

# 拡張現実と大規模言語モデルを用いた認知症ケア訓練システム

## HEARTS 5 powered by ChatGPT の開発

倉爪 亮（九州大学） ○西浦 悠生（九州大学）

### Development of Dementia Care Training System using AR and LLM

Ryo KURAZUME (Kyushu Univ.) and ○ Yuuki NISHIURA (Kyushu Univ.)

Abstract: We have developed HEARTS, a dementia care training system using augmented reality based on humanity. Humanity is a multimodal comprehensive care technique for dementia, and has attracted attention as a method to reduce the burden on both caregivers and patients. However, the HEARTS developed so far could not evaluate "speaking" skills based on the content of conversations among "seeing," "touching," and "speaking," all of which are fundamental skills in humanity. Therefore, we attempted a new quantitative evaluation of trainees' "speaking" skills by estimating the emotional value of conversational content using ChatGPT.

#### 1. 緒言

近年、医療や介護の現場では、認知症など認知機能の低下した高齢者の介護負担が増加している。認知機能の低下した高齢者の介護は過酷であり、介護士や看護師の不足、疲弊、バーンアウトや被介護者の身体機能の低下などの多くの問題が存在する。これらの問題に対して、近年注目されているケア技法であるユマニチュード（Humanitude）は、知覚、感情、言語による包括的コミュニケーションに基づいたケア技法であり、「見る」「触れる」「話す」「立つ」の4つのスキルを基本とし、150の具体的なケア技法から構成されている<sup>[1]</sup>。ユマニチュードは認知症患者に対する極めて有効なケア技法であるが、その習得の困難さが普及への課題となっていた。そこで我々は、家庭を含む介護現場へのユマニチュード導入のための教育手法の確立を目指し、これまでに拡張現実を利用した認知症ケア教育システム HEARTS (HumanitudE AR Training System) を開発した<sup>[2] [3] [4] [5] [6] [7]</sup>。

HEARTS の概要を Fig.1 に示す。HEARTS は、触れることができる柔らかな人形の顔の部分に、仮想画像を拡張現実デバイス（HoloLens 1, HoloLens 2, Microsoft）で重畳表示することで、人形に触れながら、「見る」「話す」スキルが訓練できる。さらに介護者が触れた位置や力を測定できるウェアラブル全身触覚センサ<sup>[8]</sup>や圧力分布が測定できるデータグローブ<sup>[6]</sup>を用いることで、「触れる」スキルも評価可能である。また、スキルの良し悪しにより、仮想の顔の表情を変化させることで、実際の被介護者とのコミュニケーションを模擬することができる。

本システムは既に複数の国立、私立大学の医学部および看護学部での教育に導入され、これまでに300名を超える教育実績がある。更に本システムを用いたユマニチュード教育は、人形を用いた場合に比べて教育効果が高いことが示されている<sup>[9]</sup>。

しかしこれまでに開発した HEARTS (HEARTS 1 ~ HEARTS 4) では、ユマニチュードで重要な「見る」「触れる」「話す」のうち、会話の内容に基づく「話す」スキルの評価が不可能だった。そこで本稿では、新たに ChatGPT を用いた会話内容の感情値推定を導入し、「話す」スキルの定量評価を試みた新たなシステム HEARTS 5 を紹介する。

仮想現実で顔を重畳表示

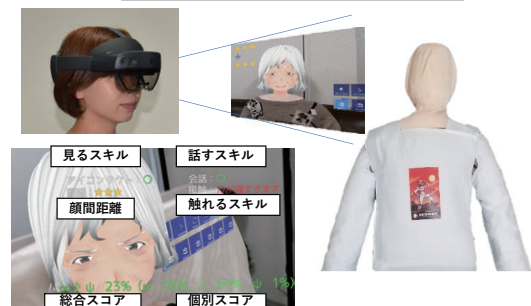


Fig.1 Dementia care training system, HEARTS

#### 2. HEARTS 4

HEARTS 4 は、2023 年 4 月に Microsoft store および github で無償公開されたシステムである<sup>[10] [11]</sup>。HEARTS 4 では、仮想現実デバイス HoloLens 2 を用いて、仮想の顔を人形の顔部分に重畳表示する。また顔のモデルには、Fig.2 に示すように、アニメ調の高齢男性、女性、若い男性、女性、およびリアルな高齢女性の5つが用意されており、訓練者は自分に適したモデルでユマニチュードを学習できる。

アバターには、顔の向きや目線の変更、瞬き、視線の追跡（アイコンタクトが取れている場合のみ）などが実装されている。更に、Fig.3 に示すように、「幸せ」「悲しみ」「怒り」「驚き」「不安」、および「ノーマル」の6つの表情が用意されており、アイコンタクトの成立や後述する会話の内容、顔間距離、突然の声かけ、声の大きさ（怒鳴り声など）などに応じて表情が変化する。Fig.4 は、顔を近づけすぎたために、怒りの表情を示している例である。

HEARTS 4 では、目線を合わせて要介護者を「見る」スキル、掌全体を使って優しく「触れる」スキル、優しく適度の声の大きさに「話す」スキルを計測し、それらスキルが個別、あるいは同時に提供できていた時間の割合をスコアとして提示する機能が実装されている。しかし、このうち「話す」スキルについては、一定時間の音圧を基に、話していたか否かのための評価であり、ユマニチュードで重要な「ポジティブな内容で話す」など、会話の内容に関する評価は不可能であった。



Fig.2 Avatars available in HEARTS



Fig.3 Six types of facial expressions

### 3. HEARTS 5

HEARTS 5 では、HEARTS 4 で実現されていない会話の内容の評価に取り組んだ。

具体的には、以下の順序で実行される会話内容のリアルタイム評価機能を開発した。

1. 会話時の音声データの切り出し
2. 1. で得られた音声データをもとに文字起こし
3. 会話文を大規模言語モデル (ChatGPT, OpenAI) へ入力し、アバターの感情  $E_{emotion} \in [0, 1]$  と会話内容のポジティブ度合い  $E_{property} \in [0, 1]$  を推定。(ただし、 $E_{emotion}, E_{property}$  はポジティブであれば 1, ネガティブであれば 0 に近づく。)

HEARTS 5 では、アバターの表情をアバターの感情  $E_{emotion}$  に基づいて決定する。Fig.5 に一例を示す。会話の内容が ChatGPT によりポジティブと判定された場合 ( $E_{property}$  が 0.5 以上) には、画面に "Positive" と表示され、さらに Positive な会話が続くと、アバターの感情値  $E_{emotion}$  が高まり、幸せの表情を示す。一方、ChatGPT によりネガティブと判定された場合 ( $E_{property}$  が 0.5 未満) には、画面に "Negative" と表示され、さらにアバターの感情値  $E_{emotion}$  が下がると悲しみの表情に変化する。

#### 3.1 会話時の音声データの切り出し

提案する評価機能では、音声データから文字起こしを行う。その際に、文字起こしのリアルタイム性や正



Fig.4 If a trainee gets too close, the avatar will have an "angry" look on his/her face.

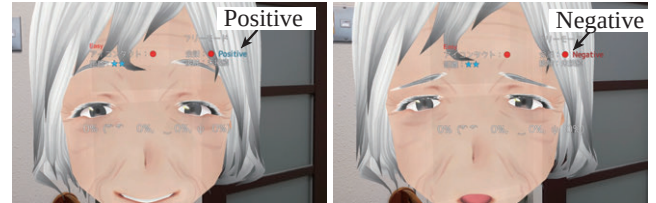


Fig.5 Examples of content evaluation in conversation.

確性を確保するため、会話の発生をシステムが認識し、会話中の音声のみを録音する必要がある。会話中および会話前後の音圧の時間推移を Fig.6 に示す。Fig.6 に示すように、発声中の音圧は会話中に環境音と同一レベルになるタイミングがあり、単なる音圧に関する閾値による会話発生の検知は困難である。

そこで、本研究では、会話の発生を音圧に関する閾値  $T_v$  に加え、無音継続時間についての閾値  $T_s$  を用いて会話検知機能を開発した。

開発した会話検知機能では、マイク入力が  $T_v$  を超えると会話が始まったと判断し録音を開始する。そして、録音中に、 $T_s$  を超える時間の無音状態を検知すると、会話が終了したと判断し録音を終了する。本システムでは、システムによる応答のリアルタイム性を確保するため、 $T_s$  を 0.5 秒とした。

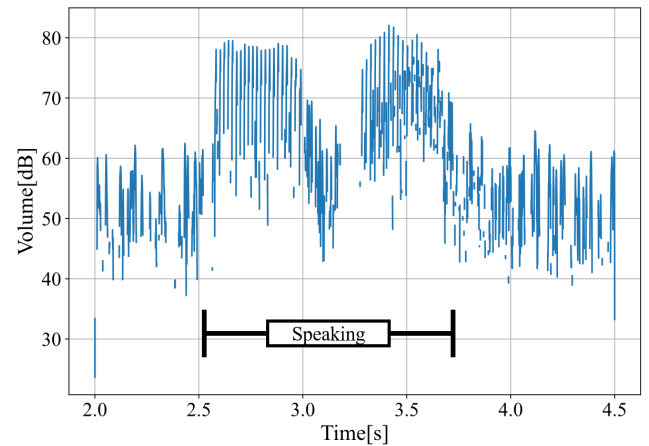


Fig.6 Example of time transition of amplitude during vocalization.

#### 3.2 会話文の文字起こし

録音データは、インターネット回線を通じディクテーションソフトウェア (Whisper, OpenAI) によって文字起こしが行われる。このディクテーションの処理の平均処理時間を Fig.7 に示す。Fig.7 に示すように、Whisper では、会話文の文字起こし処理を 0.9 秒から 1.5 秒で処理することが出来る。

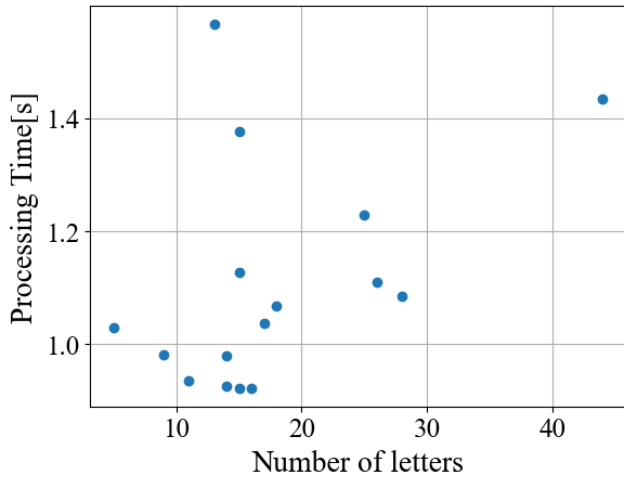


Fig.7 Figure plotting processing time for transcription by Whisper.

### 3.3 $E_{emotion}$ の推定

ディクテーションソフトウェアによって得られた文字起こし結果を ChatGPT へ入力し、アバターの感情  $E_{emotion}$ 、および直前の会話内容のポジティブ度合い  $E_{property}$  の推定結果を受け取る。

この際、事前に ChatGPT のロール設定を利用し、アバターのキャラクタ設定や返答のフォーマットを ChatGPT へ指示した。本研究で事前設定として ChatGPT へ入力した情報（一部抜粋）を Fig.8 に示す。

また、ChatGPT による  $E_{property}$  の推定時間をプロットした図を Fig.9 に示す。Fig.9 に示すように、ChatGPT による  $E_{property}$  の推定は 1.0 秒から 3.0 秒で処理することが出来る。

```
You will role-play as a pseudo-emotional chatbot according to the following conditions. You do not need to respond verbally.
<omission>
The output format of subsequent conversations shall be in the following json format. Do not generate sentences in any other format than this one.
Format:
{
  emotion: float
  property: float
}
```

Fig.8 A part of prompt given to ChatGPT.

## 4. 動作例

開発した HEARTS 5 の動作検証の一例を Figs.10, 11 に示す。Fig.10 は一例として、” 顔色がいいですね” と声掛けをした場合、Fig.11 は” 動かないください” と声掛けをした場合の表情変化である。画面右上の”Positive”、”Negative”は、直前の会話の評価値  $E_{property}$  を示し、感情の評価値  $E_{emotion}$  により表情が変化する。

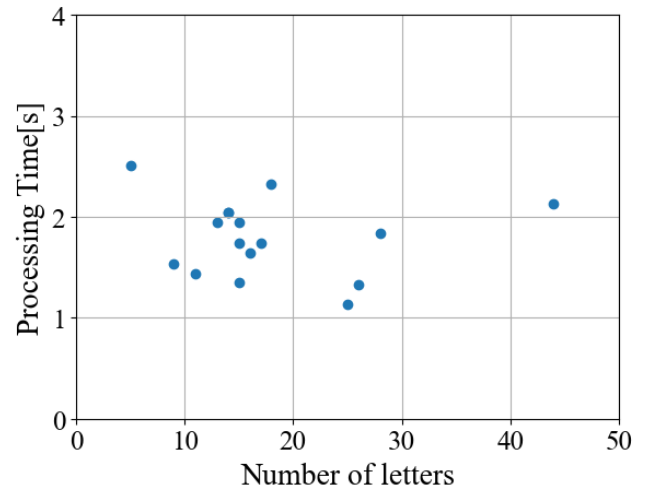


Fig.9 Figure plotting processing time for estimation of  $E_{emotion}$  by ChatGPT.



Fig.10 Examples of positive conversation.

## 5. まとめ

本稿では、新たに開発した、ChatGPT を用いて会話内容の評価が可能な認知症ケア教育システム HEARTS 5 を紹介した。提案した会話内容の評価機能は、およそ 1.9 秒から 4.5 秒の間で会話内容のポジティブ度合いを評価することが可能である。今後は、ロボットと人間間の対話において理想とされるロボットの反応時間である 2 秒以内<sup>[12]</sup>に、 $E_{property}$  の推定を行えるよう改良する。

## 謝辞

本研究は、JST CREST JPMJCR17A5 の支援を受けたものである。

## 参考文献

- [1] 本田美和子, イヴ・ジネスト, and ロゼット・マレス コッティ: ユマニチュード入門. 医学書院, (2014).
- [2] Tomoki Hiramatsu et al.: “Development of dementia care training system based on augmented reality and whole body wearable tactile sensor”. 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2020). 2020, pp. 4148–4154. URL: <http://robotics.ait.kyushu-u.ac.jp/kurazume/papers/HiramatsuIROS20.pdf>.
- [3] 平松知樹 et al.: “拡張現実と分布型接触センサを組み合わせた認知症ケア教育システムの開発”. ロボティクス・メカトロニクス講演会講演概要集 2019. 一般社団法人 日本機械学会. 2019, 1A1-Q03.
- [4] 平松知樹 et al.: “拡張現実と分布型触覚センサを組み合わせた認知症ケア教育システムの開発 -シナリオに基づく訓練システムと実証試験-”. 日本機械学会ロボティクスメカト





Fig.11 Examples of negative conversation.

- ロニクス講演会 2020. 2020.5.27-29, 1A1-D09. URL: <http://robotics.ait.kyushu-u.ac.jp/kurazume/papers/ROBOMECH20-1.pdf>.
- [5] 倉爪 亮 and Qi An: “拡張現実と分布型触覚センサを組み合わせた認知症ケア教育システムの開発 第二次試作システム HEARTS 2 の開発と評価実験”. 日本機械学会ロボティクスメカトロニクス講演会 2021. 2021.6.6-8, 1A1-C06. URL: <http://robotics.ait.kyushu-u.ac.jp/kurazume/papers/ROBOMECH21-1.pdf>.
- [6] 倉爪 亮 and Qi An: “拡張現実と分布型触覚センサを組み合わせた認知症ケア教育システムの開発 -第三次実証システム HEARTS 3 の開発-”. 第 39 回日本ロボット学会学術講演会. 2021.9.8-11, 2G1-02. URL: <http://robotics.ait.kyushu-u.ac.jp/kurazume/papers/RSJ21-4.pdf>.
- [7] Ryo Kurazume et al.: “Development of AR training systems for Humanitude dementia care”. *Advanced Robotics* 36.7, pp. 344–358, (2021). DOI: 10.1080/01691864.2021.2017342. URL: <https://doi.org/10.1080/01691864.2021.2017342>.
- [8] 亀井 雅哉 et al.: “介護技術定量化のためのウェアラブル全身触覚センサの開発”. 日本機械学会ロボティクスメカトロニクス講演会 2020. 2020.5.27-29, 1A1-D10. URL: <http://robotics.ait.kyushu-u.ac.jp/kurazume/papers/ROBOMECH20-2.pdf>.
- [9] Atsushi Nakazawa et al.: “Augmented reality-based affective training for improving care communication skill and empathy”. *PLOS ONE* 18.7, e0288175, (2023). DOI: 10.1371/journal.pone.0288175. URL: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0288175>.
- [10] Kurazume Laboratory: *HEARTS HumanitudE AR Training System*. (Accessed on 28/07/2023). (2023). URL: <https://www.microsoft.com/store/apps/9NFZ609S2JW2>.
- [11] Kurazume Laboratory: *HumanitudE AR Training System "HEARTS"*. (Accessed on 28/07/2023). (2023). URL: <https://github.com/Kurazume/HEARTS>.
- [12] 志和敏之 et al.: “対話ロボットの反応時間と反応遅延時における間投詞の効果”. 日本ロボット学会誌 27.1, pp. 87–95, (2009).