

初中級の英会話学習者を対象とする 大規模言語モデルを用いた音声対話の試み

真弓愛理^{†1} 神場知成^{†1}

概要: 初中級者の英会話学習を想定して大規模言語モデル (LLM) と簡易イラストを用いた音声対話システムによる実験を行い、その可能性や課題を示した。具体的には ChatGPT API を用いて構築したシステムで、ユーザが対話相手の性別や立場、シチュエーションを設定して英会話を行うものであり、英会話が初中級レベルの大学生ユーザを対象として、対話の自然さについて相手から受ける印象や、このようなシステムに対する将来的な期待を調査した。結果は現時点での LLM の性能に依る部分もあるが (本システムでは gpt-3.5-turbo を利用) それだけでなく、AI との対話という観点でヒューマンインタフェースの観点から本質的な知見も得られた。たとえば「相手に気を遣う必要がないので心理的障壁が低く話しやすい」という意見も「ノリや雰囲気で話せないのも、(人間を相手にする場合よりも) 却って話しにくい」というものもある。LLM は急速に進歩しており英語教育においても大きな可能性があるが、対人と比較したメリット・デメリットなどさまざまな側面から検討していく必要があるだろう。

1. はじめに

従来の義務教育における英語教育では会話シミュレーションに基づく英語学習システムが採用されており、典型的なモデル会話を繰り返すスキット学習が行われている。日本人の英語能力は世界で後れをとっており、固定化されたシチュエーションに応じた例文の暗記が最適な手段とは思えない。国際教育事業を行う EF Education First の実施する EF EPI (EF English Proficiency Index) が、試験を受験した成人の英語能力指数の平均を国ごとのランキングとして発表しているが、2023 年の日本の順位は 113 カ国中 87 位である [1]。しかも、調査が開始した 2011 年以降日本人の平均スコアは年々下降し、2022 年時点で 5 段階中 4 番目の「低い能力レベル」に属している。

一方、2022 年末に ChatGPT が公開されて以後、一般的に利用可能な大規模言語モデル (LLM) の性能が急速に向上している。筆者等は、LLM を利用することで柔軟なシチュエーションを設定した自由な英会話が可能になると考え、ユーザ自身が自由に設定した状況に応じて AI と自由会話を行うことができる Web 上の対話システムを開発し、その初期的な利用評価を行った。

2. 関連研究

コンピュータ利用を伴わない一般の中学校前半の英語教育の場面において、様々な話題に対して 1 分間自由に英会話を継続させる「1 分英会話」を導入した事例を、山岡が報告している [2]。報告によれば、取り組み開始時の生徒たちは定型的な表現に依存しており、同じ形式の表現を繰り返すことで英会話を行う傾向がみられたが、活動を続けることで定型的な応答に追加して会話を自身で発展する動き

を見せる生徒が現れた。アンケートでも、生徒たち自身も「1 分間英会話」が力の付く活動と認識しており、自由会話が英語力の伸長をもたらす可能性のある取り組みであったと結論づけている。

コンピュータを利用したものとして、中川等が音声認識技術を利用した英会話 CAI システムの有用性について述べている [3]。また岡本等は、従来の英会話学習における固定的な会話文脈によるスキット学習では、学習者は受動的な学習を余儀なくされている点を問題とし、学習者主導の対話進行の中でシステムによって対話を制御することによる教育的支援により効果的な学習を期待できると述べている [4]。

その他、すでに一般に利用可能な学習システムとして、Speakeasy Lab (<https://www.usespeak.com/jp>) による GPT-4 利用アプリ Speak, また ELSASpeak (<https://elsaspeak.com/ja/>), SpeakBUDDY によるもの (<https://www.speakbuddy.jp/>) などがある。

英語学習は初級・中級・上級などさまざまな局面において注力すべきことや必要な学習方法が異なると考えられるが、急速に広がる LLM についてはまだ効果的な活用方法の知見が少ない。たとえば初級者に対して急に「自由に英語で対話をしてください」と言ってもとまどう場合が多いであろう。本研究では、最初にユーザ自らが会話の状況を設定した上で音声による自由対話を行うシステムを作成し、初中級者を対象とし初期的な利用の効果や課題の知見を得ることを目的とする。

3. 開発システム

3.1 ユーザインタフェース

本システムは Web ページ上でユーザが対話の基本条件

^{†1} 東洋大学情報連携学部 (INIAD)

を設定した上で、音声により英語の自由対話を行うものである。画面表示にあたっては、対話相手のイメージを持ちやすくするために簡単な顔のイラストを表示し、表情に複数パターンを持たせることで対話の補助を行っている。

(1) 対話状況の設定

ユーザはまず「自身の名前」「相手の名前」「相手の性別」「相手との関係性」「会話のシチュエーション」の5項目を設定する(図1)。いずれもその時に行う対話の中で利用するためのものであり、自分の名前なども、その時に応じた任意のものを設定可能である。



図1. 対話状況の設定画面

(2) 対話画面

画面上にはまず、設定した相手の性別に応じた顔のイラストが表示されるので、ユーザはすぐに「はなす」ボタンを押して英語で話し始めることができる。

音声入力した英文は画面上に日本語訳と共に表示され、自分が発音した通りの英語が正しく音声認識されているか、またそれが自分の意図した内容として解釈されているかどうかを確認できる。その後、最初に設定された会話状況に応じて相手からの返答が音声出力され、画面上に日本語訳と共に表示される。以後、この手順を繰り返すことで疑似的な英会話を行う(図2)。



図2. システムとの対話画面

ユーザが設定した相手の性別や生成された返答文に含まれるワードによって、相手を模したキャラクターのイラストが図3のように変化する。相手の性別によって音声出力の声の高さも変化する。いずれも、ユーザ体験を実際の英会話に類似させることを目的としている。



図3. 表情の種類(上段:男性, 下段:女性)

ユーザインタフェース上は、会話はいくらかでも続けることができるが、後述のように実装上の都合により、現時点では数回の会話を続けたあと、応答が非常に遅くなるなどの限界がある。

3.2 システム構成

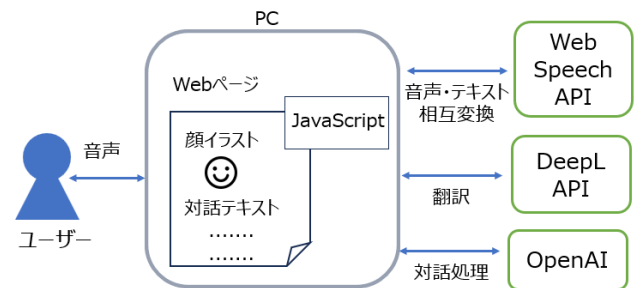


図4. システム構成

図4にシステム構成を示す。本システムはWebページ上のアプリケーションとして実装している。ユーザの音声入力はブラウザからWeb Speech APIでテキスト変換した後にOpenAIの言語モデルのChatGPT API(本稿ではgpt-3.5-turboを利用)に受け渡し、戻り値として得られるテキストをWeb Speech APIにて音声変換してブラウザから出力する。ユーザの入力とChatGPT APIの出力はいずれも、DeepL API(<https://www.deepl.com/ja/docs-api/api-access>)を使用して作成した日本語訳と共に画面上に表示される。ChatGPT APIに引き渡すsystem roleは図5のように設定している。

```
const inst =
  "あなたの名前は" + gptname + "であり、あなたはuserの" + rela + "です。\\
  成りきってください。" +
  "あなたとuserは"+situ+"という状況でこれまで会話してきました。\\
  この状況に沿う形で端的に自然な会話をuserとしてください。" +
  "自分がAIであるということは忘れてください。20ワード以内で返答をしてください。"
  messages.push({ role: "system", content: inst });
  messages.push({ role: "user", content: "My name is"+username});
```

図5. ChatGPT API に与える system role

(変数 username, gptname, rela, situ に、ユーザが設定した自分と相手の名前、関係性、シチュエーションを設定)

図 6 にシステムの処理の流れを示す。本システムは、継続中の会話内容を前提として返答を出力させるために、会話ログを毎回全て ChatGPT API に入力している。そのため会話が長期化すると入力のデータ量も増加し、API の応答を待つ時間が長くなるため、長時間の英会話をスムーズに行うことができない。会話ログを定期的に ChatGPT に要約させて system role からの指示文として会話ログにまとめる形での解決を試みたが、会話内容を前提として維持する正確性に欠けた返答が行われるようになってしまっている。

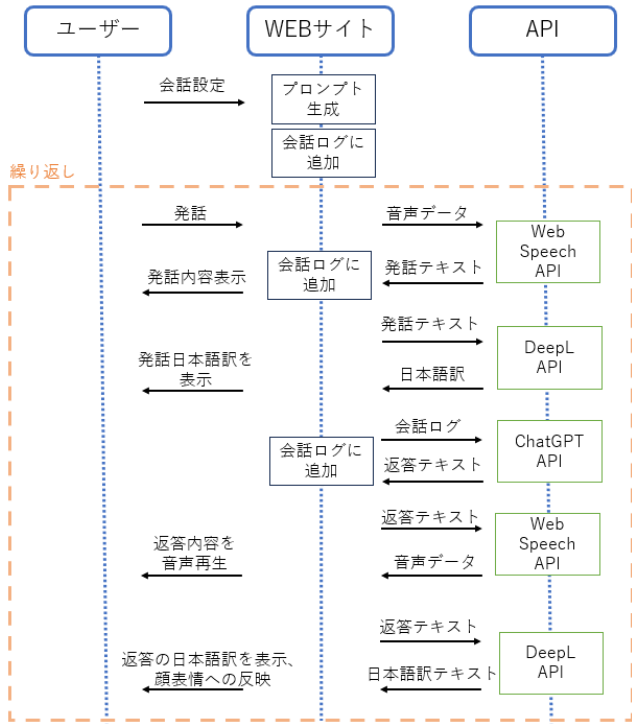


図 6. 処理の流れ

4. 対話実験



図 7. 実験風景

作成したシステムを用いて、2 回に分けて大学生 9 名に対してユーザ実験を行った。2 回目は対話の設定画面の UI に若干の改良をしているが、対話をする場面の UI は同一

である。図 7 に実験風景を示す。ここで対話用 PC の横にタブレットを置いているが、これは英会話に自信のない人が対話しやすくするために図 8 に示すような基本文例を数個表示したものである。実験では、このような環境で数分間アプリケーションを使用した後に使用感についてアンケートを行った。



図 8. 入力サポート用の例文

5. 実験結果

5.1 設定場面

図 9 に、設定された場面例を示す。ここには、実験者があらかじめ用意したサンプルをユーザが一部変更するなどして利用したものと、ユーザ自らが設定したものが含まれている。

場面 1	相手：女性，友人でキャビンアテンダント 状況：仕事の相談
場面 2	相手：女性，キャビンアテンダント 状況：飛行機内で酔った
場面 3	相手：男性，初対面 状況：日本について知りたい

図 9. 設定された場面例

5.2 アンケート結果

アンケートでは次のような質問を行った。

- 英会話への苦手意識
 - 設定したシチュエーションに沿った返答を得たか？
 - 自由会話として自然な返答を得たか？
 - 対人と比較した心理的障壁（苦手意識のある 8 名）
 - 初学者（文法・語彙不十分）に利用できると感じるか？
 - 実践（文法・語彙十分）に利用できると感じるか？
- なお、(c)から(f)については理由も聞いている。

まず基本的な集計結果を図 10 に示す。

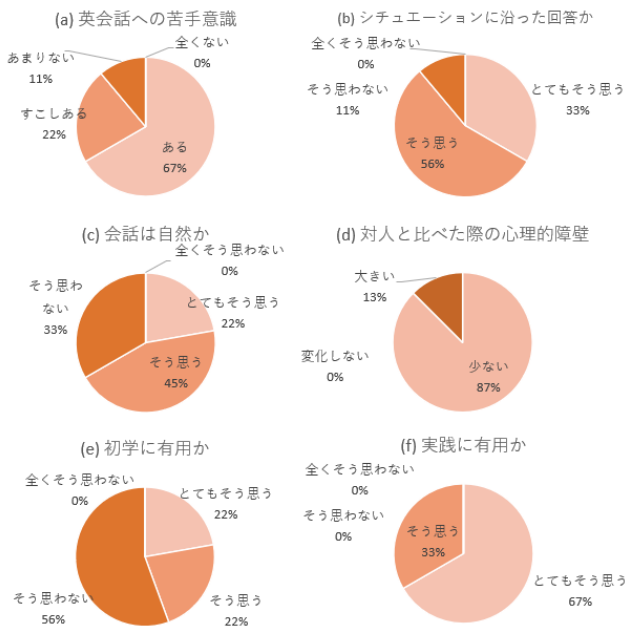


図 10. 主なアンケート結果

回答の理由等は次の通りである。

(b)で「シチュエーションに沿った回答が得られない」または(c)で「自然な返答ではない」と回答した理由

- 途中から、想定したシチュエーションからずれた
- 最初は整合性があったが、後半から「AI と会話している感」が強まった
- AI が考えた文章という感覚がある
- 発音が悪いと、AI の認識結果が異なる

(d)で「心理的な障壁が大きい」と回答した理由

- 対人だと感情やパッションを伝えることができ、ある程度ノリで会話ができるため、AI の方が障壁が高い

(e)で「初学者に対して利用できるか」に対して、「(まったく) そうは思わない」と回答した理由

- 文法が間違っているでも GPT が意図を汲んで返答してくれるので、間違いに気づかないことがありそう
- GPT の返答や会話の質が大人向けの印象
- 知識がないと会話ができないのではないかと思った

5.3 主な対話例

(1) 成功例 (テキスト部分は読みやすさのため貼り込み)



図 11. 対話例(1)

「相手女性、キャビンアテンダント、飛行機内で酔った状況」の事例。「乗り物酔い」を「アルコール酔い」と誤認したり、人間の側の英語に間違いがあるなどの問題はありますが、ユーザの設定した会話シチュエーションに沿った返答が出力されている。

(2) 失敗例 1: AI の自我が出現

(テキストは読みやすさのため貼り込み)

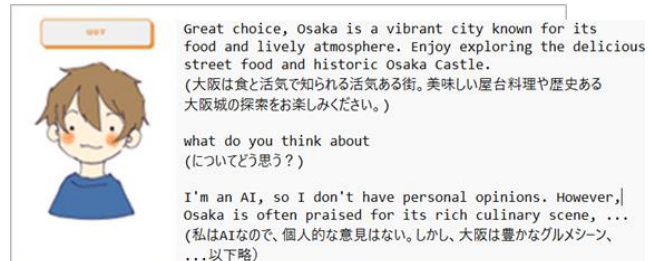


図 12. 対話例(2)

「相手男性、初対面の人、日本について知りたい」の事例。

ユーザの観光に関する質問へ適切な返答が行われる中、4 回目の返答で「私は AI」と出力された。

その他の事例でも、「相手は女性、友人でありキャビンアテンダント、シチュエーションは仕事の相談」と設定して英会話を開始した事例において、最初はキャビンアテンダントとして回答していたが途中で「AI」として返答し、そのまま戻ることはなかった。

6. 考察

6.1 対話の性能

前述のように本システムは、会話ログを毎回全て ChatGPT API に入力しているため会話が長期化すると入力のデータ量も増加し、API の応答を待つ時間が長くなるため、長時間の英会話をスムーズに行うことができない。会話ログを定期的に ChatGPT に要約させ、system role からの指示文として会話ログにまとめる形での解決を試みたが、現状では正確性に欠けた返答となっている。

その他、状況設定に含まれない相手のプロフィールを質問すると「自分は AI である」と言い始めたり、「○○について教えて」と質問をすると Wikipedia のような長文を返す事例などの問題は多発した。対話の質を上げるために、「自分が AI ということは忘れてください。」「私の英語の文法に間違いがある場合には指摘をしてください。また、私の発言から私の英語能力を推察し、私の英語能力が低いと判断できたらあなたは簡単な英語で答えるようにしてください。」等のプロンプトも試みたが、大きな改善はみられていない。

ただし前述のように ChatGPT API として現時点で gpt-3.5-turbo を用いており、より高度な API により対話の性能は向上すると思われる。

6.2 教育面

対話学習としての有効性については、このようなシステムにより英会話教室への参加や留学といった経験を得るためのコストを払うことなく英会話の実践の機会を得ることができるのではないか、という期待は高いと考えられる。

また、英語に苦手意識のあるユーザから、本システムでは失敗や間違いを恐れる気持ちや羞恥心が緩和されるという回答も多く、心理的障壁が緩和され、より気楽な実践の機会になるという効果もあると考えられる。

一方、「自然な返答が得られていない」という回答も多いほか、心理的障壁について「対人であれば感情やパッションを伝えることができ、ある程度ノリで会話ができるので、却って対話しづらい」という回答もあった点は興味深い。これは LLM の技術的な限界ではなく人間の側のマインドセットによるものと考えられ、AI と対話するときの人間側の意識として今後考えていく必要があるだろう。

6.3 マルチモーダルの活用について

本システムにおいて、単に言語による対話だけでなく、相手の顔を示す簡単なイラストを表示する、音声において男女によって声の高さを変えるなど、いくつかのメディア的な工夫を行った。生成 AI 技術はこれらの点でも急速に進歩しており、将来的にはそれらの利用の効果も含めて考える必要がある。

7. おわりに

本研究では AI との英会話を英語学習に用いることについての有用性を検討するために、AI を会話相手として音声入力・音声出力による英会話を行うシステムを作成し実験を実施した。AI との英会話の意義は、英会話の実践の機会としては有用性が見られるが、語彙や文法、発音などが不十分な初級者、ある程度スキルのある上級者の学習などの目的によっても違いがあるだろう。また、英語を話すことに対して心理的障壁を持つ人々に対して、より気楽に英語を発話する機会を与えるという効果がある一方、必ずしもそうでもないとする例もあった。

今後さらに、プロンプトにより工夫を加えることで AI を「教師」として実用化することができれば、より様々なバックグラウンドを持つ学習者に対して英語学習の機会を提供できる可能性もある。それには「間違いの指摘」「相手の英語レベルに合わせた返答」等である。

英会話の相手として AI を利用する場合、対人で行った場合と比較したデメリットについても検討する必要がある。表情やジェスチャーで感情やニュアンスを伝達するコミュニケーション能力の発達や、実際の会話の臨場感の中で発達する英語能力も考慮すれば、安易にすべてを AI に置換することには慎重になるべきであろう。実際の教育の形態を変えずに教育者をそのまま AI に置換するのではなく、

新たな価値を付与したうえで共存させていくような形を期待したい。

謝辞 本研究は東洋大学重点研究推進プログラムにより助成を受けたものです。同助成に感謝いたします。

参考文献

- [1] イー・エフ・エデュケーション・ファースト・ジャパン株式会社, 第 2022 版 世界最大の英語能力指数 ランキング, <https://www.efjapan.co.jp/epi/>
- [2] 山岡大基, 中学校前半における自由会話の指導, 中国地区英語教育学会研究紀要 No.41, pp.51-60 (2011)
- [3] 中川聖一 Allan A. Reyes 鈴木英之 谷口泰広, 音声認識技術を利用した英会話 CAI システム, 情報処理学会論文誌, Vol.38.8, p.1649 - 1658, (1997)
- [4] 岡本竜 矢野米雄, 英会話 CAI システムのための対話シミュレーションの実現, 日本教育工学雑誌, Vol.19.2, p.61-71, (1995).