

認知症者のための音声対話システムにおける 個人向けヴァーチャルエージェントの生成

中谷 将大[†] 佐伯 幸郎[†] 中村 匡秀[†] 安田 清^{††}

[†] 神戸大学 〒657-8501 神戸市灘区六甲台町 1-1

^{††} 千葉ろうさい病院 〒290-0003 市原市辰巳台東 2-16

E-mail: †shota-n@ws.cs.kobe-u.ac.jp, ††sachio@carp.kobe-u.ac.jp, †††masa-n@cs.kobe-u.ac.jp,
††††fwkk5911@mb.infoweb.ne.jp

あらまし 我々の研究グループでは、認知症者へのパーソンセンタードケアを支援することを目的とした、ヴァーチャルエージェント (VA) との対話によるコミュニケーションシステムを開発している。現在のシステムは、VA として非実在のキャラクタをモデルに利用しているが、認知症者にとって馴染みのない外見が心理的障壁となり、結果として指示を受け入れにくい、会話に消極的になるなどケアの効果が限定されてしまう。そこで本稿では、実在する人物の顔写真から VA を動的に生成するシステムを提案する。提案システムでは、顔写真から取得した顔の特徴点に基づき 3D モデルを生成する。この 3D モデルに対しリアルタイムに伸縮・変形を行うことで、感情による表情の変化や、音声と同期した口唇の動きを再現でき、臨場感のある VA との対話を実現することができる。本システムにより、認知症者個人に馴染みのある人物の写真を基に心理的障壁の低いコミュニケーションエージェントを容易に生成でき、会話を通じた様々なケアを効果的に提供できる。

キーワード パーソンセンタードケア, ヴァーチャルエージェント, Web サービス

Generating Personalized Virtual Agent in Speech Dialogue System for People with Dementia

Shota NAKATANI[†], Sachio SAIKI[†], Masahide NAKAMURA[†], and Kiyoshi YASUDA^{††}

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

^{††} Chiba Rosai Hospital 2-16 Tatsumidai-Higashi, Ichihara, Chiba, 290-0003 Japan

E-mail: †shota-n@ws.cs.kobe-u.ac.jp, ††sachio@carp.kobe-u.ac.jp, †††masa-n@cs.kobe-u.ac.jp,
††††fwkk5911@mb.infoweb.ne.jp

Abstract Our research group has been studying a speech communication system with a virtual agent (VA), to support person-centered care (PCC) of people with dementia (PWD). The current system uses the 3D model based on an unreal character for the VA. Because the unfamiliar appearance is to be a mental obstacle to PWD, PWD hardly accept advice and, which causes a limitation in the care effects. In this paper, we develop a novel system that dynamically creates a VA based on a given facial image of real person. The proposed system constructs a three-dimensional model based on facial landmarks within the image. It then stretches and transforms some portions of the 3D model to generate facial expressions. From just a given picture, the proposed system easily generates a communication agent familiar with individual PWD. Hence, it can provide effective care through conversations with familiar VA.

Key words Person-centered care, Virtual agent, Web service

1. はじめに

近年、日本は超高齢化社会に直面している。内閣府発表の高齢社会白書によると、2025年には認知症患者数が700万人を超えと言われ、これは65歳以上の高齢者のうち5人に1人が認知症に罹患する推計となっている[1]。そのため認知症者のためのケアや支援が必要とされている。

認知症者に対するケアにおいては、パーソンセンタードケア(PCC)が理想的なケアの一つとされている。PCCは認知症者一人ひとりを尊重し、その人の立場に立って理解することによってその人個人に寄り添ったケアを提供することである。PCCは、画一化されたケアとは異なり、ケア対象者個人をよく観察する必要がある、ケア提供者に精神的にも肉体的にも大きな負担をかける。そのため、人手でPCCを行うには限界がある。

そのような背景から、我々の研究グループではIoTやクラウド技術を利用して、認知症者のためのPCC支援システムを開発している[2]。システムによって認知症者の自宅からセンサデータを収集、解析することで、認知症者の行動やコンテキストを推定する[3]。そして、それをもとに対話の内容を動的に生成し、ヴァーチャルエージェントを介して認知症者に話しかける[4]。ヴァーチャルエージェント(VA)はPC上で動作する、ひとの見かけをしたチャットボットプログラムであり、音声認識、音声合成技術を利用して、実際の人間と会話するように、認知症者とコミュニケーションを取ることができる。

現在のシステムはVAとして非実在の人工的なキャラクタのモデルを利用している。社会心理学において、人同士のコミュニケーションでは外見が相手に対する印象に大きな影響があるとされている[5]。これはVAにおいても、コミュニケーションを取るうえでは同様に外見が重要と考えられる。そして、現在のシステムにおいてVAの外見を認知症者に受け入れてもらえないことが原因で、馴染みのないエージェントからのケアやアドバイスに対して興味を持ってもらえず、耳を傾けられないことがある。その結果、PCCの効果を十分に発揮させられず限界を生じさせている。

そこで本研究では、PCC支援システムによるケアに認知症者がより興味を持つことで、ケアを効果的なものとするためのVAを生成するシステムを提案する。このVAを生成する手法には顔写真1枚を利用する。これによって認知症者個人に馴染みのある人物や、認知症者に対して説得力のある人物を容易にVAとして表示させることが可能になる。

このシステムにより顔写真から生成された認知症者に合わせた人物の3DモデルはVAとしてPCC支援システムに組み込まれる。提案システムを具体的に述べると次のとおりになる。システムに与えられた顔画像に顔認識アルゴリズムを適用させることで顔の特徴点を抽出する。その顔の特徴点に基づき生成した3Dモデルの一部を伸縮・変形させることで、動的に表情を作り出す。また、エージェントがユーザに話しかける内容の音声データに合わせて口唇の動きを同調させることで、3Dモデルが認知症者に話しかけているように見せることができる。

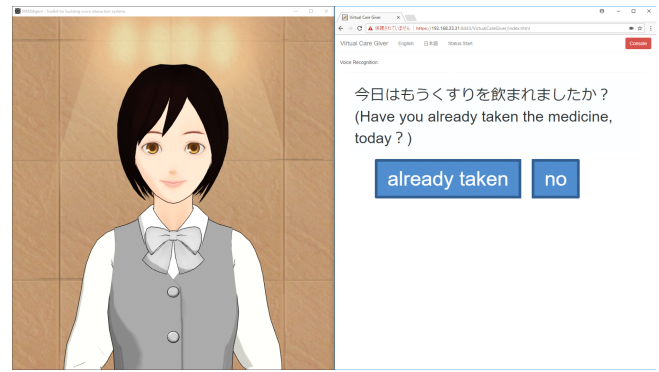


図1 VirtualCareGiver

この生成されたモデルをPCC支援システムに組み込むことで、既存のPCC支援システムによるケアやVAとの会話に対する抵抗感を、VAの外見によって軽減することが期待できる。

提案システムはたった1枚の写真から動的に認知症者個人の好みに合わせたコミュニケーションエージェントを生成することができ、例えば親戚や子供、親しい友人などのような、認知症者に馴染み深い人物との—仮想的だが、効果的な—会話も実現することができる。その結果、提案システムはより効果的なPCC支援システムの実現に寄与する。

2. 準備

2.1 ヴァーチャルエージェントを用いたパーソンセンタードケア支援システム

我々の研究グループは、ICTによって在宅認知症者を支援する方法を研究している。パーソンセンタードケア(PCC)の概念は、認知症者に対する理想的なケアとされている。PCCは個人の状況をよく観察し、理解したうえで、その個人に最適化されたケアを立案し、実行する。我々の現在の目的は、PC上で動作する人間のような見かけをしたチャットボットプログラムであるヴァーチャルエージェント(VA)の技術を用いて、認知症者に、その人個人に寄り添ったコミュニケーションを提供することである。音声認識、音声合成技術により、認知症者は実際の人間と会話するように、VAとコミュニケーションをとることができる。

我々の先行研究において、VAを利用したVirtualCareGiver(VCG)を開発している[2]。図1にVCGの画面を示す。VCGはIoTやスマートホーム、クラウドを統合するためにWebサービスで連携できるようにデザインされている。認知症者の好みに基づいて、VCGはVAを通して個人適応されたケアや会話を提供する。コミュニケーションケアをVCGが担うことで、人間の介護者はICTにはできないタスクに集中することができる。

この先行研究の実験において、VCGはあいさつなどの声かけや予定のリマインダ、服薬したかどうかの問いかけなどを含めて認知症者のコンテキストに基づいた有用なケアを実現できていることを示している。

しかしながら、VAの見た目によって生じている限界もまた

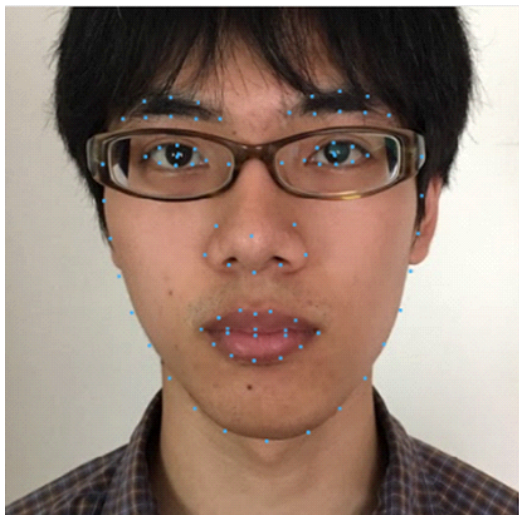


図 2 Face++によって得られた顔の特徴点

存在している。現在、VCG の VA は MMDAgent [6] を利用して実装されており、図 1 のように 3D のキャラクタモデルを表示している。その外見を認知症者が必ずしも受け入れるわけではなく、結果的に馴染みのないエージェントによるケアやアドバイスの興味を示さないことにつながる。これによって PCC への集中力や効果が限定的になってしまっている。

2.2 コグニティブコンピューティングにおける顔認識

コグニティブコンピューティングとは、人間の脳を模倣したシステムが学習によって答えを導き出せるような計算パラダイムである。従来のコンピューティングシステムでは扱うことの難しかった、言語や画像、音声のような非数値データを解析することのできる技術である。顔の認識や解析はコグニティブコンピューティングにおける主な技術であり、現在、顔認識のための Web API がいくつかの企業から公開されている。例えば、Microsoft が Azure Cloud Services において提供している Face API や Emotion API [7] が一例である。

大抵の顔認識技術は、目鼻口といった顔の特徴的なパーツを検出するために機械学習を用いている [8]~[10]。このような顔のパーツを示す点を顔の特徴点と呼ばれ、システムに画像を入力することで、画像から顔の特徴点が抽出される。図 2 は Face++ [11] による API によって顔画像から得られた特徴点を元の顔画像に重ねている。Face++ API は [9] に基づいて開発され、この API によって画像上の座標として 83 点の顔の特徴点を得ることができる。この得られた特徴点の位置を、画像を伸縮、変形させるように動かすことで、顔のパーツを動かすような操作を行うことができる。

3. 提案システム

2.1 において、現行の PCC 支援システムにおける VA の課題点を示した。そこで本研究では、現行の PCC 支援システムにおいて VA の外見が認知症者に馴染みのない外見となっていることによって、ケアが効率的に行えていない問題を解決することを目的とするシステムを提案する。

3.1 キーアイディア

認知症者に、より興味を持ってもらえるコミュニケーションケアを達成するキーアイデアとして、認知症者個人にあわせた VA を動的に生成することを目指す。そこで顔認識、解析技術を用いて、認知症者が VA として表示することを望んだ人物の写真を利用して 3D 顔モデルを生成する。そして生成したモデルを PCC 支援システムに組み込み、認知症者が選択した人物の VA と会話のできる状況を作り出す。その結果、従来の非実在のアニメーションキャラクターの外見をした VA に比べて、より質の高い PCC を実現できる。

提案システムによって置き換えられる既存システムの対象は、VA を実装するアプリケーションソフトウェアとそのアプリケーションの制御を行う Web API から構成されている。それを踏まえて、以降のサブセクションにおいては、提案システムの機能要件と、VA アプリケーションとそのアプリケーションの制御を行うためのサービス API の設計について説明する。また PCC 支援システムに組み込まれた際の動作の一例について説明する。

3.2 機能要件

既存システムから提案システムに次の 3 つの主要な機能を実装する必要がある。

機能 F1(VA の生成): 顔画像をもとにした VA の動的生成。

機能 F2(VA の発話): テキストに基づく VA の発話。

機能 F3(VA の表情変更): 表情ラベルに基づく VA の表情の変更。

F1 は認知症者個人向けの VA を作り出すという提案システムの基本となる要件を実現し、この機能によってユーザは自分だけの VA を生成することができる。F2 は VA に発話させる能力を実現し、この機能によって VA は認知症者の会話のパートナーとなることができる。F3 は VA に感情を表現する能力を実現し、この機能によって VA は人間のようにふるまうことができ、ノンバーバルなコミュニケーションを演出することができる。

3.3 サービス API とアプリケーションの設計

既存の VA のシステムとの置き換えを考慮すると、IoT やスマートホームと組み合わせることを考える必要がある。そのため、3.2 において言及した機能を実装した VA アプリケーションの制御を行うための API を Web サービスとしてデプロイする。これによって、PCC 支援システム内の外部アプリケーションが Web API を通して機能 F1, F2, F3 を利用することができる。そこで F1, F2, F3 のそれぞれの機能を利用するための 3 つの API を定義する。また、それぞれの API を利用したときの VA アプリケーションにおける動作についても説明する。

createVA(faceImageFile)

faceImageFile によって指定される画像ファイルから VA を生成する。システムは与えられた顔画像に顔認識アルゴリズムを適用することで顔の特徴点を抽出し、顔の 3D モ

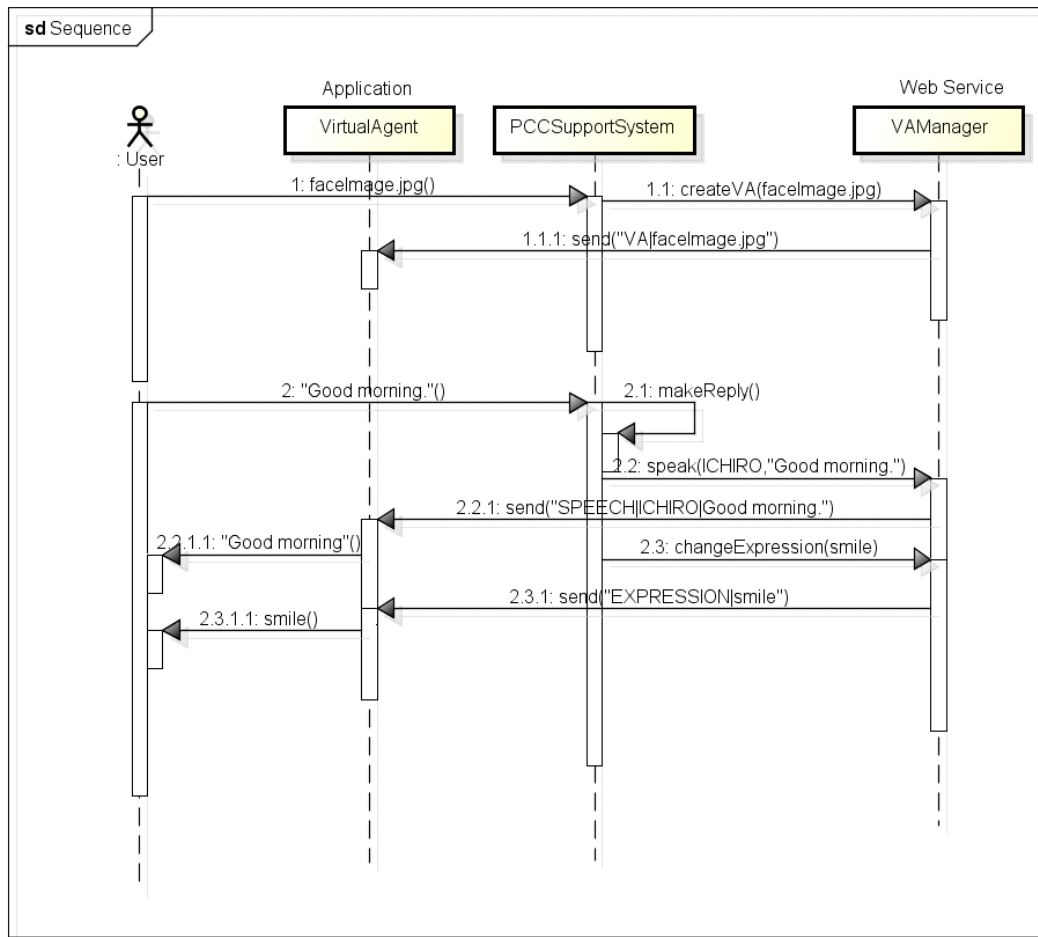


図 3 統合システムのシーケンス

デルを生成する。VA に対する操作指示が行われな
いときには、VA はまばたきや顔の揺れなどの
実際の人間らしい自然なふるまいを行う。

speak(type, text)

type によって指定された音声のタイプで、*text* によって指定されたテキストを、生成された VA が読み上げるように指示する。音声のタイプは、生成された VA の性別やユーザの使用言語に合わせて指定されるパラメータである。そして text-to-speech 技術を利用して、指定されたテキストから合成音声を生成する。生成された音声データを再生すると同時に、顔の 3D モデルの口唇を伸縮させることで、VA が音声に合わせて口唇の動きを同期させる。

changeExpression(expression)

expression によって指定された表情のラベルから、生成された VA の顔の表情を変更する。抽出された顔の特徴点に基づいて、システムが顔の 3D モデルの各所を伸縮・変形させ、顔の表情を作り出す。主な表情として、normal, happy, smile, angry, sad, surprised, fear がある。

3.4 PCC 支援システムへの組み込み

3.3 において示した Web サーバにデプロイする Web API 群を VAManager と呼ぶことにする。図 3 は PCC 支援システ

ムに提案システムを組み込んだ際の動作について示したシーケンス図である。以下に例として VA を生成し、ユーザが VA に対してあいさつするシナリオを示す。

ユーザはあらかじめ PCC 支援システムに画像ファイル *faceImage.jpg* を登録しておく。PCC 支援システムは *createVA(faceImage.jpg)* を実行し、VA を生成する。これによって、ユーザの PC の画面上に *faceImage.jpg* に写っている人物の見かけをした VA が表示される。

次に、ユーザが VA に対して“Good morning.”と話しかけたとする。PCC 支援システムは音声認識によりユーザの声を認識し、解析することで *Good morning.* という返事のテキストを生成する。そして、*speak(ICHIRO, “Good morning.”)* を実行し、*ICHIRO* という音声タイプの *Good morning.* という合成音声を、VA に発話するように指示する。このとき同時に、*changeExpression(smile)* を実行することで、VA にユーザに対して笑いかけるよう指示する。これによって、画面に表示された VA がユーザに対し、“笑顔”で“Good morning.”という返事を行うことになる。

4. プロトタイプ実装

提案システムに基づいて、システムのプロトタイプを実装した。実装に用いた技術は以下のとおりである。

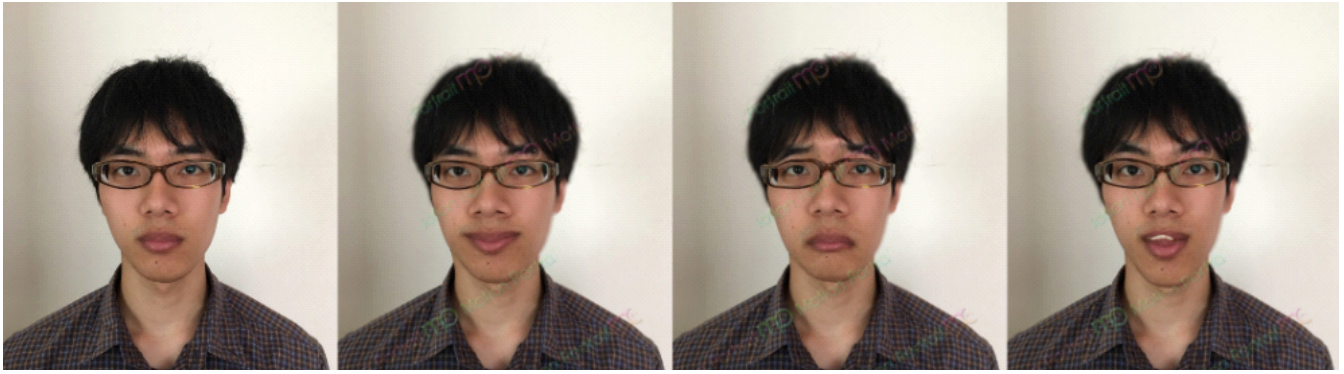


図 4 プロトタイプシステムによって生成されたヴァーチャルエージェントのイメージ

VAManager

- 開発言語 : Java
- Web サーバ : Apach Tomcat
- Web サービスフレームワーク : Apache Axis2

VirtualAgent(MPAgent)

- 開発言語 : C#
- MotionPortrait SDK [12]
- Bing Speech API [13]

VAManager は Java によって実装され、Tomcat Web サーバにデプロイされている Web API 群である。PCC 支援システムからのリクエストに基づいて、VA アプリケーションの制御を行う。

MPAgent は MotionPortrait SDK を利用し、C#によって実装されているアプリケーションソフトウェアである。MotionPortrait SDK は顔画像からの 3D 顔モデル生成、リップシンク、表情操作を行うライブラリである。このライブラリによって、3.2 における機能 F1, F2, F3 の効率的な実装を可能にしている。また音声合成のための text-to-speech 技術には、Microsoft Azure Service の Bing Speech API を利用している。Bing Speech API は、性別による声質や言語を定義した音声タイプとテキストを指定することで合成音声を生産することができる。この API によって、VA の性別やユーザの使用言語に合わせて異なる合成音声を生産し、VA に発話させることを可能にしている。機能 F1, F2, F3 の実行は、VAManager から受ける指示に従って行われている。

図 4 において、プロトタイプシステムにおいて表示される VA の例を示している。図 4 は左から順に、元の顔画像、“happy”の表情の VA、“sad”の表情の VA、喋っている VA を表している。

5. 実 験

VA の外見が与える PCC 支援システムによるケアの受容性に対する影響を、被験実験により評価する。

5.1 実験概要

デイサービス施設において、74 歳から 99 歳の要支援・要介護・認知症の罹患などを含めた高齢者 5 名を対象に提案システ

ムの評価実験を行う。

実験は、まず被験者に対し図 1 に示すキャラクタ VA を利用した VCG によるケアプログラム実施後、事前アンケートにより得られた「好きな歌手は誰か」という回答をもとに、提案システムによりその人にとって馴染み深い人物の写真を VA として生成する。被験者は VA と 3 分程度の簡単な会話の体験後、口頭によるインタビューを行い回答を得る。

5.2 実験結果

被験者 5 名中 4 名からインタビューによる回答を得られた。また残り 1 名は本システムに対し興味を示さず回答を得ることが出来なかった。以下にインタビューの質問内容と得られた回答を記す。異なる被験者から得られた同一の回答に関しては省略している。

- (1) 写真から生成された VA に対する印象（話し方、表情、音声と顔の一致度）はどうか。
 - 話し方、表情とかは気にならないが、声と顔が一致しないのは多少気になる。
 - 印象としては変な感じがしたが、話し方は不自然とは思わず、声と顔の不一致はそんなに気にならない。
 - 印象としては、びっくりした、面白い。話し方、表情はふつうである。
- (2) 写真さえあれば VA を作れるが、どんな人としやべってみたいか。
 - 話してみたい人は特にはいない。
 - いまは特に思い浮かばない。
- (3) 図 1 のような既存の VA と写真から作られたリアルな VA のどちらが話しやすいか。
 - 写真から作られた VA のほうがよい。既存の VA は話ができるが、写真から作られた VA は家族や家内と会えるという意味ではしばらく利用してみたいという気持ちはあるかもしれない。
 - 写真から作られた VA のほうがよい。
 - 昔は既存の VA のような外見の存在はいなくて、見慣れない存在なので、写真から作られた VA のほうが人間らしい見慣れた外見をしていてよい。

5.3 考 察

以下にインタビュー各項目に対する考察を述べる。

VA に対する印象

VA からの音声は、Bing Speech API により生成されているため、本来写真で利用している人物の肉声とは異なっている。また、合成音声であるため、イントネーションや発話タイミングなども人によっては違和感を感じる可能性がある。また、既存のキャラクターベースの VA と比較し写真を用いた VA について、外見・表情などをどう感じるかは個人によって異なると考えられる。この点について、実験結果より VA の見た目と音声が一致しないことについては、気になる、気にならない両方の意見があり、より違和感のない音声を利用するなど今後改善を行う必要がある。また、話し方、表情などについては全員が問題がないという意見だったため、本システムによる VA 生成は有効であるといえる。

VA として生成したい人物

特に挙げられる人物はいなかったが、他の質問に対する回答中に得られた家族や配偶者に会えるといった意見から、人によってはユーザの身近な人物を VA とすることで、一定の効果を得られることも考えられる。

VA への話しやすさ

既存の VA と提案の VA の比較において、やはり見慣れない外見であるキャラクタの VA より、写真から生成されたリアルな人間そのものの外見をした VA のほうが受け入れられることが分かった。回答を得られた被験者全員から既存の VA より提案システムによって写真から作られた VA のほうがよいとの意見を得られた。しかし、これについては被験者の回答から世代的な問題が存在すると思われる。既存の VA のような外見を見たことがなかったとの回答から、コンピュータの普及に伴い、3D のキャラクタモデルを一般的に目にする機会の増えている比較的若い世代からは異なる意見を得られる可能性がある。つまり、将来いまの若い世代が高齢者となるときには、VA の外見の受け入れ方が今の高齢者とは異なり、対照的に写真から生成された VA に対する違和感が強調されることも考えられ、提案システムは今の高齢者にのみ有効である可能性もある。

以上より、VA の外見と音声の一致については違和感を感じる場合もあるが、システムとしては写真を利用する手法の有効性が高いため、本手法はケアの現場において有効に活用できると期待される。

6. おわりに

本稿では、ヴァーチャルエージェントを用いた認知症者へのパーソンセンタードケアをより効果的にすることを目指したシステムの提案をした。顔認識技術を利用し、与えられた顔画像から 3D 顔モデルを生成し、PCC 支援システムに個人向け

ヴァーチャルエージェントとして組み込む。これによってユーザは好みの見かけをした、ユーザだけの VA を簡単に生成することができるため、より効果的な PCC を達成できると期待することができる。そして、提案システムに基づいてプロトタイプの実装と、VA の外見による影響についての実験を行った。

謝辞 この研究の一部は、科学技術研究費（基盤研究 B 16H02908, 15H02701, 基盤研究 A 17H00731, 萌芽研究 15K12020）、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] 内閣府, “平成 29 年度版高齢社会白書,” <http://www.cao.go.jp/>, June 2017.
- [2] S. Tokunaga, K. Tamamizu, S. Saiki, M. Nakamura, and K. Yasuda, “VirtualCareGiver: Personalized Smart Elderly Care,” *International Journal of Software Innovation (IJSI)*, vol.5, no.1, pp.30–43, Oct. 2016. DOI: 10.4018/IJSI.2017010103, <http://www.igi-global.com/journals/abstract-announcement/158780>.
- [3] 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,” *Digital Human Modeling 2017 (DHM 2017)*, no.LNCS 10287, pp.183–194, Springer International Publishing AG 2017, July 2017. Vancouver, Canada.
- [4] S. Sakakibara, S. Saiki, M. Nakamura, and K. Yasuda, “Generating Personalized Dialogue Towards Daily Counseling System for Home Dementia Care,” *Digital Human Modeling 2017 (DHM 2017)*, vol.LNCS 10287, pp.161–172, Springer International Publishing AG 2017, July 2017. Vancouver, Canada.
- [5] 池田謙一, 唐沢 穰, 工藤恵理子, 村本由紀子, 社会心理学 (New Liberal Arts Selection), 有斐閣, 2010.
- [6] “MMDAgent,” <http://www.mmdagent.jp>.
- [7] “Emotion API,” <https://azure.microsoft.com/en-us/services/cognitive-services/emotion/>.
- [8] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, “DeepFace: Closing the Gap to Human-Level Performance in Face Verification,” *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.1701–1708, June 2014.
- [9] H. Fan, Z. Cao, Y. Jiang, Q. Yin, and C. Doudou, “Learning Deep Face Representation,” *CoRR*, vol.abs/1403.2802, pp. •••••, 2014. <http://arxiv.org/abs/1403.2802>
- [10] Z. Zhang, P. Luo, C.C. Loy, and X. Tang, “Facial Landmark Detection by Deep Multi-task Learning,” *Computer Vision – ECCV 2014*, eds. by D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, pp.94–108, Springer International Publishing, Cham, 2014.
- [11] “Face++,” <https://www.faceplusplus.com>.
- [12] “MotionPortrait,” <https://www.motionportrait.com>.
- [13] “Bing Speech API,” <https://azure.microsoft.com/ja-jp/services/cognitive-services/speech/>.