

## ユーザは言動不一致なシステムを許せるのか？

### Can Users Live with Overconfident or Unconfident Systems? A Comparison of Artificial Subtle Expressions with Human-like Expressions

小松孝徳<sup>1</sup> 小林一樹<sup>1</sup> 山田誠二<sup>2</sup> 船越孝太郎<sup>3</sup> 中野幹生<sup>3</sup>

Takanori Komatsu<sup>1</sup>, Kazuki Kobayashi<sup>1</sup>, Seiji Yamada<sup>2</sup>, Kotaro Funakoshi<sup>3</sup>, and Mikio Nakano<sup>3</sup>

<sup>1</sup> 信州大学

<sup>1</sup>Shinshu University

<sup>2</sup> 国立情報学研究所/総合研究大学院大学/東京工業大学

<sup>2</sup> National Institute of Informatics/SOKENDAI/Tokyo Institute of Technology

<sup>3</sup> ホンダ・リサーチ・インスティテュート・ジャパン

<sup>3</sup> Honda Research Institute Japan

**Abstract:** Expressing the confidence level of a system's suggestions by using speech sounds is an important cue to users of the system for perceiving how likely it is for the suggestions to be correct. We assume that expressing confidence levels by using human-like expressions would cause users to have a poorer impression of the systems than if artificial subtle expressions (ASEs) were used when the quality of the presented information does not match the expressed confidence level. We confirmed that this assumption was correct by conducting a psychological experiment.

## はじめに

リビングルームでテレビを見ている父親に向かってなかなかうまくいかない就職活動の状況をグチりながらも、携帯電話で彼女にメールを書いている大学生。帰宅した子供に「ご飯ができるまで宿題を終わらせなさい！」と大声で叱りながらも、台所でせっせと晩御飯の支度する母親。

これらの例からも分かるように、我々はあるタスクに従事しながらでも、そのタスクとは独立に「発話」を使用し、他者とコミュニケーションをとることができる。このような発話というモダリティに注目し、音声認識や音声合成といった工学的技術を人間と人工物とのインタラクションに積極的に導入することで、発話によるインタラクションを二者間に構築しようという試みが近年は特に活発に行われている[1,2]。例えば、スマートフォンにおいてはウェブブラウザへの検索語の入力は発話音声で行われるケースが増えていると報告されたり[3]、カーナビゲーションシステムを使用する際、ユーザは目的地を発話音声によって入力し、システムもユーザへの指示を人工音声によって出力することが一般的になってきている。しかしながら音声認識や音声合成といった発話インタラクションにおけるコアな技術は、

年々その動作精度を向上させているものの、音声入力時のノイズ処理や入力データの不完全性などが原因となり、あらゆる状況にて完璧な動作が保障される技術レベルに達していないこともまた事実である[4]。このような現状を踏まえた上で Cai らは[5]、システム自身が提示する発話情報と同時にその情報自体が正確なのか否かを示す確信度を表出することが、システムに対するユーザの印象や行動を改善すると報告している。

具体的にどのような方法によってシステムがユーザに対して確信度を表出するのかということを考えて場合、おそらく多くの人が「たぶん」「完璧です」「確信度 83%です」といった人間らしい発話情報をシステムから表出する方法を思いつくのではないだろうか。しかしながら筆者らは、「システムがユーザに提示した情報」と「表出した確信度」とが現実の状況と一致しない状態（＝言動不一致状態）においてこのような人間らしい表現による確信度の表出が行われると、ユーザはシステムに対して大きな不快感を抱くのではと危惧している。例えば、カーナビゲーションシステムが指定された目的地までのルートを人工音声で提示した後、「このルートが最も早く到着できるルートです」と提示情報に関する確信度を表出していながら、実際にはその予定よりも大幅

に遅れて目的に到着した場合、ユーザはこのシステムに対して「余計なことを言いやがって...」「このウツキシステムが...」などと、システムに対して不快感を抱くことは容易に想像できる。このような現象は、人間らしい表現を表出する人工物に対して多くのユーザは人間らしい機能を想像してしまい[6,7]、その想像が打ち崩されるような現実を見てしまうと、その人工物に失望感を感じるという実験的な知見にて説明できよう[8,9]。

その一方、システムがユーザに対して確信度を表出するという枠組みに関連して、筆者らはすでにシステムの確信度といった内部状態を直観的かつ正確にユーザに伝達する手法である Artificial Subtle Expressions (ASE)を提案している[10,11]。具体的には、ビープ音のような人工音として実装された ASE が、提示した情報に関してシステムが持っている確信度（例。自信がある/ない）を直観的かつ正確にユーザに伝達できることが実験的に確認された。よって筆者らは、人間らしい表現による確信度の表出ではなく、ASEによる確信度の表出を行うことで、言動不一致状態におけるシステムに対するユーザの不快感を軽減できるのではと考えた。

そこで本研究では、「ASE によって確信度を表出する場合よりも人間らしい表現によって確信度を表出する場合の方が、言動不一致なシステムへの評価が低くなる」との仮説を検証することを目的とする。一見すると言動不一致なシステムを研究の対象とすることは非現実的のようにみえるが、「システムが提示する情報」とその際に表出される「確信度」とが常に現実の状況と一致しているとは限らず、むしろこのような状況が起こり得ないとは決して断言できない状況にあるといえる。よって、このような状況に陥ってもユーザに不快感を与えないようなインタラクションのデザインを模索することは、HAI/HCI/HRI 研究として大きな意義があると言えよう。

## 実験

### 実験環境

システムが提示する情報と表出する確信度との関連においてユーザがシステムに対してどのような印象を抱いているのかを把握するための実験環境として「ドライブ宝探しテレビゲーム」を使用した(図1)。このゲームでは、カーナビゲーションシステムが実装された車に乗って画面の奥に設定されたゴールに向かって道を進んでいくと、三つの盛り土が現れる。この三つの盛り土のうち一つには金貨が隠さ

れており、その金貨をできるだけ多く集めることが実験参加者のタスクとして与えられる。三つの盛り土のいずれかに隠された金貨の位置は、ランダムに設定される。三つの盛り土はゴール到着まで合計20回現れ、スタートからゴールまでの所要時間はおよそ三分程度であった。

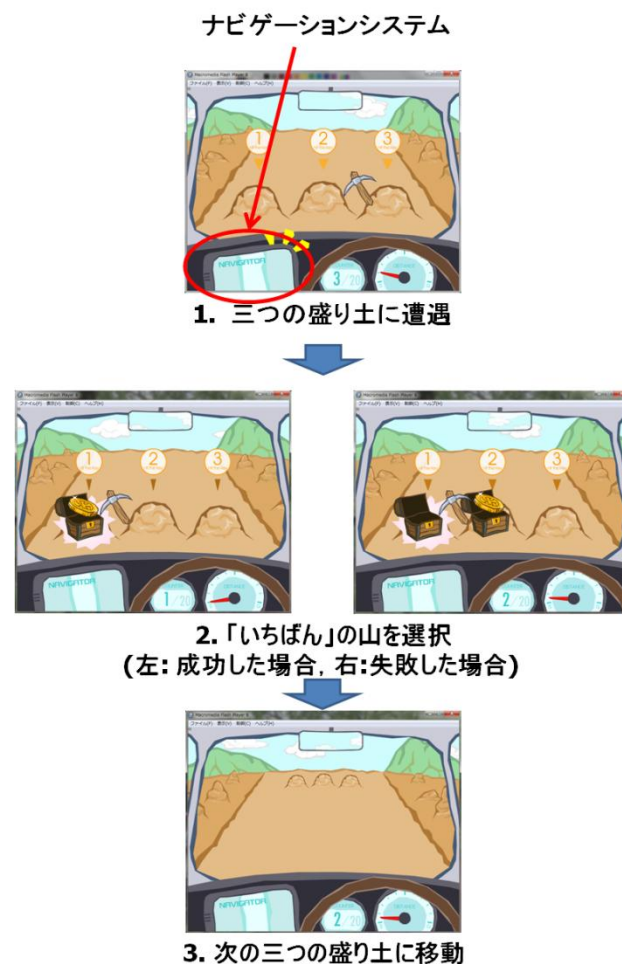


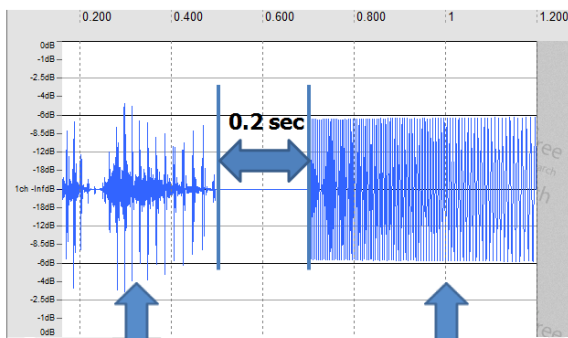
図1: ドライブ宝探しゲーム

ゲーム中に画面が三つの盛り土の前に到着すると、運転席の左側に配置したカーナビゲーションシステムから、そのシステムが予想した金貨が隠された盛り土の位置が、人工音声にて参加者に伝えられる。参加者はこのエージェントの予想通りの盛り土を選んでも、予想と違うものを選んでもよいと教示される。また参加者は、自分が選択した盛り土に金貨が実際に隠されていたのかどうか、また隠されいなかった場合にはどの盛り土に金貨が隠されていたのかというフィードバックを盛り土の選択直後に理解できるように設定された。

## 使用音声

本実験では、ナビゲーションシステムがユーザに対して金貨が含まれる盛り土の位置の予測を「いちばん」「にばん」「さんばん」という人工音声として提示する。これらの音声は、筆者のうち一人の音声を録音し、それに対してロボット音声のようなエフェクトを付与したものである。この人工音声に関する確信度の表出方法として、以下の三つの方法を実験条件として設定した。

- **ASE 条件**：金貨の位置を予想する人工音声の 0.2 秒後に ASE 二種類のうちいずれかの一つを付与することでシステムの確信度を伝達する条件。一つは音の長さが 0.5 秒で、その間の基本周波数が 250 Hz 一定のもの (Flat ASE)、もう一つは音の長さが 0.5 秒で、その間の基本周波数が 250 Hz から 100 Hz に様に減少するもの (Decreasing ASE) である (図 2)。筆者らの先行研究において、Decreasing ASE が付与された提示情報はシステムの低い確信度をユーザに伝達し、Flat ASE が付与された提示情報は、高い確信度をユーザに伝達できることがすでに示されている [10]。



“いちばん”という人工音声  
(提示時間: およそ0.45 sec)

Decreasing ASE  
(提示時間: 0.5 sec)

図 2：ASE と人工音声との関係

- **パラ言語条件**：人間らしい表現の一つの実装方法として、システムが提示する人工音声における発話速度およびイントネーションといったパラ言語情報を二通りの方法で修正することで人工音声と同時に確信度を表出する条件。一つは人工音声の発話速度を上昇させると同時にイントネーションを下がり調子にしたもの (例. 「いちばん!」と強く断定しているような口調)、もう一つは、発話速度を落とし、イントネーションを上がり調子にしたもの (例. 「いちばん?」と聞き返すような口調) である。前者は高い確信度

をユーザに伝達すること、また後者は低い確信度をユーザに伝達することを意図して設計された。

- **言語条件**：人間らしい表現のもう一つの実装方法として、システムが提示する人工音声の 0.1 秒後に二種類の接尾辞のうち一つを付与することで確信度を伝達する条件。一つは「です」という断定のニュアンスを示すもの、もう一つは「かもしれません」という逡巡のニュアンスを示すものである。前者の接尾辞の付与は高い確信度をユーザに伝達すること、また後者は低い確信度を伝達することを意図して設計された。

各実験条件内にて、ドライブ宝探しゲームの 20 回の試行のうち、高い確信度を意図した情報の表出をともなった人工音声は 10 回、低い確信度の表出をともなった人工音声は 10 回表出される。これら二種類の確信度表現が提示される順番については参加者間でカウンタバランスがとられるよう配慮した。

## 参加者

実験には 20 人の日本人の大学生が参加した (男性 15 人、女性 5 人; 21~28 歳)。これらの参加者は無作為に、ナビゲーションシステムの提示する情報と表出する確信度とが現実の状況に一致しているか否かについて、以下の二つのグループに配置された。

- **言動一致グループ (10 人)**：このグループの参加者は、提示情報と確信度とが現実の状況と一致した状態 (=言動が一致) のシステムから金貨が隠された盛り土の位置についての情報を伝達される。つまり、高い確信度をともなった情報の時には、その盛り土には 100%の確率で金貨が入っており、低い確信度をともなった情報の時には、その盛り土には 100%の確率で金貨が入っていない状況となる。
- **言動不一致グループ (10 人)**：このグループの参加者は、言動が不一致なシステムから情報を伝達される。つまり、高い確信度をともなった情報の時であっても、その盛り土には 50%の確率でしか金貨は入っておらず、逆に低い確信度をともなった情報の時でも、その盛り土には 50%の確率で金貨が入っている状況となる。

本実験は、全ての参加者が ASE 条件、パラ言語条件、言語条件の三つの実験条件を経験するように設

定されているため、最終的には2×3の混合計画となる(被験者間配置：二水準、言動一致/言動不一致グループ、被験者内配置：三水準、ASE/パラ言語/言語条件)。

## 手順

本実験は、ドライブ宝探しゲームにおける参加者の行動およびナビゲーションシステムに対する印象を把握することができるように筆者の一人が開発したオンライン実験システムにて実施された。参加者にはこのシステムのURLが送付され、そのURLにアクセスすると、まず実験への参加許諾書が現れる。それを許諾すると実験に関する教示情報が表示される。そして、ドライブ宝探しゲームで流される音声に類似した音声を再生することで参加者各自にとって聞きやすい音量にボリュームを調整をもらった後、三つの実験条件のうちのいずれかの条件にてドライブ宝探しゲームがスタートする。そしてそのゲームが終了すると、金貨の予想位置という情報を提示していたナビゲーションシステムに関する7段階リッカートスケール(最高点7点、最低点1点)で構成された合計18問の質問への回答を記入してもらう。その回答が終わり次第、残り二つの実験条件を同様の手順にて経験する。

本実験では、ゲーム終了時に回答を求められるリッカートスケールの合計点を参加者のシステムに対する参加者の「主観的印象点」(最高点126点、最低点18点)として扱う。この18問のリッカートスケールは、Love-Liking尺度7項目[12]と筆者らが独自で用意した11項目を加えたものによって構成された(表1、クロンバックの $\alpha$ 係数：0.86)。

表1：質問項目

1	このカーナビと一緒にいても、いつもと変わらない気持ちのまま。		
2	このカーナビは適応能力のあるシステムだと思う。		
3	このカーナビに責任のある仕事を任せてもよい。		
4	このカーナビの判断には全面的な信頼をおいている。		
5	このカーナビは色々な人に好かれる存在だと思う。		
6	知っているカーナビの中でもこれは最も好ましいものだ。		
7	このカーナビのことが好きで、自分に近い存在だ。		
8	このカーナビは人間らしい。		
9	このカーナビは良いサービスを提供できる。		
10	このサービスに満足である。		
11	もう一度このサービスを利用したい。		
12	このカーナビの間違いは許せない。		
13	このカーナビは礼儀正しい。		
14	このカーナビは信頼できる。		
15	このカーナビはあなたにとって役に立つ。		
16	このカーナビは憎めない存在だ。		
17	このカーナビと一緒にだと楽しい。		
18	このカーナビと一緒にだと疲れを感じる。		

## 仮説

先にも述べたように本実験では、「ASEによって確信度を表出する場合よりも人間らしい情報によって確信度を表出する場合の方が、言動不一致なシステムへの評価が低くなる」と仮定した。つまり、言動不一致グループにおいて、ASE条件よりも、パラ言語条件および言語条件におけるシステムに対する主観的印象点が低い場合、この仮説が検証されたといえる。

## 実験操作

本実験を行うにあたり、ASE/パラ言語/言語情報といった確信度を表出する情報の種類は、参加者のシステムに対する印象に影響を及ぼすものの、ゲームにおけるパフォーマンスやシステムの提示情報を受け入れるのか否かといったゲームにおける行動のものには影響を及ぼさないと考えた。そこでこの前提を確認するために、ドライブ宝探しゲームにおける20試行中において何枚の金貨を獲得できたのかという「獲得数」を各実験条件ごとに調査することで、確信度を表出する情報の種類と参加者のゲームにおけるパフォーマンスとの関係を確認することとした。さらに、10回の高確信度の表出および10回の低確信度の表出のそれぞれに対して参加者が何回システムの情報を受け入れたのかという「許諾数」を実験条件ごとに調査することで、確信度を表出する情報の種類と参加者の行動との関係を確認することとした。

まず「獲得数」に関して2×3混合計画(被験者間計画独立変数：参加者グループ(二水準：言動一致/言動不一致グループ)、被験者内計画独立変数：確信度を表出する情報の種類(三水準：ASE/パラ言語/言語条件)、従属変数：獲得数)による分散分析を行ったところ、被験者内計画の独立変数(確信度を表出する情報の種類)に関しての主効果には有意差は観察されなかった[F(2,36)=0.04, n.s.]。つまり確信度を表出する情報の種類は、ゲームにおける獲得数には影響を及ぼしていないことが確認された(表2)。

表2：参加者グループおよび確信度を表出する情報ごとの獲得数

	言動不一致	言動一致
ASE	6.8	11.6
パラ言語	8.3	12.7
言語	7	11.8

また「許諾数」に関して  $2 \times 3 \times 2$  混合計画（被験者間計画独立変数：参加者グループ（二水準：言動一致/言動不一致グループ）、被験者内計画独立変数その1：確信度を表出する情報の種類（三水準：ASE/パラ言語/言語条件）、被験者内独立変数その2：確信度（二水準：高/低確信度）、従属変数：許諾数）による分散分析を行ったところ、被験者内計画その1の独立変数（確信度を表出する情報の種類）に関しての主効果には有意差は観察されなかった [ $F(2,36)=0.03, n.s.$ ]. つまり、確信度を表出する情報の種類は、ゲームにおける参加者の許諾数にも影響を及ぼしていないことが確認された（表3）。

ここから、ASE/パラ言語情報/言語情報といった確信度を表出する情報の種類の違いは、参加者のゲームのパフォーマンスおよびゲームにおける行動そのものには影響を及ぼしていないことが確認されたため、これらの情報の種類の違いが純粹に参加者のシステムに対する主観的に印象に及ぼす影響について考察することができる。

表3：参加者グループ、確信度を表出する情報および確信度の高低ごとの許諾数

	言動不一致	言動一致
ASE 低確信度	5.4	2.4
ASE 高確信度	7.8	9.1
パラ言語 低確信度	5.2	1.9
パラ言語 高確信度	7.8	9.4
言語 低確信度	5.1	2.5
言語 高確信度	7.3	8.8

## 結果

各参加者グループおよび各実験条件における参加者のシステムに対する主観的印象点を図3に示す。言動一致グループ10人の参加者のうち、ASE条件における参加者の主観的印象点の平均は58.2点 ( $SD = 15.53$ )、パラ言語条件では75.1点 ( $SD = 9.27$ )、言語条件では68.1点 ( $SD = 11.85$ )であった。一方、言動不一致グループの10人の参加者のうち、ASE条件における主観的印象点の平均は60.0点 ( $SD = 11.40$ )、パラ言語条件では49.4点 ( $SD = 11.94$ )、言語条件では48.6点 ( $SD = 12.31$ )であった。

この主観的印象点に関して  $2 \times 3$  混合計画（被験者間計画独立変数：参加者グループ（二水準：言動一致/言動不一致グループ）、被験者内計画独立変数：確信度を表出する情報の種類（三水準：ASE/パラ言語/言語条件）、従属変数：主観的印象点）による分散分析を行ったところ、交互作用 [ $F(2,36)=11.50, p<.01, effect\ size: \eta^2=0.15$ ]および被験者間要因の主効果に有意差が観察された [ $F(1,18)=9.99, p<.01, effect$

$size: \eta^2=0.22$ ]. そこで、それぞれの独立変数に関する単純主効果に関する分散分析を行ったところ、参加者グループ間におけるパラ言語条件および言語条件に有意差が観察された[パラ言語条件： $F(1,18)=26.03, p<.01$ , 言語条件： $F(1,18)=11.71, p<.01$ ]. さらには、参加者グループ内における各実験条件間にも有意差が観察され[言動一致グループ： $F(2,36)=7.97, p<.01$ , 言動不一致グループ： $F(2,36)=4.48, p<.05$ ], LSD法による下位検定の結果、言動一致グループではパラ言語条件および言語条件の方がASE条件よりも高い評価を受けている一方 ( $MSe=90.4882, 5\% \text{ level}$ )、言動不一致グループではパラ言語条件および言語条件の方がASE条件よりも低い評価を受けていることが明らかになった ( $MSe=90.4882, 5\% \text{ level}$ ).

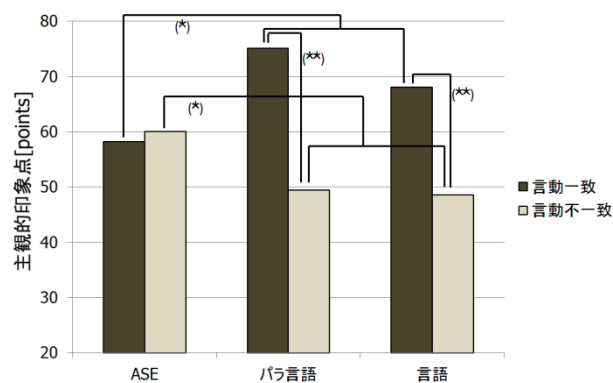


図3：参加者グループおよび実験条件ごとの主観的印象点

これより、言動不一致グループにおいて、ASE条件よりも、パラ言語条件および言語条件におけるシステムの印象点が低いことが実験的に実証されたため、「ASEによって確信度を表出する場合よりも人間らしい表現によって確信度を表出する場合の方が、言動不一致なシステムへの評価が低くなる」という我々の仮説が検証された。

さらに興味深いのが、パラ言語条件および言語条件における主観的印象点は実験グループ間にて有意に異なっていた一方で、ASE条件における主観的印象点は実験グループ間にてほとんど同じ値を示していたことである。この結果は、システムの言動一致状態がどのようなものであってもASEによる確信度表出は一樣に解釈されるという、ASEに対する解釈のロバスト性を示しているといえよう。

## おわりに

本実験の結果からは、言動一致グループにおいては、ASE条件よりもパラ言語条件および言語条件の

方がシステムに対する主観的印象点が高くなっているという現象も同時に観察することができた。このことから、パラ言語情報や言語情報といったいわゆる「人間らしい表現」の使用は排除する必要はなく、システムからの提示情報と確信度との関係を踏まえた上で、どのような確信度の表現がその状況に適しているのかといった選択の余地を与えているようにも見える。しかしながら、現在の技術水準では、常に正しい提示情報と確信度を表出できるシステムを実現することは非常に難しいと考えられる。よって、「システムがユーザに提示した情報」と「表出した確信度」とが現実の状況と一致しない状態は常に起こり得ることを踏まえた上で、ユーザに不快感を与えないような情報および確信度の提示方法を確立することは非常に重要な研究課題であるといえる。この点からも、本実験で得られた知見は非常に有用であり、さらには HAI/HCI/HRI の各分野においてその速やかな応用が期待できるものであるといえよう。

## 参考文献

- [ 1 ] Nass, C., and Brave, S. *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*, The MIT Press, MA, USA, 2005.
- [ 2 ] Cohen, M. H., Giangola, J. P., and Balogh, J. *Voice User Interface Design*, Addison-Wesley, MA, USA, 2004.
- [ 3 ] Google: 25% Of Queries From Android 2.0 Devices Use Voice Search: <http://techcrunch.com/2010/08/12/googles-hugo-barra-25-of-android-queries-are-voice-based/>
- [ 4 ] Benzeghibaa, M., De Moria, R., Derooa, O., Dupont S., Erbesa, T., Jouveta, D., Fissorea, F., Lafacea, P., Mertinsa, A., Risa, C., Rosea, R., Tyagia, V., and Wellekensa, C. Automatic speech recognition and speech variability: A review, *Speech Communication* 49, 10-11 (2007), 763-786.
- [ 5 ] Cai, H., and Lin, Y. Tuning Trust Using Cognitive Cues for Better Human-Machine Collaboration, In Proc. HFES2010, pp. 2437-2441(5).
- [ 6 ] Scholtz, J., and Bahrami, S. Human-robot interaction: development of an evaluation methodology for the bystander role of interaction. In Proc. 2003 IEEE SMC, IEEE (2003), pp. 3212-3217.
- [ 7 ] Kanda, T., Miyashita, T., Osada, T., Haikawa, Y., and Ishiguro H. Analysis of Humanoid Appearances in Human-robot Interaction, *IEEE Transactions on Robotics* 24, 3 (2008), 725-735.
- [ 8 ] Aronson, E., and Linder, D. Gain and loss of esteem as determinants of interpersonal attractiveness, *Journal of Experimental Social Psychology* 1, (2) 1965, 156-171.
- [ 9 ] Komatsu, T., Kurosawa, R., and Yamada, S. Difference between Users' Expectations and Perceptions about a Robotic Agent (Adaptation Gap) Affect Their Behaviors, In Proc. HRI 2011 workshop Expectations in intuitive human-robot interaction, published as PDF.
- [ 1 0 ] Komatsu, T., Yamada, S., Kobayashi, K., Funakoshi, K., and Nakano, M. Artificial Subtle Expressions: Intuitive notification methodology for artifacts, In Proc. CHI2010, ACM Press (2010), 1941-1944.
- [ 1 1 ] Funakoshi, K., Kobayashi, K., Nakano, M., Yamada, S., and Komatsu, T. Non-human like spoken dialogue: A design perspective, In Proc. SIGDIAL 2010, ACM Press (2010), 176-184
- [ 1 2 ] Rubin, Z. Measurement of romantic love, *Journal of Personality and Social Psychology* 16, (2) 1970, 265-273.