ASE ベース相槌によるロボットとの対話体験の向上

User Experience Improvement in Robot Talking with ASE based back-channel feedback

小林 一樹 [1*]　船越 孝太郎 [2]　小松 孝徳 [3]　山田 誠二 [4]　中野 幹生 [2]
Kazuki Kobayashi[1]　Kotaro Funakoshi[2]　Takanori Komatsu[3]
Seiji Yamada[4]　Mikio Nakano[2]

[1] 信州大学 工学部
[1] Faculty of Engineering, Shinshu University
[2] (株) ホンダ・リサーチ・インスティチュート・ジャパン
[2] Honda Research Institute Japan Co., Ltd.
[3] 明治大学 総合数理学部
[3] Department of Frontier Media Science, Meiji University
[4] 国立情報学研究所／総合研究大学院大学／東京工業大学
[4] National Institute of Informatics / SOKENDAI / Tokyo Institute of Technology

Abstract:
In this paper, we describe an investigation into users' experiences of a simple talking robot with back-channel feedbacks that is designed based on an artificial subtle expression (ASE). In the experiments with participants, they are divided into six conditions based on an expression factor (three levels; human-like speech, blinking light, and beeping sound) and a timing decision method factors (two levels; decision tree based method and sound volume based method) for investigating participants' impressions on the dialogue experience. We developed an electric pedestal to show the blinking expression, on which the simple cubic robot was fixed. Participants engaged in a task of explaining a cooking procedure with a spoken dialogue system coupled with the robot on the pedestal. The robots responded to them by making the back-channel feedbacks in accordance with the expression factor. The results of questionnaire analyses suggested that the ASE based expressions of back-channel feedback provide positive experiences for users.

1　はじめに

スムーズな会話には，発話，相槌，ポーズなどがタイミングよく出現する必要がある．特に相槌は，聞き手から話し手に対するコミュニケーション手段であり，円滑なコミュニケーションに果たす役割は大きい．声の高さ，タイミング，非言語行動も相槌として重要な要素だと指摘されており [1]，人間同士の対話研究でも相槌が分析対象の 1 つとなっている [2, 3, 4]．また，相槌のパラ言語的側面に着目した研究もあり [5]，相槌が打たれるまで 300 ミリ秒以上あると不適切と感じるという知見 [6] もある．

人間とコミュニケーションを行うロボットやエージェントの開発においても相槌は重要な要素の 1 つであり，適切な相槌やうなずきの生成を試みる研究がある [7, 8]．決定木学習を用いた話者交替・継続の判別実験 [9] をはじめ，相槌タイミングを決定する手法 [10] のほか，ユーザ発話中の終助詞（「よ」,「ね」など）に応じて相槌の発話テンプレートを使い分ける方法 [11] やユーザの話題に対する関心度に応じて単純相槌や反復相槌を使い分ける方法 [12] などが提案されている．

しかし，相槌のように短時間の発話で比較的シンプルな表現であっても，ロボットやエージェントに自然で人間らしい振る舞いを行わせるためには，タイミングの決定や表現の選択に複雑な手法を実装する必要がある．人間的な相槌表現を追求して，円滑なコミュニケーションを実現することは重要な課題であるが，次節で説明する Artificial Subtle Expression (ASE) に関する研究例のように，過度に複雑な制御を行わずとも

*連絡先：信州大学 工学部
〒 380-8553 長野県長野市若里 4-17-1
E-mail: kby@shinshu-u.ac.jp

コミュニケーションの質を向上できる可能性がある．

そこで，本研究ではロボットやエージェントによる相槌に着目し，人間的な表現を用いない相槌生成手法を提案する．提案手法では，箱型のロボットが光の明滅やビープ音を提示して相槌を表現し，それが対話体験に対して与える影響を調査する．以降では，ASE とそれを用いた相槌表現の説明を行い，続いて対話実験とその結果について報告する．

2　ASE ベース相槌

2.1　Artifical Subtle Expression

相槌に限らず，顔の表情や視線，身振りなどの非言語情報は非常に些細な変化であっても心的状態を伝達しており [13]，そのような表出は Subtle Expression と呼ばれている [14]．人間と同じように振る舞うことが要求されるロボットや擬人化エージェントの開発において，Subtle Expression を実装して人間とのコミュニケーションを円滑化する取り組みがある [15, 16, 17]．しかし，わずかな動作や変化を行うにも多くの関節を必要としたり，自然でスムーズな動作が求められるために制御が複雑になる問題がある．

この問題に対し，Artificial Subtle Expression (ASE) が提案されている．光の明滅や音声をわずかに変化させる人工的な表出であってもロボットやエージェントの内部状態を事前説明なしにユーザに伝達できることが示されている．たとえば，発光ダイオードの明滅によってロボットが「音声を認識している／考えている」という内部状態を伝達し，ユーザとの円滑な話者交替が実現されている [18]．また，ロボットがユーザにアドバイスを与える場面において，ビープ音を付加することでロボット自身が抱く確信度を伝達できることが示されている [19, 20, 21]．さらに，抑揚を変化させたビープ音を提示することにより，ロボットの態度（肯定的／否定的）を人間に伝達可能であることが報告されている [22]．対話ロボットに関しては，発光ダイオードの明滅による ASE を提示することで，ユーザは対話やロボットの印象を肯定的に評価する傾向のあることが示されている [23]．

このように ASE の利点は，コストを抑えた実装が可能であり，ユーザが事前知識なしに直観的に理解できることである．ASE をベースにした相槌表現を用いることで，シンプルな制御であってもロボットとの対話体験を向上させられる可能性がある．

図 1: 明滅光源台座に設置した箱型ロボット

2.2　ASE に基づく相槌表現

本研究では，ASE に関する従来研究で採用されている明滅光源とビープ音を用いて相槌を表現する．

明滅光源による相槌表現は，円形をした台座型の明滅デバイスを用い，この上にロボットを配置して行われる．明滅による相槌は，ユーザの発話の後 330 ミリ秒の間 15Hz の赤色の明滅によって表現される．図 1 に台座に設置した箱型ロボットを示す．箱型ロボットは，先行研究において人間型ロボットよりも肯定的に評価されることが示されている [24]．台座型明滅デバイスは，図 2 に示すとおり，6 個のフルカラー LED と LED コントローラ，透明円形アクリルで構成されている．LED コントローラに制御信号を送ることで LED が明滅する．LED は光線が放射状に広がるように円形アクリルの中心部分にに配置されている．また，円形アクリルは表面が研磨処理されており，明滅時には内部からの光が散乱してアクリル全体が光っているように見える．

ビープ音による相槌表現は，ユーザの発話の後 128 ミリ秒の間音声が再生される．このビープ音は「プ」という発音に近い音声である．図 3 にビープ音の波形を示す．この音声は，ユーザに威圧感や不快感を与えず，ロボットの反応として分かりやすいと考えられるものを，複数の候補の中から著者らで検討して選定した．

3　対話実験

対話実験では，ASE をベースとした明滅光源による相槌表現とビープ音による相槌表現の効果を確認するために，ロボットとの対話体験についてユーザの印象を調査する．

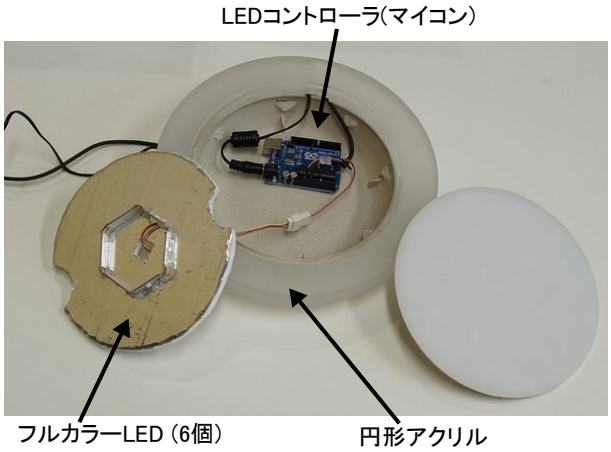

図 2: 明滅光源台座の内部構造

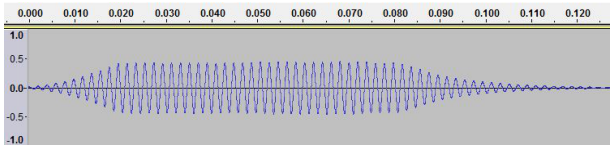

図 3: ビープ音の波形

3.1 要因と水準

対話体験の印象に与える要因として，相槌表現のほかに，相槌タイミングを採用した．相槌表現要因は明滅光源，ビープ音，人間音声の 3 水準であり，相槌タイミング要因は決定木ベースと音圧ベースの 2 水準とした．

明滅光源とビープ音は先に示したものと同一であり，人間音声は合成音で構成した「はい」という女性の声である．図 4 に合成音声「はい」の波形を示す．この音声はユーザに違和感を与えないように，人間の発話のように聞こえるように調整した．発話時間は 330 ミリ秒であり，明滅光源が明滅を行う時間と同一である．

相槌タイミングについては，対話中に WoZ によって実験者が決定することもできる．しかし，ユーザの発話末から 300 ミリ秒以上経過するとユーザが不適切に感じる [6] ことから，実験者に高度な訓練を施すのは現実的でないと判断し，ソフトウェアによって自動生成

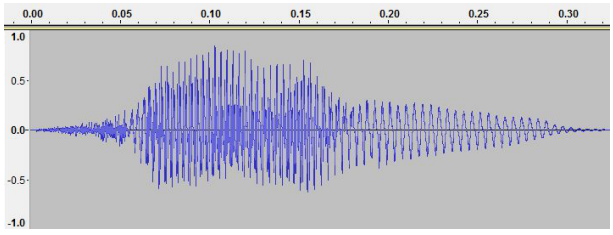

図 4: 合成音声「はい」の波形

図 5: 実験環境

する手法を採用した．決定木ベースのタイミング生成は，ユーザの発話長や文脈などを入力して決定木学習で相槌タイミングを生成する手法である．また，音圧ベースのタイミング生成は，単純にユーザ発話の音圧を監視し，200 ミリ秒以上続くユーザ発話長のあとに，200 ミリ秒以上の無音区間を検出した場合に相槌を行う方法である．

3.2 実験タスクと参加者

実験タスクは，ロボットによる相槌の効果を確認するために，参加者がロボットに対して作業の手順を口頭で教えるものを採用した．

実験参加者は，カレーの作り方の手順を説明した 1 分 42 秒の動画を一度視聴する．カレーの作り方の動画は YouTube で公開されているもので，説明が分かりやすいものを選定した．次に防音ブース内に置かれたロボットの前に移動し，図 5 に示すように椅子に腰掛けた状態でヘッドホンを装着し，動画の内容を参考にしてロボット対して口頭で作業手順を説明する．参加者には事前に，説明を行う時間は 3 分間であり，厳密に動画の手順を再現する必要はなく，個人の経験に基づいて自由にアレンジして良いことを教示してある．ロボットはその説明を聞きながら，応答（相槌）を行う．応答の返し方は，相槌表現要因 3 水準とタイミング要因 2 水準の組み合わせで構成される 6 種類であり，1 人の参加者はそのうちのどれか 1 つだけを経験する（参加者間配置）．ロボットの応答が音声による場合はヘッドホンから出力される．参加者の音声はスタンドマイクによって収録する．ロボットへの説明タスク終了後，参加者はロボットの応答の様子についてのアンケートに答える．アンケート内容は表 1 に示すとおりであり，参加者はすべての項目について 7 段階で評価を行う．

参加者は 20 歳から 64 歳までの男女 90 名（男性 40 名，女性 50 名）であり，平均年齢は 36.8（標準偏差 12.3）であった．

表 1: アンケート項目

記号	質問内容
Q1	このロボットは反応の仕方が人間らしい
Q2	このロボットの反応の仕方は違和感がある
Q3	このロボットと話をするのは楽しい
Q4	このロボットは反応を話の途中で返してくる
Q5	このロボットは話を聞くのが上手い
Q6	このロボットには話しやすい
Q7	このロボットは反応が多すぎる
Q8	このロボットは話をよく聞いてくれている感じがする
Q9	このロボットにまた話を聞いてもらいたい

表 2: アンケート評価値と分散分析の結果

タイミング	決定木ベース			音圧ベース			分散分析		
表現	合成はい	ビープ音	明滅台座	合成はい	ビープ音	明滅台座	タイミング要因の主効果	表現要因の主効果	交互作用
項目	平均 S.D.	平均 S.D.	平均 S.D.	平均 S.D.	平均 S.D.	平均 S.D.	F(1,72)	F(2,72)	F(2,72)
Q1 +	2.47 0.81	2.47 1.26	2.75 1.20	3.80 1.28	2.31 1.38	3.08 1.32	2.93+ 音圧＞決定木	2.28 n.s.	2.24 n.s.
Q2	2.87 1.75	3.13 1.78	4.38 1.65	4.27 1.73	3.31 2.09	3.67 1.49	0.46 n.s.	1.18 n.s.	2.05 n.s.
Q3 +	3.13 1.36	3.07 1.65	2.63 1.11	3.87 1.31	3.15 1.79	3.75 1.83	3.02+ 音圧＞決定木	0.41 n.s.	0.66 n.s.
Q4 **	2.00 1.26	4.93 2.26	4.50 2.24	3.00 1.79	4.31 2.23	4.75 2.28	0.18 n.s.	8.46** 下位検定 1	0.93 n.s.
Q5	2.93 1.12	3.67 1.62	3.88 1.90	3.87 1.15	3.54 1.60	3.75 1.79	0.38 n.s.	0.42 n.s.	0.93 n.s.
Q6	3.13 1.36	3.33 1.45	3.13 1.83	3.87 1.20	3.15 1.61	3.58 1.66	0.87 n.s.	0.17 n.s.	0.56 n.s.
Q7 **	2.60 1.40	4.80 2.37	5.63 1.22	4.07 2.02	4.77 2.36	4.83 1.99	0.2 n.s.	5.71** 下位検定 2	1.92 n.s.
Q8	4.00 1.71	4.07 1.91	5.25 1.71	4.67 1.19	4.46 1.99	4.58 1.80	0.1 n.s.	0.98 n.s.	0.94 n.s.
Q9	3.07 1.57	3.07 1.53	3.38 1.87	4.13 1.71	3.08 1.69	4.33 1.97	2.71 n.s.	1.25 n.s.	0.66 n.s.
平均	2.91 0.86	3.61 1.10	3.94 1.10	3.95 1.02	3.56 1.03	4.04 1.19	2.02 n.s.	1.74 n.s.	1.82 n.s.

下位検定 1(Tukey の HSD 法): ビープ音 > 合成はい (HSD= 1.32*)，明滅台座 > 合成はい (HSD= 1.45*)
ビープ音と明滅台座は，それぞれ合成はいと比較して話の途中で反応を返さない．(値が小さいほど反応を途中で返す)
下位検定 2(Tukey の HSD 法): ビープ音 > 合成はい (HSD= 1.30*)，明滅台座 > 合成はい (HSD= 1.42*)
ビープ音と明滅台座は，それぞれ合成はいと比較して反応の数が適切．(値が大きいほど反応数が適切)
+ : $p < .1$, * : $p < .05$, ** : $p < .01$

3.3 実験結果

表 2 にアンケートの分析結果を示す．評価値はポジティブな内容ほど数値が高い．具体的には，Q2,Q4,Q7 についてはネガティブな質問内容であるため，数値が高いほど質問内容にあてはまらない．それ以外の項目は，数値が高いほど質問内容にあてはまる．よって，最下段の平均値においては，各条件について数値が大きいほど肯定的な評価になっている．

アンケート内容には，表 1 に示す以外にも，戸惑ったこと，ロボットの振る舞いに望むこと，ロボット反応の仕方で良かった点について自由記述を求めている．この内容をもとに，明らかにロボットの相槌に気付いていないと判断される参加者を除外して 2 要因の分散分析を行った．分析対象の人数は，決定木＆明滅台座条件の分析対象は 8 名，音圧＆ビープ音条件は 13 名，音圧＆明滅台座条件は 12 名となり，他の 3 条件は 15 名である．

質問項目 Q1 から Q9，および全項目の平均値に対して分散分析を行ったところ，すべての項目で交互作用は認められなかった．質問項目 Q1,Q3 においてタイミング要因の主効果に有意傾向が認められ，音圧ベースによる相槌タイミング決定方法のほうが，決定木ベースに比較して反応の仕方が人間らしく，話をするのが楽しいという印象を与えられることが示唆された．

また，質問項目 Q4 と Q7 において表現要因の主効果が認められ，Tukey の HSD 法による下位検定を行ったところ，ビープ音と明滅台座による相槌表現は合成音声「はい」に比較して，それぞれ話の途中で反応を返さず，反応の数が適切である印象を与えることが示唆された．

4　考察

本研究の目的は，ASEをベースとした相槌表現がユーザの対話体験の質にどのように影響するのかを明らかにすることである．したがって，ここでは参加者実験における相槌表現要因と相槌タイミング要因との関係に着目し，その効果について検討する．

4.1　人間らしさと相槌の適切さ

相槌タイミング決定方法の違いは，人間らしさや対話の楽しさに対して影響を与えることが示唆された．ASEによる相槌は人間的な表現ではないため表現要因の影響が強いことも予想されたが，今回の実験では，表現の差異よりも相槌タイミングのほうが人間らしさや対話の楽しさに強い影響力が示された．つまり，相槌が人間のような音声であっても，光の明滅であっても関係なく，人間らしさの印象を与えるにはタイミングが重要である可能性がある．

一方で，相槌表現の違いは，ユーザが感じる反応の頻度や割り込みの度合いに影響を与えることが示唆された．これは，同じタイミングで相槌を行っていても，表現方法によって相槌が過多であったり，話の途中で割り込まれた感覚を与えてしまうことを意味する．特に，人間的な合成音声「はい」の場合には相槌が多すぎたり，話の途中で相槌を返す不適切な印象を強く与えている．

4.2　相槌タイミングの正確さと対話体験

各条件においてロボットが実際に行った相槌回数は調査中であるが，アンケートの自由記述部分を確認すると，決定木ベースの相槌生成方法によるロボットと対話した参加者の中には，話の区切りで相槌を打ってほしい，タイミングがずれていた，まだ話しているのに相槌があった，欲しいところで相槌がなかった，とする意見が散見された．決定木ベースの相槌生成方法は，本来は通常の会話で用いられる手法を相槌に限定して使用しているため，結果的に音圧ベースの相槌生成方法と比較して適切なタイミングの生成に失敗している可能性がある．

そこで，仮に，相槌生成タイミングについて，相槌タイミングの正確さが高い／低いの2水準であると考えてみる．その上で，相槌表現要因の主効果が認められた相槌の頻度や割り込み度合いの印象を検討すると，相槌タイミングの正確さには関係なく，明滅光源やビープ音による相槌の有効性が示唆される．システム開発の立場から考えると，相槌生成タイミングを細かく調整しなくてよいため，システムに人間らしさを求めない場合には明滅光源やビープ音によるASEを用いた相槌を行う利点は大きい．特に，明滅光源やビープ音は，同じパターンを何度も繰り返しているだけでも一定の効果が期待できる．合成音声「はい」を提示された参加者の中には，「はい」以外の返事が欲しいとする意見があったが，ASEによる相槌の場合には，そのような工夫がなくても相槌の頻度や割り込み度合いについて適切だとみなされる可能性がある．

5　まとめ

本研究では，人間とロボットとの音声対話において，ASE(Artificial Subtle Expressions) に基づいた相槌がユーザの対話体験に及ぼす影響を調査した．対話実験では，相槌の表出要因として，人間の音声「はい」と，ASEベースである明滅光源とビープ音とによるものの3水準を設定し，タイミング決定要因としてユーザの入力音声から決定木でタイミングを決めるものと，音圧で決めるものとの2水準を設定した．ロボットの外見はシンプルな箱型とし，台座型の明滅光源デバイスの上に配置した．実験参加者には，対話ロボットに対して料理の作り方の説明を行う課題与え，課題後にロボットの傾聴態度に関するアンケート調査を実施した．調査の結果，相槌タイミング決定手法にかかわらず，明滅光源とビープ音は合成音声「はい」に比較して，ユーザの話の途中で反応を返さず，反応の数は適切だという対話体験を与えることが示唆された．

ASEによる相槌表現は，ユーザに同じパターンを繰り返し提示するシンプルなものであり，人間が行う相槌のように「はい」や「うん」などを織り交ぜた多彩な表現ではない．それにもかかわらず，相槌タイミングの正確さに関係なく相槌としての適切さが示唆された点は興味深い．今後，様々なパターンを駆使した人間的な相槌と比較して，ASEによる相槌がユーザにどこまで受け入れられるか，その効果の範囲を調査する予定である．

参考文献

[1] 堀口純子. 日本語教育と会話分析. くろしお出版, 1997.

[2] 長岡千賀, 小森政嗣. 心理面接におけるカウンセラーの応答：話者交替時のカウンセラーの発話冒頭を指標とした事例研究. *Cognitive Studies*, Vol. 16, No. 1, pp. 24–38, 2009.

[3] 大森晃, 土井晃一. あいづちが発想数に与える影響–その実験と分析—. *Cognitive Studies*, Vol. 7, No. 4, pp. 292–302, 2009.

[4] 豊田薫, 宮越喜浩, 山西良典, 加藤昇平. 発話状態時間長に着目した対話雰囲気推定. 人工知能学会論文誌, Vol. 27, No. 2, pp. 16–21, 2012.

[5] 戸田貴子. パラ言語的側面から見たあいづちに関する研究. 日本語教育方法研究会誌, Vol. 8, No. 1, pp. 12–3, 2010.

[6] 岡登洋平, 加藤佳司, 山本幹雄, 板橋秀一. 韻律情報を用いた相槌の挿入. 情報処理学会論文誌, Vol. 40, No. 2, pp. 469–478, 1999.

[7] 小林哲則, 藤江真也. マルチモーダル会話ロボット：ロボットが会話において行う「聴く」行為について. 計測と制御, Vol. 46, No. 6, pp. 466–471, 2007.

[8] 藤江真也, 小川哲司, 小林哲則. 会話ロボットとその聴覚機能. 日本ロボット学会誌, Vol. 28, No. 1, pp. 23–26, 2010.

[9] 大須賀智子, 堀内靖雄, 西田昌史, 市川熹. 音声対話での話者交替／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, pp. 1–8, 2006.

[10] N. Kitaoka, M. Takeuchi, Nishimura R., and Nakagawa S. Response timing detection using prosodic and linguistic information for human-friendly spoken dialog systems. 人工知能学会論文誌, Vol. 20, No. 3, pp. 220–228, 2005.

[11] 大竹裕也, 萩原将文. 高齢者のための発話意図を考慮した対話システム. 日本感性工学論文誌, Vol. 11, No. 2, pp. 207–214, 2012.

[12] 小林優佳, 山本大介, 土井美和子. 音声対話システムのための発話間の共起性を利用した音声認識結果からの単語取得方法. 人工知能学会論文誌, Vol. 28, No. 2, pp. 141–148, 2013.

[13] P. R. Cohen, J. Morgan, and M. E. Pollack. *Intentions in Communication*. MIT Press, 1990.

[14] K. Liu and R. Picard. Subtle expressivity in a robotic computer. In *Proc. of CHI 2003 Workshop on Subtle Expressiveness in Characters and Robots*, pp. 1–5, 2003.

[15] C. Bartneck and J. Reichenbach. Subtle emotional expressions of synthetic characters. *International Journal of Human-Computer Studies*, Vol. 62, No. 2, pp. 179–192, 2005.

[16] O. Sugiyama, T. Kanda, M. Imai, H. Ishiguro, N. Hagita, and Y. Anzai. Humanlike conversation with gestures and verbal cues based on a three-layer attention-drawing model. *Connection Science*, Vol. 18, No. 4, pp. 379–402, 2006.

[17] M. Kipp and P. Gebhard. Igaze: Studying reactive gaze behavior in semi-immersive human-avatar interactions. In *Proc. of the 8th International Conference on Intelligent Virtual Agents*, pp. 191–199, 2008.

[18] 船越孝太郎, 小林一樹, 中野幹生, 山田誠二, 北村泰彦, 辻野広司. Artificial subtle expression としての明滅光源による音声対話の円滑化. 電子情報通信学会論文誌, Vol. J92-A, No. 11, pp. 818–827, 2009.

[19] 小松孝徳, 山田誠二, 小林一樹, 船越孝太郎, 中野幹生. Artificial subtle expressions: エージェントの内部状態を直感的に伝達する手法の提案. 人工知能学会論文誌, Vol. 25, No. 6, pp. 733–741, 2010.

[20] T. Komatsu, S. Yamada, K. Kobayashi, K. Funakoshi, and M. Nakano. Artificial subtle expressions: Intuitive notification methodology of artifacts. In *Proc. of the 28th ACM Conference on Human Factors in Computing Systems (CHI2010)*, pp. 1941–1944, 2010.

[21] 小松孝徳, 小林一樹, 山田誠二, 船越孝太郎, 中野幹生. 確信度表出における人間らしい表現と artificial subtle expressions との比較. 人工知能学会論文誌, Vol. 27, No. 5, pp. 263–270, 2012.

[22] Takanori Komatsu and Seiji Yamada. How do robotic agents' appearance affect people's interpretation of the agents' attitudess? In *Extended Abstract (Work in Progress) of the ACM-CHI 2007*, pp. 2519–2525, 2007.

[23] 船越孝太郎, 小林一樹, 中野幹生, 小松孝徳, 山田誠二. 対話の低速化と artificial subtle expression による発話衝突の抑制. 人工知能学会論文誌, Vol. 26, No. 2, pp. 353–365, 2011.

[24] Kazuki Kobayashi, Kotaro Funakoshi, Seiji Yamada, Mikio Nakano, Takanori Komatsu, and Yasunori Saito. Impressions Made by Blinking Light Used to Create Artificial Subtle Expressions and by Robot Appearance in Human-Robot Speech Interaction. In *IEEE International Symposium on Robot and Human Interactive Communication*, pp. 215–220, 2012.