

エージェント対話を通じた高齢者体調確認サービスの実験的評価

西山 敦紀[†] 中田 匠哉[†] 陳 思楠[†] 佐伯 幸郎^{††} 増田 廣介^{†††}

露崎 雄太^{††††} 安田 清[†] 中村 匡秀[†]

[†] 神戸大学 〒657-8501 神戸市灘区六甲台町 1-1

^{††} 高知工科大学 〒782-8502 高知県香美氏土佐山田町宮ノ口 185

^{†††} 医療法人共生会 長崎友愛病院 〒851-0401 長崎県長崎市蚊焼町 2314-1

^{††††} おゆみの中央病院 〒266-0033 千葉市緑区おゆみ野南 6-49-9

E-mail: [†]2174391t@stu.kobe-u.ac.jp, ^{††}tnakata@es4.eedept.kobe-u.ac.jp, ^{†††}chensinan@gold.kobe-u.ac.jp,

^{††††}saiki.sachio@kochi-tech.ac.jp, ^{†††††}kskmsds0504m@gmail.com, ^{††††††}tsuyuzaki.yuuta@oyuminohp.or.jp,

^{†††††††}yasukiyo.12@outlook.jp, ^{††††††††}masa-n@cmds.kobe-u.ac.jp

あらまし 我々の研究グループでは、仮想エージェント（VA）を用いて在宅高齢者の「こころ」の内を聞き取り、様々な支援につなげる VA 傾聴サービスを開発している。また、これまでに、顔表情による感情認識の学習済みモデルを活用し、エージェント対話中の表情から体調を推論する機械学習分類モデルを構築するとともに、その推論結果を VA 傾聴サービスの対話に組み込むことで、高齢者の体調を聞き出し記録する体調確認サービスを提案・実装した。しかし、三段階すべての体調ラベルが収集されるまで推論期間へ移行できない課題が残っていた。本研究では、少人数の実環境試用を通じて、コールドスタート問題を緩和した体調確認サービスが実環境でどのように機能するかを探索的に評価した。

キーワード 高齢者見守り, 表情, 感情認識, 体調, 機械学習

Checking Older Adults' Health Condition Using Face-based Emotion Analysis and Spoken-Dialogue Agent

Atsunori NISHIYAMA[†], Takuya NAKATA[†], Sinan CHEN[†], Sachio SAIKI^{††}, Kosuke MASUDA^{†††},

Yuta TSUYUZAKI^{††††}, Kiyoshi YASUDA[†], and Masahide NAKAMURA[†]

[†] Kobe University Rokkodai-cho 1-1, Nada-ku, Kobe, Hyogo 657-8501 Japan

^{††} Kochi University of Technology 185 Miyanokuchi, Tosayamada, Kami City, Kochi 782-8502, JAPAN

^{†††} Medical Corporation Nagasaki Yuai Hospital 2314-1 Kayaki-machi, Nagasaki City, Nagasaki 851-0401 Japan

^{††††} Oyumino Central Hospital 6-49-9 Oyumino Minami, Midori-ku, Chiba-shi, Chiba 266-0033 Japan

E-mail: [†]2174391t@stu.kobe-u.ac.jp, ^{††}tnakata@es4.eedept.kobe-u.ac.jp, ^{†††}chensinan@gold.kobe-u.ac.jp,

^{††††}saiki.sachio@kochi-tech.ac.jp, ^{†††††}kskmsds0504m@gmail.com, ^{††††††}tsuyuzaki.yuuta@oyuminohp.or.jp,

^{†††††††}yasukiyo.12@outlook.jp, ^{††††††††}masa-n@cmds.kobe-u.ac.jp

Abstract Our research group is developing a VA listening service that uses a virtual agent (VA) to listen to the inner thoughts of elderly people living at home and connect them with various forms of support. Furthermore, we have constructed a machine learning classification model that infers health condition from facial expressions during agent interactions by utilizing a pre-trained model for facial emotion recognition. By incorporating these inference results into the VA listening service's dialogue, we have proposed and implemented a health check service that elicits and records the health condition of elderly individuals. However, there remained the challenge that the system could not transition to the inference phase until health condition labels for all three stages had been collected. In this study, we conducted an exploratory evaluation of how a health check service designed to mitigate the cold-start problem functions in real-world environments through small-scale field trials.

Key words watching over the elderly, facial expressions, facial emotion recognition, health condition, machine learning

1. はじめに

近年日本の高齢者人口は増加しており、独居高齢者も増加している [1]。高齢者は孤立死を身近に感じており、実際に孤立死と考えられる事例が多数発生している [2]。そのため高齢者が孤立せず気付かれずに体調が悪化しないためにも、日々高齢者への声掛けを行い体調を把握することが大切である。先行研究として仮想エージェントを用いて、在宅高齢者の日常生活の中での傾聴を行う VA 傾聴サービスを開発しており、感情表出の場として機能している [3]。また感情分析による体調推論とエージェント対話を通じた高齢者体調確認サービスの提案と実装を行い、エージェントとの対話中に表情感情データを収集できることを確認した。しかし、三段階の体調を収集できないと推論期間に移行できないことが課題として残った。そこで本研究は先行研究で残されたコールドスタート問題を緩和した高齢者体調確認サービスを日常生活環境で試用し、推論期間における体調推論、VA による確認・深掘り対話、および短期的な受容性を探索的に評価することを目的とした。なお本研究は、表情感情値のみを用いた高齢者の体調推論の精度を統計的に検証するものではなく、限られた利用事例に基づいて、推論体調と申告体調の一致・不一致や、VA による確認・深掘り対話の働き、短期利用における課題を探索的に明らかにするものである。そして本研究では以下のリサーチクエスチョンを設定した。

RQ1：実環境において、表情感情値に基づく体調推論結果は、利用者の主観的な体調申告とどのような一致・不一致を示すか

RQ2：表情の推論結果を交えた VA の深掘りは、高齢者の好不調に関する理由、自己開示を促すきっかけとなるか

RQ3：提案システムは短期間の日常利用において、利用者にどのように受容され、どのような利用上の課題が見られるか
提案システムを日常生活環境で使用した際に、表情感情値に基づく体調推論、VA による体調確認、および深掘り対話がどのように機能するかを探索的に評価し、プロトタイプシステムとしての課題を明らかにすることを目的として実験を行った。その結果、表情感情値のみを用いた体調推論では、「普通」の体調ラベルが大部分を占める不均衡なデータ環境において、「良い」や「悪い」といった少数クラスを安定して識別することが困難であることが示唆された。一方で、推論結果を VA が提示し、それをきっかけとして体調の理由を尋ねる深掘り対話は、被験者が自身の疲れや不調を言語化する契機となる可能性が示された。さらに、毎日の利用においては、体調確認のやり取りが単調になりやすく、煩わしさや飽きにつながる可能性があることも明らかになった。

2. 準備

2.1 超高齢社会

近年、日本の高齢者人口は増加しており令和 6 年度版高齢社会白書 [1] によると、総人口が減少する中で 65 歳以上の者が増加することにより高齢化率は上昇を続け、令和 19 年に 33.3 % となり国民の 3 人に 1 人が 65 歳以上の者となると見込まれている。また 65 歳以上の一人暮らしの者についても 65 歳以上の男

女それぞれの人口に占める割合が、令和 2 年には男性 15.0 %、女性 22.1 % となり令和 32 年には男性 26.1 %、女性 29.3 % となると見込まれており独居高齢者も増加している。

独居高齢者については人と話をする頻度について「毎日」と回答した人の割合がひとり暮らし以外の人の半分以下という調査もあり、独居高齢者は高齢者の中でも特に孤立しているという状況が窺える [1]。65 歳以上の者の孤立死に対する意識調査においては、孤立死を身近に感じるかという問いに対して「とても感じる」又は「まあ感じる」と回答したものは 48.7 % である [1] とあり、高齢者自身も孤立死を身近に感じている。実際に、令和 6 年度の上半期に自宅において死亡した一人暮らしの者のうち 65 歳以上は 28330 人に上り [2]、孤立死と考えられる事例が多数発生している。

2.2 注目する問題

高齢者が孤立せず、気付かれずに体調が悪化しないためにも、日々高齢者への声掛けを行い体調を把握することが大切である。東京都福祉保健局の高齢者の見守りガイドブック [4] では日常生活を送る中で、住民同士が互いにさりげなく気遣い合い、見守り合う関係を緩やかな見守りと呼び、コミュニケーションを交えながら住民同士が普段との違い、異変を感じてケアをしたり、専門機関に相談するなどのさりげない見守りが、今まで以上に重要性が高まっているとある。地域住民やそれ以上に近い家族と日々のコミュニケーションがあれば、普段との違いを気にかけてくれるが、孤立した状況では行われたいため、定期的なコミュニケーションの機会を設けることや、また見守りサービスの活用で状況を把握する必要がある。

また、高齢者の社会的孤立は課題であるが、単に対話の機会を設けるだけでは不十分である [5]。特に、健康や経済的な問題などの深刻な困難に直面していても、それらを他者に打ち明けて支援を求めることは容易ではない。このような自己開示の困難さは、適切な社会的支援を受ける機会を逸することに繋がり、結果として孤立をさらに深める要因となる。これに対し、野口らは、人間同士の対話をロボットが仲介する「社会的媒介ロボット」により、喪失経験のような通常は話しにくい話題において自己開示が促進されうること示した [5]。ロボットは人間と社会的利害を共有しない相手として捉えられやすく、対面のコミュニケーションに比べて評価懸念や心理的抵抗が緩和されう。

2.3 先行研究

2.3.1 VA 傾聴サービス

我々の研究グループでは、在宅で生活する高齢者の日常における傾聴を目的として、仮想エージェントを活用した「VA 傾聴サービス」を開発している [6] [7] [8]。本サービスは、高齢者が日常生活で生じる不安や悩みを仮想エージェントに話すことを通じて、心理面のケアを行うことを目的とする。実証実験の結果、本サービスは感情表出を促す場として機能し、孤独感および感情状態の変化に対して有効性が示唆されている [3]。

しかし VA 傾聴サービスは対話内容のテキストを活用することが主であり、対話中の表情などの非言語的な情報が十分に活用されていない。対話中に表情を取得して分析し、VA 傾聴サー

ビスの発話に組み込むことで見守りに生かすことができる可能性がある。

2.3.2 高齢者体調確認サービスの開発

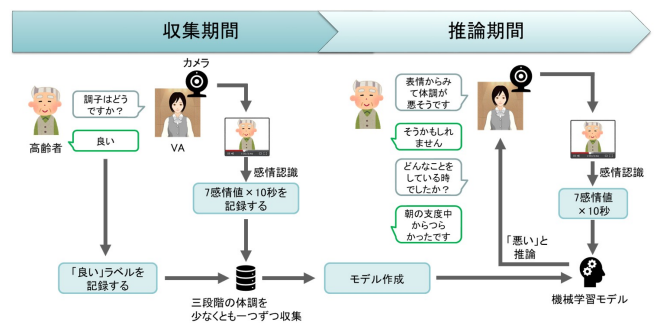
先行研究として感情分析による体調推論とエージェント対話を通した高齢者体調確認サービスの提案と実装を行った[9]。本サービスは、対話中に取得した表情感情値から利用者の体調を推論し、その推論結果に基づいてVAが体調に関する深掘りの質問を行うものである。対話中の表情感情値を利用者の体調を推論するための手がかりとして用いており、これは体調不良や疲労が表情に一定程度反映される可能性を前提としたものであり、先行研究では演技的に表出された体調差を対象として検討してきた[10]。

高齢者体調確認サービスの流れを図1に示す。高齢者体調確認サービスでは表情感情と体調のデータを収集する収集期間と、データから機械学習分類モデルを作成して以降に推論と深掘りを行う推論期間に分かれる。まず収集期間では、毎日最初のエージェントとの会話中に10秒の動画クリップが撮影される。その各フレームから既存の感情認識の学習済みモデルを用いて七つの感情クラス(angry, disgust, fear, happy, sad, surprise, neutral)の感情値を抽出する。抽出後にエージェントが高齢者の体調を良い、普通、悪いの三段階から聞きとる。本サービスで扱う「良い」「普通」「悪い」の体調ラベルは、医学的な診断結果ではなく、利用者がその時点で主観的に申告した体調である。そのため、「悪い」という回答には、身体的な不調だけでなく、疲労感、眠気、心理的な疲れ、ストレスなどが含まれる可能性があり、これらを厳密に区別するのではなく、利用者本人が日常的に感じる主観的な好不調として扱う。次に表情感情に主成分分析を適用し、第1、第2主成分を特徴量に、高齢者の体調を正解データとして決定木分類器により体調分類モデルを作成する。主成分分析は、7種類の表情感情値を低次元化し、少数データに対して特徴量数が過剰になることを避けるために用いた。また、決定木分類器は、少数の学習データからでも分類モデルを構築でき、分類規則を木構造として確認できるため、プロトタイプ段階のシステムにおいて推論結果を解釈しやすいという利点がある。そして推論期間では、収集期間と同様に会話中の動画クリップ撮影から体調の聞き取りまでを行った後に表情感情からモデルで体調を推論する。エージェントは推論された体調について「表情から見ると～」といった確認を取り、それに対する返答をLLMに入力して対話で採用する深掘り用の体調を決定し、体調について深掘りを行う。

またこのサービスの実環境における学習データの蓄積特性を把握するために予備的試用を行い、エージェントとの対話中に体に負担をかけずに自然な形で表情感情のデータが取得できることが確認された。

2.3.3 先行研究の課題

先行研究では、VAとの対話中に表情感情データを収集し、そのデータを用いて体調推論モデルを構築する高齢者体調確認サービスを提案・実装した。また、実環境において利用者に大きな負担をかけずに表情感情データを取得できることを確認した。一方で、先行研究のシステムでは、「良い」「普通」「悪い」の



三段階すべての体調ラベルが収集された後に体調推論モデルを構築し、推論期間へ移行する設計であった。しかし、日常生活環境では利用者の体調が大きく変化しない日が多く、特に「悪い」などの体調ラベルが長期間収集されない可能性がある。その場合、推論モデルを構築できず、推論結果を用いた体調確認や深掘り対話を開始できない。

したがって、三段階すべての体調ラベルが揃うまで推論期間へ移行できないというコールドスタート問題を緩和し、限られた体調ラベルからでも推論期間の運用を開始できる仕組みが必要である。

3. 提案手法

3.1 目的とリサーチクエスト

本研究の目的は、先行研究で残されたコールドスタート問題を緩和した高齢者体調確認サービスを日常生活環境で試用し、推論期間における体調推論、VAによる確認・深掘り対話、および短期的な受容性を探索的に評価することである。なお本研究は、表情感情値のみを用いた高齢者の体調推論の精度を統計的に検証するものではなく、限られた利用事例に基づいて、推論体調と申告体調の一致・不一致や、VAによる確認・深掘り対話の働き、短期利用における課題を探索的に明らかにするものである。提案システムの有用性を探索的に検討するため、本研究では以下のリサーチクエストを設定する。

RQ1：実環境において、表情感情値に基づく体調推論結果は、利用者の主観的な体調申告とどのような一致・不一致を示すか

RQ2：表情の推論結果を交えたVAの深掘りは、高齢者の好不調に関する理由、自己開示を促すきっかけとなるか

RQ3：提案システムは短期間の日常利用において、利用者にどのように受容され、どのような利用上の課題が見られるか

3.2 コールドスタート問題への対応

先行研究のシステムでは、「良い」「普通」「悪い」の三段階すべての体調ラベルが収集された後に体調推論モデルを構築し、推論期間へ移行する設計であった。しかし、日常生活環境では特定の体調ラベル、特に「悪い」が長期間収集されない可能性があるため、推論期間へ移行できない場合がある。そこで本研究では三段階すべての体調ラベルが揃う前であっても、推論期間へ移行できるようにシステムを変更した。具体的には、「良い」「普通」「悪い」のうち、いずれか二つの体調ラベルに対応する動画クリップが各1件以上収集された時点で、その二つの

ラベルを分類対象とする暫定的な2値分類モデルを構築する。そして、モデル構築後は推論期間として運用し、表情感情に基づく体調推論及びVAによる深掘り対話を実行する。

推論期間中に、それまで収集されていなかった体調ラベルが新たに1件以上得られた場合には、収集済みデータを用いてモデルを再構築し、3値分類モデルへ更新する。これにより、特定の体調ラベルが長期間出現しない場合でも推論期間の運用を開始でき、提案システムの推論機能及び深掘り対話を実環境で評価することが可能となる。

4. 実験

4.1 実験概要

RQ1からRQ3までを検証するために実験を行った。本実験の目的は、提案システムを日常生活環境で使用した際に、表情感情値に基づく体調推論、VAによる体調確認、および深掘り対話がどのように機能するかを探索的に評価し、プロトタイプシステムとしての課題を明らかにすることである。開発者自身(20代男性)および高齢者1名(70代男性、以下「被験者A」とする)を対象に、6月4日から6月18日まで各々の居住環境において実施した。開発者自身はPCとウェブカメラ、被験者Aはノートパソコンを用いて行った。

ここで、開発者自身はシステムの仕様を理解しているため、発話内容や回答が一般利用者と異なる可能性があり、開発者自身による試用はシステムの動作確認や対話上の問題点を抽出するための補助的な事例として扱う。また、被験者Aの背景情報として、PC利用に大きな支障はなく、本実験の手順を理解した上で参加したことが挙げられる。一方で、日常生活における物忘れへの不安を有していた。

本システムでは、VAとの対話中に10秒間の動画クリップを取得し、表情感情値を抽出した後、分析には表情感情値、体調ラベル、対話ログを用い、取得した動画クリップは表情感情値の抽出にのみ用い、分析対象として保存しない運用とした。

4.2 実験手順

手順として、毎日最初に行われるVAとの会話において表情感情の抽出と体調推論を行い、体調の確認と深掘りを実行した。具体的には、対話エージェントとの挨拶、雑談中に10秒間の動画クリップから感情認識値を抽出した。その後VAが被験者に現在の体調を尋ね、被験者がその時点で主観的に判断した体調を「良い」「普通」「悪い」の3段階で申告した。そして被験者ごとに構築した体調分類モデルで10秒間の動画クリップから得られた表情感情値に基づき体調を推論した。VAは、機械学習モデルによる推論体調を「表情から見ると～」という形で利用者に提示し、その推論結果が本人の認識と一致しているかを確認した。LLMには、VAによる確認発話、機械学習モデルによる推論体調、およびそれに対する利用者の返答を入力し、利用者の返答が「良い」「普通」「悪い」のいずれの体調状態を示しているかを分類させた。本研究では、この分類結果を深掘り用体調ラベルと呼び、医学的な体調判定ではなく、後続の追加質問を選択するための対話制御用ラベルとして用いた。

ここでそれぞれの体調分類モデルについては先行研究の収集

期間で得られたデータを用いた。開発者自身は普通ラベルのみが収集されていたため収集期間から実験を開始し、被験者Aは良いと普通のラベルが収集されていたため2値モデルで推論期間から実験を開始した。なお、被験者Aについては生活習慣上、毎朝最初に使用することが困難であったため、任意のタイミングで特定の起動キーワードを発話することでシステムを起動する運用とした。このため、本実験では、朝の起床直後の体調確認と、日中または夕方の任意時刻における体調確認が混在している。

4.3 評価方法

本実験は少人数を対象とした探索的評価であるため、推論精度を統計的に評価するのではなく、体調ラベル、対話ログ、表情感情値の推移、およびアンケート回答に基づいて事例分析を行った。RQ1については利用者が主観的に申告した体調と、表情感情値に基づいて機械学習モデルが推論した体調を比較し、一致および不一致の傾向を確認し、不一致が生じた事例では対話ログおよび表情感情値を参照して要因を検討した。RQ2は深掘り対話ログにおいて、単なる体調ラベルの回答にとどまらず、体調の理由、疲労や心理的負担、普段は他者に言いにくい可能性のある内容が表出したかを確認した。RQ3は被験者Aへのアンケート回答をもとに、短期間の日常利用における負担感、継続利用に関する意見、プライバシー面の懸念を整理した。なお、本実験は少人数かつ短期間の探索的評価であるため、各RQに対する結果は、提案システムの有効性を一般化して示すものではなく、実環境利用における事例的な示唆として扱う。

4.4 実験結果

実験期間中で提案システムによる対話が行われた際に得られた体調ラベルを表1と表2に示す。各表の体調はそれぞれ、申告体調は10秒の動画クリップ撮影後に「今の調子はどうですか?」とVAが尋ねた際の被験者の回答、推論体調は機械学習分類モデルによって10秒の動画クリップから推論された体調、深掘り用体調は推論された体調を高齢者に「表情から見ると…」と尋ねた際に答えた回答を基にLLMが判断した体調となっている。被験者Aは6月13日までは良いと普通の2値分類、6月14日からは3値分類、開発者自身は6月12日以降は普通と悪いの2値分類モデルによる推論が行われた。6月13日は申告でそれまでの2値以外、「悪い」ラベルが得られ、推論を行わなかったため推論、深掘り用体調ラベル無しとなっている。

被験者Aでは申告体調の大部分が「普通」であり、「悪い」は1回のみであり、推論では良いとされた日がある一方で悪いとされた日はなかった。開発者自身では、悪いと申告した日があったが推論と深掘り用体調が普通となった。

表1と表2から申告あるいは推論が普通以外の会話内容について、図2では問いかけにより会話以前の起床直後の不調に関する発話を得られた。この「起きた時直後は気分がちょっと良くなかったです」という内容が体調が悪いという風に捉えられ、表1の6月4日の深掘り用体調が悪いという判断につながった。図3では悪いと申告した日に様々な予定をによる疲れという理由が引き出された。図4では表情から「良い」と推論されても、被験者は「普通」と修正しており、確認質問が誤推論の

表 1 被験者 A の体調ラベル

日時	申告体調	推論体調	深掘り用体調
2026-06-04 09:32:32	普通	良い	悪い
2026-06-05 07:58:32	普通	普通	普通
2026-06-06 06:50:24	普通	普通	普通
2026-06-07 08:23:56	普通	普通	普通
2026-06-13 15:00:10	悪い	-	-
2026-06-14 10:10:59	普通	普通	普通
2026-06-15 10:06:34	普通	普通	普通
2026-06-16 18:20:08	普通	普通	普通
2026-06-17 09:54:02	普通	良い	普通
2026-06-18 10:10:36	普通	普通	普通

表 2 開発者自身の体調ラベル

日時	申告体調	推論体調	深掘り用体調
2026-06-11 07:12:01	悪い	-	-
2026-06-12 18:20:55	普通	普通	普通
2026-06-13 18:07:42	普通	普通	普通
2026-06-15 08:52:37	普通	普通	普通
2026-06-16 09:05:50	悪い	普通	普通
2026-06-17 09:23:13	普通	普通	普通
2026-06-18 07:09:41	普通	普通	普通

すみません。もう一度お願いします。(良い／普通／悪い) (09:33:41)	(09:33:49) 普通
きよしさん、体調は普通とのことです。(09:33:52)	
お顔を拝見すると良いようにも見えて、少し差があるようです。(09:33:57)	
今の気分はいかがですか？ (09:34:03)	
(09:34:15) はいいいですが 朝はちょっと気分 起きた時 直後は気分がちょっと良くなかったです	
起きた直後に気分がすぐれなかったのですね。(09:34:28)	
無理に思い出さなくて大丈夫ですが、そのとき体のだるさや痛みなどは少しありましたか？ (09:34:33)	
(09:34:54) 今日に限らず 最近起きた直後は右のお尻から太ももにかけてがちょっと痛いです これは ず性	
起きた直後に右のお尻から太ももが痛むのですね。(09:35:06)	
続くようでしたら、医師にご相談ください。(09:35:12)	
今朝のお散歩は、無理なく行けそうですか？ (09:35:17)	
(09:35:24) 終わり	
またお話ししてくださいね (09:35:25)	

図 2 被験者 A の 6 月 4 日の深掘り対話

今の調子はどうですか？(良い／普通／悪い) (15:00:51)	(15:00:57) 悪いです
おつらいですね。(15:01:01)	
もし話せそうでしたら、今いちばん気になる痛みやだるさはどんな感じか教えてください。(15:01:04)	
(15:01:31) 家 だ 事 務 所 に 今 日 は 10 時 から 物 売 れ の 会 が あ っ て お 年 寄 り を 集 め て 喋 っ て い ま し た そ の 後 ち	
よ っ と 深 刻 な 人 の 相 談 に 乗 っ て そ の 後 パ ソ コ ン の 相 談 室 も や っ て ち ょ っ と 疲 れ ま し た	
今日はたくさん人のために動かれたのですね。(15:01:39)	
相談に乗るのは気持ちも使いますし、疲れて当然だと思います。(15:01:44)	
今は少し涼しい場所で、ゆっくり座れそうですか？ (15:01:51)	
(15:02:04) はいこれから窓を開けて ちょっと部屋の空を涼しくしようと思います	
それはいいですね。(15:02:05)	
心地よい風が入るといいですね。(15:02:09)	
(15:02:16) 終わり	
またお話ししてくださいね (15:02:16)	

図 3 被験者 A の 6 月 13 日の深掘り対話

補正として働いた。図 5 では「ちょっと悪いかもしれない」を「変わるかもしれない」という音声認識誤りにより、不調表現が拾えず、深掘りにつながらなかった。

今の調子はどうですか？(良い／普通／悪い) (09:54:35)	(09:54:41) 普通です
きよしさん、昨日はいつも通りでしたね。(09:54:50)	
今日は体調を普通とおっしゃり、表情の推定では良いようにも見えます。(09:54:56)	
今の気分はいかがですか？ (09:55:02)	
(09:55:14) 表情がちょっと違うのは外が明るいからではありませんか	
そうかもしれませんね。(09:55:23)	
外の明るさで、お顔の見え方が変わることはありそうです。(09:55:27)	
暑くなる前に、少し涼しく過ごせるといいですね。(09:55:34)	
今朝は何か飲み物を用意されていますか？ (09:55:39)	
(09:55:46) 終わり	
またお話ししてくださいね (09:55:46)	

図 4 被験者 A の 6 月 17 日の深掘り対話

今の調子はどうですか？(良い／普通／悪い) (09:06:13)	(09:06:18) 悪い
お顔の推定は普通のようなので、少し良い方向に差がありそうです。(09:06:45)	
あつりのさん、昨日はいつも通り、今日は悪いとおっしゃっていましたね。(09:06:45)	
今の具合はいかがですか？ (09:06:45)	
(09:06:53) 変わるかもしれない	
そうですね、気分や体の感じは少しずつ変わりますね。(09:07:02)	
無理に良くしようとしなくても大丈夫ですよ。(09:07:02)	
今はゆっくり過ごせそうですか？ (09:07:02)	
(09:07:19) はい	
わかりました。(09:07:20)	

図 5 開発者自身の 6 月 16 日の深掘り対話

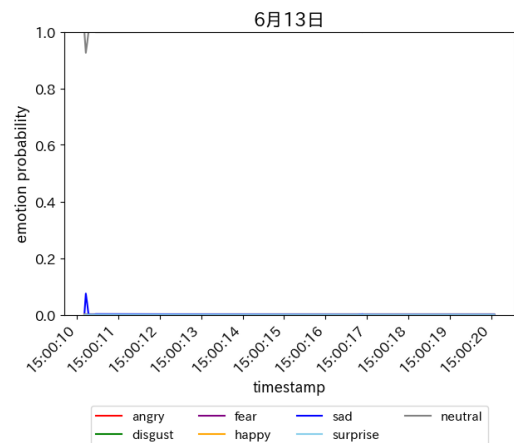


図 6 被験者 A の 6 月 13 日の感情値

被験者 A が悪いと申告した 6 月 13 日の感情値のグラフを図 6 に示す。この時は 10 秒のほとんどで neutral が支配的であることが分かった。本研究で用いた被験者 A の主成分分析では、第 1 主成分の寄与率は 77.3%，第 2 主成分までの累積寄与率は 90.3% であった。主成分負荷量は、第 1 主成分では neutral が 0.281, disgust が -0.166 と相対的に大きく、sad が -0.069 であった。第 2 主成分では disgust が 0.096, sad が -0.092 であった。

また、被験者 A にアンケートを実施し、得られた意見や改善点を以下にまとめる。

質問 1：6 月 17 日水曜日 9 時 53 分 『表情では良いようにも見えます』とシステムに言われた日がありましたが、ご自身の実際の体調と比べてどのくらい当たって

いる、またはズレていると感じましたか？

質問2：自分では『いつも通り（普通）』だと思っていたのに、システムから体調を指摘されて、自分の体調を意識したりした瞬間はありましたか？

質問3：6月13日土曜日 14時59分 「悪い」と答えた際にメイちゃんから体調を伺われて、普段より話しやすくなる気持ちの変化はありましたか？

質問4：家族や知り合いには『心配をかけたくない』と思って言わないような小さな体調の不安でも、このシステム相手なら話しやすいと感じる部分はありましたか？

質問5：毎朝、最初の会話で10秒間カメラに表情を見られ、体調を聞かれるというやり取りは、10日間続けてみて負担や煩わしさを感じましたか？『自分の表情が毎日分析されている』というプライバシーの面も含めて教えてください。

質問6：毎日、ほとんど『普通』と答えるだけの同じようなやり取りが続きましたが、途中で飽きたり、面倒くさくなったりはしませんでしたか？今後も数ヶ月、数年と毎日の生活習慣として無理なく続けられそう（持続できそう）か、負担感も含めて教えてください。

質問7：その他、毎日使ってみて気付いたこと、ご意見やご感想（もっとこうしてほしい等）がありましたら、自由にお願いたします。

質問1：実験期間中、体調にあまり変化がなかったため、体調が良いように見える、と言われても、実感がわかりません。

質問2：ありませんでした。

質問3：話しやすくなる、というよりも、自分の体調（心理的な疲れ）を素直に話せた気がします。

質問4：それはあると思います。

質問5：ちょっと煩わしい気がします。

質問6：個人的には、あまり体調に差がないので、ややあきます。改善点として、多くの人にとっては、朝方に体調、夕方や夜に疲れやストレスを聞いたほうが良い気がします。

質問7：以前から言ってることですが、これは介護者に使ってあげると、ストレスや介護疲れの表出に良いと思います。

4.5 考 察

以下にそれぞれのリサーチクエストに対する考察を示す。

RQ1: 表情推論と申告体調の一致・不一致

表1および表2が示す通り、両被験者ともに申告体調・推論体調のどちらにおいても「普通」が大部分を占めていた。このことから、本実験のように平常時のデータが多く、不調や好調のデータが少ない条件では、表情感情値のみから「良い」や「悪

い」といった少数の体調状態を安定して識別することは難しいことが示唆された。この要因として、日常環境における平時と不調時の表情の差異が極めて微細であるというデータ側の特性が挙げられる。主成分分析では、第1主成分の寄与率が77.3%と高く、第1主成分の負荷量ではneutralが相対的に大きかった。実験室環境における演技データとは異なり、図6では不調が申告された事例であってもneutralが支配的であり、実環境では体調不良、今回の場合は疲れが必ずしも顕著なsadやangryの確率といった感情表出へと直結しないことが確認された。

一方で、「普通」と申告された多くの日において推論体調も「普通」となる傾向が見られたことから、平常時の状態を継続的に記録する仕組みとしては一定の可能性があると考えられる。また、推論結果が本人の認識と異なる場合でも、VAが「表情から見ると～」という形で確認を行うことで、本人の回答に基づいて深掘り用体調を補正できることが確認された。したがって、本システムにおける表情推論は、体調を単独で断定するものではなく、体調確認や深掘り対話を開始するための補助的な手がかりとして用いることが適切であると考えられる。

また主成分分析を適用して次元削減を行い、少数のデータセットで決定木を構築するというアプローチの限界が浮き彫りとなった。「普通」クラスに極端に偏る不均衡データ環境においては、稀にしか発生しない「良い」「悪い」のデータが多数派のデータに埋没する。決定木の学習アルゴリズムにおいて不均衡データを考慮するclass.weight='balanced'などを設定しても、実データが偏っている場合は短期的に適切に識別することが難しいと考えられる。日々の対話を通じて逐次的に学習データを蓄積し、クラス比率の偏りを考慮しながらモデルを更新する方法を検討する必要がある。

RQ2: VA 深掘りによる体調理由の言語化

対話ログの分析より、システム側の体調推論の正誤に関わらず、非言語情報をきっかけにした深掘り対話そのものが自己開示に機能する可能性が示された。具体的には、図3のように2値モデル期に明示的な「悪い」の申告があった事例だけでなく、図2に示す間違いが発生した事例においても、VAから「表情から見ると～」と一歩踏み込んだ指摘を行うことで、高齢者がそれ以前の不調の理由を自発的に述べる様子が確認された。事後アンケートの質問3における、素直に話せた気がするという回答からも、VAによる深掘り対話は、小さな不調や疲れを言語化するきっかけとして機能する可能性がある。

一方で、対話制御に関する課題も明らかになった。図5の事例では、開発者自身の発話「ちょっと悪いかもしれない」という曖昧な表現が、音声認識の段階で「変わるかもしれない」と誤認識され、結果として深掘り対話につながらなかった。本システムは表情から体調を推論した後に「表情から見ると～」と確認を取ることで誤推論のまま無駄な深掘りが行われて会話の負担となることを防ぐよう実装している。しかしこの設計は音声認識誤りの影響を受けやすく、一つの認識誤りによって後続の深掘り対話が実行されない可能性があり、システム上の課題であることが確認された。

また本システムでは、VAの確認発話に対する利用者の返答

を LLM が分類し、深掘り用体調ラベルを決定している。そのため、利用者の曖昧な返答や音声認識誤りを含む入力に対して、意図と異なるラベルが付与される可能性がある。特に高齢者見守りの文脈では、誤ったラベルに基づく深掘りが、過剰な心配や不調の見逃しにつながる可能性がある。したがって今後は、音声認識の不確実性を考慮し、重要な体調確認については選択肢による回答や再確認を組み合わせた方法を検討する必要がある。

RQ3: 短期利用における受容性と課題

質問 5 および質問 6 の回答から、本システムは短期間の日常利用として試用可能であった一方で、利用者に煩わしさや飽きが生じる可能性があることが示された。被験者からは本実験期間において体調に大幅な変動がないために、毎日同じパターンのやり取りが繰り返されることに対する飽きや煩わしさが指摘された。体調に大きな変化がない日が続く場合、「普通」と答えるだけの対話になりやすく、利用者が体調確認の意義を感じにくくなる可能性がある。したがって、継続利用のためには、毎回同じ形式で体調を申告させるのではなく、体調申告を求める頻度やタイミング、質問形式を利用者の生活リズムや過去の回答傾向に応じて調整する仕組みが必要である。

また、自由記述では、本システムを介護者に用いることで、ストレスや介護疲れの表出を支援できる可能性があるとの意見も得られた。介護者は疲労感や心理的負担を他者に伝えにくい場合があるため、今後は、高齢者本人の体調確認に加えて、介護者が自身の疲労感や心理的負担を話すきっかけを提供する仕組みとしての応用も検討する。

4.6 本研究の限界

本実験の限界として、対象者数と利用期間が限られている点が挙げられる。本実験は高齢者 1 名および開発者自身 1 名を対象とした探索的評価であり、得られた結果を高齢者一般に一般化することはできない。また、開発者自身はシステム仕様を理解しているため、発話内容や体調申告、推論結果への反応が一般利用者とは異なる可能性がある。

さらに、被験者 A は毎朝決まった時刻ではなく、任意のタイミングでシステムを起動していた。そのため、本実験で得られた体調申告には、起床直後の体調だけでなく、日中の疲労や夕方などのストレスなどが含まれている可能性がある。朝の体調確認は起床時の健康状態の把握に適している一方、夕方や夜の確認は日中の疲労や心理的負担の表出に適している可能性がある。今後は、利用時刻を統制した条件と、利用者の生活リズムに応じて確認タイミングを調整する条件を比較し、体調確認に適したタイミングを検討する必要がある。

5. ま と め

本研究ではエージェント対話を通じた高齢者体調確認サービスの実験的評価を行った。実環境において、表情感情値に基づく体調推論結果は、利用者の主観的な体調申告とどのような一致・不一致を示すか、表情の推論結果を交えた VA の深掘りは、高齢者の好不調に関する理由、自己開示を促すきっかけとなるか、提案システムは短期間の日常利用において、利用者にと

うように受容され、どのような利用上の課題が見られるかという三つの RQ を立て、提案システムの VA による体調確認、および深掘り対話がどのように機能するかを探索的に評価し、課題を洗い出すことを目的とした実験を行った。

その結果、表情感情値のみを用いた体調推論では、「普通」の体調ラベルが大部分を占める不均衡なデータ環境において、「良い」や「悪い」といった少数クラスを安定して識別することが困難であることが示唆された。一方で、推論結果を VA が提示し、それをきっかけとして体調の理由を尋ねる深掘り対話は、被験者が自身の疲れや不調を言語化する契機となる可能性が示された。さらに、毎日の利用においては、体調確認のやり取りが単調になりやすく、煩わしさや飽きにつながる可能性があることも明らかになった。

今後の課題としては少数クラスに対応可能な学習方法や、音声認識誤りを考慮した確認方法を導入すること、および体調申告を求める頻度やタイミング、質問形式を利用者の生活リズムや過去の回答傾向に応じて調整する仕組みの検討が挙げられる。

謝辞 本研究の一部は JSPS 科研費 JP25H01167, JP25K02946, JP25K24389, JP24K02765, JP24K02774, JP23K17006, JP23K28091, JP23K28383 の研究助成を受けて行われている。

文 献

- [1] 内閣府, “令和 6 年度版高齢社会白書,” https://www8.cao.go.jp/kourei/whitepaper/w-2024/zenbun/slash06pdf_index.html. accessed October 17, 2024.
- [2] 警察庁刑事局捜査第一課, “警察取扱死体のうち、自宅において死亡した一人暮らしの者～令和 6 年上半期 (1 6 月) 暫定値～,” https://www.npa.go.jp/publications/statistics/shitai/hitorigurashi/240827_kenshi2.pdf. accessed October 17, 2024.
- [3] 松川晃徳, K. Nahyun, 陳思楠, 片桐恵子, 中村匡秀, “音声対話エージェントが在宅高齢者の孤独に与える影響: 対話ログと孤独感尺度による分析,” 電子情報通信学会技術研究報告, 第 124 巻, pp.13–20, Nov. 2024.
- [4] 東京福祉保健局, “第 4 版住民の皆さんのための高齢者等の見守りガイドブック,” https://www.fukushi.metro.tokyo.lg.jp/kourei/koho/jyuuminnotameno.files/j_guidebook4.pdf. accessed October 17, 2024.
- [5] Y. Noguchi, H. Kamide, and F. Tanaka, “Personality traits for a social mediator robot encouraging elderly self-disclosure on loss experiences,” ACM Trans. Hum.-Robot Interact., vol.9, no.3, pp.1–24, May 2020.
- [6] 大藺隼人, 陳思楠, 佐賀雅樹, 中村匡秀, “音声対話エージェントとウェアラブルデバイスの連携による高齢者のセルフケア促進,” 電子情報通信学会技術研究報告; 信学技報, vol.121, no.336, pp.103–108, 2022.
- [7] 陳思楠, 大藺隼人, 中村匡秀, “高齢者自助支援のための音声対話エージェントを用いたマルチモーダルダイアリーサービスの提案,” 電子情報通信学会技術研究報告; 信学技報, vol.121, no.437, pp.75–80, 2022.
- [8] 堀江寛, 陳思楠, 中村匡秀, 安田清, “動画を活用した在宅高齢者のためのストレス解消サービスの研究,” 電子情報通信学会技術研究報告; 信学技報, vol.121, no.381, pp.1–6, 2022.
- [9] 西山敦紀, 中田匠哉, 陳思楠, 佐伯幸郎, 中村匡秀, “表情による感情認識を用いた体調推定手法の開発と評価,” 電子情報通信学会技術研究報告, 第 124 巻, pp.77–84, Match 2025.
- [10] 西山敦紀, 中田匠哉, 陳思楠, 佐伯幸郎, 増田廣介, 露崎雄太, 安田清, 中村匡秀, “感情分析による体調推定とエージェント対話を通じた高齢者体調確認サービスの提案,” 電子情報通信学会技術研究報告, 第 125 巻, pp.41–47, Match 2026.