

第 58 回 東海ファジィ研究会 in 豊橋（がま研） 予稿集



主催：日本知能ファジィ学会東海支部

協賛：IEEE 名古屋支部

日本経営システム学会中部支部

日時：2025 年 8 月 4 日(月)～5 日(火)

会場：豊橋商工会議所 508 会議室

第58回東海ファジィ研究会（2025年8月4日～5日）プログラム

8月4日（月）

13:50-13:55 【開会挨拶】 副支部長(庶務幹事) 早瀬 光浩（椋山女学園大学）

13:55-15:15 セッション1 座長：加藤央昌（愛知淑徳大学）

S1-01 ロボットのデザイン要素が人の心理に与える影響

布川智義 上洞萌子 加納政芳

S1-02 運転行動教示ロボットにおける幼児感情表出モデルの基礎的な検討

成田琴美 ジメネスフェリックス 吉川大弘

S1-03 メルスペクトrogramを用いたTransformerによる音声デノイズ手法の一試み

小島巧実 高野敏明

S1-04 音響パレイドリアの研究：その後の展開

小田哲久 河辺義信 杉浦伸 山田洋巳

15:15-15:30 休憩

15:30-16:30 招待講演 司会：野崎 佑典（名城大学）

人と対話エージェントの理論指向型インタラクション研究

宮本 友樹（電気通信大学）

16:30-18:00 休憩

18:00-23:30 ナイトセッション 座長：吉川大弘（鈴鹿医療科学大学）

討論テーマ1「大学教育において『責任ある生成AI利用者』を育成する方法とは」

討論テーマ2「ファジィ理論の次の可能性と人間中心AIへの応用」

8月5日（火）

9:20-10:00 セッション2 座長：中垣雄太（三友工業株式会社）

S2-01 LTLBCの差分電力解析に対する耐タンパ性評価

野崎佑典 竹本修 吉川雅弥

S2-02 知識蒸留を用いた音声の感情推定手法の検証

森田賢太

10:00-10:05 【閉会挨拶】 東海ファジィ研究会・支部長 加納政芳（中京大学）

[会場：豊橋商工会議所 508会議室](#)

〒440-8508 愛知県豊橋市花田町字石塚42-1 TEL 0532-53-7211

ロボットのデザイン要素が人の心理に与える影響

布川智義¹ 上洞萌子² 加納政芳²

¹ 中京大学大学院 ² 中京大学

1 はじめに

近年，日本社会において精神的ストレスに関する問題が深刻化しており，メンタルヘルス対策の必要性が高まっている．このような状況の中で，新たな心理支援手法としてロボット・セラピーが注目されている [1]．ロボット・セラピーとは，動物とのふれあいを通して精神的な安らぎを与えるアニマル・セラピーに着想を得た心理療法であり，動物の代わりに動物型ロボットを用いるものである．ロボット・セラピーは，アニマル・セラピーと同様の心理的効果や生理的効果が得られるとされており [2]，加えてアニマル・セラピーよりも衛生面・安全性に優れ，場所や時間を問わずに実施できるという利点がある．その一方で，ロボットが生成する単純な振る舞いでは，動物とのコミュニケーションによって育まれるような心理的絆が培われにくく，ふれあい意欲が低下するという欠点がある [3]．ロボット・セラピーにおけるふれあい意欲についての先行研究では，柔らかい触感のロボットは硬い触感のロボットよりも心理的・生理的ストレス緩和効果が高く，ふれあい意欲の低下が抑制されることを報告している [4]．このように，ロボットの触感に関する研究はあるものの，その他の手法によるふれあい意欲への影響については不明である．そこで本研究では，ロボットのデザイン要素に着目する．

デザイン要素による印象変化について，作田らの研究 [5] では，単純な図形刺激を用いて，目の有無による印象の変化を検討したところ，目を付与することで単なる図形としてではなく顔のようなものとして捉えるようになり，印象が変化することが報告されている．中でも，直線的で規則的な図形では，目の無い方が力強く，曲線的で規則的な図形では，目の無い方が弱々しい印象をもたらすことが示されている．このように図形の形状や目の有無によって印象が変化することから，本研究では，ロボットの形状および目の有無を変化させた場合におけるふれあい意欲への影響やロボットに対する印象を調査する．

2 実験

2.1 予備実験

まず，本実験に使用するロボットの生地および目の位置を選定する予備実験を大学生 10 名を対象に行った．生地を選定する実験では，実験参加者に対して「毛並みのある生地」，「ガーゼ生地」，「ボア生地」，「タオル生

表 1: 生地選定の因子負荷量

変数	因子 1	変数	因子 2
落ち着き	0.955	かわいさ	1.070
好き	0.741	愛らしさ	0.852
快適	0.686	面白さ	0.560
安心感	0.654	暖かさ	0.448
触り心地	0.630	癒し	0.434
飽き	0.623	親しみやすさ	0.358
癒し	0.568	触り心地	0.340
親しみやすさ	0.563	安心感	0.302
暖かさ	0.449	人間味	0.272
人間味	0.139	好き	0.148
面白さ	0.108	快適	0.008
愛らしさ	0.087	飽き	-0.018
かわいさ	-0.117	落ち着き	-0.040

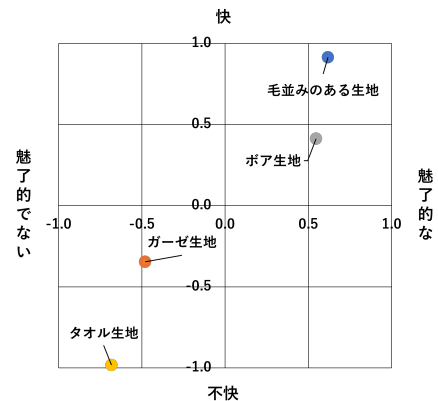


図 1: 生地選定の因子分析結果

地」の 4 種類に触れてもらい，生地の印象について，形容詞・形容動詞対 13 項目から 5 件法で回答してもらった．表 1 に因子分析の結果を示す．この結果から，第一因子を「快-不快」，第二因子を「魅力的な-魅力的でない」とした．図 1 に各生地の印象をプロットした散布図を示す．図から，最も快適で魅力的である「毛並みのある生地」を使用することとした．

目の位置を選定する実験では，1cm 間隔ごとに目の幅を変え，最も好印象のロボットを実験参加者に回答してもらった．その結果，目の幅を 8cm とした．

2.2 本実験

2.2.1 ロボット開発

予備実験の結果をもとに，本実験で使用するロボットを開発した．図 2 に開発したロボットを示す．ロボットは，ユカイ工学社製 PetitQoobo を参考に開発した．ロ



図 2: 使用したロボット

ボットのデザインは、作田らの研究 [5] を参考に、「丸い形状で目の無いロボット」(以降、丸型目無ロボット)、「四角い形状で目の無いロボット」(以降、四角型目無ロボット)、「丸い形状で目を有するロボット」(以降、丸型目有ロボット)、「四角い形状で目を有するロボット」(以降、四角型目有ロボット)の4種類とした。また、ロボットは撫でると尻尾を動かす仕様である。

2.2.2 実験手順

図 3 に実験手順を示す。

まず、実験参加者には、(1) 2 分間の計算課題を行っ

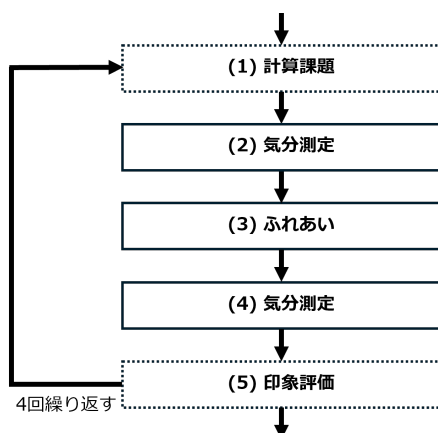


図 3: 実験手順

てもらふ。計算課題は、隣り合う 1 桁の整数をできるだけ速く、正確に加算するものとした。計算課題を使用した理由は、速さと正確性を条件に課することで精神的ストレスの要因とされる仕事量の多さと質の高さを擬似的に再現するためである。

次に、(2) ロボットとのふれあいの前に気分測定アンケートに回答する。気分測定アンケートでは、気分を表す因子として気分尺度 [6] を採用した。気分尺度は、感情状態を示す緊張、抑鬱、怒り、混乱、疲労、活気の 6 つの因子により構成されている。調査方法は、短期的な気分変化を測定するのに適した一時的気分尺度 [7] 18 項目に対して 5 件法で回答を求め、各因子の下位尺度 3 項目の合計値 (3 から 15 点) を因子得点として算出する。緊張、抑鬱、怒り、混乱、疲労が緩和した場合、因子得点は減少し、活気が向上した場合、因子得点は増加する。本実験では、心理的ストレス緩和効果を緊張、抑鬱、怒り、混乱、疲労が緩和し、活気が向上する効果として評価する。

続いて、(3) 2 分間ロボットとふれあってもらふ。ロボットとの積極的なふれあいを促すために、実験者から「よしよししてあげると尻尾が動く」や「叩くと激しく尻尾を動かす」などと教示するものとする。

そして、(4) ロボットとのふれあいの後、(2) と同様の気分測定アンケートに回答する。

最後に、(5) ロボットの印象評価アンケートに回答する。アンケートでは、森ら [8] を参考に、形容詞・形容動詞対 13 項目を SD 法を用いて 5 件法で回答してもらった。最後に自由記述にも回答してもらった。

以上の手順を 4 種類のロボットすべてで行ってもらふ。なお、実験手順は林ら [4] を参考にした。

実験参加者は大学生 26 名 (男性 13 名、女性 13 名) であった。

3 結果・考察

まず、気分測定アンケートの結果について確認する。図 4 にふれあい前後の気分測定アンケートの結果を示す。ふれあい前後における一時的気分尺度の各下位項目の平均点に対して、ウィルコクソンの符号付順位検定を行った。その結果、ロボットの形状および目の有無に関わらず、ふれあい前後で緊張、抑鬱、怒り、混乱、疲労が有意に緩和し、活気が有意に向上した (ただし、丸型目有ロボットの抑鬱については $p = 0.058$ の有意傾向である)。

次に、図 5 にふれあい前後の得点差平均 (前 - 後) を示す。ふれあい前後の得点差平均に対して一元配置分散

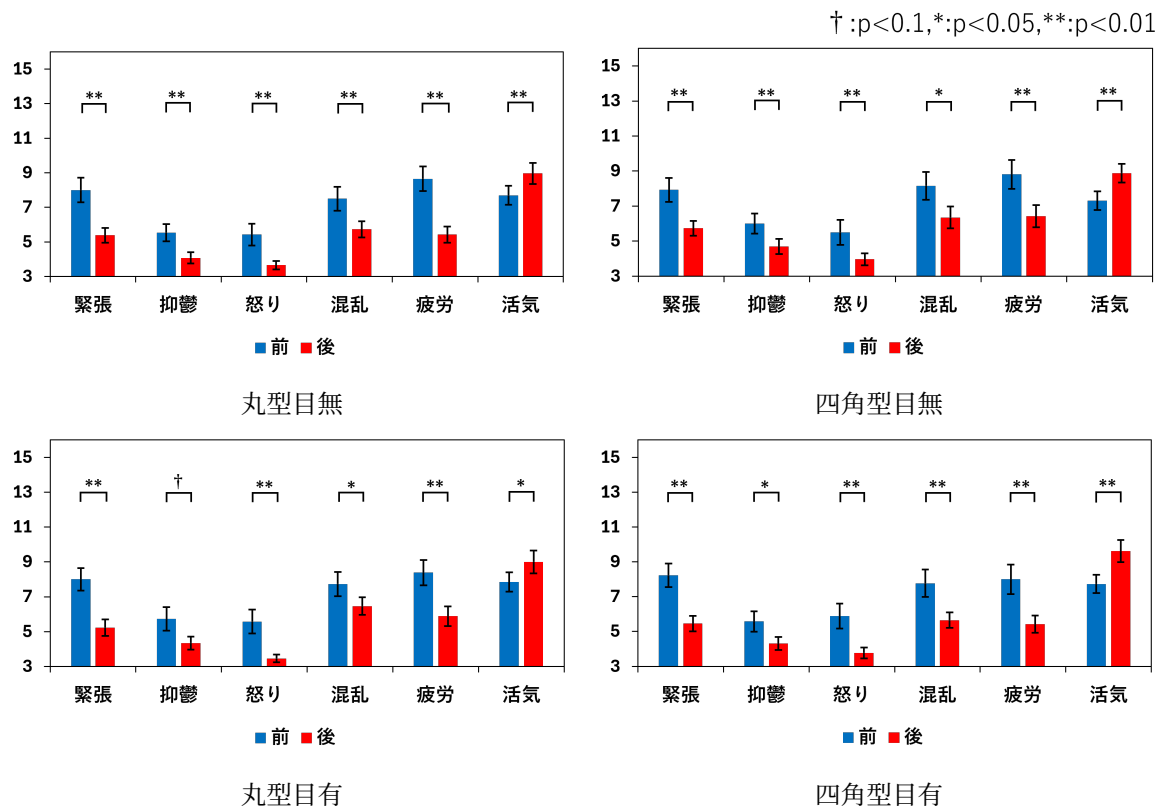


図 4: 気分測定アンケートの結果。エラーバーは標準誤差を示す。

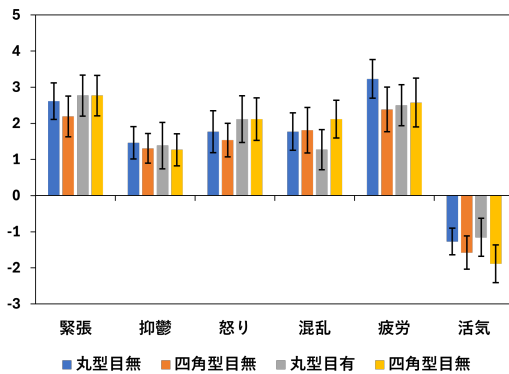


図 5: ふれあい前後の得点差平均。エラーバーは標準誤差を示す。

分析（対応あり）を行った。その結果、ロボットの種類による有意な主効果は認められなかった。ロボットの形状や目の有無によって心理的ストレス緩和効果に差が生じなかった要因として、柔らかい触感のロボットを使用したこと、予備実験によって最も好印象の生地と目の位置を選定したことが挙げられる。これらの要因により、いずれのロボットとふれあった場合においても心理的ストレス緩和効果があり、また、条件間で差が生じなかつ

表 2: 印象評価の因子負荷量

変数	因子 1	変数	因子 2
親しみやすさ	0.986	落ち着き	0.703
愛らしさ	0.745	安心感	0.658
かわいさ	0.717	飽き	0.642
あたたかさ	0.684	癒し	0.547
面白さ	0.618	うちとけやすさ	0.538
好ましさ	0.542	快適さ	0.459
生物性	0.457	好ましさ	0.293
癒し	0.439	生物性	0.208
うちとけやすさ	0.398	かわいさ	0.164
快適さ	0.186	あたたかさ	0.144
安心感	0.167	愛らしさ	0.132
飽き	-0.026	面白さ	-0.047
落ち着き	-0.089	親しみやすさ	-0.141

た可能性がある。

続いて、ロボットの印象評価の結果について確認する。まず、形容詞・形容動詞対 13 項目に対して因子分析を行った。表 2 に因子分析の結果を示す。この結果から、第一因子を「愛嬌のある-愛嬌のない」（以降、愛嬌）、第二因子を「心地よい-心地よくない」（以降、心地よさ）とした。図 6 に各ロボットの印象をプロットした散布図を、図 7 に各因子得点の平均値を示す。各因

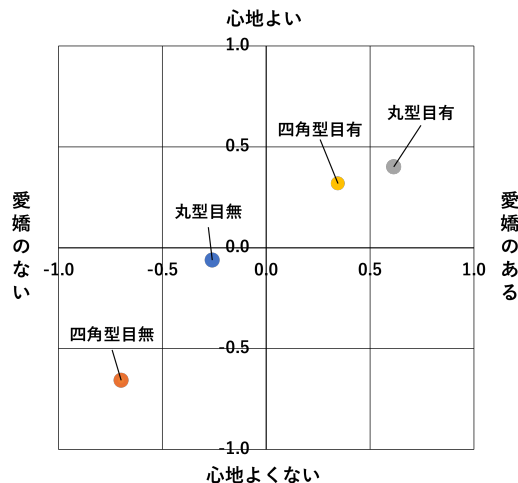


図 6: 印象評価の因子分析結果

子得点の平均値に対して一元配置分散分析（対応あり）を行い、有意差が認められた場合に Bonferroni 法で多重比較を行った。その結果、第一因子の愛嬌では、丸型目有ロボットは丸型目無ロボット、四角型目無ロボットよりも有意に評点が高かった。また、四角型目有ロボットは丸型目無ロボット、四角型目無ロボットよりも有意に評点が高かった（ただし、四角型目有ロボットと丸型目無ロボットは $p = 0.069$ の有意傾向である）。第二因子の心地よさでは、四角型目無ロボットは丸型目有ロボット、四角型目有ロボットよりも有意に評点が低かった。この結果から、ロボットの形状に関わらず、目を有する方が愛嬌があることが示唆される。また、自由記述からロボットの印象を確認すると、ロボットの形状に関して、四角い形状より丸い形状の方が「かわいい」と答えた人が全体の約 42%であり、目の有無に関して、目を有する方が「親しみやすく、かわいい」と答えた人が全体の約 58%，目を有する方が「機械らしさより生物らしさを感じる」と答えた人が全体の約 27%であった。以上の結果から、丸型目有ロボットが最も好印象である可能性が示唆される。丸型目有ロボットは 4 種類の中で最も動物に近い見た目であるため、アニマル・セラピーと似た効果を引き出すことができ、好意的な印象を高めた可能性がある。

4 おわりに

本稿では、ロボット・セラピーにおけるふれあい意欲の低下を抑制するための方法としてロボットのデザイン要素に着目し、ロボットの形状および目の有無を変化させた場合における心理的ストレス緩和効果やロボット

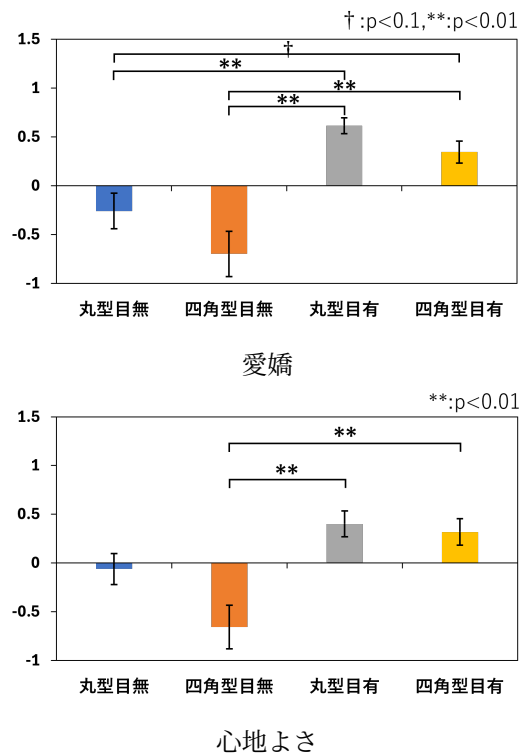


図 7: 各因子得点の平均値。エラーバーは標準誤差を示す。

に対する印象を調査した。実験の結果、心理的ストレス緩和効果には、ロボットの形状や目の有無は影響しないことが示唆された。また、ロボットの印象については、丸い形状で目を有するロボットが最も好印象である可能性が示唆された。

参考文献

- [1] 横山章光, “臨床精神医学から見た、ロボット・セラピーの未来～アニマル・セラピーの知見も鑑みて～”, 計測と制御, **51**(7), 598–602 (2012)
- [2] 浜田利満・他, “ロボット・セラピーの展開”, 計測と制御, **42**(9), 756–762 (2003)
- [3] 米岡利彦, “高齢者施設でのロボット・セラピー”, 計測と制御, **51**(7), 609–613 (2012)
- [4] 林里奈, 加藤昇平, “ロボット・セラピーにおける柔らかい触感の重要性”, 日本感性工学会論文誌, **18**(1), 23–29 (2019)
- [5] 作田由衣子, “「印象」はどこから来るか:物理的属性と性格印象の関連性の検討”, 生活科学部紀要, **57**47–54 (2020)
- [6] 横山和仁, 荒記俊一, “日本語版 poms 手引”, 金子書房 (2008)
- [7] 徳田完二, “一時的気分尺度 (tms) の妥当性”, 立命館人間科学研究, **22**1–6 (2011)
- [8] 森善一・他, “抱きつき人形の印象評価”, 日本感性工学会論文誌, **11**(1), 9–15 (2012)

運転行動教示ロボットにおける 幼児感情表出モデルの基礎的な検討

成田琴美¹ ジメネスフェリックス² 吉川大弘³

¹愛知県立大学大学院 情報科学研究科 ²愛知県立大学 情報科学部

³鈴鹿医療科学大学 医用工学部

1. はじめに

近年、高齢ドライバーによる交通事故が社会的課題となっており、高齢者に対して安全な運転行動を教示するためのロボットの開発が進められている。例えば、実運転環境においてロボットがドライバーに言葉かけを行うことで、安全運転を促進する効果が報告されている[1]。さらに、危険運転映像を用いた学習環境においても、ロボットによる教示は、従来のシステムによる提示に比べて記憶定着に優れる可能性が示唆されている[2]。このように、ロボットとの共同学習は、安全運転の習得に有効であることが明らかになりつつある。

しかしながら、既存の運転行動教示ロボットの多くは、単発的なインタラクションを前提としており、継続的な使用や長期的な関係性の構築を想定していない。しかしながら、長期的な学習効果の維持のためには、高齢者が長期間にわたって繰り返し使用したくなるような魅力を持つロボットが求められる。そのためには、まずロボットに対する印象の向上が重要と考える。

そこで本研究では、高齢者が幼児の行動に対して好意的な印象を抱きやすいという先行研究[3]に基づき、幼児のような発話や感情表出を行う「幼児感情表出モデル」を構築し、既存の運転行動教示ロボットに搭載することを目的とする。本モデルの構築にあたっては、心理学における感情理論であるアージ理論[4]の枠組みを応用する。

本稿では、その基礎的な取り組みとして、本モデルの設計手法を提示するとともに、予備実験を通じてその有効性を検証する。

2. アージ・システム

2.1 アージ理論

アージ理論は、心理学における感情理論の一つであり、感情を「高度な野生的合理性を備えた、遺伝的に規定された環境適応的な選択システム」として捉える枠組みである。この理論においては、感情は特定の目的に従って起動・制御される適応的反応と位置付けられる。

本理論では、目的を伴って起動される感情を「アージ」、およびそれを制御する環境適応的な選択システムを「アージ・システム」と定義する。アージ・システムにおいては、従来の感情概念が「アージ」と「ムード」の2つに区分されており、それぞれ異なる機能的特性を持つものとして扱われる。

2.2 アージとムード

アージとは、特定の目的を伴って起動される感情や心的反応である。たとえば、「私物が壊された」場合には、報復行動などを目的とする「怒りアージ」が生じる。また、「空腹」といった内的信号によって生じる「生理的アージ」も、行動を促す心の働きとして捉えられる。このように、従来は感情とは見なされなかった生理的機能も、アージ理論では感情的な反応として扱われる。

一方、ムードは、明確な目的を伴わない心的状態である。「喜び」や「悲しみ」は活動プランを持たないため、アージではなくムードに分類される。ただし、他者への情報伝達を目的とした場合、これらは「誇示アージ」として扱われることもある。また、ムードはアージの起動に影響を与える要因として機能する。

2.3 アージの働き 4 相

アージ・システムは、「起動相」「意思決定相」「行動相」「事後評価相」の4つの段階から構成され、すべてのアージはこれらの過程を順に経ることで機能する。具体的には、起動相において状況を認知し、意思決定相で適切な行動プランを選択する。続く行動相では、選択された行動を実行し、最後に事後評価相にてその結果を振り返り、学習・反省が行われる。

3. ロボットの概要

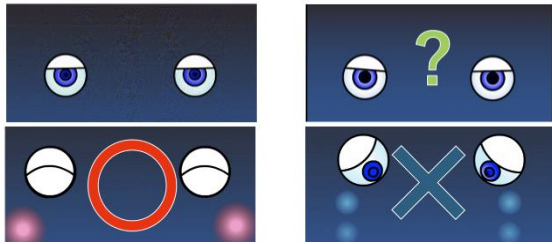
本研究で使用するロボットは、頭部にタブレットを搭載した「Tabot」である。ロボットの外観を図1に示す。上半分にはエージェント、下半分には学習システムのインターフェースが表示される。なお、本稿においては、ロボットの身体動作は実装せず、画面表示によるインタラクションのみを対象とする。

学習システムは、スタート画面から開始し、危険運転に関する映像提示、選択肢の提示、正誤判定という流れで、問題は10問用意されている。学習の最後には、内容の振り返りが表示され、学習セッションは終了する。

また、エージェントの表情は状況に応じて変化する。通常時、学習前後の質問時、正解時、不正解時のそれぞれに対応した異なる表情が表示される。ロボット通常時、質問時、正解時、不正解時における表情を図2に示す。



図 1: Tabot の外観

図 2: 従来ロボットにおける表情
(左上から, 通常時・質問時・
正解時・不正解時)

4. 幼児感情表出モデル

本研究では, 前述したロボットに幼児感情表出モデル (以下, 本モデル) を搭載した. なお, 本モデルの適用範囲は, 学習開始時と終了時のやり取り, 正誤判定時以外の通常時に限定した.

本モデルでは, アージ理論における「アージの働き 4 相」に基づき, 感情の起動から表出に至るまでの一連の処理フローを構築した. 各相における処理内容については, 以下に詳述する.

4. 1 状況認知

本モデルでは, 「学習者の進捗度 P 」と「ロボットの気分 F 」の二つのパラメータを用いて, 感情表出の変化を制御する.

学習者の進捗度 P は以下の式で表現される:

$$P(t) = \alpha \cdot P(t-1) + p(t) \quad (0 < \alpha < 1)$$

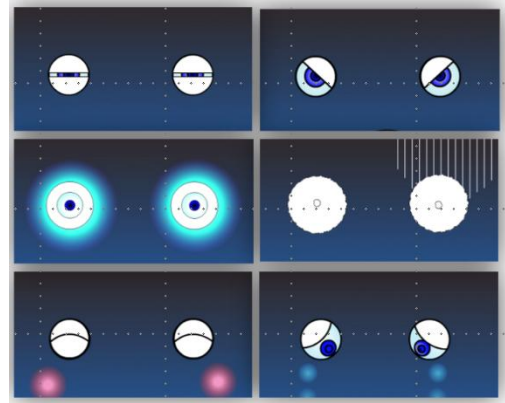
$$p(t) = \begin{cases} \text{correctScore} & \text{正答したとき} \\ 0 & \text{それ以外} \end{cases}$$

ここで, α は進捗の減衰係数, $P(t-1)$ は直前時点の進捗度, correctScore は正解時に与えられる値 (正解時に正の値, 誤答時には 0) である. つまり, 時間経過とともに進捗度が緩やかに減衰し, 学習者が正答するたびに進捗度が上昇するモデルとなっている.

ロボットの気分 F は, 以下の式で表現される:

$$F(t) = F(t-1) + K \cdot \{P(t) - P(t-1)\} - \text{danderCheck}$$

$$\text{danderCheck} = \begin{cases} 0.1 & \text{危険時間中} \\ 0 & \text{それ以外} \end{cases}$$

図 3: 提案ロボットにおいて追加する表情
(左上から, 生理的, 怒り, 驚き,
恐れ, 喜び, 悲しみ)

ここで, K は進捗度比例係数, danger Check は危険時間中に一定量減少させる項である. すなわち, ロボットの気分は学習者の進捗の増減に応じて上下し, 危険運転映像の危険時に低下するように設計されている.

4. 2 起動相

本モデルでは, Ekman が提唱する基本 6 感情 (怒り, 驚き, 恐れ, 喜び, 悲しみ, 嫌悪) [5] を基礎とし, これらをアージ・システムの枠組みに位置づけて用いる. 具体的には, 怒り・驚き・恐れは, 行動目的を伴って起動される一般的なアージとして扱うことができる. 一方, 喜びおよび悲しみは, 前述の通り, 誇示アージに分類される. また, 嫌悪は, 身体的・生理的な状態の悪化に反応して生じる感情であることから, 生理的アージとして扱う. 各表情を図 3 に示す.

また, 各アージの起動条件について, 以下に示す.

(i) 生理的アージ

生理的アージは, 身体的・心理的な負荷に対する反応として発生するとされる. 本モデルでは, 学習者の進捗度が一定時間にわたって低下し続けている状況を, 負荷の蓄積とみなし, 怒りアージの起動条件とする. 具体的には,

$$P(t-1) > P(t)$$

となった回数を N とし, 条件式は以下のとおりである:

$$N > N_{\text{physio}}$$

ここで, N_{physio} は, 進捗低下の連続回数を示す.

(ii) 怒りアージ

怒りは, 自身の「権限的縄張り」が他者に

よって侵害された際に生起する感情とされる。本モデルでは、学習者の進捗度が一定時間にわたり低下し続けている状態を、自律的な学習の妨害として解釈し、怒りアージの起動条件とする。具体的には、

$$P(t-1) > P(t)$$

となった回数を N とし、条件式は以下のとおりである：

$$N > N_{anger}$$

ここで、 N_{anger} は、進捗低下の連続回数を示す。なお、生理的アージと怒りアージの条件式は同様であるが、後述の感情の優先順位に基づき、生理的アージを優先的に表出するものとする。

(iii) 驚きアージ

驚きは、予期に反する事象が生じた際に生起する感情とされる。本モデルでは、進捗度の最大値を更新した場合に、「予想外の変化」とみなし、驚きが発現するとする。式で表すと以下の通りである：

$$P(t) > P(T_{max})$$

ここで、 T_{max} はこれまでの進捗度の最大値を記録した時刻を指す。

(iv) 恐れアージ

恐れは、危険を制御できる可能性が極めて低いときに発生する感情とされる。本モデルでは、ロボットの気分が急激に減少したときに発現するとする。式で表すと以下の通りである：

$$F(t-1) - F(t) > \theta_{fear}$$

ここで、 θ_{fear} はロボットの気分の減少幅を示す。

(v) 誇示アージ（喜び・悲しみ）

誇示アージは、自己の改善状態を他者に認知させようとするときに発生する感情とされる。本モデルでは、ロボットの気分において、基準時点と比較して変化が閾値を超えた場合に、誇示アージが発生すると定義する。式で表現すると以下の通りである。

$$\begin{aligned} F(t) - F(T_{baseline}) &> \theta_{joy} && \text{喜び} \\ F(T_{baseline}) - F(t) &> \theta_{sadness} && \text{悲しみ} \end{aligned}$$

ここで、 θ_{joy} と $\theta_{sadness}$ はそれぞれ、ロボットの気分の増加幅と減少幅を示す。

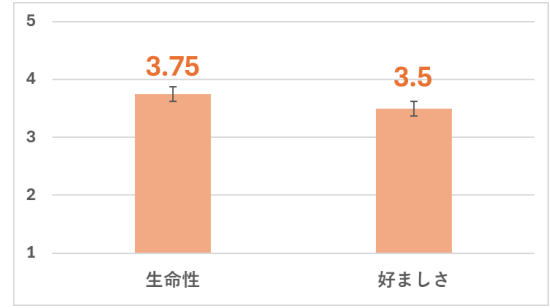


図 4: Godspeed Questionnaire による結果

4. 3 意思決定相

起動相においては、状況認知に基づき複数のアージが同時に活性化される可能性がある。そのため、本モデルでは以下に示すようにアージごとの優先順位を定め[6]、意思決定相においては、優先順位の最も高いアージを選択する。

驚き > 恐れ > 怒り > 生理的 > 誇示

4. 4 行動相

行動相は、意思決定相で選択された行動プランに基づき、実際の行動を開始および終了するフェーズである。本モデルにおける行動プランは、ロボットの表情および発話によって構成される。

4. 5 事後評価相

事後評価相では、直前に収束したアージ活動の成否に関する評価が行われ、それに基づき今後のアージ活動の調整がなされる。本モデルにおいては、学習者の進捗度の変化が随時ロボットの気分（喜び・悲しみ）に反映されるよう設計されている。

5. 予備実験

5. 1 概要

6名の大学院生および大学生を対象に、提案ロボットに関する予備実験を実施した。予備実験では、今後高齢者を対象として実施する本実験と同様の状況下で体験させた。

評価にはGodspeed Questionnaire[7][8]を用いた。本調査票は、ロボットに対する人間の印象を五つの構成概念に基づき、形容詞対を用いた印象評価（1～5点）によって測定するものである。本稿では、このうち「生命性」と「好ましさ」の2指標に着目し結果を収集した。

5. 2 結果

図 4 に、提案ロボットに対する Godspeed Questionnaire の「生命性」と「好ましさ」の結果を示す。図 4 から、「生命性」が 3.75、「好ましさ」が 3.5 と、「どちらもでない」（3 点）を超えていることが分かる。

6. おわりに

本稿では、学習者と共に運転学習を行うロボットへの実装を想定し、幼児の感情表出モデルの基礎的

検討を行った。また、本モデルを搭載したロボットに対する印象を把握するため、Godspeed Questionnaire を用いた予備実験を実施した。実験結果から、提案ロボットにおける「生命性」と「好ましさ」は、Godspeed Questionnaire の中央値の点数よりも高いことが示唆された。これにより、提案ロボットは高齢者に対しても良い印象を与える可能性があると考えられる。

今後は、提案モデルを搭載したロボットと高齢者のインタラクションを通じて、ロボットに対する印象や学習継続意欲への影響を検証する。また、提案モデルを用いないロボットとの比較を行うことで、本モデルの有効性を明らかにする予定である。

参考文献

- [1] 藤掛和広, 田中貴紘, 吉原佑器, 米川隆, 稲上誠, 青木宏文, 金森等: “ドライバエージェントの運転支援及び振り返り支援による運転行動改善の効果”, 自動車技術会論文集, vol.0, no.1, pp.134-141, 2019.
- [2] ジメネスフェリックス, 加納政芳, 早瀬光浩, 吉川大弘, 田中貴紘, 金森等: “危険運転動画を用いて運転行動を教示するロボットが与える効果”, 知能と情報, vol.33, no.4, pp.819-826, 2021.
- [3] 大庭光, 松田侑子, 小河妙子: “計量テキスト分析による一般市民の世代間交流の認識に関する検討”, 老年臨床心理学研究, vol.5, pp.35-47, 2024.
- [4] 戸田正直: “感情～人を動かしている適応プログラム～”, 東京大学出版会, 1992.
- [5] P. Ekman: “表情分析入門”, 誠信書房, 1987
- [6] 加納政芳: “アージ・システムのコミュニケーションロボットへの適用”, 認知科学, vol.28, no.3, pp.390-397, 2021.
- [7] C. Bartneck et al.: “Measurement Instruments for the Anthropomorphism, Animacy, Likability, Perceived Intelligence, and Perceived Safety of Robots”, International Journal of Social Robotics, vol.1, pp.71-81, 2009.
- [8] 野村竜也: “Human-Agent Interaction(HAI) における人の主観評価”, Journal of the Japanese Society for Artificial Intelligence, vol.31, no.2, pp.224-229, 2016.

メルスペクトログラムを用いた Transformer による 音声デノイズ手法の一試み

An Approach to Speech Denoising Using Transformer with Mel Spectrogram

小島 巧実 高野 敏明

Kojima Takumi Takano Toshiaki

静岡理科大学

Shizuoka Institute of Science and Technology

Keyword: Transformer, Mel Spectrogram, HiFi-GAN

1 はじめに

音声信号には、環境雑音や通信劣化などによって品質が低下する問題がある。これに対処する音声デノイズ技術は、音声認識の品質向上などにおいて不可欠である。従来、この分野では U-Net や RNN などの畳み込み型・再帰型ネットワークが広く用いられてきた [1]。これらは、時間的・空間的特徴の抽出に優れる一方で、長期依存関係の学習が困難であるという課題があった。

これに対し、近年注目されている Transformer は、自己注意機構 (Self-Attention) により、系列中の長距離依存性を効果的に捉えることができる。これによりデノイズタスクにおいて有用である可能性がある [2]。また、並列処理に適しており、訓練効率やスケーラビリティの面でも優れている。

一方で、音声信号の時間周波数構造をどのように Transformer に適切に入力するかという点は依然として課題である。特に、STFT (短時間フーリエ変換) スペクトルは高次元であるため、計算負荷が大きい。そのため、より圧縮された表現であるメルスペクトログラムの活用が注目されている [3]。

そこで本研究では、入力特徴量としてメルスペクトログラムを用い、Transformer によって雑音除去を行う手法を検討する。特に、低い SNR 条件下において Transformer がどのような利点を持つのかを明らかにすることを目的とする。

2 提案手法

本研究では、音声デノイズタスクに対して、メルスペクトログラムを入力とする Transformer モデルを提案する (図 1)。

メルスペクトログラムとは、音声信号を時間と周波数の情報に変換した特徴量であり、人間の聴覚特性に基づいた周波数スケールで表現される。本研究では、STFT に比べて特徴量の次元が圧縮されており、Transformer による効率的な学習が可能であることから、入力特徴量としてメルスペクトログラムを採用した。

Transformer は、自己注意機構に基づくニューラルネットワークモデルである。このモデルはエンコーダ構造を

持ち、マルチヘッドセルフアテンションにより長期的な時間依存性を効果的に捉える。位置エンコーディングを付加することで、時系列情報を保持しつつ処理を行う。出力はクリーン音声のメルスペクトログラムの予測を目的とし、学習時には MSE 損失関数でクリーン音声のメルスペクトログラムとの差異を最小化する。

HiFi-GAN は、メルスペクトログラムから高音質な音声波形を生成するニューラルボコーダである [5]。推論時には、モデルの出力メルスペクトログラムを HiFi-GAN の事前学習済みモデルによって波形に変換し、実際の音声として復元する。これにより、ノイズ成分を抑制した高音質な音声の復元が期待される。

3 実験

本研究では、公開データセットである VoiceBank-DEMAND にあるクリーン音声を用いて、Transformer による音声デノイズ手法の評価を行った。トレーニングにはクリーン音声に SNR 値が -20dB から 15dB まで 5dB ごとになるようにホワイトノイズを重ねたものを使用した。VoiceBank-DEMAND の音声データには 23,075 個の音声データがあり、それに対して各 SNR 値ごとにホワイトノイズを重ねているため、学習データの総数は 184,600 個である。そのうちの 90 % をトレーニングデータとして使用した。入力としたメルスペクトログラムは、FFT サイズ 1024、ホップ長 256、メルバンド数 80 で算出し、ノイズの混入した音声の時間周波数特徴を抽出する。評価には VoiceBank-DEMAND のクリーン音声 824 個を学習データと同様にノイズを重ねたものを使用した。

表 1 に、ノイズ混合音とデノイズ後の音声に対する SI-SNR, PESQ, STOI の評価結果を示す。全 SNR を用いて学習したモデルは、PESQ および STOI の各指標において、-5dB のみを学習したモデルよりも全体的に高い値を示しており、平均的な性能は優れているといえる。ただし、SI-SNR は両モデルとも -23dB 前後の異常な値を示しており、これは波形レベルでの復元精度に課題があることを示唆している。特に、メルスペクトログラムの構造を強く学習しすぎたことによって、元の時間波形

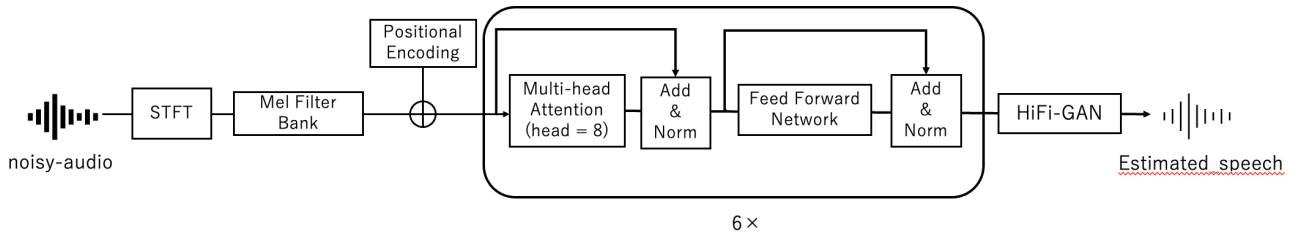


図 1: モデル構造

表 1: 入力（ノイズ混合音）とデノイズ音声の評価指標（SI-SNR, PESQ, STOI）

SNR (dB)	入力音声			全 SNR で学習			-5dB のみ学習		
	SI-SNR	PESQ	STOI	SI-SNR	PESQ	STOI	SI-SNR	PESQ	STOI
-10	-8.39	1.04	0.61	-23.53	1.33	0.70	-23.02	1.28	0.69
-5	-3.39	1.04	0.68	-23.31	1.42	0.73	-23.75	1.37	0.73
0	1.61	1.05	0.75	-23.22	1.50	0.76	-24.13	1.40	0.74
5	6.61	1.09	0.81	-23.88	1.58	0.78	-25.43	1.38	0.75
10	11.61	1.16	0.87	-24.13	1.66	0.79	-26.50	1.33	0.74
15	16.61	1.31	0.91	-24.38	1.72	0.81	-27.29	1.25	0.72

との整合性が損なわれた可能性がある。

全体として、Transformer モデルにより評価指標には小幅な変化が生じたが、定量的に明確なデノイズ効果を確認するには至らなかった。今後は、モデル構造の見直しや、時間領域の損失関数、または位相情報の復元を含む処理の導入などが、より高精度な音声復元に向けて重要であると考えられる。

4 まとめ

本研究では、Transformer アーキテクチャを用いた音声デノイズングモデルを構築し、従来の畳み込み型・再帰型ネットワークに代わる手法としての有効性を検証した。入力特徴量にはメルスペクトログラムを採用し、VoiceBank-DEMAND データセットを用いて多様な SNR 条件下で学習・評価を行った。

実験の結果、全 SNR 条件を用いて学習したモデルは PESQ の向上が見られたものの、入力音声と比較して大きな指標値の改善は見られなかった。

参考文献

- [1] S. Latif *et al.*, “Transformers in Speech Processing: A Survey,” *arXiv preprint arXiv:2211.06677*, 2022. <https://arxiv.org/abs/2211.06677>
- [2] Z. Kong *et al.*, “Speech Denoising in the Waveform Domain With Self-Attention,” ICASSP 2022

- 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp.7867-7871, 2022

- [3] N. Shao *et al.*, “CleanMel: Mel-Spectrogram Enhancement for Improving Both Speech Quality and ASR,” *arXiv preprint arXiv:2502.20040*, 2025. <https://arxiv.org/abs/2502.20040>
- [4] A. Vaswani *et al.*, “Attention Is All You Need,” *Advances in neural information processing systems* 30(NIPS2017), 2017.
- [5] K. Jungil *et al.*, “HiFi-GAN: Generative Adversarial Networks for Efficient and High Fidelity Speech Synthesis,” *Advances in neural information processing systems* 33(NIPS2020), 17022–17033, 2020.

連絡先

静岡理科大学 情報学部 高野敏明
E-mail: takano.toshiaki@ieee.org

音響パレイドリアの研究：その後の展開

小田哲久¹ 河辺義信¹ 杉浦 伸² 山田洋巳³

¹愛知工業大学 ²名城大学 ³豊橋創造大学

1. はじめに

本研究は、音響パレイドリアに関する複数の研究の続報である。既報[1],[2]では、特定のレコードの音響が、オーケストラの音響なのに、教会の鐘のように聴こえる人と、そのようには聴こえない人がいる現象の、心理学実験を通じた確認と、その個人差の分析を中心としていた。その実験の測定方法として、始めは4項目から該当する1項目を選択させる方式を採用したが、やがて、この問題用にアレンジした4項目 FCR 法を採用するようになった。アレンジされた FCR 法で得られた結果の分析は、(1) Min-Max-重心法で得られた重心値 G を元の設問にデファジィした結果を統計処理、(2) Mode-Median 法の結果を統計処理、の2種を検討した。

鐘のような音に聴こえるという話題を文献記述として残したのは、音楽評論家で指揮者の宇野功芳氏以外に無い。このことを奇異に感じて、生成 AI (ChatGPT, Gemini) を援用して文献調査を行ったが、国内、国外とも、一向に見当たらない。

そこで、幾つかの仮説を提案・考察した後、さらにもう一つの仮説として、鐘のような音に聴こえるのは、日本人(ここでは、国籍や人種概念ではなく、日本語を母語とする者を示す)だけで、英語、ドイツ語などの印欧語、あるいは中国語を母語とする者には、オーケストラの音にしか聴こえない、という可能性が浮上した。その仮説のアイデアは、角田らの研究で、日本語を母語とする者には、虫の鳴き声が左脳で処理され、その結果「虫の声」という情緒的表現がされるが、英語や中国語を母語とする者は、右脳で処理し、雑音でしかない、という研究結果からもたらされた。

本研究は、その他、既報以後に得られた様々な情報を説明する。

2. 実験課題

往年の名指揮者、カール・シューリヒトがウィーンフィルを振って英 DECCA レーベルに残した、シューベルト作曲の未完成交響曲(交響曲第7番)の録音が、宇野功芳氏によって、オーケストラの音響でありながら、鐘のような音に聴こえるという名人芸的現象が見出され、それがスコアに無いトランペットの追加によって生じている所以が LP ジャケットのライナーノートに記載されている。

その後、40年以上の年月を経て、筆者らは、その現象が全ての聴取者に同様に認識されるわけではなく、個人差が極めて大きく、鐘のように聴こえる人は装置や音量が変わっても同じように聴こえ、そうで無い人は、同じく環境が変わっても一向に鐘のようには聴こえない、ということを見出した

この現象は、音源が初発の LP レコード[3]から CD[4]に変わっても、また、最近のオンデマンド方式によるストリーミングでも、変わることなく生じている。

この現象を研究課題として取り上げる理由は、(1) 実験者にとってわかりやすい、(2) 音源の入手が容易… レコード発売以来、途切れることなく、再発売されてきた。メディアが時代とともに変わっても、常に提供されてきた。最近では SACD でも発売[5]されている。(3) 特殊な実験装置がなくても実験ができる(4) 環境変化に極めてロバストである(5) 現象の生起確率が稀ではなく、十数名という少人数グループへの実験でも、現象を認知する人とそうで無い人がほぼ確実にいること。などである。

3. これまでに行った実験

3.1 実験手続き

実験は、大学生の被験者群に教室内の拡声装置から音響刺激を呈示し、応答紙上へ応答、または、PC でオンライン応答させる。

3.2 応答方法は次の2種

- (1) 選択式の回答方式を採用した実験。具体的には、A,B,C,D,E,F 5種類の表現からどれか1つを選択させる。

A: 教会の鐘の音が聴こえた。

B: オーケストラの音だが、教会の鐘の音に似た音が聴こえた。

C: オーケストラの音だが、教会の鐘の音に聞こえなくもない音が聴こえた。

D: オーケストラの音(まとまった音)しか聞こえない。

E: 個々の楽器の音が分離して聴こえた。

- (2) 選択式で採用した5種類の表現を示し、それら全てに、該当する度合いを応答させる。

(FCR 法の適用) 具体的には、5本の連続尺度ないし10段階の尺度を用意し、その全てに応答させる。

3.3 応答紙の場合のデータの前処理

選択式は、応答結果をそのまま利用できるが、FCR 法の場合の連続尺度は、左端からの寸法をミリ単位で計測して、尺度全長で徐算して有効数字2桁の数値を算出する。

3.4 オンライン応答の場合のデータの前処理

FCR 法として Google または Moodle を採用したオンライン応答では、9段階法または11段階法の応答が得られるので、それを、最小0、最大1になる

ように線形変換して応答値とする。

3.5 FCR 法尺度値のデータ処理

応答紙，オンラインとも，5 項目の応答値 $\{a, b, c, d, e, f\}$ が得られる。E, F は，ともに，「鐘のような音には聴こえない」という意味では同じなので，まとめて X 項目とし，応答値 e, f は， $x = \min(e, f)$ として集約する。

結果の分析では次の 2 つの方法が提案されている。以下の図では， $(a, b, c, d, e) = (0.75, 1, 0.75, 0.375, 0.875)$ の場合の例を示す。

- (1) 図 1 に示すように， $x \approx 0, c \approx \frac{1}{3}, b \approx \frac{2}{3}, a \approx 1$ とファジィ数を割り付け，各ファジィ数を，幅 $2/3$ の二等辺三角ファジィ数に仮定し，代数積—加算—重心法による推論結果をデファジィして，各被験者の応答を A, B, C, X のいずれに該当するかを決定する。

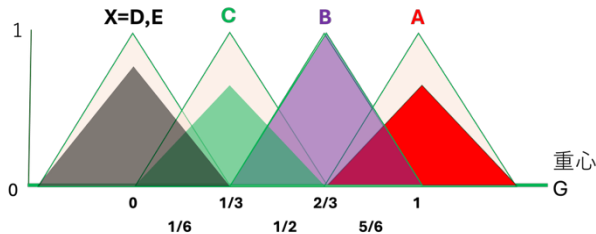


図 1： 代数積—加算—重心法 (1)
各項目への応答値を，元のファジィ数に掛け算する

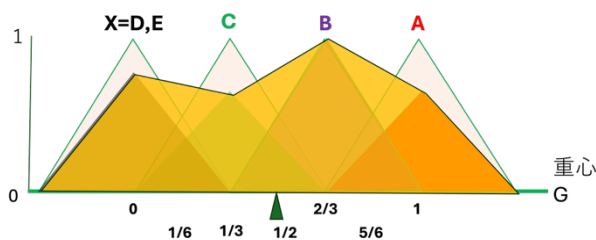


図 2： 代数積—加算—重心法 (2)
出力ファジィ集合の重心を代表値とする
今回は 0.494 となる

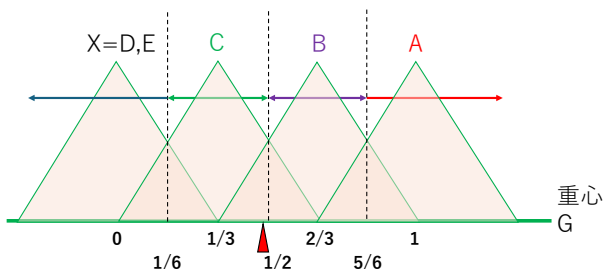


図 3： デファジィ過程の図
 $G=0.494$ は，C にデファジィされる

- (2) 応答の Mode 値を使う。ただし，測定が有限段階の場合は複数の Mode 値が得られることがあるので，その場合は，Mode-Median 法で，Mode 値の Median を用いる。Median 値が A, B, C, X に収まらない場合は，個人応答のデファジィ結果は得られず，集約して分析する。

Mode-Median 法の詳細

- (i) 1 項目のみに最大値が出現した場合
その項目を卓越ルールとする

- (ii) 2 項目が同値の最大値の場合

例えば，A と B に同じ最大値が出現した場合，(A, B) と表記する。Median が求まるか，中央付近の 1 項目に集約できる場合は → 記号で集約方法を示す。集約には， $x \leq c \leq b \leq a$ の順序関係を使う。

(X, C), (X, B) → C, (X, A) → (C, B),
(C, B), (C, A) → B, (B, A)

- (iii) 3 項目が同値の最大値の場合

(X, C, B) → C, (X, C, A) → C, (X, B, A) → B,
(C, B, A) → B

- (iv) 4 項目が同値の場合

(X, C, B, A) → (C, B)

被験者全体の集約をする場合は，2 項目の場合は各要素を $1/2$ のウェイトで集約する。3 項目なら $1/3$ のウェイトで集約。

個人単位での評価が必要な場合は形式的に被験者数を 2 倍，3 倍あるいは 6 倍に増やして対応する。

(ブートストラップ法を念頭に置く)

そうして増大させた結果の表示には，比率で表現するなどの工夫で形式的増大の影響を無くするようにする。

4. これまでに行なった実験

2023 年までの実験の概要と一部の結果を示す。
(会場にて詳しい結果を示す。)

表 1：実験の概要

NO.	時期	実験者	場所	専攻	応答	人数
1	2020.9	小田	AIT	スポマネ	択一	18
2	2020.12	山田	AIT	スポマネ	択一	84
3	2021.3	小田	SSWS	研究者	FCR	13
4	2021.5	河辺	AIT	経営情報	FCR	130
5	2021.5	河辺	AIT	経営情報	IR	130
5	2021.11	小田	AIT	スポマネ	FCR	21
6	2022.9	小田	AIT	スポマネ	FCR	20
7	2022.12	小田	AIT	経営情報	FCR	15
8	2023.7	山田	西山短大	-	FCR	21
9	2023.6	山田	AIT	経営情報 & スポマネ	FCR	21
10	2023.1	小田	AIT	スポマネ	FCR	17

表 2：第 7 実験と第 8 実験の比較

(鐘のような音に聴こえた人の比率)

環境	統計分析法	信頼区間 (95%)
正規近似法		[0.360, 0.783]
愛工大	Wilson 法	[0.365, 0.755]
	Clopper-Pearson 法	[0.340, 0.782]
正規近似法		[0.465, 0.868]
西山短大	Wilson 法	[0.454, 0.828]
	Clopper-Pearson 法	[0.430, 0.854]

5. 擬音, 聴覚パレイドリア, 錯聴

オーケストラの音なのに鐘のような音に聴こえる現象は、擬音・聴覚パレイドリア・錯聴、のいずれにあたるのか？3者の違いを示す。

(1) [擬音] Foley

定義：実音に似せて、別の物質・手段で模倣的に作られた音。リアリズムや演出のために用いられる。
特徴：明確な制作意図があり、再現性が高い、制作の主体は製作者（作曲家、または演奏家）、応用対象は、映画、演劇、放送、ゲームなど。

(2) [聴覚パレイドリア] Auditory Pareidolia

定義：本来無意味な音（環境音・雑音など）に、意味のある音（言語・音楽など）を知覚してしまう現象。
制作者は意図せず、知覚側が意味付けをする。再現性は低い（個人差が大きい）、制作の主体は聴取者（鑑賞者、被験者）、実験音楽、ノイズアート等での応用例あり

(3) [錯聴] Auditory Illusion

定義：実際の音響刺激に対する知覚のズレ。物理的には存在しない音を聴くこともあるが、多くは音の構造や位置、ピッチ、時間に関する誤認識。
特徴：制作意図は無い。再現性が高い。例はマクガーク効果（視覚情報によって聴覚が錯覚する）など。

6. 考察

上で挙げた3種の概念のうち、次に示す2つの仮説のように、擬音の可能性が皆無とは言えない。しかし、再現性が高いとは言えないので、もし擬音として意図されていたなら、不完全である。

6.1 擬音仮説1（作曲者の意図によるケース）

作曲者は、未完成交響曲の作曲と同時期に、自作の散文詩「私の夢」を書いている。最近では、その詩が、未完成交響曲のプロットになっているという説が有力視されている。その中に、母親の葬儀に遠方から駆けつける場面がある。オーストリアの郊外（リヒテンタール）での葬儀の情景描写として、ローカルな教会の鐘を音像として挿入することは自然な作法とも言える。（録音から聴き取ることでできる鐘のような音は、大聖堂のような、多数の鐘が響き

合う壮麗な音ではなく、小規模な教会の、単一の鐘による素朴な音色であり、該当リヒテンタール教会は、小教区教会で、その内容に合致する）

6.2 擬音仮説2（指揮者シューリヒトによる意図的ケース）

彼は、スコアを深読みし、ここには、鐘の音の擬音が潜んでいると考えた。ただ、そのためには、クラリネットでは音色が弱いので、敢えてトランペットに（も）クラリネットの楽譜を演奏させた。

6.3 聴覚パレイドリア仮説

聴取者が、心の中の「鐘の音」と同様の音形を聴き取ったケース。ただし、それは個人差が大きく、鐘には聴こえず、例えばトランペットが鳴ったと認識したケースも多い。これが、もっとも現実でありそうなケース

6.4 錯聴仮説

鐘のような音、あるいは、スコアを無視して楽器を追加あるいは変更したというような記述は、宇野功芳氏の書いたライナーノート[3],[4]にしか見出せない。

非常に不思議なのは、担当した録音プロデューサー、Calshowの著書[6]に、11通りもテンポを変えて録音セッションを行なっていて編集に苦労したという話はあるが、楽器の変更や鐘のような音についての言及が全く無いことである。

この事実からは、該当の音響は、レコード編集時のミキシング操作の加減で偶然マスターテープの録音に残っただけで、空気音響としてホールに鳴り響いたわけでは無い、という可能性も浮上する。その場合、現実には存在しない音響が偶然録音音源として出現したことになるが、これは錯聴とは少し異なるものとみなせる。というのは、錯聴であれば、ほぼ全員が同じように認知するはずだからである。

7. まとめ

これまでの、日本人を被験者に実験を行い、かなりの割合で鐘のような音もしくはトランペットの音が聴き取られていた。音楽は国際的な普遍性があると考えていたので、鐘のような音、（あるいはトランペット）に聴こえる現象は、世界共通と信じていた。

しかし、諸外国、特にシューリヒトと縁の深い、ドイツ・オーストリア・スイスを詳細に調査したが、音楽雑誌やレコード盤のライナーノートでの音楽評論家による演奏評などには、演奏の質や芸風などが語られるだけで、宇野氏のような記述は一切見当たらなかった。

シューリヒトは、フルトヴェングラーやカラヤンのようなスター指揮者ではないが、戦前から戦後にかけて活躍した名指揮者で、今もファンが多く、レコードが再発され続けており、その演奏に関する評論は再発売のたびに新しい評者によって更新されてきた。したがって、多数の評論家が時代を超えてその時代ごとのリマスター録音を聴いては自分の言葉

で評論を書いてきたはずである。

しかしながら、どこにも、鐘やトランペットの記述が見当たらない。これは非常に不可解な現象である。このことから、この現象は、「本格的に楽器を学んだ人は、鐘のように聴こえない」という、筆者らの一部の実験で推定された傾向が正しいという仮説を立てることができる。しかし、もう一步踏み込んで、「日本人にしか鐘のような音は聴こえない」という仮説を立てることもできる。

8. 今後の課題

今回、新たに、鐘のような音に聴こえるのは日本人特有の現象ではないか？との疑問が浮上した。これに関連して想起されるのは、角田氏の脳に関する研究結果[7], [8]である。そこでは、日本人は虫の音を左脳（言語脳）で処理し、「虫の声」と表現するが、西欧人や中国人は、虫の音を右脳（音楽脳）で処理し、声ではなく、単なる雑音に過ぎないことが示されている。この研究自体は、批判も多いが、実験によって反証されているわけではない。この研究を敷衍すれば、オーケストラの音の処理も、欧米人や中国人と、日本人とでは、脳の中の音響の処理部位が異なる、ということに由来するのかもしれない。

意外な方向に研究が進んで行きそうだが、被験者の幅を国際的に広げることが必要不可欠と思われるようになった。さらに研究を深めてゆきたい。

9. 謝辞

本研究は、音楽関係者と被験者各位の協力によって可能となりました。皆様に感謝します。

参考文献

- [1] 小田 哲久, 河辺 義信, 山田 洋巳, カール・シューリヒト指揮「未完成」交響曲録音における鐘の音の認知について, 日本心理学会大会発表論文集, 85 (0), PI-004-PI-004, 2021
- [2] 小田 哲久, 河辺 義信, 山田 洋巳, 杉浦 伸, 大山口 菜都美, ファジィ推論における出力のデファジィ法に関する新提案: 聴覚パレイドリア課題を用いた実験, ファジィシステムシンポジウム講演論文集, 38 176-181, 2022
- [3] カール・シューリヒト指揮, ウィーン・フィルハーモニー管弦楽団, キングレコード, GT9104, (LP 盤), 宇野功芳によるライナーノート付き, 1977
- [4] カール・シューリヒト指揮, ウィーン・フィルハーモニー管弦楽団, タワーレコード, PROC-1627 (24bit リマスターによる CD 盤) 宇野功芳によるライナーノート付き, 2014
- [5] カール・シューリヒト指揮, ウィーン・フィルハーモニー管弦楽団, タワーレコード, PROC-2407 (ハイブリッド SACD 盤, STEREO と MONORAL を別収録), 鈴木淳史によるライナーノート付き, 2024
- [6] ジョン・カルショー, レコードは真っ直ぐに, あるプロデューサーの回想, 学研プラス, 2005
- [7] 角田 忠信, 喜久池 良明 (Kikuchi Yoshiaki),

聴性誘発反応の左右差と角田法との比較研究, AUDIOLOGY JAPAN, 28(5), 725-738 (日本聴覚医学会誌, 1985)

[8] Tsunoda T., Sekimoto S., Itoh K., Near Infrared-spectroscopic study on processing of sounds in the brain; a comparison between native and non-native speakers of Japanese, Acta Oto-Laryngologica, 136(6), 568-574, 2016

連絡先: 小田哲久
TEL: 090-4408-4956
E-mail: oda@aitech.ac.jp

LTLBC の差分電力解析に対する耐タンパ性評価

○野崎 佑典¹, 竹本 修², 吉川 雅弥³

○Yusuke Nozaki¹, Shu Takemoto², Masaya Yoshikawa³

^{1, 2, 3} 名城大学

^{1, 2, 3} Meijo University

Abstract: リアルタイム処理を指向した低遅延軽量暗号として LTLBC が注目されている. ここで, 暗号技術は計算量的安全性が確保されている一方で, 実装時の情報を利用したサイドチャネル解析に脆弱であることが知られている. そのため, これらの解析への安全性である耐タンパ性の評価が重要である. また, この耐タンパ性評価では, 様々な多様な解析に対して網羅的な評価が必要である. そこで本研究では, LTLBC の差分電力解析に対する耐タンパ性について定量的に評価する.

1. はじめに

リアルタイム処理を指向した暗号技術として, 低遅延軽量暗号が注目されている [1]–[3]. これまでに様々な低遅延軽量暗号が提案されており, 本研究で対象とする LTLBC [1] は 2024 年に提案された最新の暗号技術である. LTLBC は, ASIC 実装での評価において, 従来の低遅延軽量暗号と比較して, 小回路規模・低レイテンシーで実装が可能であることが報告されている. ここで, これらの暗号技術は計算量的に安全性が確保されているが, 実装時の情報を利用したサイドチャネル解析 [4][6] に脆弱であることが知られている. そのため, サイドチャネル解析などの実装解析に対する安全性である耐タンパ性について評価することが重要である. これまでに LTLBC を対象に相関電力解析による安全性評価が行われている [6]. 一方で, この耐タンパ性評価では, 様々な多様な解析に対して網羅的な評価が必要である.

そこで本研究では, LTLBC の耐タンパ性の評価に必要な差分電力解析手法を提案する. そして, Field Programmable Gate Array (FPGA) を用いた評価実験を通して, 差分電力解析に対する耐タンパ性を定量的に評価する.

2. LTLBC

LTLBC [1] は, 低遅延動作を指向した共通鍵暗号であり, ブロック単位で暗号化するブロック暗号である. LTLBC の概要を図 1 に示す. ブロック長は 64bit,

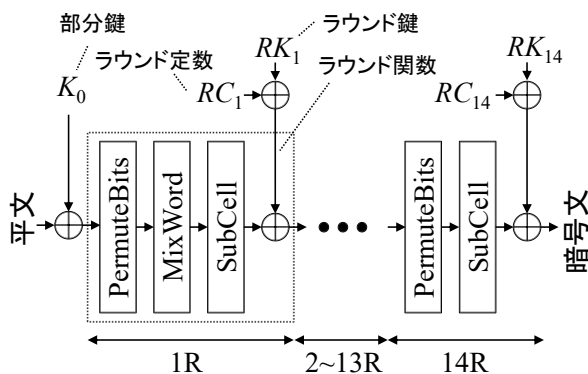


図 1 LTLBC

鍵長は 128bit である. 暗号化で繰り返し処理するラウンド関数は, PermuteBits, MixWord, SubCell, 定数値やラウンド鍵との XOR 演算で構成し, 合計で 14 回の処理を行う. PermuteBits はビット単位の転置処理を, MixWord はワード単位での XOR 演算や左巡回シフト処理などの線形処理を行う. SubCell は 16 個の 4bit 単位の S-BOX で構成しており, それぞれ非線形置換の処理を行う.

LTLBC に関連する研究としてこれまでに, FPGA を用いた実装評価 [5] が行われている. また, 相関電力解析に基づく耐タンパ性評価 [6] なども行われている.

3. 提案手法

本研究では, LTLBC の耐タンパ性評価のための差分電力解析手法を提案する. 提案手法の概要を図 2 に示す. 提案手法では, 暗号実装時にデータ（暗号中間値）を保持するレジスタ間のビット遷移と消費電力波形との線形相関を解析に利用する. レジスタ間のビット遷移には, LTLBC の暗号中間値間のハミング距離を使用する.

具体的には, 図 2 に示すように LTLBC の最終ラウンドである 14R を解析対象として, 13R 終了後と 14R 終了後のレジスタ間ハミング距離を使用する. ここで, 14R 終了後の値である暗号文 c は既知の値であるため, 13R 終了後の値 x^{13} を計算する. この計算では, 既知の暗号文 c に対してラウンド定数 RC_{14} やラウンド鍵 RK_{14} との XOR 演算, SubCell の逆関数, PermuteBits の逆関数による処理を実施することで, 13R 終了後の値 x^{13} を計算する. このとき, ラウンド鍵 RK_{14} は解析で推定したい値であるため, 候補値 k_c を利用する. これらの計算に関して, PermuteBits はビット単位, SubCell はニブル単位で計算を行っているため, 鍵の候補値はニブル単位の値を用いる. つまり, $2^4 = 16$ 通りの候補値を使用した計算を行う. そして, 計算した暗号中間値と暗号文とのハミング距離 h を導出する.

次に, 計算したハミング距離 h の値によって消費電力波形を 2 つのグループに分類する. 具体的にはある閾値 th よりも小さい場合 ($h < th$) は, グループ A へ, 閾値 th よりも大きい場合 ($h > th$) は, グループ

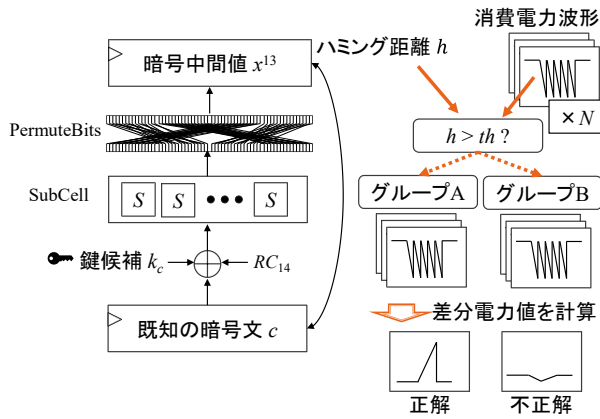


図2 提案手法の概要

グループBへと分類する。この操作を使用する消費電力波形の数 N だけ繰り返す。そして、各グループでの平均値を計算して、グループ間の平均値の差分を算出し、これを差分電力値とする。解析では、この差分電力値が最大になる鍵の候補値を正解鍵として推定する。

3. 評価実験

実験では、評価ボード SASEBO-GII 上の FPGA (Xilinx XC5VLX30) に LTLBC を実装した。対象とする LTLBC の実装アーキテクチャはループ実装であり、Verilog HDL を用いて設計した。また、実装では Xilinx ISE Design Suite 14.7 を用いて論理合成・配置配線を行った。そして、消費電力波形の測定ではオシロスコープ (Keysight DSOX 1204G) を使用し、2,000 個の波形データを取得した。

実験結果を図3に示す。図3の横軸は解析に使用した消費電力波形の数を、縦軸は解析に成功した部分鍵の数をそれぞれ示している。ここで、差分電力解析でハミング距離の分類方法（選択関数）によって耐タンパ性が変化することが知られている。そこで、本研究では4種類の選択関数を用いて評価した。図3からハミング距離の閾値を2とした場合 (HD: 01_34) が、全ての部分鍵の解析に必要な消費電力波形数が少ないことが分かる。

また、解析精度が最も高い選択関数（図3の HD: 01_34）と、先行研究 [6] の相関電力解析の結果を比較した。文献 [6] から相関電力解析では 1,400 波形を用いることで全部分鍵の解析に成功しており、提案手法は同程度の解析精度を持つことを明らかにした。以上のように、提案手法を用いることで LTLBC の耐タンパ性を多面的に評価することが可能だと考えられる。

4. まとめ

本研究では、LTLBC の耐タンパ性を評価するための差分電力解析手法を提案した。実験では、LTLBC を FPGA に実装し、実測した消費電力波形を用いた耐タンパ性評価を行った。その結果、最も解析精度の高い選択関数では、1,500 個以上の消費電力波形を

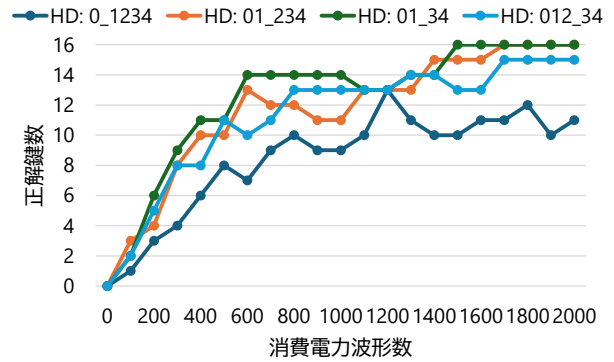


図3 提案手法による耐タンパ性の評価結果

用いることで全ての部分鍵を解析することに成功し、提案手法によって LTLBC の耐タンパ性を評価できることを明らかにした。

今後は、電磁波解析などの他のサイドチャネル解析手法に対する耐タンパ性の評価に加えて、耐タンパ実装に関する検討も進める予定である。

謝辞

本研究の一部は、JSPS 科研費 25K21202 の助成を受けたものです。

参考文献

- [1] W. Sun, L. Li, and X. Huang, "LTLBC: a low-latency lightweight block cipher for internet of things," Cluster Computing, pp. 1–12, Springer, 2024.
- [2] J. Borghoff, A. Canteaut, T. Güneysu, E. B. Kavum, M. Knežević, L. R. Knudsen, G. Leander, V. Nikov, C. Paar, C. Rechberger, P. Rombouts, S. S. Thomsen, and T. Yalçın, "PRINCE – A low-latency block cipher for pervasive computing applications," Proc. of ASIACRYPT 2012, LNCS vol. 7658, pp. 208–225, Springer, Dec. 2012.
- [3] C. Beierle, J. Jean, S. Kölbl, G. Leander, A. Moradi, T. Peyrin, Y. Sasaki, P. Sasdrich, and S. M. Sim, "The SKINNY Family of Block Ciphers and Its Low-Latency Variant MANTIS," Proc. CRYPTO 2016, LNCS vol. 9815, pp. 123–153, Springer, Aug. 2016.
- [4] P. Kocher, J. Jaffe, and B. Jun, "Differential Power Analysis," Proc. CRYPTO'99, LNCS 1666, pp. 388–397, Springer-Verlag, Dec. 1999.
- [5] 野崎佑典, 吉川雅弥, "低遅延軽量暗号 LTLBC の FPGA への実装評価," 第 56 回東海ファジィ研究会予稿集, no. S2-03, pp. 1–2, 2024.
- [6] Y. Nozaki, S. Takemoto, and M. Yoshikawa, "Power based Side-Channel Analysis for Lightweight Cipher LTLBC," Proc. 2025 17th Int. Conf. on Computer and Automation Engineering (ICCAE 2025), pp. 380–384, Mar. 2025.

連絡先

野崎佑典

E-mail: 143430019@ccalumni.meijo-u.ac.jp

知識蒸留をいた声の感情推定法の検証

森田 賢太¹

¹ 鈴鹿医療科学大学

1 はじめに

人は日常生活の中で、仕事や学業、人間関係などさまざまなストレスを経験します。このようなストレスを軽減する方法の一つとして、「誰かに話を聞いてもらうこと」が挙げられる。心理学の研究では、ネガティブな感情を言語化すること（Emotion Labeling）は、ストレスの軽減や感情の整理に効果があるとされている [1]。また、カタルシス効果（心の浄化作用）により、愚痴を話すことで心理的な負担が軽くなることが示されている [2]。一方で、現代社会では「忙しくて話を聞いてくれる人がいない」「対面で愚痴をこぼすのが苦手」といった理由から、気軽に相談できる機会が限られているのが現状である。特に、SNS やオンラインコミュニケーションの発展に伴い、対面での会話の機会が減少していることも指摘されている [3]。この問題を解決する手段の一つとして、愚痴を聴くことに特化した音声対話システムが注目されている。近年の音声認識技術や自然言語処理（NLP）の進展により、人間と AI が自然に対話できるシステムの開発が進んでいる [4]。しかし、既存の対話システムは応答が機械的で、共感的なコミュニケーションが難しいという課題がある [5]。ユーザーが安心して話せる対話システムを実現するためには、より自然で親しみのある応答が求められる。

本研究では、音声対話システムにおけるリアルタイムの感情推定を高速かつ高精度に実現することを目的とする。愚痴を聴く音声対話システムでは、ユーザーの感情を適切に把握し、それに応じた表現を行うことが重要である。感情推定に関する研究は近年盛んに行われており、大きく分けて辞書ベース手法、機械学習ベース手法、深層学習ベース手法の3つに分類される。

辞書ベース手法は、感情極性辞書を用いて単語レベルで感情を推定する手法である [6]。この手法は計算コストが低く、解釈性が高いという利点があるが、文脈を考慮できないため精度が低いという問題がある。

機械学習ベースの手法では、特徴量を抽出し、それを SVM や決定木などの分類器に入力して感情を推定する [7]。これにより、文脈をある程度考慮した推定が可能になるが、特徴量の設計が重要となる。深層学習ベース手法** 近年では、BERT や RoBERTa などの事前学習モデルを用いた感情推定が主流となっている [8]。従来の感情推定モデルは計算コストが高く、リアルタイム処理に適していないという課題がある。

そこで本研究では、従来の感情推定モデルの精度を維持しつつ、高速化を図る手法を提案する。これにより、感情推定の処理時間を短縮し、スムーズな対話体験を提供できるシステムの実現を目指す。

2 既存の感情推定モデル

従来の感情推定モデルの代表例として、rinna/japanese-roberta-base を WRIME の感情データセットでファインチューニングしたモデルを紹介する。

RoBERTa は BERT を改良したモデルであり、事前学習時のデータ拡張や動的マスキング手法を用いることで高い精度を実現している。WRIME データセットは、日本語の SNS やレビューに基づき、感情ラベルが付与されたデータセットであり、感情分析タスクに適用される。rinna/japanese-roberta-base を用いた感情推定は高精度である一方、計算コストが高く、推論速度が遅いため、リアルタイム対話システムには不向きである。この課題を解決するために、本研究では知識蒸留を用いたモデル圧縮を行う。

3 知識蒸留による感情推定手法

本研究の目的は、従来の感情推定モデルの精度を維持しつつ、推論速度を向上させ、リアルタイム対話システムへの応用を可能にすることである。本研究では、従来の感情推定モデルの精度を維持しながら、高速化を実現するために、知識蒸留（Knowledge Distillation）を活用する手法を提案する。

知識蒸留とは、大規模な事前トレーニング済みモデルの学習内容を、より小規模なモデルに転送することを目的とした機械学習アプローチである。大規模な事前トレーニング済みモデルを教師モデルと呼び、小規模なモデルを生徒モデル呼ぶ。

本手法では、WRIME データセットでファインチューニングされた高精度な rinna/japanese-roberta-base を教師モデルとし、より軽量の line-distilbert-base-japanese を生徒モデルとして学習させる。教師モデルである rinna/japanese-roberta-base は高精度な感情推定が可能である。生徒モデルである line-distilbert-base-japanese はパラメータ数が少なく、高速な推論が可能でこれにより、リアルタイム対話システムに適した高速な感情推定を可能にする。

提案手法では、以下のように知識蒸留を適用する。ロ

ジット蒸留 (Logit Distillation) は、教師モデルの出力確率分布を学習します。特徴量蒸留 (Feature-Based Distillation) は、教師モデルの中間表現を生徒モデルに転送することで、情報損失を最小限に抑えます。アテンション蒸留 (Attention Distillation) は 教師モデルのアテンションマップを生徒モデルに学習させ、類似の注意機構を持たせる。

モデルの学習手順は以下のとおりです。まず、ハイパーパラメータ、最適化手法、バッチサイズを設定して教師モデルを学習します。次に、損失関数、学習率、データ分割を行って生徒モデルを学習します。最後に、教師モデルの出力を用いて損失計算を行い、知識蒸留を行います。

4 実験

4.1 実験準備

本研究では、知識蒸留を用いた軽量な感情推定モデルの構築を目的とする。以下のモデルを比較対象として、推論速度、モデルサイズを評価する。推論速度は複数のテキストの平均推論時間であり、モデルサイズはメモリ使用量です。

- 提案手法: 知識蒸留を適用した line-distilbert-base-japanese
- WRIME データセットでファインチューニングした rinna/japanese-roberta-base
- ベースライン: 知識蒸留を適用しない line-distilbert-base-japanese

本研究では、日本語における感情推定のためのデータセットとして WRIME (Writers' and Readers' Intensities of eMotion for their Estimation) を用いる。WRIME は、日本の研究チームによって 2022 年に構築された感情推定用データセットであり、日本語テキストに対して詳細な感情ラベルが付与されている。2025 年 3 月時点で、日本語において最大規模の感情データセットである。

本データセットには、合計 35,000 件の SNS 投稿が含まれており、60 人の投稿者によるデータが収録されている。各投稿には、ロバート・プルチックの基本 8 感情 (喜び, 悲しみ, 期待, 驚き, 怒り, 恐れ, 嫌悪, 信頼) が 4 段階の強度でラベル付けされている。さらに、3 人の客観的アノテーターによる追加ラベル付けも行われており、感情推定のための質の高いアノテーションデータを提供している。

日本語の感情ラベルが付与された約 2 万件のツイートを含む WRIME データセットを使用しました。データは、トレーニングデータ、検証データ、テストデータに 8:1:1 の比率でランダムに分割されました。RoBERTa と DistilBERT のトレーニング、検証、テストには同じデータが使用されました。

実験環境は表 1 に、使用したフレームワークとライブラリは表 2 に示されています。学習中に使用したハイパーパラメータは表 3 に示されています。

表 1: マシンスペック

GPU	NVIDIA A5000(24GB)
Memory	32GB
CPU	intel(R) Xeon(R) W-2223 CPU @ 3.60GHz

表 2: フレームワークとライブラリ

Transformers	Hugging Face
Datasets	Hugging Face
Torch	PyTorch
Tokenizer & Model	Loaded from Hugging Face

表 3: ハイパーパラメータ

Batch Size	16
Epochs	12
Optimization Method	Default Setting (AdamW)
Learning Rate	Default Setting

4.2 Result

本研究では、感情推定モデルの推論速度と使用メモリを確認した。表 4 は結果を示している。

推論速度を RoBERTa を基準として比較する。提案手法 (知識蒸留した DistilBERT) の平均推論時間は 34.39ms であり、ファインチューニングした RoBERTa の 62.56ms に比べて約 45% 高速化された。また、ファインチューニングした DistilBERT とほぼ同等の推論速度を維持している。

入力データに対するメモリ使用量を測定した結果、提

表 4: 実験結果

	Knowledge-distilled DistilBERT	Fine-tuned RoBERTa	Fine-tuned DistilBERT
Inference Speed	34.39ms	62.56ms	35.13ms
Memory Usage	876.39MB	1056.27MB	876.57MB

案手法の使用量は 876.39MB であり、ファインチューニングした RoBERTa に比べて約 17%削減された。

また、ファインチューニングした DistilBERT と同程度であり、知識蒸留によるメモリ使用量の増加はほぼ見られなかった。

4.3 議論

本研究では、日本語感情推定において知識蒸留を用いた DistilBERT モデルを提案し、既存のファインチューニング手法と比較した。結果として、提案手法の推論速度は RoBERTa の約 45%の高速を達成しており、知識蒸留が推論速度に悪影響を及ぼさず、むしろ軽量化のメリットを活かせることが示された。

5 おわりに

本研究では、知識蒸留を用いた DistilBERT モデルを日本語感情推定タスクに提案し、WRIME データセットを用いて評価した。提案手法の推論速度、およびリソース使用量を従来のファインチューニング手法と比較して分析した。

実験の結果、提案手法は以下の結果を得た。推論速度の向上に関して、推論時間は RoBERTa と比較して約 45%短縮され、62.56ms から 34.39ms となった。その結果、知識蒸留によりリアルタイム処理に適した軽量の感情推定モデルが構築できることがわかった。モデルサイズとリソース使用量の削減という点では、メモリ使用量が RoBERTa に比べて約 17%削減され、低リソース環境での動作が可能であることが示された。

今後の課題として、以下の 2 つの課題に取り組む。1 つ目は、精度の評価です。知識蒸留を最適化することで、RoBERTa に匹敵する精度の実現を目指す。2 つ目は、他のデータセットでの検証で、WRIME 以外のデータセットでも評価を行い、提案手法の汎化性能を検証する。

参考文献

- [1] Lieberman, M. D., Inagaki, T. K., Tabibnia, G., & Crockett, M. J. (2011) “Affect labeling: A form of emotion regulation.” *Psychological Science* **22** (10), 1210–1218.
- [2] Pennebaker, J. W. (1997) “Writing about emotional experiences as a therapeutic process.” *Psychological Science* **8** (3), 162–166.
- [3] Turkle, S. (2011) “Alone Together: Why We Expect More from Technology and Less from Each Other.” *Basic Books*.
- [4] Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., ... & Kingsbury, B. (2012) “Deep neural networks for acoustic modeling in speech recognition.” *IEEE Signal Processing Magazine* **29** (6), 82–97.
- [5] Zhou, L., Gao, J., Li, D., & Shum, H. Y. (2020) “The Design and Implementation of XiaoIce, an Empathetic Social Chatbot.” *Computational Linguistics* **46** (1), 53–93.
- [6] Yoshihiro Adachi, Tomohiro Kondo, Takamitsu Kobayashi, Nao Etani & Kaito Ishii (2021) “Emotion Analysis of Japanese Sentences Using an Emotion-word Dictionary.” *Journal of the Visualization Society of Japan* **41** (161), 21–27.
- [7] Pang, B., Lee, L., & Vaithyanathan, S. (2002) “Thumbs up?: Sentiment classification using machine learning techniques.” In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing (EMNLP)*, 79–86.
- [8] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019) “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.” In *Proceedings of NAACL-HLT*, 4171–4186.
- [1] Lieberman, M. D., Inagaki, T. K., Tabibnia, G., & Crockett, M. J. (2011) “Affect labeling: A form