Vol.2020-HCI-188 No.5
Vol.2020-EC-56 No.5
2020/6/1

日常ヘルスケアのための音声対話3Dエージェントの開発

能勢 隆[1,a)]　三田 昌輝[1]　遠藤 勇[1]　柳澤 佑介[1]　佐藤 勇人[1]

概要：本稿では，日常生活において利用者がヘルスケアを継続的に行うことができる音声対話型の 3D キャラクターエージェントシステム「HOME@IDOL」の開発について述べる．普段の生活において簡易的な問診を定期的に受けることは健康維持のために重要であるが，実際には時間的・金銭的コストの問題などで医療機関での継続的な受診は容易ではない．これに対し我々は独自の 3D キャラクターエージェントと音声で対話をしながら毎日気軽に簡単な問診を行うシステムを提案する．これに加え，起床から就寝までの日々の生活をサポートするホームアシスタント機能を実装することにより，利用者の日常に寄り添い利用継続を促す効果が期待できる．提案システムのコンセプトと概要を述べ，各機能の詳細と実装について説明する．

キーワード：日常ヘルスケア，簡易医療診断，音声対話システム，3D エージェント，ホームアイドル

Development of spoken dialogue 3D agent for daily healthcare

Abstract: This paper describes the development of a spoken dialogue 3D chracter agent system "HOME@IDOL" that provides continuous healthcare to users in daily life. Taking a simple doctor check regularly in the daily life is important for us to keep healthy. However, the continuous medical check is practically not easy due to the problem of time and financial costs for the consulting of a medical institute. For this issue, we propose a simple medical consulting system where users take a simple daily healthcare check by having a speech conversation with our original 3D character agent. In addition, we expect that the system can give the effect to encourage a user's continous use of the system by begin close to the user. We first describe the concept and the outline of propose system. Then we give detailed explanations of each function and implementation.

Keywords: Daily healthcare, simple medical diagnosis, spoken dialogue system, 3D agent, home idol

1. はじめに

少子高齢化の加速により，高齢者を中心として一人暮らしを強いられることで日々の健康管理に不安を感じる人々が増加している [1]．このような人々は家族や友人による健康サポートが期待できないため，医療機関において定期的な問診・診察を受けることが望ましいが，中には金銭的問題等で実際の健康管理が十分でない場合も多い．近年では自宅にいながらビデオ通話により問診や診察，処方を受けることができる遠隔診療 (オンライン診療) も登場しているが [2]，スマートフォンやタブレットなどの機器を使いこなす必要があり，高齢者に対しては機器操作についての問題も存在する．またこのサービスは医療従事者の数が限られているという課題もある．

日々のヘルスケア・健康管理を考えた場合，理想的には家族や介護者が対象者のそばに寄り添い，日常的に簡易的な問診などを行うことが望ましい．近年は計算機の処理能力の向上に加え，音声認識，テキスト音声合成，対話処理の進歩により人間と音声によりコミュニケーションやインタラクションを行うことができるロボットやコンピューター上の仮想エージェントが次々と登場している [3], [4]．例えば富士ソフトの会話ロボット「PALRO」は老人ホームやデイサービスなどの高齢者福祉施設での利用により認知症など厚生労働省の介護予防項目 (生活機能低下予防項目) に適用されている [5]．

一方で，このようなコミュニケーションロボットは一般的に導入時の金銭的コストが高く，介護施設などでの団体

[1] 東北大学
Tohoku University, Sendai, Miyagi 980–8579, Japan
[a)] takashi.nose.b7@tohoku.ac.jp

利用が前提であり，個人での利用はほとんど無いのが現状である．これに対し最近では Amazon Echo や Google Home に代表されるスマートスピーカーが登場し，家庭に設置することで利用者が音声を用いてスケジュールの確認や音楽再生，簡単な会話などを行うことができ，日常に浸透しつつある [6]．しかし，これらはロボットのように対面してではなく音声のみでコミュニケーションを行うため実体が意識しづらいだけでなく，人間とは異なり表情や仕草などのマルチモーダルな情報を用いた対話を行うことができない．

音声だけでなく PC のディスプレイ上の人間やキャラクターなどと会話ができるシステムについては，これまでに様々なシステムが提案されている [4]．これらの中でも日本語向けのシステムとして Galatea [7] や MMDAgent[8] などはフリーソフトウェアとして公開もされており，規則ベースで簡単な会話を行うことができるため研究用途も含めて利用されている．このようなシステムはロボットとは異なりパソコンとディスプレイがあればエージェントを画面に映し出し，音声によりエージェントとの会話を行うことができるため導入コストが非常に低いのが特徴である．MMDAgent についてはスマートフォンへの移植も進められている [9]．

これらの従来の取り組みに対し，我々は日々のヘルスケア機能を持ち，利用者の日常生活にさらに寄り添うことのできる音声対話エージェントシステム「HOME@IDOL」の開発を進めている．このシステムでは 3D モデルによるオリジナルの対話エージェントが家庭において利用者の起床から就寝まで寄り添い，日々の健康に関する簡単な問診も含めた肉体的・精神的な健康管理をサポートすることで，利用者がシステムを負担なく継続的に利用できる「さりげないヘルスケア」を実現する．具体的には，メインの機能である音声対話による簡易医療診断と利用者の日常生活をサポートするホームアシスタント機能の 2 つの機能を提供する．本稿では，最初に提案システムの開発コンセプトと概要を述べ，次に各機能の詳細と実装について説明する．

2. 提案システムのコンセプトと概要

本研究は文科省・JST 主体のセンター・オブ・イノベーション (COI) プログラムにおける「さりげないセンシングと日常人間ドックで実現する自助と共助の社会創生拠点」(COI 東北拠点 [*1]) の取り組みの一貫として行われている．この拠点では家庭における日常生活において自然な形で健康管理・ヘルスケアを行うことを目的としており，我々は音声対話エージェントの導入によって，利用者のヘルスケア意欲の向上などの行動変容を狙っている．従来のスマートスピーカーではどうしても人としての実体化 (embody) が無く，人と向き合う・コミュニケーションを取る，という感覚が希薄となる．これに対して人型あるいはキャラクターロボットの導入が望ましいが，高価であること，あるいは表情・ジェスチャなどにおける制約から代替として人間やキャラクターのバーチャルエージェントが現状では有効であると考えられる．

バーチャルエージェントは 3D モデルを導入することによって表情やジェスチャなどを柔軟に変更することができる．さらにディスプレイ上にエージェントだけでなく様々なテキストあるいは画像情報を表示することができ，総合的な情報案内やインタラクティブ・コミュニケーションを容易に実現できる利点がある．我々はバーチャルエージェントをデザインするにあたり，以下のコンセプトを設定した．

(1) 特定の人間ではなく親しみやすいバーチャルキャラクターを新たにデザインする [*2]．
(2) 利用者が日常生活において継続的にヘルスケアを行うために役立つ簡易的な問診機能を有する
(3) 高精度な音声認識により利用者の言語情報を正しく把握する
(4) 柔軟なテキスト音声合成により豊かな対話を目指す
(5) ヘルスケアだけでなく利用者の家庭での日常に寄り添うようなアシスタント機能・簡単なエンターテイメント機能を有する
(6) 開発のベースとなる環境としてはゲームや VR 用のエンジンとして広く用いられている Unity を利用し開発コストの削減を図る

(1) については，Galatea など実際のキャラクターの顔画像を用いる例もあるが 3D 化や自然性の点で問題がある．一方，特に日本ではアニメに代表されるようにオリジナルのキャラクター文化が存在し日常に浸透しているという現状がある．これらの理由から本研究では新たにキャラクターをデザインし [*3]，この 3D モデルを作成する．詳細については 3.2.3 節で述べる．(2) については本システムのメインとなる機能であり，特定の病気であるかどうかを判別する機能と利用者が質問に答えることで最も可能性の高い病名を回答する機能に別れる．このように提案システムは従来の情報案内や Q&A 中心の音声対話システムとは異なる特徴をもつ．詳細については 3.3 節で述べる．(3) については，従来の MMDAgent 等では独自の音声認識エンジンが用いられていたが，認識精度の面で問題があったため，大規模データベースと最新の深層学習技術による高精度音声認識 API を用いることで，利用者の音声入力を正確に把握する (3.2.1 節)．

[*1] http://www.coi.tohoku.ac.jp/

[*2] このようなキャラクターの代表としては初音ミク [10] が知られている．

[*3] デザインは東北大学人工知能エレクトロニクス卓越大学院プログラム (http://www.aie.tohoku.ac.jp/) と連携して作成した．

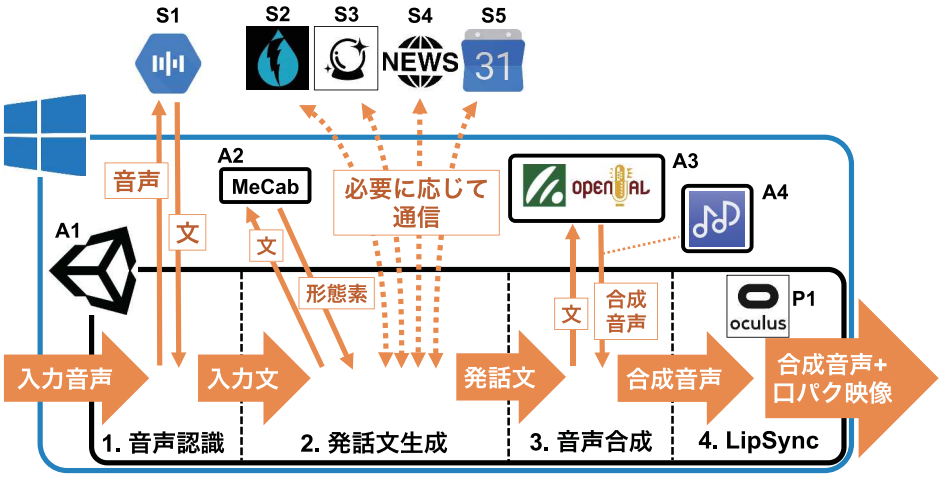

図 1 システムの大まなか流れと構成

(4) については従来，多くの音声対話システムにおいては既存の読み上げ用音声合成エンジンが用いられており対話が単調になり日常生活における会話としては不自然であった．本システムでは専用の対話用音声データベースを用い統計的テキスト音声合成によって，より自然で利用者にとって違和感の少ない音声応答を実現する (3.2.2 節)．(5) についてはヘルスケア機能だけではどうしても毎日継続して利用する動機が希薄であるため，エージェント自身が利用者にとってより身近な存在として認知されるようにするため，日常生活をサポートする機能や占いなどのエンターテイメント機能を提供する．(6) については，MMDAgent などの全体を独自のコードにより開発するのではなく，代表的なマルチプラットフォーム 3D ゲームエンジンである Unity を用いることで，アセットと呼ばれる様々な機能や 3D 描画機能を容易に用いることで，全体の開発コストの削減を実現することができる．

3. 各機能の詳細と実装

本節では提案システムである HOME@IDOL のシステムの全体構成および各機能の詳細と実装について述べる．具体的には，まず全体構成について述べた後，音声認識，簡易医療診断，スケジュール・時間管理，ニュース検索・表示，テキスト音声合成，3D キャラクターモデル，対話空間デザインについて説明を行う．

3.1 システムの全体構成

今回の対話システムを構築するにあたり，汎用 3D ゲームエンジンである Unity を用い，エージェントの AI，およびユーザインターフェース (UI) を製作した．エージェントの AI システム全体の大まかな流れと構成を図 1 に示す．図の S1–S5 は Web サーバ，A1–A4 はアプリケーション，P1 はプラグインを示す．本研究で開発のベースとなる Unity アプリケーション部分 (A1) は，利用者がエージェントに向かって話しかけてからエージェントが応答するまで，1. 音声認識フェーズ，2. 発話文生成，3. 音声合成，4. LipSync の 4 つのフェーズを経る．

まず，利用者の入力音声は音声認識フェーズにより処理される．利用者が PC のマイクへ発声すると，アプリケーションはストリーミング音声認識が可能な Google Speech-to-Text API[*4] を用いて入力音声を HTTP でサーバ (S1) へ逐次送信し，認識された文章を受信する．ここで，アプリケーションの状態によっては，利用者の入力文に依らず自発的に発話文を生成する必要があるため，このフェーズはスキップされる場合がある．発話文生成フェーズでは，自発的に，もしくは入力文を基にエージェントの発話文を生成する．入力文がある場合は必要に応じて形態素解析器 MeCab[11](A2) を用いてそれを形態素解析し，結果に応じて各種機能の処理を実行する．ここで，天気予報・占い・Google カレンダーの参照を行う際は，それぞれ Dark Sky API[*5]，Web ad Fortune 無料版 API[*6]，Google Calender API[*7] を利用する．また，ニュースサイトからスクレイピングによりニュースを取得する．これらは各種サーバ (S2–S5) と HTTP で通信することにより実現する．

音声合成フェーズでは，生成された発話文を基に音声合成とその逐次再生を行う．発話文はあらかじめ別のプロセスとして起動しておいた Rapid Text-To-Speech[*8] (A3) へプロセス間通信により送信される．通常のテキスト音声合成エンジンを用いた場合，与えられたテキストを全て波形に変換した後でのみ音声の再生が可能であるが，この場合利用者にとっては不自然な長い待ち時間が生まれる．この問題を抑制するため，Rapid Text-To-Speech では Open JTalk[12] を用いて音声合成すると同時に，OpenAL[*9] を用いて合

*4 https://cloud.google.com/speech-to-text/
*5 https://darksky.net/dev
*6 http://jugemkey.jp/api/waf/api_free.php
*7 https://developers.google.com/calendar
*8 https://github.com/ll0s0ll/Rapid_Text_To_Speech
*9 https://www.openal.org/

成音声の逐次再生を行う．逐次再生される合成音声は仮想オーディオドライバ (A4, Yamaha NETDUETTO[*10]) を介して Unity アプリケーションで取得される．最後の LipSync フェーズでは，合成音声の波形を基にエージェントの LipSync を行う．これは Oculus Lipsync Unity プラグイン [*11](P1) により実現される．

3.2 システム基幹機能

3.2.1 音声入力

音声認識には Google Speech-to-Text を利用した．これはニューラルネットワークを用いて機械学習された音声認識システムである．クラウド上で処理するためクライアント側のアプリケーションの処理負荷が少ないことが特徴である．同様の音声認識システムはオフラインで実行できるものとオンラインで実行されるものの二種類に大別される．オフラインで実行されるものとしては Julius[13][*12] が挙げられる．これはネットワークの整備されていない環境で稼働できるため，可搬性が高いという利点がある．しかし，オンラインで実行されるシステムに比べると認識精度が低い．

オンラインで実行されるものには Google Speech-to-Text の他に Watson Speech to Text[*13] や Azure Speech Recognition[*14] がある．これらは全て機械学習モデルがクラウドサーバ上で稼働しており，音声ファイルを送信することで認識結果がテキスト形式で返る仕組みとなっている．オンライン環境でしか使えないという問題はあるが，高い認識精度を得ることができ，API 形式で配布されているため利便性が高い．これらのシステムを比較した結果，このアプリケーションでは Google Speech-to-Text を採用した．理由としては，他にもネットワーク接続が必要な機能が存在するため，オンラインで実行されるもので良いということや，Unity で利用する際の利便性，各システムの予備的な性能比較の結果最も良い精度で認識ができたという点が挙げられる．

3.2.2 音声出力

本システムではテキスト音声合成に隠れマルコフモデル (HMM) に基づく合成方式 [14] を用いた．近年では研究分野を中心に深層学習に基づく音声合成 [15] が主流となっているが，本システムは Unity による 3D 描画に GPU を用いていることもあり，計算コストの低い HMM 音声合成を用いる．MMDAgent や Galatea Talk など，従来の音声対話システムでは読み上げ調の音声を用いて音響モデルの学習を行うことが多く，日常会話に適用した場合には単調な印象を与える傾向があった．このため我々は文献 [16] において構築した話し言葉音声コーパスを学習に用いた．このコーパスは，読み上げ調の ATR 音素バランス 503 文 [17] と，そのうちの 50 文の文末を話し言葉調と疑問文にそれぞれ変更した文セット，および独自に用意した対話文 92 文の計 695 文をプロの女性話者 1 名が読み上げた音声である．

HMM の学習には文献 [14] と同様に 5 状態の left-to-right HMM を用い，音韻・韻律コンテキストに対する決定木クラスタリングにより状態共有を行った．音声合成エンジンには Open JTalk を用い，これを利用した Rapid Text-To-Speech，および OpenAL を用いることで逐次合成・再生を行った．なお，3.4.3 節の音楽再生のときなどに形態素解析における読み誤りによって英単語などの発音が適切に行われない問題がある．これに対しては予め発音辞書を登録する形で対応した．

図 2 3D キャラクターモデル「Saki」

3.2.3 3D キャラクターモデル

3D キャラクターモデルは事前にデザインした女性キャラクター「Saki」を 3D モデル化した (図 2)．このキャラクターのコンセプトとしては清潔感，未来感，聡明性，人工知能などをイメージし，イラストレータの朱シオ氏 [*15] と相談の上デザインを決定した．図 3 に三面図を示す．3D モデルの形式は FBX であり，Unity に Humanoid モデルとしてインポートされる．利用者にとって親しみやすいキャラクターとなるよう，なるべく人間らしい自然な動きをすることを目標としている．そのために，アバターの服装やアクセサリーに対し物理シミュレーションを適用することで自然に揺れるよう見せている．またアニメーションといった音声以外のリアクションを取ることも親しみやすさにとって重要な要素である [4]．そのため，利用者の行動に対しジェスチャー等のアニメーションを行う機能を開発

[*10] https://www.netduetto.net/
[*11] https://developer.oculus.com/downloads/package/oculus-lipsync-unity
[*12] https://julius.osdn.jp/
[*13] https://www.ibm.com/watson/jp-ja/developercloud/speech-to-text.html
[*14] https://azure.microsoft.com/en-us/services/cognitive-services/speech-to-text/
[*15] https://www.pixiv.net/users/341747

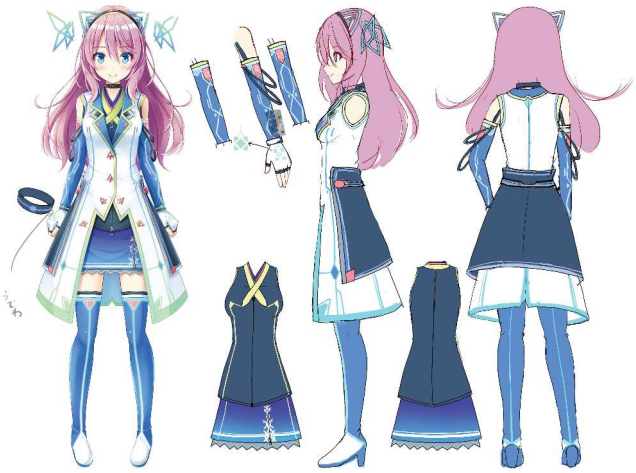

図 3 Saki の三面図

しており，現段階では利用者の寝る前や外出前に手を振る機能が実装されている．他にもキャラクターと利用者が目を合わせて会話ができるよう，利用者の顔をキャラクターが追う機能を作成した．モニタの近くに設置された Web カメラの入力画像に対し顔認識を行い，画像内に写る人物の位置を Unity の 3D 空間内でシミュレートすることで実現させている．画像からの顔認識技術には OpenCV[*16] を利用した．

3.2.4 対話空間デザイン

エージェントの背景として，図 4 や図 5 のような空間をデザインした．エージェントが気軽に受け入れられるような未来をイメージしており，医療診断を行うことを踏まえ，清潔感・透明感のあるデザインを意識した．図 4 は現在開発中の画面であるが，エージェントの頭上にある白板部分に天気予報やニュースといった情報を掲示することを考えている．これは利用者が頻繁に知りたいと思われる情報を常に掲示するためで，これにより従来のスマートスピーカを利用する際の煩雑なやり取りを省略することができる．これらの背景デザイン及び UI デザインは，2019 年度東北大学オープンキャンパスにおいて本システムを展示した際に利用者から寄せられた意見を参考に，フォントや配置を調整するといったブラッシュアップを行ったものである．背景を作成するために HLSL(High Level Shading Language) で記述されるシェーダー及び C#スクリプトを独自に作成して用いている．これによって背景や UI をアニメーションさせることが可能となり，飽きの来ない空間を作成できる．

3.3 簡易医療診断機能

問診による簡易的な医療診断を行うことのできるシステムとして，音声入力またはタッチデバイスによる項目の選択が可能な問診システムを構築した．診断には機械学習に

[*16] https://opencv.org/

図 4 対話空間デザイン

図 5 簡易医療診断の例

よる症状の分類や医療情報データベース [*17] を活用することを検討したが，アプリケーション全体の動作を優先したため，今回は同サイトにおける病気の検索ページ [*18] を参考に XML で記述される簡易的な開発用データベースを作成した．開発したデータベースを利用することで二種類の診断が可能である．図 5 に診断中の画面の例を示す．

一つ目の機能は特定の病気であるかを判別するものである．各病気ごとに作成された質問リストを基に利用者に「はい」または「いいえ」で答えられる質問をしていき，その応答から利用者がその病気にり患しているかを判断する．N 個の質問の i 番目の質問に対して応答が「はい」の時のスコアを s_i^y，「いいえ」の時のスコアを s_i^n とおくと，り患しているかの総合スコア S は下記の式によって表される．

$$S = \sum_{i=0}^{N} s_i,\ s_i \in s_i^y, s_i^n \tag{1}$$

このスコア S がその病気の閾値 S_{thresh} 以上の場合にその病気であると診断される．現在は風邪・インフルエンザ・ノロウイルスの診断が可能であり，質問とスコアのデータは XML 形式にて図 6 のようなフォーマットにより記述されている．

二つ目の診断は，利用者が感じる体の不調を回答していくことで最も可能性の高い病気の名前を示すものである．開発用データベースを基にノードの幅が 2 から 6，深さが

[*17] https://medley.life/pages/database/
[*18] https://medley.life/diseases/

```
<?xml version="1.0" encoding="utf-8"?>
<DiagnosisJudge>
  <Sick name="風邪">
    <Keywords>
      <Name>風邪</Name>
      <Name>かぜ</Name>
    </Keywords>
    <Questions score ="6">
      <Sentence yes ="2" no="0">熱はありますか？</Sentence>
      <Sentence yes ="2" no="0">鼻水は出ますか？</Sentence>
      <Sentence yes ="2" no="0">咳や痰は出ますか？</Sentence>
      <Sentence yes ="2" no="0">頭痛はありますか？</Sentence>
      <Sentence yes ="2" no="0">倦怠感はありますか？</Sentence>
    </Questions>
  </Sick>
  <Sick name="インフルエンザ">
    <Keywords>
      <Name>インフルエンザ</Name>
    </Keywords>
    <Questions score="5">
      <Sentence yes ="3" no="-5">高熱(38度以上)が出ていますか?</Sentence>
      <Sentence yes ="2" no="0">関節痛はありますか?</Sentence>
      <Sentence yes ="2" no="-1">筋肉痛はありますか?</Sentence>
      <Sentence yes ="2" no="-1">倦怠感はありますか?</Sentence>
    </Questions>
  </Sick>
  <Sick name="ノロウイルス症候群">
    <Keywords>
      <Name>ノロウイルス症候群</Name>
      <Name>ノロウイルス</Name>
      <Name>ノロ</Name>
    </Keywords>
    <Questions score="5">
      <Sentence yes ="2" no="-2">熱が出ていますか?</Sentence>
      <Sentence yes ="3" no="-3">腹痛はありますか?</Sentence>
      <Sentence yes ="2" no="-5">吐き気はありますか?</Sentence>
      <Sentence yes ="3" no="-5">下痢の症状は出ていますか?</Sentence>
    </Questions>
  </Sick>
</DiagnosisJudge>
```

図 6　簡易医療診断に用いる問診データベースの例

4 から 5 程度の単純な多分木を作成しすることで実現している．現状の実装は全ての病気の診断に使えるものではないが，先述した機械学習モデルを用いた手法に置き換えられるような実装をしており，柔軟な拡張が可能である．

3.4　ホームアシスタント機能

日常的に利用してもらうための機能としてニュースの検索やスケジュール・時間管理などの機能を実装した．また，その他の機能として音楽再生，しりとりなどのエンターテイメント機能も追加した．

3.4.1　ニュースの検索・表示

ニュース機能には特定のニュースの表示，天気予報，占い機能がある．また，より具体的なニュースの検索機能の一例として，サッカーの試合結果についての検索・対話機能も実装した．ニュース情報はニュースサイトからのスクレイピングにより取得した．利用者の入力文に「ニュース」というキーワードが含まれていた場合は MeCab による形態素解析結果から，別の名詞が含まれていないか判別する．ここで，他の名詞が含まれていた場合は利用者がその言葉を検索しようとしていると類推し，ニュースサイトからそれに関する記事のタイトルと URL を複数件取得する．キーワード以外の名詞が含まれていない場合はトップページから主要ニュースの取得を行い，タイトルの一覧を図 7 のように表示する．タイトルの一覧から特定のニュースを選択するにはクリックと音声の二種類の指定方法がある．タイトルが選択されると URL から再度スクレイピングを行い，画像と本文を取得する．取得した画像は図 8 のように表示をし，本文の最初の一文について語尾を口語調

図 7　ニュースの一覧表示

図 8　ニュース画像の表示

に変換し読み上げを行う．

天気予報の取得に関しては Dark Sky API を利用した．利用者が「天気」「気温」「温度」などのキーワードを用いて話しかけることにより，昨日，今日，明日，今週の天気を知ることが可能であり，エージェントは天気に応じて簡単な一言を発する．これに加えて，占い情報は Web ad Fortune API により取得した．利用者が「占い」「運勢」「占って」のいずれかのキーワードを入力したとき，エージェントは利用者の星座を尋ね，これを元にその日の総合運 (5 段階評価) や運勢などを取得する．続いて総合運に応じた一言 (例：いいですね！) と運勢を読み上げ，それ以外の情報 (金運・仕事運・ラッキーアイテムなど) を図 9 のように表示する．ニュース検索機能のより具体的な例としてサッカーの試合結果を教えてくれる機能を実装した．J1, J2, J3 のサッカーの試合結果をニュースサイトをスクレイピングし取得する．利用者の入力から J1, J2, J3 という区分や日時，チーム名，場所を抜き出し，一致している要素が多い試合の結果を読み上げる．

図 9 占い結果の表示

図 10 アラーム機能

図 11 タイマー機能

3.4.2 スケジュール・時間管理

スケジュール管理機能として，予定の確認や追加が可能なカレンダー機能とアラーム及びタイマー機能を追加した．カレンダー機能は Google Calender を利用している．Google Calender API[*19] を利用し，直近 24 時間の予定取得を行うほか新規の予定追加を行うことが可能である．利用者は「予定」「スケジュール」といったキーワードと「登録」「追加」といったキーワードを同時に用いることにより予定の追加が可能である（例：「明日の 16 時に会議の予定を登録して」）．その際，利用者の入力文に時間及びスケジュールの概要（「会議」など）が存在しない場合は不足している要素をエージェントが聞き返す（例：「開始時間を教えてください」や「場所を教えてください」など）．

利用者の入力文に「アラーム」という単語が存在した場合（例：「11 時 30 分にアラームを追加して」），入力文から時刻を取得してアラームを登録する．指定した時刻になると利用者への通知を行う．アラームの追加の段階で利用者の入力に特定の単語が存在した場合（例：「11 時 30 分に寝るからアラームを追加して」），エージェントはその内容に応じた通知を行う（例：「そろそろ寝ましょう」）．目覚まし時計として利用者がアラームを登録した場合（例：利用者が「6 時に起こして」などと発言する）は，利用者が「止めて」等の発言を行うまでスヌーズさせる．また「アラーム」と「見せて」「知りたい」といったキーワードを同時に用いることで (例：アラームの一覧を見せて) 登録されているアラームの一覧を表示する．その際，指定したアラームを削除することができる (図 10)．また，利用者が「タイマー」というキーワードを含む発言を行った際は発話に含まれる時刻をタイマーとして設定する．タイマー機能では指定した時刻になるまで画面に残り時間を表示し，カウントダウンを行う (図 11)．

[*19] https://developers.google.com/calendar

3.4.3 エンターテイメント機能

利用者が楽しめる機能として雑談，ミニゲームおよび音楽の再生の各機能を実装した．雑談は用例応答辞書（利用者の入力候補文とそれに対応する応答文）に基づいて行われる．適切な応答を行うために，利用者の発言と辞書中の入力候補文との正規化レーベンシュタイン距離を算出する．算出された距離に対し，あらかじめ設定した閾値 [*20] を下回る候補文の中で最も距離が近い文に対する応答文をエージェントの発言とする．閾値を超える文が存在しない場合エージェントは応答しない．

音楽の再生機能では，予め用意した曲の中から利用者の入力に応じて適する曲を選択でき，曲の再生や変更，停止が可能である．利用者の発言に「音楽」「ミュージック」などのキーワードが含まれていた場合，エージェントは利用者に希望する曲のジャンルを尋ねる．その後利用者の回答に含まれるジャンルに最も該当する曲を再生する．「ランダム」「自由」といったキーワードが含まれる場合，曲をランダムに再生する．

[*20] 本研究では予備実験により 0.6 とした．

また，ミニゲーム機能として，エージェントとしりとりができる機能を実装した．エージェントが持つ語彙として，読み仮名を含む単語辞書をあらかじめ用意しておく．利用者がしりとりを行うことを提案する（例：利用者が「しりとりをやろう」と発言する）と，エージェントはしりとりモードに移行する．エージェントはそれ以降に利用者の発した文章を形態素解析によって読み仮名に変換し，その最後の文字を先頭に持つ単語をランダムに返す．ゲームが進行し単語辞書に返せる単語が存在しない場合，または利用者の発言の末尾が「ん」であるとき，続行が不可能と判断されしりとりは終了する．

4. おわりに

本稿では現在開発中の日常ヘルスケアを促進する音声対話 3D エージェントシステム「HOME@IDOL」について述べた．従来の情報案内やチャット機能のみをもつ対話システムとは異なり，ヘルスケアを中心として日常に寄り添うことを目指し，二種類の簡易的な問診機能を実装した．利用者は日々の生活の中でこのような問診により人間との対話に近い自然な形で自分の健康状態を確認することができる．また，利用者の日常をサポートし寄り添うことができるホームアシスタント機能により，単なる医療目的だけではなく，日常に溶け込んだ形でエージェントと接することが可能となる．システム全体は Unity をベースとして構築されており，Unity の様々な機能やアセットを利用することで，コンテンツやモーションの多様化・高度化を低い作業コストで行うことが可能である．今後は，実際に利用者の 1 日を想定したシナリオを作成し，それに沿った簡単な動作検証を行う．

謝辞　本研究開発は JST COI（課題番号 JPMJCE1303）の助成を得て行ったものである．

参考文献

[1] 白波瀬佐和子：少子高齢社会のみえない格差 (2005).
[2] 飯田真美：禁煙支援における遠隔診療―情報通信機器を用いたオンライン診療，医学のあゆみ， Vol. 265, No. 10, pp. 885–888 (2018).
[3] 小林敦他：ソーシャル・ロボットのアーキテクチャ試論: 人間とロボットが共生する未来社会の実現に向けて，21 世紀社会デザイン研究: Rikkyo journal of social design studies， Vol. 13, pp. 55–69 (2014).
[4] André, E. and Pelachaud, C.: Interacting with embodied conversational agents, *Speech technology*, pp. 123–149 (2010).
[5] 二宮恒樹：コミュニケーションロボット 「PALRO (パルロ)」の紹介とさがみロボット産業特区における取り組み，日本ロボット学会誌， Vol. 33, No. 8, pp. 607–610 (2015).
[6] Kepuska, V. and Bohouta, G.: Next-generation of virtual personal assistants (microsoft cortana, apple siri, amazon alexa and google home), *Proc. CCWC*, pp. 99–103 (2018).
[7] Katsurada, K., Lee, A., Kawahara, T., Yotsukura, T., Morishima, S., Nishimoto, T., Yamashita, Y. and Nitta, T.: Development of a toolkit for spoken dialog systems with an anthropomorphic agent: Galatea, *Proc. APSIPA ASC*, pp. 148–153 (2009).
[8] Lee, A., Oura, K. and Tokuda, K.: MMDAgent—A fully open-source toolkit for voice interaction systems, *Proc. ICASSP*, pp. 8382–8385 (2013).
[9] Yamamoto, D., Oura, K., Nishimura, R., Uchiya, T., Lee, A., Takumi, I. and Tokuda, K.: Voice interaction system with 3D-CG virtual agent for stand-alone smartphones, *Proc. the second international conference on Human-agent interaction (HAI)*, pp. 323–330 (2014).
[10] 剣持秀紀：歌声合成技術の動向:「初音ミク」 を支える技術 (＜小特集＞音声合成に関する研究の動向)，日本音響学会誌， Vol. 67, No. 1, pp. 46–50 (2010).
[11] Kudo, T.: Mecab: Yet another part-of-speech and morphological analyzer, *http://mecab. sourceforge. net/* (2005).
[12] 大浦圭一郎，酒向慎司，徳田恵一：日本語テキスト音声合成システム Open JTalk，日本音響学会春季研究発表会講演論文集，pp. 343–344 (2010).
[13] Lee, A. and Kawahara, T.: Recent development of open-source speech recognition engine julius, *Proc. APSIPA ASC*, pp. 131–137 (2009).
[14] Yoshimura, T., Tokuda, K., Masuko, T., Kobayashi, T. and Kitamura, T.: Simultaneous modeling of spectrum, pitch and duration in HMM-based speech synthesis, *Proc. Eurospeech*, pp. 2347–2350 (1999).
[15] Ling, Z.-H., Kang, S.-Y., Zen, H., Senior, A., Schuster, M., Qian, X.-J., Meng, H. M. and Deng, L.: Deep learning for acoustic modeling in parametric speech generation: A systematic review of existing techniques and future trends, *IEEE Signal Processing Magazine*, Vol. 32, No. 3, pp. 35–52 (2015).
[16] 山田修平，能勢隆，伊藤彰則：多様な対話音声合成のための話し言葉音声コーパスの構築と評価，情報処理学会研究報告，Vol. 2015-MUS-107, No. 72, pp. 1–6 (2015).
[17] Kurematsu, A., Takeda, K., Sagisaka, Y., Katagiri, S., Kuwabara, H. and Shikano, K.: ATR Japanese speech database as a tool of speech recognition and synthesis, *Speech Communication*, Vol. 9, No. 4, pp. 357–363 (1990).