6ZL-07

沈黙認識に基づく音声対話システムと感性評価

前土佐 勇仁† 三枝 亮†
†神奈川工科大学 創造工学部

1 はじめに

自然言語処理の発展により音声対話システムの応答性能が向上しているが，音声対話システムはユーザの発話を起点として発話を開始する場合が多く，ユーザが沈黙を続けた場合は対話が停止してしまう．一方，高齢者と介助者の日常的な会話では，介助者は高齢者の発話を十分に待ち，発話がない場合は次の発話を続けて開始する場面がある．本研究では，人間とロボットの自然な会話行動の実現を目的として，沈黙認識に基づく音声対話システムを提案する．本システムではユーザの感情評価と事前に設定した文章群を用いて発話文章を生成し，連続した沈黙を認識した場合は，システムが直前に発話した文章に続く文章を生成する．本システムと被験者の音声対話を行い,「楽しさ」や「親しみやすさ」などの項目について感性評価を行った．

2 音声対話システム

音声対話システムの構成と実装環境を図 1に示す．音声認識には Julius を用い，応答文の生成には Transformer と GPT-2 を用いた．システムの応答にはLINE社の人工知能「りんな」の事前学習モデルに対して，独自に収集した文章群を用いて Fine-Tuning した学習済みモデルを使用した．会話のセッションでは,「バイバイ」,「さよなら」,「さようなら」のいずれかを認識することで応答を終了する．応答文の生成過程を図 2に示す．

本研究では，応答状況に基づいて沈黙を「無言」と「間」に区別する．システム A と対話者 B の二者による直前の会話において A(または B) の発話の終了時刻を起点とし， 双方とも短い時間発話せずに A(または B) が発話した場合，その開始時刻を終点とする時間帯を「無言」，同様に B(または A) が発話した場合，その開始時刻を終点とする時間帯を「間」と定義する．

図 1: 音声対話システム．上：実装環境，下：構成図.

発話は音量レベルに基づいて検知されて録音する．発話開始は入力音が予め定めた閾値を超えることで検知し，以降は 1[s] 単位で音量レベル値を確認する．検出値が閾値を下回った場合は発話の終了と見なし，録音を終了する．

沈黙認識は，直前の会話においてシステムが音声出力を終了した時刻を起点とする．システムが無言を認識するまでの待機時間は，正規分布 [1] により確率的に決定する．平均 $\mu = 2.8$, 偏差 $\sigma = 6.14$ の実現値の分布を図 3に示す．なお，2.5[s] 未満の場合は除く．待機時間を超過するとシステムは発話の検知を中断し，ユーザの応答文は「…」が 3〜5 個ほど記された文章として表現される．

システムが間を認識するまでの待機時間は，2[s] とした [2]．待機時間を超過すると，直後に認識したユーザの応答文の文頭に「…」が 2 個付与される．なお，文中の間には対応していない．応答文の生成には「ユーザ文章＋ユーザ保有感情＋ 1〜3 のランダム数字」を判定文字列とするパターンマッチングを用いた．ユーザの発話内容は 9 種類の感情に分類され，10〜30 通りの応答文を生成する．無言が連続する場合は，システムの直近の応答文を用いて新たな応答文 (独り言) を生成する．無言認識による独り言の生成過程を図 4に示す．

Voice Interaction System based on Silence Detection for Natural Dialogs
Yuto MAETOSA†, Ryo SAEGUSA†
†Faculty of Creative Engineering, Kanagawa Institute of Technology, 243-0203, Shimoogino 1030, Atsugi, Japan
{yuto.maetosa, ryo.saegusa}@syblab.org

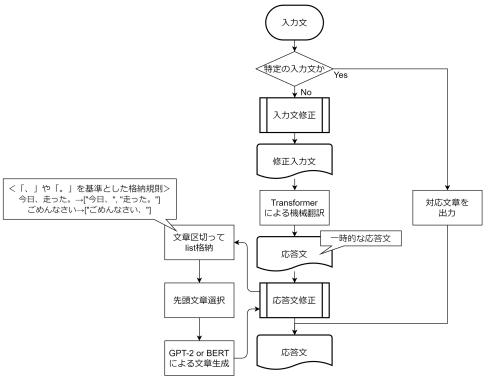

図 2: 応答文の生成過程

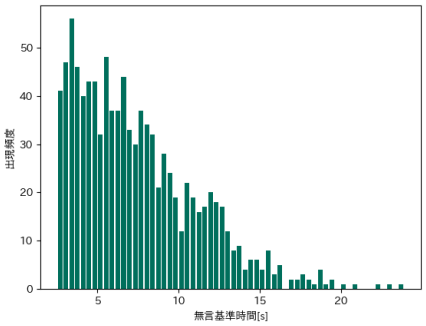

図 3: 無言認識における待機時間の実現値の分布

3　実験

本システムと被験者の二者間で自由に対話させる実験を行った．実験環境を図 5に示す．応答文の生成には，相手を励ます会話を中心に集めた約 1600 文のデータを用いて Fine-Tuning した学習済みモデルを使用した．実験の条件として，被験者には会話中に沈黙を連続で入れることと，会話を自由に終了してよいことを指示した．なお，会話中に違和感を生じさせないように被験者には文章結果を見せずに行った．

実験終了後に被験者に対してアンケート調査を実施し，システムの感性評価を行った．アンケートには違和感に関して「システム応答時間・自然さ・文の連続性」，印象に関して「楽しさ・親しみやすさ」の合計 5 項目の質問を設け，各項目の良し悪しに関する 5 段階の選択肢で回答させた．

アンケートの集計結果を表 1 に示す．被験者は 10 代から 20 代の男性 13 名を対象とし，各項目の回答を 5 点満点に換算した．結果では文の連続性の評価は低かったが親しみやすさの評価は高く，沈黙への対応によりシステムへの親和性が向上することが示唆された．

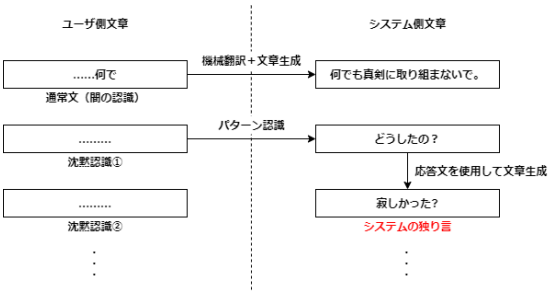

図 4: 無言認識によるシステムの独り言

図 5: 実験環境．左：被験者席，右：実験者席．

4　まとめ

本研究では，沈黙認識に基づく音声対話システムを提案した．沈黙認識によりシステムがユーザの応答を待ち，状況に応じて発話を続けることで会話を展開することが可能になった．感性評価の実験ではシステムへの印象は高評価だったが，文の連続性やシステム応答時間については課題が残った．今後は発話の終了認識の精度改善や無言と間の分布傾向に関する調査検討を行う．

謝辞

本研究を進めるにあたり，神奈川工科大学の高川俊輔氏より助言をいただいた．本研究は同大学先進技術研究所の助成を受けた．

参考文献

[1] 小川一美, 会話セッションの伸展に伴う発話の変化:Verbal Response Modes の観点から, 社会心理学研究, 第 23 巻 第 3 号, pp.269-280, 2008.

[2] 大藤聖菜, 馮建美, 小山大幾, 今井倫太, 人間とロボットにおける「間」のデザイン, 情報処理学会　第 79 回全国大会講演論文, pp.243-244, 3 月 16 日, 2017.

表 1: システム評価実験結果

項目	評価平均	標準偏差
システム応答時間	2.77	1.10
自然さ	2.92	1.91
文の連続性	1.77	0.48
親しみやすさ	4.15	0.90
楽しさ	3.92	0.67