

コンピュータ画像処理による母子相互作用の研究

石井 威望 (東京大学工学部産業機械工学科)
小林 登 (東大小児科)
高橋 悦二郎 (総合母子保健センター)

序 論

人間と機械の境界面であるマン・マシン・インターフェースは、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、益々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・インター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。本研究においては、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交換を研究対象に選定し、コミュニケーションにおける音声-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・インターフェースに応用し、人間と機械との円滑な情報交換を図ることを目的としている。

本研究では、既に第1年度において、コンピュータ画像分析の手法を導入し、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。その分析の結果、新生児が母親の呼びかけを識別し、四肢を同期的に動かすことが明らかになった。

第2年度においては、上記の分析手法を用いて、成人同士の会話における音声-体動同期現象を分析・評価するとともに、母子間コミュニケーションにおける音声-体動同期現象の分析結果との比較など、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。

本年度においては、前年度までの音声-体動同期現象の分析結果を基にして、システム論的にコミュニケーションモデルを構築し、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ている。更に、理論の実際的応用の一例として、上記のモデルをマン・マシン・インターフェースに適用し、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、音声応答システムを開発し、理論の妥当性を裏付けている。

コミュニケーションモデル

母子間並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・評価した結果、音声-体動同期現象は原始的かつ本能的な情報交換形態であることが明らかになった。従って、その同期現象の本質的メカニズムを解明するためのモデル化には、母子間の情報交換を中心としたモデル化が有効である。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1に新生児の反応モデルを示す。このモデルは習熟度零の状態を仮定している。ここで入力 y は音声刺激、出力 x は新生児の体動を表わす。識別関数は主として、

- (i) 母親の音声と他人の音声(ノイズ等)を識別する。
- (ii) 外界からの音声刺激に対する新生児の注意の度合を決定づける。

の役割を果たすフィルタである。新生児の出力信号は、次の3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。このモデルにおいて、係数 T 、 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。即ち、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、

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

となる。ここで各係数は次の意味をもつ。

K : 身体を動かすことにより内部状態を高めていると考えられる場合には、 $K < 0$ 、即ち

正のフィードバックであり、逆に身体を動かすことで内部状態を安定させると考えられる場合には、 $K \geq 0$ 、即ち負のフィードバックが働くと考えられる。

T: 積分要素 $\frac{1}{S}$ に対する時定数である。

C: 入力に対する出力のゲインを表わす。

T₁: 刺激を受けてから反応するまでのむだ時間である。

T₂: 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。

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2に示す。ここでは出来るだけ簡略化された、本質的なコミュニケーション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。図中に示したSW1, SW2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 (自由な語りかけ) の場合は、SW1とSW2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 (決まった言葉を書いたカードを母親に見せて発声してもらう) の場合はSW2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、新生児からの情報は母親に伝達されない。いま、母親の音声刺激 y と新生児の体動 x の実験データを基にして、モデルの推定パラメータ値から合成した $C(\tau)$ 線図^(註) の一例を図3に示す。図3-AはPattern Talkにおける典型的な $C(\tau)$ 線図である。また図3-BはFree Talkにおける典型的な $C(\tau)$ 線図である。このようなモデルの各パラメータ値を推定することにより、各種の音声刺激に対する $C(\tau)$ 線図を合成することが可能である。従って、本モデルはコミュニケーションにおける音声-体動同期現象の本質的メカニズムを有するモデルとして有効であると考えられる。

マン・マシン・インターフェースへの応用

前章で構築したコミュニケーションモデルは、1対の反応モデルの内、一方の反応モデルを情報機械に置き換えれば、人間と情報機械とのコミュニケーションモデルと見なすことができる。

本章ではそのモデルをマン・マシン・インター

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、音声応答システム: VR (Voice Response) システムを開発し、その有効性を検証している。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を研究対象に選定し、VRシステムの有効性について、官能検査(アンケート)による主観的評価、並びに生体情報(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・検討している。

VRシステムの構成概略図を図4に示す。本システムは入力音声に対して、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が点滅する音声応答システムである。VRシステムの有効性についてアンケートした結果、VRシステムを用いた方が話し易いことが明らかになった。

図5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.の代表例を示す。この図からVRシステムを用いると眼球運動が少く、視線が安定していることがわかる。この傾向は全ての被験者(25人)に示された。従って、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。

個人別にVRシステムを用いた場合と用いない場合とについて交互に音声入力を2分間づつ15回繰り返した場合のR-R間隔(心拍間隔)の分散の推移の代表例を図6に示す。この図から明らかかなように、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方が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示されず、音声入力に集中しやすいことがわかる。

結 論

本研究で得られた結論を要約すると以下の通りである。

(1) 前年度までの分析結果を基にして、システム論的にコミュニケーションモデルを構築し、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。

(2) 理論の実際的応用の一例として、上記のモデルをマン・マシン・ソリターフェースに適用し、その有効性を検証して、理論の妥当性を裏付けた。人間と情報機械との円滑な情報伝達を図ることを

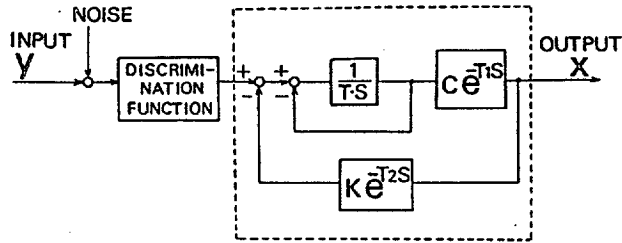


図1. 新生児の反応モデル

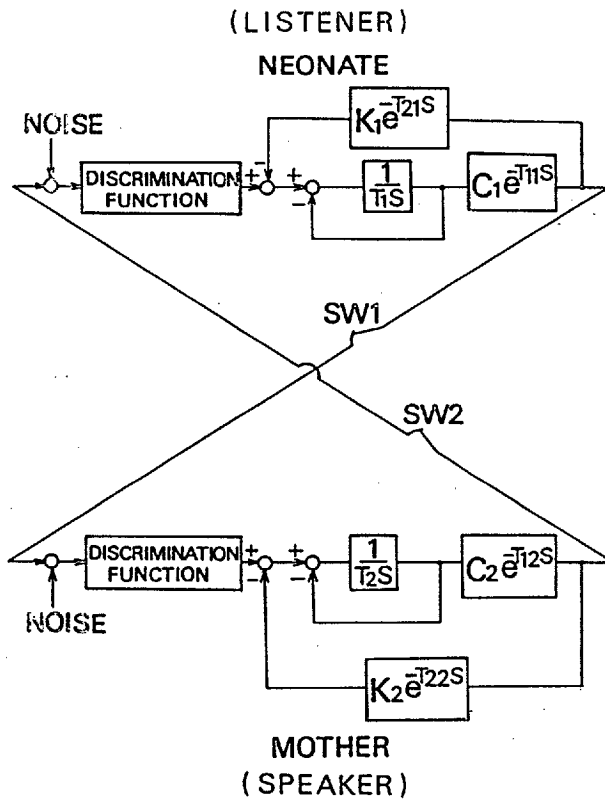
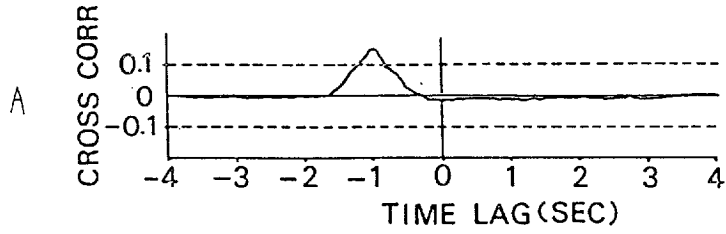
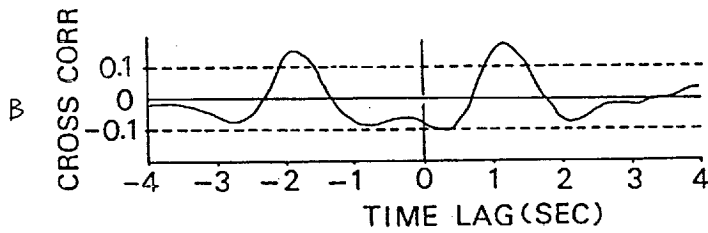


図2. 母子間コミュニケーションモデル



$$T = 0.147, K = -0.184, C = 1.14,$$

$$T_1 = 0.816, T_2 = 0.412$$



$$T_1 = 0.135, K_1 = 0.07, C_1 = 1.06, T_{11} = 1.75, T_{21} = 0.20$$

$$T_2 = 0.14, K_2 = 0.03, C_2 = 1.05, T_{12} = 1.12, T_{22} = 0.28$$

図3. モデルの推定したパラメータ値から合成したC(φ)線図

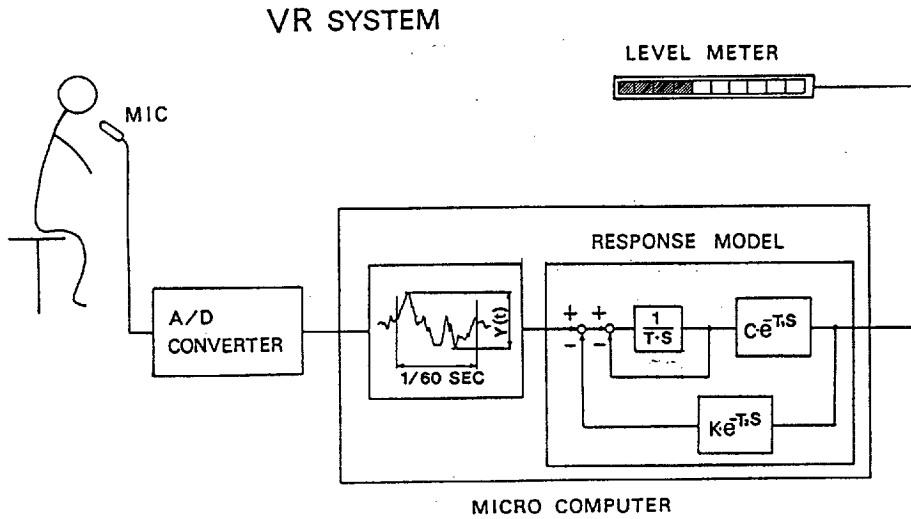
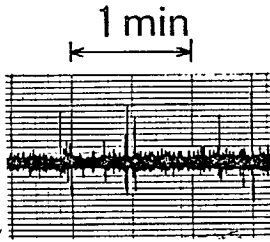
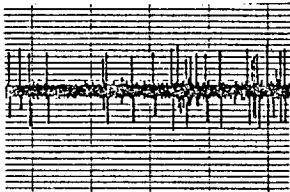


図4. VRシステムの概略構成図

E.O.G.



with VR SYSTEM



without VR SYSTEM

図5. E.O.G. の典型例

注) $x(t)$ と $y(t)$ の相互相関係数 $C(\tau)$ を区間
 $-4 \leq \tau \leq 4$ 秒について算出した図。

$$C(\tau) = \frac{Rxy(\tau)}{\sqrt{Rx(0) \cdot Ry(0)}}$$

$$Rxy(\tau) = \frac{1}{T} \int_0^T (x(t) - \mu x)(y(t + \tau) - \mu y) dt$$

ここで、

T : 記録時間長 τ : 時間遅れ
 μx : x の平均値 μy : y の平均値
 $Rx(0)$: x の分散 $Ry(0)$: y の分散

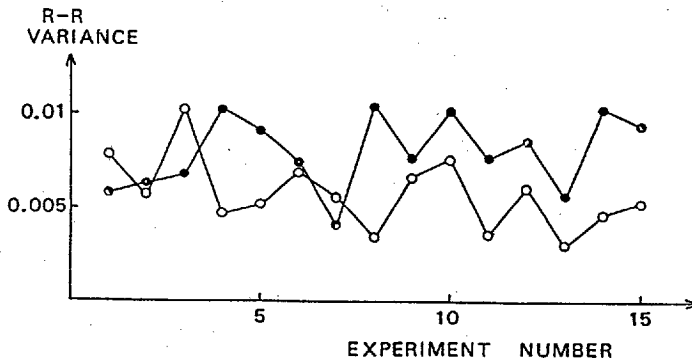
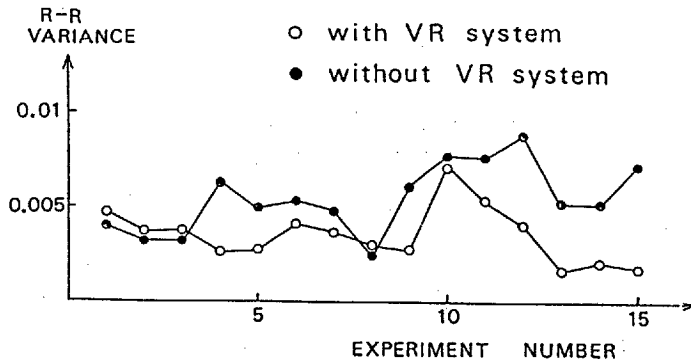
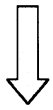
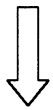


図6. 個人別に見た, R-R間隔の分散の推移の典型例



検索用テキスト OCR(光学的文字認識)ソフト使用

論文の一部ですが、認識率の関係で誤字が含まれる場合があります



序 論

人間と機械の境界面であるマン・マシン・インターフェースは、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、益々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・インター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、この音声一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。本研究においては、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交換を研究対象に選定し、コミュニケーションにおける音声一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・インターフェースに応用し、人間と機械との円滑な情報交換を図ることを目的としている。

本研究では、既に第1年度において、コンピュータ画像分析の手法を導入し、音声一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。その分析の結果、新生児が母親の呼びかけを識別し、四肢を同期的に動かすことが明らかになった。

第2年度においては、上記の分析手法を用いて、成人同士の会話における音声一体動同期現象を分析・評価するとともに、母子間コミュニケーションにおける音声一体動同期現象の分析結果との比較など、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。

本年度においては、前年度までの音声一体動同期現象の分析結果を基にして、システム論的にコミュニケーションモデルを構築し、コミュニケーションにおける音声一体動同期現象の基本的メカニズムを明らかにしている。更に、理論の実際的応用の一例として、上記のモデルをマン・マシン・インターフェースに適用し、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、音声応答システムを開発し、理論の妥当性を裏付けている。