

# マルチモーダル情報に基づく就職面接練習システムの開発

竹内 直<sup>†</sup> 神田 智子<sup>‡</sup>

<sup>†</sup> 大阪工業大学大学院情報科学研究科 〒573-0196 大阪府枚方市北山 1-79-1

<sup>‡</sup> 大阪工業大学情報科学部 〒573-0196 大阪府枚方市北山 1-79-1

E-mail: <sup>†</sup> m1m21a29@st.oit.ac.jp <sup>‡</sup> tomoko.koda@oit.ac.jp

**あらまし** 本研究では、被面接者の視線・表情・姿勢の非言語情報を視線追跡装置、カメラ画像を用いて認識し、被面接者の模範的な非言語行動モデルとの比較を行うことで、振り返りを行いながら改善点を指摘する面接練習システムの開発を行った。開発の方針は、普及型 HW とオープンソースを用いた安価かつ容易なシステム構築が可能であること、および CG エージェントによる振り返りが可能であることである。これは、従来の面接練習システムは高価な HW, SW が必要であるものが多く、一般ユーザ向けになっていないと考えた為である。また CG エージェントによる振り返りを用いたシステムは、我々が知る限り少ない為である。本システムの初期評価結果として、現段階では、精度の向上や CG エージェントによるフィードバックの改良などの改善点が挙げられた。

**キーワード** 視線認識, 姿勢認識, 表情認識, 就職面接練習, 非言語情報, CG エージェント, マルチモーダル

## Development of a Job Interview Training System with Multi-modal Behavior Analysis

Nao Takeuchi<sup>†</sup> Tomoko Koda<sup>‡</sup>

<sup>†</sup> Graduate School of Information Science and Technology, Osaka Institute of Technology

1-79-1 Kitayama, Hirakata-shi, Osaka, 573-0196 Japan

<sup>‡</sup> Faculty of Information Science and Technology, Osaka Institute of Technology

1-79-1 Kitayama, Hirakata-shi, Osaka, 573-0196 Japan

E-mail: <sup>†</sup> m1m21a29@st.oit.ac.jp <sup>‡</sup> tomoko.koda@oit.ac.jp

**Abstract** This paper introduces a job interview training system that recognizes the nonverbal behaviors of the interviewee, namely, gaze, facial expression, and posture. The proposed system uses a Tobii eye tracker for gaze recognition, and camera images for facial expression and posture recognition. The system compares the recognition results with the models of the interviewee's exemplary nonverbal behaviors and point out the non-verbal behaviors that need improvement while playing back the interview recording. Our development policy of the system is to construct an inexpensive and easy system using commercially available HWs and open source, and utilization of a CG agent that gives feedback to the interviewee. Most of the current interview training systems require high-end HW and SW and are not designed for general users, and there are few systems using CG agents to give feedback. The results of the initial evaluation of the system indicate improvements are needed in terms of recognition accuracy of non-verbal behaviors and quality of the interaction with the CG agent.

**Keywords** gaze recognition, posture recognition, facial expression recognition, job interview training, nonverbal behaviors, CG agent, multi-modal interaction

### 1. はじめに

コロナ禍の影響が就職活動に及んでおり、人と人との接触をできるだけ控えるという観点から、これまで就職支援の一環として行われていた対面での模擬面接が避けられようになった。このことから、一人で就職面接練習が可能なシステムの必要性が増していると考ええる。

対話中に視覚的な非言語情報によって伝えられるメッセージは全体の 55%とされており[1], Washburn らは面接における非言語的コミュニケーションが言語的コミュニケーションよりも被面接者の評価に影響すると指摘している[2]. また、近年マルチモーダル情報を用いた社会的信号処理技術が対話分析に用いられており[3], AI を用いた面接採用システム[4]や面接練習システム[5]にも応用されている。しかし、視線・表情を

認識するシステムはあるが姿勢を認識対象にしておらず、フィードバックにCGエージェントを用いたシステムは我々が知る限り存在しない。また、[4, 5]は高価なシステムである為、多くの人が使える普及に適したシステムではない。これらのことから、本研究では視線・表情に姿勢を加えた3つの非言語情報に特化し、安価な普及型HWやオープンソース、そしてCGエージェントを使用した面接練習システムの開発を目的とする。

## 2. 就職面接練習システム

### 2.1. システム概要

開発は、Unity 2019.4.11f1 をベースに、FaceAPI[6], OpenPose[7], TobiiEyeTracker4C[8], Web カメラを組み合わせて、システム構築を行った。本システムは、属性入力フェーズ、模擬面接フェーズ、フィードバックフェーズの三部構成となっている。属性入力フェーズ(図1)では、使用回数・性別入力を行う。模擬面接フェーズ(図2)では、実際に模擬面接を約1~2分間受け、被面接者の視線・表情・姿勢の情報を取得し、模擬面接の様子を斜めから撮影する。フィードバックフェーズ(図3)では、撮影された動画を再生し、取得した情報に従ってCGエージェントが随時動画を一時停止しながら指摘内容に対するフィードバックを行う。



図1 属性入力フェーズ



図2 模擬面接フェーズ

(左:実験の様子, 右:右ディスプレイに表示される画面)



図3 フィードバックフェーズ

(左:実験の様子, 右:右ディスプレイに表示される画面  
右上:待機状態, 右下:フィードバックを行う状態)

### 2.2. 面接官

本システムの模擬面接における面接官は、実際の人間が面接を行う動画を使用した。面接官の動画は、大阪工業大学就職課にご協力いただき撮影を行った。動画は、質問内容2種類×面接パターン5種類の計10種類である。質問内容は「自己PR」または「大学時代に力を入れたこと」について1分間話してもらいものである。面接パターンは、面接官に演技わけてもらい「通常バージョン」、「相槌多めバージョン」、「厳しめバージョン」、「メモをとるバージョン」、「履歴書よく見るバージョン」を用意した。これらから、ランダムで再生が行われる。図4に面接官の動画の一例を示す。

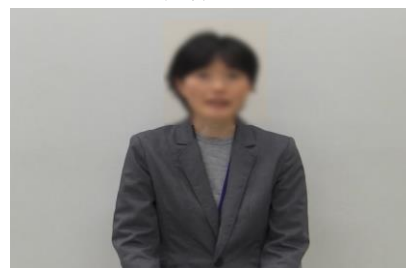


図4 面接官 (大阪工業大学就職部)

### 2.3. フィードバックCGエージェント

本システムのフィードバックを行うために、CGエージェントを使用した。使用したCGエージェント(図5)は、Unity Asset "NATASHA" [9, 10, 11]である。






図5 使用CGエージェント"NATASHA"

## 2.4. フィードバック方法

模擬面接中撮影した動画を再生しながら、検出したときの秒数に従って、随時動画を一時停止し、「この時、姿勢が崩れていませんでしたか。注意しましょう」などとCGエージェントがアドバイスをを行う。この方法により、模擬面接中のどの部分で自身のどこが良かったか、悪かったかを明確に理解できる。実際にどのようなフィードバックを行っていたかの一例を、表1に示す。

表1 フィードバック例

		フィードバック内容	
検出情報	【視線】 面接官の顔から 5秒以上注視を外した	「このとき一定期間面接官から視線がはずれています、注意しましょう」 「集中力がとぎれていませんか？」	
	【表情】 真顔度	「この時表情が硬いかもれません」	
	【姿勢】 足が開きすぎ	「この時足が開きすぎじゃないですか？気をつけましょう」	

## 2.5. 検出可能な非言語情報

非言語情報に関して、視線は1秒間隔、表情・姿勢は3秒間隔で取得する。取得情報から凝視割合、面接官の顔から5秒以上注視を外したタイミング、右上・左上に注視点が移動した回数、笑顔度、真顔度、表情（怒り、軽蔑、嫌悪、恐怖、悲しみ、驚きの6種類）、姿勢（前傾・後傾）、足の開き、足が徐々に開いたタイミング、首のぶれ、肘の張り出しを検出する。

## 2.6. 検出方法

姿勢検出に関しては大阪工業大学就職部の指導のもと、OpenPoseを用いて正しい姿勢モデル（図6）を男女別で作成し、そのモデルと被面接者から取得した姿勢の対比で判断する。また、視線検出に関しては、面接官の顔を対象とした当たり判定、表情検出に関してはFaceAPIによる笑顔度の閾値判定、および表情認識結果より判定する。

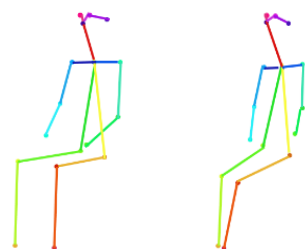


図6 基準姿勢モデル（左:男性, 右:女性）

## 2.7. フィードバックアルゴリズム

検出した情報は、秒数ごとに視線・表情・姿勢配列それぞれに格納し、一つの配列にまとめ、CGエージェントがフィードバックする際に参照する。一つの配列にまとめる際、同じ秒数で複数の検出が重複してしまうことや、同じ種類に偏った指摘を避けるため、変化の多い①視線、②表情、③姿勢の順で重みづけを加えた優先順位付けを行った。以上のように、取得データがどのような流れでフィードバックされるかを図7に示す。

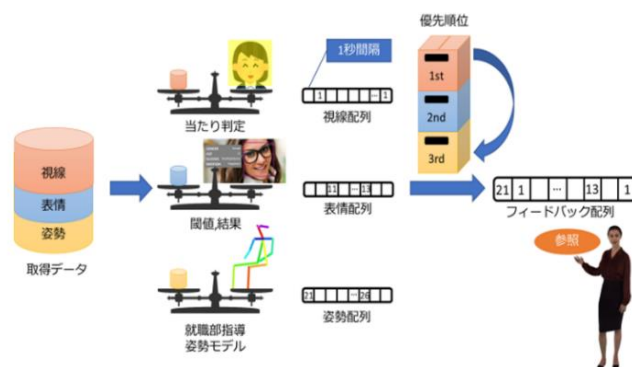


図7 フィードバックアルゴリズム

## 3. 初期評価実験

本システムの初期評価を行うために、21~22歳の大学生5名(男性3名, 女性2名)に実際に本システムを使用してもらい、インタビュー形式でヒアリングを行った。なお、被験者実験にあたり、大阪工業大学ライフサイエンス委員会よりヒト対象実験実施の承認済みである。

実験環境は、実際の面接の場面を想定しているもので、ディスプレイとユーザの距離が近すぎず、EyeTrackerが動作するかを考慮し、被験者の位置を設定した。また、OpenPoseを使用した座位姿勢における姿勢推定において、正面からでは前傾・後傾姿勢、足の角度が取得できない為、被験者に対して斜め45度にWebカメラを配置し、姿勢推定を行った。図8に実験環境を示す。

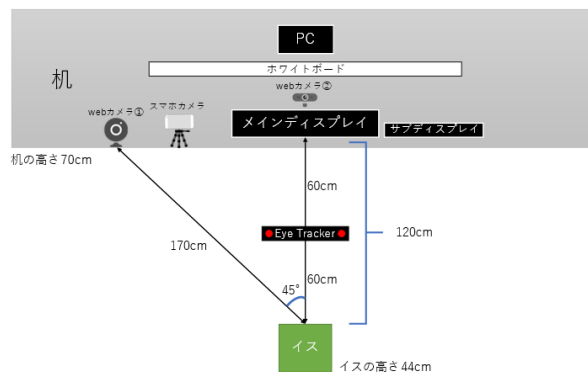


図8 実験環境

### 3.1. CG エージェントの印象に関する結果と考察

ヒアリング結果として、「面接専門のエージェント・秘書のような外見」、「しっかりちゃんと指摘してくれるイメージ」、「違和感なく素直に聞けそう」というような良い点が挙げられたのに対し、「実際の人間との練習の方が詳しく見てくれる」、「ためになるのは人間との練習」というような否定的なコメントも挙げられた。このような否定的なコメントをした被験者に対し、人間と CG エージェントでフィードバックする内容が同じであるなら、どちらがよいかと聞いた結果、「フィードバック内容が同じであればどちらでもいい」、「人間だと緊張する、恥ずかしいから CG エージェントの方が練習しやすい」という返答があった。これらのことから、CG エージェントの方が練習に取り組むのには気軽である一方、人間との対面面接の方が真剣に練習できる可能性も示している。従って、今後の展望として、実際に人間によるフィードバックを行う条件と、CG エージェントによるフィードバックを行う条件を用意し、傾聴効果を対象とした対比実験を行う。

### 3.2. CG エージェントによるフィードバックに関する結果と考察

ヒアリング結果として、「動画が止まって（CG エージェントが）指摘やアドバイスを都度言うってくれるので、どの場面でどこが良かった/悪かったのが明確に分かる」という良い点が挙げられたのに対し、「CG エージェントの場合、元々用意されたものしか再生されていないので、個人に合わせて助言している感じがあまりしない」、「気軽さは CG エージェントだが、人間の方が重く突き刺さる、受け止めやすい」というような悪い点が挙げられた。

この原因として、CG エージェントとのインタラクションが一方的かつ画一的であり、対人間に比べ劣ることが考えられる。その対策として、同じ指摘内容でも複数音声パターンを用意、また CG エージェントとのインタラクションの機会を増やす必要があると考える。例えば、模擬面接の感想や自己評価を話すフェーズを作成し、CG エージェントがその内容に対して、褒める・勇気づける・励ますといった応答をするなどの実装が考えられる。

### 3.3. 指摘精度に関する結果と考察

ヒアリング結果として、「(姿勢に対する) 指摘の誤りがあったと思う」、「指摘されると思ったが指摘されなかった」というような指摘精度に関するコメントが挙げられた。また、コメントより総合的に判断すると 50%程度の指摘精度であったと考えられる。

この原因として、まず OpenPose の検出精度が低か

ったのではないかと考える。本システムでは、姿勢(前傾・後継)の角度を取得するために、斜め 45° に Web カメラを配置した。OpenPose の性質上、斜めからだけでは椅子に座った際の腰掛け具合や、座位姿勢が直角であるかどうかなどが検出できず精度が下がる。その対策として、正面や横に Web カメラの増設もしくは Kinect での代用で、精度向上が見込めるのではないかと考える。

次に、姿勢に関する検出誤差があったと考える。姿勢検出に関して、被面接者の姿勢と基準モデルを対比し、逸脱した箇所を閾値判定により検出する方法を用いた。その基準モデルが男女 1 パターンずつで個人差を配慮していなかった為に、姿勢の逸脱箇所の検出の誤差が見られたのではないかと考える。その対策として、基準モデルを数パターン用意し、ユーザの身長や体格にできるだけ近いものを使用することで精度向上が見込めるのではないかと考える。

最後に、姿勢・表情に関する検出漏れがあったと考える。表情・姿勢検出に関して、FaceAPI、OpenPose の FPS を上げると処理が重くなることから、3 秒間隔で取得を行った。その為、3 秒の間で基本モデルから逸脱している箇所の検出ができていない部分があったことが考えられる。

### 3.4. フィードバック内容に関する結果と考察

ヒアリング結果として、「フィードバック内容(指摘の種類、回数)が少ない」というコメントが挙げられ、実際に一人あたり 2~6 回程度であった。この原因は、検出精度とも関わりはあるが、他にフィードバック配列を作る際の重みづけを加えた優先順位アルゴリズムにも原因があると考えられる。本システムでは、面接時重要視され、時間経過で変化の多い①視線②表情③姿勢の順で優先順位付けを行い、検出された種類の順位を下げる形式に、重みづけ(順位がしばらく更新されなかった時に順位付けを再構築する)を加えた。実際検出はしているが、フィードバック配列に反映されなかったというケースも考えられるため、重みづけを加えたアルゴリズムの改良が必要である。

また、検出する箇所の追加も必要であると考えられる。「手、指の動き、体の揺れ、声の大きさ、接続詞、質問に対する回答の内容、話し方の指摘があればいい」というコメントも挙げられたように、非言語情報に加え、言語情報の指摘も必要であることが指摘されている。

### 3.5. ユーザビリティに関する結果

ヒアリング結果として、「自身の面接動画を見る機会がないので、フィードバックの時に客観的に動画を見ながらアドバイスしてくれるのは良い」、「また練習ができるならしてみたい」、「面接官がリアルだった。本番のようで緊張感があり、見られている感があった」、「一人でもできる（積極的に就職課に行って練習できない消極的な人でも使えそう）」というような好評価のコメントが挙げられた。

### 4. おわりに

本研究では、被面接者の視線・表情・姿勢の非言語情報を視線追跡装置、カメラ画像を用いて認識し、被面接者の模範的な非言語行動モデルとの比較を行うことで、振り返りを行いながら改善点を指摘する面接練習システムの開発を行った。開発の方針は、普及型 HW とオープンソースを用いた安価かつ容易なシステム構築が可能であること、および CG エージェントによる振り返りが可能であることである。

初期評価実験を行い、現段階では、精度が 50%程度で十分でないこと、CG エージェントとのインタラクションが一方向的で対人間に比べ物足りないこと、フィードバックする回数が一人につき 2~6 回程度で指摘回数や内容が少ないことなどの改善点が挙げられた。

今後の展望として、初期評価実験で挙げられた改善点を解決し、視線・表情・姿勢以外の非言語情報や言語情報の取得、指摘され発覚した修正点をピンポイントに練習できる機能や被面接者が自身の面接動画をシステム使用後振り返ることのできる機能の追加、フィードバック内容の優先度設定を行うアルゴリズムの改良、2 つ以上の非言語情報を組み合わせた検出方法の追加が必要であると考えます。また、CG エージェントによるフィードバックと人によるフィードバックの傾聴効果を対比し、検証する必要がある。

### 謝辞

本研究の一部は、科研費「基盤(C)20K11926」の交付を受けて実施した。

### 文 献

- [1] Mehrabian A, Silent messages: Implicit communication of emotions and attitudes, Wadworth Publishing Co., California, 1981
- [2] Washburn P.V., Hakel M.D. , Visual cues and verbal content as influences on impressions after simulated employment interviews. Journal of Applied Psychology, pp.58,137-140, 1973
- [3] 岡田将吾, 松儀良広, 中野有紀子, 林佑樹, 黄宏軒, 高瀬裕, 新田克己, マルチモーダル情報に基づくグループ会話におけるコミュニケーション能力の推定, 人工知能学会論文誌, Vol.31, No.6

A130-E, 2016

- [4] 株式会社マイダスアイティジャパン, <https://www.inair.co.jp/>
- [5] 合田七穂, 石原圭太郎, 小尻智子, 非言語情報の特徴分析に基づいた就職面接練習支援システム, 信学技報, Vol.116, No.517, ET2016-98, 25-30, 2017
- [6] Microsoft Azure, <https://azure.microsoft.com/ja-jp/services/cognitive-services/face/>
- [7] tf-pose-estimation, <https://github.com/ildoonet/tf-pose-estimation>
- [8] トビー・テクノロジー株式会社, <https://www.tobipro.com/ja/>
- [9] Metastage: "NATASHA" - Seated Listening, <https://assetstore.unity.com/packages/3d/characters/humanoids/humans/natasha-seated-listening-153787>
- [10] Metastage: "NATASHA" - Presentation, <https://assetstore.unity.com/packages/3d/characters/humanoids/humans/natasha-presentation-153909>
- [11] Metastage: "NATASHA" - Serious Talking, <https://assetstore.unity.com/packages/3d/characters/humanoids/humans/natasha-serious-talking-153785>