音声対話システムにおける発話意図と対話の齟齬による発話タイミングへの影響の分析

豊倉正佳　　翠輝久　　河原達也

京都大学 情報学研究科 知能情報学専攻

toyokura@ar.media.kyoto-u.ac.jp

1　はじめに

音声認識技術・音声合成技術の向上により，様々な音声対話システムが研究されている．音声による情報伝達はユーザを問わず手軽で自然なモダリティであることから，人間どうしのように自然に対話ができる音声対話システムが期待される．しかし，現状ではこのようなシステムは実現されていない．

音声対話システムの自然性を改善するためには，適切な発話タイミングの制御が必要であることが指摘されており [2]，人間どうしの対話の特徴に基づいて，あいづちや割り込みのタイミングを制御する音声対話システム [3] や，発話タイミングを含めた音響的特徴をユーザ発話に同調させるシステム [4] が研究されている．また，藤原らは人間どうしのタスク指向対話において，発話意図や発話の重要度が発話タイミングに強い影響を与えることを示している [1].

しかし，実ユーザを対象としたシステムとの音声対話の大規模なコーパスを用いた分析は十分に行われていない．現状では，ユーザがシステムを人間と同じように認識するとは考えにくく，システムとの対話におけるユーザ発話に見られる傾向を分析する必要がある．そこで本研究では，音声対話システムにおける対話コーパスを用いて分析を行う．

2　京都版ダイアログナビ

京都版ダイアログナビは，音声対話により観光情報の検索と提示を行うシステムである [5].

これは，ユーザ・システム双方が対話の主導権をとりながら，インタラクティブに情報検索・提示するものである．そのために，ユーザ主導の検索・質問応答 (pull) モードと，システム主導の提示 (push) モードを用意して，ユーザの状態に応じてこれらを切り替える．検索・質問応答モードでは，漠然とした検索要求に対して文書を要約して提示したり，特定の情報・事実を求める質問に応答する機能を実装した．また，提示モードにおいては，システム側から，ユーザにとって有用な話題を動的に選択し，質問形式を用いて提示する．

Ｕ１：	金閣寺について教えて下さい.	[情報要求]
Ｓ１：	金閣寺でよろしいですか.	[確認]
Ｕ２：	はい.	[肯定 (確認)]
Ｓ２：	金閣寺こと鹿苑寺は，京都市北区にある臨済宗相国寺派の寺院です. 通称の由来となる金閣は，漆地に金箔を押した三層の舎利殿です. 金閣のそれぞれの階が何を表しているでしょうか. 知っていますか？	[質問]
Ｕ３：	知りません.	[質問応答]
Ｓ３：	一説では，一階を公家階級，二階を武士，三階が義満を表しているのではないかといいます. それでは，足利義満について説明しましょうか？	[情報提示]
Ｕ４：	お願いします	[受諾 (情報提示)]
	.....	

図 1: 京都版ダイアログナビの対話例

システムとの対話例を図 1 に示す.

本システムは，京都大学総合博物館企画展で 2006 年 6 月から 8 月までの 3ヶ月間運用された．運用に際して，情報通信研究機構 (NICT) で開発されたロボット Phyno を使用した．このため，京都観光情報だけでなくエージェント自身に対する質問も数多くなされた.

本研究では，27 開館日分（6 月 23 日から 8 月 3 日）の対話データを使用し，ユーザ発話に発話意図タグと発話タイミングの情報を付与した．誤認識などにより不適切なシステム発話が連続することがあるため，システムによる応答成功率が 50%以上の対話を用いて，幼児を除くユーザによる対話を分析対象とした．対話数は 390，発話数は 2509 であり，発話意図が付与された数は 1268 となった．本システムは，バージインを許さず，システムの発話を遮ってユーザが発話しても，合成音声を中断しない．そのためオーバーラップがあった場合に発話タイミングは負の値をとる．分析では 3 秒までのオーバーラップを考慮している．システムからユーザに働きかける発話 (Initiate) には「情報推薦」,「質問」,「確認」の 3 種類がある.「情報推

薦」とは「○○について説明しましょうか？」のようにシステムから情報を提示する発話である.「質問」は「○○について知っていますか？」のような質問形でユーザの興味を引き出す発話である.「確認」は音声認識結果の信頼度が低い場合に，ユーザの検索内容を確認する発話である.

図 1 の例のように，各ユーザ発話に発話意図をタグ付けした．相手に応答を求める発話を Initiate とし，システムからの Initiate への応答を Response とした．Initiate に属する発話は，京都観光またはエージェントに関する情報を要求する「情報要求」の 1 種類である．Response は「受諾（情報推薦）」「拒否（情報推薦）」「肯定（確認）」「否定（確認）」「質問応答」の 5 種類からなる．括弧内は先行システム発話の内容を示している.「質問応答」はシステムからの質問に対する応答である.

3　発話意図によるタイミングの分析

図 2 に，京都版ダイアログナビにおけるユーザの各発話意図の平均発話タイミングを示す.

Initiate に属する発話は Response に属する発話よりもタイミングが遅く，Initiate が Response よりも思考時間を要することによる遅延が見られる．また,「受諾（情報推薦）」,「拒否（情報推薦）」は，それぞれ「肯定（確認）」,「否定（確認）」よりもタイミングが遅い．これは，システムによる確認に対しては応答が容易であるためである．ホテルの検索・予約やバス運行情報案内における未知情報応答は，話者があらかじめ決定した内容を応答するのに対し，観光案内タスクにおけるシステムからの情報推薦では，ユーザは提示された話題が自分の興味に合うものであるか思考する時間を要する．これらの結果は，人間どうしの対話 [1] およびバス情報案内システムとの対話を分析した結果とおおむね符合した.

しかし,「受諾（情報推薦）」の方が「拒否（情報推薦）」よりもタイミングが遅く,「肯定（確認）」は「否定（確認）」よりもタイミングが遅い．この原因の 1 つとして，ユーザがバージインをしたものの，システムが認識できなかったために，再度発話が行われた場合が頻出したことが考えられる．また，システムによる情報推薦があまり興味に合わなかった場合や，誤った内容の確認が行われた場合など，本来否定すべき場面でも，躊躇した上で肯定を行う例もあった.

4　エージェントに対する接し方の相違による比較

人間どうしの対話では，対話相手への配慮から許容できる時間や，発話内容などにより応答すべきタイミングに制約が生じる．エージェントを用いた対話システムにおいて，ユーザがエージェントを対話の相手として認識している程度により，発話の時間的制約に現れる影響を調べた.

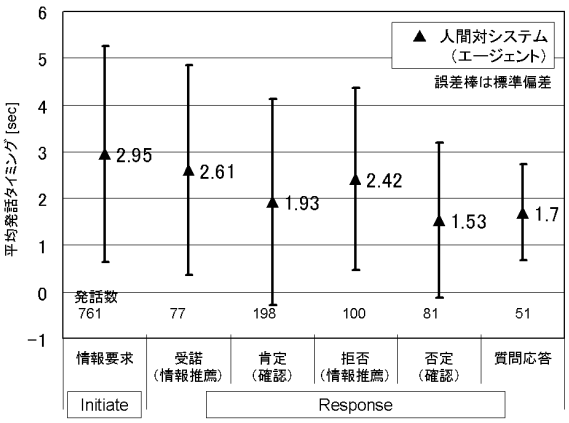

図 2: 各発話意図の発話タイミング

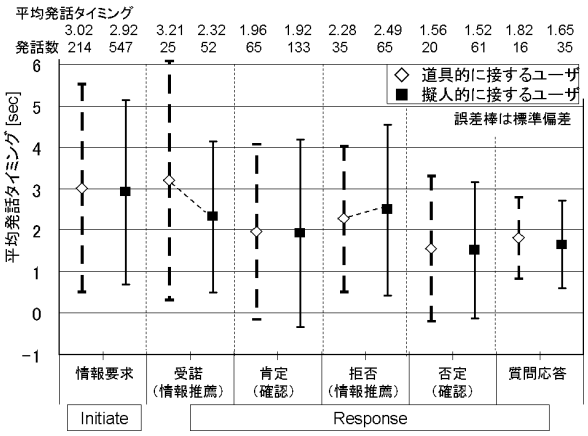

図 3: 道具的／擬人的に接するユーザの比較

本研究では，挨拶，名前の呼びかけ，エージェントに関する質問のいずれかの発話を一度でも行ったユーザを「擬人的に接するユーザ」とした．逆に，これらを一度も行わなかったユーザを「道具的に接するユーザ」とした．図 3 に，道具的に接するユーザと擬人的に接するユーザの発話タイミングを示す.

システムからの情報推薦に対する Response に着目すると，道具的に捉えているユーザは「受諾（情報推薦）」の方が「拒否（情報推薦）」よりもタイミングが遅い．一方，擬人的に捉えているユーザは,「拒否（情報推薦）」の方が「受諾（情報推薦）」よりもタイミングが遅い.

エージェントに対する接し方による,「受諾（情報推薦）」および「拒否（情報推薦）」における発話タイミングの分布の比較を，それぞれ図 4 と図 5 に示す．発話数が 10 %以上を占める頻度の大きい区間に着目す

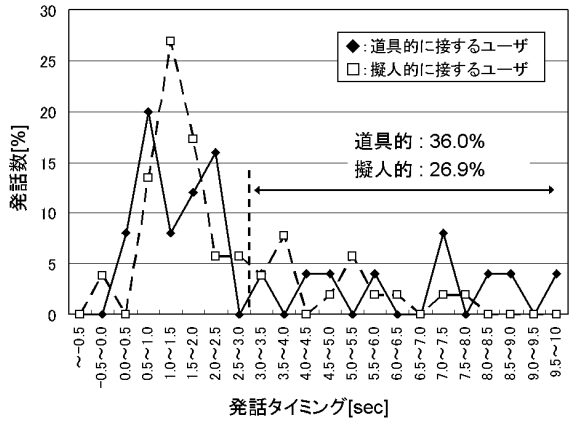

図 4: 受諾 (情報推薦) の分布の比較

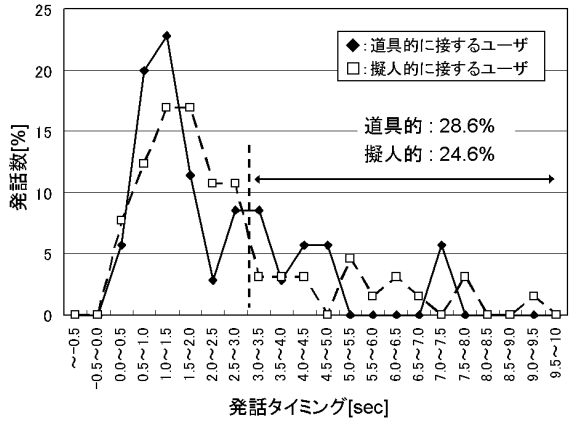

図 5: 拒否 (情報推薦) の分布の比較

ると、図 4 において，道具的に接するユーザの発話タイミングは、0.5 から 2.5 秒まで分布している．一方，擬人的に接するユーザでは，0.5 から 2.0 秒に集中しており，擬人的に捉えているユーザのほうが，応答すべき時間の制約を強く受けていることが分かる．さらに，図 5 では，擬人的に捉えるユーザの「拒否（情報推薦）」の発話タイミングは 0.5 から 3.0 秒までの発話タイミングにおいて多く観測されており,「受諾（情報推薦）」に比べて遅い発話タイミングで発話される傾向があることが分かる．

また,「受諾（情報推薦）」と「拒否（情報推薦）」の両方において，発話タイミングが 3.0 秒以上の不自然な発話は、道具的に接するユーザのほうが多く観測されている．

一般に，人間どうしの対話では，肯定よりも否定の方がタイミングが遅いことが知られており [6][7]，擬人的に捉えているユーザの方が，人間どうしに近い振る舞いをしているといえる．

【不適切な DA が行われる対話例】
S 1: それでは，上賀茂神社の祭事について説明しましょうか.
U 1: はい.
S 2: それでは，下賀茂神社の祭事について説明しましょうか.
……

【唐突な話題の推薦の対話例】
S 1: 平安神宮は，京都市左京区にある神社です.
（中略）
平安神宮について何か知りたいことはありますか？
U 1: ない.
S 2: 平野神社について説明しましょうか？
U 2: いいえ.
……

図 6: 齟齬の起きている対話例

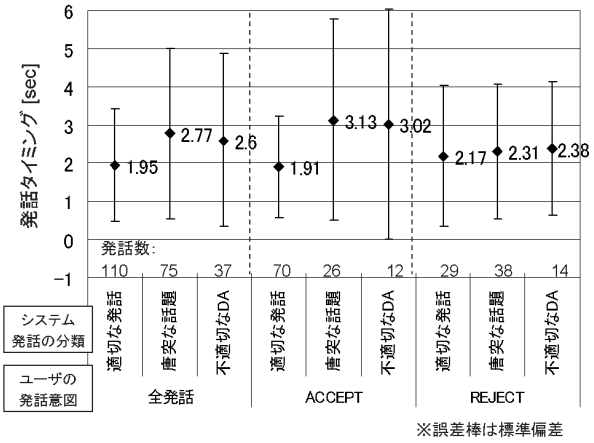

図 7: 対話の齟齬による発話タイミングへの影響

5 対話の齟齬による発話タイミングへの影響と検出

システムの発話において，音声認識誤りなどにより直前のユーザ発話に対して整合性のない発話行為 (DA: Dialogue Act) を行ったり，唐突な話題の推薦をしたりすることがある．このような発話は，ユーザとの対話に齟齬が生じる原因となる．その対話例を図 6 に示す.

そこで，システムからの情報推薦において，(a) 適切な場合，(b) 唐突な話題遷移をした場合，(c) 不適切な DA の場合，に分類して，直後のユーザ発話のタイミングを分析した．平均発話タイミングの比較を図 7 に，発話タイミングの分布を図 8 に示す．特に，適切でない情報推薦以外で，ユーザが「受諾（情報推薦）」した場合には平均発話タイミングは遅く，標準偏差も大きいことがわかる．具体的に，図 7 において，ユーザの発話意図によらず「唐突な話題の推薦」および「不適切な DA」の発話タイミングが適切な発話よりも遅くなっており，これらの発話がユーザの発

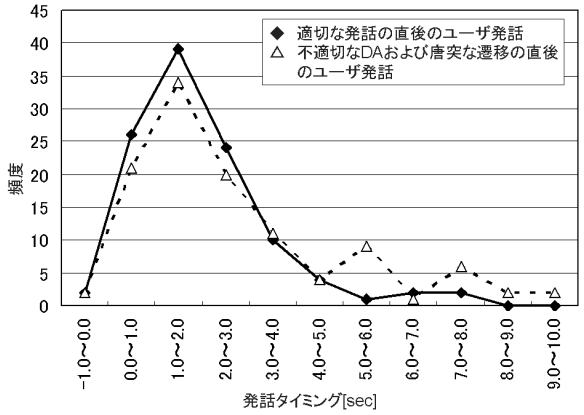

図 8: 対話の齟齬による発話タイミングの分布

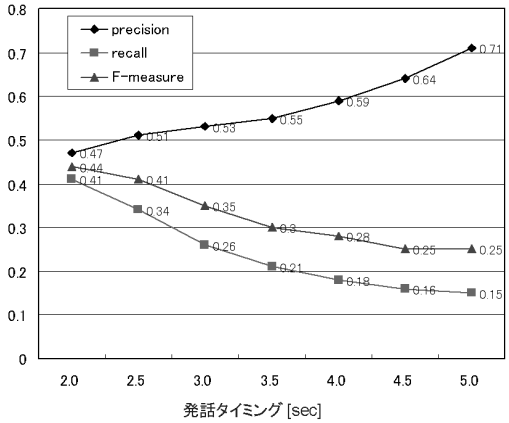

図 9: 発話タイミングによる対話の齟齬の検出

話タイミングを遅延させていることがわかる．また，図 8 では，タイミングが 5.0 秒以上の発話は，適切なシステム発話の直後にはほとんど見られず，不適切な DA および唐突な話題の直後のユーザ発話においてのみ発生している．

これらの分析結果に基づいて，ある程度以上長い発話タイミングの場合に，対話の齟齬が起きたことが検出できると考えられる．発話タイミング長のしきい値を変えた場合にどの程度検出できるかを調べた結果を図 9 に示す．

6 おわりに

本稿では，システムとの音声対話におけるユーザの発話タイミングを分析した．その結果，発話タイミングは思考時間と予測しやすさの影響を強く受け，人間どうしの対話と似た傾向を示すことがわかった．また，ユーザが対話エージェントをどの程度人間のように捉えているかを分類して，分析を行った．その結果，擬人的に捉えるユーザの方が人間どうしの対話に近い傾向を示すことがわかった．

人間どうしの対話では，発話タイミングを含めた音声のパターンは対話者間で同調傾向を示すことが知られている [8]．また,「間」の共有には物理的同時性に加えて認知的な同時性や予測性が重要であることや [9]，交互発話が意識的側面と非意識的側面の二重性を持つことが指摘されている [10]．

さらに，人間どうしの対話においては，相手の反応 (興味や知識レベル) をみながら，説明の方法や話題を制御している．このような対話を分析し，よりプロアクティブな「気が利く」システムに向けて検討をしていきたいと考えている．

参考文献

[1] 藤原敬記, 伊藤敏彦, 荒木健治. タスク指向対話における相互の対話意図を考慮した対話リズムの分析. 言語・音声理解と対話処理研究会, SIG-SLUD-A701, pp. 45–50, 2007.

[2] 伊藤敏彦, 山田真也, 荒木健治. 音声認識・言語理解性能や状況の違いによるタスク指向音声対話の言語的・音響的特徴の比較. 日本音響学会誌, pp. 43–48, 2007.

[3] Norihide Kitaoka, Masashi Takeuchi, Ryota Nishimura, and Seiichi Nakagawa. Response timing detection using prosodic and linguistic information for human-friendly spoken dialog systems. *Journal of The Japanese Society for Artificial Intellignece*, Vol. 20, No. 3 SP-E, pp. 220–228, 2005.

[4] 東海林圭輔, 高橋美佳, 井原誠也, 伊藤敏彦, 荒木健治. 対話に関するリズムや同調作用を考慮した音声対話システム. 情報処理学会研究報告, 2006-SLP-61, Vol. 2006, No. 40, pp. 43–48, 2006.

[5] 翠輝久, 河原達也. 限定されたドメインにおける質問応答機能を備えた文書検索・提示型対話システム. 情報処理学会研究報告, SLP-62-13, Vol. 2006, No. 73, pp. 69–74, 2006.

[6] 川嶋宏彰, スコギンズ・リーバイ, 松山隆司. 漫才の動的構造の分析—間の合った発話タイミング制御を目指して—. ヒューマンインタフェース学会, Vol. 9, No. 3, pp. 379–390, 2007.

[7] A. Pomerantz. Agreeing and disagreeing with assessments:some features of preferred/dispreferred turn shapes, j.m.atkinson andj.heritage. *Cambridge University Press*, pp. 57–101, 1984.

[8] 長岡千賀, 小森政嗣, 中村敏枝. 交替潜時の対話者間影響. ヒューマンインタフェースシンポジウム 2001 論文集, pp. 221–224, 2001.

[9] 今誉, 三宅美博. 協調タッピングにおける相互同調過程の解析とモデル化. ヒューマンインタフェース学会論文誌, Vol. 7, No. 4, pp. 61–70, 2005.

[10] 三宅美博, 辰巳勇臣, 杉原史郎. 交互発話における発話長と発話間隔の時間的階層性. 計測自動制御学会論文集, Vol. 40, No. 6, pp. 670–678, 2004.

[11] 竹内真士, 北岡教英, 中川聖一. 韻律・表層的言語情報を発話タイミング制御に用いた雑談対話システム. 情報処理学会研究報告, 2004-SLP-50, Vol. 2004, No. 15, pp. 87–92, 2004.