音声対話可能な2Dキャラクターを用いた仮想共食システム

鈴木瑞帆[1,a)]　塚田浩二[1]

概要：近年，一人で食事を行う「孤食」傾向が増加しており，コミュニケーション能力やストレス，食習慣等の問題に繋がる社会的課題として注目されている．こうした課題を軽減するために，本研究では音声対話可能な 2D キャラクターとの仮想共食システムを提案する．食事中のユーザの行動を認識することで，状況に応じて 2D キャラクターが反応する．さらにユーザの発話を音声認識し，AI を用いた文章生成や音声合成と組み合わせることで，キャラクターと様々な会話を行うことができる．このようにしてユーザに仮想的な共食環境を提供することで，孤食による問題の軽減を目指す．

1. 背景と目的

近年，孤食（一人で食事をすること）を行う頻度が高い人が増加してきており，社会的に問題視されている．例えば，コミュニケーション能力の減少 [1] や，ストレスの蓄積 [2]，栄養の偏りや早食いによる食習慣の悪化 [3] 等が懸念されている．

平成 23 年度から 27 年度まで施行された第 2 次食育推進基本計画では，食育の推進目標として「朝食又は夕食を家族と一緒に食べる『共食』の増加」が掲げられた [4]．さらに平成 28 年度から 32 年度（令和 2 年度）まで施行された第 3 次計画では新たに「地域等で共食したいと思う人が共食する割合を増やす」が掲げられた [5]．このように，日本では政府方針としても，共食の重要性が打ち出されている．

本研究では「食事中の会話によるコミュニケーション」に着目し，音声対話が可能な 2D キャラクターを用いて共食環境を提供するシステムを提案する．ユーザの移動や，食品を口に運ぶ動作等の食事中の行動を検出することで，状況に応じた反応を 2D キャラクターが行う．さらにユーザの発話を音声認識して，AI を用いた文章生成や音声合成と組み合わせることで，ユーザはキャラクターと様々な会話を楽しむことができる．これらの機能を用いてユーザに仮想的な共食環境を提供することで，孤食により引き起こされる問題の軽減を目指す．

1　公立はこだて未来大学
a)　b1020242@fun.ac.jp

2. 関連研究

Nakamura ら [6] は，センサを組み込んだ箸を使用して，食材の色や大きさ，さらにユーザの食べる速度等を検出し，デジタルキャンバスの色に反映させることで，食事の質を向上させるシステムを提案した．評価実験を行い，ユーザにゆっくり食事をすることを促すと同時に，よりバランスのとれた食生活の動機づけができたという結果を示した．

Kadomura ら [7] は，ユーザが食品を摂取する際に様々な音を奏でるフォーク型デバイスを提案した．聴覚的なフィードバックによって食事時の行動を促すことで，食事をより遊び心のあるものにし，日々の食習慣を改善させる．フォークの持ち手と先端が別の電極になっており，食べ物を食べる瞬間に，食材と人体を介して微弱な電流が流れる．これにより，フォークに差した食べ物の種類や食べ方（抵抗値の変化）に応じて音を奏でることができる．

さらに Kadomura ら [8] は，電極や様々なセンサを搭載して，食事中の行動や食材の種類を認識し，Bluetooth 経由でスマートフォンに送信するフォーク型デバイスを提案した．スマートフォン上の食育ゲーム「腹ペコパンダ」等と連動することで，子供の食育を支援することを目指す．母子を被験者として，偏食や食事中に気が散りやすいといった子供の食行動を軽減できる可能性を示した．

共食状態を疑似的に生成・再現する研究事例として，中田ら [9] は，対面にある鏡を見ながら食事を行うことで，共食環境を疑似的に生成する研究を行った．鏡に映っている自分自身を，同じ空間内・同じペースで食事を行っている仮想的な共食の相手として食事を行う．鏡のあり（共食状

態）・無し（孤食状態）の 2 パターンに分けて調査を行い，鏡ありの方が，食事摂取量が増加し，より食品がおいしく感じるという調査結果を示した．

山崎ら [10] は，現実世界にデジタル情報を重ねる AR 技術を用いて，共食環境を疑似的に再現するシステムを提案した．机の中央に設置したタブレットの外カメラで向かい側の空席を撮影し，画面上で 3D キャラクターを重ねることで，同じ空間内に共食相手がいるように見せる．キャラクターは，食べるモーションを一定のタイミングで行ったり，内カメラでユーザの表情（笑顔）を認識し，併せて表情を変化させる機能を備える．

本研究では，主に音声を用いた会話によるコミュニケーションに着目している．孤食と共食の違いの一つとして会話の発生が挙げられるが，これらの先行研究では，会話の支援までは行われていなかった．そのためユーザが共食中にキャラクターに話しかけやすくする工夫を行い，仮想的な共食体験を向上させることで，ユーザの食習慣の改善を図る．

3. 提案

本研究では，音声対話が可能な 2D キャラクターを用いて，ユーザに疑似的な共食環境を提供する．使用イメージとしては食事中，テーブルの手前に食品を用意し，奥にノート PC（もしくは液晶モニター）を設置する．ハードウェアとしては，ユーザの挙動を正面から検出するための PC 内蔵（または外付け）のカメラと，発話検出用のマイク，音声再生用のスピーカーを利用する．システム使用時の様子を図 1 に示す．

図 1　システム使用時の様子

本システムは，大きく分けると 2 つの機能で構成されている．

1 つ目はキャラクターとの音声会話機能である．音声認識，応答文生成，音声合成の 3 つの要素から構成されており，ユーザの発話に対して，AI で動的に応答文を作成した上で，キャラクターが音声で応答する．ユーザはキャラクターと会話しながら食事を行うことでコミュニケーションが取れるほか，会話を挟むことで飲食のタイミングを調整し，早食い防止等の効果も狙う．

2 つ目はユーザの顔の位置や手の位置を検出（トラッキング）して，2D キャラクターの動作をリアルタイムに変更する機能である．トラッキングとアニメーションの 2 つの要素から構成されており，ユーザの動作に対して，キャラクターが対応したリアクションを取る．ユーザの行動に応じたリアクションを返すことで，親近感を持たせたり，話しかけやすくする効果を狙う．

音声会話機能とトラッキング／アニメーション機能を含んだシステム構成図を図 2 に示す．

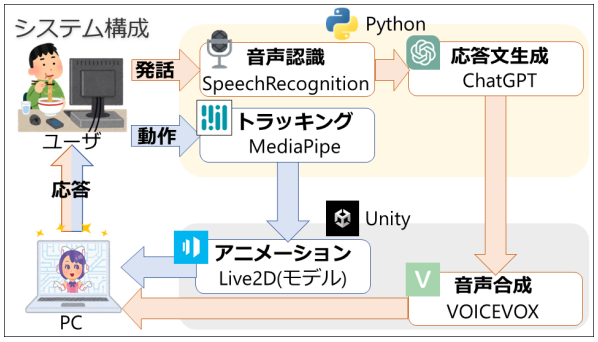

図 2　システム構成図

4. 実装

システム構成図に沿って，音声会話機能の構成要素の説明（4.1–4.3 節）と，トラッキング／アニメーション機能の構成要素（4.4–4.5 節）の説明を行う．

4.1 音声認識

ユーザが入力待ち状態のマイクに向かって発言を行うと，音声認識を用いた文字起こしを行う．Python と音声認識ライブラリ「SpeechRecognition」を利用している．

音声認識を開始すると，マイクの入力を一時停止し，4.3 節の音声合成終了後，再度入力待ち状態となる．

4.2 応答文生成

音声認識機能で文字起こしされた発話内容を元にして，応答文の生成を行う．4.1 節と同一の Python プログラムとして実装しており，AI チャットシステム「ChatGPT」と，その ChatGPT を Python 上で使用可能にするための API「ChatGPT API」を利用する．

生成した応答文は，音声合成にかかる時間を短縮するため，1 文ごとに分割を行い，テキストファイルを介して Unity に送信される．

また，応答文の傾向を調整するために，プロンプトで事

前に性格設定を行う．現在のプロトタイプでは，以下のような設定を行っている．なお実際には，設定文は一つの文章としてまとめて送信する．

- 敬語を使用しない砕けた口調のキャラクターにするため「ギャルの口調で回答する」
- 生成される文章が長くなりすぎないようにするため「128 文字以内かつ 3 文以内で生成する」
- 会話の継続を促すため「食事やレシピに関する知識がある」「ユーザの出来事や食べているものについて質問する」

実際にシステムを使用した際の，ユーザの発言と応答の内容の一例を以下に示す．

ユーザ：こんにちは

AI：おっす！　元気？　今日のご飯は何食べるの？　教えてよ～！

ユーザ：今日はこの後チャーハンを食べるよ

AI：おお，チャーハンか．それって絶対おいしいやつだよね！　具材は何入れるの？

ユーザ：いただきます

AI：いっただきまーす！ 美味しいご飯を一緒に食べるのって最高だよね．キミのご飯，美味しいかな？

4.3　音声合成

Python からテキストファイル経由で受信した応答文を Unity 上のプログラムで受け取り，音声合成ライブラリ「VOICEVOX」を介して WAV ファイルを作成し，自動的に再生する．

音声データの再生が完了したら，再生完了フラグをテキストファイルを介して，4.1 節の Python プログラムに送信する．

VOICEVOX の起動方法としては，ソフトウェアの起動を行う Web サーバー形式の「VOICEVOX エンジン」と，Unity のプロジェクト内に組み込むことができるライブラリ形式の「VOICEVOX コア」の 2 種類がある．今回は Unity 内だけでシステムが解決でき，音声合成にかかる時間もより短い，「VOICEVOX コア」を採用した．

音声データを収納するオブジェクト（AudioClip）は合成用と再生用で分けて用意し，音声合成と音声再生のプログラムは並列に実行している．これにより，音声の再生中に次の音声合成を行うことができ，ユーザの待機時間を短縮することができる．

なお 4.5 節の早食い警告機能のキャラクター音声も，事前に VOICEVOX で音声合成した数パターンの WAV ファイルを使用している．

4.4　トラッキング

PC カメラを使用してユーザの顔や手の位置，相対的な距離等を検出するトラッキング機能を，Python と MediaPipe ライブラリを用いて実装した．取得したデータはテキストファイルを介して，Unity に送信される．

MediaPipe は顔や手の位置や形状等をリアルタイムに検出する機能を提供しており，本研究ではその中でも以下のデータを利用している．

(1) 片手の人差し指の先端の座標（利き手の位置）
(2) 両目の中間点の座標（顔の位置）
(3) 片手から唇までの距離（食品を口に運ぶ動作の判定）
(4) 唇の開閉度（会話や咀嚼動作の判定）
(5) 唇の左右両端の距離（会話や咀嚼動作の判定）

図 3 に，実際にトラッキング機能を実行した際のカメラ映像を示す．図中の番号は，上記箇条書きに対応する．

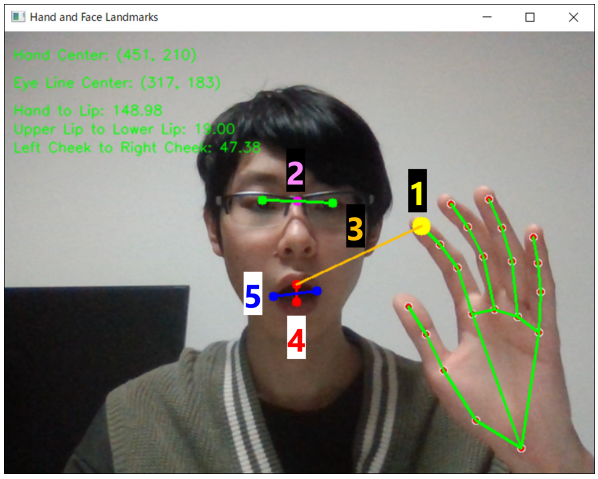

図 3　トラッキング中のカメラ映像

4.5　アニメーション

本機能は Unity 上で Live2D の公式サンプルモデル「桃瀬ひより」[*1]を利用する．

現時点では，大きく 2 つのアニメーション機能を実現している．

1 つ目は，音声合成データの再生に合わせてキャラクターの口を開閉し，喋っているようにみせる機能である．音声データの音量の大小に合わせて，口の開閉度も変化する．

2 つ目は，MediaPipe で検出したトラッキングデータを Live2D のパラメーターに合わせて正規化し，キャラクターの動きを変更する機能である．現在実装できている機能は，ユーザの顔の位置に応じてキャラクターが顔の向きや目線で追随する機能と，ユーザの飲食速度に反応してキャラクターが早食いを音声でやんわり警告し，表情が曇る機能の 2 種類である．

目線追随機能では，ユーザの両目の中間点の座標データを正規化し，ユーザの顔座標からキャラクターの顔の角度

[*1] 株式会社 Live2D，Live2D サンプルデータ集（無料配布），https://www.live2d.com/learn/sample/

と目玉の角度を求める．図 4 に本機能実行時の動作例を示す．

図 4　目線追随機能の動作例

早食い警告機能では，ユーザの片手から唇までの距離データを飲食動作の判定に使用し，ユーザが一定時間内に多くの飲食動作を行った場合[*2]は早食いを行ったとみなし，事前に 4.3 節の VOICEVOX で生成した警告用音声ファイルを再生する．キャラクターは音声に合わせて口の開閉を行い，更に警告音声の再生中はキャラクターの目の開閉度と眉の角度，口の角度を変更して，困っているような表情に変化させる．

警告音声は以下のような複数種類を用意しており，この中からランダムで再生される．

- 「ちょっとちょっと！ 食べるペース早くない？」
- 「あっ！ もうちょっとゆっくり味わって食べようね！」
- 「ねぇねぇ，もっとしっかり噛もうよ！」

図 5 に通常時と警告音声再生中の表情の違いを示す．

図 5　通常時と早食い警告中の表情変化

*2　現在の実装では，15 秒以内に 5 回

5.　議論

ここでは，現在のシステムの課題と改善案について述べる．

現時点では，ユーザの食事に対するフィードバック機能や音声会話機能等に課題がある．

まず，検出できているユーザの咀嚼動作のトラッキングデータを使用し，ユーザとタイミングを少しずらして 2D キャラクターにも咀嚼アニメーションを加えることで，共に食べている感覚を増強する．

次に，現在の音声会話機能では，ユーザの音声認識から応答文 1 文目の音声再生までに，10 秒程度の待機時間があり，スムーズな対話が難しい．こうした無言時間対策として，キャラクターの動きや音声の設計を検討すると共に，更なる処理時間の短縮を図る．また，現在の音声会話はユーザから話しかける仕様となっているが，キャラクター側から会話を始めるような調整も進める．さらに，ChatGPT へのプロンプトによる AI のキャラクター設定についても，指定より文章量が長くなったり，敬語を使ってくる等，意図通りの挙動をしない場合があるため，設定を工夫していく．

6.　まとめと展望

本研究では，音声対話が可能な 2D キャラクターを用いて，ユーザに疑似的な共食環境を提供するシステムを提案した．ユーザの発話を音声認識し，AI を用いた応答文生成と音声合成を組み合わせることで，画面上のキャラクターと様々な会話を行うことができる機能を開発した．また，PC カメラを使用してユーザの顔や手の位置，相対的な距離等を検出し，取得したデータを基にキャラクターにアニメーションのイベントを付与する機能も構築した．今後は議論で述べたようなシステムの改善を進めつつ，システムの評価実験を行う．具体的には，トラッキング性能や音声会話時間等の性能評価に加えて，ユーザにシステムを利用してもらい，孤独感の感じ方や食べる速度の変化等を調査するユーザ評価を行う．

謝辞

本研究の一部は，JST 未来社会創造事業 JPMJMI21J6 の支援を受けた．

参考文献

[1]　徳永弘子, 武川直樹, 木村敦, 孤食と共食における食事動作のメカニズム ―食事の形態がもたらす心理的影響との関連に照らして―, 日本食生活学会誌, 2016, 27 巻 3 号,

pp.167-174.
[2] 赤利吉弘, 小林知未, 小林千鶴, 植杉優一, 内藤義彦, 成人における年代別・性別の共食頻度と生活習慣, 社会参加および精神的健康状態との関連, 栄養学雑誌, 2015, 73 巻 6 号, pp.243-252.
[3] Amber J. Hammons, Barbara H. Fiese, Is Frequency of Shared Family Meals Related to the Nutritional Health of Children and Adolescents?, Pediatrics, 2011, 127(6), pp.1565–1574.
[4] 内閣府, 第 2 次食育推進基本計画, http://www8.cao.go.jp/syokuiku/about/plan/pdf/2kihonkaiteihonbun.pdf （2023-11-1 参照）.
[5] 農林水産省, 第 3 次食育推進基本計画, https://www.mhlw.go.jp/file/06-Seisakujouhou-10900000-Kenkoukyoku/0000129496.pdf （2023-11-1 参照）.
[6] Yugo Nakamura, Rei Nakaoka, Yuki Matsuda, Keiichi Yasumoto, eat2pic: An Eating-Painting Interactive System to Nudge Users into Making Healthier Diet Choices, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT), Vol. 7, No. 1, pp. 1-23, 2023.
[7] Azusa Kadomura, Koji Tsukada, Itiro Siio, EducaTableware: Sound Emitting Tableware for Encouraging Dietary Education, Journal of Information Processing, Vol.22 No.2
[8] Azusa Kadomura, Cheng-Yuan Li, Koji Tsukada, Hao-hua Chu and Itiro Siio, Persuasive Technology to Improve Eating Behavior using a Sensor-Embedded Fork, Proceedings of UbiComp2014, pp. 319-330,(Sept, 2014)
[9] 中田龍三郎, 川合伸幸, 鏡で自分を見ると食事をおいしく感じる―大学生と高齢者の比較―, 日本認知科学会第 32 回大会, 2015, pp.51-56.
[10] 山﨑柚季, 鈴木陽菜, 西森千珠, 栗原渉, 有山大地, 韓旭, 串山久美子（東京都立大）, AisekiR：AR キャラクタとの共食システム, インタラクション 2022 論文集 インタラクティブ発表 2D17, 2022, pp.329-332.