

成人間コミュニケーションにおけるエントレインメント (音声-体動同期現象)の分析†

渡 辺 富 夫††

生体リズムが相互に同期化する現象をエントレインメントと呼ぶ。人間同士のコミュニケーションにおいて、音声と動作・表情が同期する現象、すなわちコミュニケーションにおけるエントレインメントが存在し、円滑な情報交換に重要な役割を果たしている。しかしながら、従来この現象に対しては定性的研究の色彩が強く、定量的・客観的評価が困難であった。本論文では、このコミュニケーションにおけるエントレインメントを客観的に定量化する分析手法を提案し、成人同士の会話におけるエントレインメントを分析評価した。その分析の結果、コミュニケーションにおいて、話し手の音声に対する聞き手のうなずく動作の相互相関係数上の時間遅れは、0.7秒から1.4秒の範囲に、また聞き手のうなずく動作に対する話し手の音声の相互相関係数上の時間遅れは、0.8秒から2.0秒の範囲に、話し手と聞き手が相互に同期化するエントレインメントが存在することが明らかになった。さらに、官能検査を通して、本分析手法の有効性を検証した。

1. ま え が き

生命現象の一つの特徴としてリズム的变化が存在し、外界の条件に対しても何らかの関係をもっているのではないかと考えられている。このような生体リズムが相互に同期化する現象をエントレインメント(entrainment)と呼ぶ。

人間同士のコミュニケーションにおいても、音声と動作・表情が同期する現象、すなわちコミュニケーションにおけるエントレインメント(以後、たんにエントレインメントと呼ぶ)が存在し、円滑な情報交換に重要な役割を果たしている。しかしながら、従来この現象に対しては定性的研究の色彩が強く、研究者による主観的行動観察が中心であり、定量的・客観的評価が困難であった^{1),2)}。

著者はすでに、成人間の会話の原始的形態であると推定される母子間のエントレインメントを研究対象に選定し、おもに母親の語り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した³⁾⁻⁵⁾。その分析の結果、新生児が母親の語りかけをコンピュータ合成音(雑音)と識別し、その語りかけに対して四肢を同期して動かすことが明らかになった。このように、エントレインメントは人間にとって本質的な情報交換形態であると考えられる。したがって、このエントレインメントのメカニズムが人間と機械の境界部(マン・マシン・インタフェース)に導入されるならば、

人間と機械の円滑な情報交換が図られ、人間性を尊重したヒューマンインタフェースの実現に役立つと期待される⁶⁾。

たとえば、人間と知能ロボットとの協同作業において、人間の語りかけに対して、知能ロボットが適切なタイミングで反応すれば、まず話しやすいばかりでなく、そのタイミングによって、話者は知能ロボットへの情報伝達度を逐次把握することができる。一方、知能ロボット側も、音声に対する人間の反応動作をモニタして、伝達情報に対する人間側の理解度を逐次判断し、人間にとって最適な情報を伝達することが可能になる。

本論文では、以上の観点に立脚して、まず人間同士のエントレインメントを客観的に定量化する分析手法を提案し、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エントレインメントを分析評価している。さらに、官能検査を通して、本分析手法の有効性を検証している。

2. 実験方法

実験系統図を図1に示す。話し手の音声と聞き手のうなずく動作は、ビデオタイマ(最小時間単位1/100秒)を介してビデオテープに収録すると同時に、聞き手の内部状態(生理的・心理的状态)を客観的に把握するための生体情報として、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)を用いて被験者の心電図データを計測用マイクロコンピュータ(H 68/TR)に取り込み、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。

† An Analysis of Entrainment (Voice-movement Synchronization) in Communication among Adults by TOMIO WATANABE (Department of Information Engineering, Faculty of Engineering, Yamagata University).

†† 山形大学工学部情報工学科

表 1 実験の構成
Table 1 Experimental subjects.

Speaker	Listener	Contents	Place
S1 Student	L1	Explanation of voice-movement synchronization in mother-neonate communication	Laboratory
	L2 Student		
	L3		
S1 Student S2	L4 Student	Explanation of electronic instrument	
S1 Student	L5	Explanation of experiment on mother-neonate interaction	Sickroom
	L6 Mother		
	L7		
	L8		
S3 Doctor	L8 Mother		

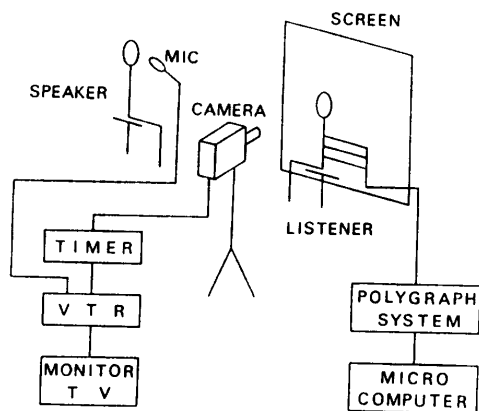


図 1 実験系統図
Fig. 1 Set-up of the experiment.

本実験では、実験条件の統制を図るために、あらかじめ話し手と聞き手を設定して、一方から他方への情報伝達におけるエントレインメントに着目した。表 1 に実験の概要を示す。被験者はすべて視覚・聴覚ともに正常な成人である。実験は同一人物の話し手に対して聞き手を換えた場合、あるいは同一人物の聞き手に対して話し手を換えた場合など、多種にわたっている。また実験環境としては、病室で新生児を囲んでの母親と医師との会話など、比較的自然的な環境での実験と、外部からの聴覚刺激等の外乱を排除した実験室内でのモデル実験とを設定した。

3. 分析方法⁷⁾

3.1 動きと音声の定量化

まず動きについては、正確なコマ送り機構を有するモーションアナライザ(ソニー社製, SUM 101)を用いて、フィールド^{*}(1/60 秒間)ごとにうなずく動

作の有無により、うなずく動作期間の動作の大きさ $M(t)$ を 1, それ以外の期間の $M(t)$ を 0 として 2 値化した。

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) 間内の音声の電圧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 t の音声の大きさ $V(t)$ と定義した。本研究では、コミュニケーションにおけるエントレインメントのリズムの要素に着目し、言葉の意味的内容とは無関係な定量化を行った^{**}。

3.2 画像データと音声データの同期

音声データの取得はリアルタイムで行うために、画像データ $M(t)$ と音声データ $V(t)$ との時間軸上の同期をとるには、画面上での音声データの取得開始時間を知る必要がある。音声データの取得を開始すると同時に、タイミングジェネレータ出力による輝線がモニタ画面上から消える。これは画面上に輝線を出力していた信号が、マイクロコンピュータに対する割込みに

* 本論文でいうコマ送りは、1フィールド(1/60 秒間)の送りを意味する。飛越走査により 2フィールドで 1フレーム(1/30 秒間)が構成される。

** 言葉のわからない新生児が母親の語りかけに対して腕を同期させて動かすエントレインメントが存在するという研究報告(文献 1), 3)は、このエントレインメントが後で発達してくる言語によるコミュニケーションよりも、むしろ人間生物学的にみて、より普遍的であり、本質的重要性をもっていることを示すものであると、著者は考えている。

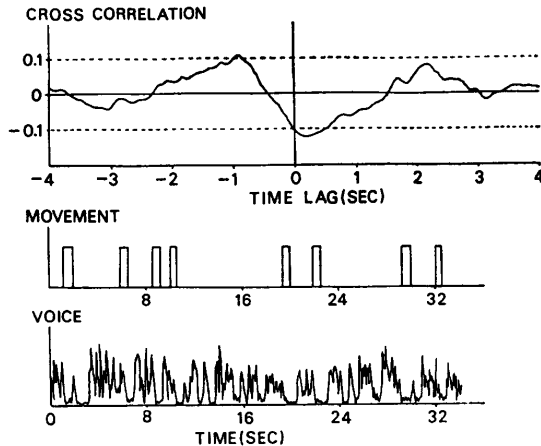


図2 動きの大きさと音声の大きさの時系列的変化ならびに両者のエントレノグラム, $n=2,048$

Fig. 2 The data of movement $M(t)$ and of voice $V(t)$, and cross-correlation curve of entrainment (entrainogram). $n=2,048$.

使用されるためである。したがって、輝線の消去時間を計測することにより、音声データの取得開始時間を知ることができる。具体的には、画面上の時間情報と輝線をビデオカメラを通して VTR に録画し、音声データ取得後、ビデオテープを再生して、輝線の消去時間を計測している。

3.3 評価法

$M(t)$ と $V(t)$ の時間的变化の例を図2下図に示す。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。

$$C(\tau) = \frac{\sum_{i=1}^n \{V(t+i/60+\tau) - \mu_V\} \{M(t+i/60) - \mu_M\}}{\sqrt{\sum_{i=1}^n \{V(t+i/60+\tau) - \mu_V\}^2} \sqrt{\sum_{i=1}^n \{M(t+i/60) - \mu_M\}^2}}$$

μ_V, μ_M は、 $V(t), M(t)$ の平均値である。分析対象時間幅は 2,048 フィールド ($n=2,048$)、すなわち $2,048 \div 60 \approx 34$ 秒間である。図2下図に示した $M(t)$ と $V(t)$ の時系列的変化について、 τ が -4 秒から 4 秒まで $1/60$ 秒ごとに $C(\tau)$ を算出した結果 (以下、エントレノグラムと呼ぶ) を図2上図に示す。 $\tau=0$ は、 $M(t)$ と $V(t)$ を時間軸上で全然ずらさずに $C(\tau)$ を計算することを意味する。 $\tau < 0$ の領域は、うなづく動作に対し音声が発生する領域である。一方、 $\tau > 0$ の領域は、音声に対しうなづく動作が発生する領域である。

本論文では、 $C(\tau)$ の最大値が 0.06 以上 (2変量正規分布を仮定すれば、 $n=2,048$ のとき $1/\sqrt{n-3} \approx$

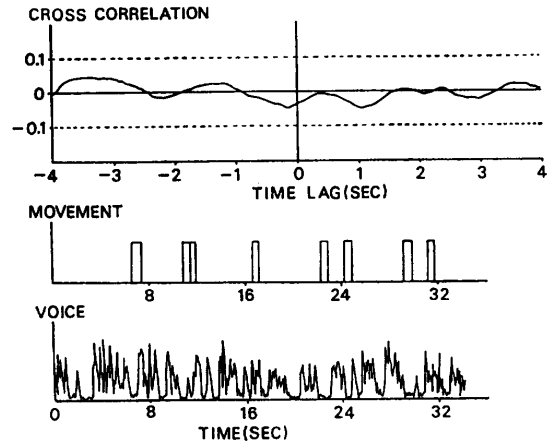


図3 図2のうなづく動作の時間間隔のみをシミュレーションによって指数分布させた場合

Fig. 3 Simulation of time interval of nodding (movement) in Fig. 2 (exponential distribution).

0.0211 であり、危険率 1% の両側検定では、域値は $\tau > 0.0211 \times 2.58 \approx 0.057$ となる) で高い山を示すとき、音声と動作に有意な相関があると判定した。したがって、図2のエントレノグラムでは両者は明らかに有意な相関があるといえる。

4. 結果および考察

図2に示したエントレノグラムは、成人間のエントレインメントの典型的な例である。 $\tau < 0$ では -1 秒付近に高い山があり、聞き手が話し手の音声に対して約 1 秒遅れてうなづくしている。一方、 $\tau > 0$ では 2 秒付近に高い山があり、話し手が聞き手のうなづく動作をみて約 2 秒後に語りかけている。つまり、話し手と聞き手は相互に同期している。このエントレノグラムの評価法を検証するために、図2に示した $M(t)$ について、うなづく動作の時間間隔がシミュレーションによって指数分布するように乱数を発生させて作成した $M(t)$ によりエントレノグラムを求めたのが図3である。上記のようなエントレインメントの特徴的な傾向はみられず、音声とうなづく動作とは無相関である。

分析総数 22 例について、有意な相関の存在を示すと考えられる $C(\tau)$ の極大値での時間遅れ τ を表2に示す。モデル実験においては、ほとんどすべての分析結果に上記と同様のエントレインメントが現れている。しかしながら、病室での母親と医師あるいは学生との実験では、モデル実験に比べて、顕著なエントレ

表 2 分析結果
Table 2 Results.

SPEAKER	LISTENER	τ at peaks
S1	L1	-0.8 , 0.8 -0.8 , 0.9
	L2	-1.4 , 1.5 -1.0 , 1.1
	L3	-0.9 , 2.0 -1.4 , 1.6
S2	L4	————— 1.2
S1		-0.6* -0.8 , 0.8 -1.0 , 1.0
S1	L5	-0.6* 0.7*
	L8	-0.3 , -1.3 -0.7* , 1.4 -1.1 , 1.6
	L7	-0.3* -1.2*
	L6	————— ————— —————
S3	L8	————— ————— —————

————— NO CORRELATION
★ $C(\tau) < 0.1$

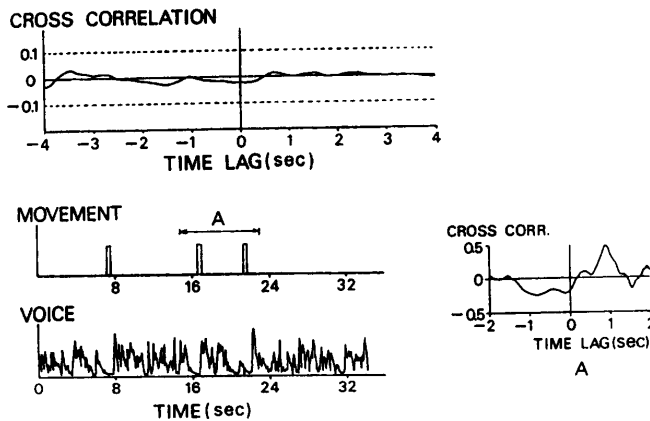


図 4 うなずく動作回数が少ない場合
Fig. 4 In the case of unfrequent nodding.

インメントの出現割合が低くなっている。これは、母親の注意が新生児に注がれ、話しに対する集中度が低下しているためであると考えられる。話し手 S2 と聞き手 L4 の情報伝達において、聞き手のうなずく動作回数が分析対象時間幅 34 秒間で 3 回と少ない場合のエントレノグラムと、うなずく動作を中心にその前後に時間軸 (分析対象時間幅 $T=8.5$ 秒, サンプル数 $n=512$) を設けて分析した結果を図 4 に示す。音声に対する聞き手のうなずく動作の同期現象は認められないが、話し手は聞き手のうなずく動作に対応して音声

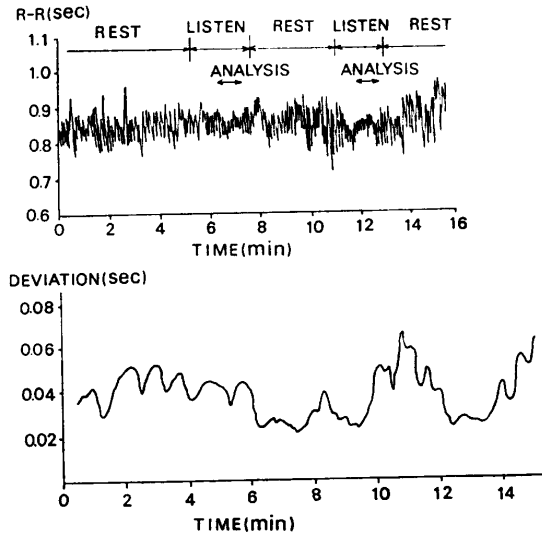


図 5 R-R 間隔とその偏差 (過去 40 個の R-R 間隔) の時系列変化
Fig. 5 Time change in R-R interval and its deviation (data 40).

を発生していることがわかる。

このように、図 2 に示したエントレノグラムが絶えず発現するわけではない。この理由として、聞き手の内部状態の影響が考えられる。ここでは、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考えられるので、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。

一般に注意が集中しているときは、R-R 間隔(心拍間隔)の分散が減少することが認められている⁸⁾⁻¹⁰⁾。実験時の聞き手の R-R 間隔とその偏差(過去 40 個の R-R 間隔)の時系列的変化の例を図 5 に示す。会話時には安静時に比べて、R-R 間隔はあまり変わらないが、その分散が小さくなり、聞き手が会話に注意を集中していることがわかる。本研究では、この R-R 間隔の分散が最も小さくなった区間についてエントレインメント分析を行った。このように、R-R 間隔の分散をリアルタイムでモニタすることにより、実験条件の安定化を図ることが可能である。

また分析例 22 例のうち、典型的なエントレノグラムを示した分析例 11 例について、 $C(\tau)$ の極大値での τ の関係 (音声に対するうなずく動作の時間遅れと、うなずく動作に対する音声の時間遅れとの関係) を図 6 に示す。 $\tau < 0$ の順域では、 $-1.4 \sim -0.7$ 秒の範囲

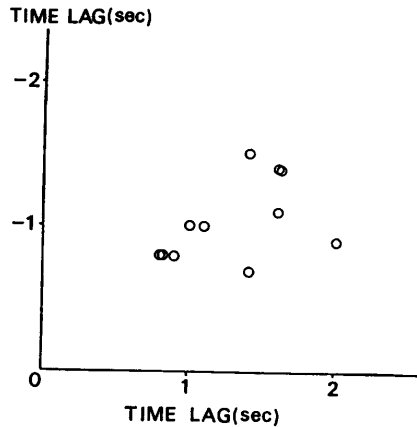


図6 典型的なエントレノグラムにおける $C(\tau)$ の極大値での τ の関係
 Fig. 6 Relation of τ at peaks in representative entrainograms.

に, $\tau > 0$ の領域では, 0.8~2.0 秒の範囲に入っている. 話し手と聞き手は, 互いにこの範囲のタイミングで同期して, 情報交換を円滑に進めていることがわかる.

5. 官能検査によるエントレインメントの評価

5.1 実験方法

ここでは, エントレノグラムの分析結果を官能検査(人間の感覚による評価・判定)による判定結果と比較検討する. 評価方法としては, 本実験から複数種類の成人の会話のビデオテープを選定し, 任意の2種類の成人の会話について, どちらがより円滑に情報交換しているかを比較対照して判定する, 一対比較法を採用した. エントレインメントの良否を定量的に決定す

表3 対比較判定順序

Table 3 Order of paired comparisons.

NO.	COMPARISON
1	① - ④
2	③ - ②
3	③ - ④
4	① - ②
5	③ - ①
6	④ - ②

るために, 以下の Bradley-Terry モデル¹¹⁾

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

π_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

を想定し, 対比較による順序づけを行った. 被験者に見せる判定用のビデオテープとしては, 話し手 S1, S3 と聞き手 L8 の会話の分析区間(約 34 秒間) 4 種類を選定した. 4 種類の分析区間のエントレノグラムを図7の①~④に示す.

①は, $\tau > 0$ の領域に極大値が存在し, 話し手が同期する現象が見られるが, 有意な相関の存在は示されていない. ②は, $\tau < 0$ の領域にのみ極大値が存在し, 聞き手が同期し, 話し手が同期する現象は見られないが, 最大の $C(\tau)$ 値が 0.1 で有意な相関の存在が示されている. ③は, 典型的なエントレノグラムである. ④は, 典型的な無相関のエントレノグラムである.

実験は, 4 種類のすべての対 ($4C_2 = 6$ 通り) について, 表3の順序で編集したビデオテープを再生して被験者に見せ, 1 対のうち, どちらがより円滑に情報交換しているかを二者択一させた. 1 対の比較判定時間は約 70 秒間であり, すべての対の判定時間は約 8 分

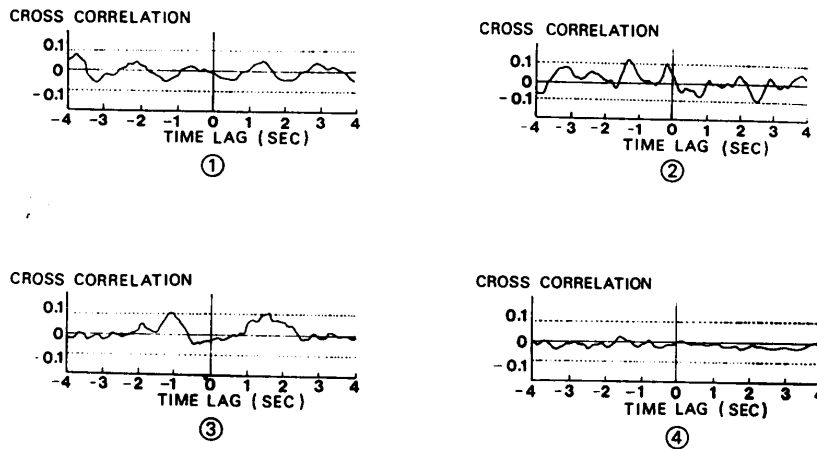


図7 官能検査用に選定した4種類の分析区間のエントレノグラム
 Fig. 7 Four kinds of entrainograms for sensory evaluation.

表 4 判定結果
Table 4 Results of sensory evaluation.

i \ j	①	②	③	④	TOTAL	π
①		6	7	18	31	0.129
②	20		15	24	59	0.477
③	19	11		22	52	0.339
④	8	2	4		14	0.055

間である。被験者は視覚・聴覚ともに正常な21~24歳の男女26人である。

5.2 実験結果

被験者26人による判定結果と Bradley-Terry モデルを想定した場合の π を最尤推定した結果を表4に示す。モデルの全体としての整合性を、有意水準5%で尤度比検定した結果、モデルが成立し、特定な苦手関係はなかった。次に、四つの π 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5%で表5に示すような(2,3)、(1,4)の二つのグループに分けられた。したがって、エントレノグラムによる評価と官能検査による評価とが一致しており、本論文で提案したエントレインメントの分析手法が有効であることがわかる。

6. むすび

人間にとって、言語等の記号情報と体動等の記号的情報行動とみなしうるものが何らかの相関を有しているという事実は、きわめて興味深いものである。本論文では、従来定性的研究の色彩が強い、このコミュニケーションにおけるエントレインメントを客観的に定量化する分析手法を提案し、成人同士の会話におけるエントレインメントを分析評価した。その分析の結果、コミュニケーションにおいて、話し手の音声に対する聞き手のうなずく動作の相互相関係数上の時間遅れは、0.7秒から1.4秒の範囲に、また聞き手のうなずく動作に対する話し手の音声の相互相関係数上の時間遅れは、0.8秒から2.0秒の範囲に、話し手と聞き手が相互に同期化するエントレインメントが存在することが明らかになった。さらに、官能検査を通して、本分析手法の有効性を検証した。

本論文で得られた知見が知能ロボットや CAI システムに生かされるならば、人間と機械の円滑な情報交換が図られ、人間側に立脚したシステムの実現に役立つと期待される。

表 5 順序のグループ分け
Table 5 Classified order.

i \ j	(2,3)	(1,4)	TOTAL	π
(2,3)		85	85	0.817
(1,4)	19		19	0.183

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。

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) Kato, T. et al.: A Computer Analysis of Infants Movements Synchronized with Adult Speech, *Pediatr. Res.*, Vol. 18, No. 8, pp. 625-628 (1983).
- 4) Watanabe, T.: A Study of Motion-Voice Synchronization, *Bull. JSME*, Vol. 26, No. 222, pp. 2244-2250 (1983).
- 5) 渡辺富夫: 動作と音声の同期に関する研究, 日本機械学会論文集C編, Vol. 49, No. 440, pp. 595-602 (1983).
- 6) 渡辺富夫: 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, 情報処理学会論文誌, Vol. 25, No. 2, pp. 251-259 (1984).
- 7) 渡辺富夫: 成人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, 情報処理学会第28回全国大会講演論文集, pp. 1555-1556 (1984).
- 8) Luczak, H. et al.: An Analysis of Heart Rate Variability, *Ergonomics*, Vol. 16, No. 1, pp. 85-97 (1973).
- 9) Sekiguchi, C.: Frequency Analysis of Heart Rate Variability under Flight Conditions, *Aviat. Space Environ. Med.*, Vol. 50, No. 6, pp. 625-634 (1979).
- 10) 渡辺富夫: 画像と音の同期に関する研究, 日本機械学会論文集C編, Vol. 49, No. 446, pp. 1728-1734 (1983).
- 11) Bradley, R. A. and Terry, M. E.: The Method of Paired Comparisons, *Biometrika*, Vol. 39, pp. 324-345 (1952).

(昭和59年4月25日受付)

(昭和59年9月20日採録)