

高齢者対話インタフェース

—ユーザの聴き手になる音声対話インタフェース—

小林 優佳[†]

山本 大介[†]

土井 美和子[†]

日本では現在高齢者が急増しており，孤独死や認知症治療が問題となっている．人との対話コミュニケーションがこれらの予防に有効と言われているが，介護分野の人材不足などにより十分な“話し相手”を得られない場合が多い．このような現状に対し，我々は高齢者の“話し相手”となる対話インタフェースの研究開発に取り組んでいる．対話インタフェースの戦略はユーザにさまざまな話題を提供する「話題提示モード」とユーザの話の聴き手となる「傾聴モード」に分かれる．本稿では「傾聴モード」における発話文生成について検討を行った．話題辞書を生成し，話題辞書中の単語に返答することで，音声認識成功時に 67%の自然な発話を生成することができ，83%の適切な発話を生成することができるという結果となった．

Development of Dialogue Interface for Elderly People

—Voice Dialogue Interface Capable of Listening Actively—

YUKA KOBAYASHI[†]

DAISUKE YAMAMOTO[†]

MIWAKO DOI[†]

The EU, South Korea and Japan have a pressing need to compensate for growing labor shortages in their aging societies. There is growing awareness that robotic technology has the potential to ameliorate this problem in terms of both physical and mental labor. To take an example of mental labor, a human therapist dealing with elderly people must be an active listener. In order to realize a robot capable of active listening, we adopt Ivey's basic listening sequence skills in microcounseling. In this paper, we describe a voice interface robot that realizes simple feedback, repeat feedback and questions for Ivey's basic listening sequence. We conducted an experiment, whose results show that 67% of incidences of feedback have adequate reflective words for spoken sentences when voice recognition is succeeded, and 83% of questions are adequate for spoken sentences.

1. はじめに

日本では人口の 22.7%が 65 歳以上であり，そのうちの 22%は独居高齢者であり，孤独死や認知症治療が問題となっている．人との対話コミュニケーションがこれらの予防に有効と言われているが，介護分野の人材不足などにより十分な“話し相手”を得られない場合が多い．このような現状に対し，我々は卓上インタフェースロボット ApriPocoTM（図 1 左）を用いて，

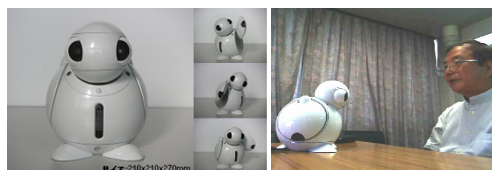


図1 インタフェースロボット ApriPocoTM 外観

と被験者実験風景

高齢者の“話し相手”となる対話インタフェースの研究開発に取り組んでいる¹⁾．

高齢者向けの対話インタフェースとしては松山ら²⁾や神山ら³⁾の研究など高齢者同士の対話を支援するシステムは多く提案されている．これに対し，人間とロボットの1対1の対話システムを提案する．

本システムはロボットが“話し手”となり話題を模索する「話題提示モード」と，話題決定後にユーザの話の“聞き手”となる「傾聴モード」の2つのモードを持つ．この2つをユーザの対話に対する関心度合に応じて切り替える対話法によって対話継続を図る．

本稿では対話戦略の中でも傾聴モードの発話文生成を検討し，実験によって評価を行った．その検討内容を説明する．

2. 音声対話インタフェース

2.1 対話戦略

図2は対話システムの構成図である．システムは話

[†] (株) 東芝 研究開発センター

Corporate Research & Development Center, Toshiba Corp.

話題提示モードと傾聴モードの2つのモードを持っている。話題提示モードでは特定の話題に関する発話を行い、ユーザに話題を持ちかける。傾聴モードではユーザの聴き手に徹し、ユーザの話を盛り上げる。話題提示モード中にユーザの関心度が高ければ話題に興味があると判断して傾聴モードに切り替え、関心度が下がると話題提示モードに切り替え他の話題を提示する。

この2つのモード切り替えを行う関心度は、カメラ・マイクを使用してユーザの話し方・表情などの情報から、4段階で計測する。現在の関心度の計測性能は58%である⁴⁾。

話題提示モードは話題を選択し、発話文テンプレートと対象の話題の話題辞書⁵⁾を組み合わせ発話文を生成する。話題辞書は図2のように単語とその概念が記述された辞書である。概念には分類語彙表⁶⁾の95個の概念を使用している。話題辞書中の単語はその話題に関連性が高い単語に限定されている。話題辞書の詳細については後述する。図2の概念辞書は話題辞書と同様に単語と概念が記述された辞書だが、概念辞書は話題に依存しない一般的な大規模辞書である。発話文テンプレートは図2のように発話文の一部が概念に置き換えられたテンプレートである。テンプレートに入れる単語を変えることによって、異なる発話文が生成される。話題提示モードでは話題辞書中の単語を適宜選択して発話を行う。

傾聴モードでは音声認識結果からユーザ発話中で重要なキーワードを抽出し、概念辞書で検索したキーワードの概念と傾聴モード用の発話文テンプレートを組み合わせ発話文を生成する。

2.2 傾聴モードの対話戦略

次に傾聴モードの詳細構造を述べる。傾聴モードはユーザの話の聞き手になるモードである。

相手の話をうまく聞くための手法として Ivy⁷⁾の提唱するマイクロカウンセリング中の「基本的傾聴連鎖」の発話方法に着目した。また、宮崎⁸⁾も相手の話を聴くための発話方法について記述している。これらを元に音声対話システムによる聴き手の発話方法の初期検討として以下の3種類の発話を行う。

- 単純相槌：「うんうん」（言語内容を含まない）
- 反復相槌：「うんうん〇〇ね」（オウム返し）
- 質問：ユーザの話に対する質問

上記3種類について発話文テンプレートを作成した。

単純相槌は言語情報を含まないでユーザの発話内容に依存しないが、反復相槌、質問はユーザの発話内容が必要になり、音声認識結果を使用してユーザの発

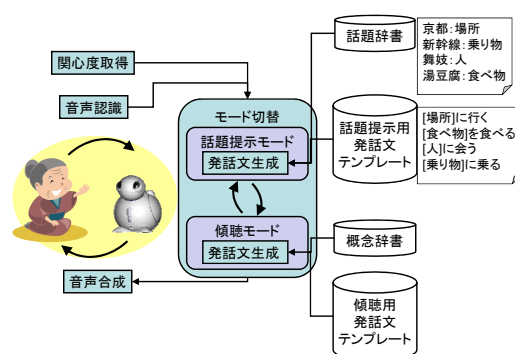


図2 システム構成図

話内容に適切な反応をする必要がある。

ユーザが特定の話題について発話する際、特に興味のある言葉については繰り返し発話され、音声認識で繰り返し認識されている可能性が高い。そこで、システムは音声認識結果を蓄積し、認識回数の多い単語を反復相槌・質問文に使用する。

3. 実験 1

3.1 実験概要

このシステムを使用してユーザの話に対して適切な発話を行えるかどうかの検証をするために、疑似的な実験を行った。まず、被験者14人（若者男性2人女性3人／高齢者男性5人女性4人）で京都旅行に関してロボットと対話実験を行った。被験者はロボットと図1右のように対面して発話し、ロボットの発話内容はWoz方式でスタッフが入力する。ただし、ロボットの発話は2.2の3種類に限定し、ロボットは話題を変更しない。

この対話実験のデータを使用して発話文を生成し、評価を行う。ただし、基本性能を評価するために音声誤認識の影響を除外し、被験者の発話内容書き起こし文章中で複数回発話された重複単語を元に上述したシステムでロボットの発話文を生成した。

実験データから104個の重複発話単語が抽出された。104個の重複発話単語を使用して質問・反復相槌を生成し、生成された発話文を被験者以外の評価者4人で評価した。

3.2 反復相槌の適切性・自然性

3.2.1 評価

まず、反復相槌について評価を行った。評価者に下記のようなロボット・ユーザの3連続の発話履歴を見せ、下線のロボットの発話の適切性・自然性について5段階（5：そう思う 4：どちらかといえばそう思う 3：わからない 2：どちらかといえばそう思わ

表 1 実験 1—反復相槌評価

	適切性	自然性
Good	69%	57%
Bad	27%	37%

表 2 実験 1—質問評価

	割合
適切	56%
わからない	1%
不適切	43%

表 3 話題辞書の一例

表層語	読み	品詞
お寺	オテラ	名詞—一般
新幹線	シンカンセン	名詞—一般
景色	ケシキ	名詞—一般
観光	カンコウ	名詞—サ変接続
豆腐	トウフ	名詞—一般

ない 1：そう思わない) で評価した。1 個目のロボット発話は実験時のものだが、2 個目の反復相槌は実験後に生成したものである。

Robot: 五重塔って、どんな景色が見られるのかな?
 User: 五重塔って言うのは建物で、周りの自然がきれいな
 Robot: うんうん、建物ね

表 1 は評価者全員の平均点が中間点より高いときを Good、低いときを Bad として割合で表したものである。まず、適切性については 27%が Bad となった。また、自然性については適切性が Good で自然性が Bad だった発話文はわずか 6 文だった。自然性は適切性に依存していることがわかる。

3.2.2 考察

適切性が低かった反復相槌を調べたところ主に 2 つの要因があることがわかった。その 2 つを述べる。

① 否定的な意味で使用された単語への相槌

User: 祇園とか、えー、東京にない町並みが好きです
 Robot: うんうん、東京ね

ユーザが否定的な意味で使用している単語に対して肯定的な相槌を返しており、ユーザの気持ちと一致しないため、評価が低くなった。現在のシステムでは否定的か肯定的かを判定することができないので、このような問題点が発生する。

② 基本単語への相槌

User: 一番京都が行きやすい場所だから
 Robot: うんうん、場所ね

対話ではどんな話題でも同じように頻繁に使用される基本的な単語と、その話題特有の単語が存在する。たとえば京都旅行の対話をしているときは「祇園」「紅葉」「新幹線」などは京都旅行の話題特有の単語である。一方、「場所」「食べ物」「建物」などはどんな話題でも同じ程度に使用される。基本単語は文法上多く使用されるだけで、話の中心とは考えられない。そのため、基本単語に対して反復相槌をすると、話の中心ではないものについて相槌を行うことになり、例のように違和感のあるものになる。

3.3 質問文の適切性

3.3.1 評価

次に質問文について評価を行った。発話文テンプレ

ートから 211 文の質問文が生成された。そこで、日本語の文章として適切かどうかを「適切」「不適切」「わからない」の 3 段階に分類した (表 2)。約半数の質問が日本語として不適切であるという結果になった。

3.3.1 考察

不適切になった文章について主に 2 つの要因があることがわかった。その 2 つについて例を使って述べる。

①「分類語彙表の概念区分ミス

②「例 「お茶はどうやって食べるの?」

このシステムでは 95 個に分類された概念を使用して発話文を作成しているが、概念の数が少ないために単語に適した文章を作成できない場合がある。

② 基本単語への質問

例 「川ってお勧めの食べ物は何?」

反復相槌と同じで、基本単語は話の中心にはならない。そのため、例のように基本単語に質問するのは違和感があり、おかしい文章だとみなされる。

不適切と判定された質問文を①②のどちらに当てはまるかを手動で分類したところ①に起因するものが 8 文、②に起因するものが 78 文、それ以外が 4 文あった。基本単語と話題特有の単語の区別ができないことが大きな問題だと考えられる。

4. 話題辞書の適用

4.1 話題辞書

実験 1 の結果から、傾聴モードでは基本単語と話題特有の単語の区別ができず適切な発話文を生成できないことがわかった。そこで、傾聴モードでも話題辞書を使用する。話題辞書は各話題に関して文書 (たとえば Web サイト) を集めてきて、他の文書と比較して特定の話題に関する文書中で頻度が高い単語を収集したものである。表 3 に「京都旅行」に関する話題辞書の一例を示す。話題辞書中の単語はその話題で特に頻繁に使用される単語であり、これらの単語に対して反復相槌や質問を行えば実験 1 の問題は起きないと

考えられる。

4.2 話題辞書を使用した対話文生成システム

遷移時に話題提示モードから引きついだ話題辞書を使用する。まず、音声認識結果中の単語を話題辞書中で検索し、見つかった単語を使用して、今までと同様に発話文テンプレートで発話を行う。この方法を使用すればロボットは話題辞書中の言葉でのみ反復相槌・質問を行うので、基本単語に対して発話することができる。

5. 実験 2

5.1 実験概要

実験 1 で使用した対話データのユーザ発話について 4.2 のシステムで反復相槌・質問文を生成し、印象評価を行った。ただし、今回は実験時に認識した音声認識結果に対して発話文生成を行った。そのため誤認識も含まれる。

5.2 反復相槌・質問文の自然性評価

ユーザ発話と生成したロボット発話（反復相槌・質問）のペアを見せ、発話の自然性を評価した（表 4）。音声認識率は話題辞書中の単語に限定すると適合率 91%、再現率 23%となった。認識成功時、自然性は 67%となり、実験 1 より向上したことがわかる。

一方認識失敗時はほとんど不自然と評価されている。話題辞書中の単語でもユーザの発話内容と異なる単語は違和感があることがわかった。ただし、適合率から誤認識による発話はわずか 9%であることがわかる。ユーザの話を聞き逃すことは多いが、不自然な発話はほとんど行われなことがわかる。

雑談対話システムではユーザの話全てに反応して発話する必要はなく、システムに都合のよい部分だけ発話に利用することができる。システム発話に使用する語彙を話題辞書中に限定することで誤認識の問題を回避できるようになった。

5.3 質問文の適切性評価

次に、質問文について実験 1 と同じように日本語として正しいかどうかを評価した（表 5）。実験 1 の 56%からかなり向上したことがわかる。話題辞書中の単語に絞ることで、意図しない発話を防げることがわかった。

6. まとめ

話題提示モード・傾聴モードの二つのモードに切り分けて対話を行う対話戦略を提案し、その詳細なシステム設計を行った。傾聴モードの対話文生成について

表 4 実験 2-反復相槌・質問評価

	認識 成功	認識 失敗
Good	67%	2%
Bad	24%	94%

表 5 実験 2-質問評価

	割合
適切	83%
わからない	4%
不適切	13%

話題辞書を使用した対話文生成システムを構築し、音声認識を話題辞書中の単語に限定することで、認識失敗による影響をわずか 9%に抑えることができた。また、音声認識に成功した場合はユーザの発話に対して自然な発話が 67%できたことを確認した。また、日本語として適切な質問文が 83%できたことを確認した。

本稿では発話文の静的な評価を行った。今後は実際に実験を行い、話の流れに即しているかなどより実践的な評価を行い、よりよい対話システムを目指す。

謝辞 本研究の一部は総務省の委託研究により実施したものである。

参 考 文 献

- 1) 山本大介, 小林優佳, 横山祥恵, 土井美和子: 高齢者対話インタフェース『話し相手』となって、お年寄りの生活を豊かに。一, HCS2009-56, pp.47-51(2009)。
- 2) 松山洋一, 谷山輝, 藤江真也: グループコミュニケーション活性化ロボットの印象評価, 人工知能学会言語・音声理解と対話処理研究会, No.59, pp.7-12(2010)。
- 3) 神山祐一, 米澤朋子, 山添大丈: TV 電話システムを通じた遠隔傾聴におけるロボットによるコミュニケーション支援, 電子情報通信学会技術研究報告, Vol.109, No.376, pp.103-108(2010)。
- 4) 小林優佳, 山本大介, 横山祥恵, 土井美和子: 高齢者対話インタフェース-雑談時における関心度検出方法と関心度を利用した音声対話インタフェース-, 人工知能学会言語・音声理解と対話処理研究会 No.59, pp.1-6, (2010)。
- 5) 横山祥恵, 山本大介, 小林優佳, 土井美和子: 高齢者対話インタフェース-雑談継続を目的とした話題提示・傾聴の切替式対話法-, 情報処理学会音声言語情報処理研究会資料, Vol.2010-SLP-80-4(2010)。
- 6) 国立国語研究所: 分類語彙表, 大日本図書刊(2004)。
- 7) A. E. Ivy, M. B. Ivey and C. P. Zalaquett: International Interviewing & Counseling Facilitating Client Development in a Multicultural Society. Brooks/Cole Pub Co(2009)
- 8) 宮崎聡子: 上手な聞き方が身につく技術, あさ出版(2008)。