

TLS を用いた声道伝達関数の推定精度の
声道アナログ合成器による検討

佐々木 幸司 三木 信弘 小川 吉彦

北海道大学工学部

〒 060 札幌市北区北 13 条西 8 丁目

TEL:011-706-6532

E-mail sasaki@kairo3.hudk.hokudai.ac.jp

あらまし

声道アナログモデルを用いたスペクトル推定評価は音声生成過程の実体に即したモデルに基づくものとして非常に有効なものである。しかし、従来のモデルでは声門下のモデルが近似モデルであるために、十分に生成系の物理的な特徴量を反映しているとは言い難い。特に肺の構造に着目した、肺のモデル化は十分と言えるものは少ない。今回は肺の分岐を考慮した声道アナログモデルを用いて声道伝達関数の推定を行い、肺のインピーダンスがスペクトル推定にどのような影響を及ぼすのかを探る。

キーワード 肺の分岐構造、声道アナログモデル、声道伝達関数

Evaluation of Estimator with TLS for Vocal Tract-Function

by using Vocal Tract Analog Synthesizer

Kohji SASAKI Nobuhiro MIKI Yoshihiko OGAWA

Faculty of Engineering , Hokkaido University

N13 W 8 ,Kita-ku , Sapporo-shi , 060 Japan

Phone:011-706-6532

E-mail sasaki@kairo3.hudk.hokudai.ac.jp

Abstract

Evaluation of spectrum estimation using a vocal-tract analog model is an effective method because of being based on speech production process. But, the traditional vocal tract simulators, however, employ an approximated model, it is difficult to say that this does not reflect sufficiently physical characteristics of the production system. Especially, the lung structure should be taken account in the modeling with branches. We propose a vocal-tract analog model based on the lung branches and estimate the vocal-tract transfer function by using TLS estimate, and consider an effect of the lung impedance.

key words structure of lungs , vocal-tracts analog model , vocal-tract function

1 はじめに

声道伝達関数の推定は音声合成、音声認識において重要である。そのために正確な声道伝達関数を推定するために種々の手法が研究されている。一方、この推定実験に用いられる音声生成過程のモデル化には従来より二つの手法がある。一つには、線形分離モデルといわれるもので母音ではAR型、子音ではARMA型の回路によって声道部を実現させる。これは、声帯音源部、口唇放射部とも比較的簡単な回路で表現でき、時系列モデルとして扱える。もう一つには、声道アナログモデルと言われるものがあり、これは音声の生成過程を音波の伝搬現象としてシミュレーションをするものである。

音声合成における高品質、音声認識における高い理解度の実現には「自然性」、「個人性」、「情緒性」といったパラメータ化が必要である。これらの要素を音声生成過程の中でどのようにパラメータで表現するかという事は複雑で、難しい問題である。そこで、このような問題を研究する道具として人間の音声生成過程の実体に即したモデルを用いることが必要であると思われる。このような点において声道アナログモデルは線形分離モデルに比べて声道の断面積、長さ等の個人性の特徴量を反映させやすいモデルと言える。

しかし、声道アナログモデルが実体に即したモデル化であるにしても近似モデルであり、さらにより良い改良が必要である。特に声帯から下の気管・肺の部分は簡略化されていることが多く、等価回路的にはコンデンサのみで近似されている。実際には気管下部から見込んだ肺の入力インピーダンスがあり ([2], [3])、コンデンサのみによる近似は単純すぎる近似ではないかと思われる。

今回は気管が次々と分岐していく肺の構造を含めたアナログモデルを推定実験にモデルとし、肺のインピーダンスがスペクトル推定にどのような影響を及ぼすかという点に注目し、さらに肺の分岐構造についての対称な場合と非対称な場合との比較を行い、両者がスペクトルに与える影響を考察する。

2 モデル化について

2.1 肺のモデル

我々が従来用いてきた音源部のモデルでは、音源部に含まれる声帯から下の気管・肺の電気的な等価回路は一本の分布定数線路と1個のコンデンサ (C_l) で近似され、 C_l が音源の肺圧によって駆動されるものとして肺の収縮を表現していた。しかし、実際には肺は、その下部に進むにつれて次々と分岐を繰り返していき、最後に肺胞 (sac) で終端される。また、分岐していく際には、気管

はステイフネス、断面積、長さ等のパラメータの等しい気管に分かれていく (対称分岐) のもと単純化して一本の分布定数線路と見なしていた (図2)。

しかし、その分岐の仕方は一般には非対称 (ステイフネス、長さ、断面積等の物理的な量が異なる。) なものと考えられる ([2],[3])。そこで、我々は肺の構造を非対称なものと考えてモデル化を行う。

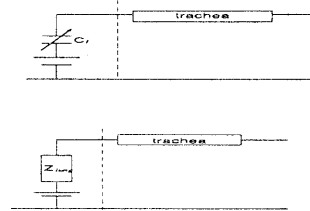


図1: 肺の等価モデル: 上図はコンデンサのみの近似、下図は入力インピーダンス Z_{lung} による近似

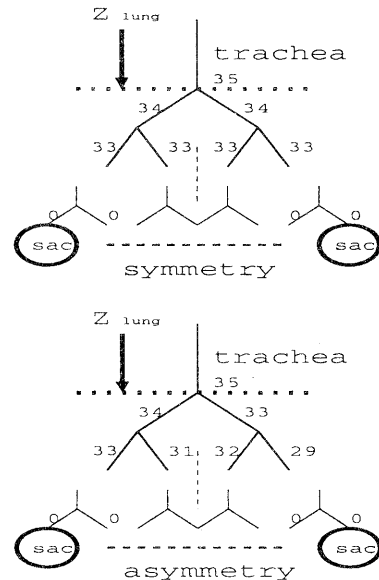


図2: 肺の分岐構造、上図は対称な分岐構造、下図は非対称な分岐構造

肺の分岐による複数の気管を分布定数線路とみなし、肺全体のインピーダンスを実測した例が既に多く報告されている ([2],[3])。また、その文献中では肺の中での各気管の断面積、直径、ステイフネス等が報告されている。我々はこれらの報告例にある肺の物理的なパラメータから壁インピーダンスを含めた粘弾性体の分布定数回路として、肺の分岐モデルを作成した。

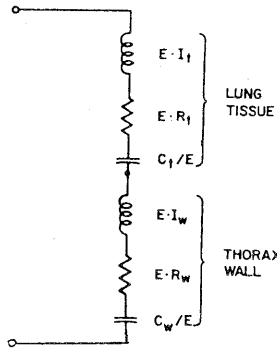


図 3: 肺胞を近似した電氣的等価回路

肺を分布定数線路と見なした場合の終端条件は、肺の最下部での気管に接続している肺胞である。肺胞は末端の気管(図2の0番)に数個接続しているものと思われる。我々は今回は肺胞の数を4($E=4$)とした。なお肺胞の数を変えた場合でも肺のインピーダンス Z_{lung} の共振周波数は変わらず、単に $|Z_{lung}(j\omega)|$ が大きくなるか、小さくなるかという違いだけであった。分布定数線路の終端条件となる肺の末端の肺胞での電氣的等価回路は図3のように表される。

この回路を用いて実測した非対称な肺の入力インピーダンスを図4に示す。また、非対称な場合の左右の肺と対称な分岐の場合の入力インピーダンスもあわせて示す。

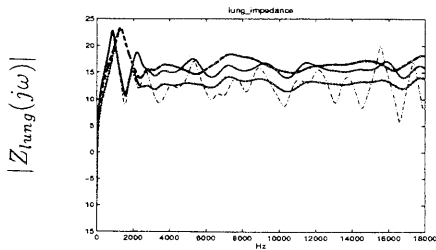


図 4: 実線(—)は35の下部から見込んだ肺のインピーダンス。点線(⋯)は34から見込んだ左肺のインピーダンス、鎖線(- -)は33から見込んだ右肺のインピーダンス、一点鎖線(- · -)は対称な分岐の場合の35の下部から見込んだ肺のインピーダンス

分岐を考慮した場合の分布定数線路を直接デジタルフィルタで表現することは困難なため、この気管下部から実測した肺の入力インピーダンス Z_{lung} を ARMA 型のデジタルフィルタで近似した入力インピーダンス

\hat{Z}_{lung} を肺の一部として挿入する。ARMA 型での近似は非対称な分岐のモデルでは AR 部 19 次、MA 部 21 次とし、対称な分岐では AR 部 31 次、MA 部 33 次とした。それぞれ対称・非対称な分岐の肺の入力インピーダンスの ARMA 型近似の様子を図5、図6に示す。また、あわせて ARMA 型による近似スペクトルの極・零の分布も示す。

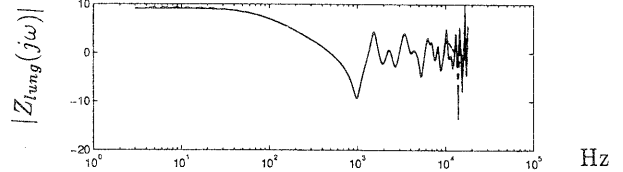
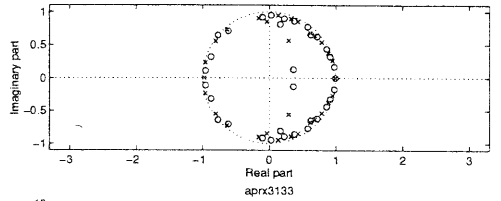


図 5: 下図: 実線(—)は実測による対称な分岐の肺のインピーダンス、鎖線(- -)は ARMA 型の近似によるスペクトル。上図: ARMA 型の近似による極・零の分布

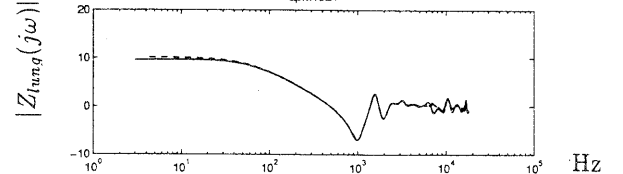
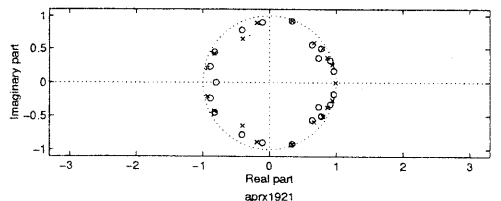


図 6: 下図: 実線(—)は実測による非対称な分岐の肺のインピーダンス、鎖線(- -)は ARMA 型の近似によるスペクトル。上図: ARMA 型の近似による極・零の分布

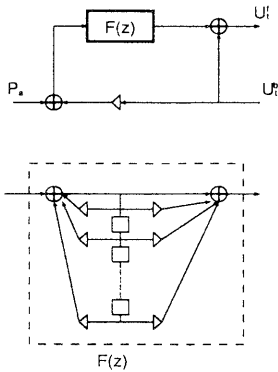


図 7: 肺の入力インピーダンスを近似した ARMA 型デジタルフィルタ

音声信号のスペクトルでは概して低周波数域にパワーが集中するので、肺のインピーダンスの近似では対称・非対称とも低周波数域でスペクトルが一致するようにした。高周波数域でわずかなずれが見られるが、近似としては十分ではないと思われる。

肺を ARMA 型のデジタルフィルタで近似したとき、肺と気管の接続モデルとして図 7 の $F(z)$ のように挿入される。

2.2 声帯音源部、声道部、口唇部のモデル

声帯音源部には Isizaka, Flanagan の提案による 2-mass モデル [1]、声道部分には Kelly 型のラティス型デジタルフィルタに基づき、空気粘性損失、熱損失を近似したものを縦続接続している ([5],[6])。口唇部には Morse らによって与えられた球形パツフルにおける放射インピーダンスに基づいた複素反射係数を AR 部を 1 次、MA 部を 2 次で近似したデジタルフィルタを用いる ([4])。

3 提案した声道シミュレータによる声道伝達関数の推定実験

3.1 推定手法

観測波形を AR パラメータに当てはめて声道伝達関数の推定を行う。推定手法としては Total Least Squares (TLS) を用いる。TLS は他の AR 型パラメータ推定法に比べて非常に良好な推定スペクトルが得られることを既に別な推定実験により確認している ([9])。

3.2 結果

以上で述べた肺を含めた声帯音源部と声道共振部、口唇放射部によるアナログ合成モデルを用い、出力信号に Kaiser 窓を掛けて切り出し、自己相関関数を求め、AR モデルに当てはめて声道伝達関数の推定実験を行った。実験には日本語の 5 母音を使用する。スペクトルの推定次数は 23 次とした。

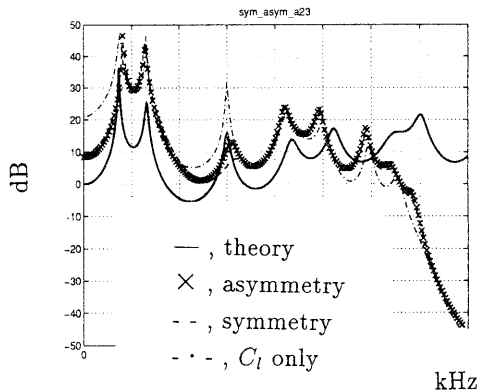


図 8: 日本語母音/a/を推定したもの、推定次数は 23 次、実線 (—) は理論値のスペクトル、× は非対称な分岐の肺を持つ場合の推定スペクトル、点線 (- -) は対称な分岐の肺を持つ場合の推定スペクトル、一点鎖線 (- · -) は肺の Z_{lung} をつけずに C_l のみで推定したもの

図 8 は日本語母音/a/を推定したものである。3kHz 付近で非対称な分岐を持つ肺の場合と対称な分岐を持つ肺の場合との相違が分かる。対称な分岐を持つ肺の推定スペクトルのパワーは低く、非対称な分岐を持つ肺のフォルマント周波数に比べて高周波数域に寄っている。また、肺のインピーダンスを単に C_l のみとした場合の推定スペクトルは 0Hz 付近でのパワーが非常に大きい。特に 3kHz 付近での第 3 フォルマントでの相違が顕著であることがわかる。 C_l のみの推定スペクトルは急峻なピークを示しているが、 Z_{lung} が存在する場合にはそのような傾向は見られず、対称な分岐での Z_{lung} ではさらに緩やかなピークを示していることがわかる。

図 9 は日本語母音/i/を推定したものである。さきほどの/a/の場合と異なり、対称な分岐の肺と非対称な分岐の肺との影響の差は見られない。 Z_{lung} をつけた合成モデルの推定スペクトルは、肺を C_l のみで近似した場合には見られなかった別なピークが 1kHz 付近に見られる。また、急峻なピークを示さず緩やかなピークとなっている。

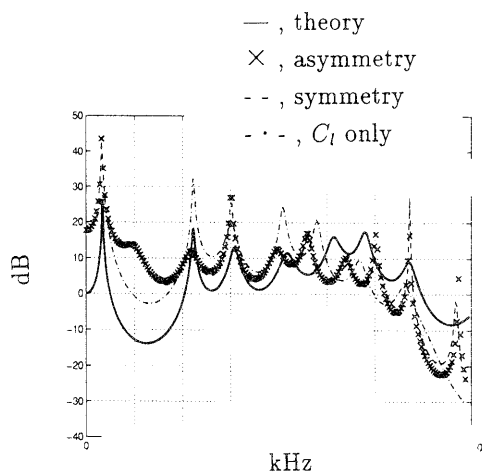


図 9: 日本語母音/i/を推定したもの、推定次数 23 次、線の表示/a/と同じ

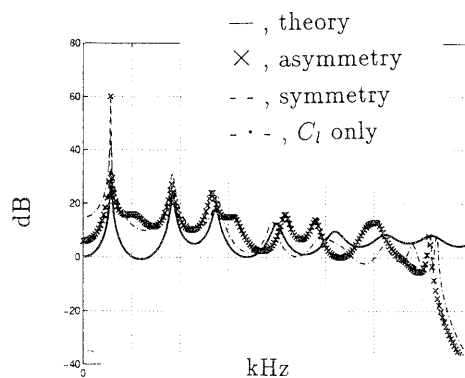


図 11: 日本語母音/e/を推定したもの、線の表示は/a/と同じ

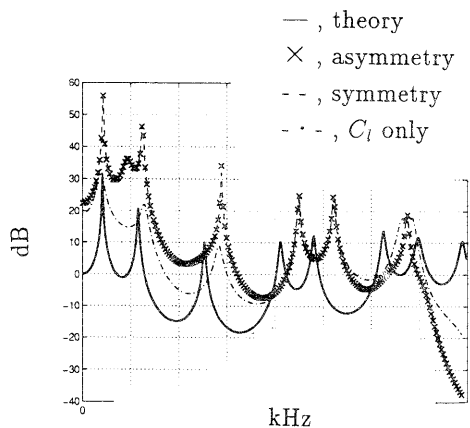


図 10: 日本語母音/u/を推定したもの、線の表示は/a/と同じ

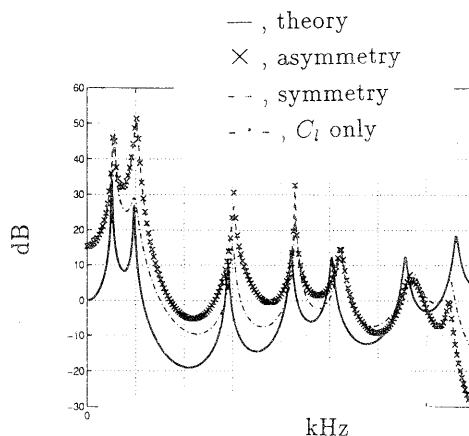


図 12: 日本語母音/o/を推定したもの、線の表示は/a/と同じ

図 10 は日本語母音/u/を推定したものであるが/i/の場合と同じように 1kHz 付近に別なピークが発生している。また、3kHz 付近での第 3 フォルマントのピークは C_1 のときに比べて、急峻なピークを示し、/a/, /i/ のときと逆になっている。

図 11 は日本語母音/e/を推定したものである。1kHz と 3kHz 付近に別なピークが見られる。また、 Z_{lung} をもつ場合の推定スペクトルは第 1 フォルマントのピークが急峻になっている。

図 12 は日本語母音/o/の推定スペクトルである。/i/, /u/, /e/とは違い、別なピークは発生していないが第 2、第 3 フォルマントで Z_{lung} をつけた合成系の推定スペクトルは C_1 のみの合成系の推定スペクトルに比べて急峻なピークを持つ。

4 まとめ

5 母音ともに共通していえることは 8kHz 付近で C_1, Z_{lung} に関わらず、推定スペクトルが下る。また、 Z_{lung} を持つ合成器での推定は擬似フォルマントが発生しやすくなる。

今回の実験では肺の分岐を考慮した声帯音源部を声道アナログモデルに組み込んだモデルで推定実験を行った。 C_1 のみの声帯音源部を持つ場合の推定スペクトルに比べて、 Z_{lung} をつけた声帯音源部での推定スペクトルの傾向が明らかになった。しかし、推定スペクトルにおいて対称な分岐と非対称な分岐構造との明確な違いは見られず、肺の構造を近似したモデルを声道アナログ

モデルに組み込む際には、どちらの構造を選んでも変わらないと思える。

我々はこれまで種々の条件下での推定実験を行ってきた。それらの結果から言えることは、比較的推定しやすいとされる母音でも、/u/のように概して推定精度が悪くなるものもある。後舌母音の/u/は口唇からの放射面積が他の母音に比べて小さいためと思われる。

今回のように肺を考慮した合成系においても推定スペクトルに与える影響は別なピークが発生したり、ピークが急峻になるなど母音ごとでさまざまな結果となった。推定精度を上げるためにはこのような母音ごとの特徴を考慮した合成系の検討や推定法が必要なのではないかと思われる。

参考文献

- [1] K.Ishizaka and J.L.Flanagan: "Synthesis of Voiced Sounds From a Two-Mass Model of the Vocal Cords", American Telephone and Telegraph Company The Bell System Technical Journal, Vol.51, No.6, July-August, 1972
- [2] J.J.Fredberg and A.Hoenig: "Mechanical Response of the Lungs at High Frequencies", Journal of Biomechanical Engineering, 1978, pp57
- [3] Herbert Hudde and Harald Slatky: "The acoustical input impedance of excised human lungs - Measurements and model matching", Journal of Acoustical Society of America, 1989, pp475
- [4] Morse and Ingard: Theoretical Acoustics - International Series in Pure and Applied Physics, chapter7, McGraw-Hill, 1968
- [5] 神山直久 三木 信弘 永井 信夫: "粘弾性体モデルを用いた声道壁インピーダンスについての考察" 信学論 (A), J75-A, No.11, pp.1600, (1992 - 11)
- [6] 佐賀 聡人 三木 信弘 永井 信夫: "連続音声合成のための動的声道シミュレータの合成法", 信学論 (A)、J68-A, No.3, pp332, (1985 - 3)
- [7] 三木 信弘: "音声生成系の時変回路モデルの研究", 平成元年度科学研究費補助金 (一般研究 B) 研究成果報告書 (1992)
- [8] Sabine Van Huffel and Joos Vanderwalle: "The Total Least Squares Problem Computational Aspects and analysis", Society for Industrial and Applied Mathematics、pp28-37 (1991)
- [9] 佐々木 幸司 三木 信弘 小川 吉彦: "信学技報、SP95、Vol.95, No.468"