

対話システムと心的負担

Mental Work Load of Dialogue System

市川 薫^{*1*2}

Akira Ichikawa

滝沢 恵子^{*2}

Keiko Takizawa

菊池 英明^{*2}

Hideaki Kikuchi

大橋 浩輝^{*3}

Hiroki Oohashi

堀内 靖雄^{*1}

Yasuo Horiuchi

黒岩 眞吾^{*1}

Shingo Kuroiwa

^{*1}千葉大学大学院融合科学研究科

The Graduate School of Advanced Integration Science, Chiba University

^{*2}早稲田大学人間科学学術院

Faculty of Human Science, Waseda University

^{*3}東京大学大学院教育学研究科

The Graduate School of Education, The University of Tokyo

This paper describes methods of easing the mental work load of the elderly people and disabled person who are the users of a dialogue system.

1. はじめに

音声対話システムが活躍する特に重要な場面に、発災時の高齢者や視覚障害者などへの情報提供がある（大地震や津波、交通遮断など）。発災時には、高齢者や視覚障害者などは緊張状態にあり心理的余裕は小さい。このような状況下では、重要な情報を提供するシステムでは、高齢者や障害者が利用する際の心的負担は、できるだけ軽いものであることが不可欠である。

しかし現状の音声システム技術は、音声品質や識率など、伝える情報に主に重点が置かれており、伝達プロセスにおける心的負担に関しては、ほとんど意識されていないように思われる。もちろん合成音声では、従来も聞きやすい抑揚の付加は不可欠とされてきた。心的負担を軽減する重要な機能を持っている。

本稿では、伝達プロセスを中心に心的負担を軽減している音声の構造について考察を行う。特に日常使い慣れている母語（ここでは主に実時間対話音声）に着目する。そこで母語の聞き易さの手がかりを探る一つの試みとして、話者交替時の重複発話に注目し、非母語話者と母語話者の実時間対話の比較を試みた。その他にも伝達プロセスの心的負担を軽減する機能が存在するが、最後にまとめて触れる。

2. 話者交替と重複現象

対話における話者交替時には、先行話者の音声に後続話者の発話が重なり、そのまま発話権を後続話者が引き継ぐ重複現象が度々観察される [Forsyth 08, Stivers 09, 岩 98]。重複発話現象は、以下のように円滑で心的負担の軽い対話を可能とする機能を持つと考えられる。

重複する後続話者にとっては、先行話者の発話の中に、発話の終了を予測可能とする情報（「予告」）が含まれていると、先行発話を最後まで聞く必要が小さくなる。先行話者にとっては、発話が終了しないうちに後続話者が発話権を引き取るため、重複が始まった段階で、継続して発話を新たに計画する必要がないことが明確になる。また、重複が生じている先行発話と後続発話には何らかの内容的な繋がりが深いと自然に判断できるため、文脈を予期するための負担が少なくなる。以上のように考えると、重複現象が適切に生じる対話は、対話参加者にとって心的負担の小さい対話である。

ユーザにとって心的負担の軽い対話システムを実現するには、応答文の発話終了を「予告」し、また「ユーザ音声の予告情報を感知」して予測し、適格な位置（話者移行適格場 Transition-Relevance Place, 以下 TRP）で応答する「対話能力」が必要だろう [市川 11]。TRP とは発話末に存在する話者交替に適した場であり、[榎本 09] は、日本語母語話者では、重複開始位置が TRP よりも早く出現すると不自然に聞こえることを示した。こうした話者交替は、予告情報が存在することで可能になると考えられる。

[大須賀 06] は韻律に話者交替を予告する情報が含まれることを示唆し、[Oohashi 10] は基本周波数パターン生成過程モデル (F_0 モデル) により精度が 70% 程度の発話終了予告が存在していることを示した。[千田 09] は、プロソディ情報のみで話者交替を予測できる可能性を認知実験により示した。

本研究では、言語獲得の臨界期以降に習得した非母語話者の音声対話における重複現象を分析した。非母語話者の話者交替研究は、これまであまり行われていない。中国人留学生の日本語対話と日本語を母国語とする日本人大学生との対話の、また臨界期以降に学習によって英語を習得した日本人の英語対話と英語を母国語とする米国人との対話の重複現象を比較した。

3. 実験方法

3.1 対話条件

千葉大学コーパスに使われた日本語地図データ [小磯 96] を使用し地図課題タスクにより収録を行った。また日本語対話では、課題達成などの負担が無い雑談形式の自由対話も行った。

3.2 対話者の選択

3.2.1 中国人による日本語対話 [市川 13]

下記の条件を満たす中国からの留学生（修士女子学生 2 名）である。

- 言語獲得期以降に日本語を学習
- 実用日本語検定中上級聴力試験 385 及び 420 点
- 日本語学習歴がほぼ同程度であり、日常は母国語（本研究では中国語）で対話している。

なお、日本語母語話者は学部女子学生 2 名。

連絡先: ichikawa@1964.jukuin.keio.ac.jp

3.2.2 日本人による英語対話 [滝沢 13]

- 言語獲得期以降に英語を学習
- TOEIC 得点 940 点以上 (満点 990 点)
- 大学時に 1 年間の留学経験を持つ英語非母語話者 2 名と、帰国子女である英語非母語話者 2 名。

なお、英語母語話者は 20 代米国人男性 2 名。

3.3 収録環境と対話データ

相互に遮音された防音室で収録した。被験者は、互いの音声および実験者の教示をヘッドセットマイクを通して聞くことができる。非母語話者同士、母語話者同士、および非母語話者と母語話者それぞれの組合せで、地図課題対話・自由対話ともに 10 分程度の対話を収録した。

本報告では、先行話者の発話が終わると同時に後続話者の発話が始まる「ラッチング」も、重複と同様のものと考え、あいつちは重複発話に含めていない。

4. 実験結果

4.1 重複率

4.1.1 日本語対話

図 1 は、対話の組み合わせごとの重複率を示している。横軸は話者交替の組合せを示す。先行話者 → 後続話者をあらわす。は自由対話、は地図課題対話である。母語話者同士の重複率を基準とすると、母語話者は非母語話者との対話において重複率が増える。自由対話では重複率は高い。

4.1.2 英語対話

地図課題のみを行なった。図 2 にそれぞれの対話の組み合わせによる重複率の結果を示す。母語話者も非母語話者との対話になると重複が増える点は 4.1.1 の日本語の場合と同じである。

4.2 重複タイミング

話者交替の重複現象が先行話者の TRP 開始よりも前で始まっていた場合は、自然な重複のタイミングを獲得していないと考えられる。この点について検討を行うために、後続話者の発話開始時点の分布を調査した。

4.2.1 日本語対話

対話の書き起こし文から判断できる発話未要素 [榎本 09] を TRP と仮定し分析を行った。図 3 は、後続話者が重複を始めたタイミングの累積図である。TRP 開始点を基点ゼロとして、

重複発話が始まるまでの秒数を計測した。後続話者の重複が先行発話の TRP よりも早く始まったものは左方向 (-)、後に始まると右方向 (+) に示している。確率分布はカーネル密度推定によって推定した。日本人対話では、地図課題と自由対話では分布に差異が見られなかったため、紙面の関係制約上双方を合わせて示す。母語話者対話 (日本人同士) では TRP より前で重複は殆ど行なわれていない。

4.2.2 英語対話

英語における TRP が何かを具体的に示した先行研究は見当たらなかったため、本研究では発話末を基準に分布を調べた。[Stivers 09] のデータでは英語では発話末 300msec 程度前から話者交替が多くなっていることが提示されている。母語話者によるデータであることを考えると、その区間は概ね TRP と考えられる。今回の実験結果でも (図 4)、その辺りから母語話者同士の対話では重複が始まっていることが確認される。

5. 重複発話に関する考察

図 1 に示したように、自由対話に比較し、地図課題対話では重複率が低い。自由対話では、地図課題対話ほど注意深く発話を聞いていないことが示唆される。重複率は対話の内容などに影響される指標といえよう。

日本語及び英語ともに、母語話者と非母語話者との対話は、母語話者同士で行う対話よりも重複率が高い。非母語話者が先行話者である場合は、発話終了を予告し、また後続話者である場合は、予告情報を感知して予測し、適格な位置 (TRP) で応答する「対話能力」に問題があることを示唆している。非母語話者の場合、タスクの有無で差が生じている。一般に対話は、相手の発話内容理解と自己の予期内容の組み合わせなどで理解が行われるが、非母語話者は後者に重みがかかっている可能性を示唆している。つまり、重複率は対話の内容や話者の個性に左右されるといえよう。

次にタイミングについてみてみる。後続話者が非母語話者の場合、図 3 (日本語) 及び図 4 (英語) とともに、先行話者が非母語話者か母語話者によらず基点より早く重複が始まっていることから、後続話者である非母語話者自身にその理由があると考えられる。予告情報を発していると思われる母語話者が先行発話の場合でも、重複タイミングのばらつきが大きい。後続話者の非母語話者は終了予告情報を適切に受取ったり、終了予測を行ったりすることができず、話したい内容を想起した時点で自由に発話を開始していると考えられる。

母語話者が重複する場合には、先行話者の違いによって重複

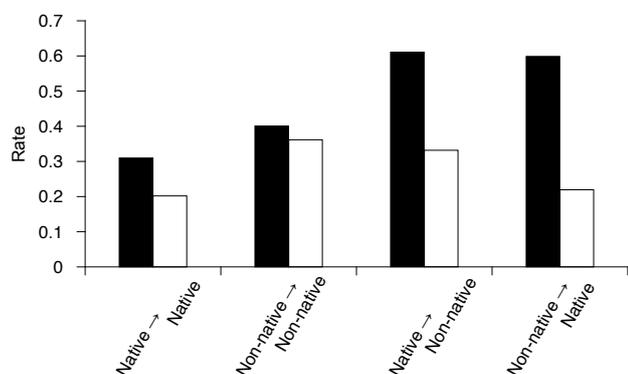


図 1: 日本語対話の重複率 [市川 13]

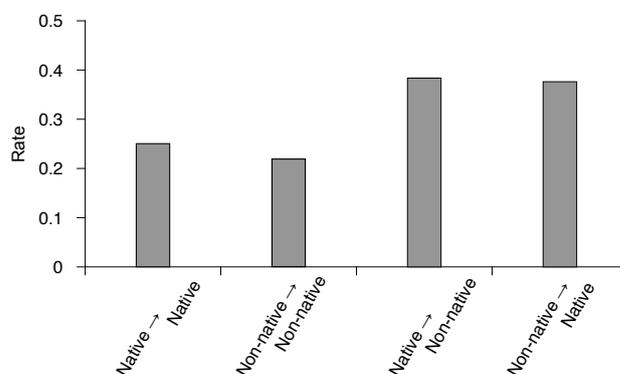


図 2: 英語対話の重複率 [滝沢 13]

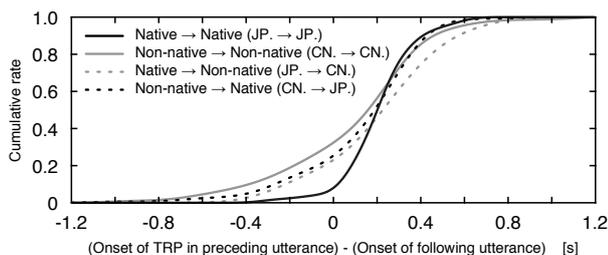


図 3: 日本語対話の重複タイミング [市川 13]

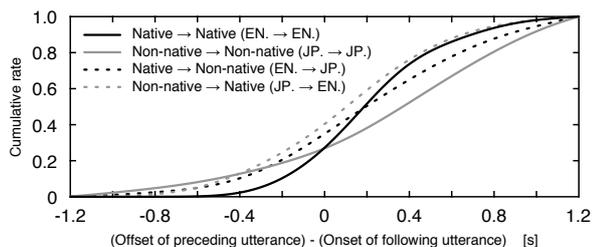


図 4: 英語対話の重複タイミング [滝沢 13]

タイミングの特徴が大きく異なる。母語話者である先行話者が発している発話終了予告情報を受け取ることで、適切なタイミングで話し始める拘束を受けているが、先行話者が非母語話者であるときには、予測することができず、不自然なタイミングで重複を始めてしまっている。非母語話者は予告情報を発することが出来ていないことが推察される。

以上の結果から、母語話者の場合にはタスクの有無に関わらず話者交替の「ルール」として、「対話能力」としての「予告-TRPの対が無意識に機能する」レベル（条件反射的）で獲得されており、それに拘束されているものと考えられる。その結果、対話は楽に（心的負担が小さく）進めることができていると思われる。話し手の発する「予告情報」に呼応して聞き手はTRP以降に発話を開始し、これによりあたかも鎖のように円滑に発話が繋がっていくものと思われる。

6. 単語境界及び係り受けの予告

その他の伝達プロセスの心的負担を軽減する機能には、話者交替の場合と同様に、日本語については、単語（正確にはアクセント句）境界の予告情報や文の構造を示す係り受けの構造を予告する情報が60-80%の精度で存在していることが判っている [Oohashi 10, 市川 11]。単語境界の予告情報は±1モラの誤差を許容すれば90%以上であった。

7. 心的負担の軽い対話システム

以上を参考に実時間対話言語の持つ情報を表1に示す [市川 11]。音声に存在する情報としては、従来は「言語的情報」「周辺言語的情報」「非言語的情報」があるとされてきた。表1では、それらを修正し、主に「伝達内容情報」とし、本報告の中心的内容（「予告情報」）は、「実時間対話の操作に必要な支援情報（実時間伝達支援情報）」として追加している。

「伝達内容情報」（対話言語が伝える情報）としては、「言語関連情報」と「発話者情報」がある。「言語関連情報」は語彙

などの情報を伝える「言語的情報」と、語彙の境界を示す情報（アクセント）や文の係り受け関係を示す「文構造情報」（イントネーション）が存在する。「周辺言語的表現」とは、たとえば疑問の助詞の代わりに文末の基本周波数を上げるような表現方法を指す。「文構造情報」は、係り受け関係などに関する認知処理の探索範囲を限定し、理解処理における心的負担を軽くする機能と、「言語的情報」の文構造の曖昧さを解消するのに必要な情報である。音声合成では不可欠なプロソディによる情報として、従来から重視されてきた情報である。相手の意図を理解するためには、「言語関連情報」として、「誰が」「どのような状態（体調・感情など）」で話しているのか、という情報が不可欠である。従来「非言語的情報」などに含まれていた情報である。人と人との実時間対話では、この情報が対話内容を的確に理解する上で重要であり、心的負担を軽減する機能としても重要である。なお「発話者情報」は、高齢者への支援ロボットなどの合成音声には、ロボットとへの親しさや信頼感を支える為に必要な情報となる。

「実時間対話の操作に必要な支援情報（実時間伝達支援情報）」は、「受け手」が「予測」可能なように「予告」する情報で、実時間対話プロセスの心的負担軽減のために重要な情報であり、本報告のポイントとなるものである。ユーザが、「対話言語が伝えることのできる情報（伝達内容情報）」を受け取ってから認知処理を始めるとは、おそらく実時間対話は間に合わない。ユーザが心的負担を軽く、且つ実時間で知覚認知可能にするためには、「予告」に基づき「発話内容」などを「予測」して準備し、受け取りながら並行して理解処理を進められるようにする必要がある。「予告」情報としては、「言語関連情報」を予告する「伝達内容予告情報」と「話者交替予告情報」がある。「発話者情報」は瞬時に変化する性質の情報ではないため「予告」は必ずしも必要としないと思われる。「伝達内容予告情報」としては、音韻を「予告」する「わたり」や「単語」や、プロソディによる「セグメント境界」を「予告」する情報、文の意味構造の「係り受け」を「予告」する情報がある。「セグメント境界を予告する」ことは、受け手（システムのユーザ）が単語の区間を「予測」し、「心的辞書」に予測的にアクセスすることを可能にする。また「係り受け」の単位を予測する上でも不可欠である。

これらの結果を対話システムという視点から見ると、ユーザ音声に発話終了の予告情報があれば、直ちに応答を始め、存在しないと判断されれば、ユーザ発話に間が生じても継続発話の思考中の可能があり、暫らく応答を待つという処理が考えられる。一方、システムの応答音声に発話終了の予告情報が含まれれば、ユーザはシステム応答が終了すると判断することが可能になり、ユーザは次の要求に自然に移行することが可能になる。予告情報は出さずに、発話継続の状態でも若干のポーズを挿入し、ユーザの理解処理に余裕をあたえるようなことも考えられるであろう。

8. おわりに

これらの情報の具体的構造を更に精細に解明し、音声対話システムに実現することにより、ユーザの心的負担を更に軽減することが期待される。今後の課題としては、心的負担を軽減する予告構造の具体的モデル化が必要である。

謝辞

本研究にあたり、実験音声の収録にご協力いただいた皆さまに心より感謝申し上げます。本研究の一部は科学研究費補助金

表 1: 実時間理解型対話言語の持つ情報の構造 ([市川 11] 修正)

| 実時間理解型言語の持つ情報の種類 | | | 物理特性 | 心的負担軽減機能 | |
|------------------|---|----------------------------|--------------------------|--------------------------|-------------------|
| 伝達内容情報 | 言語関連情報 | 語彙的信息 | 言語的信息 | スペクトル | |
| | | | 周辺言語的表現 | | |
| | | 文構造情報 | セグメント境界 (単語, アクセント節等) | | |
| | 発話者情報 | | 係り受け構造 | プロソディ | 理解探求の 心的負担軽減支援 |
| | | | 個人性 (男女, 個人差等) | | |
| | | | 感情情報 | | |
| | | 体調 | | | |
| | | 社会的関係, その他 | | | |
| 実時間伝達支援 | 伝達内容予告情報 (音韻・語彙・文型などにより それぞれ自動付加) | 後続単位の予告情報 (音韻, 単語等) | スペクトル (わたり) | 対話伝達実時間プロセスの 心的負担軽減支援 | |
| | | セグメント境界予告 (単語, アクセント節等) | プロソディ | | |
| | | 係り受け構造予告 | | | |
| | | 話者交替予告情報 (発話態度で付加) | | | |

24650075 及び早稲田大学重点領域研究機構応用脳科学研究所の支援による。

参考文献

[Forsyth 08] Forsyth, R., David, C., and Phoenix, L.: Timelines, talk and transcription: A chronometric approach to simultaneous speech, *International Journal of Corpus Linguistics*, Vol. 13, pp. 225–250 (2008)

[Oohashi 10] Oohashi, H., Ohsuga, T., Horiuchi, Y., Kikuchi, H., and Ichikawa, A.: Prosody, supporting real-time conversation, in *Speech Prosody*, P2b-07 (2010)

[Stivers 09] Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., Ruiter, de J. P., Yoon, K.-E., and Levinson, S. C.: Universals and cultural variation in turn-taking in conversation, *Proceedings of the National Academy of Science of the United States of America*, Vol. 106, pp. 10587–10592 (2009)

[榎本 09] 榎本 美香: 日本語における聞き手の話者移行適格場の認知メカニズム, ひつじ書房 (2009)

[岩 98] 岩 惇子, 榎本 美香, 大谷 京子, 嶋野 健, 土屋 俊: 日本語地図課題対話における相手話者発話中の発話開始現象について, 人工知能学会研究会資料 SIG-SLUD-9802, pp. 15–21 (1998)

[市川 11] 市川 薫: 対話の科学—プロソディが支えるコミュニケーション—, 早稲田大学学術叢書 (2011)

[市川 13] 市川 薫, 大橋 浩輝, 大川 万里子: 非母語話者の音声対話—中国留学生の日本語音声対話—, 日本音響学会春季講演論文集, 1–Q–23c (2013)

[小磯 96] 小磯 花絵, 堀内 靖雄, 土屋 俊, 市川 薫: 地図課題対話における重複発話の分析, 人工知能学会研究会資料 SIG-SLUD-9601, pp. 47–54 (1996)

[千田 09] 千田 みのり, 大橋 浩輝, 大須賀 智子, 菊池 英明, 市川 薫: 話者交替に対するプロソディ情報を利用した聞き手による予測認知の検討, 人工知能学会研究会資料 SIG-SLUD-0803, pp. 57–62 (2009)

[大須賀 06] 大須賀 智子: 音声対話での話者交替/継続の予測における韻律情報の有効性, 人工知能学会論文誌, Vol. 21, pp. 1–8 (2006)

[滝沢 13] 滝沢 恵子, 菊池 英明, 大橋 浩輝, 堀内 靖雄, 黒岩 眞吾, 市川 薫: 日本人学習者の英語対話における流暢性に関する研究, 日本音響学会春季講演論文集, 1–Q–49c (2013)