.* 1, 99-111.

Fujimura, O., Baer, T. and Niimi, S. (1979) "A stereo-fiberscope with a magnetic interlens bridge for laryngeal observation." *J. Acoust. Soc. Am.* 65, 478-480.

Honda, K., Hibi, S. R., Kiritani, S., Niimi, S. and Hirose, H. (1980) "Stereoendoscopic measurement of the laryngeal structure." *Ann. Bull. RILP* 14, 73-78.

Sawashima, M. and Miyazaki, S. (1974) "Stereo-fiberscopic measurement of the larynx: a preliminary experiment by use of ordinary laryngeal fiberoscopes." *Ann. Bull. RILP* 8, 7-10.

Sawashima, M., Hirose, H., Hibi, S., Yoshioka, H., Kawase, N. and Yamada, M. (1981) "Measurements of the vocal fold length by use of stereoendoscope—a preliminary study." *Ann. Bull. RILP* 15, 9-16.

Sawashima, M., Hirose, H., Honda, K., Yoshioka, H., Hibi, S. R., Kawase, N. and Yamada, M. (1983) "Stereoendoscopic Measurement of the Laryngeal Structure." In Bless, M. D. and Abbs, H. J. (eds.) *Vocal Fold Physiology, Contemporary research and Clinical issues* (pp.264-276). San Diego, CA: College-Hill Press.

Schuster, M., Lohscheller, J., Kummer, P., Eysholdt, U. and Hoppe, U. (2005) "Laser projection in high-speed lottography for high-precision measurements of laryngeal dimensions and dynamics." *European Archives of Oto-Rhino-Laryngology* 262, 477-481.

Timcke, R., von Leden, H., and Moore, P. (1959) "Laryngeal vibrations: measurements of the glottic wave, Part I." *AMA Arch. Otolaryngol.* 69, 438-444.

(Received Aug. 24, 2010, Accepted Aug. 25, 2010)