

# 対話の低速化と Artificial Subtle Expression の適用による 発話衝突の抑制

## Reducing Speech Collisions by Using an Artificial Subtle Expression in a Decelerated Spoken Dialog

船越 孝太郎<sup>1\*</sup> 小林 一樹<sup>2</sup> 中野 幹生<sup>1</sup> 小松 孝徳<sup>2</sup> 山田 誠二<sup>3</sup>  
Kotaro Funakoshi<sup>1</sup> Kazuki Kobayashi<sup>2</sup> Mikio Nakano<sup>1</sup> Takanori Komatsu<sup>2</sup> Seiji Yamada<sup>3</sup>

<sup>1</sup> (株)ホンダ・リサーチ・インスティテュート・ジャパン

<sup>1</sup> Honda Research Institute Japan Co., Ltd.

<sup>2</sup> 信州大学

<sup>2</sup> Shinshu University

<sup>3</sup> 国立情報学研究所/総合研究大学院大学

<sup>3</sup> National Institute of Informatics / The Graduate University for Advanced Studies

**Abstract:** This paper describes the experiments in which participants engaged in reservation tasks with a spoken dialog system coupled with an interface robot using a blinking light expression. The blinking light expression is designed as a subtle expression to intuitively notify a user about a robot's internal states (such as processing or busy) for the sake of reducing speech collisions as consequences of turn-taking failures due to end-of-turn misdetection. Speech collisions harm smooth speech communication and degrade system usability. Two experimental factors were set up: the blinking light factor (with or without a blinking light) and the reply speed factor (moderate or slow reply speed), resulting in four experimental conditions. The results suggest that the blinking light expression can reduce speech collisions and improve user impression, and surprisingly that users do not care about slow replies.

## 1 はじめに

話者交替 (turn-taking) の失敗, 特に対話する二者が同時に発話してしまう「発話衝突」は, 対話エージェントとユーザ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. 発話衝突が起き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 [1].

発話衝突の原因の一つは, 対話エージェントがユーザの発話末を誤検出することである. これに対する一つのアプローチは, 機械学習によりマルチモーダル情報を利用した高精度な発話末検出器を構成しようとするものである. もう一つのアプローチは, ユーザの音声信号が途絶えてからシステムが応答するまでの時間を単に長くするというものである. ある程度の間待っても次の音声がおこなければ, ユーザの発話は終了したと解釈できる.

多くの従来研究が一つ目のアプローチを選んでいる. これはユーザがより応答性の高いシステムを好むという前提に基づいている. この前提は, システムの能力が人間と同等の水準にあれば真であろう. しかしながら, システムの能力が人間を下回る場合には, 高い応答性は重要ではないばかりか有害でさえあり得ると我々は考える. 例えば, 相槌を返すシステムに対してユーザが感じる親密さが, 全体的にはかなりよい性能であってもわずかなエラーの為に, 相槌を返さないシステムに対して感じる親密さよりも劣ったという報告がある [2]. 音声対話技術は向上を続けているが, 依然として人間の水準には届いていない. 一方, 二つ目のアプローチは単純なので, 対話システムの構築も容易であるし, 構築したシステムはユーザや環境の変化に対して安定した振る舞いを示す. さらに, このアプローチでは応答までに余裕があるので, より計算量のかかる音声処理を導入したり, 資源の限られた小さなデバイスの上にシステム導入することができる. そこで本研究では二つ目のアプローチ, つまり対話を「低速化」するアプローチの可能性を追求する.

\* (株)ホンダ・リサーチ・インスティテュート・ジャパン  
〒351-0188 埼玉県和光市本町 8-1  
E-mail: funakoshi@jp.honda-ri.com

この「低速化」アプローチには、応答性の悪化の他にもう一つの問題が存在する。それは、システムの応答が遅いことでユーザが繰り返しや言い換えといった不必要な発話をしてしまい、それが原因で発話衝突が起きるという問題である。発話衝突を減らそうとして別の形で発話衝突を増やしてしまつては意味がない。これに対し我々は以前の実験で、明滅光源による非言語フィードバックを用いることで、発話衝突の原因となるユーザの不必要な繰り返し発話を抑制できる可能性を示した [3]。しかしながら、実験は人間のオペレータがユーザに隠れてシステムを操作する Wizard-of-Oz (WOZ) 方式の模擬対話で行っており、全自動の音声対話システムを用いてはいなかった。そして、明滅光源により繰り返し発話を抑制できることは示したが、実際に発話衝突が減ることを示してはいない。また、明滅光源の導入によりシステムの印象が改善できることを示したが、低速化による印象の悪化がどれくらいのものであり、それが明滅光源によりどれくらい補償されているのかまでは示していなかった。

本稿では、48人の参加者が明滅光源を備えたインターフェースロボットと連結された音声対話システムとのホテル予約課題に取り組んだ実験の結果を示す。我々が用いた明滅光源は次節で述べるように Artificial Subtle Expression (ASE) の考えに基づいており、システムの内部状態（処理中かどうか）をユーザに直感的に伝える。実験の結果、対話の低速化と明滅光源の導入は共に発話衝突の抑制に有効であることが示された。また、明滅光源の導入により、ユーザが抱く印象が改善する傾向が認められた。一方、予想に反して、対話の低速化による評価の悪化は認められなかった。これらのことから、対話の低速化は、音声対話システム設計に関する有効なアプローチであると我々は考える。

## 2 ASE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 [4, 5]。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 [6]。

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、いくつかの研究が行われている [7, 8]。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 [9]。山田らは周波数が変化するビーブ音などにより、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ビーブ音を実装したとき、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。また、同じくビーブ音を用いて、ロボットがユーザに与えるアドバイスに対してロボット自信が抱く自信の強弱を伝達できることを示すとともに、人工物らしい表現方法を活用した subtle expression という概念を Artificial Subtle Expression (ASE) として提案している [10]。

我々は以前の実験で、ロボットの内部状態（処理中かどうか）をユーザに直感的に伝える ASE として明滅光源を採用し、明滅光源が発話衝突を抑制できる可能性を示した [3]。しかしながら、あくまで WOZ によるしりとり対話という限定された状況で繰り返しの数が減ることを通じて間接的に示しただけであり、発話衝突が実際に減ることを確認したわけではなかった。本稿では、より実地的な課題指向対話を行う全自動の音声対話システムを用いて、発話衝突の抑制に ASE としての明滅光源が有効であることを直接的に示す。

## 3 実験

会話ロボットと対話してホテルの予約をするという課題設定で実験を行った。音声対話システム、ロボット、明滅光源、実験条件及び手法について以下で説明する。

### 3.1 音声対話システム

ホテル予約ドメインで動作する音声対話システムを構築した。対話システムは予約内容を表示する液晶モニタと胸部に発光ダイオードを備えたインターフェースロボットを備えている（図1参照）。参加者の発話の



図 1: 液晶モニタとロボット

音声認識には Julius<sup>1</sup>を用い、ロボットの発声には市販の音声合成ソフト (NTT-IT FineVoice) を用いた。ユーザ発話の言語理解には KNOLU [11] を用いた。液晶モニタはシステムが予約の詳細を読み上げる時間を短くして、システムの発話を聞き取る際の参加者の認知的負荷を軽減する目的で使用した。従って、システムがユーザに確認を求める時にだけ予約内容の詳細をモニタ上に表示し、それ以外の場合は最初に歓迎メッセージを表示する他は何も表示しない。

システムの音声言語理解の能力は可能な限り高くなるように事前に調整した。例えば、Julius が使用する言語モデルと辞書は、事前に同様の課題で収集したデータ (47 人からの 4223 発話, 140 分) から作成した。

Julius の各種パラメータの値にはデフォルトの値を使用した。Julius は、入力音声の終端を検出してから 400 ミリ秒後に、認識結果の単語列を対話システムに出力する。我々が構築した対話システムは Julius からの入力を受け取った後、さらに  $\alpha$  ミリ秒の間入力を待ち受ける。この入力音声の終端検出から待ち受け終了までの時間の長さ ( $400 + \alpha$  ミリ秒) は実験要因の一つであり、以降「待機時間」とよぶ。もしこの待機時間の間に更なる入力を受け取ったときは、システムは同じ長さの時間だけ待機する。この待機時間のうちに入力を受け取らなかった場合は、システムは全ての認識結果の単語列を受け取った順に連結し、それに対して解釈を行い、応答する。

### 3.2 ロボットと明滅光源

ロボット (WowWee RS-Media) と明滅光源 (赤色・直径 4 ミリの発光ダイオード) は、[3] と同じものを用いた。明滅パターンも同じである (1/30 秒の等間隔での明滅)。システムは音声信号を検出すると発光ダイオードの明滅を始め、応答を開始した時点で明滅を止める。

明滅光源の基本的な機能は GUI で用いられる砂時計アイコンのそれに近い。但し、大きな違いは、GUI では砂時計アイコンを表示している間に来る全ての入力

<sup>1</sup><http://julius.sourceforge.jp/>

を無視すればよいが、対話システムではダイオードを明滅させている間に来る音声も受け入れなければならないことである。我々が明滅光源を用いて行おうとしていることは、発話衝突の原因となる繰り返しや言い換えのような発話 (以降「補足発話」とよぶ) だけを抑制することである。補足発話は殆どの場合無視して問題ないのだが、実際に発話衝突が起きた時には意図的な割り込み発話 (バージン) と区別することが難しい。一方、対話システムはバージン発話を無視できない。そこで、不必要な補足発話だけを抑制する。

### 3.3 実験条件と参加者

実験要因として、明滅光源要因 (あり/なし) と応答速度要因 (中速/低速) の二つの要因を設定した。これらを組み合わせると、次の四つの実験条件になる。

**条件 A:** 明滅光源あり, 低速

**条件 B:** 明滅光源なし, 低速

**条件 C:** 明滅光源あり, 中速

**条件 D:** 明滅光源なし, 中速

48 人の実験参加者を四つの実験条件にランダムに振り分けた。但し、性別と年代は条件間でなるべく均等に分布するようにした。参加者の年齢は 18 歳から 55 歳で、平均年齢は 30.9 歳 (SD = 10.2) であった。

応答速度 (ユーザの発話が終了してからロボットが発話するまでの時間) は前述の待機時間 (入力音声途切れた後に対話システムが後続発話を待ち受ける時間) に依存する。Shiwa らが行った実験では、会話ロボットの最適な応答タイミングは発話終了直後ではなく約 1 秒後であるという結果が得られている [12]。そこで、実際の応答までには処理のための遅延が加わるため、中速条件での待機時間として 800 ミリ秒 ( $\alpha = 400$ ) を設定した。また、低速条件での待機時間として 4000 ミリ秒 ( $\alpha = 3600$ ) を設定した。

### 3.4 手法

実験は一人ずつ静かな部屋の中で行った。参加者は部屋に入ると図 1 に示すように机の前の椅子に座る。椅子は動かないように固定した。システムの音声と参加者の音声を分離して収録するため、参加者にはヘッドフォンを着用させた。

実験者は参加者に目の前のロボットと話しをしてホテルの部屋を予約するように指示した。参加者には全員同じ課題五つを同じ順序で与えた。各課題では参加者は 1 つから 3 つの部屋を予約する。明滅光源の存在と意味は参加者には説明しない。指示を与えた後に実

験者は参加者を残して退去し、対話システムを始動させた。参加者はロボットからの話しかけによって予約対話を始める。各課題は最大3分間に制限し、制限時間内に課題が完了しなかった場合は実験者が対話を中断した。参加者は五つの課題に休憩無しで取り組んだ。

課題終了後、参加者は対話とロボットそれぞれの印象とその他の質問についてのアンケートに答え、実験参加に対する謝礼を受け取って退室した。

## 4 結果

先に述べたように、繰り返し、言い換えなど、発話衝突の原因となる全ての発話を補足発話とよぶ。我々が以前行った実験 [3] では特に繰り返しの数を数えて分析を行ったが、それは参加者の発話が1語だけになる単純なしりとり対話であったため可能であった。補足発話の厳密な定義を与えることは容易ではないので、今回の実験で収録した対話データの中でその数を数えることは難しい。そこでここでは、補足発話の数は数えずに発話衝突の数だけを数えて分析を行う。

### 4.1 実際の応答速度

全ての対話を発話毎のタイムスタンプ付きのテキストに書き起こした。このタイムスタンプをもとに実際の平均応答速度を算出したところ、低速条件（条件 A と B）で 4.53 秒（SD = .75）、中速条件（条件 C と D）で 1.42 秒（SD = .43）であった。

### 4.2 課題達成数

A, B, C, D の各条件での平均課題達成数は、それぞれ 4.00 (SD = 1.28), 3.83 (SD = 1.34), 3.83 (SD = 1.27), and 4.33 (SD = .98) であった。これら 4 条件間で課題達成数に関する 2 要因 2 水準の分散分析を行ったが、有意な差は見られなかった。

### 4.3 発話衝突

参加者がまだ話し続けているのにロボットが話し始めてしまった場合の発話衝突だけを数えた。勿論、まだロボットが話しているのに参加者が話し始めることで起こる発話衝突もあるが、これらは本研究の対象ではないのでここでは数えない。つまり、先に述べたバージョンの場合や参加者が話し始めるのが単に早すぎた場合などは数えない。同様に、システムが発話中に参加者が行う相槌もここでは数えない。

表 1: 被験者毎の発話衝突の回数

条件	各条件内での参加者番号												計
	1	2	3	4	5	6	7	8	9	10	11	12	
A	0	1	2	0	0	0	0	0	2	0	0	0	5
B	1	0	0	3	0	1	0	2	1	1	1	1	11
C	2	5	0	0	5	6	3	3	1	14	5	1	45
D	7	0	2	0	6	6	0	2	0	5	0	2	30

表 2: 低速条件で発話衝突が起きた参加者数の分割表

条件	発話衝突あり	発話衝突なし	計
明滅光源あり (A)	3	9	12
明滅光源なし (B)	8	4	12

発話衝突は実験中に実験補助スタッフによって数えられ、また実験後に別の研究補助スタッフによって音声聞きながら書き起こし上でも数えられた。著者のうちの一人が二つの計数結果を確認・総合し、最終的な計数結果とした。参加者毎の発話衝突の回数を表 1 に示す。

この結果に対して、まず 2 要因 2 水準の分散分析を行った。その結果、応答速度要因、すなわち低速（条件 A と B）と中速（条件 C と D）の間に有意差が認められた ( $F_{1,44} = 4.06, p < 0.005$ )。これは、ユーザの発話が終了してからシステムの応答までの間隔を長くすることで、実際に発話衝突を削減できることを示している。

この分散分析の結果では、明滅光源要因に関して有意な差は認められなかった。しかしこの結果は、明滅光源が有効となるのは主に低速条件であることと、その低速条件では発話衝突の数が中速条件よりも少ないことを考えると妥当である。

そこで、次に低速条件（条件 A と B）だけに限って、発話衝突が起きた参加者の数についてフィッシャーの正確確率検定（片側）を行った。発話衝突が起きた参加者の数の分割表を表 2 に示す。検定の結果、有意差が認められた ( $p < 0.05$ )。これは、低速化した対話において明滅光源が発話衝突の抑制に有効であることを示している。

### 4.4 対話の印象

対話の印象に関する形容詞対を用いたアンケートの結果を分析した。表 3 に対話の印象についての形容詞対に対する参加者の評価結果を示す。評価には 7 段階のリッカート尺度（1 がネガティブな形容詞に対する強い同意、7 がポジティブな形容詞に対する強い同意を表す）を用いた。表中では、各形容詞対毎に、4 条件の中で最も高いスコアに下線を引いてある。

因子分析（主因子法・バリマックス回転）を行い、スクリープロットから 5 因子を抽出した。この因子を形

表 3: 対話の印象についての形容詞対と評価結果

実験条件		A	B	C	D
ポジティブ	ネガティブ	平均 (SD)	平均 (SD)	平均 (SD)	平均 (SD)
気軽な	重厚な	3.42 (1.24)	3.33 (1.07)	3.50 (1.31)	4.00 (1.54)
円滑な	ぎこちない	3.25 (1.29)	2.58 (1.00)	3.17 (1.59)	3.00 (1.35)
感じのよい	感じのわるい	4.42 (1.00)	4.08 (0.79)	4.08 (1.00)	4.42 (1.08)
面白い	つまらない	4.58 (0.67)	4.17 (1.03)	3.83 (1.95)	4.17 (1.11)
リラックスする	緊張する	3.25 (1.06)	3.25 (1.48)	3.42 (1.16)	3.75 (1.48)
安心な	不安な	3.33 (1.37)	2.42 (1.24)	2.75 (1.06)	2.58 (1.08)
暖かい	冷たい	3.25 (1.48)	3.25 (0.97)	3.67 (0.98)	3.67 (1.15)
愉快的な	不愉快的な	4.33 (0.89)	3.67 (1.15)	3.17 (1.03)	4.25 (0.97)
ゆったりとした	せわしない	4.75 (1.06)	4.92 (1.51)	4.83 (1.27)	4.67 (0.89)
うちとけた	かたくなるしい	3.08 (1.16)	2.67 (0.89)	2.67 (1.44)	3.17 (1.03)
明るい	暗い	3.42 (1.00)	3.00 (0.85)	3.17 (0.94)	3.67 (0.78)
わかりやすい	わかりにくい	4.00 (1.76)	3.58 (1.51)	4.17 (1.59)	4.50 (0.80)
好きな	嫌いな	4.42 (0.67)	3.75 (0.87)	4.08 (1.16)	4.33 (0.78)
得意な	苦手な	3.75 (1.22)	3.17 (1.11)	3.17 (1.19)	3.58 (1.24)
おだやかな	いらいらする	4.33 (1.37)	3.08 (1.56)	3.83 (1.95)	3.75 (0.87)
興味深い	退屈な	4.58 (1.24)	4.25 (1.14)	4.33 (1.67)	4.67 (1.30)
元気がでる	疲れる	3.42 (0.67)	2.92 (1.00)	3.00 (0.85)	3.25 (0.87)
落ち着いた	落ち着かない	4.50 (1.31)	3.50 (1.38)	3.83 (1.11)	3.58 (0.79)

容詞対に基づいて解釈し、それぞれ「親和性因子」、「安心性因子」、「娯楽性因子」、「リラックス因子」、「温和性因子」と名付けた。回帰法によって因子得点を求め、条件間で比較を行った。2 要因 2 水準の分散分析を行ったところ、2 番目の因子に関して明滅光源要因について有意傾向が認められ ( $F_{3,44} = 3.53, p = .07$ )、明滅光源を用いることでユーザに安心な印象を与えられる可能性が示唆された。

#### 4.5 ロボットの印象

ロボットの印象を分析した。分析方法は対話の印象の分析と同じである。表 4 にロボットの印象についての形容詞対に対する参加者の評価結果を示す。

ロボットの印象についても 5 因子が抽出され、それぞれ「親和性因子」、「信頼性因子」、「活動性因子」、「積極性因子」、「恥ずかしがり因子」と名付けた。4 条件間の分散分析の結果、5 番目の因子について応答速度要因について有意傾向が認められ ( $F_{3,44} = 3.39, p = .07$ )、低速化することでロボットに恥ずかしがりな印象が与えられる可能性が示唆された（ここでの「恥ずかしがり」は「恥知らず」と対比されておりポジティブな意味合いを持つ）。

#### 4.6 システムの実用性

参加者に対話システムを二つの指標で評価させた。一つは「システムを便利だと思うか（便利度）」で、もう一つは「システムをまた利用したいと思うか（利用希望度）」である。評価には印象評価と同じ 7 段階のリッカート尺度を用い、数値が大きいほど、便利だ／利用したいと思うこと表す。

便利度についての平均スコアは、各 4 条件でそれぞれ、3.50 (SD = 2.02), 3.17 (SD = 1.53), 3.17

表 4: ロボットの印象についての形容詞対と評価結果

実験条件		A	B	C	D
ポジティブ	ネガティブ	平均 (SD)	平均 (SD)	平均 (SD)	平均 (SD)
積極的な	消極的な	3.67 (0.78)	3.58 (0.79)	3.92 (1.16)	4.17 (1.19)
人のよい	人のわるい	4.08 (1.08)	3.92 (1.31)	4.17 (0.94)	4.33 (0.78)
なまいきでない	なまいきな	4.75 (0.87)	4.17 (1.59)	4.42 (0.90)	5.25 (1.36)
ひとつつつこい	近づきかた	3.08 (1.08)	3.42 (1.31)	3.50 (1.09)	4.00 (0.95)
かわいらしい	にくらしい	4.17 (0.39)	4.17 (0.58)	4.17 (0.83)	4.08 (0.79)
心のひろい	心のせまい	4.17 (0.58)	3.75 (0.75)	3.92 (0.79)	4.08 (0.79)
社交的な	非社交的な	3.75 (1.22)	3.50 (1.17)	3.33 (1.15)	3.92 (0.90)
責任感のある	責任感のない	4.42 (1.08)	4.17 (1.34)	4.08 (1.24)	4.50 (1.00)
慎重な	軽率な	5.42 (1.08)	4.92 (0.90)	4.92 (1.24)	4.50 (1.31)
恥ずかしがりの	恥知らずの	4.08 (0.51)	4.08 (0.51)	3.83 (0.72)	3.83 (0.39)
重厚な	軽薄な	5.00 (1.28)	4.67 (1.07)	5.00 (1.35)	4.50 (1.17)
うきうきとした	沈んだ	3.50 (0.52)	3.67 (1.07)	3.33 (0.78)	4.00 (1.04)
堂々とした	卑屈な	4.92 (1.51)	5.00 (1.21)	4.75 (0.97)	4.83 (0.72)
感じのよい	感じのわるい	4.25 (1.06)	3.67 (1.23)	3.58 (0.90)	4.17 (0.72)
分別のある	無分別な	4.67 (1.23)	4.33 (1.07)	3.92 (1.38)	4.33 (0.78)
親しみやすい	親しみにくい	3.75 (1.29)	3.33 (1.50)	3.17 (1.40)	4.42 (1.00)
意欲的な	無気力な	3.83 (0.23)	3.58 (0.67)	3.25 (1.48)	4.17 (0.94)
自信のある	自信のない	4.50 (1.00)	4.25 (0.87)	4.00 (0.95)	4.67 (0.89)
気長な	短気な	4.42 (0.79)	4.92 (1.31)	4.00 (1.04)	4.33 (1.07)
親切的な	不親切的な	3.67 (1.30)	4.08 (1.44)	3.33 (1.15)	4.00 (1.04)

(SD = 1.47), 3.92 (SD = 1.62) であった。利用希望度についての平均スコアは、各 4 条件でそれぞれ、3.58 (SD = 1.73), 2.58 (SD = 1.31), 2.83 (SD = 1.34), 3.42 (SD = 1.56) であった。分散分析を行ったが、いずれの指標についても有意差は認められなかった。

#### 4.7 ユーザの評価についての考察

印象評価の結果、明滅光源は対話に関してポジティブな印象をユーザに与えることが示唆された。更に、条件 A と B の間で形容詞対及び実用性評価の平均スコアを比較すると、条件 A に高いスコアが与えられている傾向が読み取れる。これらのことから、明滅光源の導入が発話衝突の抑制を通じてユーザの印象改善に効果があることが伺える。

しかしながら、条件 C と D の間のスコアを比較すると、条件 D に高いスコアが与えられる傾向が読み取れる。このことから、今回使用した明滅光源には好ましくない効果も含まれており、発話衝突の抑制に有効でない状況では、かえって明滅光源が対話システムの評価を悪化させているのではないかと推測される。今後、明滅光源の発光パターンなどを改良することで、この負の要素を取り除くことができないうか検討したい。

一般に HCI においてはシステムの応答はすぐに返すべきとされており、[12] でも GUI を用いた実験でそれを支持する結果が示されていた。しかし今回の実験では、対話を低速化することによる評価の低下は、印象評価においても実用性評価においても認められなかった。このことは、一般的な HCI 設計の指針とは異なり、音声対話においては応答を遅延させることが必ずしも悪くないこと、つまり人間のように素早く応答するシステムを目指すよりも、対話を遅くすることで得られるメリット（より高度な処理の導入による対話理解精



度の向上など)を追求する方が有意義であることを示唆している。

## 5 おわりに

本稿では、48人の参加者が発光ダイオードを備えたインターフェースロボットと連結された音声対話システムとのホテル予約課題に取り組んだ実験の結果を示した。実験の目的は、Artificial Subtle Expression (ASE) [10]として実装された明滅光源が、音声対話を阻害する発話衝突を抑制できることを示すことである。実験では、明滅光源要因(光源あり/なし)と応答速度要因(中速/低速)の二つの要因を設定した。分析により、システムの応答を遅くすることで対話を低速化させると発話衝突が減少し、さらに明滅光源を導入することで発話衝突が減少したことが統計的有意性をもって示された。アンケートの分析結果からは、明滅光源が参加者に安心感を与えることが示唆された。

また、対話を低速化させることはシステムに対するユーザの評価に悪影響を及ぼすと予測したが、それに反して分析結果からそのような事実は確認できなかった。これは、対話の低速化が音声対話システム設計における有効なアプローチであり得ることを示す一つの証拠と考える。

## 参考文献

- [1] M. Nakano, Y. Nagano, K. Funakoshi, T. Ito, K. Araki, Y. Hasegawa and H. Tsujino: "Analysis of user reactions to turn-taking failures in spoken dialogue systems", Proc. SIGdial 2007 (2007).
- [2] N. Kitaoka, M. Takeuchi, R. Nishimura and S. Nakagawa: "Response timing detection using prosodic and linguistic information for human-friendly spoken dialog systems", J. The Japanese Society for AI, **20**, 3, pp. 220–228 (2005).
- [3] K. Funakoshi, K. Kobayashi, M. Nakano, S. Yamada, Y. Kitamura and H. Tsujino: "Smoothing human-robot speech interactions by using a blinking-light as subtle expression", Proc. ICMI 2008, pp. 293–296 (2008).
- [4] A. Kendon: "Some functions of gaze direction in social interaction", Acta Psychologica, **26**, pp. 1–47 (1967).
- [5] A. Kendon: "Do gestures communicate?", A Review. Research in Language and Social Interaction, **27**, 3, pp. 175–200 (1994).
- [6] W. Rogers: "The contribution of kinesic illustrators towards the comprehension of verbal behavior within utterances", Human Communication Research, **5**, pp. 54–62 (1978).
- [7] C. Bartneck and J. Reichenbach: "Subtle emotional expressions of synthetic characters", Int'l J. Human-Computer Studies, **62**, 2, pp. 179–192 (2005).
- [8] H. Prendinger, J. Mori and M. Ishizuka: "Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game", International Journal of Human-Computer Studies, **62**, 2, pp. 231–245 (2005).
- [9] S. Yamada and T. Komatsu: "Designing simple and effective expression of robot's primitive minds to a human", Proc. 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2614–2619 (2006).
- [10] 小松, 山田, 小林, 船越, 中野: "Artificial Subtle Expressions: 人工物のための直感的情報伝達手法の提案", HAI シンポジウム 2009 (2009).
- [11] 船越, 中野, 長谷川, 辻野: "概念体系を中心とした自然言語インターフェース構築フレームワーク", 人工知能学会論文誌, **23**, 6, pp. 437–446 (2008).
- [12] T. Shiwa, T. Kanda, M. Imai, H. Ishiguro and N. Hagita: "How quickly should communication robots respond?", Proc. HRI 2008, pp. 153–160 (2008).