ヒューマンインタフェース 65－6
音声言語情報処理 10－6
（1996. 2. 29）

音声対話システムにおけるヒューマン・インタフェース
― 引き込みを中心として ―

渡辺富夫
watanabe@cse.oka-pu.ac.jp
岡山県立大学情報工学部情報システム工学科
〒719-11 岡山県総社市窪木１１１

本報告では、まずヒューマン・フレンドリーな音声対話システムを実現するのに重要と考えられるノンバーバル情報を整理している。次にノンバーバル情報のヒューマン・インタフェースへの応用として、コミュニケーションにおける引き込み現象と韻律情報適応化に着目した研究を報告している。引き込み現象の研究は、対話者相互の引き込み原理に基づく音声対話システムの開発を目指したもので、話し手の音声と聞き手のうなずき・表情反応との引き込み現象だけでなく、対話者相互の生理的・心理的側面での引き込み現象について検討している。韻律情報適応化の研究は、入力音声の韻律情報に基づいて音声入力者個人に最適な韻律情報で応答する聴き易い音声出力システムの開発を目指したものである。両システムを結合すれば、話し易く、聴き易い音声入出力システムが実現され、人間-機械間の音声対話の円滑化が可能となる。

Human Interface for Improved Human-Computer Interaction
— Entrainment in Communication —

Tomio Watanabe

Faculty of Computer Science and System Engineering, Okayama Prefectural University
111 Kuboki, Soja, Okayama, 719-11 JAPAN

The entrainment which forms the biological relation between the speaker and the listener, plays an important role in the smooth exchange of information. In this report, first nonverbal means of communication in face-to-face interaction is summarized from the viewpoint of developing more user-friendly computer systems. Then the entrainment in communication as the representative nonverbal means is focused to make the human-to-computer speech input environment seem warmer, more interpersonal, and more natural for the speaker. Finally, effects of prosodic adaptation on human-computer verbal communication are discussed with the object of developing a machine which adapts to the optimal speech characteristics of the speaker. Combined the entrainment system with the prosodic adaptation system, a speech input/output system for improved human-computer dialogues would be realized.

1．　はじめに

通常の人間同士の音声対話では、ことばによる情報、すなわちバーバル情報によりコミュニケーションしていると思われがちである。しかし、ことばを発声して音声情報として伝達するには、ことばそれ自体だけでなく、必然的にピッチや抑揚などの韻律情報が付随することになり、コミュニケーションを進める上でことば以上に重要な役割を果たしている。この韻律情報のように音声情報からことばの属性を切り放した言語を周辺言語と呼び、バーバル情報に対して「ことばによらない」情報という意味でノンバーバル情報と呼んでいる。また電話でのコミュニケーションに比べて、直に対面してのコミュニケーションでは伝わる情報の円滑さも理解の度合いも異なるのは、日常よく経験することである。これは、ノンバーバル情報として周辺言語以外にも表情、身振り・手振りといった身体動作による情報、対話者同士の距離、関係、対話者相互に同期化する引き込み現象などの場の情報がバーバル情報では伝達不可能な情報の伝達を可能にしているからである。確かに人間社会システムの出発点とも言うべき乳児と育児者とのコミュニケーションにおいては、ノンバーバルなインタラクションによって成立している。したがって、このノンバーバルなインタラクションのメカニズムが人間機械系に導入されるならば、真に人間に優しいヒューマン・インタフェースが実現できるものと期待される。

本報告では、まずヒューマン・フレンドリーな音声対話システムを実現するのに重要と考えられるノンバーバル情報を整理している。次にノンバーバル情報のヒューマン・インタフェースへの応用として、対面コミュニケーションにおける引き込み現象と韻律情報適応化について検討している。

2．コミュニケーションにおけるノンバーバル情報

ノンバーバル情報の特徴は、バーバル情報の補助的、相補的役割を担うことはもちろんであるが、感性情報が直接伝わること、逆に場合によっては明示的表現がさけられること、マルチモードで同時に双方向の情報交換が可能であること、かつモードの選択をはじめ、種々の約束事を意識せずにすむこと、場の情報のように関係が成立する等である。ここでは表1に示すようにコミュニケーションにおけるノンバーバル情報を周辺言語、身体動作、場の情報に分けて解説する。

表1 コミュニケーションにおけるノンバーバル情報

分類	例　　[機能]
周辺言語	韻律情報 ピッチ構造（声の高さ、イントネーション） 時間構造　（テンポ、リズム） 振幅構造　（アクセント、ストレス） 声質 非言語的音声 [感性情報伝達]
身体動作	表情（顔色） 視線、まばたき、瞳孔 うなずき 身振り・手振り 口唇 姿勢（構え） [表象、例示子、情感表示、調整子、適応子]
場	引き込み現象 対人距離、空間、接触 対人関係など [関係の成立]

2.1　周辺言語

音声情報からことばの属性を除いたものは、すべて周辺言語である。音声を特徴づけるピッチ、抑揚、ストレスなどの韻律情報がその代表で、感情の表現、強調、会話の調整など、広く感性情報の伝達に不可欠である。韻律情報は、ピッチ構造、時間構造、振幅構造からなり、音声対話システムにおける韻律情報の個人への適応化においては、とくにON-OFFパターンや間など時間構造の適応化が有効である（4章参照）。また調音器官により決定される声質は、音声メッセージの内容とは無関係で、個人への依存性が高く、個人、性の同定だけでなく、その人柄の推定など、種々の印象の手がかりになっている。

相づちなどの会話を制御する非言語的音声(vocal segregate)も周辺言語である。その機能として受託、納得、疑問、驚き、拒否それにスピーチの継続機能が選定され、非言語的音声のヒューマン・インタフェースの導入可能性が検討されている[1]。

2.2　身体動作

表情、身振り・手振り、姿勢など、ノンバーバル

情報を表現する身体の動きが身体動作である。バーバル情報が交互のメーセージ交換を前提にしているのに対し、身体動作は同時に双方向のメーセージ交換が可能であることが特徴で、この同時情報交換性が会話の制御、理解に重要な役割を果たしている。P.Ekmanは身体動作を機能に基づいて表象、例示子、情感表示、調整子、適応子に分類している[2]。これら5種類の機能分類に基づき、ノンバーバル・インタフェース設計の観点から、身体動作によるコミュニケーションモードを機能的にサイン、指示、例示、操作、情感表示、調整、関心に分類し、身振りを手の形状、全身形態、運動に分けてコンピュータで処理するコーディング法、身振り表示、身振り辞書等、身振りインタフェースの開発が進められている[3]。

2.3 場

場の雰囲気などコミュニケーションでの場所的状況をいう。場の情報は、対話者との関係の中に成立する情報で、対面コミュニケーションにおいて対話者相互に音声と動作・表情が同期する引き込み現象などは、その代表である。対話者の生体リズムが相互に引き込むことで、関係が成立し、円滑なインタラクションが図られる。この引き込み現象の成否に関連する対人距離、空間、接触、対人関係なども場の情報に含まれる。身体動作、周辺言語も場の情報を生成する要因ではあるが、それらは対話者を独立に分析して情報抽出可能であり、対話者とのインタラクション、関係を切り放しては分析不可能な場の情報とは区別して分類されている。

場の情報をヒューマン・インタフェース設計に応用するには、これまでの身体動作や周辺言語を独立して扱うのでは不十分で、例えば、聞き手のうなずきなどの身体動作の調整子と話し手の周辺言語とが同期化する引き込み現象として、その関係として捉えることが大切である。情報機械との円滑なインタラクションには、この引き込み現象の存在は不可欠であり、次の章で紹介するように引き込み現象のヒューマン・インタフェースへの導入が試みられている。

3．コミュニケーションにおける引き込み現象

人間同士の対面コミュニケーションにおいては、対話者相互に音声と動作・表情が同期化する引き込み現象が存在し、円滑なコミュニケーションに重要な役割を果たしている。生体リズム間で引き込みが生起することで、関係が形成され、これが人間生物学的コミュニケーションの本質と考えられる。したがって、この引き込み現象のメカニズムが人間-機械系に導入されるならば、人間と情報機械との円滑なコミュニケーションが図られ、人間に適合したヒューマン・インタフェースの実現に役立つと期待される。

著者は、既に対話者相互の引き込み現象として、とくに話し手の音声と聞き手のうなずき反応、表情との引き込み現象を分析評価し、システム論的にモデル化して、音声対話の円滑化を図る引き込み化システムの開発を進めている[4]-[6]。

ここでは、これまでの音声・画像分析による視聴覚情報に基づく引き込み現象だけでなく、対話者相互の生理的・心理的側面での引き込み現象について、心拍間隔変動の生理指標に基づき分析評価している。

3.1 意識的なうなずきの引き込み現象

学生の聞き手2人が教員の話し手に合わせて意識的にうなずいた場合の話し手の音声と聞き手のうなずきの引き込み現象の典型例を図1に示す。音声V(t)もうなずきM(t)もビデオフレーム（1/30秒）ごとにその有無が判定され、ON-OFFパターンとして処理されている。これより意識的にうなずいた場合には個人差が少ないことがわかる。図1(c)は両者の相互相関関数である。うなずきに対する音声のずれ時間τが-1.2秒で有意なピークを持ち、うなずきが音声に対し1.2秒遅れて聞き手が合わせている。またτが0.3秒での負のピークは、聞き手が話の区切りを予測し、話の区切りの前にうなずき反応が開始されいることに起因するものである。したがって、円滑なコミュニケーションでは話が区切れてからうなずきが開始されたのでは既に遅く、この予測反応の引き込みが円滑なコミュニケーションに重要な役割を果たしていると考えられる。

この結果から、うなずき反応の存在区間を過去の音声の呼気段落区分でのON-OFFパターンに基づき推定するマクロ層のMA(Moving Average)モデルと、その区間内でのうなずき開始時点を1/30秒毎の音声時系列の線形結合で推定するミクロ層のMAモデルからなる音声-うなずき反応モデルを提案し、その有効性を示した[4]。

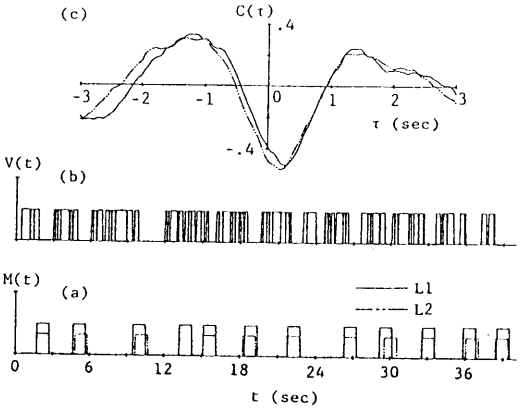

図1　話し手の音声と2人の聞き手(L1,L2)のうなずきの引き込み現象の評価
(a) うなずきM(t)の時系列、(b) 音声V(t)の時系列
(c) ずれ時間τでの相互相関関数C(τ)

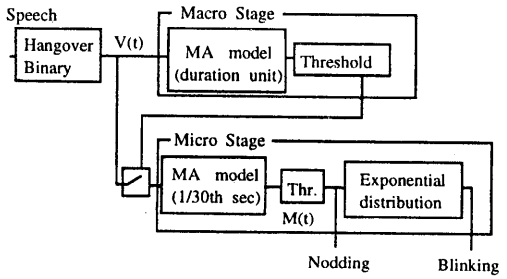

図2　音声-うなずき・まばたき反応モデル

また、対面コミュニケーションの観点から対話の円滑化を図る聞き手の表情を分析し、6種類（受諾、納得、疑問、驚き、拒否、無視）の基本的な表情（動作）を選定した。これら6種類の表情アニメーションを抽象的でアニメティックな顔線画とリアルな人物顔画像で構築し、音声入力の円滑化を図るビジュアルフィードバックとしての表情アニメーションの有効性を官能検査により示した[5]。

表情アニメーションの合成的解析法による効果的な引き込み現象の誘導法を確立する過程において、表情アニメーションの反応がビジュアルフィードバックとして静的な印象を与えることがわかり、まばたき反応と表情のゆらぎに着目して再検討した。その結果、対話時のまばたき反応は音声のON区間では抑制され、OFF区間に生起することが相対的に多いことが示され、音声時系列とまばたき反応とは有意な相関があること、またうなずいている時にはまばたきすることが多いことなど、音声とうなずき・表情反応との引き込み現象だけでなく、まばたき反応との引き込み現象の相互関連性が判明した。これらの分析結果を基に、図2に示すように音声-うなずき反応モデルの階層モデルに指数分布モデルを組み込んだ音声-まばたき反応モデルを提案し、シミュレーション実験によりその有効性を示した[6]。

3.2　生理的側面での引き込み現象

音声対話における対話者相互の情動（内部状態）を客観的にかつ定量的に計測するには、生体が示す生理的反応を利用することが考えられる。ここでは、そのような生体情報として心拍間隔（R-R間隔）の時系列変化、とくにそのばらつきに着目した。心拍間隔のばらつきと情動との関係については、心拍間隔のばらつきが大きければリラックス状態を、小さければ緊張状態を表すことが知られている[7]。

学生の対話者相互について約30秒間の心拍間隔の標準偏差の時系列変化を180秒間について計測した結果の一例を図3に示す。これは、笑い等のリラックス状態から話の集中状態まで、情動の変動が顕著な箇所であり、引き込み現象がみられる。

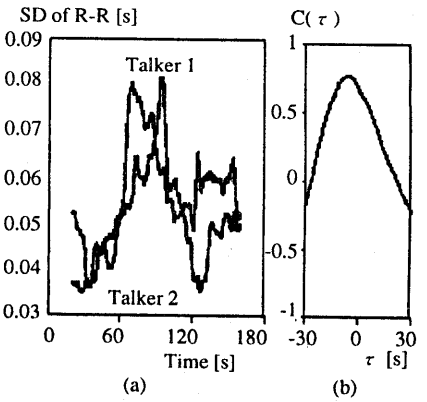

図3　成人間の対話時におきる心拍間隔の標準偏差の時系列変化　(a) 30秒毎の標準偏差の時系列
(b) ずれ時間τでの相互相関関数C(τ)

また母親と乳児（5カ月）との対面コミュニケーションにおいて、約30秒間の心拍間隔の標準偏差の時系列変化が180秒間にわたり相互に同調する結果の一例を図4に示す。さらに図5に示すように、乳児が覚醒状態から睡眠状態に移行する過程において、母子共に心拍間隔変動の周期性（呼吸成分）が

検出され、迷走神経系支配への移行過程の情動面での引き込み現象が観察された。

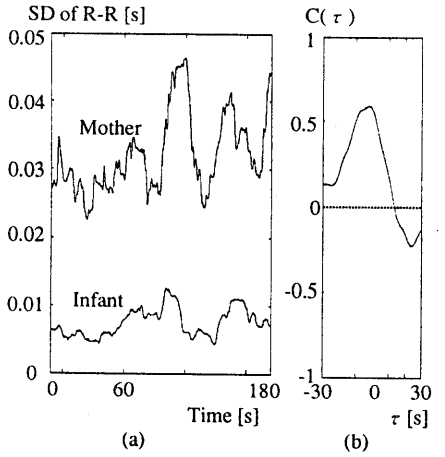

図4　母親と乳児の対話時におきる心拍間隔の標準偏差の時系列変化　(a) 30秒毎の標準偏差の時系列 (b) ずれ時間τでの相互相関関数C(τ)

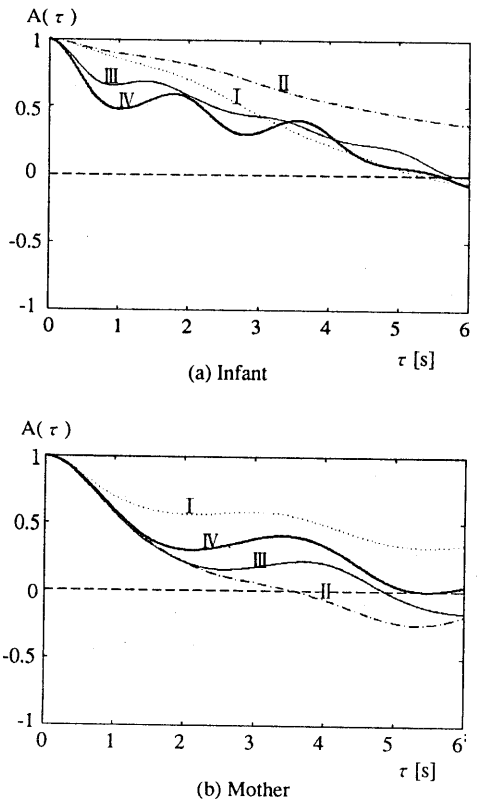

図5　母親と乳児の迷走神経系支配への移行過程における心拍間隔の自己相関関数A(τ)の一分毎の変化

これらの結果は、円滑なコミュニケーションが図られるには、こういった情動まで含めた生理的側面での引き込み現象が生じていることを示唆するものである。

4．韻律情報適応化

人間-機械間の音声対話において、人間に聴き易い音声の手がかりはその人自身の発話する周辺言語、とくに韻律情報に含まれると考えられる。ここでは、音声対話システムにおける韻律情報適応化の有効性について時間率と基本周期に着目して検討する。

4.1 時間率適応化

著者はこれまでに個人の発話速度を時間率（speech activity）［平均ON区間／（平均ON区間＋平均OFF区間）］に基づいて推定し、機械応答音声のOFF区間を伸縮してその音声入力された時間率に一致させることにより、音声応答速度を適応化させる手法を提案した[8]。

応答音声と入力音声の呼気段落区分（フィルイン130 msec: 130msec以下のOFF区間をON区間に置換）での時間率が各々$\alpha 1 = T1/(T1+S1)$と$\alpha 2 = T2/(T2+S2)$のとき、両者の時間率が一致するように、

$$T1/(T1+S') = T2/(T2+S2)$$
$$S' = T1\,S2/T2 \qquad (1)$$

T1:応答音声の平均ON区間
S1:応答音声の平均OFF区間
T2:入力音声の平均ON区間
S2:入力音声の平均OFF区間
S':適応化後の応答音声の平均OFF区間

S1をT1S2／T2に伸縮することによって、即ち、フィルイン130msecを施した（呼気段落区分での）各OFF区間を　(T1S2)／(T2S1)の比率で伸縮すれば適応化が図られる。

応答音声を個人の時間率に一致させた場合と発話速度に一致させた場合について対話の円滑化を官能検査により比較検討した結果、時間率が平均値以下の場合には、個人の時間率そのものへの適応化が発話速度への適応化よりも有効であることが明らかになった[9]。

被験者46人の録音音声について、OFF区間の回数iまでの時間率αiの時系列的推移を分性した結果、αiはi＝8以後では最終の時間率に近い値で安

定していることが判明した。この知見に基づいて、予め録音されている音声を個人の時間率に適応化させる、音声対話のための時間率適応化システムを開発した。本システムは入力音声に対して8回のOFF区間を含むON区間の終了時点で時間率を計算し、その時間率が録音音声の時間率(0.64)以下ならば時間率適応化を図り、録音音声の時間率以上ならばそのままの録音音声を、入力音声終了後に音声出力するシステムである。

4.2 基本周期適応化

ここではピッチ構造の中のとくに基本周期の適応化に着目し、評価検討した。聴き手の発話時基本周期と聴き易い音声の基本周期との関係について、音声の基本周期を各種変化させた音声資料の一対比較による官能検査により評価した結果、個人の基本周期と聴き易い音声の基本周期との間には有意な相関関係はみられなかった。しかし、音声の基本周期が平均基本周期から標準偏差の2倍の範囲を越えている被験者について適応化の有効性を検討した結果、高音の被験者は個人の基本周期より平均基本周期が、低音の被験者は平均基本周期以上の基本周期が聴き易い傾向がみられた[10]。従って、音声対話システムにおける韻律情報適応化においては、個人への時間率適応化が有効であり、ピッチ構造については機械側が平均的な基本周期の音声に設定してあれば十分であることを明らかにした。

5．おわりに

本報告では、人間同士の対面コミュニケーションにおけるインタラクションをヒューマン・インタフェース設計に生かす上で不可欠なノンバーバル情報を整理するとともに、ノンバーバル情報のヒューマン・インタフェースへの応用について、引き込み現象と韻律情報適応化を中心に紹介した。これらの特徴を生かしたヒューマン・インタフェースが構築されるならば、人間と情報機械との自然なコミュニケーションが図られると期待される。とくに種々の約束事を意識せずに、引き込み現象のように関係が成立するノンバーバルなインタラクションは、異文化、異民族間のコミュニケーションにおいても通用する共通語ともいえ、人間生物学的に本質的なコミュニケーションであると考えられる。このインタラクションを成立させる基本原理の解明、すなわち関係を成立させる場の創出原理の解明が次世代ヒューマン・インタフェースの課題である。

参考文献

[1] Avons, S.E.et al.: "Paralanguage and Human-Computer Interaction", Behavior & Information Technology,Vol.8, pp.13-21 (1989).

[2]Ekman, P.:"Three Classes of nonverbal behavior", Aspects of Nonverbal Communication, ed. Swets and Zeitlinger（本名信行他訳、ノンバーバル・コミュニケーション, pp.3-26, 大修館書店）(1989).

[3] 黒川隆夫:ノンバーバルインタフェース、オーム社 (1994).

[4] Watanabe, T. and Yuuki, N.:"A Voice Reaction System with a Visualized Response Equivalent to Nodding", Advances in Human Factors/Ergonomics, Vol.12A , pp.396-403 (1989).

[5] Watanabe, T. and Higuchi, A.: "Facial Expression Graphics Feedback for Improving the Smoothness of Human Speech Input to Computers", Advances in Human Factors/Ergonomics, Vol.18A, pp.491-497 (1991).

[6] Watanabe, T. : "Voice-Responsive Eye-Blinking Feedback for Improved Human-to-Machine Speech Input", Advances in Human Factors/Ergonomics, Vol.19B, pp.1091-1096 (1993).

[7] 渡辺富夫: "画像と音の同期に関する研究"、日本機会学会論文集, 50巻, pp.1728-1734 (1983).

[8] Watanabe, T.:"The Adaptation of Machine Conversational Speed to Speaker Utterance Speed in Human-Machine Communication", IEEE Trans. of Systems, Man, and Cybernetics, Vol.20, pp.502-507 (1990).

[9] 渡辺富夫:"人間-機械間の音声対話のための時間率適応化システム"、計測自動制御学会論文集, 26巻, pp.902-907 (1990).

[10] Watanabe, T.: "Effects of Pitch Adaptation in Prosody on Human-Machine Verbal Communication", Advances in Human Factors/Ergonomics, Vol.20A, pp.269-274 (1995).