

明滅光源を用いた内部状態表出による 音声コミュニケーションの円滑化

Smoothing Human-Robot Speech Communication by Using a Blinking-Light

船越 孝太郎*1 小林 一樹*2 中野 幹生*1 山田 誠二*3 北村 泰彦*2 辻野 広司*1
Kotaro Funakoshi Kazuki Kobayashi Mikio Nakano Seiji Yamada Yasuhiko Kitamura Hiroshi Tsujino

*1 (株) ホンダ・リサーチ・インスティテュート・ジャパン
Honda Research Institute Japan Co., Ltd.

*2 関西学院大学
Kwansei Gakuin University

*3 国立情報学研究所
National Institute of Informatics

In this paper, we propose a method to realize smooth communication in speech interaction between a user and a robot. Our method is based on subtle expression by a robot that blinks a small LED attached to its chest. In experiments, participants played a last and first game. We analyzed the number of their speech-repairs and impression for communication and the robot, and obtained promising results.

1. はじめに

話者交代 (turn-taking) の失敗, 特に複数の対話者が同時に発話してしまう「発話の衝突」は, 対話エージェントと人との間の円滑なコミュニケーションを阻害する大きな要因の一つである。発話が衝突するとユーザは自身の発話を中断する傾向があるが, 中断された音声を自動認識することは難しく, 対話エージェントが適切な応答を行えなくなる [Nakano 07]。

エージェントとユーザとの対話において発話が衝突する状況は二つある。一つは, (a) 発話内に休止区間が現れたときに, ユーザがまだ話し続けようとしているにもかかわらず, エージェントがユーザの発話が終了したと誤解して話し始めてしまう状況である。もう一つは, (b) エージェントが話し始めようとしているのにユーザが話し始めてしまう状況である。ユーザが発話を継続するつもりなのにエージェントが発話終了と誤解することを減らすためには, ユーザの発声が止まってからエージェントが話し始めるまでの間隔を長くとればよい。しかし, そうするとエージェントの応答性が悪くなるために, 対話の快適性が損なわれる。また, エージェントからの応答がないためにユーザが発話を繰り返してしまうことによって, 上記 (b) の状況を誘発してしまう恐れもある。

(a) の状況を減らすためには, 発話内容 (言語情報) をみて発話が終了したかどうかを判断する方法 [Bell 01] がある。しかし音声認識を含む不完全な言語情報からの判断には限界がある。韻律情報を使って発話の終了を判定する試み [大須賀 06] もあるが, 韻律情報を安定して取り出すためにある程度の発話長が必要であり, 発話が細切れになると対応が難しい。

(b) の状況に対応するためには, 人間がするように目線や体の微妙な動きといった様々な仕草によって発話権取得の意思をユーザに伝えることが考えられる。しかし, このアプローチでは, ロボットに人間のような繊細で自然な仕草を行わせる技術的な問題と, 経済性の問題がある。また, 身体的な仕草ではなく音声言語 (主に「えっと」などの間投詞) によって上記の意思を伝えることも考えられる。しかし, 何気なく見える間投詞の使用も目線や仕草と同じくらい難しい課題である。間投詞には多くの種類があるがそれらの間には微妙ではあるが確かな違

いがあり, それぞれの使用はランダムではなく, 適切な使用場面やタイミングが存在する。これらの表現を適切に用いるシステムはユーザに好まれるが [Ward 03], タイミングと場面を間違えた瞬間にそれらを使用しない寡黙なシステムよりもユーザの評価が悪くなる [Kitaoka 05]。

本研究では, 明滅光源を導入するという単純な方法でこれらの問題にどれだけ寄与できるかを検証する。機械を人間に似せようとする上記のアプローチに対して, 人工物特有の表現を積極的に採用するアプローチ [Yamada 06] が存在し, その有効性が示されている。このような背景から, 人間によって使用の適切性が強く規定されていない明滅光源のような人工物特有の表現であれば, 人間的な表現を用いる際の問題点を回避しつつ所望の効果を得られる可能性がある。

本稿では, 明滅光源を用いて内部状態を表出する音声対話ロボットを採用し, ユーザとの簡単な対話 (しりとり) を題材とした実験を行う。そしてユーザが言い直した回数や主観評価を通して対話の円滑さを分析する。

以降では, まず提案手法の基礎となる subtle expression について説明する。そして実験方法を説明し, 実験結果と考察を述べる。最後に本研究のまとめを行う。

2. Subtle Expression としての明滅光源

人間同士のコミュニケーションでは, 意味の伝達は発話に代表される言語情報により明示的に行われるが, それ以外の顔の表情, 視線, 身振りなどの非言語情報の重要性も指摘されている [Kendon 94]。その効果は, 単に感情の伝達などの補助的な意味をもつだけでなく, 発話の意味理解に影響を与えることが知られている。例えば, 話し手の身振りを見ることは, 発話理解の精度に影響を与える [Rogers 78]。

さらに, このような非言語情報の中でも, 表情や身振りの非常に些細な変化が人間同士のコミュニケーションに少なからぬ影響を与えていることがわかっており, そのような些細な表出が subtle expression と呼ばれている。この subtle expression が人間同士のコミュニケーションにおいて重要であることは, すなわち人間がそのような些細な表出を直観的かつ容易に理解する能力を備えていることを意味し, 人間と擬人化エージェントやロボットのコミュニケーションでも subtle expression を

利用することで、人間に負担をかけずにコミュニケーションを円滑化できる可能性がある。このような考えのもとに、いくつかの研究 ([Bartneck 05] など) が行われている。しかし、これらはエージェントに表情や身振りを持たせ、その上に subtle expression を実現するため、実装コストが非常に高い。

エージェント、あるいは広く人工物の subtle expression については、山田、小松の研究がある [Yamada 06]。山田らは周波数に変化するピープ音などにより、エージェントの単純な基本的態度の推定を促すことに成功している。また、実際に、外見の異なる2つのロボットに同様のピープ音を実装したとき、人間がロボットのポジティブ/ネガティブな2つの態度を推定できることを示している。これらの研究は、ユーザによる基本的な態度の推定を促進することに、人間的な表現によらずとも subtle expression が有効であることを示唆している。

このような背景から、本研究では、複雑な表情を表出するデバイスではなく、発光ダイオードの明滅という非常に単純な光のシグナルにより、subtle expression の考え方を実現し、ユーザに負荷をかけることなく直観的なロボットの内部状態の理解を促すことを目指す。

具体的には、ユーザとロボットとのしりとり対話タスクを用い、話者交代を subtle expression により円滑に行うことを目指す。ユーザが、ロボットが音声を聞き取れなかったのか、聞き取ったが処理中なのかの2つの内部状態を容易に区別できれば、発話衝突を回避できるため、円滑な対話が可能になると考えられる。本研究では、明滅光源によりそれらの内部状態を表出し、心理学実験を通してその効果を示す。

3. 実験

本実験の目的は、subtle expression として明滅光源を導入することで、ロボットとユーザの対話にどのような影響を与えるかを調査することである。ここでは、しりとりを題材とした対話を設定し、対話の円滑さの基準として、実験参加者の言い直し、対話およびロボットに対する印象、明滅光源の解釈について調査する。以下では、しりとり対話システム、ロボット、明滅光源の設定を説明し、実験手順について述べる。

3.1 しりとり対話システム

本実験では、音声認識器の誤認識が多発する問題を回避するため、参加者からはロボットが自律的に動作しているように見える WOZ (Wizard-of-OZ) システムを採用した。オペレータはロボットの耳の役割を担い、参加者の回答を図1に示すインタフェースを用いて入力する。しり通りの回答はシステムが自動的に選択し、発話を行う。ロボットの音声には市販の音声合成ソフト (NTT-IT FineVoice) を用いた。オペレータは必要に応じて、以下の操作を行う。

1. **参加者の回答をキーボードで入力する:** オペレータは、参加者が回答を言い終わったら、直ちにそれを入力する。回答入力された回答はしり通りのルールに違反していないかチェックされる。ルール違反があればロボットはその旨を参加者に伝え、しりとりをはじめからやり直す。ルール違反がなければ自分の次の回答を作成し、発話する。回答はあらかじめ与えた単語辞書の中から選ばれる。
2. **ロボットに再発話を指示する:** ロボットの発話は自動合成されるため、しばしば聞き取りづらい。参加者がロボットの発話が聞き取れず再発話を要求すると、オペレータの指示によってロボットは直前の発話を繰り返す。

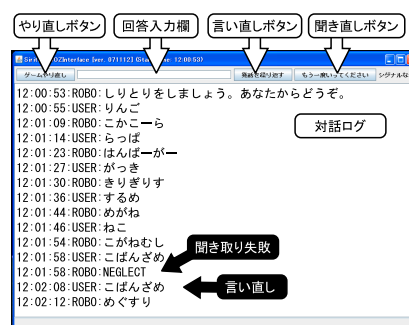


図1: オペレータ用インタフェース

3. **ロボットに聞き返しを指示する:** オペレータが参加者の発話を聞き取れない場合もある。オペレータが聞き取れなかった場合にロボットに聞き返しを指示することで、ロボットは「もう一度言ってください」と発話する。
4. **ロボットにしり通りのやり直しを指示する:** 様々なトラブルによってしりとり対話が破綻し、中断され得る。そのような場合には、トラブルの原因がどのようなものであっても強制的にしりとりをやり直す。ロボットは「また最初からしりとりをしましょう。あなたからどうぞ」と発話し、参加者にしり通りの再開を促す。
5. **参加者による発話の言い直しを記録する:** 操作1によって参加者の回答が入力されると、ロボットが自身の発話を完了するまで図1のシステムは入力を制限し、操作1から4の操作はできなくなる。この間に参加者が発話を言い直したときは、ESC キーを押すことでそれを記録する。

本実験では、観察対象となる現象とそれに対する効果をより明確に確認するために、以下のような調整を行った。

操作1によって参加者の回答が入力されると、ロボットは自分の回答を発話する。しり通りの回答の生成は一瞬で終わるので、このままではロボットの発話が遅れることはない。そこで、回答が入力されてからロボットが発話を行うまでの時間を、発話待機時間を挿入することで、人為的に調整した。発話待機時間は、0秒から15秒までの間で、回答の機会ごとに乱数によって自動で決定した。これは、実際の音声対話エージェントにおいては、内部処理の遅延による応答の遅れに対応する。

また、ロボットは4分の1の確率で参加者の発話を無視するようにした。これは、実際の音声対話エージェントにおいては、音声認識器が入力音声を誤って雑音と見なし棄却してしまうことに対応する (聞き取り失敗)。さらに、参加者が発話せずに膠着状態になるのを回避するため、ロボットが発話してから30秒経過しても参加者が発話しない場合、自動的に直前の発話を繰り返すようにした。

3.2 ロボットと明滅光源の設定

参加者がしりとりを行う対象として、図2に示す人間型ロボット RS Media (WowWee社) を採用した。実験ではロボットは手足を動かさず、静止した状態を維持した。ロボットの胸部には、直径4mmの赤色発光ダイオードを埋め込んでいる。ダイオードは、オペレータがキーボード入力を開始してからロボットが発話するまで明滅する。つまり、ロボットが返答を生成する内部処理が行われている間、ダイオードが明滅する。聞き取り失敗時はこれに相当しないため、ダイオードは消灯したままである。今回の実験では1/30秒の等間隔でダイオードを明滅させた。

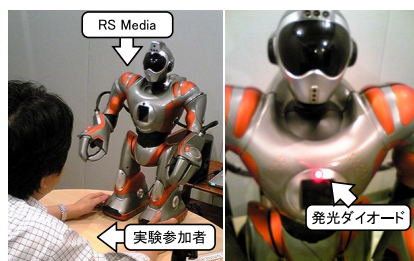


図 2: 実験に用いたロボットと発光ダイオードの取り付け位置

表 1: しりとりルール

- ・最後に「ん」のつく言葉を言ったら負け。
- ・自分がすでに使った言葉は再度使えない。
- ・濁点・半濁点とはつてもよい。(例: ぽ→ぼ)
- ・長音は削除する。(例: コーヒーであれば「ひ」からはじめる)
- ・拗音は最後の文字を使う。(例: しゃ→や)

3.3 参加者

実験への参加者は、19 歳から 62 歳の 47 名の一般人である。ロボットの振る舞いに応じて、以下の 2 つのグループにランダムに分けられた。ただし性別や年代は 2 条件間である程度均等になるよう調整した。

(1) 明滅光源あり条件:

男性 11 名, 女性 12 名, 平均年齢 32.0 歳, 標準偏差 11.0 歳,

(2) 明滅光源なし条件:

男性 11 名, 女性 13 名, 平均年齢 31.0 歳, 標準偏差 10.3 歳.

3.4 実験手順

実験は周囲から区切られた実験室内で行われた。1 回に 1 名のみが入室し、ロボットとしりとり対話を行う。ロボットは、椅子に座った状態の参加者に対し、机の上に正面を向くように設置された。実験室に案内された参加者は、実験概要を説明されたのち、事前アンケートに回答する。

アンケートに回答後、実験の注意事項が教示される。参加者は、これからロボットと 10 分程度のしりとりをすることを説明され、できるだけ長く続くようにと指示される。次に、しりとりのルールとして、表 1 に示すルールが説明され、ルールに迷った場合は気軽に試して構わないことが告げられる。また、ロボットの反応として、ロボットがすぐに返答することもあるが、時間がかかることもあり、聞こえていないこともあると説明される。以上の説明の後、ロボットが話しかけてきたら実験開始の合図だと伝え、参加者を実験室に残して実験者が退出する。実験は 10 分間継続され、実験者がインターフォンで実験終了を告げる。

参加者はしりとりの後、引き続き実験室にてしりとりに対する印象、ロボットに対する印象、実験中のその他の事項に関する 3 つのアンケートに回答する。この実験を通し、参加者の行動と印象に対して明滅光源が与える影響を明らかにする。

4. 実験結果と考察

4.1 言い直しの抑制と促進

参加者の言い直しを (1) ロボットが聞き取りに失敗した時のもの (聞き取り失敗時の言い直し) と、(2) ロボットが聞き取りに成功し、処理を行っている時のもの (処理中の言い直し) に分類した。

表 2: 平均言い直し率

条件	聞き取り失敗時		処理中	
明滅光源あり	55.3%	(SD=50.0)	1.3%	(SD=3.6)
明滅光源なし	24.3%	(SD=34.0)	5.4%	(SD=10.2)

表 3: しりとり印象評価のための形容詞対と評価結果

形容詞対	明滅光源あり		明滅光源なし	
	平均値	標準偏差	平均値	標準偏差
ポジティブ				
気軽な	4.61	1.62	4.04	1.46
円滑な	2.83	1.23	2.33	1.09
感じのよい	3.96	1.30	4.25	1.26
面白い	4.91	1.59	4.04	1.65
リラックスする	3.78	1.41	3.71	1.30
安心な	3.83	1.23	3.42	1.35
暖かい	3.96	1.11	3.67	1.05
愉快な	4.48	0.99	4.25	1.42
ゆつたりとした	5.26	0.96	5.17	1.20
うちとけた	4.40	1.40	3.71	1.23
明るい	4.04	1.07	4.25	0.99
わかりやすい	3.74	1.57	3.21	1.10
好きな	4.35	1.03	4.21	1.41
得意な	4.17	0.98	4.21	1.06
おだやかな	3.83	1.19	3.67	1.52
興味深い	4.61	1.56	4.46	1.74
元気がでる	3.17	0.72	3.50	1.44
落ち着いた	4.13	1.14	3.54	1.25
ネガティブ				
重厚な	4.61	1.62	4.04	1.46
ぎこちない	2.83	1.23	2.33	1.09
感じのわるい	3.96	1.30	4.25	1.26
つまらない	4.91	1.59	4.04	1.65
緊張する	3.78	1.41	3.71	1.30
不安な	3.83	1.23	3.42	1.35
冷たい	3.96	1.11	3.67	1.05
不愉快な	4.48	0.99	4.25	1.42
せわしない	5.26	0.96	5.17	1.20
かたくな	4.40	1.40	3.71	1.23
暗い	4.04	1.07	4.25	0.99
わかりにくい	3.74	1.57	3.21	1.10
嫌いな	4.35	1.03	4.21	1.41
苦手な	4.17	0.98	4.21	1.06
いらいらする	3.83	1.19	3.67	1.52
退屈な	4.61	1.56	4.46	1.74
疲れる	3.17	0.72	3.50	1.44
落ち着かない	4.13	1.14	3.54	1.25

表 2 に聞き取り失敗時と処理中の平均言い直しの率を示す。聞き取り失敗時の平均言い直し率は、ロボットが聞き取りに失敗したときに参加者が言い直しをした割合を参加者ごとに算出し、それを平均した値である。聞き取り失敗時の平均言い直し率に対して t 検定を行った結果、明滅光源あり条件となし条件間で有意差が認められた ($t = 2.67$, $d.f. = 45$, $p = .010$)。処理中の平均言い直し率は、ロボットが参加者の回答を処理している間に参加者が言い直しをした割合を参加者ごとに算出し、それを平均した値である。処理中の平均言い直し率に対して t 検定を行った結果、明滅光源あり条件となし条件間で有意傾向が認められた ($t = 1.80$, $d.f. = 45$, $p = .079$)。

以上から、明滅光源あり条件では、参加者はロボットが聞き取り失敗したときに言い直しを行い、ロボットが処理中の場合には発話しないことが示され、明滅光源の導入によって、発話の重複を適切に抑制できる可能性が示唆された。また、言い直しが過剰に抑制されることはなく、必要な場面 (ロボットが聞き取りに失敗した時) では言い直しが促進されることが示唆された。

4.2 印象評価

形容詞対に対する 7 件法 (非常に・かなり・やや・どちらともいえない・やや・かなり・非常に) により、しりとりに対する印象とロボットに対する印象を参加者に評価させた。表 3 と表 4 に、それぞれしりとりに対する印象評価とロボットに対する印象評価に用いた形容詞対および参加者の評価結果を示す。表はポジティブな形容詞の数値が高くなるように反転項目の処理を行ったものである。この結果に対して [小林 07] と同様の手順で因子分析 (主因子法) を行い、回帰法によって因子得点を求め、明滅光源あり条件となし条件間で比較を行った。

紙面の都合上分析の詳細は省くが、しりとりに対する印象では条件間で有意差が認められる因子が抽出されなかった。一方、ロボットに対する印象では「責任感のある/ない」と「心のひろい/せまい」の 2 つの質問項目が寄与する因子について有意差 ($t = 2.68$, $d.f. = 45$, $p = .010$) が認められた。このことから、明滅光源の導入が誠実で心が広い印象を与えることが示唆された。

表 4: ロボット印象評価のための形容詞対と評価結果

形容詞対	明滅光源あり		明滅光源なし	
	平均値	標準偏差	平均値	標準偏差
ポジティブ				
ネガティブ				
積極的な	4.22	1.31	4.21	1.02
消極的な	4.17	1.27	4.29	1.46
人のよい	4.00	1.48	4.50	1.56
人のわるい	3.96	1.19	4.17	1.31
なまいきでない	4.26	1.14	4.08	1.41
なまいきな	4.22	1.09	3.67	1.27
ひとつっこい	4.13	1.46	3.83	1.17
近づきたい	4.52	1.08	4.50	1.44
にくらしい	4.78	1.00	4.50	1.44
心のひろい	4.04	0.56	3.79	0.72
心のせまい	4.35	1.15	4.21	1.35
社交的な	3.65	1.19	3.79	0.78
非社交的な	4.96	1.15	4.83	1.13
責任感のある	4.22	1.13	3.92	1.35
責任感のない	4.52	1.04	4.46	1.18
慎重な	4.09	1.53	3.96	1.57
軽率な	4.22	1.17	4.21	0.88
恥ずかしがりの	5.00	1.21	4.46	1.22
恥しらずの	4.65	1.23	4.83	1.13
重厚な	3.52	1.34	3.71	1.33
軽薄な				
うきうきした				
沈んだ				
堂々とした				
卑屈な				
感じのよい				
感じのわるい				
分別のある				
無分別な				
親しみやすい				
親しみにくい				
意欲的な				
無気力な				
自信のある				
自信のない				
気長な				
短気な				
親切な				
不親切な				

表 5: 明滅光源の解釈

解釈	人数	(割合)
反応, 理解, 認識	14	(60.9%)
思考中	4	(17.4%)
その他	3	(13.0%)
(気づかなかった)	2	(8.7%)

4.3 明滅光源の解釈

実験後のアンケートでは、「ロボットの振る舞いで特徴的であったこと」、「言い直しをした状況と理由」、「実験中に主にどこをみていたか」を自由回答で尋ねた。これらの項目に対する回答から明滅光源ありの条件で参加者が明滅光源をどのように解釈したかを分類した結果を表5に示す。この結果から、明滅光源あり条件の23名中21名が明滅光源の存在に気づいていたことがわかった。またその意味についても「音声に反応・音声を認識していることを表している」や「考えていることを表している」など、こちらの意図したように解釈していることがわかった。

明滅光源の意味を解釈するためある程度の対話時間が必要であれば、言い直しの頻度が時間軸上で変化すると予想される。そこで前半5分と後半5分の平均言い直し率についてt検定を行った結果、光源ありの場合では処理中・失敗時ともに有意差は認められなかった。一方、光源なしの場合では処理中・失敗時ともに有意差が認められた。失敗時の前半の平均言い直し率は29.7%、後半の平均言い直し率は12.9% ($t = 2.11$, $d.f. = 23$, $p = .046$)、処理中の前半の平均言い直し率は5.9%、後半の平均言い直し率は3.6% ($t = 2.32$, $d.f. = 23$, $p = .029$)であった。

以上のことから、明滅光源の意味は対話開始後すぐに正しく解釈されたと推定でき、明滅光源による内部状態表出は解釈が容易であることが示唆された。

また、言い直しをした理由を条件別に分類した結果を表6に示す。光源なし条件では「返事が遅い」という理由をあげた参加者が3割弱いたが、光源あり条件ではそのような理由をあげる参加者はいなかった。このことから、明滅光源が存在することで返事が遅いという否定的な印象が薄められている可能性が示唆された。

表 6: 言い直した理由

条件	解釈	人数	(割合)
光源あり	光源が点滅しなかったので	16	(69.6%)
	同じ言葉を繰り返したので (言い直さなかった)	5	(21.7%)
		2	(8.7%)
光源なし	同じ言葉を繰り返したので	10	(41.7%)
	返事が遅いので	7	(29.2%)
	その他 (言い直さなかった)	5	(20.8%)
		2	(8.3%)

5. まとめ

本研究では、ロボットが対話中に明滅光源を用いて内部状態を表出することで、円滑なコミュニケーションを実現する手法を提案した。参加者実験では、しりとり対話において、実験参加者の言い直し回数と対話の印象、ロボットの印象を調査した。実験の結果、明滅光源の導入によって、ユーザの不要な言い直しや抑制されることが示された。また、大部分のユーザは明滅光源の意味を対話開始後すぐに正しく把握できることが分かった。

これらの結果から、提案手法である明滅光源による内部状態表出手法を導入することで、教示や訓練なしに人間とロボットの対話を円滑化できることが示唆された。今後は課題指向対話において提案手法の有効性を検証する予定である。

参考文献

- [Bartneck 05] Bartneck, C. and Reichenbach, J.: Subtle emotional expressions of synthetic characters, *Int. J. of Human-Computer Studies*, Vol. 62, No. 2, pp. 179–192 (2005)
- [Bell 01] Bell, L., Boye, J., and Gustafson, J.: Real-time handling of fragmented utterances, in *Proc. of NAACL-2001 workshop on Adaptation in Dialogic Systems* (2001)
- [Kendon 94] Kendon, A.: Do Gestures Communicate?, *A Review. Research in Language and Social Interaction*, Vol. 27, No. 3, pp. 175–200 (1994)
- [Kitaoka 05] Kitaoka, N., Takeuchi, M., Nishimura, R., and Nakagawa, S.: Response Timing Detection Using Prosodic and Linguistic Information for Human-friendly Spoken Dialog Systems, *Journal of The Japanese Society for Artificial Intelligence*, Vol. 20, No. 3, pp. 220–228 (2005)
- [小林 07] 小林 一樹, 船越 孝太郎, 山田 誠二, 中野 幹生, 北村 泰彦, 辻野 広司: 明滅光源を用いた内部状態表出によるロボットとの音声対話の円滑化, HAI シンポジウム 2007 (2007)
- [Nakano 07] Nakano, M., Nagano, Y., Funakoshi, K., Ito, T., Araki, K., Hasegawa, Y., and Tsujino, H.: Analysis of User Reactions to Turn-Taking Failures in Spoken Dialogue Systems, in *Proc. of SIGdial-2007* (2007)
- [大須賀 06] 大須賀 智子, 堀内 靖雄, 西田 昌史, 市川 薫: 音声対話での話者交代/継続の予測における韻律情報の有効性, *人工知能学会論文誌*, Vol. 21, No. 1 (2006)
- [Rogers 78] Rogers, W.: The Contribution of Kinesic Illustrators towards the Comprehension of Verbal Behavior within Utterances, *Human Communication Research*, Vol. 5, pp. 54–62 (1978)
- [Ward 03] Ward, N.: On the Expressive Competencies Needed for Responsive Systems, in *Proc. of the CHI2003 workshop on Subtle Expressivity for Characters and Robots* (2003)
- [Yamada 06] Yamada, S. and Komatsu, T.: Designing Simple and Effective Expression of Robot's Primitive Minds to a Human, in *Proc. of 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2614–2619 (2006)