

音声による雑談型対話システムの提案  
 Construction of chat-type dialogue system by voice control  
 上 寫 亮 佑 † Ueshima Ryosuke  
 土 屋 誠 司 ‡ Seiji Tsuchiya  
 渡 部 広 一 ‡ Hirokazu Watabe

### 1. はじめに

近年、ロボットが教育や福祉等の様々な分野で人間のパートナーとして活躍することへの期待が高まりつつある。そのパートナーロボットを実現させるためには、ロボットと人間との円滑な意思疎通の実現が必要になる。人間同士では対話や SNS、ジェスチャーなど様々な方法で意思疎通を行っているが、実際に対面したときの日常的な意思疎通の手段としては主に音声による自然言語での対話を用いている。そこで、ロボットが人間と同様に自然言語を用いた対話を行う必要がある。人間と対話を行うシステムは大きく分類すると、タスク指向型と非タスク指向型がある。前者は特定のタスク達成を目的としており、例えばユーザが「メールを送りたい」と入力するとメールのアプリケーションを起動するようなものである。このようなシステムでは、ユーザの入力内容をあらかじめ想定しやすく、また文脈を考慮する必要が低い。それに対して後者では、達成すべきタスクを特定していないため、どんな話題についての対話がなされるかをあらかじめ想定するのが難しい。また、非タスク指向型のシステムも人間が文字を入力することでその応答が文字で返ってくるものではなく、文字の入力よりも情報伝達量も多く対面しているときや一緒に部屋にいる場合などに日常的によく利用している音声による入出力をすることがパートナーロボットとして人間と社会で活躍することや導入の実現に必要な不可欠な要素のひとつになると考えられる。

そこで本稿では既存の雑談型対話システムの入出力をチャットではなく音声で行えるシステムを提案する。また、対話システムでは、入出力が文字のタイピングによるチャットと音声の入出力のみによる対話、またチャットと音声を混ぜたもののどれが一番適しているのかを検証する。

## 2. 関連技術

### 2.1. 自己開示応答を取り入れた雑談型対話システム

自己開示発話を取り入れた雑談型対話システム<sup>[1]</sup>とは人間と同様に自己開示をする発話を行うこと、また対話の展開の仕方をランダムにすることによってより多様性を持たせることで、人間とのより自然な雑談型対話の実現を目指したシステムである。自己開示応答とは、話者の発話に含まれる語についてシステム側から主観を含んだ表現を行う応答である。以下の図 1 に自己開示応答を取り入れた雑談型対話システムの全体フローを示す。

システムを起動すると、入力待機状態となり話者が文を入力すると、それが挨拶か否かで分岐する。挨拶であれば挨拶応答<sup>[2]</sup>を行い、挨拶でなければ意味理解システムを用いて会話履歴フレームに格納し、各応答手法で応答が生成できるかを確認する。応答可能なものが複数ある場合、各応答手法に対応する発話の種類によって確率を付与し、それに基づいてランダムに選択する。

† 同志社大学大学院理工学研究科

Graduate School of Science and Engineering, Doshisha University

‡ 同志社大学理工学部

Faculty of Science and Engineering, Doshisha University

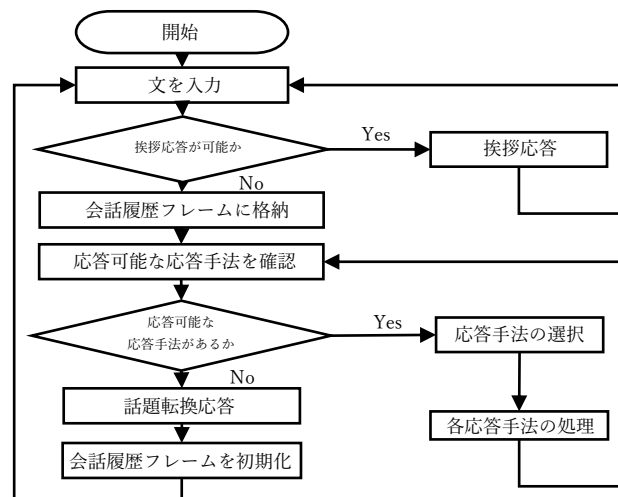


図 1 雑談型対話システムの全体フロー

### 2.2. AITalk

AITalk<sup>[3]</sup>とはテキスト音声合成のソフトウェアで、文章に応じて音響的特徴が最も滑らかにつながる音素を選び出し、それらの音素を結合して合成音声出力する。また必要であれば手動で音素の音響的特徴を変更することが出来る。AITalk の仕様を以下の図 2 に示す。

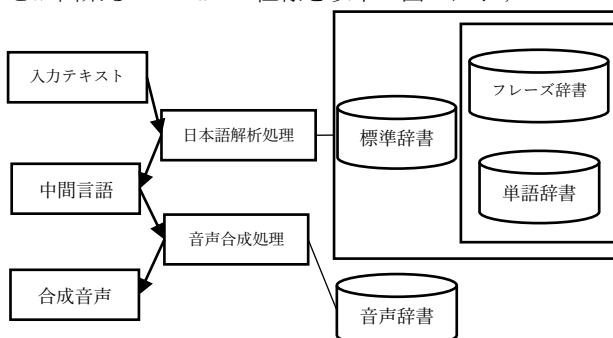


図 2 AITalk の仕様

### 2.3. Julius

音声認識装置 Julius<sup>[4]</sup>とは、ディクテーションと呼ばれる口述筆記などの大語彙連続音声認識を主な目的として開発されたソフトウェアである。Julius の流れを以下の図 3 に示す。

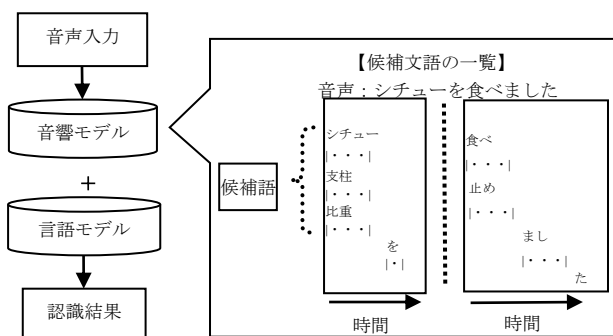


図 3 Julius の認識処理の流れ

図 3 は、音声「シチューを食べました」を Julius が認識した際の処理の流れである。第一に Julius が音響モデルを用いて音声の音響的特徴を分析し、音声区間毎の認識結果の候補となる単語（以下候補語とする）を取得する。次に、言語モデルを用いて単語間の繋がりを考慮して複数の認識結果の候補となる文（以下候補文とする）を作成する。そして、それらの候補文から音響的、言語的に最も相応しいと考えられる候補文を選出して出力する。図 3 の例では、「シチュー」と同じ音声区間の候補語として「支柱」、「比重」が存在し、「食べ」と同じ音声区間の候補語として「止め」が存在する。

### 3. 提案手法

本稿では、ロボットと音声による対話を行うために、自己開示発話を取り入れた雑談型対話システムに Julius と AITalk の機能を利用して音声による入出力を提案する。提案システムのフロー図を以下の図 4 に示す。

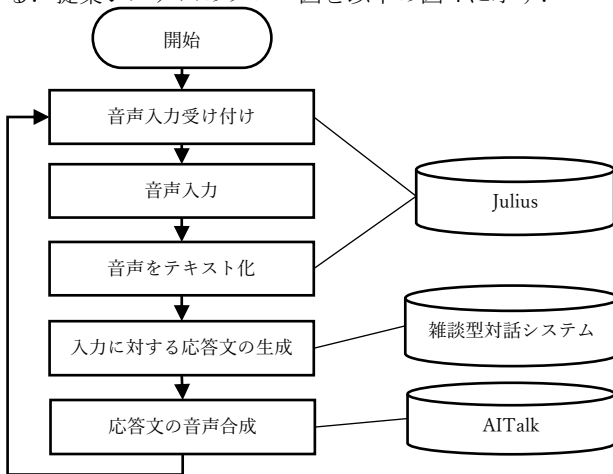


図 4 提案システムのフロー図

#### 3.1. 音声入力受け付け

Julius を利用してマイクからの音声入力を受け付ける。Julius を起動し、音声入力の受け付けを始めると「Audio Input Ready」と表示される。この状態でマイクに話しかけることで音声入力が完了する。音声入力が完了した後はシステムが入力に対する応答を音声出力するため、一時音声入力を停止した状態となる。「Audio Input Ready」が再び表示されれば次の音声入力が可能であり、その再表示は 10 秒後に設定している。

この流れで音声の入力は続くが、手動で Command から Pause を選択することで一時停止、Resume を選択することで再始動することもできる。

#### 3.2. 音声をテキスト化

音声入力を受け付けている状態でマイクに向かって発話することにより、2.3 節で述べたように Julius が認識結果を返す。この認識結果をテキストファイル出力する。2 回目の音声入力の出力結果は新しいテキストファイルに出力し、1 回目のテキストファイルを削除する。常に新しいテキストファイルを作成することで以前までの入力結果を誤って雑談型対話システムに送ることを防いでいる。

#### 3.3. 入力に対する応答文の生成

自己開示を取り入れた雑談型対話システムを利用して入力に対する応答文を生成する。既存システムは起動すると入力受け付け状態となりキーボードによる入力を待機している状態であったが、音声による入力をするために音声認識結果を格納したテキストファイルを自動で読み込むように変更した。つまり、音声入力をまだ行っていない状態では音声入力受け付けの待機状態となり、マ

イクから音声入力を行うと入力の音声認識結果が自動的に対話システムへの入力となる。そしてこの入力に対して 2.1 節で述べたように応答文を出力する。

#### 3.4. 応答文の音声合成

対話システムによって生成された応答文を AITalk を利用して音声合成する。この合成音声をスピーカーで再生することにより音声出力となる。なお合成音声の出力は生成されると同時にスピーカーで出力するように設定した。

### 4. 評価

被験者 5 人に対話システムが音声の入出力の場合、チャットの入出力の場合、入力はチャットで出力は音声の場合、入力は音声で出力はチャットの場合の 4 パターンを検証してもらい点数付けした。最も良いものから順に 3 点、2 点、1 点、0 点とした。結果を以下に示す。

表 1 評価結果

|              | 点数 |
|--------------|----|
| 音声のみ         | 6  |
| チャットのみ       | 9  |
| 入力チャット, 出力音声 | 15 |
| 入力音声, 出力チャット | 0  |

最も良いと思われるものは入力がチャットで出力が音声、次に良いものはチャットのみであった。音声入力の場合に良い評価が得られなかったのは、はっきり丁寧に話さないと誤認識が目立ち、また入力をしてから次の入力まで 10 秒かかってしまうということが原因と考えられる。パートナーロボットを想定する場合にはチャットの入出力には違和感があり、もっと精度が高くリアルタイム性があり円滑に音声の入出力ができれば音声の入出力が最も良いと思うという意見もあった。

### 5. おわりに

本稿では人間のパートナーとして活躍するロボットに必要な不可欠であると思われる音声による対話を実現することを目指した。既存研究の対話システムではキーボード入力を行うことで応答文が文字で出力されていたが、Julius と AITalk を利用することでその入出力を音声に変更することができた。しかし音声入力では雑音のない静かな場所ではっきりと発話を行う必要があった。また入力に誤った認識結果を採用することがあり、思ったとおりの対話ができない場合があった。さらに、音声入力の場合には入力してから次の入力まで 10 秒置いているためリアルタイム性に欠けると考えられる。よって多少雑音がある環境でも扱えるようにすること、音声の誤入力を減らすこと、出力音声を入力として受け入れないようにすることでよりよいシステムになると考えられる。

#### 謝辞

本研究の一部は、JSPS 科研費 16K00311 の助成を受けた。

#### 参考文献

- 北川智裕, 土屋誠司, 渡部広一, “自己開示発話を取り入れた雑談型対話システムの提案”, 第 113 回知識ベースシステム研究会, 13-18, 2018-03-01, <https://ci.nii.ac.jp/naid/40021483385>, 2019/6/19
- 吉村枝里子, 渡部広一, 河岡司, “知能ロボットにおける挨拶文の自動生成方式”, 第 18 回人工知能学会全国大会論文集, 2D1-09, 2004, <https://doi.org/10.1241/johokanri.59.658>, 2018/6/14.
- AITalk®とは? 株式会社 AI (エーアイ) - 音声合成 エーアイ, <https://www.ai-j.jp/about/>, 2018/7/11
- 京都大学, 情報処理振興事業協会(IPA), “大語彙連続音声認識デモコード Julius”, <http://julius.sourceforge.jp/>, 2018/12/18
- 情報処理学会 音声言語処理研究会, “連続音声コンソーシアム 2003 年度 < 最終 > 版 ソフトウェア”, <http://www.lang.astem.or.jp/CSRC/>, 2019/1/17.