

人間—エージェント間における
「自然なコミュニケーション状態」の客観的指標
An objective indicator of natural communication
between a human and an artificial agent

小松 孝徳¹, 宇都宮 淳², 植田 一博², 岡 夏樹³
Takanori Komatsu, Atsushi Utsunomiya, Kazuhiro Ueda and Natsuki Oka

1. 公立はこだて未来大学, 2. 東京大学, 3. 京都工芸繊維大学
1. Future University Hakodate, 2. The University of Tokyo, 3. Kyoto Institute of Technology

Abstract: In this paper, at first, we introduced our recent study to construct a speech interface agent which can communicate smoothly with everyday users and can be a communication partner for them. To evaluate the interaction between the users and this agent model, e.g., whether the user regards this agent as a communication partner or not, we need to refer the objective indicator which reflects users' internal status accurately and reliably. However, there were no previous studies to propose such objective indicator so far. We then proposed and explained the experimental setting to acquire the objective indicator with exploiting the users' cognitive, behavioral and physiological indexes.

1. はじめに

これまで筆者らは、ユーザとの相互適応的な関係を通じて自然なコミュニケーションを成立させることのできる「適応的な音声インターフェイス」を構築することを目的とした研究を行ってきた[4,5,8]。ここで相互適応的な関係とは、ユーザがインターフェイスに対して適応学習し、その適応に基づいてインターフェイスもユーザに適応学習するというプロセスを繰り返すこと、つまり、両者が相手のことを学習し、その学習結果に応じて自分も相手に対して適応していくことを繰り返すといった「適応のループ」が実現されている関係のことである。このような関係は、例えば、新生児と母親、さらには飼い犬と飼い主といった「コミュニケーションを成立させることのできる二者間」において共通にみられる現象の一つであると考えられる。よって、自然なコミュニケーションを成立させることができることを望まれているユーザと適応的インターフェイスとの間

にも同様に存在すべき関係であると考えられる。そこで筆者らはその実現に向けた基礎技術として、ユーザの適応を促進し、そこで得られた適応を利用することで「適応のループ」をユーザとの間に構築し、ユーザの発話意図を理解するような意味獲得モデルの実現を目指している。

本稿の前半では、ユーザと相互適応的な関係を構築し、その発話の意味を理解するエージェントモデルの開発を行った筆者らの最新の研究結果について述べる。結論からいうと、筆者らの開発したエージェントモデルとユーザとの間には、本来の目的である相互適応的な関係は、完全な形としては観察されなかった。しかしながら、最低限の目的である「ユーザの発話の意味をモデルがインタラクティブに獲得すること」を達成するには十分なコミュニケーションが二者間に成立していたことが確認された。よって、二者間のコミュニケーションを評価するには、相互適応という行動的な側面のみに注目するのではなく、被験者の内部状態も加味した上で詳細な解析を行う必要が生じてきたといえる。特に本研究のような場合、「インタラクションの対象であるエージェントをどのように認識しているのか」という人間の内部状態を把握する必要があると考えられ、そのためには人間の内部状態を正確に反映した客観的指標を参照する必要がある。しかしながら、

小松 孝徳 公立はこだて未来大学 情報アーキテクチャ学科
〒041-8655 北海道函館市龜田中野町 116-2
Tel: 0138-34-6220, Fax: 0138-34-6301,
E-mail: komatsu@fun.ac.jp

このような客観的指標を提案した研究は筆者らの知る限り皆無である。

そこで、本稿の後半では、インラクションの対象である人工的エージェントを人間がどのように認識しているのか、特に「エージェントをコミュニケーションの適切な対象として認識しているのかどうか」を客観的に示す指標を獲得するための実験設定を提案する。将来的には、提案した実験設定から得られた客観的指標の妥当性についても議論を行う予定である。

なお、本文中における「コミュニケーションの対象」という表現は、様々な文脈で使用され得るため、曖昧かつ多義的に解釈されやすい。本稿では、「継続的にインラクションをしている人間にとって、強度のストレスや覚醒・興奮を与えない対象」という意味に限定してこの表現を使用する。

2. コミュニケーション実験

人間の適応能力を利用しながらその発話の意味を理解する適応的インターフェイスの構築に向けた基礎的知見を得るために、人間同士がコミュニケーションを成立する際に、相手に対してお互いがどのように適応しているのかを観察するコミュニケーション実験を行った。具体的には、相手が何かを話していることはわかるがその意味はわからないような状況、つまり与えられる発話中の音韻情報を利用できないような環境を設定し¹、そのような状況において話し手（ゲームの教示者）の発話意味を聞き手（ゲームの操作者）がどのように理解していくのかを観察した²（図1）。この実験設定は、例えるならば「スイカワリ」のような状況に良く似ている。目隠しをして棒を持ち、スイカにめがけて歩くのが「ゲームの操作者」、その周囲で操作者がどのように動けばよいのか声で指示を出す周りの観客が「ゲームの教示者」に相当する。ただし、与えられる教示音声はその音韻情報が削除されているため、操作者にとっては未知の言語のよろなものに相当する。

¹ ゲームの操作者には、ローパスフィルタを通すことによって音韻情報を削除した教示音声が与えられた。

² コミュニケーション実験に関する詳しい説明は参考文献[4]を参照されたい。

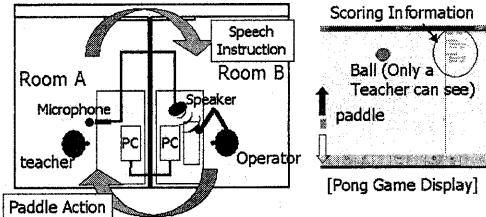


図1. コミュニケーション実験環境

この実験の結果、与えられる未知の音声の意味を獲得し、コミュニケーションを成立させることができた被験者ペアでは、お互いが相手のことを学習しながら相手に徐々に適応していくという行動レベルでの相互適応的なプロセスが観察された。具体的には、以下のような行動がそれぞれの被験者に観察された。

- 教示者

1. 使用する教示の種類を減少させ、
2. 警告韻律を用いることで操作者の行動に注意を与えていた。

- 操作者

1. 発話が与えられてから行動することで、現在の教示に対する理解度を教示者に示し、
2. 異なる種類の発話に対しては異なる行動をとり、
3. 得点が得られたときは自分の行動は正しかったと理解し、逆に警告韻律（教示者が声を荒げた時に観察される韻律的特徴）が与えられると、それをを利用して自分の行動を修正していた（教示者が怒ったような声を出すと「自分の行動は間違っていたようだ」と推測して、自らの行動を修正していた：図2参照）。

このような教示者と操作者の行動は、それぞれ独立して観察されていたのではなく、操作者(1)→教示者(1)→操作者(2)→教示者(2)→操作者(3)、というような順番で観察されていた。つまり、コミュニケーションを成立させることができたペアにおいては、操作者が教示者の自然な適応を促し、同時にその適応を利用していくことで、教示者の発話意図を理解していたと考えられた。また、その際に、韻律情報における普遍的な機能の一つとみなすことができた警告韻律が、操作者に対して自己の行動への注意として解釈されていたことから、意味獲得プロセスの効率化の点で特徴的な韻律情報（この場合は警告韻律）が重要な役割を果たしていることが確認された。

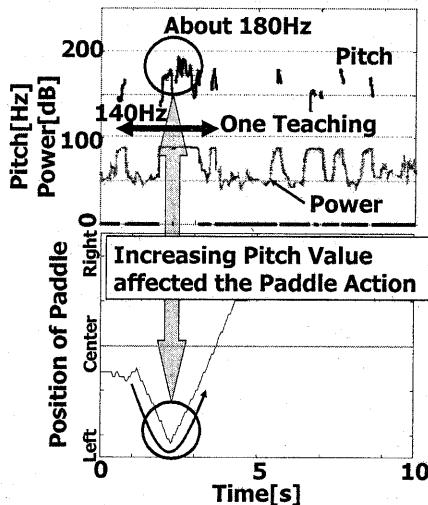


図2. 警告韻律がラケット操作に与える影響

3. 意味獲得モデルの構築

前節のコミュニケーション実験の結果から、コミュニケーションを成立させることのできた被験者ペアにおける操作者の行動を実現できるように意味獲得モデルを設計することで、ユーザの発話意味を理解する学習モデルが実現できることと考えられた。具体的には、以下のような条件を満たす意味獲得モデルを提案した。

1. 発話が与えられると、その発話はなんらかの行動を示していると認識する。逆に、発話が与えられなければ、行動は起こさない。
2. 教示区別の決め手となる音響的・韻律的特徴を見つけ出す。
3. 警告韻律を抽出し、意味学習に利用する。

そして実際には、以下のような概要のモデルを構築した。

- 自分の行動に対して正の報酬を受けたとき（行動が成功したとき=ボールを当てることができたとき）、自分の行動の直前に発せられた教示音声の意味は、自分のとった行動だと認識し、負の報酬を受けたとき（行動が失敗したとき=警告韻律を与えられたとき）、教示音声の意味は、自分の行動ではないと認識する。
- 教示音声はある程度の誤差を持って発せられると仮定する。報酬を受けたときの教示音声一行動のセットは蓄積され（音声一行動データ），その

蓄積されたデータはいくつかのクラスタに分類される。一つのクラスタが一つの教示の意味に相当する。

本モデルでは、正規混合分布から音声一行動データが生成されると仮定したため、それぞれの正規分布のパラメータ（平均値・分散）を求めることが、教示の意味学習に相当する。本モデルでは、EMアルゴリズム[1]を用いて、混合分布中の各分布のパラメータを推定することとした。しかし、従来のEMアルゴリズムでは負の報酬を得た際のデータを学習に利用することはできない。そこで本研究では、従来型のEMアルゴリズムのEstepを拡張することで、負の報酬を受けた音声一行動データを学習に利用できる拡張型EMアルゴリズムを新たに提案した[8]。具体的には、音声データ i が与えられた時、混合分布中の分布 j に属する行動を出力することで負の報酬を受けた場合（教示者から警告韻律を与えられた場合）、音声データ i が分布 j に属する確率を示した値（隠れ値）を $Z_{ij}=0$ とし、残りの分布の隠れ値を $Z=1/(分布数 \cdot 1)$ と修正することとした（成功例の場合には、従来の方法と同様に現在のパラメータから隠れ値を推定する）。この拡張型EMアルゴリズムを実装した、本意味獲得モデルのアルゴリズムを図3に示す。

4. 意味獲得モデルの評価

次に、前節で提案された意味獲得モデルが、教示者とインタラクションしながら発話の意味を学習していく操作者のモデルとして適しているのかどうかを実験的に検討した。具体的には、コミュニケーション実験でゲームの操作者が操作していたゲーム上のラケットに、提案された意味学習モデルを実装し、実際に一般のユーザからリアルタイムで教示を受けた際のラケットの挙動、ゲームの得点状況からモデルの学習状況を評価した。

近年、コンピュータ上のエージェントやコンピュータそのものに対して人間が伝言を行うという実験を行った研究[3]では、その際に人間が「恥ずかしさ」「違和感」を感じてしまうため、機械に対して与える発話は人間に与えるものと比べて不自然なものになると報告している。ユーザが音声インターフェイスを使用する際に、この「不自然な発話」を使用し続けると、ユーザに精神的な負担が生じてしまうとも考えられるため、その不自然さを解消する方法を検討することは必要不可欠である。この先行研究では、エージェントに

1. 音声（ピッチ），行動（ラケット位置），報酬データ読み込み
2. 突然0になる，あるいは突然大きな値になっているピッチデータを補正
3. 補正データをさらに0.05[s]間隔幅の移動平均で円滑化
→このデータに対して以下の処理を実行
4. 差分，2次差分，極端な変化頻度のデータ，有声率を計算
5. 教示の開始，停止のポイントをチェック（1.0[s]の無音区間があれば別教示とみなす）
6. 報酬があった場合
 - 6-1 教示区間の行動をサーチして，現在に近い行動に重みづけをしてラケットの左右行動値を計算
 - 6-2 それをもとに加重速度を計算
 - 6-3 教示区間の音声データの平均を計算（6-2, 6-3がクラスタリングの元データとなる）
 - 6-4 混合ガウス分布を仮定して拡張型EMアルゴリズムの開始
 - 6-5 計算が一定回数を越える，または平均・分散データの変化が一定の値以下になるまで計算をする
 - 6-6 各パラメータを更新
7. 1. に戻る

図3. 意味獲得モデルの学習アルゴリズム

ロボットのような物理的実態を持たせることで，エージェントに対するユーザの自然な発話を誘発できると報告しているが，人間と対話するエージェント全てが物理的実態を持つとは限らないと考えられるため，そのような物理的実態に依存せずに自然な教示を誘発できるような条件を考察することは重要な課題である。そこで本評価実験では，その条件の一つとして，被験者に対して教示対象をイメージさせるようなインストラクションを与えることとした。このインストラクションは「このモデルは，まるで〇〇の様に扱うことができます」といった，キャッチフレーズのようなものに相当するものである。

この評価実験には10人の被験者が参加し，そのうち4人（グループ1）には「コンピュータが動かすラケットにボールがうまく当たるように，このコンピュータに対して教示を与えてください」，そして残りの6人（グループ2）には「コンピュータが動かすラケットにボールがうまく当たるように，このコンピュータに対して人間と接するように教示を与えてください」というインストラクションを与えた。つまり，グループ1の被験者には，教示対象はコンピュータであるとのイメージが与えられ，グループ2の被験者には，教示対象は単純なコンピュータではないとのイメージが与えられたことになる。

この実験の結果，ほとんどの被験者（グループ1：4人，グループ2：4人の計8人）が自分の意図どおりにラケットを操作できるようになったことが観察された。しかし，グループ1の被験者は，警告韻律を使用することなく，自分の用いたい教示を一方的にモデルに対して使用し続けていたのに対し，グループ2の

被験者は，警告韻律を使用しながら，教示の種類を徐々に減少させるといった，コミュニケーション実験の教示者と同じような行動をしていたことが観察された³。

ここから，このモデルは，グループ2のようなインストラクションを受けた被験者の適応を促し，かつ同時にその適応を利用して，自然に与えられる発話の意図通りに行動できたことが確認できた。またその際に，韻律情報における普遍的な機能の一つとみなすことができた警告韻律を，意味学習において負の報酬として利用していたことが確認できた。適応学習した意味獲得モデルに対して被験者がさらに適応していくことを繰り返す「適応のループ」を構成することで，この二者間に真の意味での相互適応が実現できると考えられるが，ここでは，被験者の適応を促進し，その適応を利用した意味獲得モデルが被験者の発話意図を理解するという適応のループの一部分のみが被験者とモデルの間に実現されていた。

5. 被験者はモデルをどのようにみなしていたのか？

以上，コミュニケーション実験の結果を踏まえて構築した意味獲得モデルは，実験者から与えられるインストラクションによっては，一般的なユーザとして参加した被験者との間に，相互適応の一部を構築し，その

³ モデルの評価実験に関する詳しい説明は，参考文献[5,8]を参照されたい。

発話意味を理解できることが確認された。この場合、被験者の行動のみに注目すると、人間同士のコミュニケーション実験の被験者間に観察されていた行動と、ほぼ同様の行動がモデルと被験者との間にも観察されていたといえる。しかしながら「人間が自然にインタラクションできる適応的インターフェイス」の実現のためには、被験者はモデルに対して自然に教示を行うことができたのか、といった被験者の主観的な情報を加味した上でコミュニケーションを評価することが不可欠である。特に本研究のような場合では、インタラクションの相手であるエージェントを「コミュニケーションの適切な対象」として被験者が認識していたのか否かということを把握する必要がある。

これまで、被験者の内部状態を客観的に観測しようとした研究の多くは、被験者の内観報告に重点をおいたものがほとんどであった。しかしながら近年、人間の内部状態を把握するためには、先のコミュニケーション実験行ったような「行動的観察データ」や「被験者の内観報告」に、「被験者の生理的情報」を加えた合計三種類の指標を同時に測定・解析することが、もっとも信頼性の高い方法であると報告されている[9]。

- 認知的指標 (Cognitive Index)：アンケートなどの内観報告による被験者の主観的情報。
- 行動的指標 (Behavioral Index)：ゲームの総得点や刺激に対する反応速度といった、従事している課題における被験者のパフォーマンス。
- 生理的指標 (Physiological Index)：心拍・皮膚電位などの自律反応、脳波・脳磁気電位などの中枢反応などの生理的反応といった被験者の生理的情報。

これまで、これら三種類の指標を同時に測定・解析することで、人間の内部状態を客観的に把握することを提案した研究は存在しておらず、またこれらの客観的な指標を用いてインタラクションの対象を「コミュニケーションの対象」とみなしているのか議論を行った研究も存在していない。

近い将来、本研究が目指しているような「人間にとてコミュニケーションパートナーとなり得るインターフェイス・ロボットなどの人工物（エージェント）の開発」はより盛んになると考えられ、そのようなコミュニケーションエージェントの能力を評価するためには、そのエージェントとインタラクションしている人間の内部状態を客観的に把握する必要がある。よって、本稿で提案するような客観的指標の実現は非常に重要性の高い課題であると考えられる。そこで本稿の残りでは、そのような客観的指標を獲得するための実

験設定を提案する。

6. 人間の内部状態を客観的に把握するため

の実験の提案

6.1 実験手順・環境

実験環境としては、コミュニケーション実験やモデルの評価実験と同様の Pong ゲームを使用する。具体的な実験手順としては以下のようなものを検討している。

1. まず、先述のコミュニケーション実験のように、人間同士が円滑なコミュニケーションを成立させている状態を観察する（実験 A）。具体的には、実験に参加した被験者（一人）に対して、次のようなインストラクションを与える：「隣室にいる他の被験者がラケットでボールを当てられるように音声で指示を出してください」。このインストラクションによって、この被験者には「ゲームの教示者」としてのタスクが与えられる。一方、隣室でラケットを操作するのは教示理解のエキスパートである実験者の一人であるため、この二者は数分のインタラクションの後、スムーズに得点を獲得できるようになり、被験者はインタラクションの対象をコミュニケーションの対象として認識するようになると考えられる。その際の、被験者の認知的・行動的・生理的指標を測定し、これを所望の客観的指標として一時的に定義する。
2. 続いて、インタラクションの対象と人間とがスムーズにコミュニケーションができていない状態を観察する（実験 B）。参加した被験者に対しては、実験 A と同様のインストラクションが与えられる。しかし、実際にラケットを動かすのは実験 A のような実験者ではなく、ランダムにラケットを動かすコンピュータである。この場合、被験者はインタラクションを積み重ねてもスムーズに得点を獲得できるようにならないため、インタラクションの対象をコミュニケーションの対象として認識しない状態になると考えられる。このような状態にある被験者の三種類の指標を実験 A と同様にして測定する。
3. 複数の被験者にて実験 A, B を行い、そこで得られた指標を比較する。そして、それらの指標間に統計的有意差が確認された場合、実験 A で得られた指標を「インタラクションの対象をコミュニケーションの適切な対象と認識している状態を示す客観的指標」として定義することができると思

えられる。

6.2 測定指標

実際に測定を行う三種類の指標は、以下のようなものを想定している。

1. 認知的指標：

NASA-TLX 法[2]に代表されるような、いわゆる SD 法に準拠したアンケート手法を用いて、被験者のタスク (Pong ゲーム) に対する認知的負担度を測定する。被験者が、インタラクションの対象をコミュニケーションの対象として認識するようになれば、認知的な負担が減少していくと考えられる。

2. 行動的指標：

Pong ゲームにおけるパフォーマンスを被験者の行動的指標とする。具体的には、AP 率（総教示数における警告韻律の割合：被験者が操作モデルに対して自然な発話をしていたか否か）、使用教示種類の変遷（被験者が操作モデルの理解度に応じて教示種類を適応させていたか否か）の二つの指標を測定する。被験者が、インタラクションの対象をコミュニケーションの対象として認識するようになれば、AP 率は人間同士のコミュニケーション実験で観察されていた約 5% 程度とほぼ同じくらいの値を示すようになり、使用する教示の種類も実験が進むにつれて徐々に減少していくと考えられる。

3. 生理的指標：

この実験で測定すべき生理指標は、人間の精神的負担・ストレス・覚醒度といった各状態を明確に反映するものを選択する必要がある。また、生理的指標を測定するためには、程度の差こそあれ、電極などの測定端子によって被験者を肉体的に拘束する必要があるため、測定方法によっては被験者に精神的負担を与える場合がある（例：暗室で長時間拘束、頭部を固定した上で多数の電極を接続）。よって、人間の自然な状態を把握する目的の当研究では、被験者に対して肉体的・精神的負担を与えるような測定方法・装置の使用は極力控えるべきである。そこで本研究では、実験中には被験者の心電図 (Electrocardiogram: ECG) のみを測定し、実験後にその心電図データから、心拍率 (Heart Rate: HR)、呼吸性変動 (Respiratory Sinus Arrhythmia: RSA)、心拍性変動 (Mayer Wave Related Sinus Arrhythmia: MWSA) といった各生理指標を測定することとした。具体的には、心電図データから得られる心拍

数から HR と心拍間変動 (R-R Wave Interval: RRI) を求め、RRI を周波数分析すると、0.3, 0.1, 0.25 [1/beat] 付近にそれぞれピークが観察される。そのうち、最初のピーク (0.3 [1/beat]) が RSA で、二番目のピークが MWSA である（図4）。これら HR, RSA, MWSA は人間の以下のような内部状態を反映することが知られている[9,10]。

- HR: 緊張感が増したときに増加。
- RSA: リラックスしている状態の時に増加。
- MWSA: 覚醒度が低下している時に減少し、緊張感が増したときに増加。

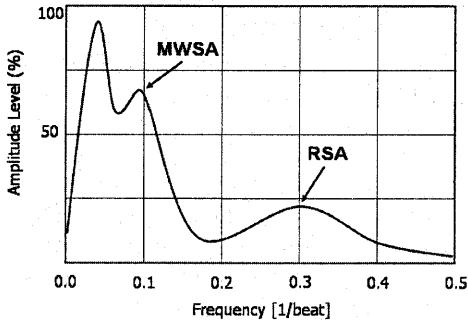


図4. 心拍間変動 (RRI) の周波数分析の例

このような三種類の指標（認知的・行動的・生理的指標）を実験 A, B においてそれぞれ測定し、その指標間に統計的有意差が存在した場合には、実験 A において測定された三種類の指標を、「インタラクションの相手をコミュニケーションの対象と認識している」状態を示した客観的な指標として提案できると考えられる。もっとも、提案した実験を行って実際に得られた指標の分析を行うまでは、これらの指標の有効性を強く主張することはできないが、このような実験の結果が HAI (Human-Agent Interaction) 研究全般において活用されることは十分に期待できる。

7. おわりに

「インタラクションの相手がコミュニケーションの相手としてふさわしいと認識されているのか」という疑問に対しては、認知発達の分野においても興味深い研究が行われている。例えば、Legerstee ら[6]は、新生児はどのような対象をインタラクションの相手としてふさわしいと認識しているかを観察する実験を行っている。このように、新生児を被験者として用いる実験では、「興味を持った対象を好んで注視する」という新

生児の性質を利用して、与えられた対象をどのくらいの長さにわたって注目しているかを測定した選好注視時間 (looking time to target)が、「新生児の興味対象を示す客観的指標」として広く使用されている。認知発達研究における選好注視時間のように、本稿で筆者らが提案した被験者の認知的・行動的・生理的指標を組み合わせた客観的指標は、特に成人が被験者の場合に有効な指標として、主に HAI 研究分野で利用できるものになると期待される。

本稿の前半では、ユーザと相互適応的な関係を構築することで発話の意味を理解するエージェントモデルの開発を行った筆者らの最新の研究結果について述べた。その際、筆者らの開発したエージェントモデルと一般的な被験者との関係をより詳細に解析するために、「被験者はインテラクションの対象（この場合は、構築したエージェントモデル）をコミュニケーションの適切な対象として認識していたのかどうか」ということを客観的に把握することが必要であると考えられた。そこで、インテラクションの対象である人工的エージェントを人間がどのように認識しているのか、特に、「エージェントをコミュニケーションの対象として認識しているのかどうか」を客観的に示す指標を得るための実験設定を提案した。将来的には、ここで提案した実験から得られた指標の妥当性について議論を行う予定である。

- [5] Komatsu, T., Utsunomiya, A., Suzuki, K., Ueda, K., Hiraki, K., & Oka, N. (2003). Toward a mutual adaptive interface: An interface induces a user's adaptation and utilizes this induced adaptation, and vice versa. *Proceedings of the 25th Annual Conference of the Cognitive Science Society*, (CD-ROM).
- [6] Legerstee, M., Barna, J., & DiAdamo, C. (2000). Precursors to the development at intention of 6 months: Understanding people and their actions. *Developmental Psychology*, 36(5), 627-634.
- [7] Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies*, 45, 669-678.
- [8] 小松・鈴木・植田・開・岡(2003). 発話理解学習を利用した適応的インターフェイス--人間同士のコミュニケーション成立過程からの知見.『システム制御情報学会論文誌』,16(6), 260-269.
- [9] 宮田編(1998). 生理心理学. 東京: 北大路書房.
- [10] 渥美(1994). 心拍計測によるドライバーの意識状態評価・自律神経指標を用いた緊張感と眠気の計測. 自動車技術会学術講演会前刷集 946.

参考文献

- [1] Dempster, A., Laird, N., & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM Algorithm. *Journal of Royal Statistical Society B*, 39, 1-28.
- [2] Hard, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock, and N. Meshkati (Eds.), *Human Mental Workload*, Amsterdam: Elsevier Science Publishers.
- [3] 原田(2002). 音声インターフェイスにおけるエイジェント性と社会的文脈の効果:伝言実験による検討.日本認知科学会第19回大会予稿集, pp. 14-15.
- [4] Komatsu, T., Suzuki, K., Ueda, K., Hiraki, K., & Oka, N. (2002). Mutual adaptive meaning acquisition by paralanguage information: Experimental analysis of communication establishing process. *Proceedings of the 24th Annual Conference of the Cognitive Science Society*, pp.548-553.