チャットボットにおける時間要素の設計について

神場　知成[†1]

概要：テキストまたは音声によって利用者と対話を行うチャットボットが広がり始めており，以前は論理的整合性のある対話の実現手法に研究の主眼があったが，最近では，ボットの性格づけや表現手段，それにより利用者が受け取る印象の相違などの検討も進みつつある．本稿では，これまであまり検討されていなかった，ボットの発話タイミングに注目し，その微妙な相違が人間側に与える印象について基本的な論考を行う．大学一年生 160 人に対するアンケートによれば，チャットボットの用途としては，たとえば天気等の現実的な情報取得だけでなく，自分が退屈なときなどの話し相手としての期待も高く，利用者の印象に影響を与える発話タイミングの微妙な制御はチャットボットの有用性に大きく影響する可能性が高い．従来研究のなかでは，ロボットによる演劇に関しても発話タイミングに関する検討が行われているが，自由な対話における発話タイミングについては，さらなる検討が必要と考える．

Temporal Design of Chatbot Interactions

TOMONARI KAMBA[†1]

Abstract: Chatbots which interact with users by text or voice are starting to spread, and in the past there was a focus on research on realization method of dialogue with logical consistency, but in recent years, bot's characterization and expression means, and the difference of the impression received by the user is being studied. In this paper, we focus on the timing of utterance of bot, which has not been studied much so far, and make a basic discussion on the impression that subtle difference gives to human side. According to the questionnaire for about 160 college freshman students regarding the use of chatbots, there are high expectations for not only realistic information such as weather but also for easy chatting when they are tedious, and delicate control of utterance timing is likely to greatly affect the usefulness of the chatbots. Studies on utterance timing are conventionally being conducted with respect to robotic drama too, but we think further discussion is necessary for utterance timing in free dialogue.

1. はじめに

人工知能技術の発展に伴い，自然言語を用いて人間とコンピュータとが対話するシステムの実用化が進んでいる．対話の方法としては，音声を用いるもの，タイプされるテキストを用いるものとに大別され，最近では，前者はスピーカー型のもの，後者は PC やスマートフォン上でのチャットボットが典型的である．本稿では，このような音声またはテキストの形態での自然言語により人間と対話するコンピュータをチャットボットと総称し，これまであまり検討がされてこなかった，チャットボットの発話タイミングという課題を中心に論じる．

2. チャットボットと従来の研究

自然言語により人間と対等に対話するコンピュータを作りたいという考えはコンピュータの歴史の初期段階からあり，相手がコンピュータか人間かわからない状態で人間が対話を行い，コンピュータを見抜けるかというチューリング・テストはよく知られている [1]. 自然言語によって人間と対話するシステムの実装という点では，1970 年に開発された Winograd による SHRDLU が有名である [2]．ここでは積み木の操作という限定された領域ではあるものの，その時点でどのように積み木が積まれているといった文脈情報まで含めてコンピュータが適切に扱い，人間と自然言語による対話ができることが示された．以後，対話領域を限定しつつもそのようなシステムの研究は進み，最近では，ディープラーニングを用いて学習をする対話システム等の研究が進んでいる [3].

これらは人工知能の処理能力を向上させるという研究目的のシステムであるが，最近では，必ずしも自由に知的な対話ができるというわけではないものの，現状の技術を用いた対話システムの実用化が進んでいる．音声による対話が可能なスピーカー型の装置を Google, Amazon, Apple 等，さまざまな会社が発売している他，チャットシステム上でテキストによる対話を行うチャットボットも多数実用化されている．チャットボットは，たとえば企業が利用者からの問い合わせに答えるカスタマーサポートの場などでもよく使われている．これは，一般に企業に対する問い合わせには，内容的に共通のものが多く，それらに対してはチャットボットが定型的なパターンで回答するだけで短時間に

†1 東洋大学　情報連携学部
Toyo University, INIAD

適切かつ正確な回答を返すことができる場が多いためであり，それにより企業にも利用者にもメリットがある．その他にも，たとえば列車の乗り換え案内をチャットボット形式で行い，目的とする駅まで短時間で到達する乗り換え方法を回答してくれるものなどが，実際に利用されている．

人間との対話を音声で行うものとテキストで行うものとは異なる性質を持つが，いずれも自然言語による対話という点では共通する部分も多い．以下，本稿ではこの 2 つをまとめて「チャットボット」と呼ぶこととし，メディア（音声であるかテキストであるか）の違いは意識しつつも，共通に関連する事項を中心に論じる．

さて，実際に多くの利用者がこれらのチャットボットに接するようになると，コンピュータに知的な会話を行わせるだけでなく，それに対して人間がどのような印象を持ち，場合によっては感情的な影響を受けるかという点が重要となる．

たとえば，Zamora [4]は，ヴァーチャル・アシスタントの実現にあたり，音声とテキストのどちらを利用者は好むか，どのようなシーンに適するか等の実験を行い，①複雑なタスクは，人間側にとって正確に表現可能な，テキストの方が適する．②たとえば銀行口座の話題など，パーソナルな情報に音声には適さない，等の結果を示している．Xu 等 [5]はソーシャルメディア上における，顧客とカスタマーサポート（実際の人間）との会話を分析し，顧客からのリクエストの 40%以上は，明確な課題やほしい情報がある informational な問い合わせではなく，emotional（感情的）なものであることを示し，ディープラーニングの一種である LSTM を用いて学習したヴァーチャルエージェントが，informational，emotional いずれのタイプの問い合わせに対しても，人間と同等の対応が可能であったと述べている．Portela 等 [6]は，人間がボットとの対話で心理的な信頼関係や愛情のようなものを感じるようになるかという実験を行い，現時点の技術ではそのレベルまでは想定できないものの，ユーモアの取り込みや社交的な態度など，今後の検討の可能性を示唆している．Li 等 [7]は，2 つのチャットボットに，顔のイラストも含める形で「穏やかで明るい女性」「厳格で断定的な物言いの男性」という性格づけを行ったうえで，法律事務所のアシスタント募集のジョブインタビューに投入し，それが実際に採用に役立ったこと，および，人格づけによって利用者（この場合はインタビューを受けた人）への影響が異なること，しかも影響は利用者自身の人格によっても異なるので，チャットボットを利用者の人格に合わせてカスタマイズすることが望ましいと述べている．Wei 等 [8]は，人間がソーシャルメディアを利用する中での言葉，アバター，感情アイコンの使い方などを多角的に分析することで人格を指標化する手法を開発し，利用者との対話の中でその指標を積極的に取得するパーソナライズドチャットボットを作成している．また，Candell 等 [9]は，異なる書体で表示されたファイナンシャルアドバイザ（FA：人間の場合と，チャットボットの場合とがある）と人間との対話スクリプトを第三者が見て，その FA が実際には人間とチャットボットのどちらであるかを見分けるという実験をし，FA 側の発話表示に用いた書体が結果に影響したことを述べている．ただし，書体のせいで人間をボットと思ってしまうことはあっても，逆（つまり，手書き的な書体を用いているせいで，ボットを人間と思ってしまう例）はなかったと述べている．

3. 予備実験と考察

3.1 対話内容について

前述したように，チャットと言っても，さまざまな内容がある．ここでは参考として，大学一年生約 160 名に，アンケートをとった結果を示す．質問は「最近，音声で対話するスピーカー型デバイスが発売され始めているが，自分が利用者としてそのようなデバイスで提供してほしいもの，または自分がサービス提供者だったと仮定して提供したいことを記入せよ．」である．学生は，全員，プログラミング等のコンピュータサイエンスを学習中でロボットやチャットボットに関してもある程度の知識を持つ 1 年生であるが，2 年次以降は，エンジニアリング，デザイン，ビジネス等のコースに分かれる予定であり，バックグラウンドや興味領域は広範囲に渡る．利用者とサービス提供者のどちらの視点で回答するかは，学籍番号の偶数奇数で分けているため，回答者の性質は統計的には共通である．視点を 2 つに分けている理由は，「利用者視点」の方は自分自身がどのようなことがしたいかという主観的な回答が増え，「サービス側視点」の方は，社会的ニーズを想像する視点で回答する傾向が強まると考えたからである．図 1 に結果を示す．

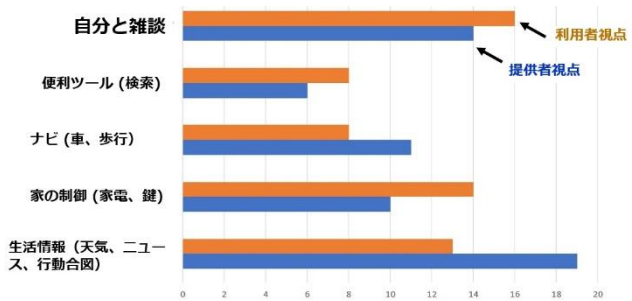


図 1 スピーカー型音声でやりたいこと

グラフの解釈に当たって留意する点は，

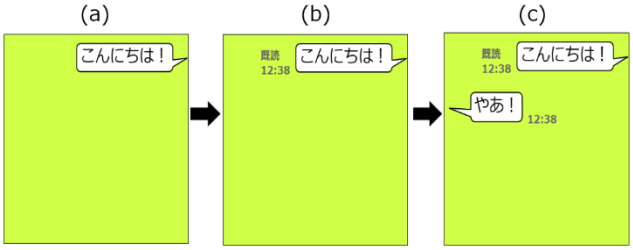
- 音声対話を想定した質問になっており，テキスト対話のチャットボットであれば，異なる結果となる可能性があること．
- 50 文字程度で自由回答したものを，筆者が読み取って分類，整理したものであり，分類には若干の主観が入っていること．

である．たとえば，「電車の時刻表を簡単に知りたい」などの回答は，「便利ツール」にも「生活情報」にも分類可能だが，「生活情報」に分類している．

上記のような留意点はあるものの，筆者の予想と反して興味深い 1 つの点は，利用者視点でも，サービス提供者視点でも，「単に雑談したい」という回答が，回答者の約 20%を占めることである．実際のコメント文では，学生が「退屈なときに」「部屋に一人でいるときなど寂しいので」などとの理由とともに書いていることが多い．

3.2 ボットの発話タイミングについて

前述と同じ大学生約 160 名に，テキスト型のボットの発話タイミングに関する簡易実験を行った．



パターン(1)　(a)と同時に(c)まで表示
パターン(2)　(a)→(b)に 1.5秒、(b)→(c)に 1秒

図 2 チャットボットの発話タイミングに関する予備実験

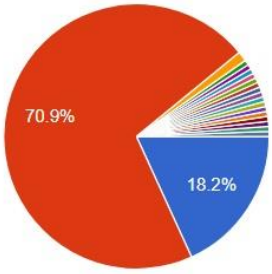
実施の方法としては，正面に大きなプロジェクタがある講堂に，全員の学生が集まっている状態で，スマートフォン上でチャットを行っているイメージでプロジェクタ上に画面を大きく表示した上で，テキストをアニメーション表示し，それに対する印象をアンケート形式で確認したものである．図 2 にそのイメージを示す．この図は，プロジェクタに時系列的に表示した画面を，横に並べたものである．具体的には，コミュニケーションツールである LINE で 1 対 1 の対話をする際の画面を意識したものとなっており，自分の入力するテキストが画面右側から吹き出しの形で表示され，それに対する相手の応答が画面左側からの吹き出しで表示されることを想定している．自分が入力したテキストを相手が読んだ時に，自分の入力エリアのすぐ左側に「既読」の文字と閲覧時刻が表示される．それに対する相手の応答にも，応答時刻が表示される．なお，LINE はボットを実現するインタフェースを公開しており (https://developers.line.me/ja/services/messaging-api/)，これとほぼ同様の機能をスマートフォン上で実現することは容易であるが，現時点の LINE の API を利用した場合には，利用者側の入力と同時にボット側が「既読」をつけてしまい，既読になるまでの時間を制御できないため，上記のように，類似する画面をアニメーションで模擬する方法をとった．

表示方法は 2 パターン用意した．

パターン 1)(即座に返事を表示)：右側から「こんにちは！」の吹き出しが表示されると，**直後に**その横に「既読」と表示され，さらに左側から「やあ！」の吹き出しが表示される．

パターン 2)（一呼吸おいて返事を表示）：右側から「こんにちは！」の吹き出しが報じされた後，**1.5 秒経過してから**「既読」と表示され，**さらに 1 秒経過してから**，左側から「やあ！」の吹き出しが表示される．なお，ここで既読，および吹き出しが表示されるまでの時間は，人間がすぐに操作をしていると想定した場合の時間を大雑把に模擬したものであり，厳密な根拠のもとに選択した数値ではない．

この 2 パターンの表示を，交互に 2 回表示し，学生への質問文は「どちらを好ましいと感じるか？その理由は何か？」とし，回答としては①パターン 1)，②パターン 2)，③その他，を選択肢とした．結果は以下の図 3 に示す通りである．



■パターン 1)　（即座に表示する方式）が好ましいとした者：約 18%
■パターン 2)　（一呼吸おく方式）が好ましいとした者：約 71%
□その他：約 11%

(a)　全体（大学一年生 166 名）

	即座	一呼吸	その他	合計
男子	26 (21.3%)	77 (63.1%)	19 (15.6%)	122 (100.0%)
女子	4 (9.1%)	40 (90.9%)	0	44 (100.0%)
合計	30	117	19	166

(b)　男女別

図 3 チャットボットの表示方式に関するアンケート結果

全体としては，一呼吸おく方式が望ましいとする方がかなり多い結果となった．ただし，男女別にみると「一呼吸置く方が望ましい」とする回答は，女子の方が有意に高い結果が出ているが（p 値 ＜0.01），理由は現時点では不明である．

下記は，それぞれの選択に関する，各自からの補足コメントの例である．

1)　即座表示が望ましい

- 相手がボットとわかっていれば，即座表示が良い．遅いと通信障害なども想定してしまう．
- 最初は一呼吸ある方が良いかも知れないが，会話が続けばすぐに返してほしい．
- ボットにリアルさは不要．公式 LINE もすぐに表示さ

れるので，慣れている．

2) 一呼吸置く表示が好ましい

- 即座に返事があると驚き，威圧感を感じる．ボットであっても，こちらの返事を急がされている感じがする．
- ふだん友人と行っている対話をイメージできるシステムが望ましい
- 即座に返事がくると，監視されている感じがして驚く
- チャットボットの良いところは，情報がスムーズに手に入ることだけでなく，会話しているような気楽さや親しみやすいやりとりもある．

3) その他

- 心地よいと感じるタイミングは人によって違うので設定できると良い．
- 何を対話するものかによって異なる
- 音声かテキストかによっても異なる．音声であれば，即座に返ってくる方がよい
- ランダムな表示が良い．人間に類似したものをつくるなら，感情表現も類似しているほうが良い．

4. 対話と時間要素

前章で述べたように，大学一年生という年齢層にチャットボットに対するニーズを聞いてみると，約 20%が「寂しいときや暇なときに雑談をする」という，きわめて感情的なニーズを回答している．このような場合は，利用者からの入力に対してボットが返事をする場合，単に返事の内容だけでなく，表現方法などさまざまな要素が，利用者にとっての印象に関わってくるであろう．また，必ずしもこのような感情的なニーズに対応する場合だけでなく，単にボットが「今夜の天気や傘の必要性を回答してくれる」というような，情報取得のための場合であっても，利用者アンケートの結果として「即座に回答があると，監視されている気がして驚く」というようなものがある通り，必ずしもコンピュータが処理可能な最短時間で回答するのが最適とは限らない．つまり，利用者のニーズに応える対話内容を実現するための人工知能技術の実現は重要な課題であるが，それだけではなく，利用者の感情に微妙に影響する発話タイミングの制御は，重要な 1 つの研究テーマと考える．

さて，本稿では利用者の発話に対して，その場で柔軟な対応を返すチャットボットの応答時間制御を論じているが，ロボット演劇の分野で，すでに石黒・平田 [10]が，興味深い考察と実験を行っている．石黒等は，演劇の分野において「人間が動作と発話を同じタイミングで行うと不自然である」との演劇分野の知見をもとに，ロボットにおいても「ずれ」を意図的に組み込むことで，ロボットの人間に対するコミュニケーションの質が向上すると述べ，その実現に当たっては，ロボットのそれぞれの発話や動作など 1 つのタスクと定義し，タスク間は 500ms 刻みで制御可能にしている．ここで元となっているのは，平田の現代口語演劇理論と演出の方法論である [11] [12].

石黒等が示しているロボット演劇の場合と，本稿で述べているようなチャットボットの場合とでは,「人間にとって，より自然に受け入れることができる対話の実現」という点では共通している．一方，チャットボットの場合は，ロボットのような身体的動作との関係性は考慮せず，音声またはテキストだけによる対話を想定している点で，設計のパラメータが減る一方，あらかじめ想定しない発話を利用者が自由に行うという点での困難さもある．

5. 考察

5.1 チャットボットのシステム構成

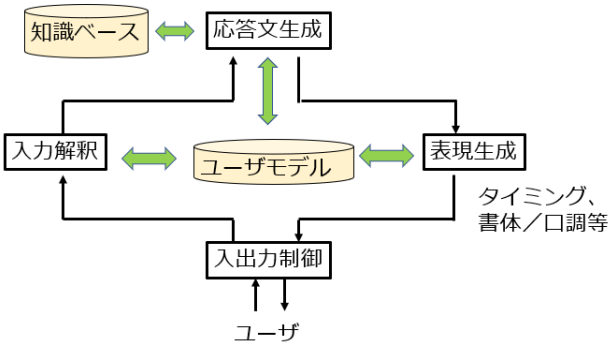


図 4 チャットボットのシステム構成案

チャットボットの発話タイミングを制御することも想定しが，システム構成案を図 4 に示す．ユーザがテキストを入力すると，システムはその入力を解釈したうえで応答文を生成し，ユーザに提示する．この図は，以下のような点を考慮したものとなっている．

・ユーザモデル：システムは入力の解釈や応答文生成をするために，何らかの知識ベースを必要とする．これは，記号処理モデルの場合も，ディープラーニングシステムで各ノードに蓄積された重み等のモデルの場合もあるだろう．しかしいずれにせよ一般に対話においては，対話相手に関する何らかのユーザモデルは保持する必要がある．相手の知識レベルに応じて説明のレベルや用いる単語を適切に選択するためである．さらにチャットボットは，その時の対話ですでに出た話題などに関する文脈の知識も必要とする．これは対話の中でダイナミックに更新されていくものであり，それを知識モデルとして扱うか，ユーザモデルとして扱うかは議論の余地がある．

・表現生成：システムが出力するにあたり，同じ内容を表現するに当たっても口調（丁寧・敬語表現など），書体をどのように選択するかにより，利用者側の印象は異なる．表示タイミングの制御もその 1 つであり，本稿で述べてきたように,微妙な違いでさえも利用者にとまどいを与えたり，異なる印象を与えたりする．図中に示したように，これはユーザモデルを参照して制御されるべきであろう．前述の

既存研究等が示す通り，利用者の人格によっても，適切な表現方法が異なるからである．

さらに，応答が表示されるタイミングの制御という観点に絞って考えると，利用者が入力をしてから回答が行われるまでの時間は，実際には下記のようになる．

Tr = Ti + Tp + Tc

ただし，
Tr: 応答(response)までの時間
Ti: 入力解釈（interpretation）にかかる時間．入力解釈モジュールで発生
Tp: 応答生成（production）にかかる時間．応答文生成モジュールで発生
Tc: 意図的に制御（control）する時間．表現生成モジュールで発生
である．

この中で，一般には Ti, Tp についてはコンピュータ技術の進歩により短縮されていく．もちろん，たとえば外部情報（たとえば外部の Web サイト）を参照して回答を生成する場面では外部サイトの遅延により時間がかかる可能性もあるし，その他の理由により時間がかかる場合もある．たとえばユーザが「6 時 10 分になったら教えてくれ」というような発言をチャットボットに対して行う場合もあり，その場合の表示時間制御をするのは，応答文制御モジュールである．

Tc は，それとは別に，意図的に応答を遅らせることを示している．これは，チャットボット側の発話の内容に関わるものではなく，利用者が受け取る印象を制御するためのものであるが，他研究でも述べられていたり，本稿で示した簡易実験でもわかる通り，利用者に応じてカスタマイズすることが望ましく，それを考慮して，ユーザモデルを参照する設計を予定している．

5.2 音声とテキストの相違

本稿では，チャットボットとして音声を用いるものと，タイプされるテキストを用いるものとを併せて論じてきたが，この 2 つの間に相違点もあることは明らかである．一般に，タイプされたテキストと比較し，音声は声の高さなどさまざまな要素を含み複雑であるが，時間の制御だけを取り上げても，音声について論じる方が，より複雑であろう．テキストであれば，一般にシステムからの出力は一度にすべて表示可能であるのに対し（もちろん，あたかもタイピングをするように，最初から順に文字を一部ずつ表示していくという手法もある），音声であれば必然的に，発話内容を時系列的に示していくしか方法がない．その際に，発話の途中でいったん時間を空ける等，さまざまな制御要素がある．今後，まずはタイプされたテキストを中心にチャットボットの時間制御について検討し，そのうえで音声の要素についても検討をしていきたい．

6. おわりに

チャットボットの発話時間制御に関する考察を示した．チャットボットは，人工知能を用いた対話システムの実験ではなく，実際に日常生活に利用されるものとなってきた．これにより，対話内容を知的で広範囲なものにする「知」の処理能力の向上だけでなく，人間側が自然に受け入れることができたり，良い印象を受け取ったりするための「情」を扱う能力の向上が，重要となってくるであろう．本稿はそのなかで，発話タイミングを制御する必要性，およびその方法について論じた．

参考文献

[1] A. Turing, “Computing machinery and intelligence,” Mind, vol. 59, no. 236, pp. 433‐460, 1950（原文は http://www.loebner.net/Prizef/TuringArticle.html）
[2] Terry Winograd, Understanding Natural Language (Academic Press, 1972 (システムは https://en.wikipedia.org/wiki/SHRDLU)
[3] Zi Yin et al.: DeepProbe: Information Directed Sequence Understanding and Chatbot Design via Recurrent Neural Networks, KDD’17, pp.2131-2139
[4] J. Zamora, I’m Sorry, Dave, I’m Afraid I Can’t Do That: Chatbot Perception and Expectations, HAI’17, pp.253-260, 2017
[5] A. Xu, et al. A New Chatbot for Customer Service on Social Media, CHI 2016, pp.3507-3510, 2017
[6] M. Portela, C. Granell-Canut: A new friend in our Smartphone? Observing Interactions with Chatbots in the search of emotional engagement, Interaction '17, pp.25-27, 2017
[7] J. Li et al.: Confiding in and Listening to Virtual Agents: The Effect of Personality, IUI’17, pp.275-286, 2017
[8] H. Wei et al. Beyond the Words: Predicting User Personality from Heterogeneous Information, WSDM’17, pp.305-314, 2017
[9] H. Candell et al. Typefaces and the Perception of Humanness in Natural Language Chatbots, CHI’17, pp.3476-3487, 2017[10]
[10] 石黒浩，平田オリザ：ロボット演劇，日本ロボット学会誌, Vol.29, No.1, pp.35-38, 2011
[11] 平田オリザ：演劇入門，講談社現代新書，208p, 1998
[12] 平田オリザ：演技と演出，講談社現代新書，220p, 2004
[13] Linda Segar: Japanese Edition of “Making a Good Script Great (1994)”, 330p, 2000（邦題：ハリウッドリライティングバイブル 愛育社）