

他者理解を伴う発話における間合いの解析

小島 敬*¹ 川嶋 宏彰*¹ 松山 隆司*¹

Timing Analysis of Response with Understanding Others

Kei Kojima,*¹ Hiroaki Kawashima,*¹ and Takashi Matsuyama*¹

Abstract – Until now, few approaches considered the effect of timing of utterance on empathy, while most of them deal with a smooth conversation. In this work, we discuss how differences in timing show up between a dialog where empathy is important and a dialog where it is not. For this purpose we rely on the experiment of psychology by using audio response system. As a result, in a dialog where empathy is not important the timing is usually comprised between 100ms and 400ms, on the other hand, in a dialog where it is important the timing is comprised between 400ms and 800ms, that is to say, in the latter people like the timing slower.

Keywords : 対話システム, 発話タイミング, 認知プロセス, 他者理解, 一対比較法

1. はじめに

人が自然に利用できる音声対話システムの実現を目指して、その設計に関する様々な研究が行われており、その中でも発話タイミングは頻繁に扱われている。これらの研究の大半はタスク指向型の対話を扱っており、発話者の発話内容をシステムが把握し、円滑な対話によって、情報検索、予約をするといったタスク目標を達成することを中心に考えている。しかしながら人が日常行う対話の中にはタスクの達成とは違った目的を持つ対話も存在する。例えば、「悲しいです」という発話をする発話者の発話意図は、相手に対して何らかの行動を求めるのではなく、「悲しい」という自己の心的状態を相手に理解してもらうことを目的としている。本研究は、このような「他者理解」を発話者が求めるような対話状況において、聞き手のどのような反応が好まれるかを、特にその発話タイミングの観点から考察・解析する。これによって、人の相談や悩みを聞き、単に情報提供をするのではなく、その心的負担や不安を軽減できるような対話システムの実現のための知見を得ることを目的とする。

発話タイミングの決定要因に関する先行研究

人間の発話タイミングの決定要因についての研究は様々な観点からアプローチされている。竹内ら^[1]は、相槌、話者交替のタイミング決定の要因を先行話者の言語情報、韻律情報によって決定されと考え、決定木により（先行話者の発話終了から）100ms ごとに応答を決定するシステムを構築した。その結果、応答を決定づける要因は先行話者の韻律情報が主要であるこ

とを示した。また藤原ら^[2]は、発話者の発話意図が対話のリズムに大きく影響すると考え、発話タグワーキンググループが提案した発話単位タグ^[3]ごとに発話のタイミングを解析し、発話タイミングは思考時間に関係するという仮説を示している。長岡ら^[4]は、非協調的対話と協調的対話のそれぞれの対話での話者交替のタイミングを解析し、非協調的対話に比べ協調的対話においては、2者間のタイミングが同調する傾向があることを示している。

以上の研究は、タスク指向型の対話を扱っているため、対話のコーパスとしては、2者間の認知情報の交換が中心となっているものを用いている。しかし、コーパスから発話タイミングを解析する方法を取る場合、話者の心的状態に関して考えることは困難である。そこで本研究では、コーパスの解析ではなく、心理実験により被験者にある程度の心的対話タスクを想定してもらい、他者理解を求めるという心的要因が、許容される発話タイミングにどう影響するかを考察する。

2. 認知プロセスによる他者理解のモデル化

2.1 二者間における認知モデル

認知モデルに関しては古くから数多くのモデルが存在するが、他者理解がある場合と無い場合とで異なる認知プロセスを表現している図1の対人認知モデルをもとにして、対話における他者理解というものを考える。このモデルは William James の主我（「見る自己」）と客我（「見られる自己」）という対人認知の分け方から、山本ら^[6]が提案したものである。

この認知モデルには、次の要素が存在する。

1. A1, B1 : 見る自己
2. A2, B2 : 見られる自己

*1: 京都大学大学院 情報学研究科

*1: Graduate School of Informatics, Kyoto University

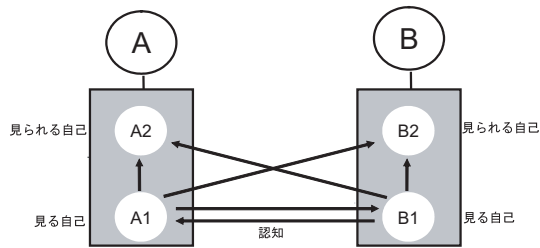


図1 認知モデル

また、以下では

- 認知主体：A1, B1
- 認知情報：A1, A2, B1, B2 の情報（例，A1 は器用である）
- 認知対象：A2, B2 などの，矢印（意識）が向く対象
- 認知状態：A1 A2, A1 B2 などの，認知主体が認知対象を意識する状態

という言葉を用いて説明する。

図1におけるAを主体として考えた場合，認知状態としては，以下のようなパターンなどが考えられる。

- A1 A2：自分に対する認知（例：私は不器用だ）
- A1 B2：相手に対する認知（例：彼は器用だ）
- A1 (B1 B2)：相手の認知状態に対する認知（例：彼は自分が不器用であると認知している，ということを私は知っている）

この最後の例のように，B1 B2 という認知状態も認知対象となりうるとする。以下では，これらの認知状態から他者理解について検討する。

2.2 発話者が求める認知プロセス

発話者がAである場合，Aが自己（A自身）に対して言及したような場合，AはBに対して期待する認知には2パターンある。以下に具体的な例を示す。

(1) 他者理解が求められない場合：

この場合，AはBに対して自己の表層的イメージをBに認知してもらうことを目的とした発話を行う。図1で言えば，AはまずA1 A2の認知状態（私は大阪人である）を表現する発話（「私は大阪人なんですよ〜」）をする。その後BにはB1 A2の認知状態（Aは大阪人である）が生まれ，認知状態の変化を表現する発話（「そうなんですか〜」）をする。このようなプロセスを繰り返し情報共有を行うことで，両者は互いのイメージを形成し，関係が親密になる。

(2) 他者理解が求められる場合：

Aが自己の心的状態を相手と共有したい場合は，上のモデルでは不十分である。例えば，Aが自己の心的状態に対するA1 A2の認知状態（私は今傷ついている）を表現した発話（「悲しかったんですよ〜」）の

場合，BにはB1 A2の認知状態（Aは傷ついている）が求められるのではなく，この場合はB1 (A1 A2)のような認知状態（Aは自己の傷ついた感情を意識している，ということを知っている）が求められる。このようなプロセスにより，BはAの認知世界を自己の中に広げることができ，Aの心的状態を自己の心的状態のように感じることができるのである。このような感情の共有（共感）を行うことで，人と人はより親密な関係を築いているといえる。

2.3 認知プロセスの違いが発話タイミングに与える影響

以上の認知プロセスの違いが，返答する側（先程の例では話者Bにあたる）の発話タイミングに以下のような違い与えようとする。

(1) B1 A2：

2者間の認知情報の交換が目的である場合が多いので，対話のテンポの良さが重要となる。そのため比較的早いタイミングで発話が行われ，聞き手も，同様に比較的早いタイミングを好む。

(2) B1 (A1 A2)：

早いタイミングでの発話は少なく，遅いタイミングでの発話が行われる。これは，発話者の心的状態を自己に投影する認知プロセスに，一定の思考時間を要するため，タイミングとしては自然にゆっくりとなる場合，もしくは，相手が一呼吸つくのを待つ場合がある。また聞き手としても，このような比較的ゆっくりとしたタイミングを好む。

3. 心理実験による発話タイミングの違いの検証

3.1 発話者から見た場合のタイミングの評価

発話者による発話意図が相手に他者理解を求める場合と求めない場合とで，発話者が返答者に対して親近感が持てるような返答者の返答タイミングの違いが現れるかどうかを，音声応答システムを用いた心理実験により評価する。

3.2 実験設定

100, 200, 400, 800, 1600msの返答タイミングについて，親近感に対する感覚的尺度を求めるため，一対比較を用いて実験を行った。なお，違和感なく心理評価してもらうために，一回の対話パターンをいずれの場合も，（未知情報要求）-（未知情報応答）-（受諾）のような共通の発話単位タグ^[3]に設定し，被験者に評価してもらった。

(1) 他者理解を発話者が求めない場合：

1. システムが被験者に対して発話（未知情報要求）をする：出身地はどこなん？
2. 被験者はシステムに対して発話（未知情報応答）をする：大阪やで〜

表1 他者理解を求めない場合の判断比率(番号kの返答タイミングが番号jの返答タイミングよりも親近感が持てると評価された割合)

j \ k	100ms	200ms	400ms	800ms	1600ms
100ms		0.38	0.72	0.22	0.17
200ms	0.62		0.62	0.39	0.01
400ms	0.28	0.38		0.22	0.0
800ms	0.78	0.61	0.78		0.17
1600ms	0.83	0.99	1.00	0.83	

表2 他者理解を求める場合の判断比率(番号kの返答タイミングが番号jの返答タイミングよりも親近感が持てると評価された割合)

j \ k	100ms	200ms	400ms	800ms	1600ms
100ms		0.55	0.61	0.78	0.39
200ms	0.45		0.78	0.62	0.45
400ms	0.39	0.22		0.39	0.50
800ms	0.22	0.38	0.61		0.17
1600ms	0.61	0.55	0.50	0.83	



図2 他者理解を求めない場合の親近感に対する一次元尺度



図3 他者理解を求める場合の親近感に対する一次元尺度

3. システムは被験者の発話を受けて返答(受諾)する: そうなんや~
 4. 3の発話タイミングが異なるような1~3のパターンを2つ聞いてもらい、どちらが親近感を持つ返答かを評価してもらおう
- (2) 他者理解を発話者が求める場合:
1. システムが被験者に対して発話(未知情報要求)をする: どうしたん?
 2. 被験者はシステムに対して発話(未知情報応答)をする: 最近...なことがあって悲しかったん~
 3. システムは被験者の発話を受けて返答(受諾)する: そうなんや~
 4. 3の発話タイミングが異なるような1~3のパターンを2つ聞いてもらい、どちらが親近感を持つ返答かを評価してもらおう

以上を、被験者の学生8人にそれぞれの場合で各20対、合計40対に関して評価してもらった。

3.3 考察

表1と表2に、他者理解を求めない場合、求める場合の発話タイミングの違いの判断比率表をそれぞれ示す。また、図2と図3に、それらから得られた一次元尺度をそれぞれ示す。

図2と図3より、他者理解を求めない場合と求める場合を比較すると、求めない場合については400msのタイミングがピークであり、800msはそれほど高い評価は得ていないが、他者理解を求める場合に関しては、400msと800msがもっとも親近感が持てると評価されている。100ms, 200msは両者に共通して感覚的尺度は近い位置にあり、他者理解を求めない場合では400ms, 800msの間に位置している。また、他者理解を求める場合では、これらは1600msに近い関係にあり、100ms, 200ms, 1600msが一つのグループになっている。

以上をまとめると、他者理解を求めない場合では早いタイミング(100ms~400ms)が好まれており、他者理解を求める場合に関しては少し間をおいたタイミング(400ms~800ms)が好まれているといえる。

3.4 被験者の発話における韻律情報

心理実験の結果に関しては、他者理解を求めない対話に比べて、他者理解を求める対話の方が返答までのポーズが長い方を良いと評価しているが、被験者の発話(未知情報応答)における韻律情報を調べてみたところ、発話者の話速(mora/s)に関しては他者理解を求めない場合と求める場合とではほとんど相関が見られなかった。

一方、発話長に関しては被験者8人のうち8人ともが発話の長さが他者理解を求める場合に長くなっている。これは、他者理解を求めない場合の対話では、被験者が自分の出身地を聞かれて返答するという状況であるので、「出身地はどこなん?」というシステムの問題に対して、「大阪やで」や「奈良」などの、ほぼ単語レベルでの発話になっているが、他者理解を求める場合では、「どうしたん?」というシステムの問題に対して、「研究会で先生にきついこと言われて、悲しいねん~」というような、文章レベルでの発話になっていることが原因として考えられる。実際、3.3節での実験での被験者の発話長の平均が他者理解を求めない場合で0.64s、他者理解を求める場合で3.00sであった。

そこで他者理解を求めない場合に関しては、被験者が2s以上の発話長を持つ文章レベルでの発話を行うような設定にして同様の実験を行った。なおシステム側からの未知情報要求発話は、発話長が長くなるよう、「出身地はどこなん?」という発話から「正月とか何してたん?」という発話に変更した。

表 3 発話長が 2s 以上の場合での親近感に対する比率 (番号 k の返答タイミングが番号 j の返答タイミングよりも親近感が持てると評価された割合)

j \ k	100ms	200ms	400ms	800ms	1600ms
100ms		0.64	0.58	0.29	0.22
200ms	0.34		0.65	0.35	0.07
400ms	0.42	0.35		0.42	0.07
800ms	0.71	0.65	0.58		0.07
1600ms	0.78	0.93	0.93	0.93	

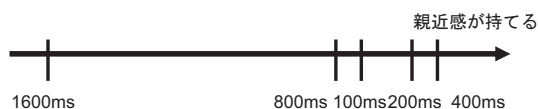


図 4 発話長が 2s 以上の場合の親近感に対する一次元尺度

3.5 被験者の発話長が長い場合のタイミング評価

表 3, 図 4 に結果を示す。発話長が長くなれば, 短い発話での結果 (図 2) に比べて, ゆっくりとしたタイミングで返答されることが好まれているが, 依然, 100ms, 200ms という早いタイミングでの返答は, 比較的高い評価を得ている。

これに比べると, 他者理解を求める場合での結果は, 400ms, 800ms という狭い範囲での返答が好まれていて, 今回の実験と違い 100ms, 200ms という早いタイミングは評価として低い。なお, いずれの実験においても, 1600ms という返答のタイミングは被験者にとって低い評価になっている。

4. 人間との情報共有と発話タイミング

4.1 発話タイミング決定要因について

以上より, 2 者間の対話における発話タイミングの決定要因を考察すると, 先行研究では, 先行話者の韻律情報^[1] や発話意図のタグ情報^[2]などを挙げているが, 前者に関しては, 3 節の実験において, 好まれる発話タイミングが先行話者の発話長や話速によって一概に決定できるわけではないことを示した。また後者に関しても, 共通のタグで実験を行ったが, 好まれる発話タイミングは異なっていた。つまり, 発話タイミングを考える場合, 従来の発話タグでは表現できない発話者の意図があり, 本研究ではそのような意図として自己の心的状態の理解を他者に求める場合と, 求めている場合で違いが現れることを示した。特に, 心的状態の理解を求める場合では一呼吸おいて返答するようなタイミング (400ms ~ 800ms) が好まれてるが, これは, 相手における他者理解の認知プロセスの有無を, 発話者が相手の取ったポーズ時間から判断したことが一因と考えられる。

4.2 対話システム設計と発話タイミングについて

従来の対話システムの設計に関する目的は, 円滑な対話を目指すものが多く, 大半の研究は早いタイミングでのシステムの発話を考えている。しかしながら本研究で扱った対話のように, 状況によっては早いタイミングの発話ではなく, 一呼吸 "間" を空けた発話为好まれる場面があることも考慮に入れなくてはならない。このような早いタイミングと間を空けたタイミングとの違いの要因の一つには, 発話者が相手と認知情報を共有することを目的としているか, 相手と心的状態を共有することを目的としているかの違いがあるといえる。前者の目的であれば, どれだけ相手とスムーズに情報交換ができるかで相手との認知情報の一致具合を判断し, 相手との関係性を判断する。一方で心的状態を共有することを目的としている場合では, 自己の心的状態をスムーズに判断して欲しいということは無く, 空いた間合いから, その人がどの程度自分の立場に立っているかを見て, 自分と感情の共有ができたかを判断する。つまり, 対話システムが発話者に対してどちらの情報共有の目的に応じて, 早めに返答するか, 間をおいて返答するかを決定することで, より人間が親近感を持てる対話システムを実現することができる。

今後は, 発話者が認知情報の共有を求めているのか, また心的状態の共有を求めているのかを, 発話者の心的状態を表現する語の頻度や, 発話者の韻律情報 (強度, F0 パターン, 発話タイミング) から判定する方法を検討し, 状況に応じて発話者の要求に応え, 発話タイミングを制御する対話システムの実現を目指す。謝辞

本研究の一部は, 科学研究費補助金 18049046 の補助を受けて行った。

参考文献

- [1] 竹内, 北岡, 中川: 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; 情報処理学会研究会報告. SLP, 音声言語情報処理, Vol.2004, No.15, pp.87-92 (2004).
- [2] 藤原, 伊藤, 荒木: タスク指向対話における発話意図の対話リズムへの影響; 情報処理学会研究会報告. 2007-SLP-66, 自然言語処理研究会, pp.37-42 (2007).
- [3] 荒木, 伊藤, 熊谷, 石崎: 発話単位タグ標準化案の作成; 人工知能学会論文誌, Vol.14, No.2, pp.251-260 (1999).
- [4] 長岡, 小森, Draguna, 河瀬, 結城, 片岡, 中村: 協調的対話における音声行動の 2 者間の一致・意見固持型と聞き入れ型対話の比較; ヒューマンインタフェースシンポジウム 2003 論文集, pp.167-170 (2003).
- [5] 榎本, 土屋: オーバーラップ発話の評価方法とその基礎統計 ~ 日本語地図課題対話を通して ~; 情報処理学会研究会報告. SLP, 音声言語情報処理, Vol.99, No.99, pp.145-150 (1999).
- [6] 片口, 大山, 他: 医学のための心理学; 誠信書房, pp.247-274, (1962).