



# HOKKAIDO UNIVERSITY

Title	発声時における声道伝達特性の推定
Author(s)	大場, 良次; Ohba, Ryoji; 土肥, 英明 他
Citation	北海道大學工學部研究報告, 74, 41-51
Issue Date	1975-03-05
Doc URL	<a href="https://hdl.handle.net/2115/41256">https://hdl.handle.net/2115/41256</a>
Type	departmental bulletin paper
File Information	74_41-52.pdf



## 発声時における声道伝達特性の推定

大場良次\* 土肥英明\*† 井戸川 徹\*

(昭和49年7月25日受理)

### Transfer Characteristics Estimation for Vocal Tracts under Utterance

Ryoji OHBA Hideaki DOHI Tohoru IDOGAWA

(Received July 25, 1974)

#### Abstract

A new method to estimate transfer characteristics of vocal tracts in utterance is presented. The method is based on a simple model of the voicing process and a new fact that the known vocal cord waveform may be approximated with a mean waveform of those of five vowel sounds from the same subject. The accuracy of estimation is also discussed in detail. It is a special feature of the present method that very short observations of a voice waveform, namely only a single pitch period, is required. This is applied to determine formant-patterns of the Japanese vowels and it was shown to work as well as conventional methods. Using several features of the estimated transfer characteristics, a voice recognition test was carried out on 500 samples of five Japanese vowels from 100 male adults and the recognition rate of 94.4% is obtained by the test.

#### 1. はじめに

いわゆる音声タイプライタをはじめとする音声の機械による自動認識の問題は、人間—機械系における情報媒体の一つとして音声を用いる目的で、かなり古くから研究されている<sup>1)</sup>。しかし、電算機への音声による入力あるいは電話の自動ダイヤリングなど、種々の応用が挙げられながら、いずれもいまだ実用の域には至っていない。

従来の研究では、ほとんど例外なく、ピッチ周期に比して比較的長時間に亘る音声波形標本をスペクトル解析して求められる、いわゆるフォルマント構造(フォルマントパターン)に基づき音声の識別が行なわれている。しかしこのようにして求められるフォルマントパターンは、音声波形標本の個体差の影響を受け変動が激しく、時に、当然現われるべき周波数帯域にフォルマントを見出し得ないこともある。これはフォルマントが、個体差の大きな声帯波と本来時間的に変化する発声時における声道伝達特性との双方の影響を受けることに起因する。音声識別に利用する特徴としては、個体差の比較的少ない声道伝達特性がフォルマントに勝ると考えられている。従って、何らかの方法により声帯波の影響を受けずに発声時の声道伝達特性が測定し得れば、フォルマントパターンに基づく従来の音声識別法の短所が補える可能性がある。

\* 応用物理学科

† 現在富士通(株)勤務

ところで、発声に与る声帯や声道は、人間の“生体”の中に位置している。そのため声帯ないし声道の、単体としての特性を測定するのはかなりむづかしい。たとえば声道の伝達特性を実験的に求めた例として次のようなものが報告されている。

1) 喉頭部に穿った小孔を通じ、外部から音響振動を与えて声道を励振し、口唇部で検出される出力音響振動と励振振動との関係から伝達特性を求めた例<sup>2)</sup>。

2) 喉頭部外壁を振動子で加振して間接的に声道を励振し、口唇部で検出される出力音響振動と振動子加振信号との関係から伝達特性を求めた例<sup>3)</sup>。

3) 発声時声道の形状を X 線写真で連続撮影し、それをもとに伝達特性を推定した例<sup>1),4)</sup>。

1)および2)は、単に発声時と同様な緊張状態に保たれた声道の、いわば静的な特性を求めるに過ぎず、発声時における声道のいわば動的な特性を求められない。しかも1)は外科手術を要する特殊な方法であり、2)では喉頭部外壁の影響も考慮しなければならない。3)は、前2例と異なり、発声時における声道の動的な特性を推定しうる。しかしこの方法には推定の精度、実験の困難な点など問題が少くない。

一方声帯の特性を求める実験としては、口唇部から挿入したファイバースコープなどにより、声帯の振動状況を連続撮影し、声門開口面積の時間変化から声帯の振動波形を推定したり<sup>5)</sup>、声道伝達特性のちょうど逆特性となるように試行錯誤的に構成・調節されるフィルタを用いて声帯波形を求める<sup>6)</sup>などの実験が行なわれている。しかし、どの実験も簡単には実行し得ない点が欠点である。

本論文では、発声時声道の伝達特性を近似的に求めるきわめて簡単な方法について述べる。2章では、実験事実と単純な発声過程模型に基づき、発声時における声道伝達特性を近似的に推定する一簡便法を提案し、その際の推定誤差の程度について検討する。3章では、2章で提案される方法を用いて、実際に声道伝達特性を測定した結果と、その2・3の応用例を述べる。

## 2. 近似的声道伝達特性の推定

### 2.1 発声過程模型

有声音は、声帯波が声道なる可変音響管を励振して発生させると考えられる。ここで声帯波は肺を定圧源とする呼气流が声門部で声帯により断続されて生ずる大略周期的な圧力波である。従来の研究によれば、声帯波による声道励振の過程は、少くとも声道の形状が時間的に不変に保たれる間は、線型現象と考えて良いとされている<sup>1)</sup>。また Miller<sup>6)</sup>を始めとする多くの研究者によって、声帯波形は図1のような略周期的繰返し間欠非対称三角波であることが明らかにされている。しかもこの非対称三角形部分の形は、ピッチ周期の多少の変動や発声音韻の種類に依らず同一話者について略一定で、ただ間欠部のみがピッチ周期の変動に伴い伸縮するとされている。

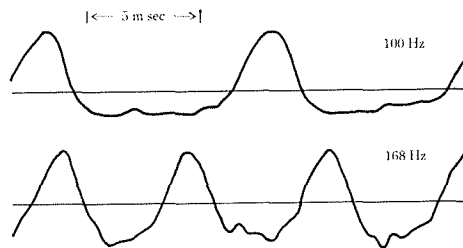


図1 声帯波形 (Miller<sup>6)</sup>による)

このような知見に基づき、発声過程について次の仮定を行ない、模型を構成する。

1) 同一話者の声帯波  $v(t)$  は音韻の種類と関係なく (1) 式で表わせるものとする。

$$v(t) = \int_{-\infty}^t v_0(\tau) \times \left\{ \sum_{n=-\infty}^{\infty} \delta(t - \tau - nT) \right\} d\tau \quad (1)$$

ただし、 $v_0(\tau)$  は非対称三角波、 $T$  はピッチ周期である。

2) 声道の励振過程は線型とし、音声波形  $x_l(t)$  が (2) 式で表わせるものとする。

$$x_l(t) = \int_{-\infty}^{\infty} h_l(\tau) \cdot v(t-\tau) d\tau \quad (2)$$

ただし、 $h_l(t)$  は音韻  $l \in \Omega \equiv \{/a/, /i/, /u/, /e/, /o/\}$  を発声する際の声道のインパルス応答、 $v(t)$  は声帯波である。

声道は鼻腔という枝管のついた両端開口の可変音響管で、発声する音韻に応じて連続的に管断面積を変え、特性を変化させる。このため会話音声など通常の連続音声には、いわゆる“ワタリ”などの特性変更時の過渡状態が存在する。この過渡状態は、音声波の解析に際して大きな障碍となるので、以下では対象を一音一音区切って発声されるいわゆる単一音節の母音に限り、連続音節や子音は扱わないものとする。このとき次に述べる方法によって、発音時における声道の伝達特性が近似的に推定できる。

## 2.2 推定原理

(2) 式をフーリエ変換し、(1) 式の関係を用いれば音声波  $l$  に対して次の表式が得られる。

$$\begin{aligned} X_l(j\omega) &= V(j\omega) \cdot H_l(j\omega) \\ &= \int_{-\infty}^{\infty} V_0(j\omega) \cdot H_l(j\omega) \times \left\{ \sum_{n=-\infty}^{\infty} \delta(j\omega - jn\Delta\omega) \right\} dj\omega \end{aligned} \quad (3)$$

従って、

$$\begin{aligned} X_l(jn\Delta\omega) &= V_0(jn\Delta\omega) \cdot H_l(jn\Delta\omega) \\ l \in \Omega, \quad n &= \dots, -2, -1, 0, 1, 2, \dots \end{aligned} \quad (4)$$

ただし、 $\Delta\omega = 2\pi/T$ 。

(4) 式から、 $V_0(jn\Delta\omega) \neq 0$  の時、

$$H_l(jn\Delta\omega) = X_l(jn\Delta\omega) / V_0(jn\Delta\omega) \quad (5)$$

ここで  $X_l(jn\Delta\omega)$ 、 $V_0(jn\Delta\omega)$  は、おのおの音声波  $x_l(t)$ 、声帯波  $v_0(t)$  のフーリエ係数。 $X_l(jn\Delta\omega)$  は測定可能であるから、 $V_0(jn\Delta\omega)$  が知れば  $H_l(jn\Delta\omega)$ 、つまり音韻  $l$  を発声する際の声道伝達特性が推定できる。

一方著者等の観察によると、同一話者により略等しいピッチで発声された日本語の5種の母音を同期的に加え平均した波形は、既に知られている声帯波形とかなりの類似が認められる<sup>7)</sup>。図2はそのような平均波形の一例である。これは、声帯波のピークに対応して現われる音声波形のピークを目安にピッチ周期を定め、音声波から1ピッチ周期分を選び標本化し、周期および分散を規格化後(6)式により同一話者について求めた平均波形である。

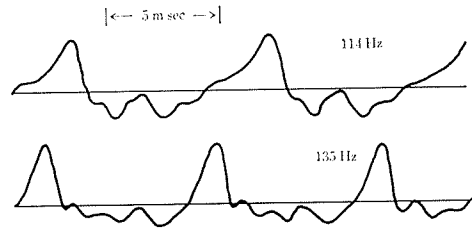


図2 平均波形,  $\bar{x}(t)$

$$\bar{x}(t) = \frac{1}{5} \sum_{l \in \Omega} x_l(t), \quad 0 \leq t \leq T. \quad (6)$$

そこでこの事実に基づき、声帯波  $v_0(t)$  を (6) 式の平均波形  $\bar{x}(t)$  で代用することによれば、(5) 式から近似的に声道伝達特性が求められる。

## 2.3 推定誤差

上述の方法を要約すれば、1ピッチ周期間の声帯波  $v_0(t)$ 、 $0 \leq t \leq T$  を (6) 式の平均波形  $\bar{x}(t)$

で近似し、(5)式により声道伝達特性を推定することである。つまり同一ピッチで発声された各母音の一周期分の波形を、声帯波に同期して加え平均波形を求め、それを近似的に声帯波形と見做す訳である。しかし十分に注意して実験しても、実際に得られる波形標本が雑音により汚損されたりあるいは各標本間でピッチや位相の微小な差異が生じたりするのは避け得ない。一般にそのような差異や雑音の存在は、推定の結果に誤差を生ずる。従って、この方法を実地に用いる前に推定に伴う誤差の程度を評価しておく必要がある。

推定に伴う誤差の主たる要因は次の2つである。その1は、連続量を離散値化する際に必ず混入する、測定雑音としての量子化誤差であり、その2は、波形標本相互間のピッチや位相の微小な差異に起因するこの方法に特有の誤差である。

まず第一の量子化誤差について検討する。最小量子化幅  $q$  が量子化定理を満たすよう選ばれるならば、量子化誤差は振幅分布が図3で表わされるような信号とは独立な雑音と考えて良いことが知られている<sup>8)</sup>。すなわちこの誤差は、信号に、それとは独立で振幅が区間  $[-q/2, q/2]$  の間に一様分布する雑音を加えることと等価である。従って音声波  $x_l$  の量子化された波形標本を  $x_{li}$ ,  $i=1\sim N$  とすれば、分散を1に規格化後の音声波における量子化誤差のパワースペクトル  $P_{li}(\omega)$  は次のようになる。

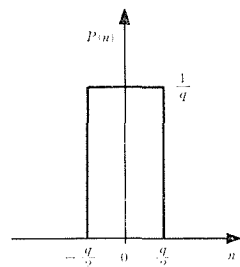


図3 量子化誤差の分布

$$P_{li}(n\Delta\omega) \simeq \frac{q^2}{6 \left\{ \sum_{i=1}^N (x_{li} - m_l)^2 \right\}} \quad (7)$$

$$n=1, 2, \dots, N/2, \quad l \in \Omega.$$

ただし  $\Delta\omega = 2\pi/T$ ,  $m_l = \sum_{i=1}^N x_{li}/N$ .

また平均波形  $\bar{x}_l = \sum_{i \in \Omega} x_{li}/5$ ,  $i=1\sim N$  における量子化誤差の影響は、互に独立な5個の誤差が平均されることに注意すれば、次式のように求まる。

$$P_{\bar{x}_l}(n\Delta\omega) \simeq \frac{1}{25} \cdot \frac{q^2}{6} \left\{ \sum_{l \in \Omega} \frac{1}{\sum_{i=1}^N (x_{li} - m_l)^2} \right\}$$

$$= \frac{1}{5} \left\{ \frac{1}{5} \sum_{l \in \Omega} P_{li}(n\Delta\omega) \right\}$$

$$\simeq \frac{1}{5} P_{li}(n\Delta\omega) \quad (8)$$

$$n=1, 2, \dots, N/2, \quad l \in \Omega$$

次に第2の誤差について考える。既に述べたように各音声波形標本間でピッチ、位相が少しずつ異なるので、現実には(6)式の  $\bar{x}(t)$  を(9)式の  $\hat{x}(t)$  で代用させている。

$$\hat{x}(t) = \frac{1}{5} \sum_{l \in \Omega} x_{ls} \left( (1 + \Delta_l) t + \theta_l T \right) \quad (9)$$

ここで  $x_{ls}(t)$  は周期、位相についておのおの  $\Delta_l T$ ,  $\theta_l T$  だけ偏倚する音声  $l$  の標本である。(9)式は  $\hat{x}(t)$  が各標本の時間軸を、

$$t \rightarrow (1 + \Delta_l) t + \theta_l T, \quad l \in \Omega \quad (10)$$

のように正規化後平均した波形であることを示している。従って、 $\bar{x}(t)$  の代りに  $\hat{x}(t)$  を用いる時の残差  $R(t)$ ,

$$R(t) = \bar{x}(t) - \hat{x}(t) \quad (11)$$

がこの場合の誤差となる。

2.1 節の仮定から、 $x_{ls}(t)$  は周期に関して (1) 式の声帯波  $v(t)$  と  $\mathcal{J}_l T$  だけ偏倚するような、周期  $T_l$  の声帯波  $v_l(t)$  により励振される声道の共鳴出力と考えられる。すなわち、

$$x_{ls}(t) = \int_0^\infty h_l(\tau) v_l(t-\tau) d\tau. \quad (12)$$

ただし、

$$v_l(t) = \int_0^\infty v_0(\tau) \cdot \left\{ \sum_{n=-\infty}^{\infty} \delta(t-\tau-nT_l) \right\} d\tau, \quad T_l = \frac{T}{1+\mathcal{J}_l}. \quad (13)$$

(12), (13) 式から、

$$\begin{aligned} x_{ls}(t) &= \int_0^\infty h_l(\tau_1) d\tau_1 \int_0^\infty v_0(\tau_2) \left\{ \sum_{n=-\infty}^{\infty} \delta(t-\tau_1-\tau_2-nT_l) \right\} d\tau_2 \\ &= \sum_{n=-\infty}^{\infty} \int_0^\infty h_l(\tau_1) d\tau_1 \int_0^\infty v_0(\tau_2) \delta(t-\tau_1-\tau_2-nT_l) d\tau_2 \\ &= \sum_{n=-\infty}^{\infty} \int_0^\infty h_l(\tau_1) \cdot v_0(t-\tau_1-nT_l) d\tau_1. \end{aligned} \quad (14)$$

従って時間軸を正規化後の標本波形は、

$$\begin{aligned} x_{ls}\left((1+\mathcal{J}_l)t + \theta_l T\right) &= \sum_{n=-\infty}^{\infty} \int_0^\infty h_l\left((1+\mathcal{J}_l)\tau_1\right) \cdot v_0\left((1+\mathcal{J}_l)t - (1+\mathcal{J}_l)\tau_1 - n(1+\mathcal{J}_l)T_l + \theta_l T\right) d\tau_1 \\ &= \int_0^\infty \int_0^\infty \frac{1}{1+\mathcal{J}_l} h_l(\tau) v_0\left((1+\mathcal{J}_l)t + \theta_l T - \tau - \sigma\right) \left\{ \sum_{n=-\infty}^{\infty} \delta(\sigma - nT) \right\} d\sigma d\tau. \end{aligned} \quad (15)$$

ここで  $v_0$  を  $t-\tau-\sigma$  の周りで展開すれば、

$$v_0\left((1+\mathcal{J}_l)t + \theta_l T - \tau - \sigma\right) = \sum_{m=0}^{\infty} \frac{1}{m!} v_0^{(m)}(t-\tau-\sigma) \cdot (\mathcal{J}_l t + \theta_l T)^m, \quad (16)$$

ただし、

$$v_0^{(m)} = \frac{d^m v_0}{dt^m}.$$

従って (6), (9), (11), (15), (16) 式から、

$$\begin{aligned} R(t) &= \frac{1}{5} \sum_{l \in \Omega} \int_0^\infty \left[ \sum_{m=1}^{\infty} \frac{(\mathcal{J}_l t + \theta_l T)^m}{1+\mathcal{J}_l} \cdot \frac{1}{m!} \int_0^\infty v_0^{(m)}(t-\tau-\sigma) \cdot h_l(\tau) d\tau \right. \\ &\quad \left. - \frac{\mathcal{J}_l}{1+\mathcal{J}_l} \int_0^\infty v_0(t-\tau-\sigma) \cdot h_l(\tau) d\tau \right] \left\{ \sum_{n=-\infty}^{\infty} \delta(\sigma - nT) \right\} d\sigma \\ &= -\frac{1}{5} \sum_{l \in \Omega} \left[ \frac{\mathcal{J}_l}{1+\mathcal{J}_l} x_l(t) - \sum_{m=1}^{\infty} \frac{(\mathcal{J}_l t + \theta_l T)^m}{1+\mathcal{J}_l} \cdot \frac{1}{m!} \frac{d^m}{dt^m} x_l(t) \right] \end{aligned} \quad (17)$$

ただし、 $x_l(t)$  はピッチ周期、位相とも偏倚のない理想的音声波形。

$\mathcal{J}_l, \theta_l$  が共に小さな時には、(17) 式はそれ等に関する 2 次以上の項を省略して次のように近似できる。

$$R(t) \simeq -\frac{1}{5} \sum_{l \in \Omega} \left[ \mathcal{J}_l x_l(t) - (\mathcal{J}_l t + \theta_l T) \frac{d}{dt} x_l(t) \right]$$

この式から残差のパワースペクトルが、

$$|R(j\omega)|^2 \simeq \frac{1}{25} \left| \sum_{l \in \Omega} \left\{ 2\mathcal{J}_l X_l(j\omega) + \mathcal{J}_l \left[ j\omega X_l'(j\omega) \right] + \theta_l T \left[ j\omega X_l(j\omega) \right] \right\} \right|^2 \quad (18)$$

と求められる。ただし  $X'_l(j\omega) = dX_l(j\omega)/d\omega$ 。

さて、 $J_l, \theta_l$  は統計量であるから  $|R(j\omega)|^2$  も統計量である。いま  $\{J_l\}, \{\theta_l\}, l \in \Omega$  のおのが、互に独立でかつ平均値 0、分散  $J^2, \theta^2$  なる同一の確率密度函数に従うとすれば、残差のパワースペクトルの期待値は次のように求められる。

$$\begin{aligned} E\left[|R(j\omega)|^2\right] &\leq \frac{1}{25} E\left[\sum_{l \in \Omega} J_l^2 \omega^2 |X_l(j\omega)|^2 + \sum_{l \in \Omega} (4J_l^2 + \theta_l^2 T^2 \omega^2) |X_l(j\omega)|^2\right] \\ &= \frac{1}{25} (4J^2 + \theta^2 T^2 \omega^2) \sum_{l \in \Omega} |X_l(j\omega)|^2 + \frac{J^2}{25} \sum_{l \in \Omega} |X'_l(j\omega)|^2 \end{aligned} \quad (19)$$

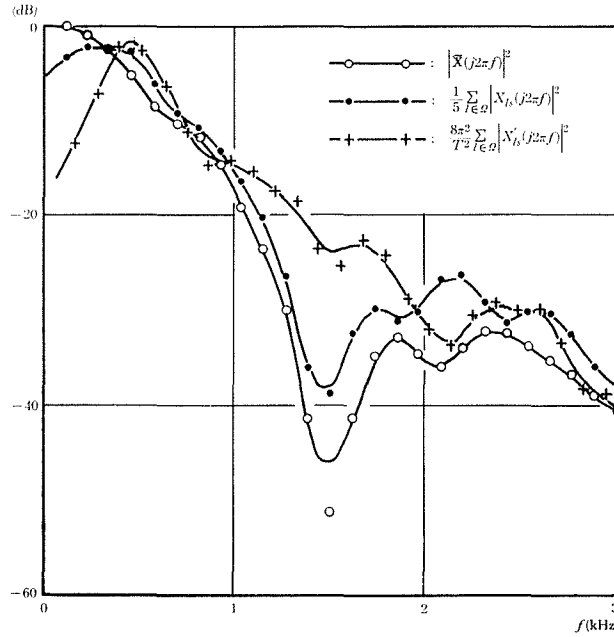


図4 実際のデータにおける  $\sum_{l \in \Omega} |X_{ls}(j\omega)|^2$ ,  $\sum_{l \in \Omega} |X'_s(j\omega)|^2$  と  $|\tilde{X}(j\omega)|^2$  の関係

ここで実際のデータから  $\sum_{l \in \Omega} |X_{ls}(j\omega)|^2$ ,  $\sum_{l \in \Omega} |X'_s(j\omega)|^2$  を求めてみると、図4に示すように、近似的に次の関係が成立する。

$$\left. \begin{aligned} \frac{1}{5} \sum_{l \in \Omega} |X_{ls}(j\omega)|^2 &\simeq |\tilde{X}(j\omega)|^2 \\ \frac{1}{5} \sum_{l \in \Omega} |X'_s(j\omega)|^2 &\simeq \frac{T^2}{40\pi^2} |\tilde{X}(j\omega)|^2 \end{aligned} \right\} \quad (20)$$

そこで  $X_l(j\omega)$  を  $X_{ls}(j\omega)$  で代用させることにして、(19)式に(20)式の関係代入すれば次の関係が得られる。

$$E\left[|R(j\omega)|^2\right] \leq \left[\frac{4}{5} J^2 + \frac{1}{5} \left(\theta^2 + \frac{J^2}{40\pi^2}\right) T^2 \omega^2\right] |\tilde{X}(j\omega)|^2 \quad (21)$$

結局(21)式から相対誤差の程度が、

$$\frac{E[|R(j\omega)|^2]}{|\tilde{X}(j\omega)|^2} \simeq \frac{4}{5} J^2 + \frac{1}{5} \left(\theta^2 + \frac{J^2}{40\pi^2}\right) T^2 \omega^2 \quad (22)$$

と求められる。

(22) 式は、残差のパワースペクトルの期待値が、平均波形  $\tilde{x}(t)$  を  $4d^2/5 + (\theta^2 + d^2/40\pi^2) T^2 \omega^2/5$  なるパワー伝達特性の線型回路 (一種の微分回路) に通じた時の出力のパワースペクトルと見做し得ることを示している。これは  $f_0 = d/\sqrt{\pi^2 \theta^2 + d^2/40T}$  以上の周波数域では、残差  $R$  に対する  $\theta$  の影響が、相対誤差で 20 dB/decade の率で増加することを示している。今仮に、ピッチ周期  $1/T = 100$  Hz で、周期および位相の偏差が略等しい ( $d \approx \theta$ ) とすれば、 $f_0$  は約 30 Hz となる。つまり誤差の大部分が位相偏倚に支配されることが分る。

### 3. 近似的声道伝達特性の測定と応用<sup>10)</sup>

2.2 節に述べた方法によって実際に近似的声道伝達特性を測定した。実験は、まず被験者に {/a/, /i/, /u/, /e/, /o/} の 5 種の日本語母音を一音ずつ区切って略等しいピッチで発声させ、データレコーダ (TEAC 社製, 410 型) に記録した。次いで、その記録を再生しつつ各母音波形中の定常的な部分を選び、1ピッチ周期分を切り出して 1/100 周期毎に標準化し離散値化した\*。こうして得られた音声波形標本を電算機でデータ処理して、最終的に伝達特性  $|H_l(j\omega)|$  を求めた。図 5 は被験者の当該母音発声時における声道の近似的伝達特性を求める際のデータ処理の流れ図である。

測定される伝達特性の誤差の程度は、2.3 節の結果を用いて推定される。実験に用いたデータでは  $N=100$ ,  $q=1$ ,  $\sum_{i=1}^N (x_{li} - m_l)^2 \approx 2 \times 10^4$  である。従って、(7), (8) 式から、音声波形標本および平均波形における量子化誤差の影響がおのおの次のように求められる。

$$\left. \begin{aligned} P_{ix}(n\Delta\omega) &\leq 8.33 \times 10^{-6} \sim -51 \text{ dB} \\ P_{i\tilde{x}}(n\Delta\omega) &\leq 1.67 \times 10^{-7} \sim -58 \text{ dB} \end{aligned} \right\} \quad (23)$$

この式から音声波形  $x_l(t)$  または平均波形  $\tilde{x}(t)$  を用いる際には、スペクトルの値がおのおの  $-51$  dB,  $-58$  dB 程度となる周波数帯域で注意が必要であることが知られる。

実地にパワースペクトルを求めた結果では、取扱う周波数帯域 0~4 kHz ではほとんどの場合音声波形・平均波形共  $-15 \sim -45$  dB の範囲にあった。従って量子化誤差に関する限り、少なくとも 6~13 dB 以上の信号対雑音比が保障される。

一方標本波形相互間の位相、周期の偏倚に由る誤差は、(22) 式で  $1/T \sim 100$  Hz,  $d = \theta \sim 0.02$  として求められる。結果は 100 Hz で約 0.05%, 1 kHz で約 0.8% の相対誤差となることが分る。我々が扱う周波数帯域では 4 kHz の時の誤差が最も大きく、それは相対誤差で約 16% 程度と見込まれる。

#### 3.1 フォルマントパタンの決定

図 6 は実際に推定された近似的伝達特性  $|H_l(j\omega)|$ ,  $l \in \Omega$  の例である。おのおの特有のパターンを呈しているのが見られる。伝達特性は、音響管としての声道の共振特性を表わすものであるか

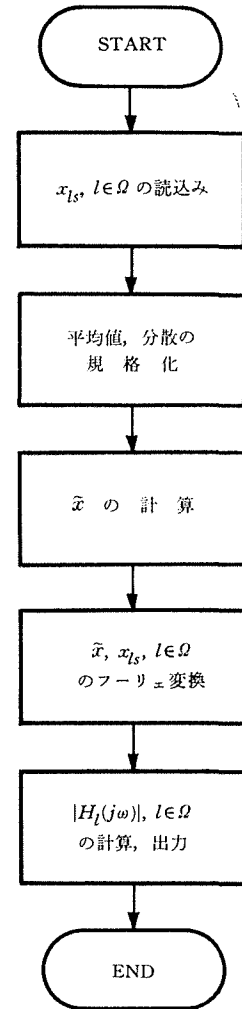


図 5 近似的声道伝達特性を求める際のデータ処理の流れ図

\* ピッチ周波数は通常 100~200 Hz であるから、この標本化による folding 周波数は 5 KHz 以上となる。これは音声の主要な帯域を十分掩蔽していると考えて良く、いわゆるエイリアジング効果は無視し得る。

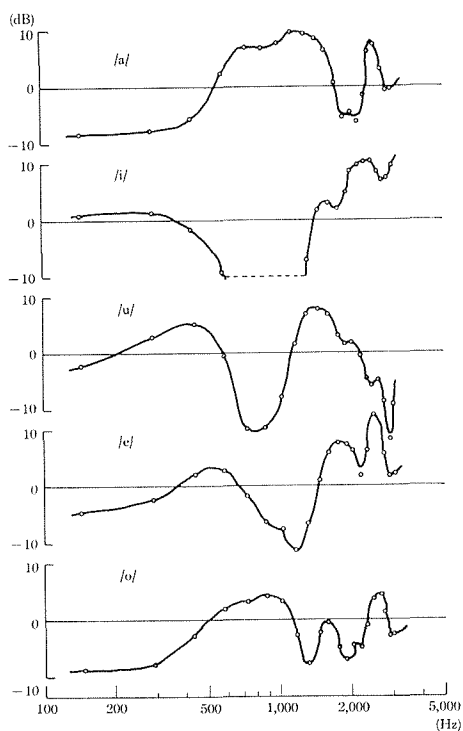


図6 近似的声道伝達特性,  $|H_t(j\omega)|$

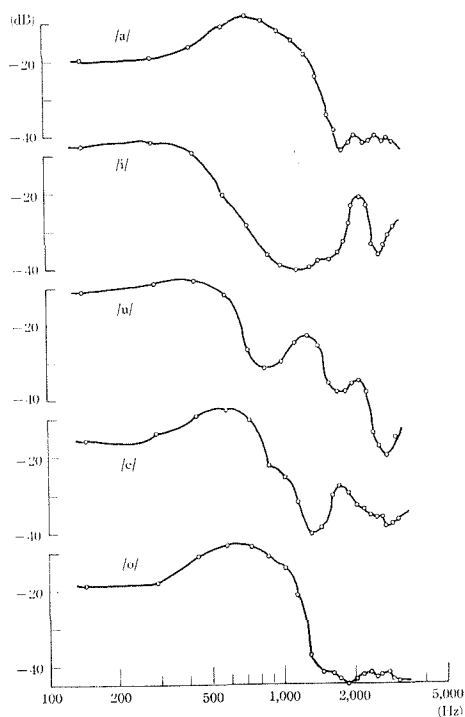


図7 図6と同じ音声波形標本から求めた音声波形のスペクトル,  $|X_t(j\omega)|$

ら, 図中に見られるピークはフォルマントに一致すると考えられる。そこでこの近似的伝達特性におけるピークの構造と, 従来の方法——すなわち, 音声波形  $x_t(t)$  のパワースペクトルのピーク構造から求める方法——により求めたフォルマントパターンとを比較して見た。

図7は, 図6を求める時に用いたのと同じの波形標本から求めた音声波形のスペクトル  $|X_t(j\omega)|$  である。この図ではフォルマントを示すピークが顕著に現われていないので, フォルマント周波数, 特に低周波域でのそれが判別し難い。これは, 音声波スペクトルに含まれる声帯波の影響が, 特に低周波域で大きなためである。しかるに一方図6の近似的伝達特性を見ると, 声帯波の影響が除かれているため, ピークが低周波域においても顕著に現われており容易に識別できる。表1, 2は2通りの方法で求めたフォルマント周波数を示したもので, 表1は近似的伝達特性  $|H_t(j\omega)|$  を用いる新しい方法による結果, 表2は従来の方法による結果である。表中の F1, F2 および F3 は, おのおの第1, 第2および第3フォルマントを示し, (—) の記号は対応する

表1 近似的声道伝達特性から求めたフォルマント周波数

母音	フォルマント		
	F1	F2	F3
/a/	700	1,150	2,300
/i/	350*	2,300	3,200
/u/	350	1,500	2,300
/e/	450	1,500	2,300
/o/	500	850	1,400

表2 表1と同じ波形標本のスペクトルから求めたフォルマント周波数

母音	フォルマント		
	F1	F2	F3
/a/	700	—	2,300
/i/	—	1,950	3,200
/u/	350*	1,300*	1,900
/e/	—	1,700	2,500
/o/	450	800	1,250

ピークの認められないもの、(\*)はピーク位置が判然としないものを示している。

いずれの方法によっても略した値が求まっているが、従来の方法では決定し得ない場合でも新しい方法により決定できたこと、しかも新しい方法では単に1ピッチ周期分の波形標本しか要しない点に注意したい。このようにして求めた近似的声道伝達特性のピーク構造、換言すれば極の配置が、従来のフォルマントパターンと良く一致することから、逆に、2.2節で提案した方法により求まる声道伝達特性が、真の伝達特性の良い近似となっていると言うことが出来よう。

### 3.2 母音の識別

図8は10名の被験者について近似的声道伝達特性を求め平均した結果を示している。平均操作のために図6の個々の特性に見られた細かな構造は消失しているが、各母音に特有の平均的構造が強調されて現われている。表3は各母音毎に特徴を調べた結果である。図8から更に次の事実が知られる。すなわち、各伝達特性では各母音に特有の比較的狭い帯域にパワー通過域が集中していることである。この集中の度合いを見るために、パワー伝達関数  $|H_i(j\omega)|^2$  の累積度を次式で定義し計算してみた。

$$IP(f) = \int_0^f |H_i(j2\pi f)|^2 df / \int_0^{3,000} |H_i(j2\pi f)|^2 df, \quad f \leq 3,000 \text{ Hz} \quad (24)$$

(24)式で定義される累積度  $IP(f)$  は、伝達特性が0~3 KHzの範囲に有する全パワーのうち、周波数0~ $f$  Hzまでのパワーの占める割合を示している。

図9は累積度の例で、(a)~(e)はおのこの母音 /a/, /i/, /u/, /e/, /o/ に対応する。累積度曲線の勾配が急峻な部分は、その帯域にパワー通過域が集中していることを示している。これらの図

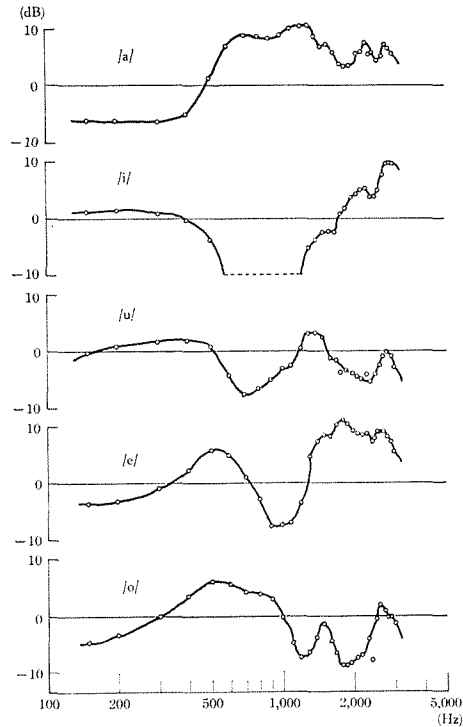


図8 10名の被験者から求めた近似的声道伝達特性の平均パターン

表3 図8の各パターンの特徴

$l$	$ H_l(j\omega) $ の特徴
/a/	700 Hz 付近に小ピーク、1,200 Hz 付近に大きなピークを有し、両ピークを含む帯域での利得が極めて大。
/i/i	600~1,200 Hz 帯域での利得が非常に小さいのに対して、500 Hz より低域でのそれは他の母音に比して大。高周波域での利得が極めて大。
/u/	600~1,000 Hz 帯域での利得が小さいのに対して、それより低域における利得が比較的大。400 Hz 付近に小ピークを有する。
/e/	500 Hz 付近に小ピーク、1,800 Hz 付近に極めて大きなピークを有する。高周波域での利得が大。
/o/	500 Hz 付近にピークを有する。

から、各母音にはパワー通過域の集中する特有の帯域が存在することが結論される。多くの累積度曲線を各母音毎に平均した結果、母音 /a/, /u/ および /o/ では 1.3 KHz 以下の帯域に過半のパワーが集中しているのに対し、/i/ および /e/ ではそれが約 2.6 KHz 以上の帯域になることが分った。

そこで上述の伝達特性およびパワー累積度の特徴に基づき、次の手順で5種類の母音の識別を試みた。図10は識別手順を示す流れ図である。すなわち、まず 200 Hz における伝達特性の値  $|H_i(j2\pi \cdot 200)|$  を閾値  $\varphi_1$  と比較し、その大小により未知音韻  $l$  を  $\{i/, /u/\}$  と  $\{a/, /e/, /o/\}$  とに分ける。それが  $\varphi_1$  より大きい場合には、次に、2,000 Hz における累積度  $IP(2,000)$  を求めて閾値  $\varphi_2$  と比らべ  $\{i/, /u/\}$  を分離する。一方  $|H_i(j2\pi \cdot 200)|$  が  $\varphi_1$  より小さい場合には、更に 1.1 KHz における伝達特性の値  $|H_i(j2\pi \cdot 1,100)|$  を閾値  $\varphi_3$  と比較し、その大小に応じて  $\{a/\}$  と  $\{e/, /o/\}$  とを分離する。最後に 1,000 Hz での累積度  $IP(1,000)$  を閾値  $\varphi_4$  と比較して、その大小により  $\{u/\}$  と  $\{o/\}$  とを分離する。

成年男子 84 名から集めた音声波形標本をもとに、図10の各分岐毎に誤り率が最小となる様に閾値  $\varphi_1 \sim \varphi_4$  を定めた。この閾値を用いて成年男子 100 名 (先の 84 名を含む) から得た 5 種類計 500 の母音の識別実験結果を表4に示す。この実験における識別率は 94.4% であった。

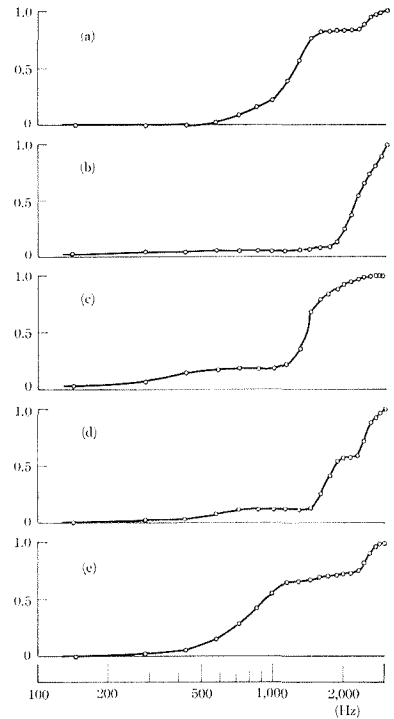


図9 累積度曲線。(a), (b), (c), (d)および(e)は、おのおの母音 /a/, /i/, /u/, /e/ および /o/ に対応する

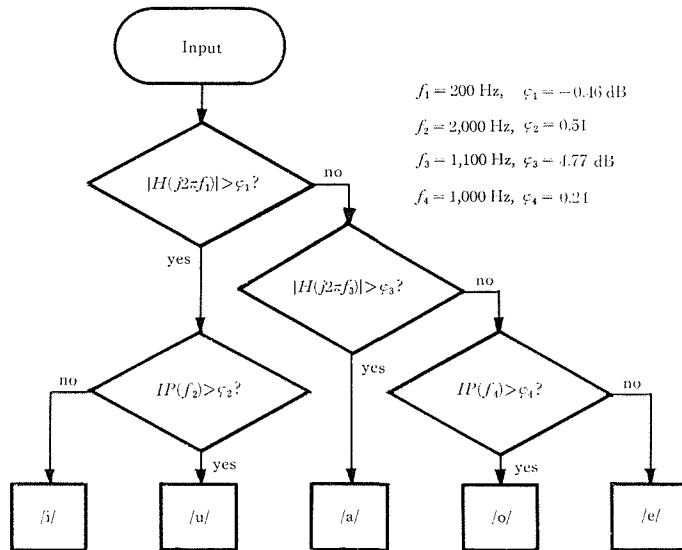


図10 日本語母音識別の流れ図

表4 識別結果

入 力	識 別 出 力				
	/a/	/i/	/u/	/e/	/o/
/a/	99	—	—	—	1
/i/	—	91	8	1	—
/u/	—	1	93	3	3
/e/	—	1	2	96	1
/o/	2	—	4	1	93

識別率 94.4%

## 4. む す び

本論文では発声時における声道伝達特性を推定する新しい方法を提案し、その推定誤差について検討した。この方法は従来の知見に基づく発声過程模型と、著者等が見出した単純な実験事実をもとに考案されたもので、極めて短時間の音声波形標本しか要しない点が特長である。同方法により声道伝達特性を実地に求め、その構造が従来のフォルマントパターンに良く一致することを確かめた。このことから、フォルマントパタンの決定という機能に関する限り、同方法が従来の方法に劣らないことが分かった。最後に同方法で推定される伝達特性の特徴の幾つかを用いて、5種類の日本語母音の識別実験を行なった。幾つかの特徴の存・否を逐次照合することにより母音の種類を識別した。100名の成年男子から得た500個の波形標本について実験を行ない、94.4%の識別率を得た。使用特徴を精選しより高度な識別手段を用いれば、識別成績をこれより向上させることは十分可能である。しかし、近似伝達特性に基づく識別法を実地に音声識別装置に適用することはできない。その最大の原因は、近似的伝達特性を求めるために、同一被験者について予め5種の母音波形標本を得ておく必要があることである。この点は、本論文で述べた方法の根元に触れる問題であり、今後の検討が必要である。

本研究におけるデータ処理には北海道大学大型計算機センターを利用した。

## 参 考 文 献

- 1) 電子通信学会(編): 聴覚と音声(昭和41年), コロナ社.
- 2) van den Berg, Jw.: Transmission of the vocal cavities, JASA, 27 (1955), pp. 161-168.
- 3) Fujimura, O.: Sweep-tone measurement of vocal tract, JASA, 49 (1971), pp. 541-558.
- 4) 大泉充郎(監修): 音声科学(昭和47年), 東大出版会.
- 5) Flanagan, J. L.: Some properties of the grottal sound source, Jour. of Speech Hear. Res. 1 (1956), pp. 99-116 あるいは Speech analysis, synthesis and perception (1965), Springer-Verlag.
- 6) Miller, R. L.: Nature of the vocal cord wave, JASA, 31 (1959), pp. 667-677.
- 7) 土肥英明, 大場良次, 井戸川 徹: 発声時に於ける声道の近似的伝達特性, 第4回計測制御に関する北海道研究集会論文集(昭和46年), pp. 43-44.
- 8) Widrow, B.: Statistical analysis of amplitude quantized sampled-data systems, AIEE. App. & Ind., 52 (1961), pp. 555-568.
- 9) 大場良次: 音声信号処理における周期・位相の微小偏倚の影響, 第12回計測自動制御学会学術講演会予稿集(昭和48年), pp. 307-308.
- 10) 大場良次, 井戸川 徹, 土肥英明: 近似的声道伝達特性とその応用, 第12回計測自動制御学会学術講演会予稿集(昭和48年), pp. 299-306.