

# 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

陳 思楠<sup>†</sup> 大藪 隼人<sup>†</sup> 中村 匡秀<sup>†,‡</sup>

<sup>†</sup> 神戸大学 〒657-8501 神戸市灘区六甲台町 1-1

<sup>‡</sup> 理化学研究所・革新知能統合研究センター 〒103-0027 東京都中央区日本橋 1-4-1

E-mail: †{chensinan,ozono}@ws.cs.kobe-u.ac.jp, ‡†masa-n@cs.kobe-u.ac.jp

**あらまし** 少子高齢化が進む日本では、在宅高齢者のセルフケアに対するニーズが高まっている。AIを活用したスマートヘルスケアが急速に進展しているが、記録・分析する在宅生活データの多様性や再利用性の不足が課題として残っている。先行研究では、高齢者の「心のうち」を傾聴するための音声対話エージェントを活用した傾聴システムを開発した。本研究の目的は、多角的な視点で対話データを記録・分析するマルチモーダル日記サービスを提案することである。テキストログだけでなく、音声や画像も取得し、マルチモーダルな可視化を行うことをキーアイデアとしている。提案手法では、異種データ抽出、Web API 開発、マルチモーダル日記生成、異種データ解析の主要なステップを説明する。また、異種データの拡張のアプローチについても述べる。提案手法を用いることで、高齢者自身がより良く過去を回想することができる。これによって、在宅介護や健康管理をより正しく、個人個人に提供できるようになることが期待できる。

**キーワード** セルフケア, 音声対話エージェント, 異種データ, マルチモーダル可視化, スマートヘルスケア

## Proposal of Multi-Modal Diary Service for Assisting Elderly Self-Care Using Spoken Dialogue Agent

Sinan CHEN<sup>†</sup>, Hayato OZONO<sup>†</sup>, and Masahide NAKAMURA<sup>†,‡</sup>

<sup>†</sup> Kobe University Rokkodai-cho 1-1, Nada-ku, Kobe, Hyogo 657-8501 Japan

<sup>‡</sup> Riken AIP 1-4-1 Nihon-bashi, Chuo-ku, Tokyo 103-0027 Japan

E-mail: †{chensinan,ozono}@ws.cs.kobe-u.ac.jp, ‡†masa-n@cs.kobe-u.ac.jp

**Abstract** As Japan's aging population and low birth rate, the self-care of elderly people at home has become increasingly in demand. Although the rapid progress of AI-enabled smart healthcare is developing, diversity and reusability of in-home daily living data to be recorded and analyzed are still challenging points. In our previous study, a system utilizing spoken dialogue agents for listening to the minds of elderly people was developed. The purpose of this study is to propose a multi-modal diary service that records and analyzes dialogue data with a diversified viewpoint. Our key idea is to retrieve not only the text logs but also audio and images and execute multi-modal visualization of them. In the proposed method, main steps including heterogeneous data extraction, Web API development, multi-modal diary generation, and heterogeneous data analysis are described. Moreover, the approach to extending heterogeneous data is also discussed. Using the proposed method, elderly people can recollect the past for their own better. In this way, the more correct improvement of in-home care and health management for all individuals is promising.

**Key words** Self-care, Spoken dialogue agent, Heterogeneous data, Multi-modal visualization, Smart healthcare

### 1. はじめに

世界中の人口高齢化に伴い、介護の需要は年々増加してい

る [1]. 特に在宅の高齢者向けに、工学技術に基づく *assistive technology* [2] を駆使し、セルフケアを改善することが広く研究開発されている。スマートヘルスケア領域では、日常生活行

動 (*Activities of Daily Living, ADL*) [3] を見守り、高齢者の状態とを理解するのが一般的なアプローチである [4] [5]。また、日常生活を記録し、在宅高齢者の記憶を支援するために、ロボットまたはスマートフォンを使用して対話を実行することは [6] [7] で提案されており、分析するためのライブビデオの録画も開発された [8] [9] [10]。

本研究で着目するチャレンジ課題には、**データの多様性と再利用性**の欠如が存在している。例えば、単一の形式のみを使用してデータを収集したり、完全なデータを記録したりするなど、重複のデータが含まれる場合がある。我々の趣味は、在宅高齢者の「**こころ**」を言葉で外化する [11] システムを開発・応用することである。先行研究では、**バーチャルエージェント (VA)** テクノロジーと IoT センサーを統合した**傾聴サービス** [12] が開発されました。高齢者とエージェントの間の「口頭での対話」を通して、「**心のうち**」を収集することを目的とし、*WebSocket* および *Pub/Sub* アーキテクチャ [13] を使用して、*MMDAgent* [14] を *Web Speech API* [15] に連携する汎用コンピュータと IoT センサー [16] を一般家庭に導入している。

本研究の目的は、音声対話エージェントを使用した高齢者セルフケアのためのマルチモーダル日記サービスを提案することである。日常生活データには、テキスト [17] だけでなく、音声や画像 [18] など、様々なモダリティが含まれている。我々のキーアイデアは、先行研究を踏まえ、豊富な対話データを記録および分析することである。アプローチとして、4つの主要なステップが提案された方法に分けられる：(1) 異種データ抽出。(2) Web API の開発。(3) マルチモーダル日記の生成。(4) 異種データ分析。このように、対話型対話エージェントを備えたシステムは、対話型対話を実行するだけでなく、多様な視点を持つ高齢者にとってよりよく考え直すという、より意味のあるものになることが期待できる。考察では、音声による対話データを純粋なテキストから音声や画像に拡張し、対話ログをより豊かで厚みのあるものにするが利点とされる。また、音声認識と感情分析のパフォーマンスは実験されることが不可欠であり、プライバシーとセキュリティの問題も検討する必要がある。

## 2. 準備

### 2.1 日本の高齢化社会と高齢者セルフケア

近年、世界中の先進国は、**高齢化社会** [1] での介護に対する需要の高まりに直面している。日本では、廃用症候群や認知症を持つ高齢者、一人暮らしの高齢者が年々増えている。医療、福祉の人員や施設の不足はますます深刻になっている。このような背景を受けて、日本政府は**地域包括ケアシステム**を積極的に強化している<sup>(注1)</sup>。これは自助や相互扶助、協力、および公的扶助で構成されている。さらに、少子化と COVID-19 の影響により、相互扶助、協力、公的扶助を進めることは困難である。したがって、自助は今後重要な役割を果たすと考えられる。

高齢者の自助に関連するこれまでの研究には、通常、日常生

活動 (*Activities of Daily Living, ADL*) [3] の維持とメンタルセルフケアが含まれ、自分の努力とニーズに完全に集中するように促す。特に一人暮らしの高齢者のメンタル**セルフケア**では、ストレスへの対処、健康管理、記憶力の強化は容易ではないことがよくある。上記のケースを改善するために、**スマートヘルスケア**分野で工学技術を使用する *assistive technology* [2] は世界中で急速に進歩している。最も代表的なシステムの1つは、在宅高齢者を見守ることである。機械 [4] [5] [10] によって高齢者の状態を理解するのが一般的である。

### 2.2 技術的なチャレンジ

在宅高齢者のセルフケアを便利に実現するために、我々は2つのチャレンジ課題があると考えている。

1 番目の課題は**データの多様性**である。高齢者が在宅日常生活中で発生する**忘却**を経験することは珍しいことではない。例えば、「私の財布はどこにありますか？」や「テレビで有名人の名前は何かですか？」などがある。高齢者が気付かなかった内容を記録・分析するために、既存の研究には、インタラクティブな家庭用ロボット [6] を使用したライブウォッチングサービス、スマートフォン [7] を使用した今後の記憶機能の改善、遠隔音響および臨床診療における音声録音と分析 [8] が含まれる。ただし、これまでの研究では、単一の形式（つまり、テキストまたは音声データのみ）でデータを収集している。高齢者の記憶にある特定のことを振り返り、効果的に支援することは難しい。

2 番目の課題は**データの再利用性**である。確かに、自宅で高齢者の記憶を助けるために、ライブビデオを記録するいくつかの関連する研究がある [9]。ビデオ分析技術 [10] を使用して、高齢者の日常生活の状態 (*context* と呼ぶ) を取得することも重要な研究トピックである。ただし、繰り返しのビデオクリップ (部屋の中を移動するなど) を多く含む完全なビデオを毎日蓄積するには、大きなストレージスペースが必要である。それに関する多くの場合、貴重な情報を取得するために効果的に再利用されることはない。ビデオ分析に関しては、莫大なコストとコンピューティングリソースを必要とするため、一般家庭に広く導入されることは非現実的である。したがって、データの再利用性を向上させるには、データ自体を最初は `textbf` 軽量のアプローチを行う必要がある。

### 2.3 先行研究：高齢者のための音声対話エージェント

我々の研究グループは、在宅高齢者の「**こころ**」**センシング** [11] の研究と開発を行ってきた。そのコア部分は、バーチャルエージェント (VA) 技術と IoT デバイスを利用することである。純粋な IoT センサーを使用する従来のアプローチとは異なり、「**ダイアログ**」の形式で *Human-Computer Interaction (HCI)* に焦点を当てている。最新の成果の1つは、**傾聴サービス** [12] を展開して運用することである。高齢者との会話や挨拶 (日本語の相槌 [19]) にランダムに返信することができる。図 1 は、高齢者向けの傾聴サービスの概念図を示している。

具体的には、一般的なコンピュータと IoT センサー (Phidgets [16] モーションおよび圧力センサー) の2つのハードウェアコンポーネントで構成されている。一方、*MMDAgent* [14] (エージェント素材、音声合成) をリンクする一般的なコンピュー

(注1) : [https://www.mhlw.go.jp/english/policy/care-welfare/care-welfare-elderly/dl/establish\\_e.pdf](https://www.mhlw.go.jp/english/policy/care-welfare/care-welfare-elderly/dl/establish_e.pdf)

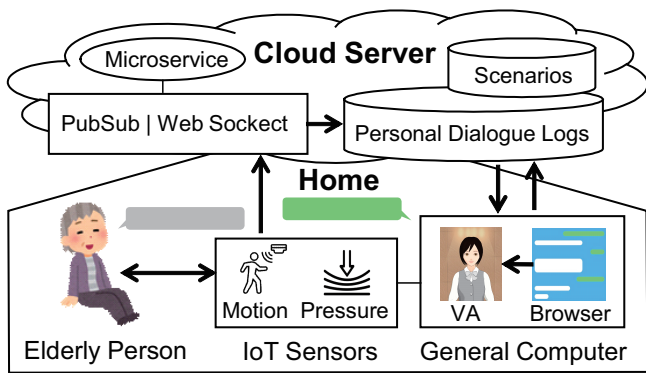


図 1 高齢者向けの傾聴サービスの概念図

表 1 エージェントが 3 時間ごとに尋ねる質問内容の例

Period of Time	Question Contents
6:00 ~ 8:59	Can you tell me how you are feeling today?
9:00 ~ 11:59	Is there anything you want to talk about now?
12:00 ~ 14:59	Can you tell me what you did this morning?
15:00 ~ 17:59	Can you talk to me about whatever your mind?
18:00 ~ 20:59	Can you tell me what you did this afternoon?
21:00 ~ 23:59	Can you tell me what you felt and did today?

ターと Web ブラウザー (Web Speech API [15] を使用した音声認識, 対話コンテンツ UI) の 2 つのソフトウェアコンポーネントで構成されている。WebSocket および Pub/Sub アーキテクチャ [13] を使用して、一般的なコンピューターと IoT センサーが接続される。このように、エージェントは自動的かつ定期的に質問する。これは、IoT センサーによって検出された自宅の高齢者が「存在」している場合にのみ実行される。表 1 に、エージェントが 1 日 3 時間ごとに尋ねる質問内容の例を示す。

## 2.4 マルチモーダルおよびマルチメディアデータ

在宅高齢者のセルフケアを支援するための重要なステップの 1 つは、日常生活データを収集することである。それらには、テキストや音声の形式の言語データ [17] から、画像や動画の形式の視覚データ [18] まで、様々なモダリティのほか、次のような環境データ (温度, 湿度, または照度 [20]) も含まれる。次世代のスマートヘルスケアでは、個人の日常生活のより豊富な説明を提供するために、時系列で異種データを収集し、マルチモーダル視覚化を提供することは非常に重要な研究トピックである。さらに、時系列の異種データをさらに分析することも有望である。例えば、テキストデータ [21] への感情分析, 顔画像データ [22] への感情推定, 音声データ [23] への韻律特徴抽出などがある。これらの時系列での結果によって、今後様々な付加価値サービスに拡張できる。

## 3. 提案手法

### 3.1 目的とキーアイデア

章 2.2 で説明されている 2 つの課題を改善するために、音

声対話エージェントに基づくマルチモーダル日記サービスを提案することを目指している。本研究の長期的な目標は、高齢者に過去の記憶をよりよく考えさせることである。例えば、Web ページのマルチモーダル日記を確認することで、気づいたり覚えたりできないことが改善できる。我々のキーアイデアは、対話データをマルチモーダル形式で記録および分析することである。テキストだけでなく、画像や音声も含まれている。これにより、音声対話エージェントを備えたシステムがより意味のあるものになり、コミュニケーションスキルが向上するだけでなく、在宅高齢者向けのマルチモーダル視覚化を使用した記憶補助も向上する。

### 3.2 全体アーキテクチャ

図 2 は、在宅高齢者向けに提案されたマルチモーダル日記サービスの全体的なアーキテクチャを示している。先行研究に基づき、タイミングコンソールは、会話エージェントによるタイムリーな質問や挨拶への返信において重要な役割を果たす。提案手法は、4 つの主要なステップに分かれている。

具体的には、Step 1 では、高齢者の対話コンテンツをテキスト, 音声, および画像の形式で収集する方法について説明する。このステップでは、クライアントから日時とともに収集されたダイアログコンテンツがクラウドデータベースに送信される。Step 2 では、クラウドデータベースとクライアントデバイス間の転送ステーションとして、異種データを取得する方法を構築するための Web API の開発について説明する。Step 3 では、Web ブラウザーでマルチモーダル日記を生成するためのアプローチを提案する。これにより、異種データのマルチモーダル視覚化が構築される。Step 4 では、テキスト感情分析, 音声感情検出, 顔の感情推定など、収集されたデータの拡張アプローチについて説明する。さらなる付加価値サービスについても説明する。

### 3.3 Step 1: 異種データの抽出

#### Step 1-1: 音声認識

Step 1-1 では、高齢者と話し言葉の対話エージェントとの間のスムーズな相互作用を構築するために、話し言葉の文脈をテキストの形で記録する必要がある。Speech-to-Text は、音声を認識し、特定の言語でテキストを抽出するためのテクノロジーである。通常、Web application programming interface (API) の形式で使用される。伝統的な手法には、Web ブラウザを使用した Web Speech API [15] 多くのクラウドサービス (Google, Microsoft, Watson など) [24] を使用した Speech to Text API が含まれる。

章 2.3 で説明した先行研究と同様に、Web 音声 API (音声認識) と MMDAgent (音声合成) を使用するシステムが提案されている。変換されたテキストログは、クラウド上の MongoDB データベース [25] に送信された。音声対話エージェントからの質問または挨拶のテキストログも MongoDB に送信された。クラウドデータベースに送信するダイアログコンテキストのデータサンプルを以下に示す。音声認識と合成の間に時間間隔が必要であることを注意する必要がある。それは個人の話し言葉の速度に適用されるべきである。

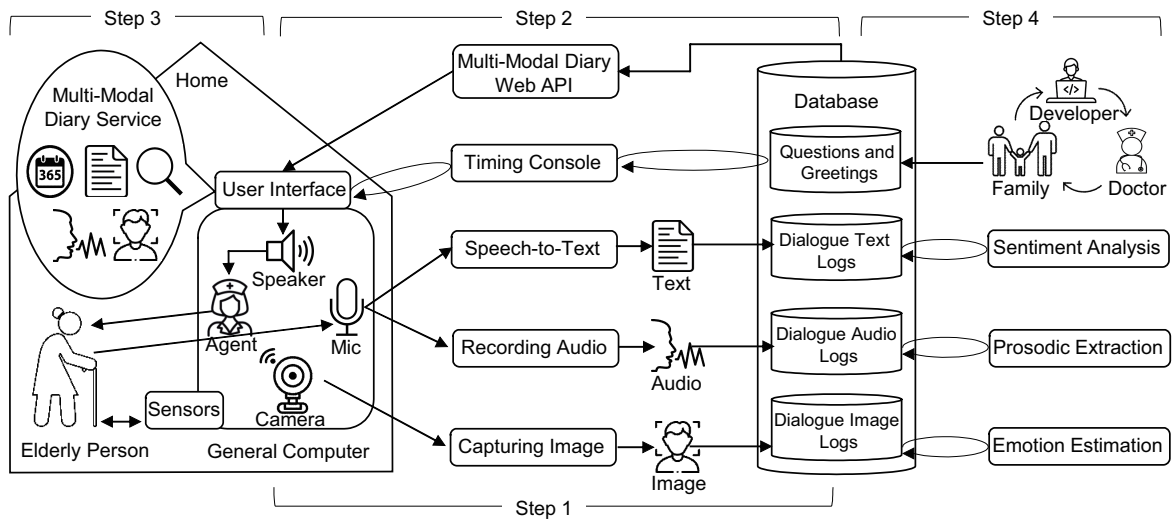


図 2 提案手法の全体アーキテクチャ

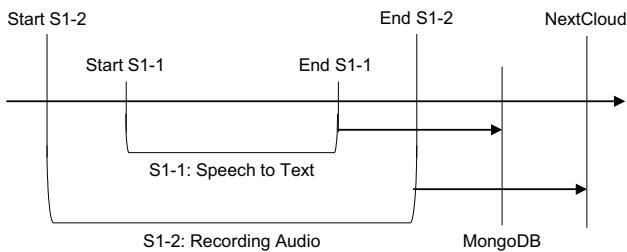


図 3 音声とテキストの録音を表す概略図

クラウドデータベースに送信するデータの例

```

{
  "_id": "data id",
  "from": "user-id",
  "to": "keicho-bot",
  "contents": "I had a backache at night",
  "dataType": "text",
  "timestamp": "2021-09-08T18:02:25+09:00",
  "tag": "va.keicho.user-id",
  "time": "2021-09-08T09:02:19.475Z"
}

```

### Step 1-2: オーディオの録音

Step 1-2 では、高齢者からの音声生データを記録するには、最も素朴なアプローチは、JavaScript プログラミングによって Web ブラウザに埋め込まれたオーディオ録音機能を構築することである。典型的な手法の 1 つは、Web Audio API [26] である。通常、関数 `navigator.getUserMedia` を使用して、オーディオバッファ配列を入力するためのメディアストリームソースを作成する。先行研究では、録音されたオーディオバッファ配列は NextCloud データベース [27] に送信され、*wav* の形式でオーディオファイルとして保存された。音声を録音するプロセスには、音声からテキストへの変換が含まれることが必要となる。その概略を図 3 に示し、チャレンジポイントは異なる特徴間の対話と傾聴のスムーズな変換である。

表 2 時系列で画像データをキャプチャするための関連パラメータ

Shooting device	Camera (Lenovo IdeaPad C340)
Shooting method	Drawing canvas from video stream
Shooting interval	One second
HTML5 canvas size (pixel)	320 × 240
Computer browser	Chrome (Version 84.0.4147.125)

### Step 1-3: 画像のキャプチャ

Step 1-3 では、高齢者との対話中に画像を収集するための最も単純なアプローチは、Web ブラウザにも組み込まれた画像キャプチャ機能を構築することである。Step 1-2 の期間で、時系列の画像データをキャプチャするために、関連パラメータを表 2 に示されている。具体的には、まず、ビデオストリームが Web ブラウザで作成される。次に、関数 `drawImage()` を使用して、ビデオスナップショットから *HTML5 canvas* が毎秒描画される。次に、*base64encode* 形式の生データが、関数 `canvas.toDataURL()` によってキャンバスから抽出される。*base64* エンコードが NextCloud データベースに送信され、*jpg* の形式で画像ファイルとして保存される。すべてのファイル名が日時を使用することが必要である。

### 3.4 Step 2: Web API 開発

#### Step2-1: 対話テキストの取得

Step 2-1 では、クラウドデータベースから過去のダイアログテキストログを取得するために、Keicho-Log という名前の WebAPI が開発される。通常、Java や Nodejs などのさまざまなプログラミング言語を使用して開発できる。ダイアログテキストデータを取得するための WebAPI 仕様の例を以下に示す。先行研究に基づき、ID (つまり、*uid*) によってさまざまなユーザーを示す *You-IDSERVICE* が提案されている。*uid* と WebAPI を統合することにより、過去のすべてのダイアログテキストログを抽出できるだけでなく、キーワード、日付、お

## Example of multi-modal diary in Japanese

```

めいちゃ
ん：
08:12:33 チンさん、おはようございます、今朝のご気分はいかがですか？
チン：
08:13:18 今日は普通です
08:13:42 午前中はライナー券のミーティングがあります
08:13:59 今日から健康を意識することをしなさいといけませんね
08:14:27 そしていんなプレゼンを準備しないといけません本当に頑張らないといけません効率よくやりましょう
08:14:39 カレンダー
08:14:46 確認
08:14:54 今日の予定
08:15:06 作成
08:15:15 今日の10時
08:15:25 共同研究のミーティング

```

## Multi-modal visualization and analysis

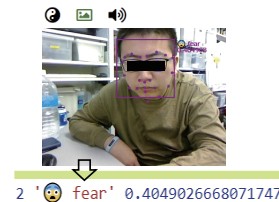
そしていんなプレゼンを準備しないといけません本当に頑張らないといけません効率よくやりましょう

```

"eachWord": [
  "準備", "ない", "と", "ない", "ない", "と", "ない", "効率", "よく"
],
"eachScore": [
  -0.075, -0.102, -0.022, -0.102, -0.102, -0.022, -0.102, -0.019, 0.922
],
"total": 0.373,
"min": ["ない", -0.102],
"max": ["よく", 0.921]

```

(a)



(b)



(c)

図4 提案手法を用いた日本語の例：(a) テキスト感情分析 (b) 顔の感情推定 (c) 音声感情検出。

よび制限された数で検索することもできる。さらに、クラウドで *Cross-Origin Resource Sharing (CORS)* を承認することにより、関数 `jQuery.ajax()` または `FetchAPI` を使用してクライアントからダイアログテキストログを抽出できる。

データを取得するための Web API 仕様の例

```

GET ~/uid={uid}/
GET ~/uid={uid}/count={count}
GET ~/uid={uid}/keyword={keyword}
GET ~/uid={uid}/date={date}
GET ~/uid={uid}/keyword={keyword}/count={count}
GET ~/uid={uid}/date={date}/count={count}
GET ~/uid={uid}/date={date}/keyword={keyword}

```

### Step 2-2：対話音声の抽出

Step 2-2 で、クラウドデータベースからオーディオデータを取得するための重要なステップは、保存されたファイル名から指定された日時を検索することである。タイミングの違いにより、ファイル名はオーディオとイメージで異なること(図3を参照)が必要である。NextCloud データベースを使用する場合、Web URL には、エンドポイントを含むファイル ID または名前(例えば、`~/nextcloud/apps/files/?dir=fileid`)が表示される。

### Step 2-3：対話画像の読み込み

Step 2-3 では、クラウドデータベースから画像データを取得するための重要なステップは、保存されたファイル名から指定された日時を検索することでもある。手順 2-2 と手順 2-3 の共通点は、音声の生データは `buffer array` の形式ですが、画像の場合は `base64encode` の形式であるということである。

## 3.5 Step 3：マルチモーダル日記の生成

### Step 3-1：質問によるテキストの表示

Step 3-1 で、話し言葉のテキストログを取得するには、ログのコンテキスト(例えば、このトピックを高齢者から言った理

由)を明確に確認する必要がある。このために、特定の質問(表1を参照)の後の高齢者の回答が取得され、特定のキーワードで終了する Web ページに表示される。また、音声対話を停止するためのキーワードは事前に設定しておく必要がある。

### Step 3-2：音声と画像をテキストに接続する

Step 3-2 では、音声と画像をテキストで接続するために、最小時間間隔の音声、画像、テキストを含む一意のグループを計算し、ファイル名で取得できる。このようにして、質問ごとに、回答テキストとそのグループをリストに表示できる。表示時間は画像ファイル名から抽出されているため、テキストや音声よりも画像が送信されやすいことである。図4は、日本語のマルチモーダル日記サービスの例を示している。

## 3.6 Step 4：異種データ分析

Step 4 では、テキスト、音声、画像の形式で対話データからさらに豊富な情報を取得するために、感情分析、感情検出、推定などの方法も提案される。出力された結果の例を図4に示す。(a) ダイアログテキストログは、ネガポジ判定<sup>(注2)</sup>にかけて、その色は程度を表すために使用された。(b) 対話画像からの感情スコアは、face API と tensorflow.js [28] を使用して元の画像に描画される。(c) 感情的な傾向は、ユーザーローカル API<sup>(注3)</sup>との対話音声から検出される。

## 4. 考察とまとめ

本論文では、音声対話エージェントに基づくマルチモーダル日記サービスを提案した。本研究の利点として、音声による対話データを純粋なテキストから音声や画像に拡張し、対話ログをより豊かで厚みのあるものにする。ただし、本研究の限界については、音声認識と感情分析のパフォーマンスは実験されていない。さらに、プライバシーとセキュリティの問題も非常に

(注2) : <https://github.com/azu/negaposi-analyzer-ja>

(注3) : <https://ir.userlocal.jp/en/>

重要であり、調査する必要がある。今後の課題として、マルチモダル対話サービスを実世帯に展開し、提案手法の有効性を評価する予定である。音声対話エージェントの多言語バージョンを拡張することも興味深い研究である。

**謝辞** 本研究の一部は JSPS 科研費 JP19H01138, JP18H03242, JP18H03342, JP19H04154, JP19K02973, JP20K11059, JP20H04014, JP20H05706 および、立石科学技術振興財団の研究助成を受けて行われている。

## 文 献

- [1] S. Takatsuki, "Are we ready for the upcoming super-aging society?," *Circulation Journal*, pp.CJ-21, 2021.
- [2] F. Embarak, N.A. Ismail, and S. Othman, "A systematic literature review: the role of assistive technology in supporting elderly social interaction with their online community," *Journal of Ambient Intelligence and Humanized Computing*, vol.12, no.7, pp.7427-7440, 2021.
- [3] M. Jahouh, J.J. González-Bernal, J. González-Santos, D. Fernández-Lázaro, R. Soto-Cámara, and J. Mielgo-Ayuso, "Impact of an intervention with wii video games on the autonomy of activities of daily living and psychological-cognitive components in the institutionalized elderly," *International Journal of Environmental Research and Public Health*, vol.18, no.4, p.1570, 2021.
- [4] S. Chen, S. Saiki, and M. Nakamura, "Proposal of home context recognition method using feature values of cognitive api," 2019 20th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)IEEE, pp.533-538 2019.
- [5] S. Chen, S. Saiki, and M. Nakamura, "Nonintrusive fine-grained home care monitoring: Characterizing quality of in-home postural changes using bone-based human sensing," *Sensors*, vol.20, no.20, p.5894, 2020.
- [6] K. Abe, H. Fuse, M. Okumura, and M. Isshiki, "Development of a life watching service for elderly people by using interactive home robot," *International Journal of Informatics Society (IJIS)*, vol.12, no.1, pp.59-67, 2020.
- [7] M.K. Scullin, W.E. Jones, R. Phenix, S. Beevers, S. Rosen, K. Dinh, A. Kiselica, F.J. Keefe, and J.F. Bengel, "Using smartphone technology to improve prospective memory functioning: A randomized controlled trial," *Journal of the American Geriatrics Society*.
- [8] S.L. Schneider, L. Habich, Z.M. Weston, and C.A. Rosen, "Observations and considerations for implementing remote acoustic voice recording and analysis in clinical practice," *Journal of Voice*, 2021.
- [9] M.S. Anderson, N. Homdee, A. Bankole, R. Alam, B. Mitchell, J. Hayes, G. Byfield, and J. Lach, "Behavioral interventions for alzheimer's management using technology: home-based monitoring," *Current Geriatrics Reports*, vol.9, no.2, pp.90-100, 2020.
- [10] E. Airola, P. Rasi, and M. Outila, "Older people as users and non-users of a video conferencing service for promoting social connectedness and well-being-a case study from finnish lapland," *Educational Gerontology*, vol.46, no.5, pp.258-269, 2020.
- [11] H. Maeda, S. Saiki, M. Nakamura, and K. Yasuda, "Rule-based inquiry service to elderly at home for efficient mind sensing," *Proceedings of the 21st International Conference on Information Integration and Web-based Applications & Services*, pp.664-668, 2019.
- [12] H. Ozono, S. Chen, and M. Nakamura, "Study of microservice execution framework using spoken dialogue agents," In 22th IEEE-ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel Distributed Computing (SNPD2021), Nov. 2021.
- [13] F. Chen, P. Liu, J. Zhu, S. Gao, Y. Zhang, M. Duan, Y. Wang, and K. Hwang, "Improving topic-based data exchanges among iot devices," *Security and Communication Networks*, vol.2020, 2020.
- [14] A. Lee, K. Oura, and K. Tokuda, "Mmdagent-a fully open-source toolkit for voice interaction systems," 2013 IEEE International Conference on Acoustics, Speech and Signal ProcessingIEEE, pp.8382-8385 2013.
- [15] J. Adorf, "Web speech API," *KTH Royal Institute of Technology*, 2013.
- [16] S. Greenberg and C. Fitchett, "Phidgets: easy development of physical interfaces through physical widgets," *Proceedings of the 14th annual ACM symposium on User interface software and technology*, pp.209-218, 2001.
- [17] H. Maeda, S. Saiki, M. Nakamura, and K. Yasuda, "Recording daily health status with chatbot on mobile phone-a preliminary study," 2019 Twelfth International Conference on Mobile Computing and Ubiquitous Network (ICMU)IEEE, pp.1-6 2019.
- [18] S. Chen, S. Saiki, and M. Nakamura, "Toward flexible and efficient home context sensing: Capability evaluation and verification of image-based cognitive apis," *Sensors*, vol.20, no.5, p.1442, 2020.
- [19] I. Imelda, A. Rahmat, and N. Nuruddin, "Development learning materials for the responding skills in speaking the basic japanese language-beginning step," *Journal of Educational Science and Technology (EST)*, vol.7, no.1, pp.26-31, 2021.
- [20] K. Tamamizu, S. Sakakibara, S. Saiki, M. Nakamura, and K. Yasuda, "Capturing activities of daily living for elderly at home based on environment change and speech dialog," *International Conference on Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk ManagementSpringer*, pp.183-194 2017.
- [21] S. Chen and M. Nakamura, "Generating personalized dialogues based on conversation log summarization and sentiment analysis," *The 23rd International Conference on Information Integration and Web Intelligence*, pp.217-222, 2021.
- [22] K. Hirayama, S. Chen, S. Saiki, and M. Nakamura, "Toward capturing scientific evidence in elderly care: Efficient extraction of changing facial feature points," *Sensors*, vol.21, no.20, p.6726, 2021.
- [23] K.C. Fraser and M. Komeili, "Measuring cognitive status from speech in a smart home environment," *IEEE Instrumentation & Measurement Magazine*, vol.24, no.6, pp.13-21, 2021.
- [24] I. Siegert, Y. Sinha, O. Jokisch, and A. Wendemuth, "Recognition performance of selected speech recognition apis-a longitudinal study," *International Conference on Speech and ComputerSpringer*, pp.520-529 2020.
- [25] A. Utka, et al., "Berri corpus manager: A corpus analysis tool using mongodb technology," *Human Language Technologies-The Baltic Perspective: Proceedings of the Ninth International Conference Baltic HLT 2020*, vol.328IOS Press, p.166 2020.
- [26] H. Lindetorp and K. Falkenberg, "Putting web audio api to the test: Introducing webaudioxml as a pedagogical platform," *Web Audio Conference 2021*, 2021.
- [27] N. Sari, H. Amnur, and R. Hidayat, "Monitoring next cloud sebagai private cloud storage dengan notifikasi telegram," *JITSI: Jurnal Ilmiah Teknologi Sistem Informasi*, vol.1, no.4, pp.144-149, 2020.
- [28] H. Agarwal, G. Verma, and L. Gupta, "Student attendance system based on the face recognition," *Asian Journal For Convergence In Technology (AJCT) ISSN-2350-1146*, vol.7, no.2, pp.70-73, 2021.