

ロボットの対話インタフェースにおける 動作に対する発声遅延の効果

山本 倫也[†] 渡辺 富夫^{†‡}

[†]岡山県立大学情報工学部

[‡]科学技術振興機構

1 はじめに

人と人の対面コミュニケーションでは、単に言葉によるバーバル情報だけでなく、音声に対するうなずきや身振り・手振りなどのノンバーバル情報が相互に同調して、対話者同士が互いに引き込み合うことでコミュニケーションしている [1]. この身体性の共有が、一体感を生み、人との関わりを実感させている.

近年、音声認識および知能情報処理の技術が飛躍的に向上し、インタラクション指向のロボットなど、物理的に身体をもち、人と同じ空間を共有して関わる身体的インタラクションシステムが開発されている. これらのインタフェースでは、高精度で信頼できるシステムの実現だけでなく、身体の働きを活かし、人とロボットの本質的なインタラクション支援と、円滑なコミュニケーションの実現が望まれている [2].

ロボットとのインタラクション支援においても、人と人のインタラクションと同様に、バーバルとノンバーバル情報の生成タイミングのわずかな違いが、異なるコミュニケーション効果をもたらさうる. したがって、単にロボットの動作と発声を同時に生成するのではな

く、人のインタラクション特性に基づいて、動作と発声の生成タイミングを制御する必要がある.

そこで本研究では、人の典型的な関わり行動である、あいさつを合成的に解析し、人とロボットのインタラクション (図 1) において、コミュニケーション動作に対して発声タイミングを適度に遅延させることで、自然・好き・丁寧など、好ましいコミュニケーション効果が得られることを明らかにしている. さらに、身体をもつロボットのインタラクション生成において、動作に対する発声遅延を積極的に活用できることを示している.

2 発声遅延の合成的解析

2.1 あいさつ合成システム

既に著者らは、人と人とのあいさつインタラクションの特性をモーションキャプチャシステムを用いて分析し、動作に対して発声が遅延することを明らかにしている [3]. これを踏まえて、コミュニケーション動作と発声のタイミングをマウス操作で自由に制御できるあいさつ合成システムを開発した. システム構成は図 2 の通りで、声かけあいさつの身体動作 (うなずき) を生成する USB 制御型「うなずき君¹」、女子学生の録音音声を再生するスピーカ、これらを制御する PC とソフトウェア、PC に接続したマウスから構成されている. USB 制御型「うなずき君」は、市販品に USB コントローラ (Cypress 製 CY7C63001A-PC) を組み込み、独自に開発した.

2.2 方法

あいさつ合成システムを用いて被験者実験を行い、発声遅延の大きさがどのようなコミュニケーション効果を与えるかを合成的に解析した. 実験では、図 2 のあいさつ合成システムを用いて、「こんにちは」「よろしくお願ひします」「さようなら」の 3 種類のあいさつに

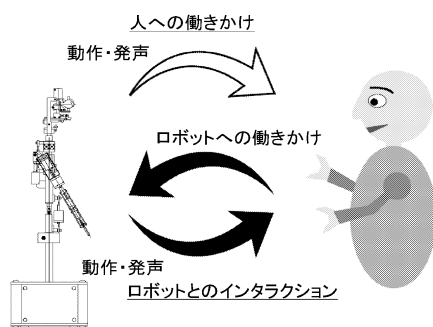


図 1: 人とロボットのインタラクション

Time Lag Effects of Utterance to Communicative Actions on Conversational Interface of Robots

Michiya Yamamoto[†] and Tomio Watanabe^{†‡}

[†]Faculty of Computer Science and System Engineering, Okayama Prefectural University

[‡]CREST of JST

¹ 音声リズムを解析して身振り手振りで応える身体的インタラクション玩具. 著者の渡辺らが開発した iRT (音声から身体動作を自動生成する技術) を導入している.

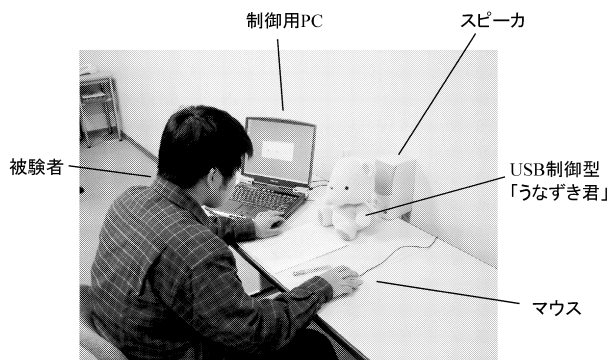


図 2: あいさつ合成システム

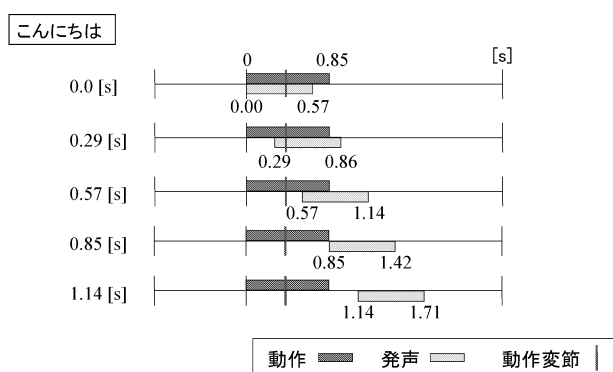


図 3: 合成した遅延タイミング

対して、「うなずき君」の動作時間 0.85 秒とあいさつ発声時間の平均値を組み合わせ、発声遅延の時間を 0 秒から 1.14 秒の間で変化させた 5 パターンを合成した。図 3 は「こんにちは」の例である。

被験者には、まず「こんにちは」の 5 パターンを「自然」「丁寧」「好き」を基準に順位付けするよう指示し、同時に、それぞれがどれくらい「自然」か、「丁寧」か「好き」かを 7 段階で評価させた。その後、「よろしくお願ひします」の 5 パターン、「さようなら」の 5 パターンも同様に評価させた。

この際、制御用 PC のソフトウェアに「A」「B」「C」「D」「E」の文字を書いた 5 つの操作ボタンを配置し、マウスでクリックして何度でも各パターンを試せるようにした。実験中は、「うなずき君」を見ながらクリックするよう指示した。各パターンにどのような差があるかは伝えなかった。被験者は、18~25 歳の男女学生各 25 名、計 50 名であった。

2.3 結果

「こんにちは」で 1 位に選ばれた発声遅延を図 4 に示す。「自然」と「好き」では、人の平均あいさつとほ

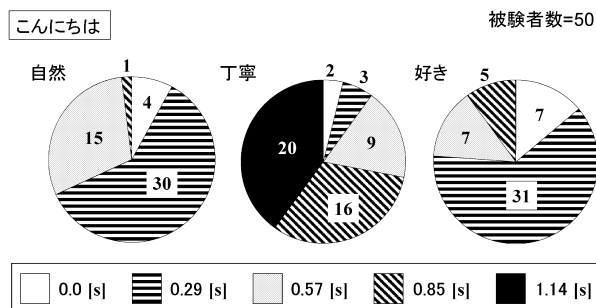


図 4: 1 位に選ばれた発声遅延

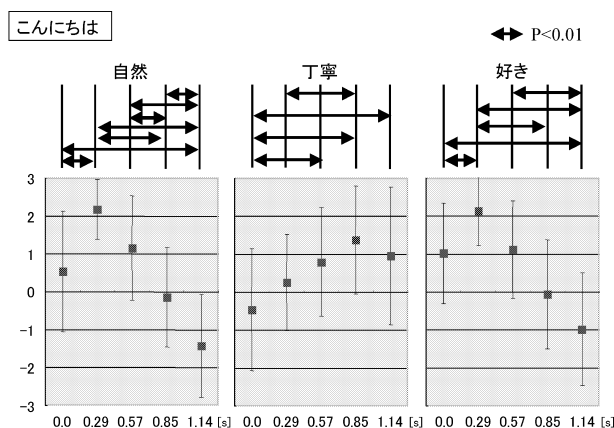


図 5: 発声遅延の 7 段階評価

ぼ同じ 0.29 秒のパターンが最も支持された。「丁寧」では、より発声遅延が大きいパターン、つまり 0.57 秒~1.14 秒の遅延を伴うパターンが好まれた。発声遅延に対する評価は、他のあいさつでも同様であった。

「こんにちは」の 7 段階評価の結果を図 5 に示す。図中の矢印は、この結果を Friedman の分散分析法で検定した結果で、有意水準 1% の有意差を示す。他のあいさつでも、評価結果に有意差が認められ、「自然」と「好き」は平均的な 0.29 秒のパターンが、「丁寧」はより遅いパターンが高く評価された。

2.4 考察

ロボットから人へのあいさつで、発声遅延の大きさが変わること、そのコミュニケーション効果が変わる結果が示された。遅延なしの 0 秒のパターンと比較して、0.29 秒の発声遅延を伴うあいさつタイミングが「自然」であり「好き」と評価され、また、0.85 秒の発声遅延を伴うあいさつタイミングが「丁寧」に見えるなど、発声遅延を伴うあいさつが高く評価された。動作終了と発声開始に大きな間がある 1.14 秒のパターンは、評価が非常に低かった。

以上より、コミュニケーション動作に対して発声タイミングを適度に遅延させることで自然、好き、丁寧など、好ましいコミュニケーション効果が得られることが分かった。

3 インタラクションの合成的解析

3.1 あいさつインタラクション合成システム

うなずき君に「こんにちは」と音声入力すれば、「こんにちは」と動作と録音音声で返答する、あいさつインタラクション合成システムを開発した。システムは、2.1節のあいさつ合成システムに、手持ちマイク (IBM製 ViaVoice 用手持ちマイク) を接続した構成である。

インタラクションの生成では、手持ちマイクの入力信号を 40Hz でサンプリングした後、OFF 区間の音圧レベルの 3 倍を ON 区間とし、その後、短い OFF 区間を ON 区間に交換するために、短い ON 区間を延長するハングオーバー処理を行った。これらの処理により、声かけあいさつの発声開始と発声終了を正確に検出するとともに、応答あいさつの合成精度を向上させている。

3.2 方法

あいさつインタラクション合成システムを用いて被験者実験を行い、インタラクションの時間とタイミングが与えるコミュニケーション効果を合成的に解析した。

実験では、図 6 に示す①～⑤のパターンで「こんにちは」のあいさつを合成した。各パターンは、音声対話システムを想定したもので、「うなずき君」の動作時間 0.85 秒と「こんにちは」の発声時間 0.57 秒の組み合わせである。各パターンでは、声かけあいさつの開始時点を検出して時刻 0 とし、返答あいさつの動作と発声の開始タイミングを 1 秒から 1.6 秒の範囲で変化させている。

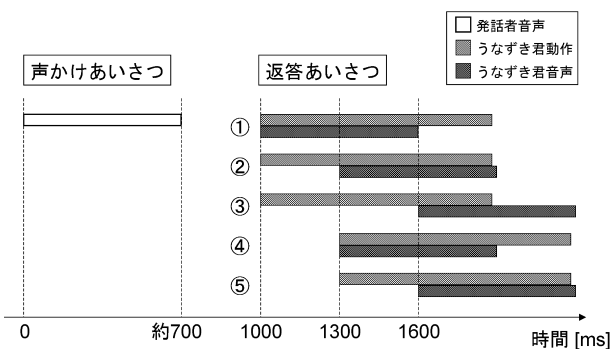


図 6: 合成したあいさつパターン

被験者には、①～⑤の各パターンを一対比較させた。まず、制御用 PC のソフトウェアに「A」「B」の文字を書いた 2 つのボタンを配置し、①～⑤のパターンの中から 2 パターンをランダムに選び、2 つのボタンに割り当てた。被験者には、何度か「うなずき君」と「こんにちは」のあいさつをした後、「A」と「B」のどちらのパターンが「好き」かを二者択一させた。この間は、手持ちマイクを握り、「うなずき君」を正面に見て座るよう指示した。続いて、同様に「対話しやすさ」「一体感」「丁寧」の各観点から該当するパターンを決めさせた。これが終わると「次へ」と書かれたボタンを押し (次の 2 パターンが「A」「B」に割り当てられる)、同様に 4 つの観点から該当するパターンを決めさせた。比較回数は ${}_5C_2 = 10$ 回であった。

実験時間は約 30 分で、被験者は 18～25 歳の男女学生各 25 名の計 50 名であった。

3.3 結果

「好き」と「丁寧」の一対比較の結果を表 1 に示す。表中の数字は各列のあいさつパターンの勝数、つまり各パターンを選択した被験者の数を表している。

これらの結果に対して、一対比較に基づく評価を一義的に定めるために、以下に示す Bradley-Terry モデルを想定した。

$$P_{ij} = \frac{\pi_i}{\pi_i + \pi_j} \quad (1)$$

$$\sum_i \pi_i = const. (= \text{平均 } 50) \quad (2)$$

(π_i : i の強さの量, P_{ij} : i が j に勝つ確率)

得られた強さ π の値を図 7 に示す。有意水準 5% でまず適合度検定を行い、さらに尤度比検定を行ってモデルの整合性を検定した結果、モデルは却下されず、強さ π の妥当性が保証された。これは、あいさつインタラクションにおいて、動作に対する発声遅延が効果を与えることを示す。

例えば、声かけあいさつに対する応答時間は同じでも、動作と音声と同時に反応する場合と、動作に対し

表 1: 一対比較の結果

好き	①	②	③	④	⑤	合計	順位
①	-	19	33	31	31	114	2
②	31	-	35	36	31	133	1
③	17	15	-	16	14	62	5
④	19	14	34	-	27	94	4
⑤	19	19	36	23	-	97	3

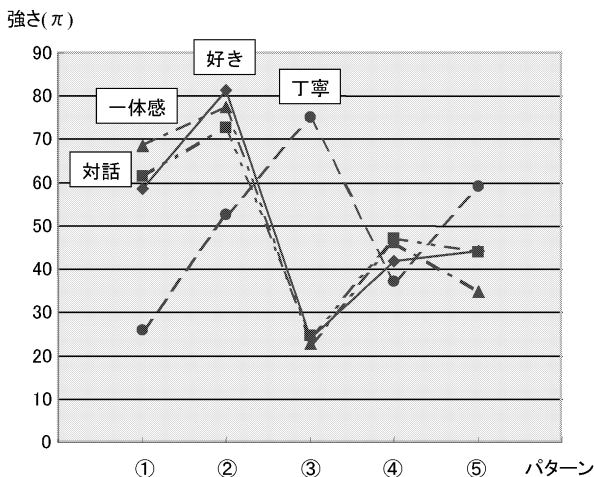


図 7: 一対比較に基づく評価

て発声が遅延するパターンとで、インタラクションの効果が異なっている。とくに、動作と発声が同時の①と人のあいさつに近い②のパターン比較では、②がいずれの項目でも高く評価されている。

また、各観点での強さに対して適合度検定を行った結果、「好き」「対話しやすさ」「一体感」に対して「丁寧」のみ有意差があった。つまり、発声遅延は同じでも「丁寧」の観点では、他と異なるコミュニケーション効果を与えている。

4 対話インタフェースへの導入

あいさつインタラクションの合成的解析により、動作に対して発声を遅延させることで、自然あるいは丁寧など好ましいコミュニケーション効果をもたらすことが分かった。この効果は、身体をもつロボットとのインタラクションにおいて成立するものであり、ロボットと人の対話インタフェースにおいて、積極的に発声遅延を活用できることを示す。

実際、iRT[2]などの、高速なコミュニケーション動作生成技術と音声対話技術を併用することで、自然な動作を生成しながら、音声出力までの時間を利用して高度な音声認識処理・対話生成処理の処理を進めることができる。処理が多少遅れた場合にも、丁寧な対話を生成することで、ロボットのインタラクションを多様化できる(図8)。

今後、家庭用ロボットの普及が予想されるが、そのインタラクションでは、音声認識処理だけではなく、各種検索・予約など、インターネットへのインタフェースとしての機能が重要である。この場合は、不可避なネッ



図 8: 対話インタフェースへの導入例(名古屋市科学館)

トワーク遅延を好適な発声遅延として取り扱うことで、より円滑なコミュニケーション支援が可能となる。

5 おわりに

人とロボットのインタラクション支援の観点から、典型的な人との関わり行動として、あいさつインタラクションにおける動作と発声の生成タイミングを合成的に解析した。

まず、身体的インタラクションロボットを用いたあいさつ合成システムを開発し、人への働きかけを合成的に解析し、動作に対して発声が0.3秒程度遅延するあいさつが好ましく、より遅延が大きくても丁寧に感じられるなど、動作に対する発声遅延が好ましいコミュニケーション効果をもたらすことを明らかにした。

さらに、あいさつインタラクション合成システムを開発し、ロボットとのインタラクションにおいても、適度な発声遅延が好ましいインタラクション効果をもたらすことを明らかにした。

以上は、動作に対して発声の遅延を十分に考慮することで、人とロボットとの円滑なコミュニケーション・インタラクション支援が可能となることを示している。

参考文献

- [1] 渡辺 富夫: コミュニケーションにおける身体性, ヒューマンインタフェース学会誌, Vol. 1, No. 2, pp. 14-18, (1999).
- [2] H. Ogawa and T. Watanabe: Interrobot: Speech driven embodied interaction robot, Advanced Robotics, Vol. 15, No. 3, pp. 371-377, (2001).
- [3] 山本 倫也, 池谷 晴生, 渡辺 富夫: 対面あいさつのコミュニケーション動作における発声遅延の分析, ヒューマンインタフェース学会研究報告集, Vol. 5, No. 1, pp. 83-86, (2003).