

音声対話システム利用時における感情表出誘因性に関する考察 A study of incentive to express emotions on spoken dialogue system

宮島 崇浩[†] 菊池 英明[‡] 榎松 明[†] 白井 克彦[†]
Takahiro Miyajima Hideaki Kikuchi Akira Kurematsu Katsuhiko Shirai

1. はじめに

認識精度やノイズ耐性の向上に伴い、音声画像認識技術の利用が拡大している。そして、従来の感情認識技術などと融合し、音声や画像から人間の状態を推測して、それに従ったサービスを提供することが期待されている。例えば、デジタルカメラに「笑い顔検知機能」が搭載されたりすることなどから、その一端を垣間見ることが出来る。また、近年のロボット研究の分野では、ロボットが感情を表出・理解し、それに応じたサービスを提供するシステムが提案されるなど、その利用は今後さらに広がっていくことが期待される。

ここで我々は、ひとつの疑問を投げかけたい。それは、「人間は、機械と人間と対峙したときと同レベルの感情表出を行うものなのだろうか」ということである。直観的には、同レベルでは感情表出を行わないことは想像できるが、要因なども含めてこの仮説は明らかではない。そこで本稿では、音声対話システムにおいて、人間の感情表出に影響を与える要因（感情表出誘因性）の存在を明らかにするための予備的な考察を行う。

2. 感情表出誘因性

2.1 仮定

本研究では、コミュニケーションに関係するすべての物事（モダリティ、外的環境、システムへの信頼、リアルタイム性、相手が人間かどうか、など様々）に、人間の感情表出の度合へ影響を及ぼす要素が存在すると考え（表 1）、それを「感情表出誘因性」と仮定する。

感情表出誘因性の存在そのものは、従来の様々なヒューマンインターフェース利用において、間接的に示唆されてきたと考えられる。例えば、音声認識を行う際、ユーザが音声認識に失敗したとき、ユーザは大きい声で言葉を言い直しその韻律はある意味不自然なものになる可能性が高いだろう。また、ロボットと会話するシチュエーションで、生身の人間とは異なるロボットの挙動（自分の発話に対する反応速度）が人間を戸惑わせ、不自然な発話を誘導することは容易に想像できる。この点について別の見方をすると、音声対話システムが人間と異なる挙動をすればするほど、人間の応答は不自然なものになっていくと予想することができる。具体的に言い換えれば、「音声認識率」と、「感情表出誘因性」は相関があると予測できるし、「システムからの応答速度」も関連があると推測できる。

さらに、別の例を考えてみる。音声や文字によるチャットシステムも対話システムの一環であるとする、文字を

表 1: 人間の感情表出と関連があると推測される要素

要素	具体例	システムに対する感情表出量
リアルタイム性 (レスポンス性)	チャットシステム、掲示板、音声対話における時間差、システムからの応答速度	低いほど低下
信頼性	音声認識率、システムエラー率	低いほど低下
対話相手種別	相手がロボットだという意識	人間のモデルから遠いほど低下
モダリティ種別	音声、画像、テキスト	情報量が少ないほど低下
慣性	IT への慣れ、システムへの慣れ	不慣れなほど低下

一文字打つごとに表示する場合と、一行ごとに表示する場合、さらにそれぞれに対してタイピング速度を変えた場合で比較したときに差異が見られれば、リアルタイム性には感情表出誘因性があることになる。

2.2 従来研究からの考察

人間の発話に対して、音声対話システムが誤解応答した場合におけるユーザ発話の特徴量の変化の分析[1]においては、ユーザの困惑状態や怒りの状態の検出のためにポーズ長、パワー、F0 の分析が行われた。そして、ポーズ長に関しては、システムの誤解（エラー）が生じた場合に長くなること、パワーに関しては個人差が大きい、誤解を解くための強調として極端な増大が見られるケースがあること、そして F0 に関しても個人差が大きい、7 割以上のケースで F0 の指標における値の増大がみられるなどの変化があることが報告されている。

また、音声インターフェースに対する慣れの分析[2]では、慣れが大きくなるほどシステムへの応答時間が短縮することが報告されている。

2.3 定義

従来研究からすれば、システムエラーによる感情表出量への影響が存在するのは明らかである。また、慣れの問題に関しては、応答時間と感情表出が比例するとするならば、やはり影響があると考えることができる。これらの点から共通で言えることは、人間とシステム間の対話には、人間同士の対話では存在しない外的要因が発生し得て、それが本来あるべき感情とは異なる感情を誘発する可能性があるという事である。

以上を踏まえて感情表出誘因性について再考したい。「システムエラーに対する怒り」に見られるような、単に感情を誘発する性質というだけでは、本稿で明らかにしたい性質とは異なる。そこで、感情表出誘因性とは、人間にとって、対人間と対システムで感情の表出が同等となるような性質を指すことと定義する。この定義のもと、感情表出誘因性が高いインターフェースを構築することが、人間の感情情報を利用した次世代の対話システムを高度化させていくために必要になると考え、感情表出誘因性に関する予備実験を行うことにした。

[†] 早稲田大学理工学術院

Faculty of Science and Engineering, Waseda University

[‡] 早稲田大学人間科学学術院

Faculty of Human Sciences, Waseda University

表 2：本稿におけるモダリティ毎の特徴の位置付け

モダリティ	距離感	顔情報	音声情報	即時性
FtF	近い	あり	あり	あり
Video Chat	遠い	あり	あり	あり
Voice Chat	遠い	なし	あり	あり
Text Chat	遠い	なし	なし	なし

表 3：被験者 A のモダリティ毎の F0 平均値の比較

モダリティ	最大値	最小値	レンジ	平均値
FtF	388.9	161.7	227.2	268.0
Video Chat	334.7	156.3	178.4	247.7
Voice Chat	357.5	161.6	195.9	240.2
Text Chat	278.6	162.2	116.4	213.3

単位：Hz

表 4：被験者 B のモダリティ毎の F0 平均値の比較

モダリティ	最大値	最小値	レンジ	平均値
FtF	288.0	156.6	131.4	210.8
Video Chat	290.5	144.0	146.5	200.6
Voice Chat	320.0	146.4	173.6	215.0
Text Chat	237.6	134.9	102.7	180.0

単位：Hz

3. モダリティの違いによる感情表出量の比較

先行研究では、システムへの怒りなどを検証するものとしてパラ言語情報の分析が行われたが、怒りを誘因する要素などはバイアスであると考えられる。そこで、バイアスができるだけ少なくなる状況を設定したうえで、パラ言語情報の分析を行うことにした。

3.1 実験

対面状況（Face to Face：FtF）、ビデオチャット、ボイスチャット、（音声情報を取得するために、音声入力からの）テキストチャットと、モダリティを変更しながら対話を継続することで、相互に伝達される情報の種類を変化させていき（表 2）、被験者の感情表出がどの程度減少していくのかを調べ、感情表出誘因性を比較する。このとき、音声入力からのテキストチャットでは Wizard of OZ 方式を用いるため、システムエラーに対するストレスは存在しないものとする。

この設定のもと、20 代女性の被験者 2 名に、FtF、ビデオチャット、ボイスチャット、音声入力方式テキストチャットの順で 3 分づつ対話を行ってもらった。

被験者同士は日ごろから仲が良く、頻繁に様々な話題を交わしている。これらの対話中における音声を録音し、分析を行うこととした。

3.2 分析

一般的に、感情表出を示す指標は多数あるが、今回は 1～2 秒の発声を 1 発話単位として、F0 の最大値、最小値、レンジ、平均値を用いた。そして、各 3 分間のデータから 10 個づつ取り出して平均した。その結果を表 3、4 に示す。

被験者 A、B ともに、モダリティが変化することによりパラメータが変化している。特に顕著な違いは F0 の最大値に表れている。レンジや平均値は、発話内容のイントネーションなどの影響が大きいせいから、2 者間で変化の仕方に差が見られたが、特にテキストチャットにおいて小さくなる傾向が見られた。最小値には目立った傾向は見られなかった。

表 5：ボイスメールによる追加実験の F0 値比較

被験者	最大値	最小値	レンジ	平均値
A	294.2	148.0	146.2	206.8
B	259.6	133.6	126.0	193.8

単位：Hz

3.3 考察

実験結果から単純に考えれば、音声によるパラ言語・非言語情報の相互伝達量および即時（リアルタイム）性と、F0 の変化量には相関があると推測できる。また、FtF、ビデオチャット、ボイスチャットには明確な差があるとは言えない結果になったが、これはボイスチャットと電話が似通っている等の人間にとって慣れきった違いであるからだと考えることができる。

F0 は感情研究における最も重要な指標のひとつであり [3]、この値の変化が見られたことで、前述の各パラメータに感情表出誘因性の存在が期待できる結果になったと考えられる。

3.4 検証

前述の結果の検証を行うために、追加実験を行った。先の実験と同じ被験者で、録音器に収録した音声メッセージを交換し合うという、ボイスメールのようなシチュエーションを用意した。これは、ボイスチャットからリアルタイム性を除いたと考えることもできるし、テキストチャットにパラ言語情報を付加したと捉えることもできる。

この設定のもと、同様に F0 値の解析を行った結果を表 5 に示す。表 3、4 と比較すると、ボイスチャットとテキストチャットの間あるいは同程度の値を示していることが分かり、少なくとも今回の被験者においては、前述したリアルタイム性やパラ言語情報量と感情表出誘因性との関係が裏付けされる結果となった。

4. まとめ

本稿では、人間とシステム、あるいはシステムを介した人間同士のコミュニケーションの様々な要素に、感情の表出量を左右するパラメータが存在すると仮定し、それを「感情表出誘因性」と定義した。そして、F0 を指標としたとき、リアルタイム性や感情の相互伝達量に感情表出誘因性がある可能性が示唆された。

今後の課題として、被験者の数をさらに増やすことと、スピーチレート等の様々な指標から分析を行うことが挙げられる。また、相手が機械だからなのか、正確に聞き取ってくれることに不安があるのか、単に緊張からなのか等、他にも考えられる様々な理由との共通の因子を明らかにする必要がある。

参考文献

- [1] 菊池 英明, 林 美穂子, 青山 一美, 白井 克彦, “音声対話システムの誤解状態におけるユーザ応答の分析”, 特定領域研究(B)「韻律に着目した音声言語情報処理の高度化」研究成果報告書(平成 12 年度), pp.133-138 (2001).
- [2] 菊池 英明, 北村 泰司, 本島 嘉朗, 山田 哲史, 宮澤 幸希, 白瀬 綾子 “音声インタフェース評価における慣れの影響の分析”, 情報処理学会音声言語処理研究会(SLP), pp.977-102 (2007).
- [3] 濱 治世, 鈴木 直人, 濱 保久, “感情心理学への招待”, 新心理学ライブラリ 17, サイエンス社(2001).