音声言語情報処理 21－7
（1998. 5. 29）

人間とロボットのコミュニケーションにおける非言語情報の利用

横山真男　青山一美　菊池英明　白井克彦
早稲田大学理工学部
東京都新宿区大久保 3-4-1
kazumi@shirai.info.waseda.ac.jp

あらまし　人間型ロボットのコミュニケーション能力を人間のそれに近付ける為に、人間同士のコミュニケーションにおいて重要な役割を持つ非言語情報の利用を検討している。本稿では、まず人間同士の対話において、各種非言語情報の出現タイミングについての分析を行なう。さらにその結果を踏まえて、ロボット側の非言語情報の出力タイミングによる対話への影響を分析する。分析の結果、非言語情報の種類による発話交替における制約としての強さや自然性の違いが明確になった。さらに非言語情報の出力タイミングとして、人間同士と同様に発話開始直後あるいは終了時が自然かつ円滑な対話の実現にとって適切であることが確かめられた。

キーワード　非言語情報、ロボット、音声対話、発話交替、マルチモーダル

Use of Non-verbal Information in Communication between Human and Robot

Masao Yokoyama　Kazumi Aoyama　Hideaki Kikuchi　Katsuhiko Shirai
School of Science and Engineering, Waseda University
3-4-1, Okubo, Shinjuku-ku, Tokyo, 169-8555, JAPAN
kazumi@shirai.info.waseda.ac.jp

Abstract　In this research, we consider the use of non-verbal information in dialogue between human and robot to draw the communication ability of the robot to the one of human beings at all and near. This paper describes analysis of output timing of non-verbal information in the dialogues between human beings. Moreover, we analyze influence on dialogues by controlling output timing of non-verbal on the CG robot. As the result, we clarified the strength of constraints and naturalness in speaker-changes. Also, we confirmed that simultaneous with or end of utterance is appropriate for output timing of non-verbal information as well as on model of human beings.

Key Words　Non-verbal Information, Robot, Spoken Dialogue, Speaker Change, Multimodal

1 はじめに

近年、高齢化社会に向けて福祉ロボットへの期待が高まっており、ロボットが人間の補助を行ない、人間と協同で、時には人間に代わって作業をするような関係の実現が望まれている。加藤 [1] はこうした「人にやさしいロボット」とは、第2次産業の分野で発展してきた生産効率向上のための産業用ロボットではなく、人間に近い容姿・能力を備えており、人間と協調し、安心して使える行動主体であるべきと指摘している。

そのような人間型のロボット（ヒューマノイド）を実現する際、そのコミュニケーション能力において多くの課題がある。そのうちの一つとして、ユーザとなる不特定多数の人間にとって共通であり従って負担の少ないユーザインタフェースとして、人間が持つコミュニケーション能力をロボットに実装することが要求される。その場合、人間の能力を全て実装することは困難であり、現状ではそのうちの音声のみをコミュニケーションの手段として実装するのが一般的である。しかし、ロボットとのコミュニケーションにおいては、実空間を主に視覚によって共有しているため、音声のみでは不自然なものとなる恐れがある。

従って、ロボットのコミュニケーション能力を少しでも人間のコミュニケーション能力に近付けるため、音声に伴う非言語情報の利用が考えられる。人間同士のコミュニケーションにおける「非言語情報」には、ジェスチャーなどの身体動作の他、姿勢、対人距離、服装・装飾等の外見、身体的特徴など、幅広い概念が含まれており、人間のメッセージ伝達の 65%[2] あるいは 93%[3] が非言語情報で占められているとも言われる。ここでは、非言語情報のうち、ヒューマノイドに実装可能な身体動作に注目する。

人間の身体動作は Ekman[4] により表 1の様に分類されている。このうち、対象物を指差すような例示子は、発話内容と直接結びつく動作である。ユーザの意図が摑めずに再発話を要求する場合に人差し指を立てる動作などは、標識に分類され、コミュニケーションの促進に直接影響を与える。また、調整子は、話す順番を決めたり、会話の流れを円滑にする機能を持つ動作である。相手の発話を促すうなずきや視線一致などは調整子の例である。

これらの非言語情報をロボットの身体動作として適用することで、より人間に近いコミュニケーション能力を持たせることが可能になると考えられる。これまでに、標識と例示子に分類される身体動作をロボットの応答音声に加えた結果、ユーザの理解度や親和性を向上する効果が得られた [5]。本研究では、さらに調整子の利用を目指している。円滑なコミュニケーションの実現には、話者間の発話権の移動をスムーズにする必要があるが [6]、調整子はその役割を果たす重要な情報である。これまでに、人間同士の対話の分析により、うなずきなどの頭の動作が発話の番（以降、発話権と呼ぶ）と密接に関係していることが確認されている [7]。また、対話において、聴者が発話権を獲得しようとする場合には相手と視線を合わせ、そうでない場合には視線を合わせない傾向があることが報告されている [8]。

表 1: Ekman による身体動作の分類

標識 (emblem)	音声語句に翻訳可能で表象、サインとも呼ばれる。Vサイン、お金、など。
例示子 (illustrator)	発話の内容や流れと結び付き、発話内容を強調、精緻化、補足する。思考動作、指示動作 など
情感表示 (affect display)	情動に伴う表情や身振り。握り拳など。
調整子 (regulator)	発話権の授受を制御したり対話の流れを円滑にする動作。
適応子 (adaptor)	状況に適応するための動作。頭をかく、貧乏ゆすり、など。

うなずきや視線移動のみならず、調整子に分類される種々の非言語情報を対話インタフェースにおいて有効に利用するためには、人間同士のコミュニケーションにおける各非言語情報の特徴を捉え、人間の動作モデルをロボット制御に適用する方針が適当であると考える。

従って本研究では、まず人間同士の対話における各種の非言語情報について、主にその出現タイミングについての分析を行ない、その結果をふまえて、非言語情報の出力タイミングの違いによるユーザへの影響やコミュニケーション場への影響についての分析を行なう。

以下、2章では、人間同士の対話における非言語情報の出現タイミングの分析、3章では、制御方法の違いによるユーザおよび対話への影響の分析について述べる。

2 人間同士の対話における非言語情報の出現タイミング

ここでは、人間の身体動作を伴った発話の様子を収録し、人間が言語情報にともなってどのように非言語情報を相手に提示しているかを分析する。さらに非言語情報がどの出現タイミングによって機能しているのかを考察する。

2.1 対面対話データ収録

前述した調整子の役割である円滑な発話交替の実現に関係する状況として、システムがユーザに発話を促す（発話権譲渡）状況を設定する。被験者（大学生１０名）に、システム役としてユーザ役の別の被験者に発話を促すという設定で収録を行なった。その際、発話内容のバランスを整えるため、問いかけ→応答、といった対話の一部分を抜き出した形の発話対 [10] 単位を設定した。発話権譲渡の発話対として、表 2に示すような５種類１０文を用意し、それぞれ均等に発話を収録した。

表 2: ユーザに発話を促す発話対の例

発話対の種類	例
挨拶-挨拶	「どうも、こんにちは」-「こんにちは」
依頼-承諾／拒否	「えーと手を挙げて下さい」-「はい」/「いいえ」
呼びかけ-応答	「それでは、始めましょう」-「はい、始めましょう」
yes/no 質問-応答	「あなたは日本人ですか」-「いいえ」
属性質問-属性伝達	「あなたのお住いはどちらですか」-「＊＊＊です」

それぞれの被験者にカメラを向けＶＴＲに録画した。音声・画像データをワークステーションに取り込み、対話分析ツールを用いて言語情報および非言語情報出現の時間関係を分析した。非言語情報としては、調整子として機能すると考えられる視線、まばたき、うなずき、手振りに着目した。

2.2 非言語情報の出現タイミング

非言語情報の出現タイミングの傾向を調べる為に、各非言語情報の発話開始、キーワード発声開始、発話末尾からの相対的な出現時刻の分布を調べた。そのうち、最も傾向のはっきりした発話開始からの相対時刻の分布を図 1に示す。

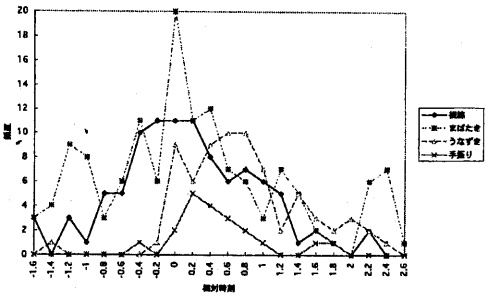

図 1: 発話開始時に対する非言語情報の出現頻度

図 1において、各非言語情報の出現タイミングとして、発話開始直後のピークが特徴的であることがわかる。さらに詳細に見ると、視線・まばたきに関しては発話開始と同時に出現することが最も多いが、うなずき・手振りに関しては発話開始時よりも発話開始後 0.5 秒程度の出現が多いことがわかる。これは、これらの非言語情報はいずれも調整子としての役割である円滑な発話交替の実現に関した機能を持つが、機能する適切なタイミングが種類により若干異なるためと考えられる。

以上より、ロボットに人間の動作モデルを適用する場合、ロボットは非言語情報を、種類およびその機能に応じて発話の開始前後の適切なタイミングで出力すれば良いと言える。

しかし、人間の動作モデルをロボットにそのまま適用して良いかどうかは疑問である。したがって、次章では、実際に被験者に対話システムと向きあって対話してもらい、どのようなインタフェースなら自然でスムーズに対話が出来るかの検証を行う。具体的には、ＣＧで表現されたロボットが出力する各非言語情報（以降、モダリティと呼ぶ）の種類およびタイミングを様々に変え、自然かつ円滑なコミュニケーションの実現において、どのタイミングでの出力が適切であるかを調べる。

3 非言語情報の制御とその効果

前章と同様に、発話権譲渡の状況を設定する。ＣＧロボットが相手に発話を促す合い図を表 3にあげる。ここでは、表3にあげた合い図のうち、6種類の非言語情報について比較評価を行なう。

表 3: ユーザに発話を促す合い図

モダリティ				利用例
非言語情報	人間にない モダリティ	ビープ音	聴覚	特定の音を短時間鳴らす
		図表示	視覚	口のイラストや文字の表示
	人間にある モダリティ	視線	視覚	視線を向ける
		まばたき	視覚	まぶたの開閉, うなずきの代用
		うなずき	視覚	縦振り
		手振り	視覚	手を差し出す
パラ言語		韻律	聴覚	イントネーション, ポーズ, 発話速度の強調
言語情報		発話内容	聴覚	疑問文, 確認文, あいづち

3.1 発話交替における制約

システムに対してユーザが自由なタイミングで発話できるということは対話インタフェースとして重要なことである。この発話タイミングが制限されると、ユーザの思考の妨げやストレスなどを引き起こす [9]。しかし、ユーザへの負担を減らすあまり、制限を弱くし過ぎると、円滑な発話交替がかえって行なわれなくなる危険も考えられる。ここではユーザに発話を促す場合の、ユーザの発話タイミングの自由度を「発話交替における制約」の強さであらわす。つまり、発話交替における制約が弱いインタフェースとは、ユーザの自由度が高くユーザへの負担も少なく、逆に制約が強いインタフェースでは、ユーザは発話のタイミングを限定されるということである。

以下に述べる実験では、各非言語情報の比較評価に、ポーズ長（システム発話終了からユーザ発話開始までの時間）を用いる。一概に、ポーズ長が短い方が良いとは言えないが、ポーズ長が長いということは、ユーザにとって発話すべきタイミングがわからず不必要に長い無音区間が生じた結果であると考えられ、コミュニケーションにおいて非効率であると言える。従って、このポーズ長の長短は制約の強弱に関係すると思われる。

3.2 非言語情報の制御方法

次に、実験システムと実験手順について説明する。実際のロボットに様々な制御方法をすべて適用して比較評価を行うのは困難であるので、ここではＣＧによるシミュレーションロボット (DoraeMan) を用いて実験を行った。

実験システムは、３次元ＣＧロボット画像表示と音声入出力を統合したマルチモーダルインタフェースを備えている。システム構成図を図 2に示す。

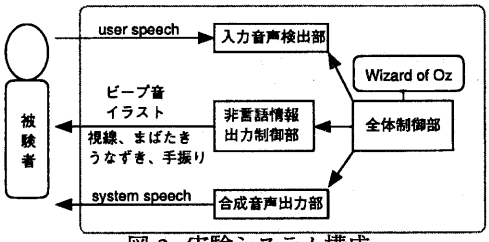

図 2: 実験システム構成

実験では、非言語情報の種類（図 3）とＣＧロボットの出力合成音声に対する出力タイミング（図 4）の制御方法を選択できるようにした。

種類 ビープ音・イラスト・視線・まばたき・うなずき・手振り

出力タイミング 発話開始・キーワード[1]・発話末尾・発話終了後（約 0.5 秒）・無し

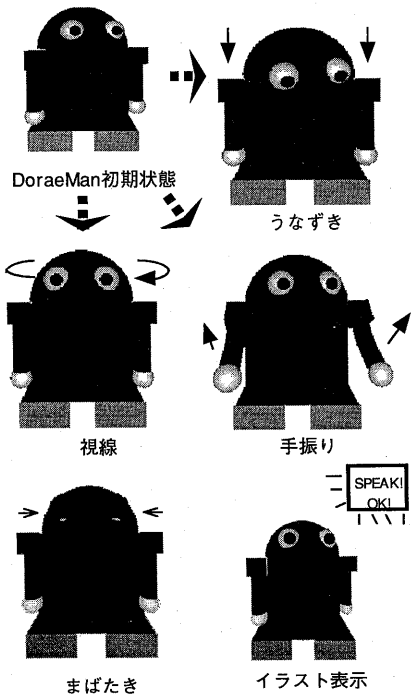

図 3: 表現する非言語情報の種類

[1]発話中のキーワードの発声と同時に非言語情報を出力させるものである。

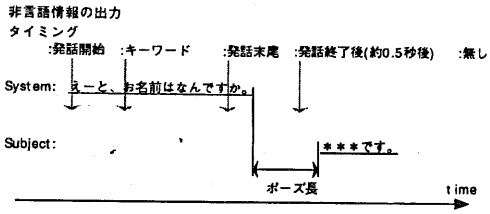

図 4: 発話と非言語情報の出力タイミング

実験では、大学生２０名の被験者が、上で述べたパラメータを組み合わせて、非言語情報の種類６通りと出力タイミング５通りの組み合せ計３０通りを比較した。

なお、対話収録にあたって、発話権譲渡の発話対として、表2に示した５種類１０文についてそれぞれ均等に１回ずつ対話を収録した。

さらに、被験者にはアンケートにより制御方法の自然さや返答のしやすさなどを評価してもらった。１種類のモダリティにつき５パターンのタイミングをランダムで出力し（１種類につき５回行い、これを１セットとする)、１セット終了後、その種類についての評価をしてもらった。また、種類とタイミングを変えて出力をしていることだけは被験者に伝えているが、次にどのタイミングで出力されるかは知らせていない。ロボットの問い掛けに対する被験者の返答内容は自由にした。

3.3 実験結果および考察

3.3.1 ポーズ長による制約の強さに関する比較

非言語情報の種類および出力タイミングによるポーズ長の変化を図5に示す。非言語情報が無しの場合、全体的にポーズ長が増加する傾向にあり、不自然な沈黙が生じ易くなっていた。これは、そもそも被験者は、同じような発話対のやり取りを繰り返すことによる慣れから、いつもシステムに何らかの動作があると思いそれを待ったためと考えられる。

特に、ビープ音・イラストが無しの場合、他に比べポーズ長が長くなっていることから、これら人間にないモダリティであるビープ音・イラストは、システムの動作に人間が合わせその合い図に影響させられる度合が高く、制約が強いモダリティといえる。逆に、視線など人間にあるモダリティの場合は、その度合が小さいことから制約が弱いモダリティと考えられる。人間にあるモダリティの制約の強さは、手振り、まばたき、視線、うなずきの順となっている。

なお、アンケートにおいても、自然さの点で、ビープ音・イラストが他に比べ劣るという結果が得られ、それ以外は、まばたき、視線、手振り、うなずきという順に評価が高かった。

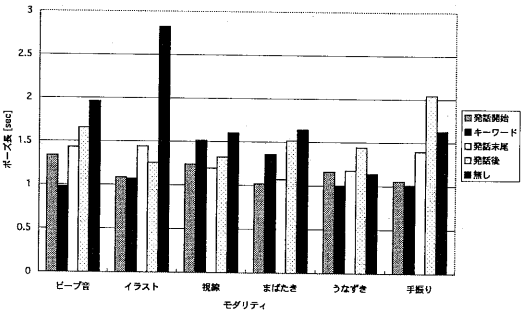

図 5: ポーズ長の変化

3.3.2 各モダリティの出力タイミング

図5において、視線・まばたきは、キーワード発声時よりも発話開始時あるいは発話末尾の方がポーズ長が短い。一方、手振り・うなずきはその逆となっている。これは、2.2節で示した、人間の動作モデルにおける非言語情報の種類による出現タイミングの違いの傾向と同一である。

アンケート結果からも、若干であるが発話開始からキーワードあたりでの早い段階で出力した場合の評価が良く、人間の動作モデルをシステムに適用することが有効であると言える。

また、ポーズ長での評価、アンケート評価ともに、発話後に出力した場合の評価が全体的に低かった。インタビューからも、システムの発話終了後の動作は発話の妨げになるという意見が得られた。

4 おわりに

本論文では、ロボットと人間の自然な対話の実現を目指し、非言語情報の利用、特に発話交替のスムーズさに大きく影響を与えるとされる調整子という身体動作に注目した。ロボットの動作における非言語情報の制御に向けて、実際の人間における非言語情報の出現タイミングの分析、ＣＧロボットにおける非言語情報出力方法の違いによるユーザに与える影響についての検討について述べた。

各々の非言語情報の評価のまとめを表4に示す。

非言語情報の出力方法については、各々の特徴が明らかになり、ビープ音・イラストなどはユーザにとって発話交替の制約が強く、視線など人間にあるモダリティは制約が弱いことが言え、一般に制約の弱いモダリティのほうが好まれることなどがわかった。人間にあるモダリティにおいて評価の良い出力タイミングが、人間の動作モデルにおける非言語情報の出現タイミングの傾向と一致し、人間の動作モデルをロボットの制御に適用することが有効であることを示した。

今後の課題として、本論文ではキーワード発声時刻との関連分析に留まっており、さらに発話内容や意図なども考慮に入れた制御方法の検討を行う必要がある。

表 4: 各モダリティの評価のまとめ

モダリティ	制約		評価
ビープ音	強い	聴覚	出力タイミングは発話後がよい システム発話と重なると聞きづらい（聴覚同士で組合せが悪い）
イラスト		視覚	出力タイミングは発話後がよい イラストに気をとられる どちらを見たらいいかわからない
視線など人間にあるモダリティ	弱い		出力タイミングは発話開始もしくはキーワード時がよい 自然、人間らしいなど

参考文献

[1] 加藤一郎: “人間ロボット論”, 日本ロボット学会誌, Vol.10 No.1, pp 76-79, 1992.

[2] Birdwhistell,R.L.: “Kinesics and Context”,Univ.of Pennsylvainia Press, 1970.

[3] Mehrabian,A., Williams,N.: “Non-verbal concomitants of perceived and intended persuasiveness” , J. of Personality and Social Psychol. 13, pp.37-58,1969.

[4] Ekman,P., Friesen,W.V. : “the repertoireof nonverbal behavior” semiotica, 1, pp.49-98, 1969

[5] 帆足啓一郎, 田中修一, 中里収, 白井克彦: “構内案内ロボットにおける音声とジェスチャーの統合に関する検討”, 日本音響学会春季講演論文集, 1-P-15, pp 191-192, 1996.

[6] H.Sacks, E.A.Schegloff, G.Jefferson, “A Simplest Systematics for the Organization of Turn-taking for Conversation”, Language, Vol. 50, No. 4, 1974.

[7] Y.Iwano, Y.Sugita, Y.Kasahara, S.Nakazato, and K.Shirai : “Difference in visual information between face to face and telephone dialogues.” Proc. of ICASSP97, vol.2, pp.1499-1502, 1997.

[8] Novick,D., Hansen,B., Ward,K.: “Coordinating turn-taking with Gaze”, Proceedings of ICSLP96, pp 1888-1891, 1996.

[9] 菊池英明, 工藤育男, 小林哲則, 白井克彦: “音声対話インタフェースにおける発話権管理による割り込みへの対処”, 電子情報通信学会論文誌, Vol.J77-D-II, No.8, pp 1502-1511, 1994.

[10] 竹下敦:“プラン認識に影響を与える対話現象”, 第4回人工知能学会全国大会論文集,pp.367-370,1990.