

音声認識を用いたキャラクター育成アプリケーションの開発

Development of Character Growing Application Using Speech Recognition

J3A 日比 瑛太 ・ J3A 本田 楓 ・ J3A 本間 晶 ・ J3A 松浦 颯

指導教諭: 辰巳 良美, 中山 智明

1. はじめに

近年、青少年の間で「ヤバイ」「エモい」といった単純な表現が多用され、国語力や表現力の低下が問題視されている。厚生労働省の調査 [1] によると、効果的なコミュニケーションには「意思疎通」「協調性」「自己表現能力」が不可欠とされている。しかし、適切な言葉を用いる力が不足しているために、日常会話で自分の意図を正確に伝えられないケースが増えていると考える。また、教育現場では社会性やコミュニケーション能力の低下が指摘されている。[2] 一方、経団連の新卒採用に関するアンケート調査 [3] では、企業が採用選考で最も重視する要素が「コミュニケーション能力」であることが分かっている。これらの結果から、企業が求める能力と若年層が現実持つスキルとの間に乖離が生じていると考えられる。こうした現状から我々は、普段から積極的にコミュニケーションを取る機会を提供することで、表現力やコミュニケーション能力の向上につながると考えた。

そこで本研究では、ユーザーがキャラクターとの対話によってコミュニケーション能力を向上させることを目的としたキャラクター育成アプリケーションの開発を行った。

2. 論理・実験

アプリケーションの概要

本アプリケーションはキャラクター育成ゲームである。ユーザーがキャラクターに対して話しかけることで、キャラクターが返答を行う。キャラクターはプルチックの感情の輪に基づいた 8 つの感情 [4] をシミュレーションし、ユーザーの発話内容に応じて感情を変化させる。感情の変化はキャラクターの表情や口調に反映される。一定回数の会話を行うとキャラクターが進化し、感情の出現割合によってその進化先が分岐する。ユーザーはキャラクターの感情を観察することで、自分の発話内容が相手にどのような影響を与えるかを学ぶことができる。

システム面について

ゲームエンジンには、Live2D との互換性の関係から Unity を採用した。音声認識には OpenAI の Whisper を採用し、ユーザーの発話内容をテキストに変換する。Whisper は他

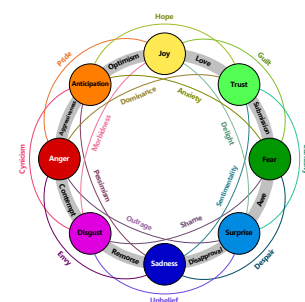


図 1: プルチックの感情の輪 [4]

の音声認識に比べてその精度が高く、また多言語に対応していることからこれを採用した。キャラクターとの会話には GPT-4o mini を使用し、テキスト生成の前に、キャラクターの感情として「喜び」「信頼」「恐れ」「驚き」「悲しみ」「嫌悪」「怒り」「期待」の 8 つを 0～9 の 10 段階（以下「感情値という」）でシミュレーションしている。感情値はキャラクターの口調や表情などに反映され、また感情値の出現割合によって 1 段階目の進化先が分岐し、キャラクターの性格に影響を与える。

音声会話の実装は次のようになっている。ユーザーの音声は AWS Lambda を通じて Whisper へ送られ、文字起こしが行われる。この文章はこれまでの会話履歴と合わせて GPT-4o mini に入力され、会話文と感情値の生成が行われる。生成された会話文は VOICEVOX により音声に変換された後、感情値とともにクライアントへ送られる。

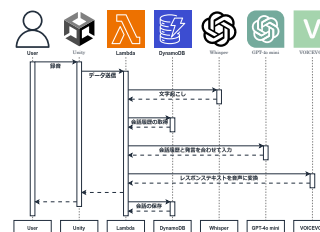


図 2: 音声会話のシーケンス図

デザイン面について

キャラクターデザインはドット絵、Live2D、3D の 3 種類を製作し、キャラクター進化に 3 つの段階を設けた。ドット絵のモデルでは、プレイヤーにキャラクターの進化先を予測されないように、表情差分を 5 種類、カラーバリエーションを 7 種類製作した。Live2D モデルでは、プレイヤーが親しみやすいように動物をモチーフとしたキャラクターデザインを行った。また、愛着を持ってもらいやすいよう、首を傾げたり尻尾が揺れるなどの可愛らしいモーションを製作した。3D モデルでは 5 種類の表情差分を製作した。

3D の段階では、ゲームをより楽しんでもらうための効果として、裸眼立体視に挑戦した。裸眼立体視は錯視効果の一種で、3D メガネや HMD（ヘッドマウントディスプレイ）を使わずに立体視を楽しむことができる。

感情シミュレーションについての実験

本番環境と同じシステムプロンプトを入力したアシスタントに対して、8 感情から「喜び」「恐れ」「悲しみ」「怒り」の 4 感情を引き出すであろうプロンプトをそれぞれ 100 パターン入力し、シミュレーションの正確性を確認した。正誤判定は、当該の感情値が最も高い値であるか否かでを行った。

レスポンスについての実験

1 文と 3 文での 2 パターンで会話を行い、返答にかかる時間を測定した。また、1 文と 3 文での会話の平均時間を比較した。

3. 実験結果

システム面

感情シミュレーションに関する実験の結果は次のようになった。「喜び」と「悲しみ」についてはそれぞれ8割以上、「恐れ」についても7割ほどの精度を確認したが、「怒り」についてはわずか9%とかなり低い値となり、ほとんどのパターンで感情値が0～1を示していた。

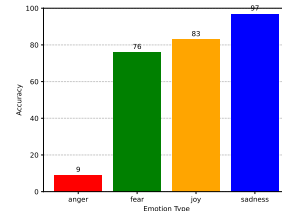


図 3: 感情シミュレーションの実験結果

レスポンスに関する実験の結果は次のようになった。

	1 回目	2 回目	3 回目	4 回目	5 回目	平均
1 文	6.72	7.32	9.27	5.93	6.23	7.09
3 文	7.99	7.62	8.70	9.26	8.96	8.51

表 1: レスポンスの実験結果

応答時間は平均して7～8秒だった。最速と最遅の差は3秒程度であり、1文と3文での会話による応答時間の差は1秒程度であった。

デザイン面

キャラクターデザインがシンプルすぎたため、視覚情報をさらに追加することでユーザーがキャラクターとの触れ合いに飽きてしまわないような工夫が必要だと感じた。裸眼立体視は街中のビジョンを参考にしたため、見上げる時に立体的になる構図になってしまっており、斜め上から立体的に見れるように変更する必要があると分かった。

4. 比較・考察

システム面

シミュレーション実験の結果の要因として、学習時のチューニングによる影響が考えられる。このようなAIを公開するにあたって、開発者は差別や情報の偏りといった課題がある程度クリアする必要がある。そのため、学習データに存在するバイアスの除去など、何らかの調整を行っていることがある。[5] このことから、今回利用した GPT-4o mini はユーザーに対立するような生成を制限されているのだと考えられる。

レスポンスについては、今回の構成ではこれ以上の高速化は厳しいと思われる。システムが外部 API に依存した設計であるため、ネットワークの遅延や API 自体の応答時間に影響される。

デザイン面

裸眼立体視は一定の場所からしか効果を得ることができないため、その位置（スイートスポット）の設定が重要となる。今回は対象年齢となる子供の座高や椅子と机の高さからスイートスポットを設定した。しかし、これは平均的な座高を基準にしているため、人によって大きく異なった見え方になってしまう恐れがある。そのため、実際のユーザーに合わせたスイートスポットの設定が必要であると感じた。

5. 結論

本研究では、キャラクター育成アプリケーションの開発を行った。感情のシミュレーションにおいては、喜びや悲しみについては高い精度を達成したが、怒りに関しては正確な表現が困難であることが分かった。また、レスポンスについては平均して7～8秒で、ある程度の自然な会話が可能であることが確認できた。今後は、感情の表現をより正確に行うためのチューニングや、レスポンスの高速化を実現するためのシステムの見直しが必要である。

6. 今後の課題

システム面

怒りに関しては、感情値の調整を行い、より正確な表現を行うためのチューニングを行いたい。また、レスポンスについては、より高速な応答を実現するために、システムの構成を見直す必要があるだろう。

また、実際に会話した時に感じたこととして、対等な立場での会話ができていないと感じた。キャラクターが受動的に返答を行うだけでなく、自発的に会話を始めるなど、より自然な会話を実現するための工夫が必要だと感じた。

デザイン面

2D および 3D のグラフィックでは、全体的に感情を表現するためのレパートリーが欠けていたため、AI を使って感情を表現するというゲームの特徴をより活かせるようにパターンを追加していきたい。3D のグラフィックでは短い距離感の場合頭の位置がズレると立体的に見えなくなってしまうという欠点があったため、Head-coupled Perspective の技術を導入し、頭の位置を追尾して立体的に見えるようにしたい。

7. 引用文献

- [1] 厚生労働省, 「「YES・プログラム」(若年者就職基礎能力支援事業)の対象講座(254講座)・対象試験(29試験)を新たに認定」. 参照: 2025年1月16日. [Online]. 入手先: <https://www.mhlw.go.jp/houdou/2009/03/h0313-4.html>

- [2] 文部科学省, 「コミュニケーション教育推進会議 資料 6 『コミュニケーション能力』に関する指摘・調査等」. 参照: 2025 年 1 月 14 日. [Online]. 入手先: https://www.mext.go.jp/b_menu/shingi/chousa/shotou/075/shiryo/_icsFiles/afieldfile/2010/06/22/1294462_03.pdf
 - [3] 一般社団法人 日本経済団体連合会, 「2018 年度 新卒採用に関するアンケート調査結果」. 参照: 2025 年 1 月 14 日. [Online]. 入手先: <https://www.keidanren.or.jp/policy/2018/110.pdf>
 - [4] フリー百科事典『ウィキペディア (Wikipedia)』, 「ロバート・プルチック - Wikipedia」. 参照: 2025 年 1 月 16 日. [Online]. 入手先: <https://ja.wikipedia.org/wiki/%E3%83%AD%E3%83%90%E3%83%BC%E3%83%88%E3%83%BB%E3%83%97%E3%83%AB%E3%83%81%E3%83%83%E3%82%AF>
 - [5] Tyna Eloundou ほか, 「First-Person Fairness in Chatbots」. 参照: 2025 年 1 月 16 日. [Online]. 入手先: <https://cdn.openai.com/papers/first-person-fairness-in-chatbots.pdf>
-

J3A 日比 瑛太

1. 自分が担当したところ

1. システム開発

2. 自分が担当したところの結果

おおむね正しい感情の表現を行うことができたが、怒りに関しては正確な表現が困難だった。レスポンスについては平均して 7~8 秒で、ある程度の自然な会話が可能であることが確認できた。

3. 自分が担当したところの今後の課題・考察

怒りに関しては、より正確な表現を行うために、感情値の調整を行いたい。また、レスポンスについては、より高速な応答を実現するために、システムの構成を見直す必要がある。

J3A 本田 楓

1. 自分が担当したところ

1. ゲーム UI、タイトルロゴ作成

2. ドットデザイン

3. 3D 小物作成

2. 自分が担当したところの結果

親しみやすいポップなデザインができたと思う。3D ではキャラクターの等身や雰囲気に合わせて小物を作成することができた。

3. 自分が担当したところの今後の課題・考察

ロゴデザインではアプリの親しみやすさを表すことができたのではないと思う。UI と

して、2D、3D を想定したものは作成できたが、ドットの際に使用するものの作成が抜けていたのでこれから追加できるように準備していきたい。

J3A 本間 晶

1. 自分が担当したところ

1. キャラクターデザイン

2. Live2D モデルの制作

2. 自分が担当したところの結果

キャラクターデザインは全8種類、モデルは3体分完成した。プレイヤーにとって親しみやすいよう、動物をモチーフとしてキャラクターのデザインを行い、飽きがこないようなモデルの動きを製作した。

3. 自分が担当したところの今後の課題・考察

キャラクターに揺れものを追加して、視覚的情報量を多くし、飽きがこないようにする。動きがどのキャラクターでも比較的変わらないので、そのキャラクターだけの特有のモーションなどを追加することでプレイヤーのプレイ意欲を高められると考える。

J3A 松浦 颯

1. 自分が担当したところ

1. 3DCG 関連のグラフィック

2. 自分が担当したところの結果

ゲームに対応した表情差分付きの3DCGキャラクターを製作し裸眼立体視を利用した立体的に見えるゲーム演出を製作することができた。また、感情を表現するために複数のアニメーションを製作した。

3. 自分が担当したところの今後の課題・考察

今回は一定の位置関係からしか立体的に見ることができないようになってしまっている。ので、**Head-coupled Perspective** を利用してどの方向からでも立体的に見えるシステムを導入したい。また、キャラクターアニメーションだけではなく演出を追加することでよりユーザーに楽しんでもらえるようにしたい。