

# 音声-体動同期現象のマン・マシン・インタフェースへの応用†

渡 辺 富 夫\*\*

音声は人間に適した、自然で容易な情報伝達手段であり、マン・マシン・インタフェースにおける入出力方式として注目を浴びている。音声による人間同士の情報交換（コミュニケーション）においては、音声と動作が同期し、この音声-体動同期現象が円滑な情報交換に重要な役割を果たしている。したがって、マン・マシン・インタフェースにこの音声-体動同期現象のメカニズムが導入されるならば、人間と機械の情報交換の円滑化が図られると考えられる。著者は、従来定性的研究の色彩が強いこの分野に、新たに画像自動分析の手法を導入し、コミュニケーションにおける音声-体動同期現象を客観的に定量化する分析手法を確立した。この分析手法を用いて、成長後の会話の原始的形態である、母子間の音声-体動同期現象、ならびに成人間の音声-体動同期現象を分析評価し、コミュニケーションにおける音声と動作の同期を現象論的に解明した。次に、上記の分析結果を基にして、マン・マシン・インタフェースに適用可能なモデルを制御論的に構築し、コミュニケーションにおける音声-体動同期現象の基本的メカニズムを明らかにした。さらに、理論の実際的应用の一例として、上記のモデルをマン・マシン・インタフェースに適用し、人間と機械の円滑な情報交換を図ることを目的とした、音声反応システムを開発し、その有効性を検証して、理論の妥当性を裏づけた。

## 1. はじめに

近年、コンピュータなどの情報機械の本格的な普及に伴い、人間-機械系が情報システムとして調和がとれたトータルシステムを形成することが望まれている。とりわけ、人間と機械の境界面であるマン・マシン・インタフェースは、人間と機械の調和の問題を具体的に解決する部分であり、システムの高度化に伴い、ますますその重要性を増している。一方、音声は人間に適した、自然で容易な情報伝達手段であり、マン・マシン・インタフェースにおける入出力方式として注目を浴びている。

人間同士の情報交換（コミュニケーション）においては、音声と動作が同期し、この音声-体動同期現象が円滑な情報交換に重要な役割を果たしている。従来この現象に対しては、定性的研究の色彩が強く、肉眼による分析が主流で、精度的にも研究の効率の面からも不十分であった<sup>1),2)</sup>。したがって、まず実験データの客観化、定量化、自動高速処理化などが必要である。

本論文では、人間同士の情報交換における音声-体動同期現象のメカニズムを解明するとともに、それをマン・マシン・インタフェースに応用し、人間と機械

の円滑な情報交換を図ることを目的としている。

## 2. 音声-体動同期現象の分析評価

### 2.1 概 要

本章では、従来定性的研究の色彩が強いこの分野に、新たに画像自動分析の手法が導入され、人間同士の情報交換（コミュニケーション）における音声-体動同期現象を客観的に定量化する分析手法が確立されている<sup>3)</sup>。研究対象としては、成長後の会話の原始的形態である、出生後まもない新生児と母親との情報伝達が選定されている。ここではおもに、母親の呼びかけに対する新生児の身体運動に現れる応答について分析評価され、母子相互の音声-体動同期現象の存在が定量的に解明されている。さらに、確認の意味で、成人同士の会話における話し手の音声と聞き手のうなずく動作との同期現象について分析評価されている。

成人は学習・習熟など各種の要因が含まれ、分析評価が複雑である。これに対し、新生児は遺伝情報だけを有する、いわば習熟度ゼロの状態と考えられる。したがって、本研究は人間の特性を把握する出発点として、すなわち適応性のゼロ点補正としての本質的重要性をもっている。

### 2.2 実験方法

実験系統図を図1に示す。研究対象の新生児は生後1～6日の健常児で、裸体で保育器に仰臥位で寝かされている。新生児に与えるテスト用の音声刺激としては、①母親が自由に語りかける場合 (free talk)、②決

† Application of Voice-Movement Synchronization Phenomenon to Man-Machine Interface by TOMIO WATANABE (Department of Information Engineering, Faculty of Engineering, Yamagata University).

\*\* 山形大学工学部情報工学科

まった言葉を書いたカードを次々に見せて発声してもらう場合 (pattern talk), ③母親以外の人間による語りかけの場合 (stranger talk), ④コンピュータで一樣乱数を使用して合成した雑音の場合 (random noise), ⑤その他 (肉声を変化させた音, テープに録音された音等), を選定した. 母親が語りかけを行う際には, 新生児に触れない. 音声と新生児の体動はビデオテープに収録する. 時間情報は 1/100 秒刻みのタイマをデジタル表示でビデオテープ画面にスーパーポーズした.

さらに, 母親と新生児の内部状態を把握することを目的として, 生体情報として心電図を同時計測した. 具体的には, ポリグラフシステム (日本光電社製, AB-620 G, RMP-6008) を用いて母子双方の心電図データを誘導してマイクロコンピュータ (日立, H 68/TR) に 1,310 Hz のサンプリングで取り込み, R 波 (心電図波形の内の棘波) をその勾配により検出し, R-R 間隔 (心拍間隔) を計測した.

### 2.3 分析方法

マイクロコンピュータ (TM 990/101 M) 制御による音声・画像自動分析のシステム構成図を図 1 に示す. 本システムは 1 フィールド (画面) を 245×312 画素に分割し, 1 画素について 12 ビットの階調で輝度信号をデジタル変換する. 画面上の任意の領域は, タイミングジェネレータ (TV モニタの水平同期信号, 垂直同期信号が入力されている) により設定され, タイミングジェネレータ出力の輝線により, その領域をモニタすることができる.

まず画像は VTR から正確なコマ送り機構を有するモーションアナライザ (SONY 社製, SUM 110) にコピーされる. 次に, マイクロコンピュータ制御により自動的にモーションアナライザの画像がコマ送りされ, 画像分析される. 具体的には, 各フィールドの特定の領域 (分析の対象にする部分, おもに腕部や脚部) を定めて, その領域内部を格子状に分割し, 一定の閾値で 0 と 1 のデジタル画像に変換する. たとえば図 2 に示すように, 1 フィールド 245×312 画素の内, 64×64 画素の領域について, 4×4 画素ずつ平均化を行い, 16×16 マトリクスの画像情報 (明暗度 1 ビット) への変換を行っている. その結果, 図に示すようなモザイク状の近似的図形ができるが, ある時刻  $t$  における画像と  $t+1/60$  秒における画像との間, すなわち相続く二つのフィールドの間で対応するマトリクスについて, その明暗度の変化のあったものの総数

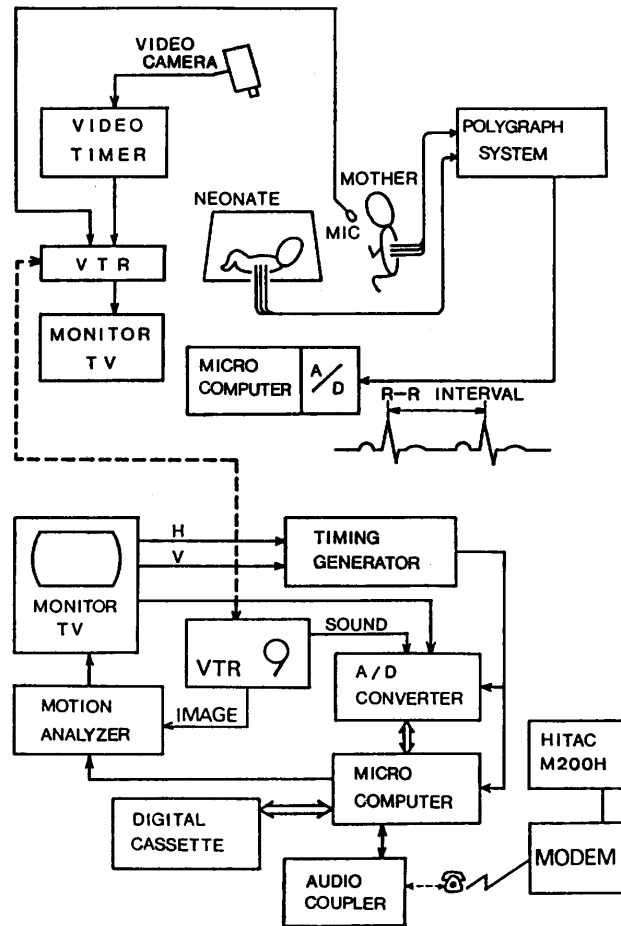


図 1 実験系統図とシステム構成図

Fig. 1 Set-up of the experiment & general arrangement and mechanism for analysis of voice and body movement by microcomputer (TM 990/101 M).

を  $x(t)$  と定義し, 動作の大きさを表す測度として定量化した. 図 2 の最下図はフィールド間の腕の動きの大きさを視覚化した図である. +印が新しく追加された部分で, -印が消えた部分である. 両者を合計すると, 明暗度の変化のあったものの総数  $x(t)$  は 30 となる.

一方, 音声については, 人間の音声帯域がたかだか 5 kHz であること, および画像データのサンプリング周期が 1/60 秒であることを考慮して以下のように定量化した. すなわち, ビデオテープ再生時に音声の電圧変換値を周波数 10 kHz でサンプリングして A/D 変換する. それと並行してフィールドごとにタイミングジェネレータから割込みをマイクロコンピュータにかけ, フィールド間 (1/60 秒間) 内の音声の変換値の最大値と最小値の差を算出し, その差をその時点の

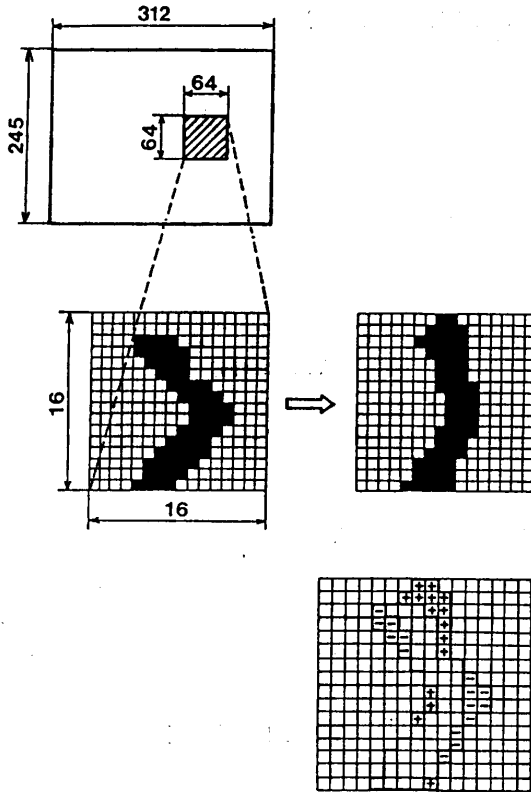


図 2 動作の大きさ  
Fig. 2 Analysis of body movement.

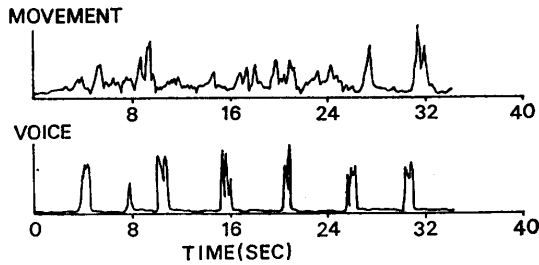
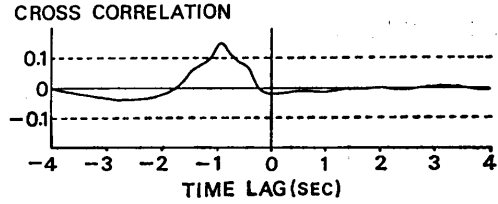


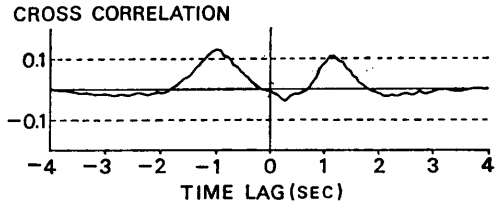
図 3 体動データと音声データの時間的变化  
Fig. 3 The data of body movement  $x(t)$  and of voice  $y(t)$ . These are derived from the mother's free talk to her 4 day-old neonate.

音声の大きさ  $y(t)$  と定義した。このように語りかけられる音声の時間的变化は、その強さ、すなわちエネルギーの大きさの変化で示している。したがって、言葉の意味内容とは無関係な定量化である。

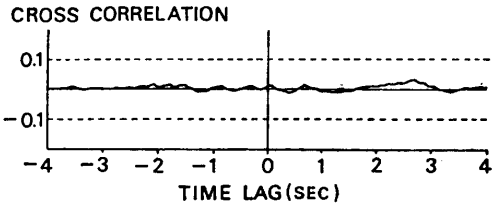
体動データ  $x(t)$  と音声データ  $y(t)$  の時間的变化の例を図 3 に示す。これら両者の相互関係を分析するために、以下の相互相関係数  $C(\tau)$  を導入した。



(a) pattern talk



(b) free talk



(c) random noise

図 4  $C(\tau)$  線図の典型例,  $N=2,048$

Fig. 4 Representative pattern of  $C(\tau)$  graph.  $C(\tau)$  graph: the result of calculated  $C(\tau)$  for the period of  $-4 < \tau < 4$  seconds.  $N=2,048$ .

$$C(\tau) = \frac{R_{xy}(\tau)}{\sqrt{R_x(0) \cdot R_y(0)}} \quad (1)$$

$$R_{xy}(\tau) = \frac{1}{T} \int_0^T (x(t) - \mu_x)(y(t+\tau) - \mu_y) dt$$

$$T = N \cdot \Delta T$$

ここで、 $\tau$ : 時間遅れ、 $T$ : 分析対象時間幅、 $N$ : データのサンプル数、 $\Delta T$ : サンプリング周期 (1/60 秒)、 $\mu_x$ :  $x$  の平均値、 $\mu_y$ :  $y$  の平均値、 $R_x(0)$ :  $x$  の分散、 $R_y(0)$ :  $y$  の分散、である。計算にあたっては FFT (Fast Fourier Transform) 法を用い、周波数領域の平滑化は Hanning ウィンドによって行っている。図 3 に示した体動データと音声データを用いて、区間  $-4 \leq \tau \leq 4$  秒について  $C(\tau)$  を算出した結果 (以下、たんに  $C(\tau)$  線図と呼ぶ) を図 4 (a) に示す。  $\tau < 0$  の領域は、体動に対し音声刺激が先行する領域である。一方、 $\tau > 0$  の領域は、音声刺激に対し体動が先行する領域である。

#### 2.4 分析評価

分析対象としては、新生児が開眼し、動いてはいるが泣いてはいない状態を選定した。新生児は睡眠状態

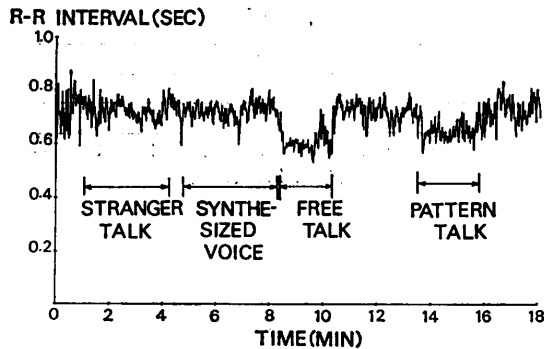


図5 母親の R-R 間隔の時系列的変化

Fig. 5 Time change in R-R interval of a mother.

あるいは啼泣状態が大部分を占め、また開眼していても意識状態が推移しやすく、音声刺激を受信するのに望ましい状態が継続されるのはたかだか60秒程度である。したがって、分析対象時間幅は約34秒(サンプル数  $N=2,048^*$ ) に設定した。

図4(a)に示した  $C(\tau)$  線は pattern talk の典型的な例である。 $\tau=-1$  秒付近にのみ高い山が存在し、 $\tau>0$  では存在しない ( $C(\tau)$  の最大値が0.06以上で、2変量正規分布を仮定、サンプル数2,048、危険率1%以下で有意な相関の存在を示す)。すなわち、新生児のみが母親の音声に反応し、母親の音声と同調する現象は見られない。一方、free talk の  $C(\tau)$  線図の典型的な例を図4(b)に示す。 $\tau<0$  では、 $\tau=-1$  秒付近に高い山がある。すなわち、新生児は母親の音声を受信してから約1秒の遅れで身体を動かしている。 $\tau>0$  では、 $\tau=1$  秒付近に高い山がある。母親も新生児の体動に誘起されて約1秒後に発声している。つまり、母と子は相互に同調している。random noise の  $C(\tau)$  線図の典型的な例を図4(c)に示す。音声刺激と体動との間に相関はなく、母親の語りかけの場合とは根本的に異なることがわかる。

しかしながら、このような典型的な例が絶えず出現するわけではない。この理由として母子双方の内部状態の影響が考えられる。その内部状態を把握する一つの指標として、心電図上の変化に着目した。心臓が神経系、内分泌系、各種の物質代謝の複雑な相互作用により支配されているという医学的知見によれば、心臓の拍動は人間の生理的・心理的状況を何らかの形で反映していると考えられる<sup>4),5)</sup>。一般にジェット機の操縦やコンピュータ端末の操作時において、操作者の

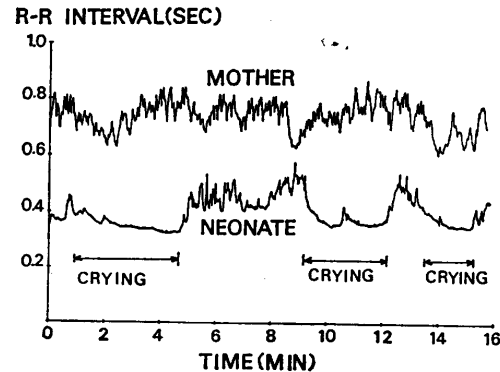


図6 母子双方の R-R 間隔の時系列的変化

Fig. 6 Time change in R-R interval of a mother and her neonate.

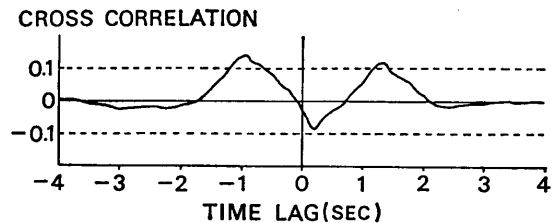


図7 成人の会話における  $C(\tau)$  線図の典型例、 $N=2,048$   
Fig. 7 Representative  $C(\tau)$  graph of synchronization of speaker's talk and listener's nodding in adult conversation.  $N=2,048$ .

R-R 間隔 (心拍間隔) は、平均値の低下に加えてその平均値まわりの分散が減少するという事実が発見されている<sup>6)-8)</sup>。実験時の母親の R-R 間隔の時系列的变化の一例を図5に示す。この図から pattern talk や free talk においては、上記と同様の変化が現れることがわかる。また母子双方の R-R 間隔の時系列的变化を同時計測した例を図6に示す。この例では、新生児が泣き出すと R-R 間隔が短くなり、その分散が小さくなる傾向が見られる。さらに、その前後で母親の R-R 間隔およびその分散が小さくなっており、母子の内部状態の同調が顕著に示されている。したがって、新生児の R-R 間隔およびその分散をリアルタイムでモニタし、新生児が音声を受信するのに望ましい内部状態と推定される時点から音声刺激を与えるなど、実験条件の安定化を図ることが可能である<sup>9)</sup>。

成人同士の会話における話し手の音声と聞き手のうなずく動作との  $C(\tau)$  線図の一例を図7に示す。free talk の  $C(\tau)$  線図と同様、 $\tau$  が  $\pm 1$  秒付近に高い山がある。つまり、話し手と聞き手とは相互に同調している。このように成人の会話における  $C(\tau)$  線図と母子

\* 相互相関係数をスペクトル解析により算出しているため、フーリエ変換のデータ数は2のべき乗であることが望ましい。

間の free talk における  $C(t)$  線図との波形パターンが類似していることから、成人の会話における本質的情報交換形態がすでに新生児期に存在することが推定される。

ここでは方法論についてのみ論じ、分析結果の詳細については別論文に譲る。

### 3. コミュニケーションモデル

#### 3.1 概要

本章では、前章の音声-体動同期現象の分析評価をもとにして、マン・マシン・インタフェースに応用可能なコミュニケーションモデルが制御論的に構築され、コミュニケーションにおける音声-体動同期現象の基本的メカニズムが明らかにされている。ここでは、スペクトル解析ならびに最小二乗解析の手法が導入され、モデルの各パラメータ値が推定されるなど、シミュレーション的研究が広範に展開され、次章の工学的応用への足掛りとされている。

#### 3.2 モデル化

母子間コミュニケーションモデルを構築するにあたり、まず新生児の反応モデルについて考える。図8に反応モデルを示す。ここで、入力  $y$  は音声刺激、出力  $x$  は新生児の体動を表す。  $x$  と  $y$  は各偏差で除して無次元化されている。識別関数は主として、①母親の音声と他人の音声 (noise 等) との識別、②音声刺激に対する新生児の注意度の決定、の役割を果たすフィルタである。したがって、新生児が音声刺激に対して注意を集中し、かつその音声刺激が母親である場合が、最も円滑に情報伝達が行われ、そのときの識別関数はゲイン1の比例要素として働くと考えられる。新生児の出力信号は次の三つの信号、①入力信号、②内部のフィードバック信号、③自己の出力信号のフィードバック信号、により決定されると仮定した。人間は感情の高まりに対して、実際に発声するか身体を動かすことにより、内部状態が変化すると考えられるので、③の信号を付加している。

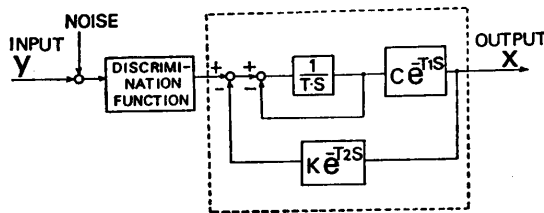


図8 反応モデル  
Fig. 8 Response model.

また、このモデルにおいて、パラメータ  $T, K, C, T_1, T_2$  は確定的な値ではなく確率変数であるが、ここではその期待値を考えることにし、定数として扱う。すなわち図中の破線で示したシステム  $H(S)$  は、定係数線形系であると仮定する。この系の伝達関数は、

$$H(S) = \frac{C \cdot \exp(-T_1 \cdot S)}{T \cdot S + 1 + K \cdot C \cdot \exp(-(T_1 + T_2)S)} \quad (2)$$

である。ここで各パラメータは次の意味をもつ。

$T$ : 積分要素  $1/S$  に対する時定数である。人間は入力信号のなかのある限度以上の高い振動数成分には追従することはできない。この現象は  $1/(T \cdot S + 1)$  によって表される。

$C$ : 入力に対する出力のゲインである。

$K$ : 身体を動かすことにより内部状態を高める場合には、正のフィードバックが働き、逆に身体を動かすことで内部状態を安定させる場合には、負のフィードバックが働くと考えられる。

$T_1$ : 刺激を受けてから反応するまでのむだ時間である。

$T_2$ : 自己の出力信号が内部状態に影響を及ぼすまでのむだ時間である。

次に、新生児の反応モデルを中心に構築した母子間コミュニケーションモデルを図9に示す。ここではできるだけ簡略化された、本質的なコミュニケーションモデルを構築することを目的として、母親の反応モデルも新生児の反応モデルと同じであると仮定した。図

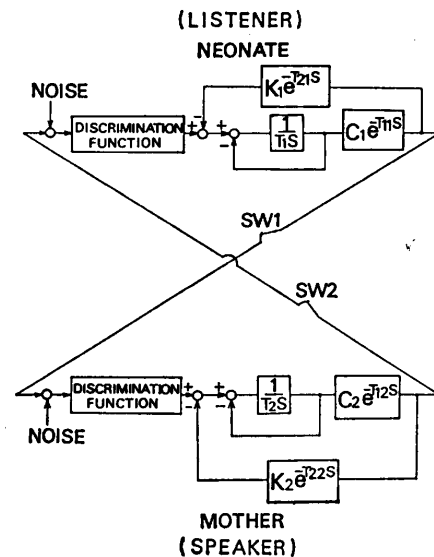


図9 コミュニケーションモデル  
Fig. 9 Communication model.

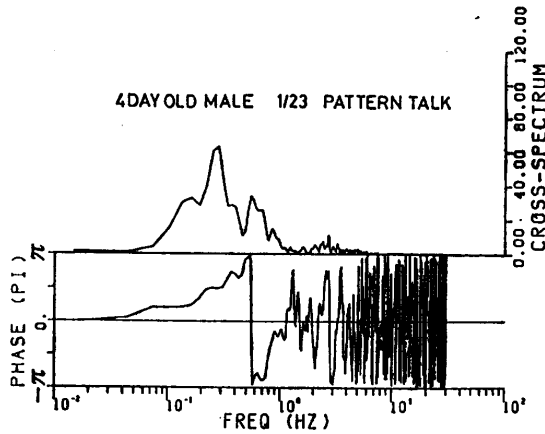


図 10 図 4(a) に示した pattern talk における相互スペクトル密度関数のゲインと位相  
Fig. 10 Gain and phase of cross-spectral density function for pattern talk as shown in Fig. 4 (a).

中に示した SW 1, SW 2 の ON-OFF により, 各種音声刺激に対する新生児の反応が説明できる。つまり, free talk は SW 1 と SW 2 が ON の場合であり, 母親と新生児とは相互に情報交換を行っている。また pattern talk は SW 2 だけが ON の場合であり, 母親が発する情報のみが新生児に伝達される。random noise は SW 2 が OFF の場合であり, 新生児は random noise を識別関数で識別し, 反応しない。

3.3 モデルの最適パラメータ値の推定

スペクトル解析の手法を用いて, 音声刺激に対する新生児の反応モデルの各パラメータ値を推定する。反応モデルの系  $H(j\omega)$  のゲイン  $|H(j\omega)|$  と位相  $\phi(\omega)$  は以下のごとく算出される。

$$|H(j\omega)| = \frac{C}{\sqrt{\zeta^2 + \eta^2}}$$

$$\phi(\omega) = \arctan \frac{\zeta \cdot \sin(T_1 \cdot \omega) + \eta \cdot \cos(T_1 \cdot \omega)}{\zeta \cdot \cos(T_1 \cdot \omega) - \eta \cdot \sin(T_1 \cdot \omega)} \quad (3)$$

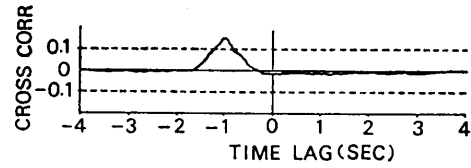
$$(-\pi < \phi(\omega) \leq \pi)$$

$$\text{where } \zeta = 1 + K \cdot C \cdot \cos((T_1 + T_2)\omega)$$

$$\eta = \omega \cdot T - K \cdot C \cdot \sin((T_1 + T_2)\omega)$$

入力  $y(t)$  のスペクトル密度関数を  $\Phi_{yy}(j\omega)$ , 系  $H(j\omega)$  を通過した入出力の相互スペクトル密度関数を,  $\Phi_{yz}(j\omega)$ , すなわち  $|\Phi_{yz}(j\omega)| \cdot e^{-j\theta(\omega)}$  とすると, 次の関係式が成立する。

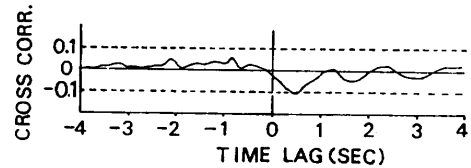
$$|\Phi_{yz}(j\omega)| \exp(-j\theta(\omega)) = |H(j\omega)| \exp(-j\phi(\omega)) \cdot \Phi_{yy}(j\omega) \quad (4)$$



$T=0.147, K=-0.184, C=1.14, T_1=0.816, T_2=0.412$

図 11 モデルの推定したパラメータ値から合成した  $C(\tau)$  線図

Fig. 11  $C(\tau)$  graph generated by fitting parameters of response model for pattern talk as shown in Fig. 4 (a).



$T=0.1, K=-0.9, C=1.5, T_1=0.82, T_2=0.40$

図 12 不安定領域にある例

Fig. 12 Example of unstable region.

$$|H(j\omega)| = \frac{|\Phi_{yz}(j\omega)|}{\Phi_{yy}(j\omega)}$$

$$\phi(\omega) = \theta(\omega)$$

一例として, 図 4 (a) に示した pattern talk について, 入力  $y(t)$  が識別関数のフィルタをそのまま通過する (識別関数がゲイン 1 の比例要素として働く) と仮定して,  $H(j\omega)$  の各パラメータ値を推定する。

図 10 に, 図 4 (a) に示した pattern talk における相互スペクトル密度関数のゲインと位相を示す。この図から fitting の対象とすべき周波数領域は, 1.2 Hz 以下の低周波数領域で十分であることがわかる。この低周波数領域において, (3) 式と (4) 式が適合するように, 最小二乗推定法により  $H(j\omega)$  の各パラメータ値を推定した ( $T=0.147, K=-0.184, C=1.14, T_1=0.816, T_2=0.412$ )<sup>\*</sup>。図 11 に推定したパラメータ値から合成した  $C(\tau)$  線図を示す。これは図 4 (a) の pattern talk の  $C(\tau)$  線図と類似している。また, 系  $H(j\omega)$  が不安定領域 ( $K < 0$  では,  $-1 > K \cdot C$ ) にある一例 ( $T=0.1, K=-0.9, C=1.5, T_1=0.82, T_2=0.40$ ) を図 12 に示す。このように音声刺激に対する新生児の反応モデルの各パラメータ値を推定することができる。またこの反応モデルを用いて, 音声刺激に対する新生児の体動をシミュレートすることが可能である<sup>10)</sup>。

\* 計算にあたっては, 東京大学大型計算機センターのライブラリ 最小二乗標準プログラム SALS (Statistical Analysis with Least-Squares Fitting) を利用した。

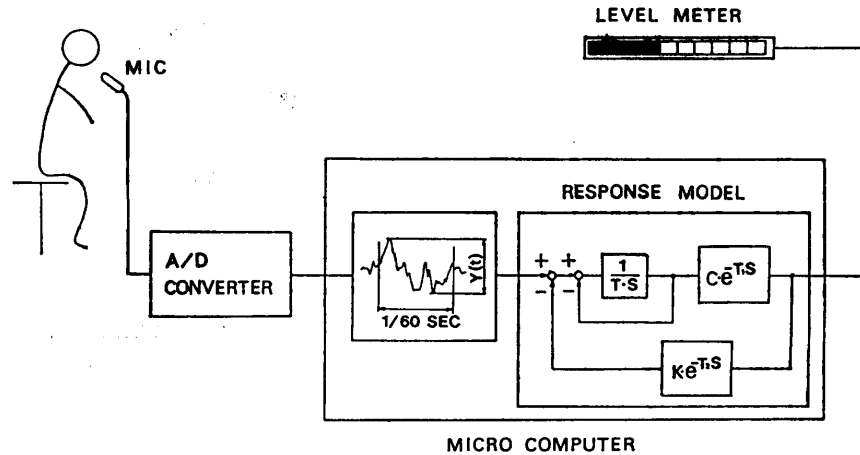


図 13 VR システムの概略構成図

Fig. 13 General arrangement and mechanism of VR (Voice Reaction) system.

#### 4. マン・マシン・インタフェースへの応用

##### 4.1 概要

本章では、理論の実際的应用の一例として、前章で構築したモデルがマン・マシン・インタフェースに適用され、人間と情報機械との円滑な情報伝達を図ることを目的とした、音声反応システム：VR (Voice Reaction) システムが開発され、理論の妥当性が裏づけられている。具体的には、VR システムを導入したビデオカメラに対面しての情報伝達が研究対象に選定され、VR システムの有効性について、官能検査による主観的評価、ならびに生体情報（心電図、E. O. G. (electro-oculography)) を利用した客観的評価に基づき、総合的に評価・検討されている。

##### 4.2 VR システム構成

VR システムの構成概略図を図 13 に示す。本システムは、前章で構築した反応モデルが組み込まれ、入力音声に対して、モデルの最適パラメータ値で、人間の反応動作に近い、適切なタイミングでレベルメータの LED が点滅する音声反応システムである。

具体的には、まずマイクロフォンからの入力音声は、10 kHz のサンプリングで A/D 変換され、1/60 秒間内でのその電圧変換値の最大値と最小値の差、すなわち 2 章で定義した音声の大きさが算出される。次に、この音声の大きさをモデルの入力  $y$  として、その出力  $x$  が 1/60 秒ごとにマイクロコンピュータ内でソフトウェア的に以下の差分方程式により算出される。

$$X(t+T_1+1) = X(t+T_1) + \frac{\Delta T}{T} (C \cdot Y(t) - X(t+T_1) - C \cdot K \cdot X(t-T_2)) \quad (5)$$

$$\text{where } \Delta T = \frac{1}{60} \text{ sec, } T_1 = \frac{T_1}{\Delta T}$$

$$T_2 = \frac{T_2}{\Delta T}$$

理想的なマン・マシン・インタフェースの情報伝達においては、モデルの識別関数は、ゲイン 1 の比例要素として働くので、図 13 の VR システムでは省略されている。最後に、出力  $x$  の大きさに対応して、任意の数の LED が点滅する。

##### 4.3 VR システムの有効性

###### 4.3.1 実験方法

VR システムの有効性を検証するために、VR システムを導入したビデオカメラに対面しての情報伝達を研究対象に選定した。実験条件としては、

- ① 前章で推定したモデルの最適パラメータ値で LED を点滅させた場合 (with VR system)
- ② VR システムを導入しない、すなわち VR システムの LED を点滅させない場合 (without VR system)
- ③ ①の対照実験条件として、入力音声に依存せず、一様乱数によりランダムに VR システムの LED を点滅させた場合 (random)

の 3 条件を選定した。

VR システムの有効性を定量的に決定するために、以下の Bradley-Terry モデル<sup>11)</sup>を想定し、対比較による順序づけを行った。

表 1 判定結果

Table 1 Result of sensory evaluation.

i \ j	①	②	③	TOTAL	$\pi$
①		24	25	49	0.961
②	1		11	12	0.018
③	0	14		14	0.021

①with VR system, ②without VR system,  
③with VR system (random).

表 2 順序のグループ分け

Table 2 Classification of order.

i \ j	①	②,3	TOTAL	$\pi$
①		49	49	0.98
②,3	1		1	0.02

$$P_{ij} = \pi_i / (\pi_i + \pi_j), \sum \pi_i = 1 \quad (6)$$

$\pi_i$ :  $i$  に対する強さ (好み) の割合

$P_{ij}$ :  $i$  が  $j$  よりもよいと判定する確率

実験は、上記の実験条件について、被験者に各3分間ずつ対比較 (①-②, ②-③, ③-①) により音声入力させた後、話しやすく、音声入力が容易な実験条件を二者択一させた。被験者は視覚・聴覚ともに正常な21~26歳の男女25人である。

さらに、被験者の官能検査による主観的評価に加えて、生体情報として心電図 (R-R 間隔) ならびに E. O. G. を同時計測し、VR システムの有効性を客観的に評価・検討した。

#### 4.3.2 官能検査による評価

被験者25人による判定結果と、Bradley-Terry モデルを想定した場合の音声入力の容易さ  $\pi$  を最尤推定した結果を表1に示す。モデルの全体としての整合性を有意水準1%で尤度比検定した結果、モデルが成立し、特定の苦手関係はなかった。次に三つの  $\pi$  のうち、任意の二つを等しく置くことができるか尤度比検定し、順序のグループ分けを行った。その結果、有意水準1%で表2に示すような①と②,3の二つのグループに分けられ、前者が後者よりも勝っていることが認められた。したがって、VR システムを導入したほうが音声入力が容易であることがわかる。また音声入力の容易さの要因として、たんに VR システムの LED の点滅の有無が問題ではなく、入力音声に対して VR システムが適切なタイミングで反応することがあげられる。

E.O.G.

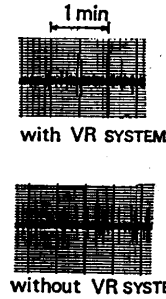


図 14 E. O. G. の典型例  
Fig. 14 Example of E. O. G.

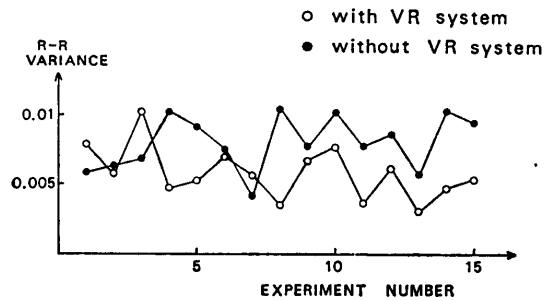


図 15 R-R 間隔の分散の推移  
Fig. 15 Time change in variance of R-R interval.

#### 4.3.3 生体情報を利用した評価

##### (1) E. O. G. (electro-oculography)

人間同士の情報交換においては、うなずく動作だけでなく、顔の表情、とくに目の動きが重要な役割を果たしている。

本研究においては、この眼球運動を、眼球の前後で生じる電位差を電氣的に誘導し、計測した。この方法は一般に E. O. G. と呼ばれる<sup>12)</sup>。図14に VR システムを用いた場合と用いない場合における E. O. G. の典型例を示す。この図から、VR システムを用いると眼球運動が少なく、視線が安定していることがわかる。この傾向はすべての被験者にみられた。したがって、視線の安定性が VR システムを用いた場合の音声入力の容易さの一要因であると考えられる。

##### (2) 心電図

個人別に VR システムを用いた場合と用いない場合について、交互に音声入力を2分間ずつ15回繰り返した場合の R-R 間隔の分散の推移の代表例を図15に示す。この図から明らかなように、前半においては、個人差により VR システムの有無の差異は明白ではないが、ある程度の集中度、緊張度を必要と



する後半においては、VRシステムを用いた場合のほうがR-R間隔の分散が小さく、しかも前半に対してR-R間隔の分散が上昇する傾向はみられず、音声入力に集中しやすいことがわかる。

### 5. むすび

本論文で得られた結論のうち、基本的な部分を要約すると次のとおりである。

(1) 従来定性的研究の色彩が強いこの分野に、新たに画像自動分析の手法を導入し、人間同士の情報交換(コミュニケーション)における音声-体動同期現象を客観的に定量化する分析手法を確立した。

(2) 上記の分析手法を用いて、成長後の会話の原始的形態である。母子間コミュニケーションにおける音声-体動同期現象、ならびに成人間コミュニケーションにおける音声-体動同期現象を分析・評価し、コミュニケーションにおける音声と動作の同期を現象論的に解明した。

(3) 上記の分析結果をもとにして、マン・マシン・インタフェースに応用可能なコミュニケーションモデルを制御論的に構築し、コミュニケーションにおける音声-体動同期現象の基本的メカニズムを明らかにした。

(4) 理論の実際の応用の一例として、上記のモデルをマン・マシン・インタフェースに適用し、人間と情報機械との円滑な情報伝達を図ることを目的とした、音声反応システム:VR (Voice Reaction) システムを開発し、その有効性を検証して、理論の妥当性を裏づけた。

**謝辞** 本研究において、適切なご指導とご助言を賜った東京大学工学部石井威望教授に深く感謝の意を表します。また、多大のご支援を賜った東京大学医学部小林登教授、ならびに愛育病院関係各位に深謝します。

### 参考文献

- 1) Condon, W.S. and Sander, L.W.: Neonate Movement is Synchronized with Adult Speech; Interactional Participation and Language Acquisition, *Science*, Vol. 183, pp. 99-101 (1974).
- 2) Condon, W.S. and Ogston, W.D.: Sound Film Analysis of Normal and Pathological Behavior Patterns, *J. Nerv. Ment. Dis.*, Vol. 143, No. 4, pp. 338-342 (1966).
- 3) 石井威望, 渡辺富夫: 画像自動分析による母子間コミュニケーションにおける音声-体動同期現象の研究, 第23回情報処理学会全国大会, pp. 1119-1120 (1981).
- 4) Luczak, H. et al.: An Analysis of Heart Rate Variability, *Ergonomics*, Vol. 16, No. 1, pp. 85-97 (1973).
- 5) Ashmead, D.H. et al.: Neonates' Heart Rate, Sucking Rhythm, and Sucking Amplitude as a Function of the Sweet Taste, *J. Exp. Child Psychol.*, Vol. 29, No. 2, pp. 264-281 (1980).
- 6) 石井威望, 渡辺富夫: 生体情報を利用したマン・マシン・インターフェースの評価に関する研究, 日本機械学会講演論文集, No. 817-1, pp. 157-159 (1981).
- 7) Sekiguchi, C.: Evaluation Method of Mental Workload under Flight Conditions, *Aviat. Space Environ. Med.*, Vol. 49, No. 7, pp. 920-925 (1978).
- 8) Sekiguchi, C.: Frequency Analysis of Heart Rate Variability under Flight Conditions, *Ibid.*, Vol. 50, No. 6, pp. 625-634 (1979).
- 9) 石井威望, 渡辺富夫: 画像自動分析による母子間コミュニケーションにおける音声-体動同期現象の研究(第2報), 第24回情報処理学会全国大会, pp. 1117-1119 (1982).
- 10) 渡辺富夫: マン・マシン・インターフェースにおける動作と音声の同期に関する研究, 東京大学博士論文(1983.3).
- 11) Bradley, R.A. and Terry, M.E.: The Method of Paired Comparisons, *Biometrika*, Vol. 39, pp. 324-345 (1952).
- 12) 池田光男: 視覚の心理物理学, 森北出版, 東京(1975).

(昭和58年5月16日受付)

(昭和58年9月13日採録)