

## 声道損失のデジタルモデル化と評価 Digital Modeling and Evaluation of Vocal Tract Losses

林 理絵† 樺澤 康夫‡  
Rie Hayashi Yasuo Kabasawa

### 1. まえがき

より自然な、人とコンピュータとの対話のために合成音声の感情表現など高度な音声合成を目指して、従来以上に柔軟な音声合成モデルが必要になってきている。録音された実音声素片の連結による方法が高品質音声合成技術として実用化されているが、柔軟性を増すためには音声素片のデータベースの規模を非常に大きくする必要があり、さらに高速の探索技術が必要となる。これに対し、音声生成の音響理論に基づく音声合成は柔軟性には富むが合成音声の品質に問題がある<sup>1)</sup>。計算の高速化のために、過度の簡単化が行われているためであると考えられる。このためここでは、後者の合成方式について、音声品質向上のために声道系のデジタルモデルの再検討を行う。

声帯・声道型動的音声合成モデル<sup>2)</sup>は、声帯の自励振動機構と声帯と声道の間の相互作用や声道の各種損失(空気の粘性損失、熱損失・声道壁振動損失、口や鼻孔からの音波の放射による損失など)を考慮に入れ、実際の音声生成の物理に則した計算機シミュレーションにより高品質の音声を合成できることが知られている。このようなモデルの精密化は計算量を大幅に増加させるため、計算速度の改善が残された問題の1つである。声道の各種損失を考慮した声道モデルとして、T型混合モデル<sup>3)4)</sup>が提案されている。声道を短い音響管の縦続接続で表し、1つの音響管を損失部分の集中定数の二端子対回路網と無損失の分布定数線路モデルで表すものである。本稿では、T型混合モデルと同程度の精度を保ち、合成の計算量の低減を図るために、新しい混合モデル(I型混合モデル)を提案し、このモデルの周波数領域における近似精度の評価とデジタルモデルの構成について検討する。

### 2. 声道の混合モデル

#### 2.1 声道I型混合モデル

声道壁振動以外の損失、すなわち粘性損失、拡散損失、熱損失を分布定数線路部に組み込むことを考える。まず伝送線路の低損失条件： $R/\omega L \ll 1$ ,  $G/\omega C \ll 1$ を満たすことを示す。音響管の断面積を  $A$ 、評価周波数を  $f$  とすると、次式が成立する。

$$\frac{R}{\omega L} = \frac{4\mu}{\rho} \frac{1}{Af} + \sqrt{\frac{\mu}{\rho}} \frac{1}{\sqrt{Af}}$$

$$\frac{G}{\omega C} = (\eta - 1) \sqrt{\frac{\lambda}{c_p \rho}} \frac{1}{\sqrt{Af}}$$

ここで、 $\mu$ は空気の粘性、 $\rho$ は空気の密度、 $\lambda$ は空気の熱伝導度、 $\eta$ は断熱比熱比、 $c_p$ は定圧比熱を表す。上式は、どちらも  $Af$  の単調減少関数である。断面積  $0.25\text{cm}^2$ 、周波数  $100\text{Hz}$  において評価すると、 $R/\omega L = 0.107$  であり、 $G/\omega C = 0.036$  である。これより通常の断面積に対して音声帯域全域で低損失条件は満たされていることが分かる。このとき、伝搬定数  $\gamma$ 、特性インピーダンス  $Z_0$  は次のように近似される。ただし、 $c_0$  は音速を表す。

$$\gamma = \sqrt{(R + j\omega L)(G + j\omega C)} = \alpha + j\beta$$

$$\alpha \approx \frac{1}{2} \left( \frac{R}{Z_0} + GZ_0 \right)$$

$$= \frac{\pi}{\rho c_0} \left[ \frac{4\mu}{A} + \left\{ \sqrt{f_0 \mu \rho} + (\eta - 1) \sqrt{f_0 \lambda \rho / c_p} \right\} \frac{1}{\sqrt{A}} \right]$$

$$\beta \approx \frac{\omega}{c_0}$$

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \approx \sqrt{\frac{L}{C}} \approx \frac{\rho c_0}{A}$$

従って低損失モデルでは、音響管は近似的に減衰を伴う遅延を与えることになる。このように損失は線路を通過する信号の単なる減衰として扱えば、声道区間の接続部における回路網から声道壁アドミタンス以外の損失項を表す音響抵抗を取り除くことができる。

次に声道壁振動について検討しよう。声道壁振動の共振周波数は  $20\text{Hz}$  程度と低く  $1\text{kHz}$  以上では十分に減衰する。また本来声道壁アドミタンスは、音響管内に分布しているが、この周波数範囲では音波の波長は十分に長いので、集中定数化できる。そこで、声道壁アドミタンスのみ集中定数回路網で表し、これを減衰のある線路と接続するモデルを新しい混合モデルとして採用することができる。接続部の回路形状よりI型混合モデルと呼ぶ。これを Fig.1 に示す。

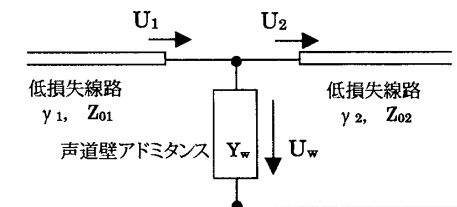


Fig.1 I型混合モデルの接続部

#### 2.2 声道I型混合モデルの周波数領域評価

分布定数モデルをターゲットとし、このモデルの周波数領

† 日本工業大学大学院工学研究科情報工学専攻

‡ 日本工業大学工学部情報工学科

域特性を伝達関数により求め、誤差を評価した。中性母音に対する結果を Fig.2 に示す。ただし粘性損失および熱損失は、本来、周波数に依存する項を含むが I 型モデルではデジタル実現化を考慮して、1 kHz における値の一定値とした。

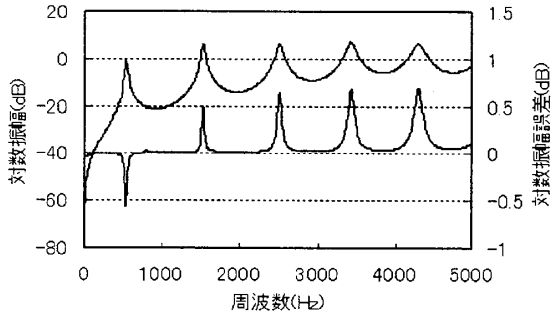


Fig.2 声道周波数特性と誤差 (連続時間モデル)

Fig.2から対数スペクトルの誤差の最大値が0.8dB以下となっていることが分かる。そこで、I 型混合モデルは十分なる精度を持っているといえる。

2.3 I 型混合モデルのデジタル実現化

このモデルのデジタル実現を検討する。声道壁振動の周波数成分は、ほぼ 1 kHz 以下に限定され、またサンプリング周波数 (16kHz) に比べて、1kHz は十分に低いので、現在の値と 1 時刻前の声道壁振動に関する値の差は十分に小さいと考えられる。このことを考慮に入れて、声道壁体積速度  $U_w$  は 1 時刻前の値を使うことにする。従って、声道区間接続部の体積速度流の進行波の関係は  $U_w$  は既知として扱うことができ、次のように与えられる。これを信号線図で表すと Fig.3 のようになる。

$$U'_0 = U_1^+ - 0.5 U_w - 0.5 U_n \quad \text{とおき、}$$

$$U_0 = U'_0 + U_2^- \quad \text{とおくと、}$$

$$U_1^- = -K U_0 + U_2^- - 0.5 U_w + 0.5 U_n$$

$$U_2^+ = K U_0 + U'_0$$

ここで、 $U_w$  は声道壁体積速度流、

$U_n$  は乱流雑音、 $K$  は反射係数を表す。

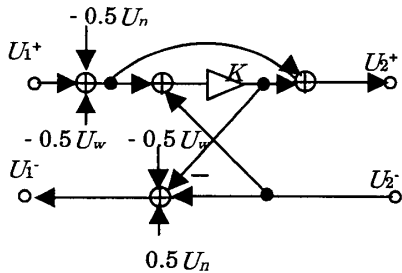


Fig.3 声道区間接続部の信号線図

$U_w$  は 1 時刻前において次の関係より計算される。

$$U_w = Y_w P_w = Y_w R_{01} (U_1^+ + U_1^-)$$

$Y_w$  のデジタルフィルタを構成するには、連続時間における声道壁アドミタンスを双 1 次変換により離散時間の伝達関数に変換する。これにより次の形式のフィルタが得られる。

$$H_w(z) = \frac{a_0 (1 - z^{-2})}{1 + b_1 z^{-1} + b_2 z^{-2}}$$

この構成では分母多項式の係数は、音響管の断面積に寄らず一定となる。従来の構成では、接続部の両側の断面積に依存して変化し一定値ではなかった。1 サンプルの計算ごとに毎回すべての区間においてこれらの係数の計算を必要としていたが、この計算を完全に除去できるため、大幅な計算量の削減ができる。

2.4 I 型混合デジタルモデルの評価

このモデルの計算精度を評価するために、声道系のインパルス応答を計算し、フーリエ変換して、周波数応答を求め、従来の T 型モデルと比較した。中性母音についての結果を Fig.4 に示す。他の母音については Table 1 に数値を示す。誤差はいずれも  $\pm 1.2$  dB 程度以下に収まっている。

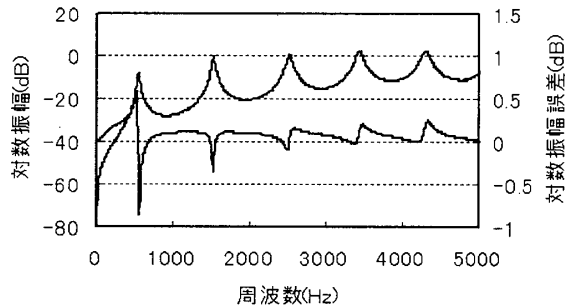


Fig.4 声道周波数特性と誤差 (離散連続時間モデル)

Table 1 典型的母音に対するスペクトル誤差

母音名	RMS 誤差 (dB)	+最大誤差 (dB)	-最大誤差 (dB)	最大絶対誤差 (dB)
中性	0.14	0.80	-0.80	0.80
/a/	0.20	1.05	-1.20	1.20
/e/	0.11	0.48	-0.55	0.55
/i/	0.15	0.52	-0.86	0.86
/o/	0.21	0.96	-0.80	0.96
/u/	0.21	1.10	-1.00	1.10

3. むすび

以上、声道の新しいモデルを提案しその評価とデジタル実現について示した。声帯モデルと組合せて新しい声帯声道型音声合成プログラムを作成することができる。今後、更なる計算量削減のため声帯モデルの検討を行う予定である。また、このモデルのための調音パラメータ推定の問題についても検討したい。

参考文献

[1] 匂坂芳典、ニック キャンベル: "音声合成のための規則とデータの表現、獲得、評価", 信学論(D-II), vol. J83-D-II, no. Nov. 2000.  
 [2] J.L.Flanagan, K.Ishizaka and K.L.Shipley: "Signal models for low bit-rate coding of speech", J. Acoust. Am. 63(3), 1980.  
 [3] 樺澤康夫ほか: "損失を考慮した声道のデジタル・モデルの簡略化", 日本音響学会音声研究会資料, 1982.  
 [4] Y.Kabasawa, et al, "Simplified Digital Model of the Lossy Vocal Tract and Vocal Cords", 11thICA, 1983.