音声対話システムの応答に対するユーザの許容範囲の調査 -パラ言語情報に着目して-

Research of Users' Allowable Range for Responses of Spoken Dialogue System -Focus on Paralinguistic Information-

菊池浩史 [1*] 楊潔 [1] 菊池英明 [1]
KIKUCHI Hirofumi[1] YANG Jie[1] and KIKUCHI Hideaki[1]

[1] 早稲田大学
[1] WASEDA University

Abstract: ユーザが許容できないパラ言語情報での応答を音声対話システムがすることによって、ユーザの対話継続欲求が下がるという問題が存在する。本研究ではこのような破綻の問題の解決を目指して予備的な調査を行う。ユーザとシステムの対話を想定して、固定したユーザ発話と汎用性の高い相槌の「そうですか」というシステム応答を接続した音声刺激を用意し、聴取評価実験を行う。分析を通してパラ言語情報に着目した応答多様性の許容範囲を探る。

1 はじめに

音声対話システムは孤独感の解消や介護・子守の現場への活用が期待されている。しかしながら、音声対話システムが同じ応答を繰り返すことによってユーザが飽きてしまう問題がある。対話を続けたい・また対話したいと思う欲求、対話継続欲求が低くなることが原因と考えられる。以上の問題を解決するために、音声対話システムの同じ応答の繰り返しを避け、多様なシステム応答を生成する必要がある。鈴木らはシステムがユーザ発話のパラ言語情報を反響的に模倣しビープ音で応答することでユーザの志向的な姿勢が誘発されることを示唆 [1] し、発話の音調の多様性をイントネーションで表現できる可能性を示した。一方、宮澤らは人と音声対話システムの対話において、ユーザの対話継続欲求を高めるにはシステムがユーザに対して「話を聞いてもらえるという実感を与えること」を示すことが有効であると示唆している [2]。このことから、ユーザ発話に対するシステム応答のイントネーションを調節し多様なシステム応答を表現することで、ユーザの対話継続欲求を高めることが期待できると推測する。さらに鈴木らは、ユーザがシステムに対し人間の成人と同等な自律的振る舞いを期待することが考えられるため、同調では必ずしもポジティブな印象を持つとは限らない [1] とも述べている。つまり音声対話システムでは同調のみではない多様な応答の実装が必要であると推測できる。

多様な応答を実現する際、ユーザが許容できないパラ言語情報での応答を音声対話システムがすることによって対話が破綻する恐れがある。そこで本研究では、ユーザ発話とシステム応答のパラ言語情報として表出された発話者の快不快状態に着目し、ユーザ発話へのシステム応答に対するユーザの許容範囲について音声聴取評価実験によって調査した。

2 音声聴取評価実験

2.1 仮説

本研究の目的を実現するために下記の仮説を立てた。

1. ユーザのシステム応答に対する許容範囲とシステム応答に表出された快不快状態には関係がある
2. 仮説 1 の関係にはユーザ発話に表出された快不快状態が影響する
3. 仮説 1 において、システム応答に表出された快不快状態の度合いと、ユーザのシステム応答に対する許容の度合いの関係は線形ではない

*連絡先： 早稲田大学
〒 359-1192 埼玉県所沢市三ヶ島 2-579-15
E-mail: hirofumi.kikuchi@toki.waseda.jp

2.2 概要

ユーザ発話「連絡を待っています」とシステム応答「そうですか」を結合した音声試料を被験者が聴取し、ユーザ発話が自分自身の発話であると仮定した時にシステム応答がどのくらい許容できるかを 7 件法（1. 非常にできない 7. 非常にできる）で回答してもらった。以後、得られた値を評価値と呼ぶ。実験の最後に、許容できる（できない）と判断したポイントと後述する背景について、アンケートによって回答してもらった。また、教示は以下の通りにした。

1. ユーザ発話は自分自身の発話であると仮定
2. システムのキャラクター性は考慮しない
3. 背景は深く考えず直感で答える

本実験は対話システムの応答が自分に向けてであることを前提としている。そのため、音声試料のユーザ発話を被験者の発話であるという仮定が必要不可欠である。また、システムの声そのものが許容できないといった、システムのキャラクター性は考慮しないこととした。最後に、対話は背景によって多様な解釈ができてしまうため、対話の背景を限定せず、被験者に素早く直感で答えてもらうことを優先とした。被験者は 10 代から 60 代で合計 10 名（男女各 5 名）の協力を得た。

2.3 快不快識別器

Fairy Devices 株式会社が提供する音声感情識別器の実装を用いた。この識別器は LLD (Low Level Descriptor) 音響特徴量の BoAW (Bag of Audio Words) 表現 [3] を用いた SVR (Support Vector Regressor) である。快不快の推定のために、UUDB[4]（宇都宮大学パラ言語情報研究向け音声対話データベース）release 1 の全データ（4840 発話、1 時間 53 分）を使用して感情識別器の訓練を行った。訓練の際、UUDB のパラ言語情報ラベルのうち ”快-不快” の全評価者の平均評価値を用いた。以下、この訓練済みの識別器を「快不快識別器」と呼ぶ。また快不快識別器が産出する値を「識別値」と呼ぶ。識別値は -1〜1 の範囲を取り、小さいほど不快、大きいほど快を表す。

2.4 音声試料

ユーザ発話として、特定の感情の影響を比較的受けにくいと思われる「連絡を待っています」を使用した。ユーザ発話の快不快状態として「発話者の気持ちや気分の良し悪し」を表現したものを収録し、その中から識別値が適度に分散するように音声を選定し、強い不快（0.218）、弱い不快（0.412）、弱い快（0.638）、強い不快（0.765）の 4 種類を用意した（括弧内は識別値）。システム応答として、汎用性が高く、発話時間が短くない相槌の「そうですか」を使用した。「そうですか」には多様な音声が必要となった。そこで、女性声優 5 名による、人物像 10 種類、シチュエーション 28 種類、平静音声 1 種類、合計 1405 種類の「あーそうですか」が収録されている「表現豊かな音声コーパス」[5] から抜粋して使用した。本実験では「あーそうですか」の「そうですか」部分のみをシステム応答として 48 種類用意した。48 種類の音声試料は不快から快までほぼ均等に分布（識別値 : -0.137〜0.807）している。以上、4 種類の「連絡を待っています」と 48 種類の「そうですか」を組み合わせ、合計 192 種類の音声試料を作成した。

2.5 手順

音声分析ソフトウェア Praat を用いて図のようなインターフェイスを作成した（図 1）。被験者 10 名全員に対し、192 種類の音声試料をランダムに全て聴取し、どのくらい許容できる（できない）かを評価してもらった。

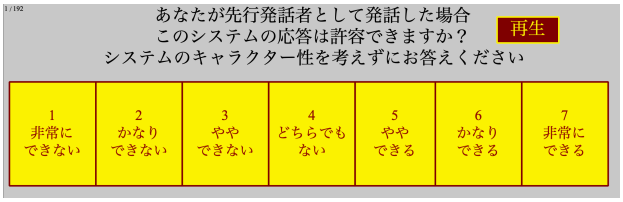


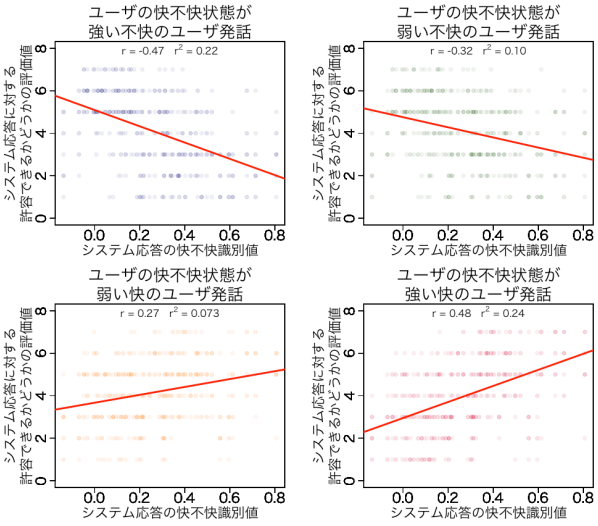
図 1: 音声聴取評価用インタフェース

3 結果

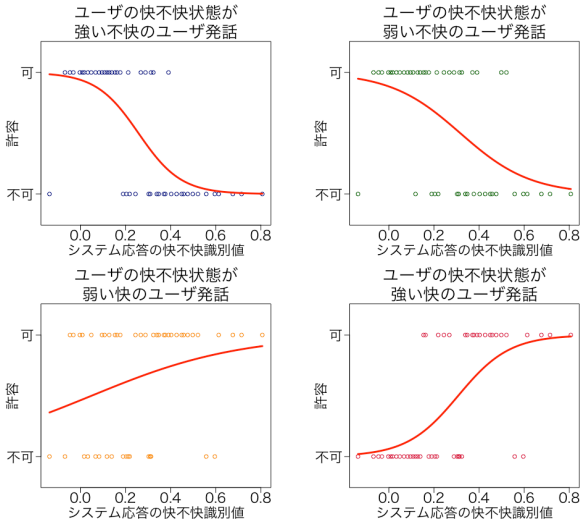
ユーザ発話 4 種類ごとに集計した評価値を散布図で示し、回帰直線を記したものを図 2 に示す。散布図の点の色の濃さは同じ評価をした人が多かったことを表している。次に、192 種類の音声ごとの評価値の平均を算出し、その値が 4 よりも大きかったときを許容できるとし、ユーザ発話 4 種類ごとにロジスティック回帰分析（図 3）を行なった。なお、閾値を 4 に設定したのは、今回、許容できないと判断されたもの以外を許容できると定義したためである。

4 おわりに

実験より、ユーザ発話とシステム応答において両者の快不快状態に一定の関係があることを示した。図 2 の回帰直線の傾きや相関係数から、ユーザの快不快状



各ユーザ発話における
図 2: システム応答に対する評価値の
散布図と回帰直線



各ユーザ発話における
図 3: システム応答に対する評価値の平均を用いた
許容可/不可のロジスティック回帰分析

態が強い不快および強い快のユーザ発話のときにより顕著な傾向が確認できる。しかし図 3 がユーザの快不快状態が弱い不快および弱い快のユーザ発話のときに許容の可不可がシステム応答に表出された快不快状態に関係のない部分があることから、仮説 2 を示唆していると推測できる。また、決定係数の低さから、回帰直線は当てはまりが良いとは言えず、仮説 3 を示唆する可能性が示された。

アンケートより、「同じテンションで応答してくれていること」が許容の判断に大きく寄与していることがわかった。これは宮澤らの「話を聞いてもらえるという実感を与えること」を示すことが有効であるという示唆 [2] を裏付けていると考えられる。一方で、「自分よりも少し快」であるシステム応答に対して印象が良い回答が散見された。これは鈴木らの、同調では必ずしもポジティブな印象を持つとは限らない [1] を裏付けていると推測できる。次に、人を馬鹿にするようなシステム応答はどの状況でも許容できないという答えも多かった。このことから、システム応答のパラ言語情報として表出された快不快状態だけがシステム応答に対するユーザの許容の判断に寄与しているといえない可能性が示された。今後は快不快状態以外のパラ言語情報にも着目し研究を進める。

謝辞

本研究は Fairy Devices 株式会社との共同研究により実施された。音声感情識別器をご提供いただいた同社に感謝する。

参考文献

[1] 鈴木紀子, 竹内勇剛, 石井和夫, 岡田美智男: 非分節音による反響的な模倣とその心理的影響, 情報処理学会論文誌, Vol.41, No.5, pp.1328-1338 (2000)

[2] 宮澤幸希, 小川義人, 松尾智信, 中山真太郎, 常世徹, 桝井祐介, 菊池英明: 音声対話システムにおける継続性向上の要因, 研究報告ヒューマンコンピュータインタラクション *(HCI)*, Vol.2011-HCI-142, No.1, pp.1-8 (2011)

[3] Schmitt, M., Ringeval, F., Schuller, B.: At the Border of Acoustics and Linguistics: Bag-of-Audio-Words for the Recognition of Emotions, *Speech. Proc. Interspeech*, pp.495-499 (2016)

[4] 「宇都宮大学パラ言語情報研究向け音声対話データベース」, NII 音声資源コンソーシアム, URL: http://research.nii.ac.jp/src/UUDB.html[閲　覧日:2020 年 2 月 13 日]

[5] 宮島崇浩, 菊池英明, 白井克彦, 大川茂樹: 演技指示の工夫が与える音声表現への影響: 表現豊かな演技音声表現の獲得を目指して, 音声研究, Vol.17, No.3, pp.10-23 (2013)