音声対話システムにおけるユーザの許容範囲を考慮したシステム応答の検討

Study of System Response Considering Users' Acceptable Range in Spoken Dialogue Systems

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: 音声対話システムによるユーザが許容できないパラ言語情報での応答による破綻により、ユーザの対話継続欲求が下がる問題が存在する。本研究では、このような破綻の問題の解決を目指す。これまでに多様なユーザ発話とシステム応答の快不快状態に着目し、システム応答に対する許容範囲の存在を確認した。本稿では、ユーザ発話の快不快状態によって許容されるシステム応答の傾向が異なることを示し、許容されるシステム応答を出力可能にするための快不快状態に基づく許容評価モデルについて述べる。

1　はじめに

情報技術の発展により音声対話システムの普及が進んでいる。近年、我が国では少子高齢化や新型コロナウイルス感染症（COVID-19）の感染拡大防止対策などの社会情勢により会話の機会が減っている。さらに、会話の頻度が減少することで健康の低下が懸念される [1]。そうしたなか、音声対話システムは対話相手として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音調での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である対話継続欲求の低下が原因の一つとして挙げられる。以上の問題を解決するために、音声対話システムの同じ音調による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。

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」が有効であると示唆している [2]。鈴木らはイントネーションの調節による多様なシステム応答によってユーザの志向的な姿勢が誘発される可能性を示唆した [3]。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、模倣では必ずしもポジティブな印象を持つとは限らない [3] とも述べている。つまり、対話の価値を提供する音声対話システムでは模倣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。そして、竹内らは、人間と人格化したエージェントとのインタラクションが人間同士のインタラクションと同様に社会的であるとの示唆を得たうえで、「一般的な社会性から逸脱したエージェントの振る舞いは，人間とエージェントによる社会的関係を形成するうえで障害となる」[4] と述べている。

このように、多様な応答を実現する際、対話の破綻を防ぐために一般的な社会性から逸脱しない振る舞いを検討する必要がある。そこで本研究は一般的な社会性から逸脱したシステムの振る舞いについて、ユーザが許容できないパラ言語情報での応答に着目し、ユーザがシステム応答に許容できる、対話の価値を提供する音声対話システムの実現を目指す。これまでに、筆者らは、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、一名のユーザ発話音声を用いた音声聴取評価実験によって調査した。その結果、被験者による共通する許容評価が広がり、分布していることが確認され、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の存在を示唆した [5]。また、許容範囲の特性として、

1. ユーザ発話とシステム応答の快不快状態が同じ快または不快のときに許容される傾向がある

*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. ユーザ発話とシステム応答の快不快状態が同じ快・不快であっても、ユーザ発話の快不快状態に対してシステム応答の快不快状態が過剰に強い快または不快を表出するときに許容されない傾向がある

3. メッセージ性の強度によって、どのユーザ発話に対しても許容されないシステム応答音声がある

を確認した。さらに、[6] では、ユーザ発話者 9 名によるユーザ発話を収録し、音声聴取評価実験を行なうことで多様なユーザ発話におけるシステム応答への許容範囲の存在を示唆し、その傾向について考察をおこなった。本稿ではこれまでに得られた多様なユーザ発話へのシステム応答に対する許容評価の分析を行い、対話システムへ実装するための許容評価のモデル構築について検討を行う。

2 研究手法

本研究では、ユーザとシステムの 1 発話ずつの対話に着目する。ユーザ発話へのシステム応答に対するユーザの許容評価を得ることで、人間の許容評価モデルを対話システムへの実装を目指す。

2.1 パラ言語情報と快不快次元

本研究は、ユーザとシステムとの対話におけるパラ言語情報に着目する。本研究のパラ言語情報は森ら [7] の定義に従う。パラ言語情報は発話音声を書き言葉にしたときに失われてしまう情報を指す。例えば、あきれた「そうですか」という発話を書き言葉にしたとき、「あきれた」という情報は失われ、どのような「そうですか」かはわからない。パラ言語情報は数多の種類の情報を有しているため、コミュニケーションの豊かさを再現するためには特定の種類の情報に限定せず多様な感情・気分などを包括した分析が必要である。そのため本研究では心的状態として、扱うパラ言語情報の種類を限定せず、多くの心理学研究で主要な感情の次元として扱われる快不快次元を扱う。

2.1.1 快不快識別器

本研究において、快不快状態を数値で扱う上では、フェアリーデバイセズ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 [8] を用いた SVR (Support Vector Regressor) である。快不快の推定のために、UUDB[9]（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。快不快識別器は主観評価により人手の評価とみなせることを確認した。

2.2 音声試料

2.2.1 発話内容

ユーザ発話は、特定の感情の影響を比較的受けにくく、直前に接続する文脈によって強い当事者意識で発話できる「連絡を待っています」を用いる。システム応答は、汎用性が高く発話時間がある程度長い相槌である「そうですか」を用いた。

2.2.2 ユーザ発話

ユーザ発話には [6] で収録した、20〜60 代の 9 名（女性 5 名男性 4 名、被験者名を A〜I と呼称）による 5 段階の快不快状態（強い不快・弱い不快・平静・弱い快・強い快）が表出された音声を用いる。9 名それぞれについて以下の要件を満たす 5 段階の快不快状態各 1 音声（合計 45 音声）を選出した。

1. 5 段階の快不快状態の音声の識別値が幅広い

2. 5 段階の快不快状態が均等であるとした時の相関が大きい

9 名ごとの用いた音声の相関係数を表 1 に示す。

表 1: 選出したユーザ発話音声の快不快状態と識別値の相関

被験者名	相関係数
A	0.976
B	0.993
C	0.984
D	0.997
E	0.993
F	0.994
G	0.986
H	0.998
I	0.998

2.2.3　システム応答

パラ言語情報は数多の種類の情報を有しており、特定の種類の情報だけではコミュニケーションの豊かさを再現することは難しい。そのため、システム応答として基本 6 感情をステレオタイプで発話した音声ではなく、表現豊かな多様な音声が必要となった。そこでシステム応答には、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」[10] から抜粋して使用した。本研究では「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 45 種類用意した。45 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：-0.069 〜0.807）している。

2.3　許容評価

ユーザ発話「連絡を待っています」45 音声とシステム応答「そうですか」45 音声を総当たりで接続し、先行発話がユーザ発話、後続発話がシステム応答となる、2025 の対話音声資料を用いて許容評価実験を行った [6]。許容評価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ため、音声試料のユーザ発話が被験者自身の発話であるという仮定が必要不可欠である。よって被験者には、作成した対話音声資料を一つずつ聴取し、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（4. をどちらでもないとし、1. に近づくほど許容できない、7. に近づくほど許容できるとした）で回答してもらった。以後、得られた値を許容評価値、2025 音声それぞれの許容評価値の平均を許容評価値平均と呼ぶ。なお、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。また、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。本実験はクラウドソーシングサイトであるクラウドワークス（https://crowdworks.co.jp/）で実施した。被験者は 20 代から 60 代の男女合計延べ 500 名である。ユーザ発話とシステム応答の識別値と許容評価値平均を図 1 に示す。横軸はユーザ発話の識別値、縦軸がシステム応答の識別値である。凡例にあるマーカの色はそれぞれユーザ発話とシステム応答の識別値が対応する音声の許容評価値の平均の大きさを示している。

3　モデル化への検討

本研究はシステム応答へのユーザの許容評価モデルを構築し、対話システムへ実装することを目指している。そのため、ユーザ発話の快不快状態に応じた許容評価モデルの検討を行う。図 2 は横軸にユーザ発話の識別値、縦軸に許容評価値平均をとり、凡例にシステム応答の識別値を示したものである。

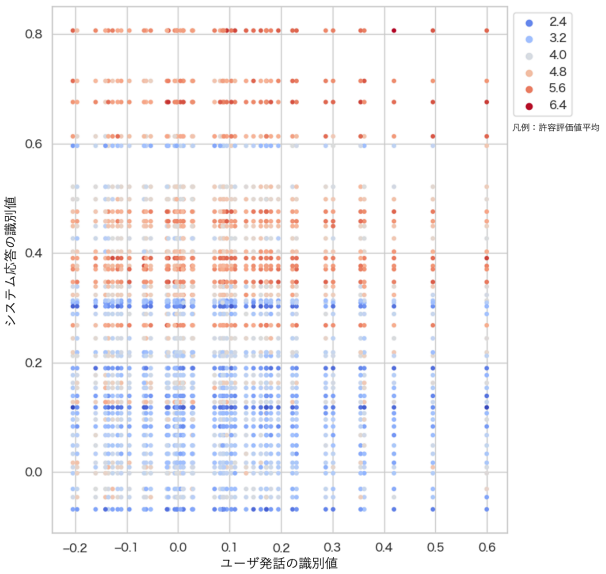

図 1: ユーザ発話とシステム応答の識別値と許容評価値平均 (1)

本稿ではユーザ発話の快不快状態を快群と不快群の 2 群に分類し、分析を行った。快不快状態を 2 群に分類するにあたり、2.1.1 で実施した主観評価を参考とした。主観評価にて快と不快のどちらでもないとされる音声の識別値の平均値（0.187）を算出し境界とすることで、ユーザ発話 45 音声を快群と不快群に分類した。図 2 に示す赤線が 2 群の境界である。2 群について、ユーザ発話数、許容評価値平均の個数および許容評価値平均が 5 以上を許容できるとしたときの音声 (「許容できる音声」と呼称) の個数を表 2 にまとめる。

表 2: ユーザ発話の快不快状態による許容評価の違い

快不快状態	不快群	快群
ユーザ発話数	35	10
許容評価値平均の個数	1575	450
許容できる音声の個数	291	109

表 2 より、許容できる音声の割合について不快群が 291/1575（約 18%）に対して快群が 109/450（約 24%）であることがわかる。このことから、ユーザ発話の快不快状態が不快のときよりも快のときの方が、ユーザはシステム応答をより許容できる傾向にあることが推測できる。さらに、許容評価値平均が 5 以上の音声について、横軸をシステム応答の識別値、縦軸を許容評価値平均とする 2 群それぞれの散布図を図 3 と図 4 に

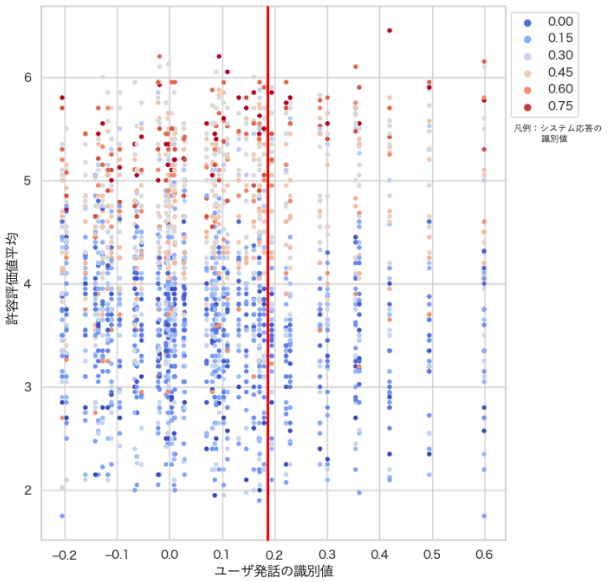

図 2: ユーザ発話とシステム応答の識別値と許容評価値平均 (2)

示す。

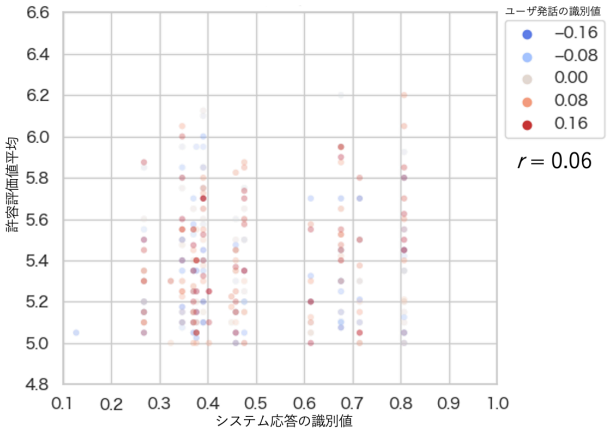

図 3: 許容できる音声：不快群

図 3 および図 4 より不快群は快群に比べて許容評価値平均が 5.5 以上の音声の割合が少ないことがわかる。また、システムの識別値と許容評価値平均の相関係数について、不快群が r=0.06（相関なし）であるのに対し、快群は r=0.39（弱い正の相関）であった。さらに、等分散を仮定（p >.05）した t 検定を行ったところ許容評価値平均の平均に有意差（p <.05）が見られた。

4　おわりに

本稿では、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の調査の結果から対話システムへ実装する許容評価モデルの検討を行った。ユーザ発話の快不快状態を不快群と快群に分類し分析を行ったところ、不快群に比べて快群の方が許容できる音声の割合が多く、許容評価値平均も有意に大きいことがわかった。また、快群においてシステム応答の識別値と許容評価値平均に弱い正の相関がみられたものの、不快群においては相関が見られなかった。このことは、ユーザ発話の快不快状態が不快のときと快のときではシステム応答の識別値が許容評価に及ぼす影響が異なることを示唆している。そして、ユーザ発話の快不快状態によって許容評価モデルが異なることが推察できる。

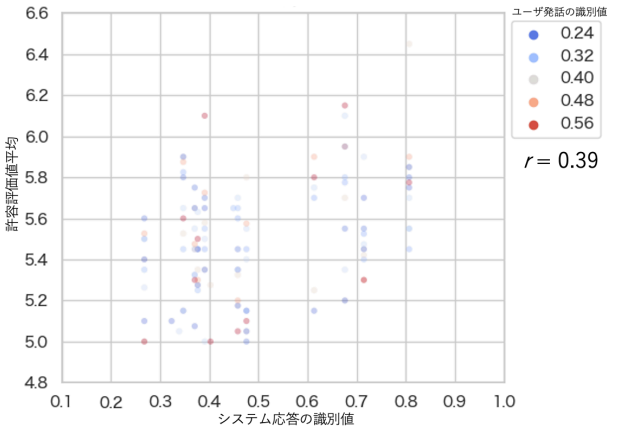

図 4: 許容できる音声：快群

本稿ではユーザ発話の快不快状態を不快群と快群に分類したが、快不快状態は快と不快の 2 群ではなくもっと複雑に分類できると推測できる。快不快状態をどのような範囲で分類するかが今後の課題である。さらに本研究は対話継続欲求のメカニズムを解明するため人間の許容に着目し、本稿では快不快次元に基づくモデルの構築を検討したが、快不快次元だけが許容判断に寄与する感情次元ではないことが明らかになった。そのため、今後課題として許容評価モデルの構築にあたり快不快次元以外の感情次元についても検討することが挙げられる。

謝辞

本研究は、JST 次世代研究者挑戦的研究プログラム JPMJSP2128 の支援を受けたものです。また、本研究は、フェアリーデバイセズ株式会社との共同研究により実施されました。音声感情識別器をご提供いただいた同社に感謝いたします。

参考文献

[1] 株式会社住環境研究所・生涯健康脳住宅研究所: 会話促進により生活改善の効果を確認—高齢者個人宅におけるコミュニケーションロボットの実証実験結果—, URL:https://www.sekisuiheim.com/info/press/20180320.html[閲覧日 2022 年 2 月 7 日]

[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] 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)

[4] 竹内勇剛, 片桐恭弘: ユーザの社会性に基づくエージェントに対する同調反応の誘発, 情報処理学会論文誌, Vol.41, No.5, pp.1257-1266 (2000)

[5] 菊池浩史, 楊潔, 菊池英明: 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–パラ言語に着目して–, *HIA* シンポジウム, *2020*, P-43 (2020)

[6] 菊池浩史, 楊潔, 菊池英明: 音声対話システムにおけるユーザの許容範囲を考慮した多様な同調応答の検討, *HIA* シンポジウム, *2021*, P-27 (2021)

[7] 森大毅, 前川喜久雄, 粕谷英樹: 音声は何を伝えているか–感情・パラ言語情報・個人性の音声科学–, コロナ社 (2014)

[8] 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)

[9] 「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]

[10] 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)