対話型鑑賞のファシリテータ養成のための複合現実 AI システム

○福田 涼太（九州大学），倉爪 亮（九州大学）

Mixed reality AI system for training VTS facilitators

○ Ryota FUKUDA (Kyushu University), and Ryo KURADUME (Kyushu University)

Abstract: Visual Thinking Strategies, or VTS for short, is a method of art appreciation in which multiple people in a group interpret a work of art through repeated dialogue. In this paper, we propose a mixed reality AI system that uses MR headsets and Large Language Models to train facilitators, who play an important role in VTS.

1. 序論

美術鑑賞の方法の一つとして、複数人のグループで対話を重ねながら美術作品を読み解いていく対話型鑑賞 [1][2] という鑑賞法がある。対話型鑑賞のイメージ図を Fig.1 に示す。この対話型鑑賞では、グループ内の会話をスムーズに進行させるファシリテータの役割が重要となる。しかし、ファシリテータの訓練、養成には、現状では美術館等で開催される絵画鑑賞会などで、実際の参加者を対象に現場で練習を繰り返すしかなく、訓練のための十分な機会が確保できないという課題があった。そこで、我々は拡張現実ヘッドセットと大規模言語モデルを用いて対話型鑑賞を仮想的に再現することで、美術館等の場所の制約なく、一人で対話型鑑賞のファシリテータ訓練を行うことができる複合現実 AI システムを開発した。

Fig. 1 対話型鑑賞のイメージ (ChatGPT-4o にて作成)

2. システム構成

2.1 対話型鑑賞についての状況設定

まず仮想的に再現する対話型鑑賞の状況を Fig.2 で示す。複合空間内には仮想的に 5 人のキャラクターが存在しており、それらは全員対話型鑑賞の参加者である。拡張現実ヘッドセットを装着している人はファシリテータとして振る舞い、Fig.2 の上部中央 (ユーザの正面)、および Fig.3 で示すように、ユーザの背面 (実際の対話型鑑賞と同じく鑑賞者の正面) にある絵画に対して対話型鑑賞を展開していくことが要求される。

Fig. 2 対話型鑑賞の状況

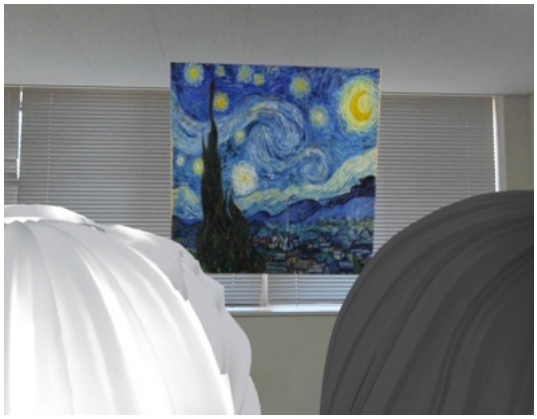


Fig. 3 ユーザの背面すなわち鑑賞者の正面にある絵画

2.2 5 人のキャラクターについて

各キャラクターはそれぞれ個別に感情の値を 0 から 1 の間の値で持っており、会話がポジティブな内容であれば値が大きくなり、ネガティブな内容であれば値が小さくなるというように、ユーザとの会話の内容に応じて値は徐々に変化していく。また、この感情の値により顔の表情も変化する。ヘッドセット (Quest3、Meta) のアイトラッキング機能を利用することでキャラクターと目が合っているかどうかを判定し、目が合っているキャラクターの表情にその時点での感情の値を反映させている。さらに、ファシリテータが各参加者に均等に話しかけるようにするために、キャラクターごとに目が合っている時間を計測して表示し、最も目が合っている時間が短いキャラクターには”short”と表示している (Fig.4)。Fig.2 の下部に表示されているように 5 人のキャラクターにはそれぞれ名前があり、年齢や出身、性格なども個別に設定されている。この

ような性格などの違いにより各キャラクターの発言内容や話し方が異なるようにしているが、詳細は 2.3 で説明する。

Fig. 4　目が合っている時間の計測

2.3　複合現実内のキャラクターとの会話の実現について

ユーザとキャラクターの会話をどのように実現しているかについて説明する。ユーザが発言してからキャラクターが返答するまでの大まかな流れを Fig.5 に示す。

まず対話型鑑賞において、目を合わせながら会話をすることが大切である。したがって目を見て発言しなければキャラクターが反応しない仕様にしており、目を見て発言することが会話を成立させる条件の一つ目となる。しかし、実際の対話型鑑賞ではずっと目を見て話すわけではなく、絵画に視線を向けたり、他の人の様子を見たりする。このような状況を再現するために、発言中はユーザは視線を動かすことができるが、発言が終わるタイミングで返答して欲しいキャラクターの目を見ている必要がある。会話を成立させる条件の二つ目は、一定の大きさ以上の音量で発言することである。わずかな音量であれば反応せず、一定の大きさ以上の音量であれば録音を開始し、無音状態が閾値以上続いた時点で録音を終了する。ここで、録音の長さが閾値以下であった場合はキャラクターは反応しない。つまり一定時間より長く発言するということが会話を成立させる条件の三つ目である。これら三つの条件を満たした場合に次のステップに移行する。

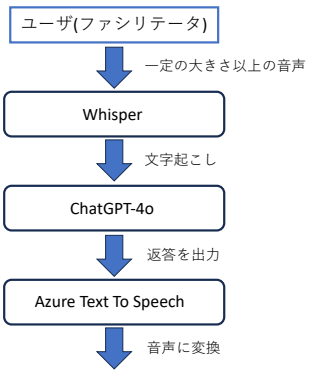


Fig. 5　ユーザの発言～キャラクターの返答までの流れ

2.3.1　音声認識モデル Whisper

音声認識モデル Whisper(OpenAI) を利用し、ユーザが発言した音声を文字起こしして、テキストデータに変換する。

2.3.2　大規模言語モデル GPT-4o

前のステップで得られたテキストデータを大規模言語モデル GPT-4o に読み込ませる。この時、GTP-4o には、対象キャラクターなどに関する設定が記述されたプロンプト (Fig.6、Fig.7)、対象とする絵画、および実際に複数人に聞き取りを行って得られた対象の絵画に関する印象などをまとめたデータ (Fig.8) を読み込ませている。なお、前述のように、プロンプトでは、キャラクターごとに異なる性格などの特徴を設定している。このプロンプトを読み込ませることで、そのキャラクターに合わせた返答を得ることができる。さらに、Fig.8 で示した印象データをポジティブな内容のものとネガティブな内容のものに分け、ポジティブな性格のキャラクターにはポジティブな印象データを反映させ、ネガティブな性格のキャラクターにはネガティブな印象データを反映させることで、よりそのキャラクターの性格に合った返答を得られるようにしている。各キャラクターにつき、一つの絵画に対して読み込ませる印象データは 100 個前後である。また、対話型鑑賞では 1 対 1 のやり取りではなく、複数人での会話であるため、誰がどのようなことを言ったのかを各キャラクターが把握する必要がある。そこで、今までに誰がどのような発言をしたのかを記録し、記録された会話の流れも GPT-4o に読み込ませるようにしている。

2.3.3　音声合成システム Azure Text To Speech

GPT-4o により出力されたテキストデータを音声機能を提供する Azure Text To Speech(Microsoft) を利用することによって、音声に変換する。この時合成される音声は年齢、性別、国籍で異なるようにしている。このような流れで音声が出力されることによって、ユーザとキャラクターとの会話が実現される。

2.3.4　キャラクターの口の動き

キャラクターが発言していることを視覚的にわかるようにするために、Oculus が提供するリップシンクライブラリである OVRLipSync を用いることによって、音声に対応して口を動かすようにしている。

```
### SETTING OF A CHATBOT (YOU) ###
Please refer to the following for character settings.
Character Settings
  Name: Satomi Hamaguchi
  Number(番号): 1
  Gender: Female
  Background: Artist
  Age: 56 years old
  Personality: Cheerful, active, kind and sensitive
  Favorite food: Orange
  Hobbies: Taking a picture
  Hometown: Hokkaido
  Habit of saying: ファンファン

Please refer to the following for setting up a conversation
situation.
Conversation situation
  Place: Dimly lit museum
  Situation: Appreciating a painting in the hallway of the
museum
  Content: Communicating impressions of the painting
  People: You(浜口サトミ), User, Character number 2(名前: 真
田マサヒロ), Character number 3(名前: 松林カイ), Character
number 4(名前: 大空ツバサ), Character number 5(名前: 安森
ウメ)
```

Fig. 6　キャラクター設定のプロンプトの一部: 例①

The following are examples of conversations in this situation.
Examples(First example, second example and third example):
First example:
Userが次のように発言しました。
「この絵についてどう思いますか。」
松林さんが次のように発言しました。
「とても綺麗やな。」
Userが次のように発言しました。
「真田さんは松林さんの意見に対してどう思いますか。」
真田さんが次のように発言しました。
「松林さんの意見とは異なり、この絵はあまり良くないと思いました。」

「あなたも真田さんの意見に賛成ですか。」(Userの発言)
「いいえ。この絵は印象的で素敵だと思います。」
(Your statement)

Second example:
Userが次のように発言しました。
「この絵についてどう思いますか。」
松林さんが次のように発言しました。
「この絵、めっちゃ好きやねんな。」

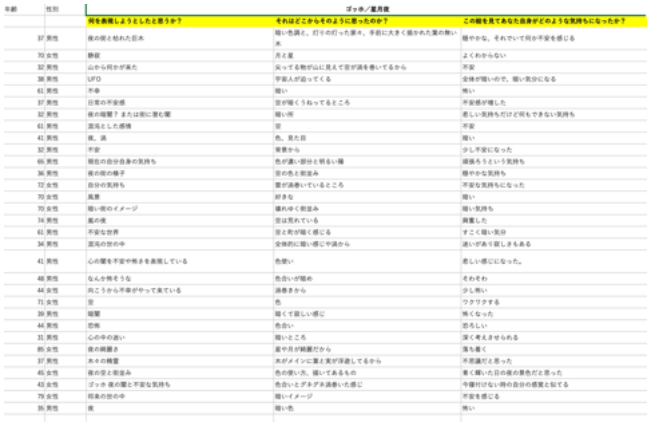
「松林さんはこのように言ってますが、浜口さんはこの絵についてどう思いますか。」(Userの発言)
「私も松林さんと同じで、この絵が好きです。」
(Your statement)

Third example:
「この絵についてどう思いますか。」(Userの発言)
「私はこの絵に良い印象を持ちます。」(Your statement)
Please voice your own thoughts after considering the flow of conversations as in these examples in this situation.
In addition, with these examples as a reference, please respond flexibly to the situation.
These example are for reference only and you do not have to answer exactly like them.

If a user asks you a question about someone other than Satomi Hamaguchi, Masahiro Sanada, Kai Matsubayashi, Tsubasa Ozora, or Ume Yasumori, please ask who that person is.

When a user says, "Say something," please speak as if you were speaking autonomously, following the flow of the conversation and expressing whose opinion you would like to hear or your opinion.
The user is the facilitator, so basically don't ask the user about their impressions of the painting. However, you may ask questions to obtain information about the paintings.

Fig. 7　キャラクター設定のプロンプトの一部: 例②



年齢	性別	ゴッホ／星月夜		
		何を表現しようとしたと思うか？	それはどこからそのように思ったのか？	この絵を見てあなた自身がどのような気持ちになったか？
	37 男性	夜の街と始れた巨木	暗い色調と、灯りの灯った家々、手前に大きく描かれた黒の暗い木	穏やかな、それでいて何か不安を感じる
	70 女性	静寂	月と星	よくわからない
	32 男性	山から何かが来た	光ってる物がはに見えて空が渦を巻いてるから	不安
	36 男性	UFO	宇宙人が迫ってくる	全体が暗いので、暗い気分になる
	61 男性	不安	暗い	怖い
	37 男性	日常の不安感	空が明くうねってるところ	不安感が増した
	32 男性	夜の暗闇？または宇宙に潜む闇	暗い所	恐しい気持ちだけど何もできない気持ち
	61 男性	混沌とした感情	空	不安
	41 男性	夜、渦	色、見た目	暗い
	32 男性	不安	背景から	少し不安になった
	66 男性	現在の自分自身の気持ち	色が濃い部分と明るい箇	頑張ろうという気持ち
	36 男性	夜の街の様子	空の色と街並み	穏やかな気持ち
	72 女性	自分の気持ち	雲が渦巻いているところ	不安な気持ちになった
	70 女性	風景	好きな	暗い
	70 女性	暗い夜のイメージ	優れゆく街並み	暗い気持ち
	74 男性	星の夜	空は荒れている	興奮した
	61 男性	不安な世界	空と月が暗く感じる	すこく暗い気分
	34 男性	混沌の世の中	全体的に暗い感じや渦から	違いが多く感じしまれる
	41 男性	心の闇を不安や恐さを表現している	色使い	恐しい感じになった。
	48 男性	なんか時そうな	色合いが暗め	それぞれ
	44 女性	向こうから不幸がやって来ている	渦巻きから	少し怖い
	71 女性	空	色	ワクワクする
	39 男性	暗雲	暗くて寂しい感じ	怖くなった
	44 男性	恐怖	色合い	恐ろしい
	31 男性	心の中の迷い	暗いところ	深く考えさせられる
	66 女性	夜の綺麗さ	星や月が綺麗だから	落ち着く
	37 男性	木々の微妙	木がメインに置き実が浮遊してるから	不思議だと思った
	45 女性	夜の空と街並み	色の使い方、描いてあるもの	青く輝いた日の夜の景色だと思った
	43 女性	ゴッホ 夜の闇と不安な気持ち	色合いとグネグネ渦巻いた感じ	今着けてない時の自分の感覚と似てる
	70 女性	将来の世の中	暗いイメージ	不安を感じる
	35 男性	夜	暗い色	怖い

Fig. 8　印象データ

2.4　その他の機能

画面内には、Fig.9 に示す 8 つのボタンを提示し、これらのボタンを押すことで選択または機能を追加できる。

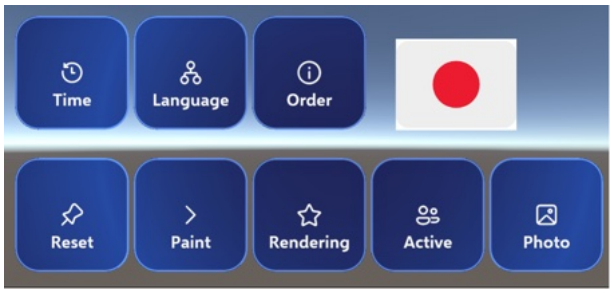


Fig. 9　ボタン

2.4.1　Language

Language ボタンにより、キャラクターが話す言語を切り替えることができる。切り替え可能な言語は日本語、英語、フランス語、ドイツ語、ノルウェー語、簡体字中国語、繁体字中国語、韓国語の 8 言語である。

2.4.2　Reset

Reset ボタンにより、キャラクターや絵画の位置を変更できる。

2.4.3　Rendering

Rendering ボタンにより、会話の内容がポジティブかネガティブか、キャラクターが反応しているかなど、様々な情報を表示できる。

2.4.4　Paint

Paint ボタンにより、6 種類の作品に切り替えることができる。6 種類の作品のうちの 3 作品は絵画、残り 3 作品は写真である。絵画 3 作品は、フィンセント・ファン・ゴッホ (Vincent van Gogh, 1853-1890) の《星月夜》、カスパー・ダーヴィト・フリードリヒ (Caspar David Friedrich, 1774-1840) の《孤独な木》、ジョン・エヴァレット・ミレイ (John Everett Millais, 1829-1896) の《狼の巣穴》である。

2.4.5　Time

このシステムは無音状態が一定時間以上続いた場合、ユーザが話しかけなくても、5 人のキャラクターのうちの誰かがランダムで発言するようになっている。Time ボタンを押すことでその一定時間を 1 分、2 分、3 分、5 分、10 分の中から選択して設定することができる。

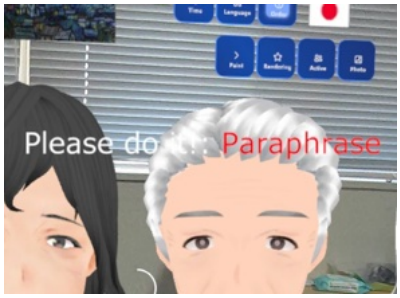
2.4.6　Active

実際の対話型鑑賞には様々な参加者がおり、例えばよく発言する人もいれば、あまり発言しない人もいる。この状況を再現するために、Active ボタンを導入している。この Active ボタンを押すことでキャラクターが自発的に発言するまでの無音状態の継続時間が短くなり、このボタンを押した状態では特定のキャラクターが積極的に発言し、それ以外のキャラクターは自発的に発言しなくなる。このため、ユーザはできるだけ発言が均等になるように会話を誘導する必要がある。

2.4.7　Order

対話型鑑賞におけるテクニックとして、参加者の発言内容を言い換えるパラフレーズ、会話の内容をまとめるサマライズ、絵画の注目する場所を指定して話題を変えるフォーカシングなどがあり、これらのテクニックをうまく使いながら対話型鑑賞を進行していくことがファシリテータに求められる。そこで Order ボタンを導入し、このボタンを押すことでユーザが次行うべき指示が表示されるようにした (Fig.10)。表示される指示は段階的に変化していく。指示の流れを Fig.11 に示す。このように指示を表示することによって、

ファシリテータがテクニックをうまく使いながら対話型鑑賞を進行するための訓練につながると期待される。



(a) パラフレーズ

(b) サマライズ

(c) フォーカシング

Fig. 10　表示される指示 (3 種類)

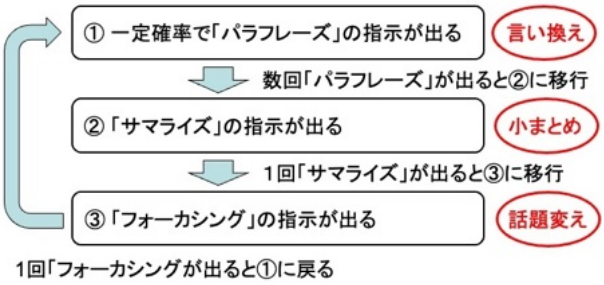


Fig. 11　表示される指示の流れ

2.4.8　Photo

この機能は前述の 6 種類の絵画・写真ではなく、現実空間に存在する絵画や風景などの写真を撮って取り込むことで、その写真に対して対話型鑑賞を行うものである。現段階ではスマートフォン等で撮った写真を専用の GitHub ページにアップロードし、アップロードした写真の URL を用いてシステムに取り込んでいる。今後は、ヘッドセット (Quest3) で直接写真を撮って取り込むなど、より使いやすい手法を検討している。

3.　今後の展開

今後このシステムの実践、評価を行い、その上で改良を行なっていく。現在は GPT-4o を利用しているが、Anthropic によって開発された AI チャットモデルである Claude と GPT-4o を比較し、どちらがこのシステムに適しているか検討したいと考えている。

4.　結論

対話型鑑賞においてファシリテータは重要な役割を担うが、ファシリテータの養成、訓練を行う機会が少ないという問題がある。そこで本稿では、現在我々が開発している対話型鑑賞のファシリテータ養成のための複合現実 AI システムを紹介した。本システムは、場所を問わず、一人でも対話型鑑賞におけるファシリテータの訓練を行うことができ、ファシリテータ養成、訓練の機会を増やすことができる。今後は、システムの評価実験を通して得られる評価、意見をもとにシステムを改良し、実際の対話型鑑賞にできるだけ近づける予定である。

謝辞

本研究は，JST 共創の場形成支援プログラム「共生社会をつくるアートコミュニケーション共創拠点」(JPMJPF2105) の支援を受けたものです。また印象データをご提供いただきました川畑秀明慶應義塾大学教授、開発にご協力いただきましたファシリテータの近藤乃梨子氏、春日美由紀氏に感謝いたします。

References

[1]　Philip Yenawine, 京都造形芸術大学アートコミュニケーション研究センター：“学力をのばす美術鑑賞 ヴィジュアル・シンキング・ストラテジーズ：どこからそう思う?”, 淡交社, -, 2015

[2]　稲庭彩和子, 伊藤達矢, 河野佑美, 鈴木智香子, 渡邉祐子：“こどもと大人のためのミュージアム思考”, 左右社, -, 2022